

Elementi d' Informatica Musicale e Acustica

Per i corsi strumentali

a cura di E. Giordani

Conservatorio Statale di Musica G. Rossini - Pesaro

CAPITOLO 1

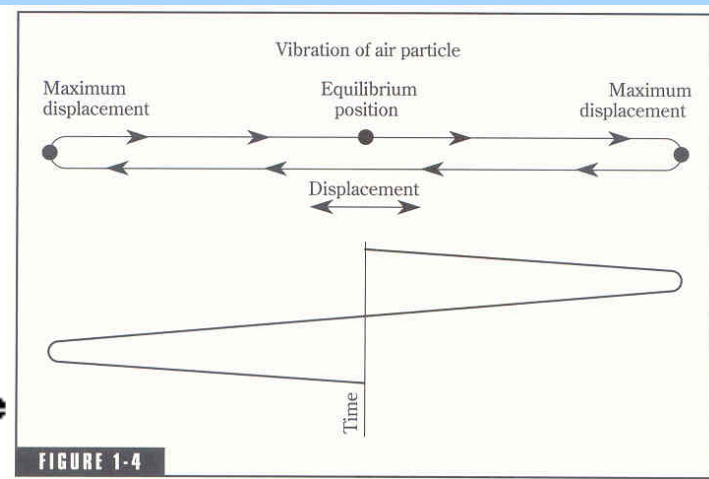
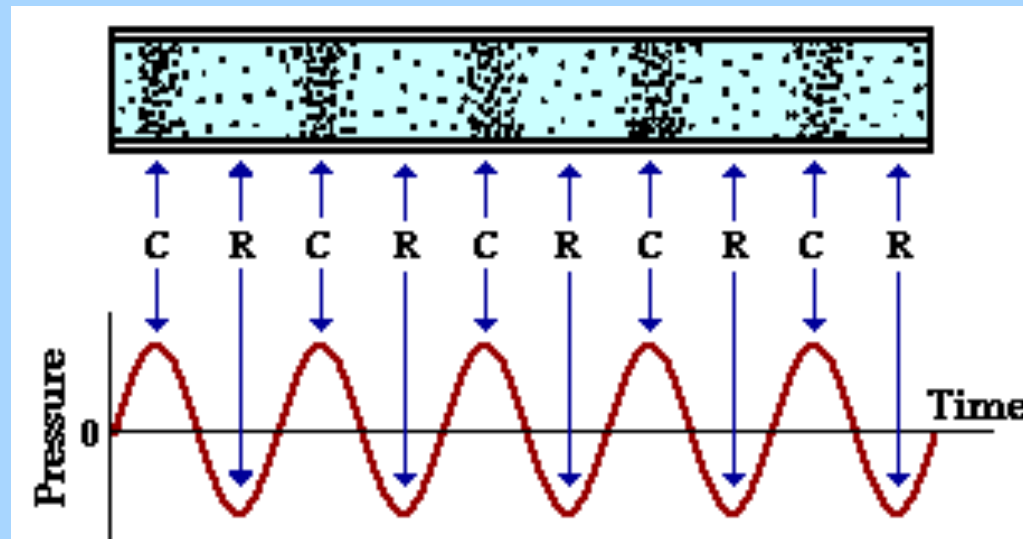
Elementi di Fisica del Suono e sue
rappresentazioni

Definizioni

- Il suono è una qualità psicologica percepita dall'apparato uditivo.
- Il suono è costituito fisicamente da un'onda che si propaga nell'aria.
- L'onda sonora è sostanzialmente una variazione periodica o aperiodica della pressione atmosferica.

Onde Acustiche

Un'onda acustica è costituita da variazioni periodiche/aperiodiche della pressione atmosferica che si propagano nello spazio. Rispetto al valore di riferimento, la pressione può aumentare (compressione) o diminuire (rarefazione)

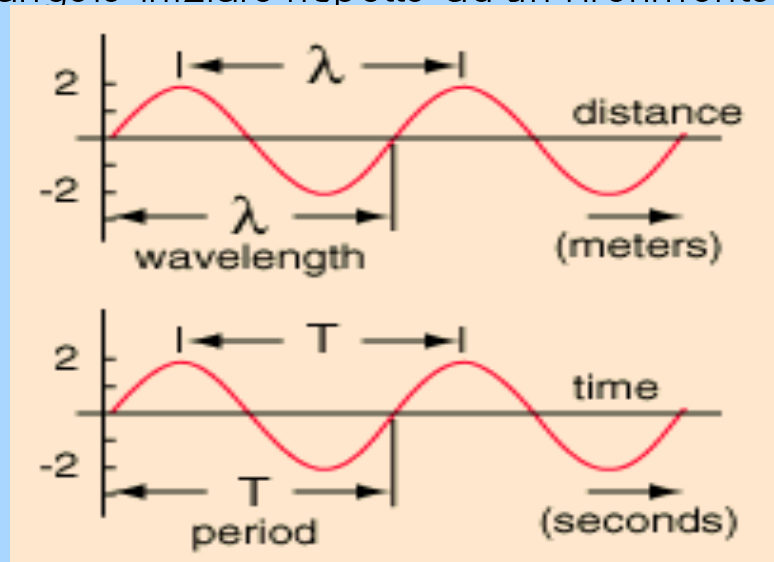


Le singole particelle si muovono solo localmente ma insieme trasferiscono un'onda in questo caso di natura **periodica**

Oscillazioni periodiche

Il moto periodico può essere descritto con i seguenti parametri fisici:

- Periodo:** il tempo richiesto per completare un ciclo intero, T in sec
- Frequenza:** il numero di cicli per secondo, f in 1/seconds or Hertz (Hz)
- Ampiezza:** il massimo spostamento dall'equilibrio
- Fase:** angolo iniziale rispetto ad un riferimento (zero)



Se il moto è in forma di un'onda viaggiante, occorre definire:

- Velocità di propagazione:** rapidità di spostamento, v (m/sec)
- Lunghezza d'onda:** spazio occupato da una oscillazione, λ (m)

Relazioni fondamentali

- Relazione di tipo “temporale”
tra frequenza (f) e periodo (T)

$$f = 1/T$$

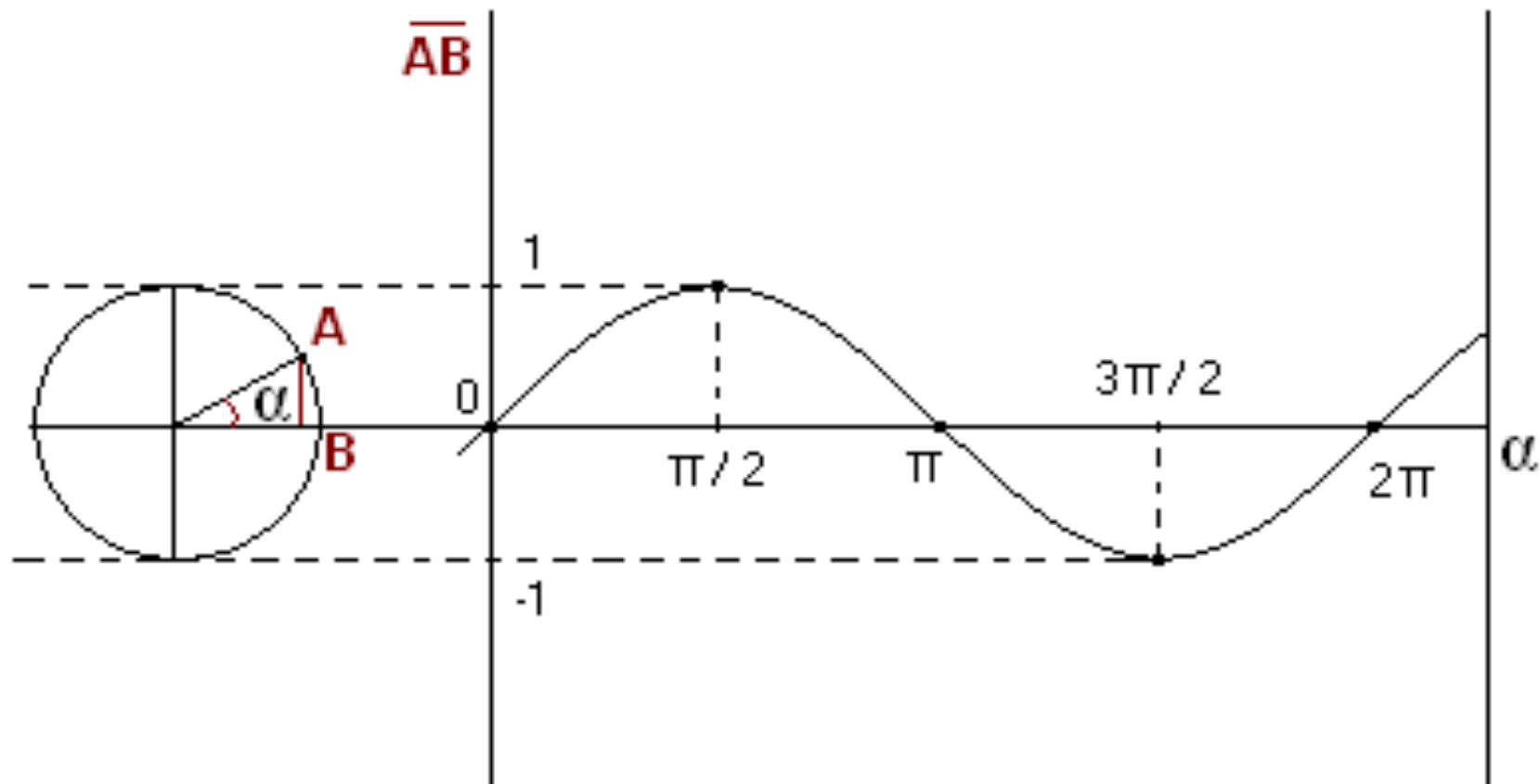
- Relazione di tipo “spazio-temporale”

$$\lambda = vT$$

Il periodo T si misura in secondi, la frequenza in Hz e la lunghezza d'onda λ in metri. (sono usati sottomultipli per T e λ)

