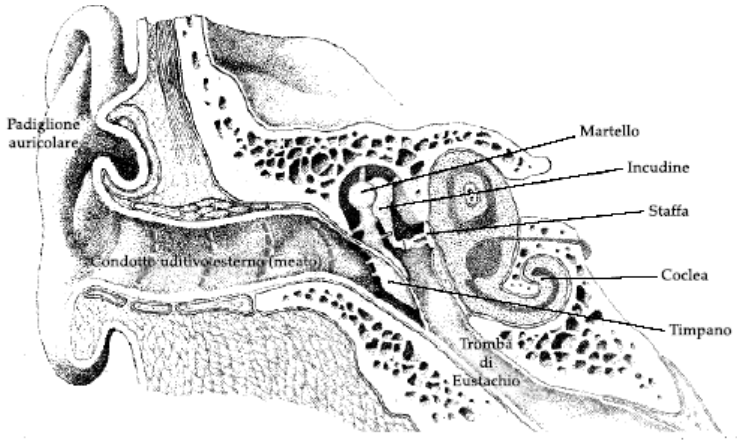
Sistemi Multimediali
I formati audio

Ombretta Gaggi
Università di Padova

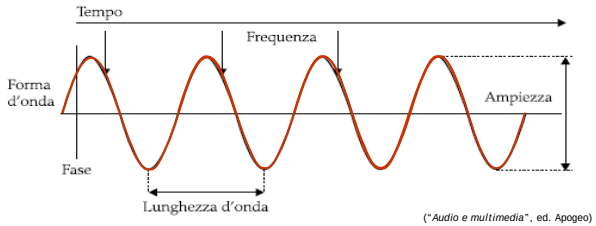
Struttura interna dell'orecchio umano



(Fonte: Vincenzo Lombardo e Andrea Valle, "*Audio e multimedia*", ed. Apogeo)

Audio: fondamenti

Il suono udibile è un'onda di pressione continua nell'intervallo di frequenza da 16 Hz a 22 kHz



("*Audio e multimedia*", ed. Apogeo)

ampiezza (dB) → volume, intensità del suono

frequenza (Hz) → altezza del suono

forma d'onda → timbro, permette di distinguere un suono da un'altro

Livelli di intensità

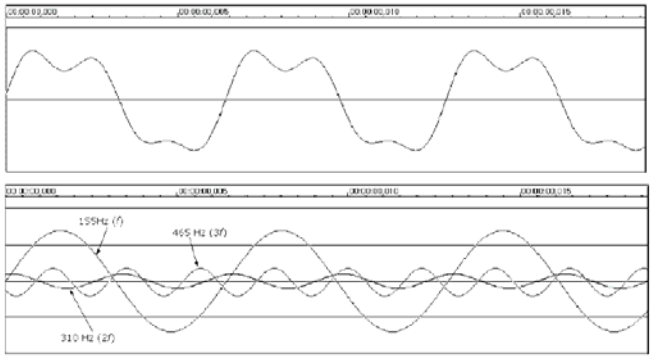
Tabella1.1 Livelli di intensità

Suono	SPL (dB)	Reazione
Massimo rumore prodotto in laboratorio	210	Suono insopportabile
Lancio di un missile (a 50 m)	200	
Rottura del timpano	160	
Jet al decollo (a 50 m)	130	
Suono al limite del dolore	120	Dolore fisico
Complesso rock in locale chiuso	110	
Schianto di fulmine	110	
Urlo (a 1,5 m)	100	
Martello pneumatico (a 3 m)	90	
Traffico cittadino diurno	70-80	
Ufficio o ristorante (affollati)	60-65	Suoni utili
Conversazione (a 1 m)	50	
Teatro o chiesa (vuoti)	25-30	
Bisbiglio (a 1 m)	15	
Fruscio di foglie	10	
Zanzara vicino all'orecchio	10	
Soglia dell'udito (a 1000 Hz)	0	Non udibile

("*Audio e multimedia*", ed. Apogeo)

Audio: analisi di Fourier

"Un segnale periodico qualsiasi è dato dalla sovrapposizione di onde sinusoidali semplici, ciascuna con la sua ampiezza e fase, le cui frequenze sono armoniche della frequenza fondamentale del segnale"



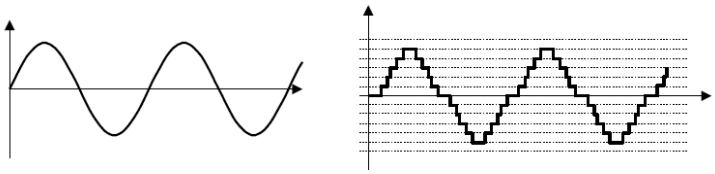
("*Audio e multimedia*", ed. Apogeo)

Rappresentazione digitale del suono

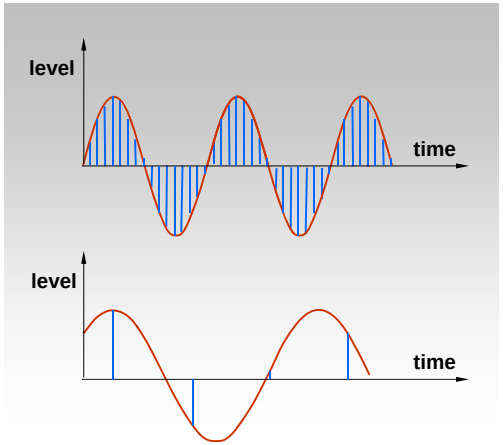
Per essere elaborato il suono deve essere digitalizzato (conversione A/D)

- *campionamento:* divisione lungo l'asse del tempo (si misura in Hz)
- *quantizzazione:* rappresentazione discreta del livello del segnale (si misura in in bit di precisione)
- es. CD audio è digitalizzato a 44.1 kHz, 16 bit

Qual è l'effetto della digitalizzazione sulla qualità del segnale?



Conversione analogica/digitale del segnale audio



Un segnale il cui spettro di frequenza è limitato superiormente può essere completamente ricostruito da un insieme di campioni se la frequenza di campionamento è almeno doppia della più alta frequenza presente nel segnale (teorema di Nyquist)

Rapporto segnale/rumore

Nei sistemi analogici il segnale è alterato dal *rumore*, una fluttuazione casuale del livello del segnale dovuto a fenomeni elettronici

- il *rapporto segnale/rumore* (*SNR*, *signal to noise ratio*) è una misura della qualità del segnale

$$SNR = 10\log\frac{V_{signal}^2}{V_{noise}^2} = 20\log\frac{V_{signal}}{V_{noise}}$$

Nei sistemi digitali il rumore compare quale differenza tra il livello del segnale reale e il livello del segnale quantizzato

$$SQNR = 20\log\frac{V_{signal}}{V_{quant-noise}} = 20\log\frac{2^{N-1}}{1/2} = 6.02N\ (dB)$$

Qualità dell'audio vs. dimensione

Qualità	*Intervallo di frequenza Hz*	*Campiona-mento kHz*	*Bit per campione*	*Mono stereo*	*Velocità dei dati kbit/s*
Telefonia	200-4000	8.0	8	mono	8
Radio AM	100-6500	11.025	8	mono	11
Radio FM	20-12000	22.050	16	stereo	705.6
CD audio	20-20000	44.1	16	stereo	1411.2
DAT	20-20000	48.0	16	stereo	1536
DVD audio	20-20000	192.0	24	stereo	9216

Problemi nella codifica dell'audio

La codifica e decodifica digitale di segnali audio presenta maggiori problemi rispetto alle immagini (e al video)

- L'audio ha una struttura temporale che non può essere modificata (frequenza)
- l'informazione audio è variabile nel tempo (non esiste il "fermo audio")
- la qualità di riproduzione richiesta di solito supera di molto la soglia di semplice comprensibilità

Audio: dimensione & tempo di trasferimento

L'audio non compresso di buona qualità supera la capacità di trasmissione delle reti convenzionali

- Radio FM > 640 Kbit/sec, CD audio > 1.2 Mbit/sec

I dati sono di grandi dimensioni

- il segnale codificato occupa molto spazio
- La lunghezza è in linea di principio non limitata (audio dal vivo)

Non è possibile trasferire tutto il file prima di iniziarne la riproduzione

- *streaming*: riproduzione durante la ricezione

Formati audio (1)

WAV, Waveform Audio File

- sviluppato congiuntamente da Microsoft e IBM
- standard *de-facto* per la codifica del suono su PC
- non compresso

AIFF, Audio Interchange File Format

- sviluppato da Apple Computer
- formato audio standard del Macintosh
- non compresso (esiste una versione compressa)

µ-LAW

- formato audio standard Unix
- standard telefonico in USA (8KHz, 8 bit)

A-LAW

- versione europea di µ-LAW

Formati audio (2)

Audio MPEG-1

- codifica le tracce audio nei video MPEG-1
- è un formato compresso per codifica a qualità variabile
- usa un algoritmo di compressione a più stadi basato su principi di psicoacustica
- sono definiti tre livelli di codifica per tre diversi bit-rate
- è uno standard cross-platform
- le applicazioni nel mercato *consumer* sono molte e commercialmente significative

La compressione audio (1)

La compressione senza perdita non fornisce buone prestazioni

- i dati audio sono molto variabili
- le configurazioni ricorrenti sono rare

E' necessario utilizzare metodi di compressione *con perdita*

- l'informazione audio è ridondante
- la qualità di compressione può essere controllata
- l'orecchio umano non ha un comportamento lineare

La compressione audio (2)

Compressione del silenzio

- il silenzio è un insieme consecutivo di campioni sotto una certa soglia
- è simile alla compressione RLE (*run-length encoding*)

Adaptive Differential Pulse Code Modulation

- codifica la differenza tra campioni consecutivi
- la differenza è quantizzata, quindi si ha perdita di informazione

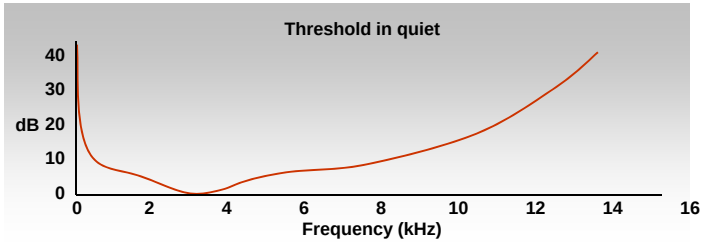
Linear Predictive Coding

- adatta il segnale ad un modello del parlato umano
- trasmette i parametri del modello e le differenze del segnale reale rispetto al modello