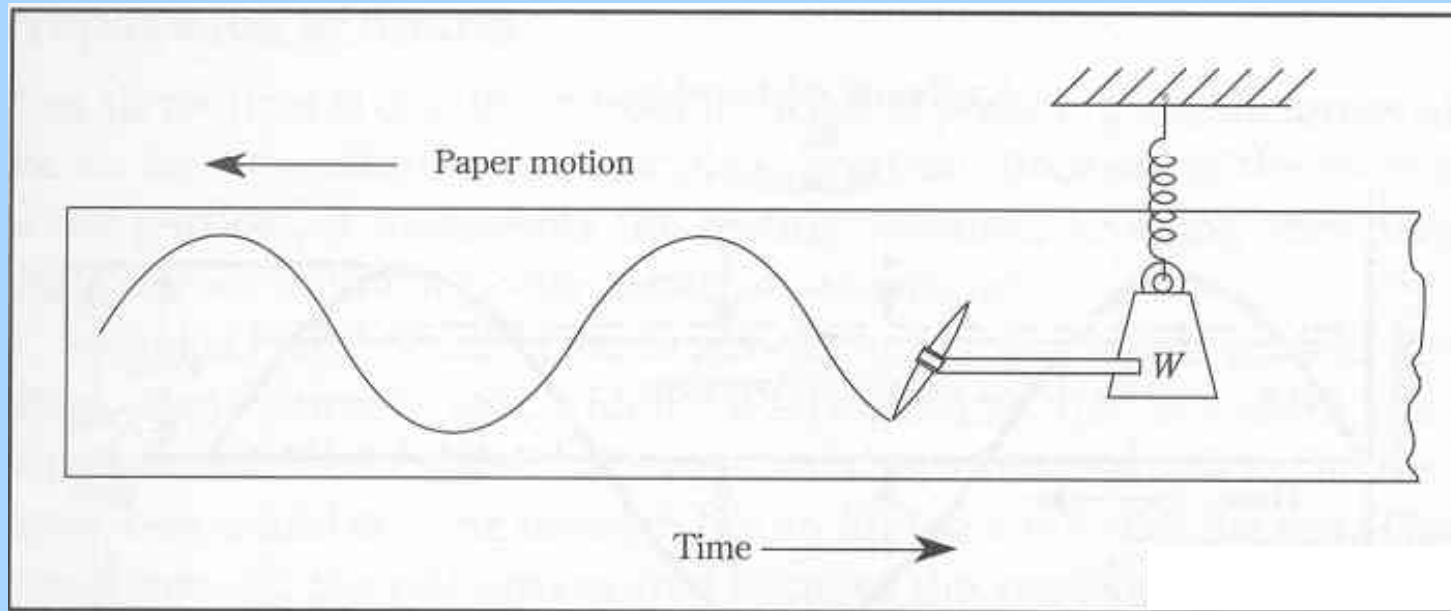
Onde Sinusoidali..

Genesi geometrica di una senoide



..Onde Sinusoidali

Genesi fisica di una senoide



Onde e Segnali

Segnale : è la rappresentazione analitica o grafica di una grandezza fisica che varia in funzione di una o più variabili indipendenti.

Esempi:

- Onda sonora la cui ampiezza varia in funzione del tempo (segnale monodimensionale)
- Immagine come segnale bidimensionale (luminosità e colore in funzione delle coordinate spaziali $(x;y)$)
- Filmato video come segnale tridimensionale (luminosità e colore in funzione delle coordinate spaziali e del tempo $(x;y; t)$)

Segnali non sinusoidali

I segnali non sinusoidali appartengono alle seguenti categorie:

1. Segnali periodici complessi (somma di sinusoidi di frequenze multiple di una fondamentale)
2. Segnali pseudo-periodici (la gran parte dei suoni strumentali)
3. Segnali aperiodici
 - continui** (rumore bianco, rosa..)
 - impulsivi** (spikes, click, pop..)

Intervalli e Sistemi d'intonazione

- Gli intervalli musicali sono rappresentati da rapporti numerici che ne indicano l'estensione. Ad esempio 2:1 identifica il rapporto d'ottava (es. 880 e 440 Hz).
- Intervalli uguali definiscono differenze frequenziali diverse a seconda della zona. Ad esempio 880 : 440 Hz (intervallo = 2:1 , differenza = 440 Hz) oppure 1760:880 Hz (intervallo = 2:1, differenza 880)
- Valori di rapporti < 2 identificano intervalli minori di una ottava mentre rapporti > 2 intervalli maggiori
- Ad esempio $3:2 = 1.5$ rappresenta un intervallo di QUINTA naturale che differisce dalla quinta temperata.
- La generazione delle scale nell'antichità greca con il metodo *ciclico* di quinte ascendenti e quarte discendenti derivanti dalla divisione della corda del monocordo. Così facendo il ciclo tende a non essere in realtà una figura a spirale e non c'è coincidenza tra i rapporti di ottava.

Sistema temperato equalizzato

Il sistema equalizzato divide in maniera uniforme l'intera ottava in 12 suoni.

Detta F1 una frequenza di partenza, il semitono successivo si ottiene moltiplicando tale valore per un fattore k come segue:

F1

$$F2 = F1 * k$$

$$F3 = F2 * k = F1 * k * k = F1 * k^2$$

$$F4 = F3 * k = F1 * k^3.$$

.

$$F13 = F12 * k^{12}$$

F13 è l'ottava di F1 cioè $F13 = 2 F1$. Allora si ha:

$$2F1 = F12 * k^{12} \text{ quindi } 2 = k^{12} \text{ da cui } k = \sqrt[12]{2} = 1,0594630943.....$$

(Corrisponde circa ad una variazione del 6 % della frequenza)

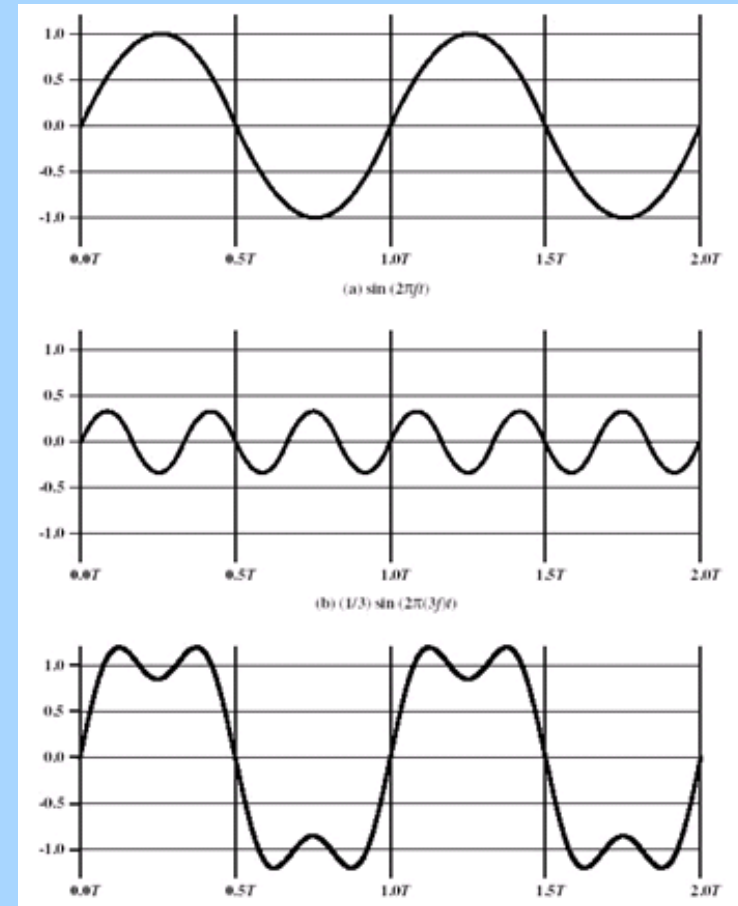
Segnali periodici non sinusoidali

La somma di segnali sinusoidali di frequenza multipla produce segnali periodici di forma non sinusoidale.

La frequenza del segnale risultante (somma algebrica) coincide con la frequenza più bassa ed è detta **fondamentale (F0)**

L' n-esima componente (**armonica**) ha frequenza pari a **$f_n = n \cdot F_0$**

Nell'esempio riportato il terzo segnale è la somma dei primi due ed ha la stessa frequenza del primo.



Armoniche e parziali

Ogni suono che sia percepito con un'altezza (pitch) definita è costituito generalmente da un insieme di suoni elementari (sinusoidali) di frequenza pari alla seguente successione di numeri interi:

1, 2, 3, 4, 5, 6, 7, 8, 9,.....

Questo significa che il secondo suono (seconda armonica) ha una frequenza doppia rispetto alla prima mentre il terzo suono 3 volte. I rapporti numerici tra armoniche diverse determinano i relativi intervalli. Ad esempio 3:2 è un intervallo di quinta naturale mentre 4:3 forma una quarta naturale.

Quando le frequenze deviano dalla successione indicata le armoniche sono più frequentemente indicate con il termine "parziali".