Elementi di psicoacustica (1)

La sensibilità dell'orecchio umano è variabile lungo lo spettro audio

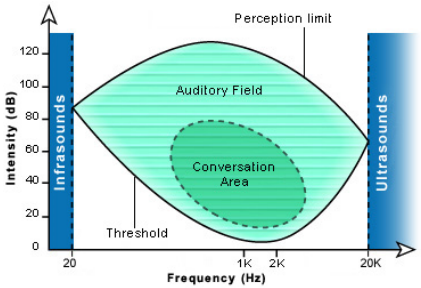
- la sensibilità massima si ha intorno ai 2-3 kHz, e decresce alle estremità dello spettro
- la sensibilità dell'orecchio cambia considerevolmente per fattori che variano da persona a persona (ad esempio l'età)



La percezione del suono nell'uomo (2)

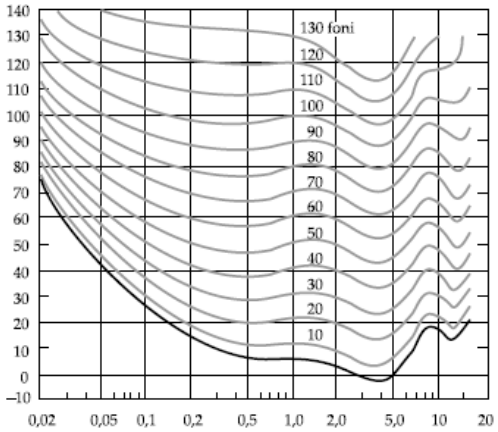
Come percepiamo il suono e la voce

- intervallo di frequenze udibili: ~ 20 Hz - 20 kHz
- l'*intervallo dinamico* riconoscibile, ossia l'intervallo dal suono più debole al più forte percepibile, è ~ 96 dB
- la voce umana ha frequenze nell'intervallo ~ 500 Hz (vocali) - 2 kHz (consonanti)



(Fonte: Institut Universitaire de Recherche Clinique - Montpellier)

Diagramma di Fletcher-Munson



$$T_q(f) = 3{,}64\,(f/1000)^{-0.8} - 6{,}5\,e^{-0.6(f/1000-3.3)^2} + 10^{-3}\,(f/1000)^4$$

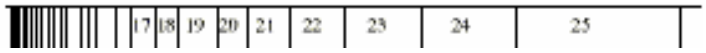
(Fonte: Vincenzo Lombardo e Andrea Valle, "*Audio e multimedia*", ed. Apogeo)

Bande critiche

Le frequenze per cui si ha una percezione uniforme dell'ampiezza possono essere riunite in *bande critiche*

- ogni banda ha un'ampiezza da 100 Hz a 4 kHz
- l'intero spettro delle frequenze udibili è suddiviso in 25 bande critiche

L'apparato uditivo umano può essere assimilato, a grandi linee, ad un banco di *filtri passa-banda* che si sovrappongono

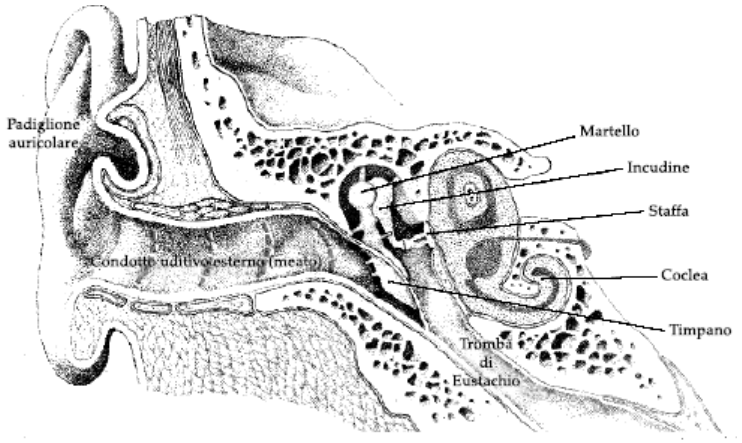


- 1 Bark = ampiezza di una banda critica (in onore del fisico tedesco **Heinrich Georg Barkhausen)**

Per frequenze minori di 500 Hz: $1\ Bark \approx freq/100$

Per frequenze maggiori di 500 Hz: $1\ Bark \approx 9 + 4\log\,(freq/1000)$

Struttura interna dell'orecchio umano

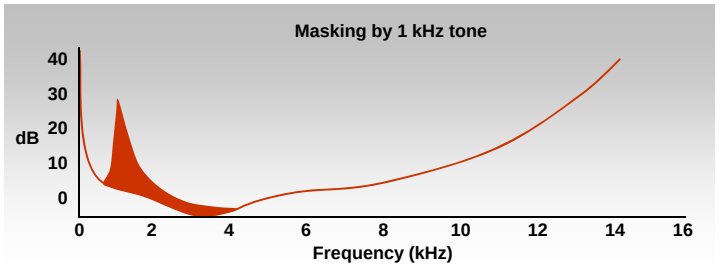


(Fonte: Vincenzo Lombardo e Andrea Valle, "*Audio e multimedia*", ed. Apogeo)