Rappresentazione notazionale dei suoni armonici

Le armoniche di un suono (o suoni armonici) possono essere rappresentate approssimativamente sul pentagramma secondo lo schema seguente

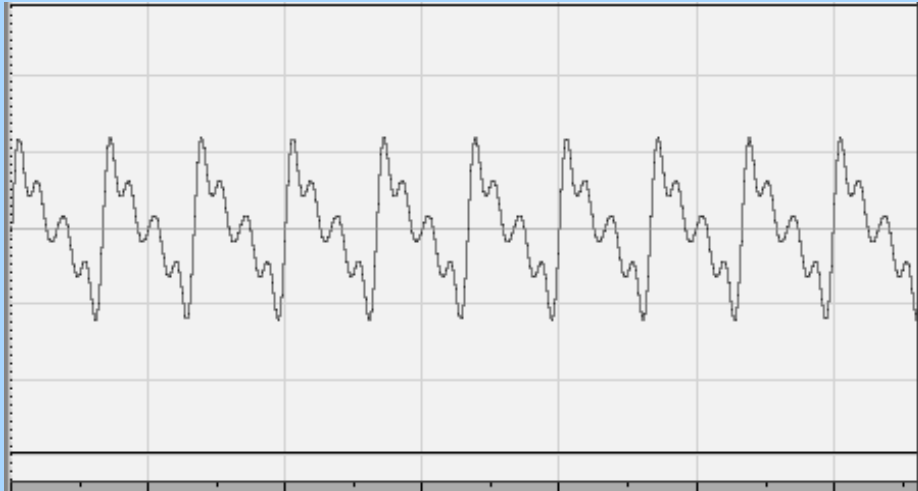
i suoni armonici

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

* in cents, confrontati con la scala temperata

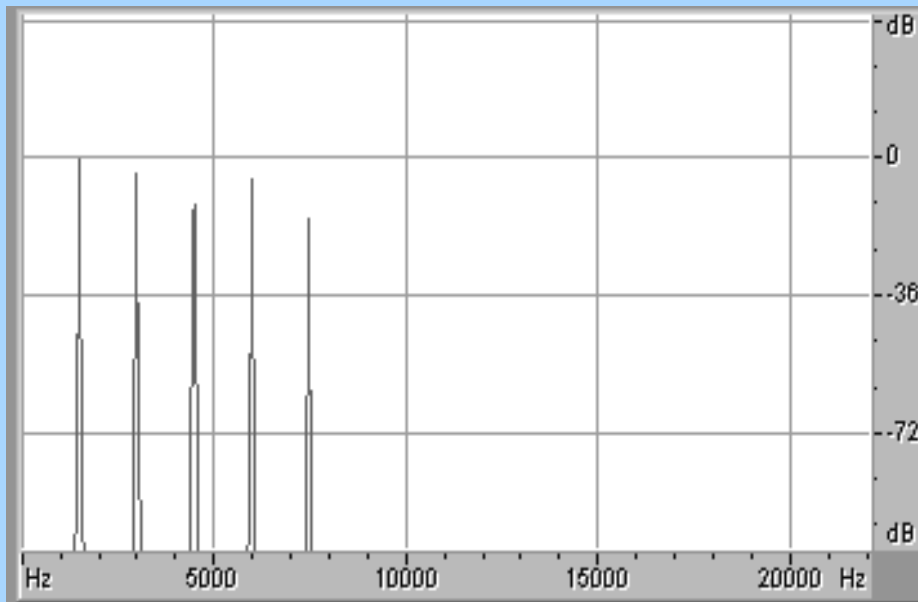
Nota: 1 cent = 1/100 di semitono temperato

Rappresentazione spettrale



Segnale rappresentato
in funzione del tempo

$S(t)$



Segnale rappresentato in
funzione della frequenza
(SPETTRO d'AMPIEZZA)

$S(f)$

Serie di Fourier

- **Definizione:** ogni segnale periodico complesso può essere rappresentato come sommatoria (serie) di segnali elementari (sinusoidali)

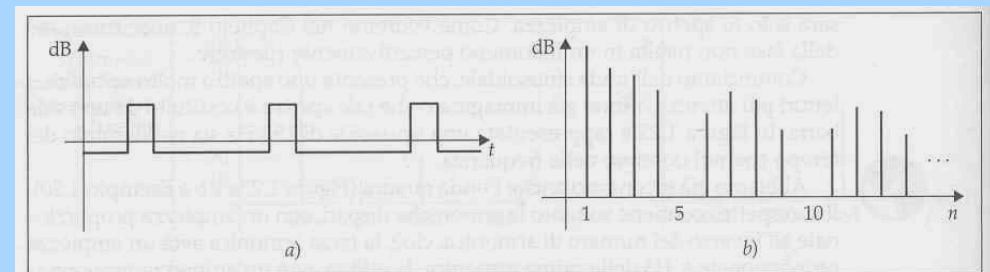
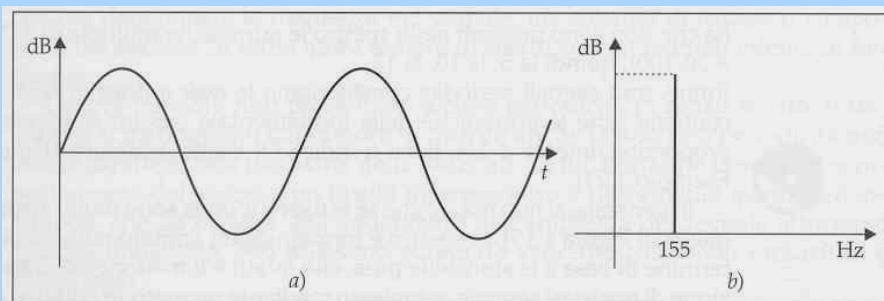
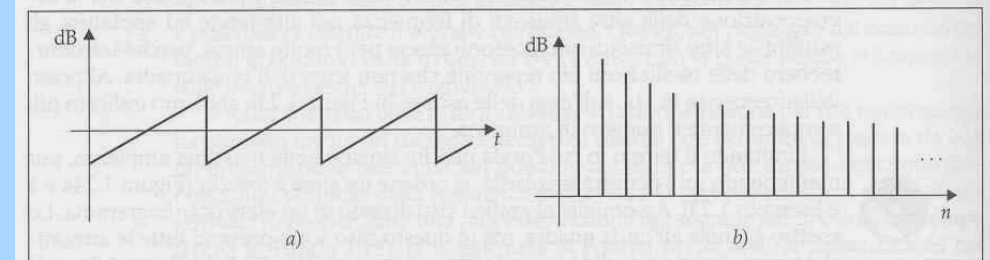
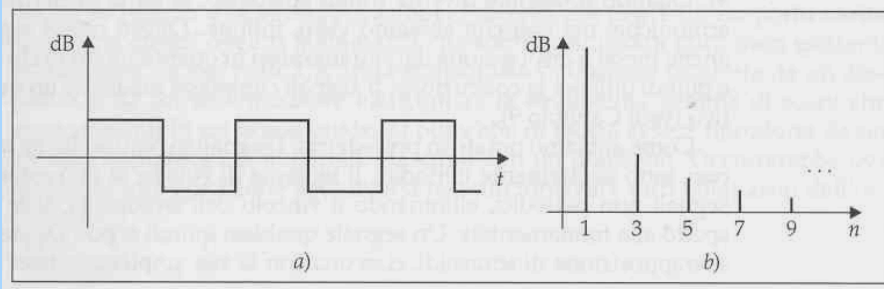
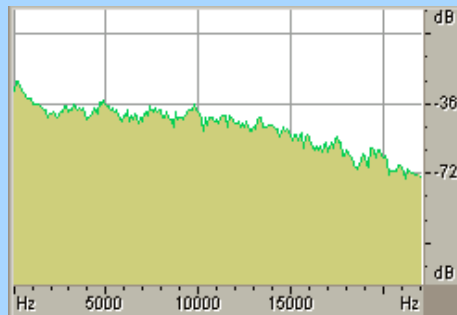
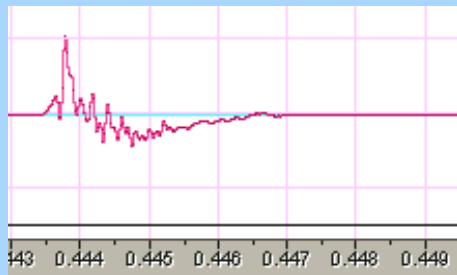


Figura 1.24 Onda a impulsi: nel dominio del tempo (a) e della frequenza (b)

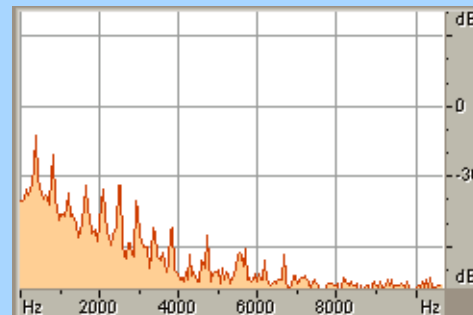
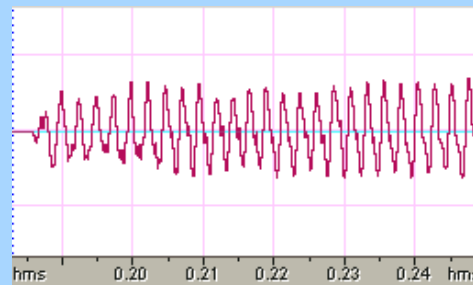


Trasformata di Fourier

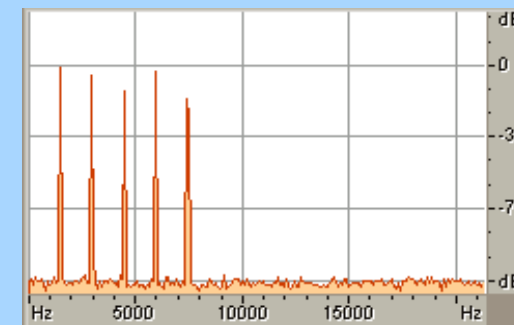
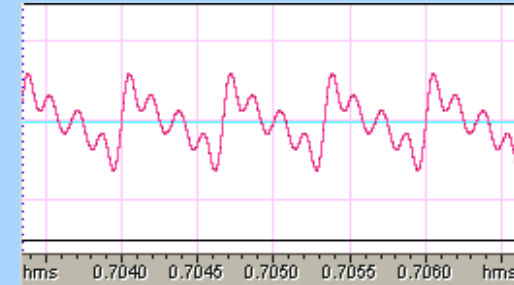
Definizione: ogni segnale, periodico e non periodico, continuo o impulsivo, può essere rappresentato da una funzione nel dominio frequenziale