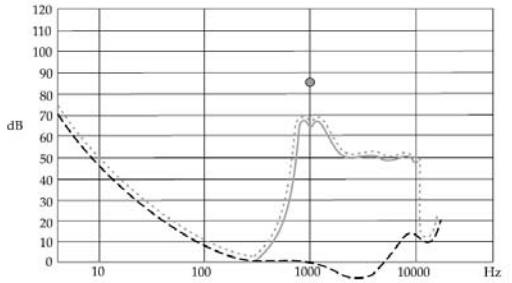
Mascheramento di frequenza (1)

Un suono puro può mascherarne un altro con frequenza vicina e livello più basso

- se riproduciamo un suono a 1 kHz, altri suoni contemporanei nell'intervallo di mascheramento non possono essere percepiti



Mascheramento tonale



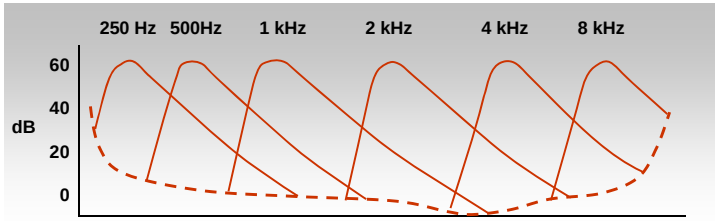
- Suono mascheratore (1000 Hz, 85 dB)
- Soglia di udibilità in quiete
- Curva di mascheramento
- Nuova soglia di udibilità

(Fonte: Vincenzo Lombardo e Andrea Valle, "*Audio e multimedia*", ed. Apogeo – pagg. 148-150)

Mascheramento di frequenza (2)

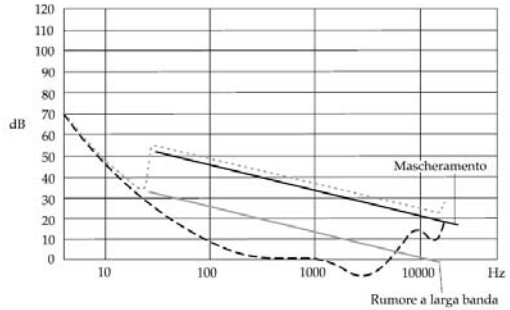
Il mascheramento

- è differente per ogni frequenza
  - ✓ può essere definito per ogni banda critica
- varia al variare dell'ampiezza del suono



Mascheramento non tonale

Il mascheramento non tonale avviene quando il suono mascheratore è una forma di rumore a banda più o meno larga in cui non è possibile individuare un tono specifico



- Soglia di udibilità in quiete
- Nuova soglia di udibilità

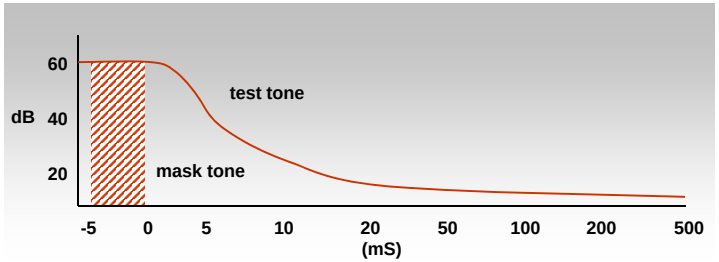
Rispetto al mascheramento tonale si nota come il mascheramento non tonale interessa addirittura suoni a un'intensità superiore a quella del mascheratore.

(Fonte: Vincenzo Lombardo e Andrea Valle, "*Audio e multimedia*", ed. Apogeo – pagg. 148-150)

Mascheramento temporale

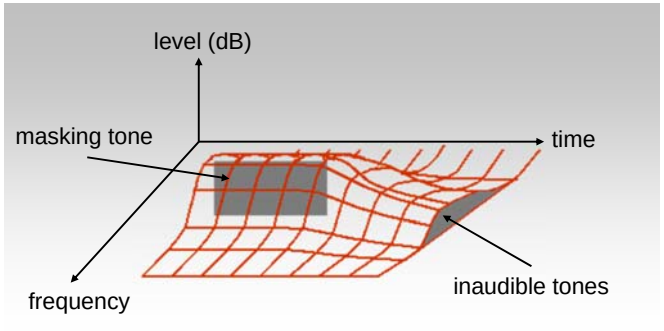
Un suono può mascherare un altro suono per un breve intervallo di tempo

- il *pre-mascheramento* nasconde i suoni precedenti il segnale di mascheramento in un intervallo che va dai 5 ms ai 40 ms
- il *post-mascheramento* nasconde i suoni più deboli successivi al segnale di mascheramento in un intervallo che va dai 50 ms ai 200 ms



Mascheramento combinato

Gli effetti del mascheramento di frequenza e di quello temporale si combinano



Proprietà dell'audio MPEG

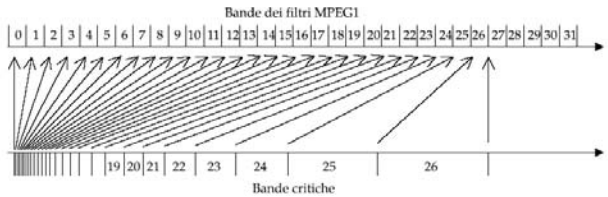
MPEG-1 layer 3 (MP3) è lo standard attuale per audio (musicale) di alta qualità con un elevato livello di compressione

- i *bit-rate* più comuni per lo standard MPEG in generale vanno da 48kbit/sec a 384 kbit/sec (CD audio non compresso è >1.4 Mbit/sec)
- il livello di compressione è nell'intervallo 2.7 - 24
- un livello di compressione di 6:1 (256 kbit/sec) è praticamente indistinguibile dal segnale originale
- da 96 a 128kbit/sec la qualità è ottimale per applicazioni domestiche (consumer)
- diverse frequenze di campionamento (32, 44.1 and 48 kHz)
- segnale monofonico, dual, stereo, joint stereo

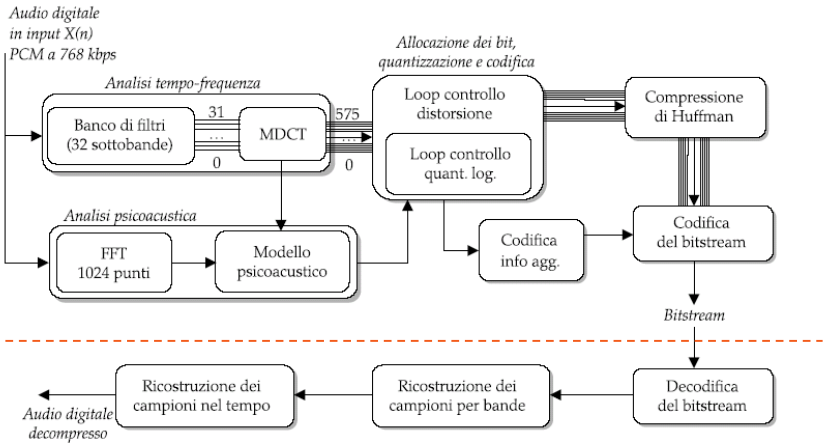
L'algoritmo di compressione audio MPEG (1)

E' diviso in quattro stadi che utilizzano le proprietà del modello psicoacustico

- divide il segnale audio in 32 sotto-bande di frequenza
- per ogni sotto-banda calcola la quantità di mascheramento
- se la potenza del segnale nella sotto-banda è inferiore alla soglia di mascheramento, il segnale non viene codificato
- altrimenti, calcola il numero di bit necessari per rappresentare il segnale (da 0 a 15) in modo che il rumore di quantizzazione sia inferiore alla soglia di mascheramento (1 bit ~ 6 dB di rumore)
- compone il flusso di bit secondo un formato standard per la trasmissione



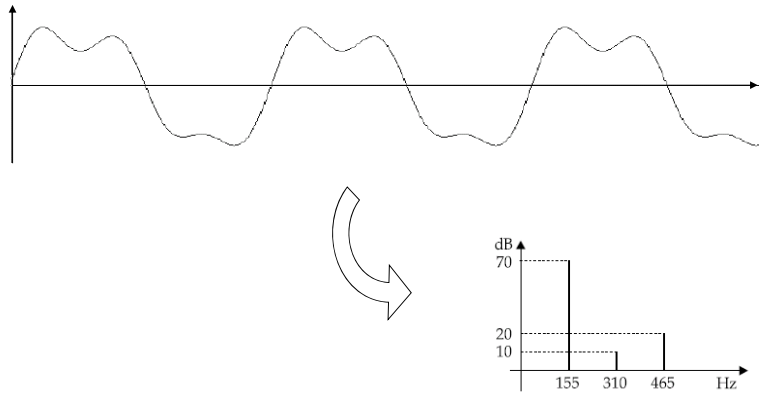
L'algoritmo di compressione audio MPEG (2)



(V. Lombardo, A.Valle, "*Audio e multimedia*", Apogeo)

Trasformata DCT nell'audio

Si passa dal dominio del tempo al dominio delle frequenze:



Audio MPEG: un esempio

Il livello in banda 8 è 60dB

- il mascheramento è di 12 dB sulla banda 7, 15dB sulla banda 9

Il livello in banda 7 è 10 dB (< 12 dB), viene ignorato

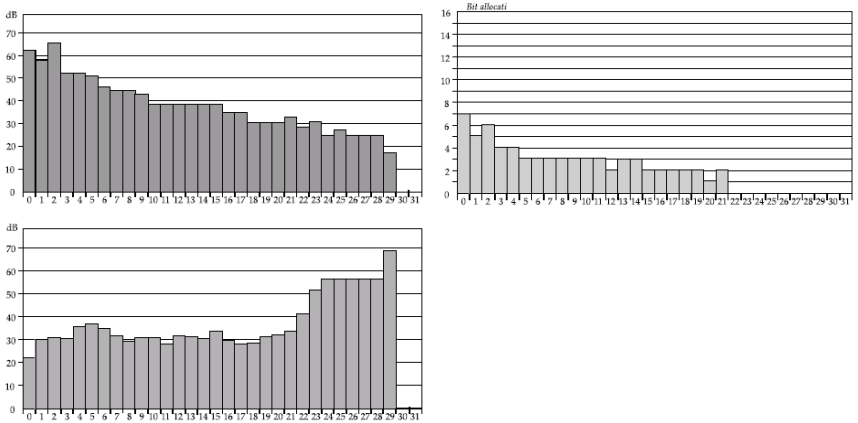
Il livello in banda 9 è 35 dB (> 15 dB), viene codificato