Impulsivo



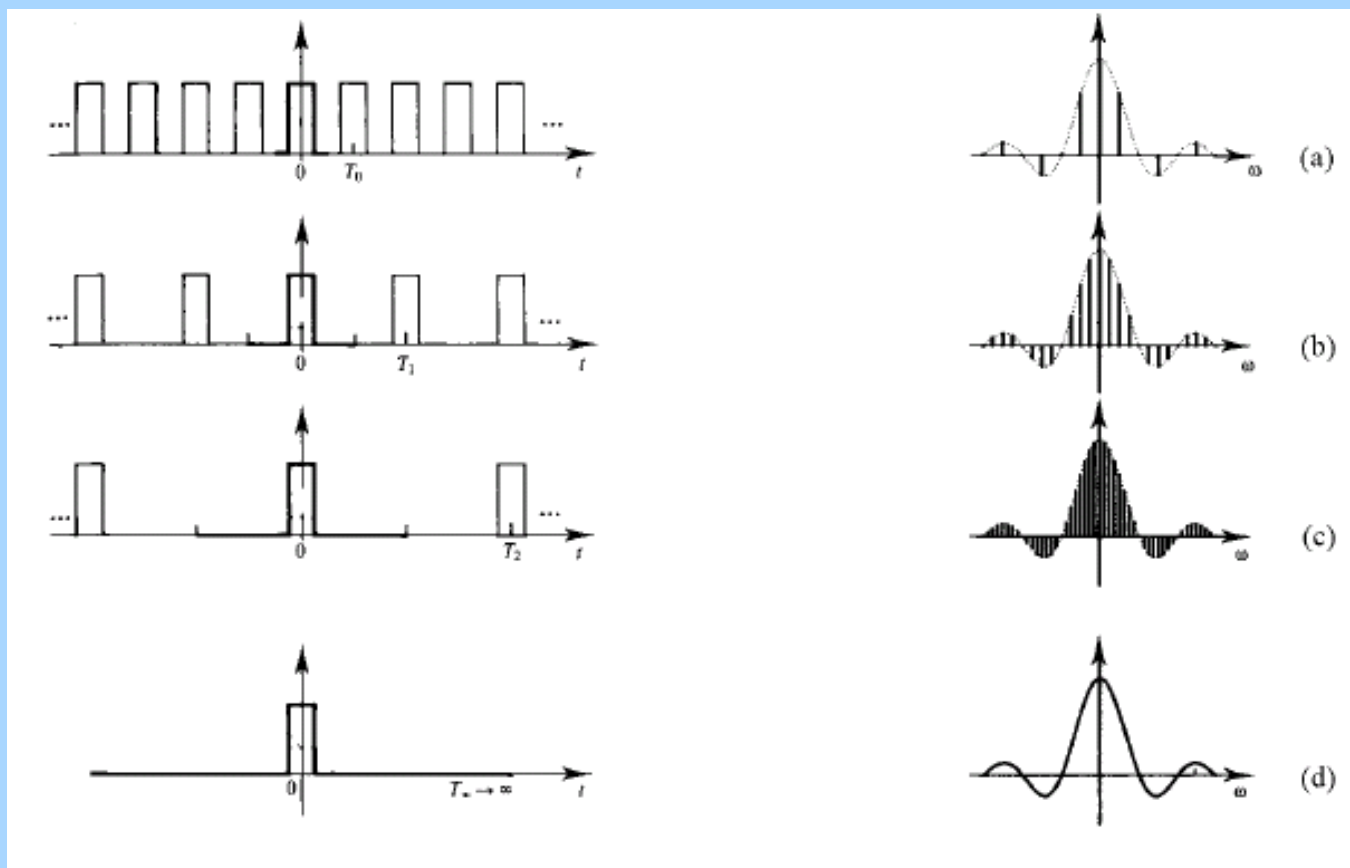
Quasi periodico



Periodico

L'impulso

Un impulso può essere considerato il caso limite di un segnale periodico di periodo infinito e frequenza tendente a zero.

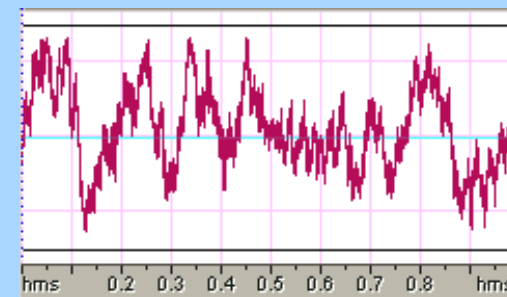
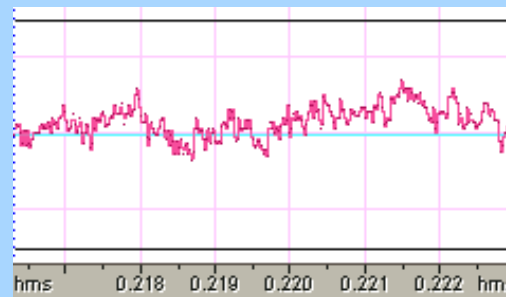
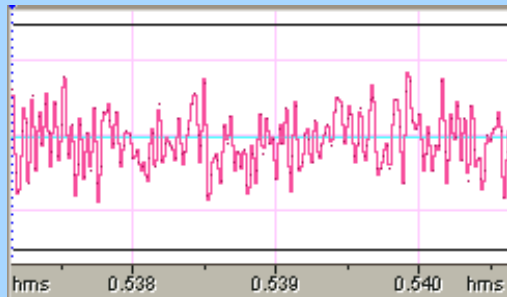


All'aumentare del periodo le linee dello spettro si fanno sempre più ravvicinate, fino alla situazione limite di un unico impulso aperiodico (d).

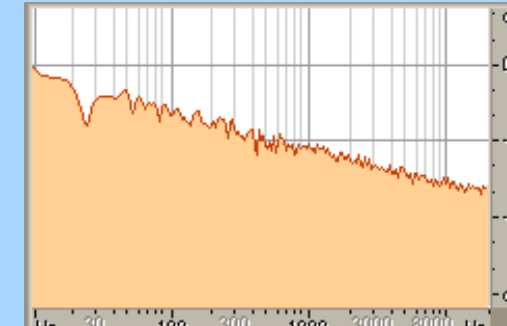
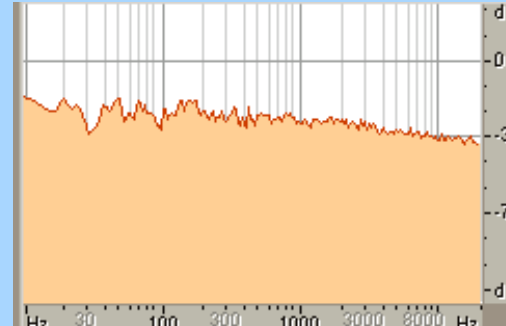
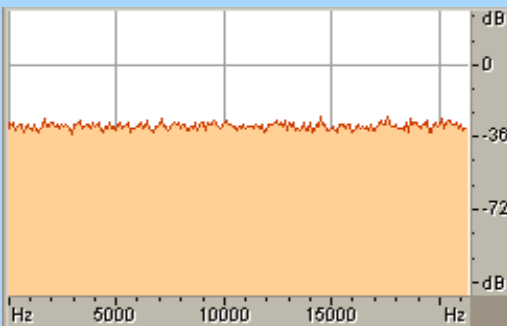
Rumore

Definizione: segnale aperiodico di ampiezza casuale contenente statisticamente tutte le frequenze. Si distingue in:

- Rumore bianco (White Noise) : distrib. Freq = 1
- Rumore rosa (Pink Noise) : “ “ = $1/f$
- Rumore marrone (Brown Noise) : “ “ = $1/f^2$



time



FFT

White Noise

Pink Noise

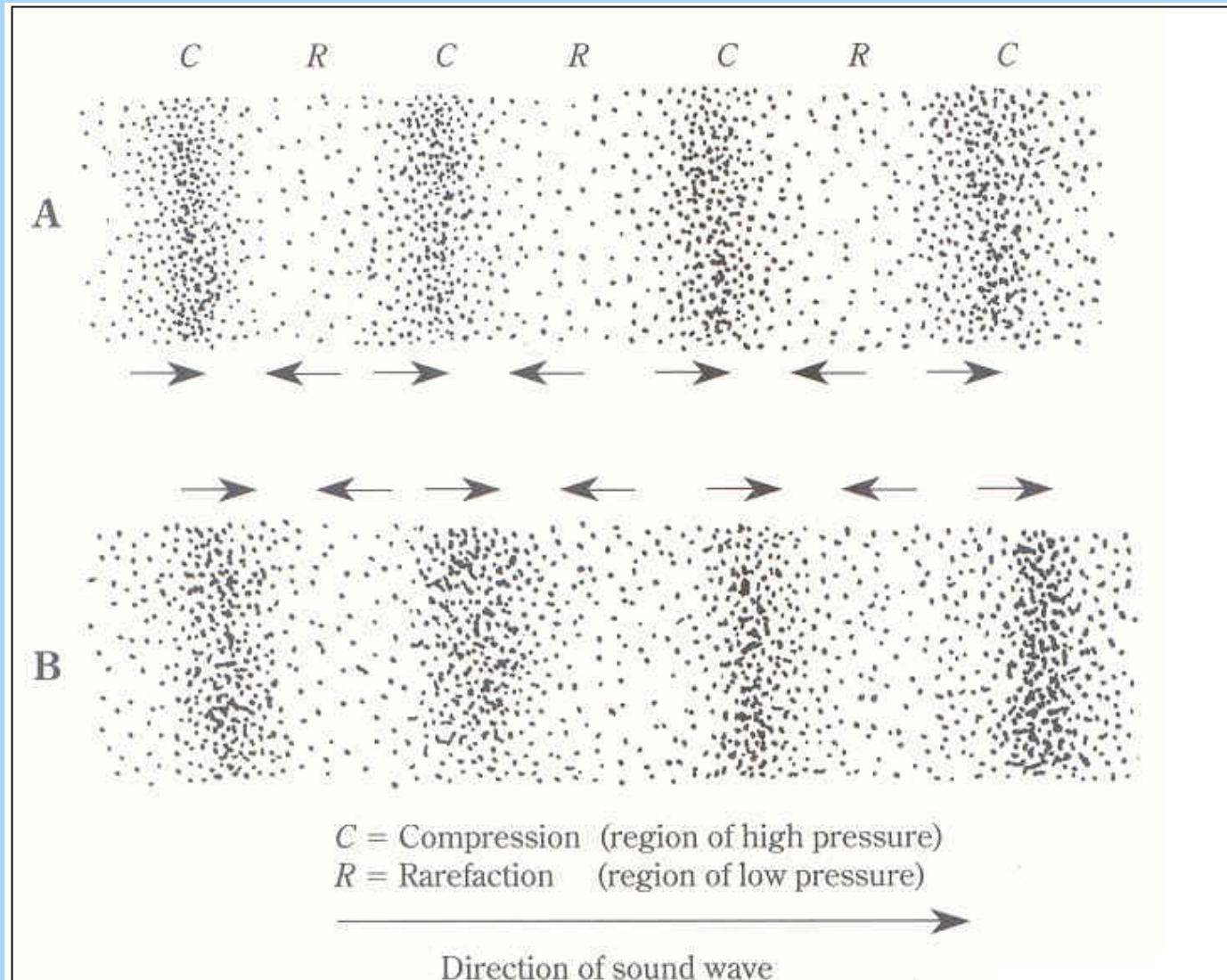
Brown Noise

Trasformata rapida di Fourier (FFT)