Solo la parte sopra la soglia di mascheramento deve essere codificata

- si usano 4 bit invece di di 6 (2 bit = 12 dB)

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

Allocazione dei bit per banda



(V. Lombardo, A.Valle, "*Audio e multimedia*", Apogeo)

Un'altra applicazione del mascheramento: il watermarking

Il *watermarking* è l'inclusione di informazioni digitali di vario genere (origine, destinazione, informazioni sul copyright, permessi di accesso, etc…) in maniera non percettibile all'interno di dati multimediali (immagini, video, audio, testo, animazioni)

Le informazioni (watermarks):

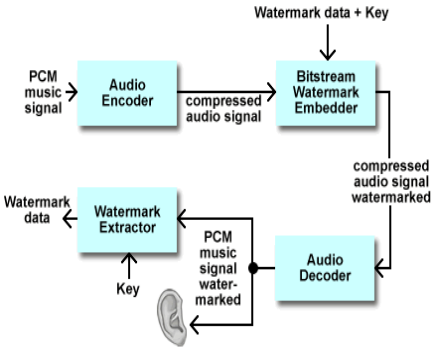
- non devono essere modificabili
- non devono modificare il dato che le ospita
- devono sopravvivere alle operazioni che vengono fatte sul segnale
- devono essere collegate direttamente ai dati (non nell'header)
- devono essere statisticamente invisibili

Bitstream watermarking

Per inserire watermarks in un bitstream audio si sfruttano gli effetti del mascheramento "al contrario" rispetto a come avviene nell'algoritmo MPEG

Affinché il segnale marcato sia inconfondibile da quello originario, il watermark viene inserito in **prossimità di segnali di livello alto**, in modo tale che esso venga mascherato da questi ultimi

Una successiva codifica MPEG eliminerebbe il watermark. Altri metodi consistono nell'inserire un segnale watermark la cui frequenza sia al di fuori dell'intervallo di frequenze udibili dall'orecchio umano



(fonte: Fraunhofer Institut Integrierte Schaltungen)

Livelli (layer) di codifica dell'audio MPEG (1)

Layer 1 (bitrate superiore a 128 Kbps): filtro DCT con un solo *frame* ed una ripartizione costante delle frequenze nelle sotto-bande

- il modello psicoacustico usa solo il mascheramento di frequenza
- ogni frame contiene 32 blocchi di 12 campioni, un intestazione, un codice di controllo degli errori (CRC) ed eventualmente informazioni aggiuntive

Layer 2 (bitrate uguale a 128 Kbps): usa tre *frame* durante il filtraggio (precedente, corrente, prossimo, per un totale di 1152 campioni)

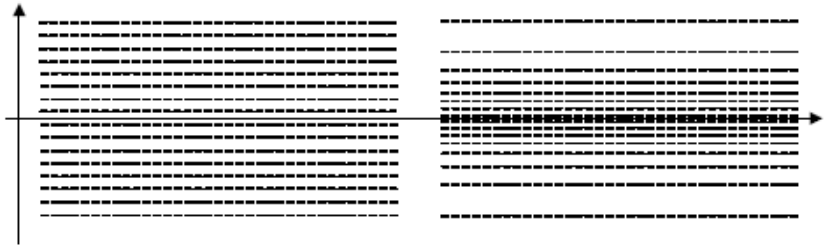
- utilizza in parte anche il mascheramento temporale
- utilizza una rappresentazione più compatta delle informazioni accessorie (intestazione, numero di bit allocati per banda, …)

Livello 3 di codifica dell'audio MPEG (MP3)

Layer 3(bitrate a 64 Kbps): ripartisce lo spettro di frequenza in sotto-bande di ampiezza non uniforme più simili alle bande critiche nelle frequenze più basse

- il modello psicoacustico comprende il mascheramento temporale
- considera anche la ridondanza stereo
- implementa un bitrate variabile:
  - ✓ utilizza un compressore di Huffman su coppie di valori
  - ✓ utilizza una "riserva" dei bit dove vengono messi i bit "avanzati" durante la codifica di un frame e quindi disponibili per le codifiche successive

Quantizzazione non uniforme



(Fonte: Vincenzo Lombardo e Andrea Valle, "*Audio e multimedia*", ed. Apogeo)

Qualità dell'audio MPEG

Layer	Target bit rate	Compres-sione	Qualità a 64 kb	Qualità a 128 kb	Ritardo
Layer I	192 kb/s	4:1	--	--	19 msec
Layer II	128 kb/s	6:1	< 3	4+	35 msec
Layer III	64 kbit/s	12:1	< 4	4+	59 msec

Fattore di qualità: 5 - perfetta, 4 - appena percettibile, 3 - leggermente fastidiosa 2 - fastidiosa, 1 - molto fastidiosa

Successivi formati di audio MPEG

MPEG2 (Novembre 1994)

- è la codifica usata nei DVD
- introduce la gestione dei canali in surround, cioè 5 canali audio (sinistro, centrale, destro, sinistro surround, destro surround), più un canale supplementare per le bassissime frequenze (subwoofer)
- lavora a 16 kHz, 22.05 kHz o 24 kHz oltre ai tassi MP3

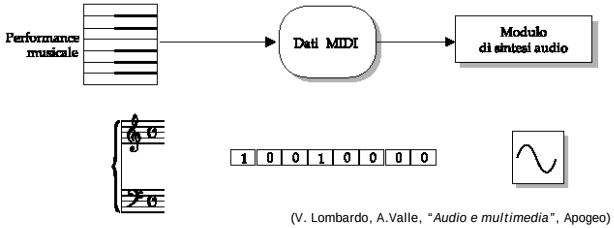
MPEG4 (Dicembre 1999)

- audio (e immagini) vengono visti come composizione di più oggetti diversi. Un utente può decidere di ascoltare un concerto ambientato in posti differenti e di mettere in risalto alcuni suoni su altri

Musical Instruments Digital Interface

Il protocollo MIDI (1983) fornisce un modo standard ed efficiente per descrivere eventi musicali

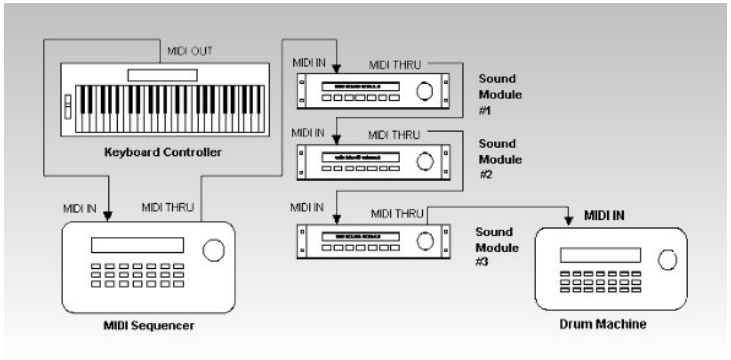
- permette a computer, sintetizzatori, tastiere e altri strumenti musicali elettronici di comunicare tra di loro
- I messaggi MIDI sono istruzioni che dicono ad un sintetizzatore "come" suonare un brano musicale
- la generazione del suono è locale al sintetizzatore
- i messaggi descrivono il tipo di strumento usato, la note suonate, il loro volume, la velocità, l'attacco, gli effetti, …



(V. Lombardo, A.Valle, "*Audio e multimedia*", Apogeo)

I sistemi MIDI

I sistemi MIDI possono essere molto complessi…



… ma la maggior parte di sound card è corredata dell'hw necessario

Il MIDI sequencer

MIDI sequencer

- un sistema di registrazione ed esecuzione dotato di una memoria programmabile, nella quale vengono memorizzati i dati di controllo operativi necessari alla generazione di eventi musicali
- riceve i dati da un dispositivo di input, ne consente l'editing, e crea la musica inviandoli al dispositivo che si occupa della sintesi (es. scheda audio)
- non influenza la qualità che dipende totalmente dal modulo di sintesi (o sintetizzatore)

Canali e tracce MIDI

Canali

- permettono di spedire e ricevere i dati musicali
- sono un metodo per differenziare i timbri e spedire informazioni indipendenti: a canale diverso corrisponde uno strumento diverso
- il protocollo MIDI ne prevede solamente 16 numerati da 1 a 16

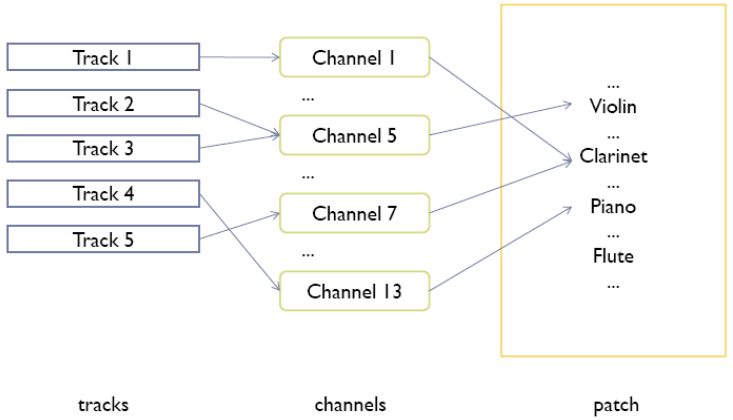
Tracce

- una traccia è un flusso strutturato e autonomo di messaggi MIDI
- esempio: in un brano suonato da un pianoforte ci sono due tracce, la melodia e l'accompagnamento
- si può considerare come un contenitore di messaggi che possono essere assegnati a canali differenti