Il procedimento matematico necessario per calcolare il contenuto spettrale di un segnale può essere velocizzato attraverso un algoritmo specifico (FFT : fast fourier transform) che trova larga applicazione nei sistemi digitali.

NB: di fatto le immagini degli spettri nella pagina precedente sono ottenuti effettivamente calcolando l'FFT

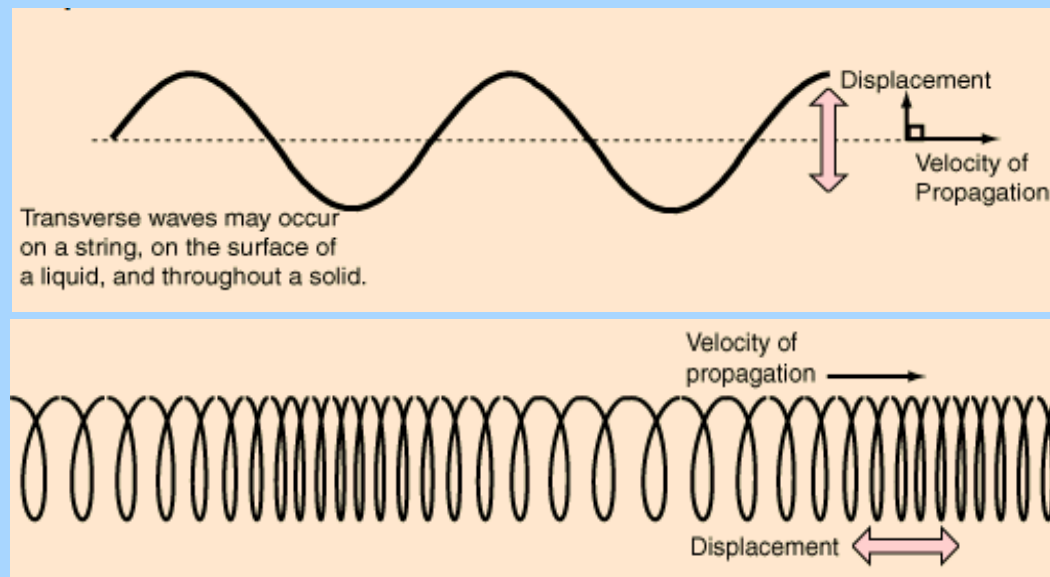
Propagazione



Tipi di onde

Il suono appartiene dal punto di vista fisico alla categoria delle onde.
Tra le varie tipologie possono essere distinte principalmente in base alla modalità di propagazione:

- A) Propagazione trasversale: il moto vibratorio è perpendicolare al verso di propagazione
- B) Propagazione longitudinale : il moto vibratorio è parallelo al verso di propagazione (onde sonore)

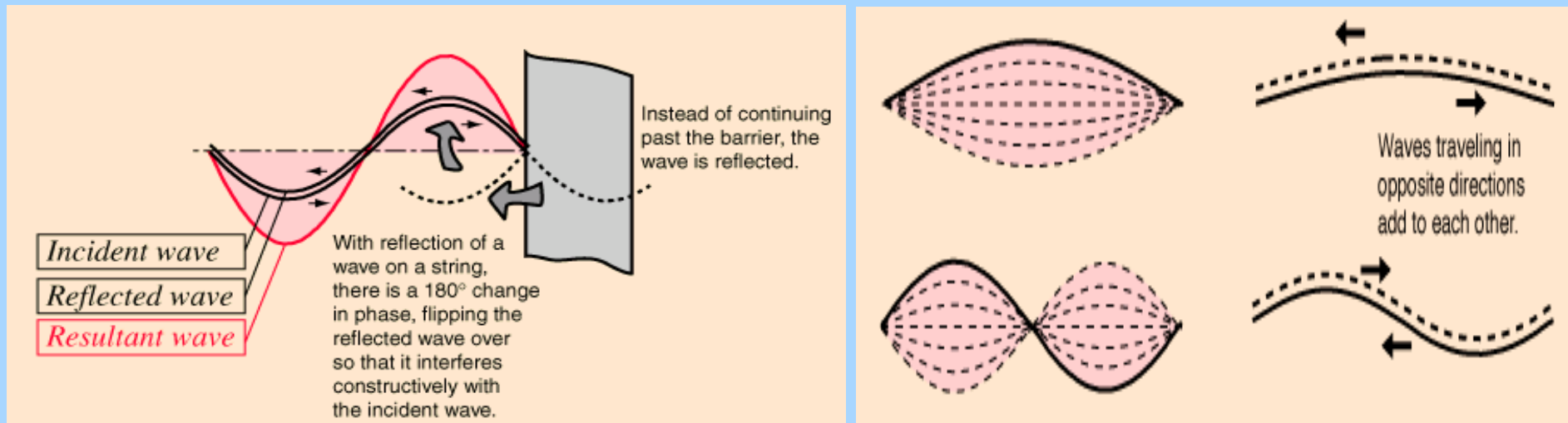


Trasversale

Longitudinale

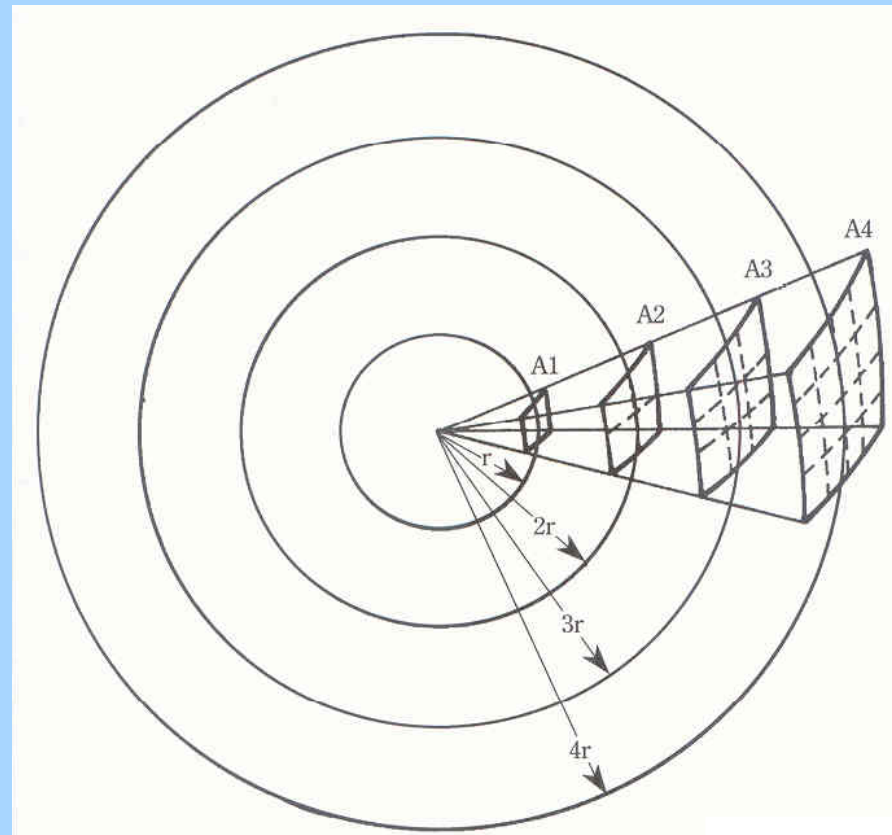
Onde stazionarie

Quando un'onda nel propagarsi incontra un ostacolo viene riflessa. In certe condizioni di simmetria le onde incidenti e riflesse possono dar luogo alle onde stazionarie (onde responsabili della generazione del suono negli strumenti musicali)



Propagazione nello spazio libero

- La propagazione del suono nello spazio a 3 dimensioni è di tipo sferica



Velocità del suono

Nell'aria, la velocità del suono è data circa dalla seguente relazione:

$$v_{\text{sound in air}} \approx 331.4 + 0.6T_C \text{ m / s}$$

dove T_C = temperatura dell'aria in gradi centigradi.

La velocità varia in funzione della temperatura e del tipo di fluido (vedi tabella)

Liquid	Temperature (°C)	Speed in m/s
Water	0	1402
Water	20	1482
Methyl alcohol	0	1130
Sea water 3.5% salinity	20	1522

Pressione Sonora

- Poiché il suono udibile consiste in onde di pressione, uno dei modi di misurare il suono è valutare la variazione della pressione atmosferica causata dal suono.
- A causa della grande sensibilità dell'udibilità umana, la soglia di udibilità corrisponde ad una variazione di pressione di circa un miliardesimo di pressione atmosferica (1 /1.000.000.000.000)
- Tale pressione (soglia) equivale alla quantità:

$$p_0 = 2 \times 10^{-5} \text{ Newtons / m}^2$$

- La pressione sonora (espressa in dB è definita livello di pressione sonora) è quindi data da:

$$L_p(\text{dB}) = 20 \log p/p_0$$

Intensità sonora

- Ogni sorgente sonora emette una certa quantità di potenza espressa Watt. L'intensità sonora è definita come la potenza sonora per unità di superficie espressa in watt/m²
- Analogamente alla pressione sonora, la soglia di udibilità in termini di intensità è pari a:

$$i_0 = 10^{-12} \text{ watt/m}^2$$

- L'intensità sonora (espressa in dB è definita livello d'intensità sonora) è quindi data da:

$$Li(\text{dB}) = 10 \log i/i_0$$

Le due quantità sono legate nel modo seguente:

$$Li(\text{dB}) = 10 \log i/i_0 = 20 \log p/p_0$$

Potenza Acustica

- La potenza erogata da una sorgente acustica è espressa in watt, ma ha senso esprimere tale valore relativamente ad una potenza di riferimento che è lo stesso valore utilizzato per l'intensità. Tale valore risulta essere:

$$w_0 = 10^{-12} \text{ watt}$$

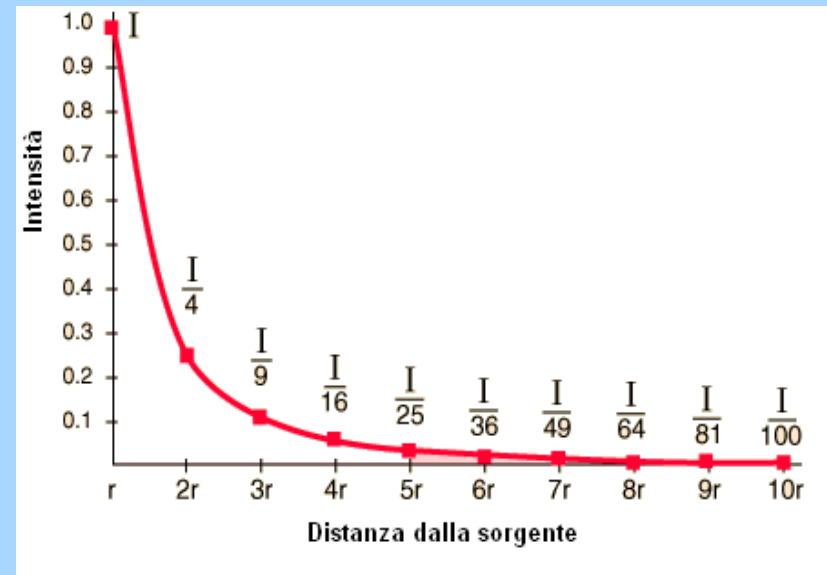
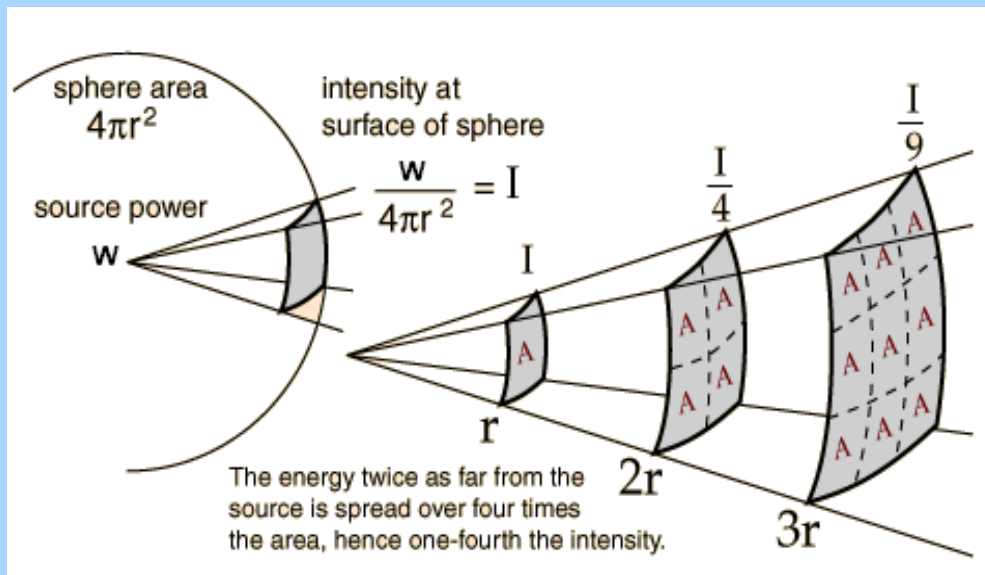
- La potenza (espressa in dB è definita livello di potenza sonora) è quindi data da:

$$Lw(\text{dB}) = 10 \log w/w_0$$

- Alla luce delle definizioni date, la potenza sonora può essere definita come l'intensità sonora che perviene su una sfera di superficie 1 m² con centro sulla sorgente

Propagazione sferica

- L'intensità sonora ad una certa distanza dalla sorgente obbedisce alla legge dell'inverso del quadrato



Livelli sonori tipici

Tabella 2.2 Livelli di suono in decibel (dB), rilevati da un fonometro.

<i>Suono al limite della percettibilità</i>	0 dB
Zanzara vicino all'orecchio	10
Fruscio di foglie	10
Bisbiglio (a 1 m)	15
Teatro o chiesa (vuoti)	25-30
Rumore di fondo notturno in città	30
Ufficio o ristorante (quieti)	50-55
Stadio	55
Conversazione (a 1 m)	50
Ufficio o ristorante (affollati)	60-65
Traffico cittadino diurno	70-80
Martello pneumatico (a 3 m)	90
Fortissimo (<i>fff</i>) di grande orchestra in sala	100
Urlo (a 1,5 m)	100
Gruppo rock in un locale chiuso	110
Schianto di fulmine	110
Martello su acciaio (a 0,5 m)	115
<i>Suono al limite del dolore</i>	120
Jet al decollo (a 50 m)	130
Rottura del timpano	160
Missile al decollo (a 50 m)	200
Massimo rumore prodotto in laboratorio	210

Esempi

- Trovare il livello di pressione sonora relativo ad una pressione sonora di 0.1 Pascal:

$$L_p(\text{dB}) = 20 \log p/p_0 = 20 \log (0.1 / 2 \times 10^{-5}) = \\ 20 \log (0.1 / 0,00002) = 20 \log 5000 = 73.9 \text{ dB} \approx 74 \text{ dB}$$

- Trovare il valore di intensità sonora relativa ad un livello d'intensità di 74 dB:

Tenuto conto che $L_i(\text{dB}) = 10 \log i/i_0$ si avrà:

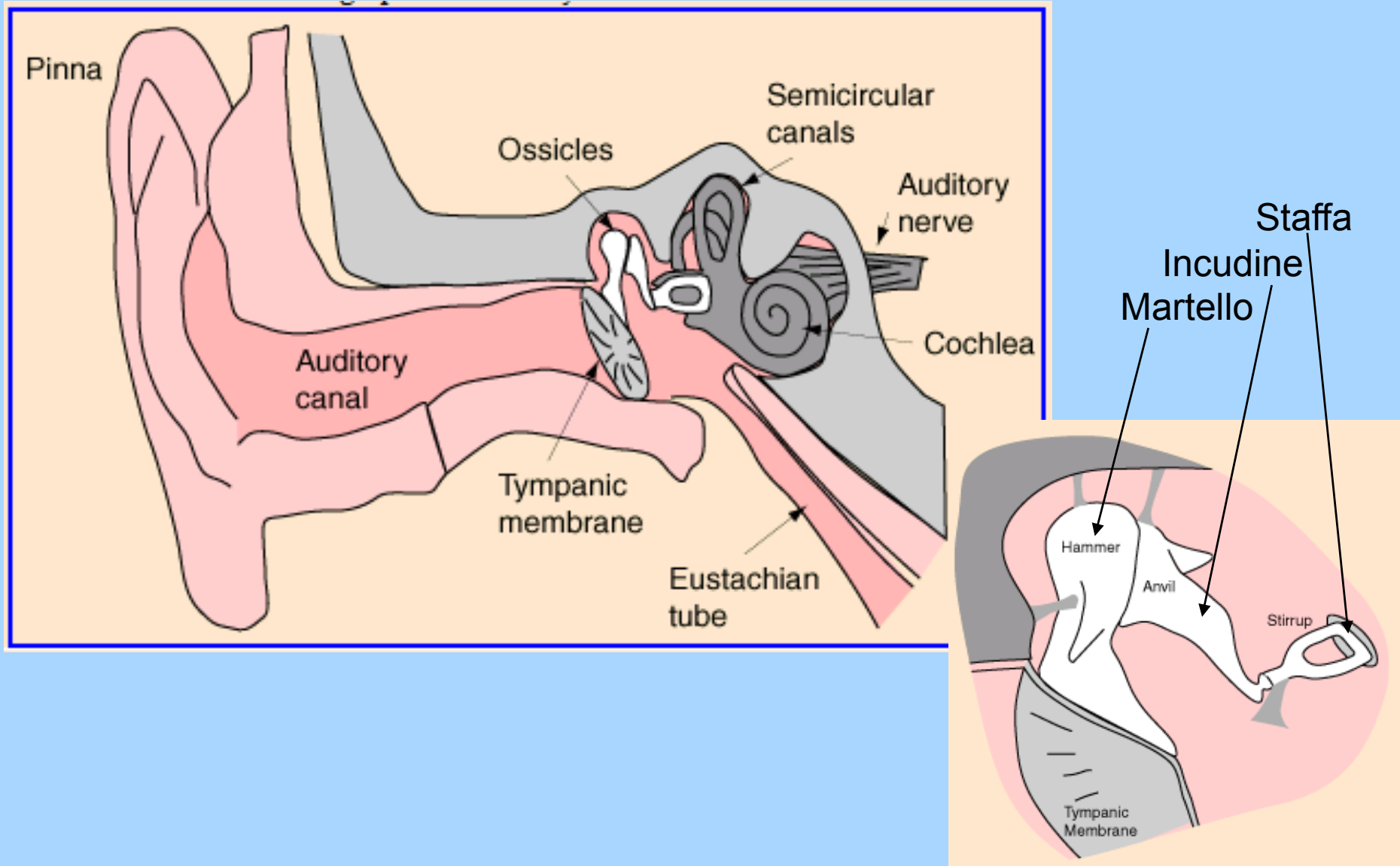
$$\log i/i_0 = L_i/10$$

$$\log i/i_0 = 7.4$$

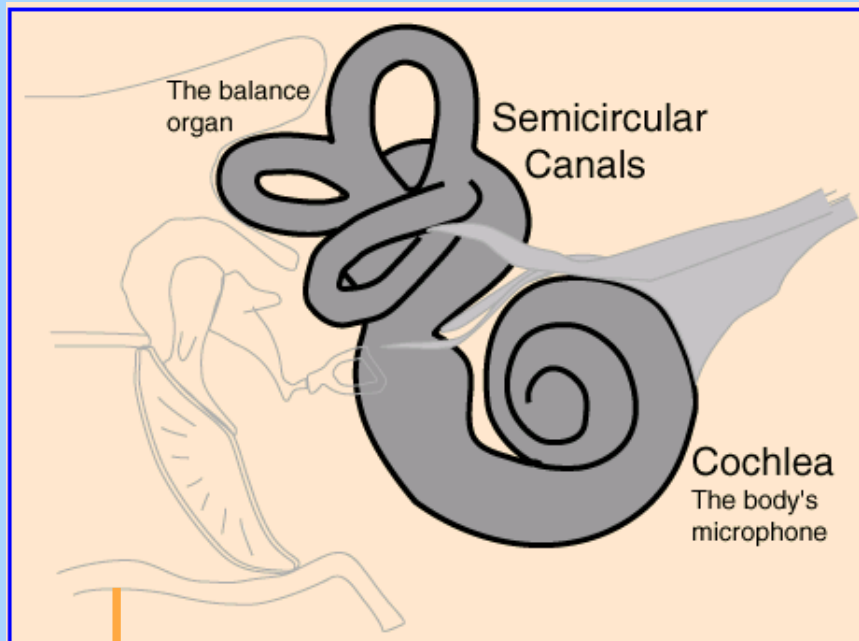
$$i/i_0 = 10^{7.4}$$

$$i = i_0 \times 10^{7.4} = 10^{-12} \times 10^{7.4} = 2.5 \times 10^{-5} \text{ watt/m}^2 \approx 25 \text{ } \mu\text{watt}$$

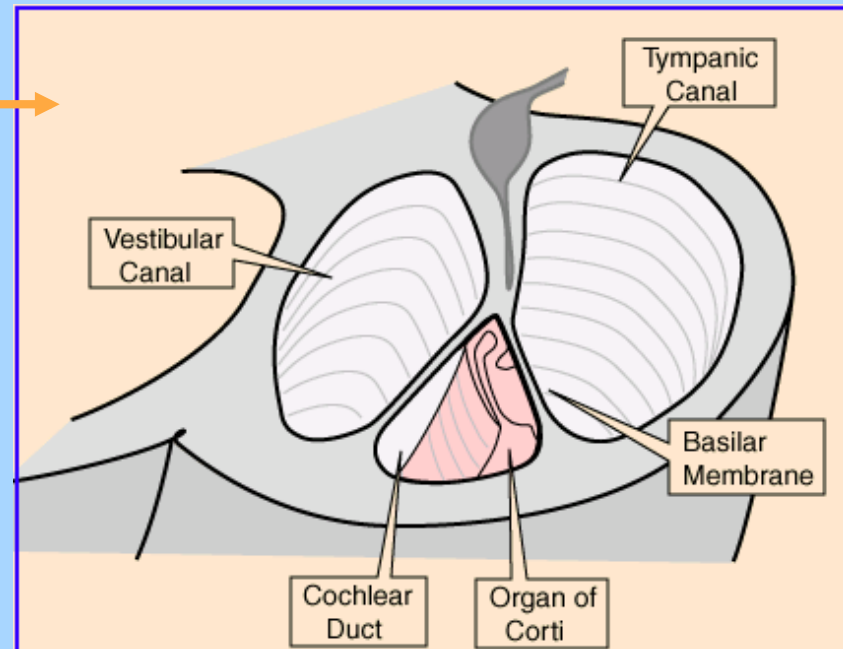
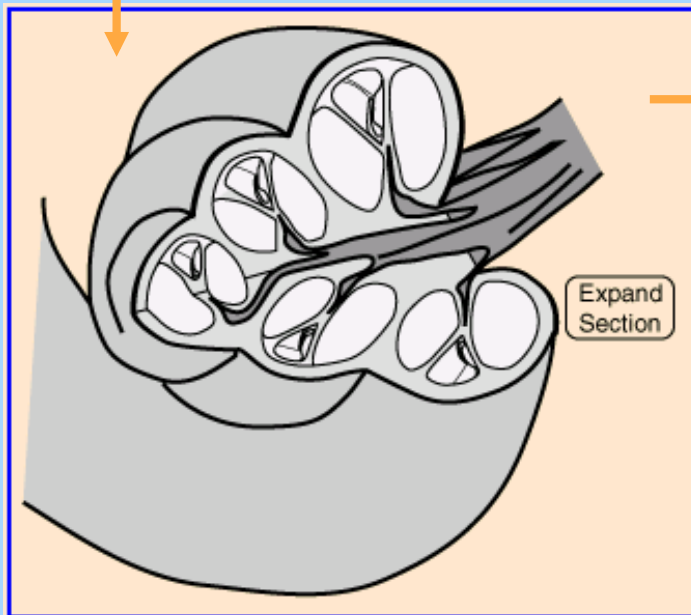
Il suono come fenomeno uditivo



Orecchio Interno

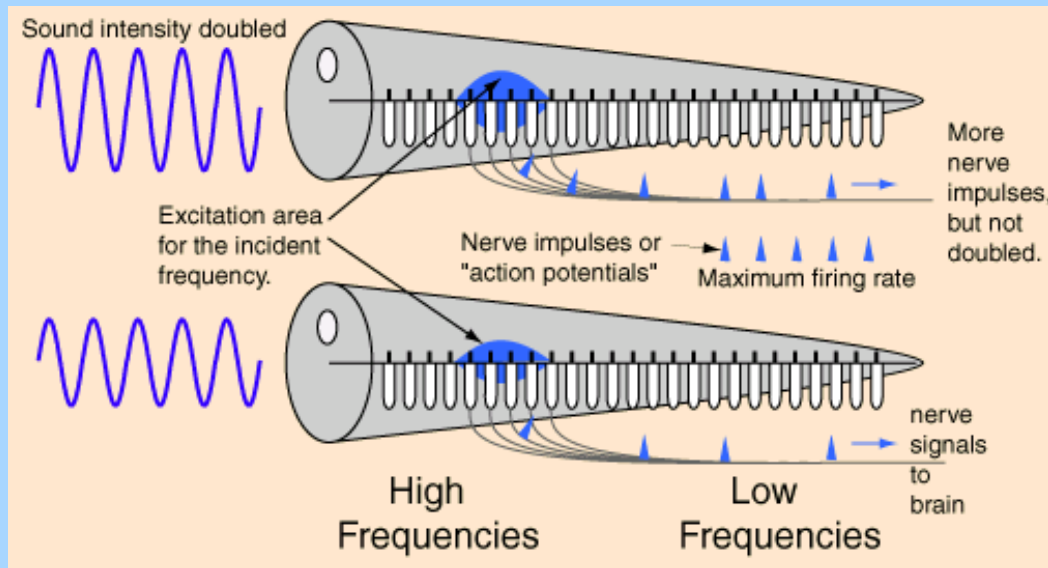


La struttura dell'orecchio interno denominata **coclea** ha la forma di una chiocciola ritorta divisa internamente in tre parti riempite di fluido. Due sono canali per la trasmissione di pressione e nel terzo è l'organo sensibile di **Corti**, che rileva gli impulsi di pressione e risponde con gli impulsi elettrici che viaggiano lungo il nervo uditivo al cervello.



Sensibilità uditiva

Schematizzazione della Coclea (srotolata)