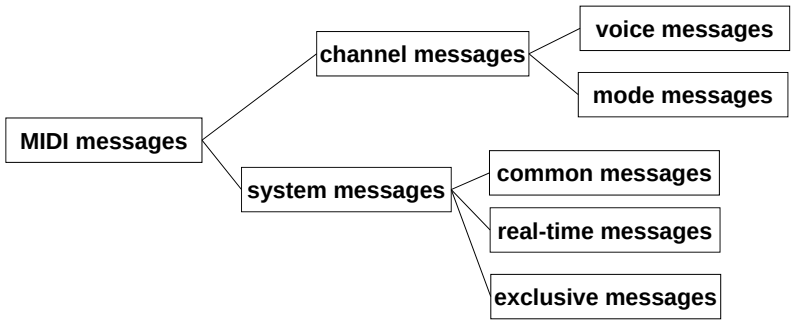
Patch

- indica il timbro prodotto da un generatore
- i MIDI possono contenere fino ad un massimo di 128 patch differenti

Rappresentazione della musica in MIDI



I messaggi MIDI



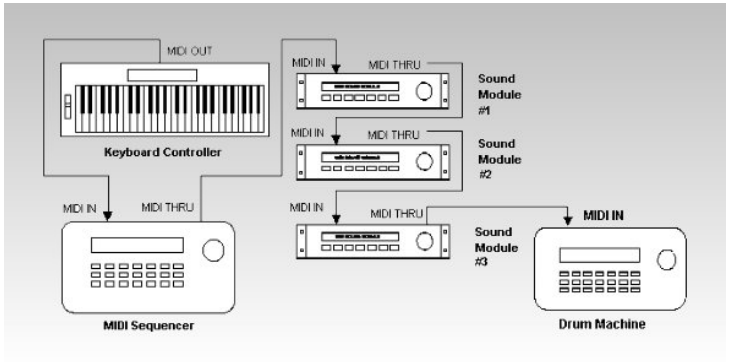
I messaggi di canale (*channel message*) descrivono le note suonate (*voice*) e il modo di suonarle (*mode*)

I messaggi di sistema (*systems message*) trasportano informazioni di set-up e di sincronizzazione

Tutti i messaggi MIDI sono costituiti da sequenze di 10 bit (1 byte di dati utili)

I sistemi MIDI

I sistemi MIDI possono essere molto complessi…



… ma la maggior parte di sound card è corredata dell'hw necessario

Struttura dei messaggi

Messaggi di canale

1° nibble				2° nibble				
1	x	x	x	n	n	n	n	canale

tipo messaggio

Messaggi di sistema

1° nibble				2° nibble				
1	1	1	1	x	x	x	x	tipo messaggio

Byte successivi

1° nibble				2° nibble			
0	x	x	x	x	x	x	x

Messaggi di canale

Indicano il numero di canale su cui è veicolata l'informazione

I *voice message* descrivono cosa deve suonare uno strumento:

- quale nota deve essere eseguita (*Note on*)
- quale nota deve essere terminata (*Note Off*)
- eventuali modificazioni espressive (es. vibrato) (*Pitch Bend Change*)
- il cambiamento di pressione sul tasto (*Channel pressure)*
- …

I *mode message* descrivono come si comporta uno strumento all'arrivo del voice message

- Omni On/Off
- Poly/Mono
- General MIDI Mode

Messaggi di Sistema (1)

Non sono indirizzati specificatamente ad un canale, ma sono rivolti a tutto il sistema

Ogni dispositivo risponde solo ai messaggi a cui è abilitato a rispondere

System common message

- svolgono funzioni generali relative a tutto il sistema (es. la sincronizzazione di un brano eseguito da più dispositivi)
- settano un clock comune
- posizionamento all'interno del brano (*Song Position Pointer*)
- selezionamento di una traccia (*Song Select*)

Messaggi di Sistema (2)

System real time message

- si occupano del funzionamento sincronizzato dei diversi moduli di un sistema in tempo reale
- sincronizzano i dispositivi sulla base di un tempo relativo (24 messaggi ogni quarto)
- iniziano o fermano la riproduzione di un dispositivo (*Start*/*Stop*/*Continue*)
- svolgono funzioni di reset

System exclusive message

- sono messaggi attraverso i quali i costruttori possono veicolare informazioni specifiche ai loro prodotti

MIDI: perché e quando (1)

MIDI è un modo efficiente per codificare suoni musicali all'interno di documenti Web

- i file MIDI sono compatti e contengono informazioni sulla temporizzazione (non ci sono vincoli *hard real-time*)
- non si rappresenta la forma d'onda del suono, ma solo eventi discreti di tipo predefinito
- brani musicali complessi occupano un piccolo spazio di memoria

Particolarmente adatto per musica di sottofondo

MIDI: perché e quando (2)

Ma...

- può essere descritta solo la musica tradizionale occidentale (scala tonale)
- non è possibile rappresentare suoni quali rumori, voce, altri fenomeni acustici
- i computer devono avere opportune schede audio (molto comuni, ma non tutte le schede audio di base sono adeguate)
- la qualità dipende dall'equipaggiamento MIDI (sintetizzatore)
- la codifica dei canali e dei messaggi non è completamente standard (es. Roland, Yamaha, ...)