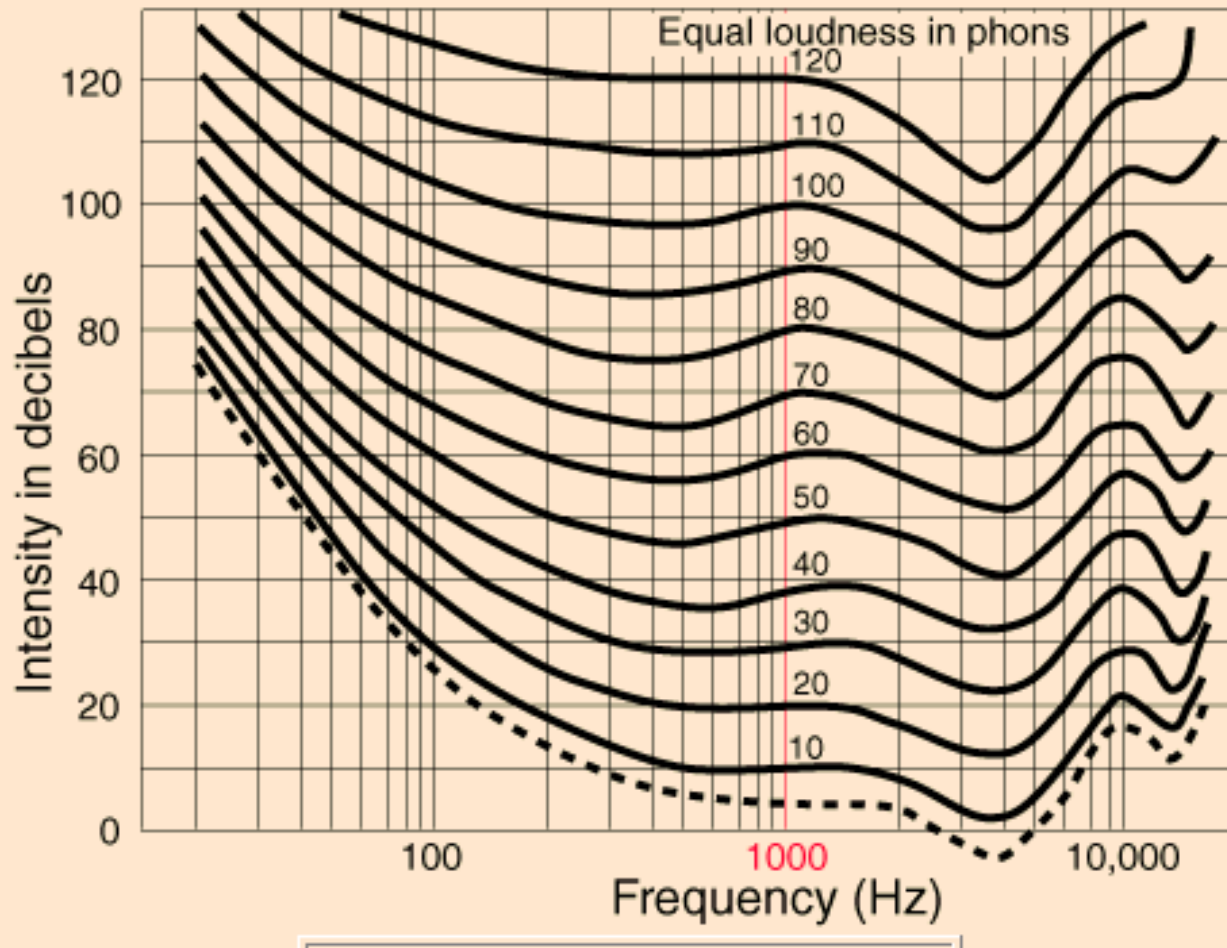


Una regola pratica ampiamente usata per la sonorità di un suono particolare è che il suono deve essere aumentato di intensità da un fattore dieci per percepire un'intensità doppia.

Due suoni hanno uguale l'intensità non hanno stesso **"loudness"** (sonorità) . Poiché la sensibilità uditiva umana varia con la frequenza, è utile tracciare le curve di uguale sonorità (**equal loudness curves**) che mostrano la variazione per l'orecchio umano medio.

Curve ad uguale sonorità (Curve di Fletcher)

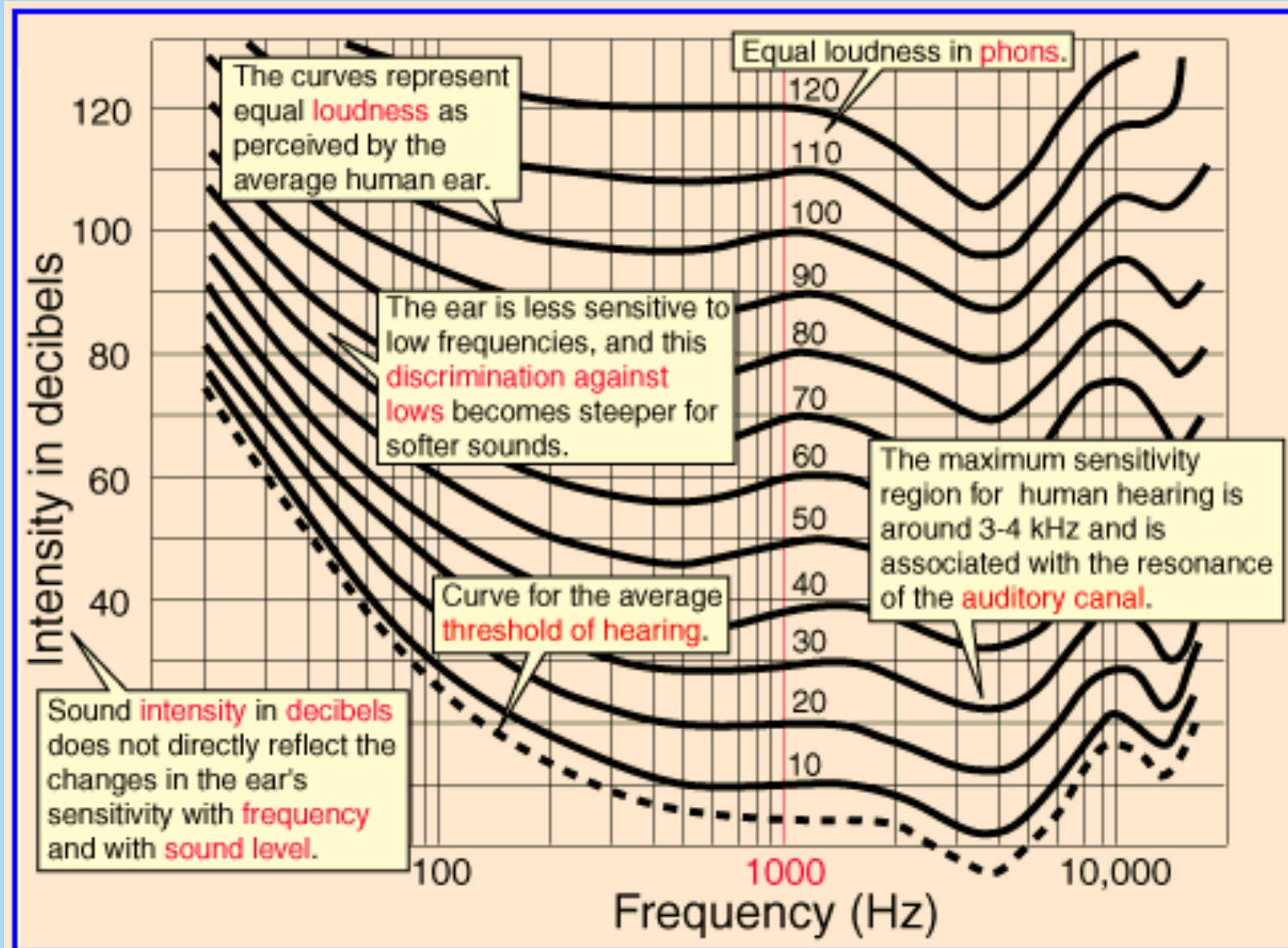
Equal Loudness Curves



Due suoni diversi di 60 dB in generale non avranno la stessa sonorità. Vengono introdotti i Phons.

60 Phons equivalgono a 60 dB, ma solo a 1000 Hz.

Significato delle curve di Fletcher

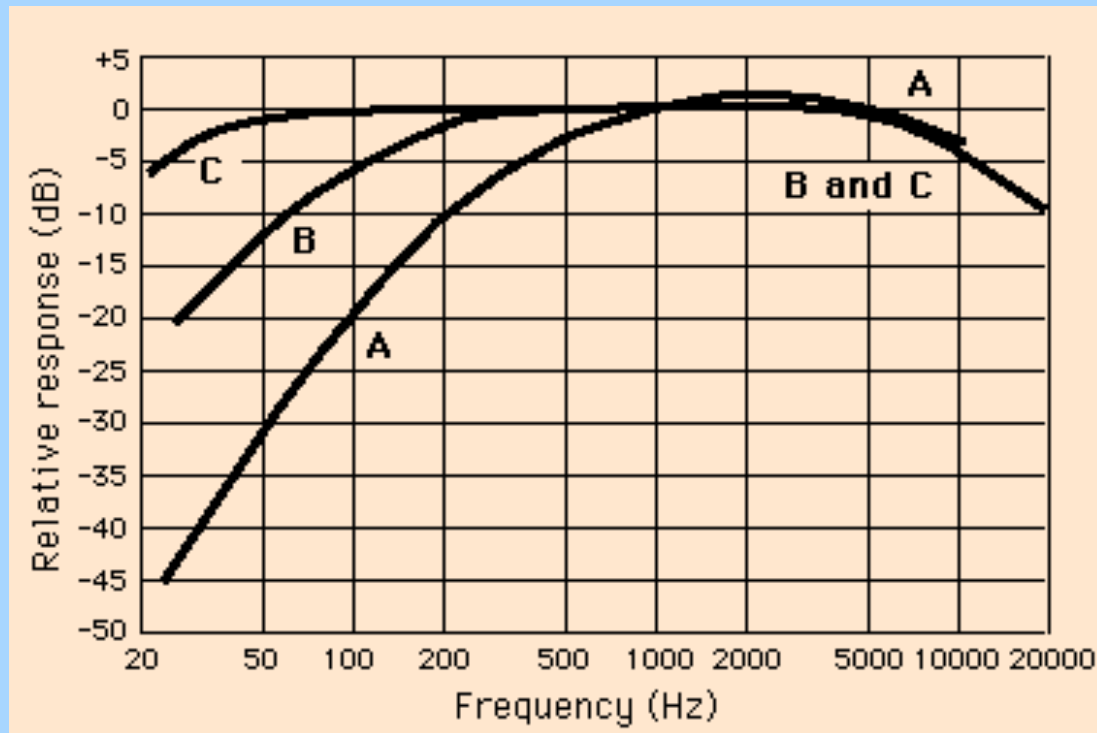


Dai Phones ai Sones

L'uso dei **Phones** come unità della sonorità è un miglioramento rispetto ai decibel, ma non è ancora una misura che sia direttamente proporzionale alla sonorità. Usando la regola pratica per la sonorità, vien introdotto il **Sone**. La scala dei sones è stata generata per fornire così scala lineare della sonorità. È presunto solitamente che la gamma standard per musica orchestrale sia di circa 40 - 100 Phones. Se all'estremità più inferiore di quella gamma è assegnata arbitrariamente una sonorità di **1 sone**, allora 50 phones avrebbero una sonorità di 2 sones, 60 phones sarebbero 4 sones, ecc.

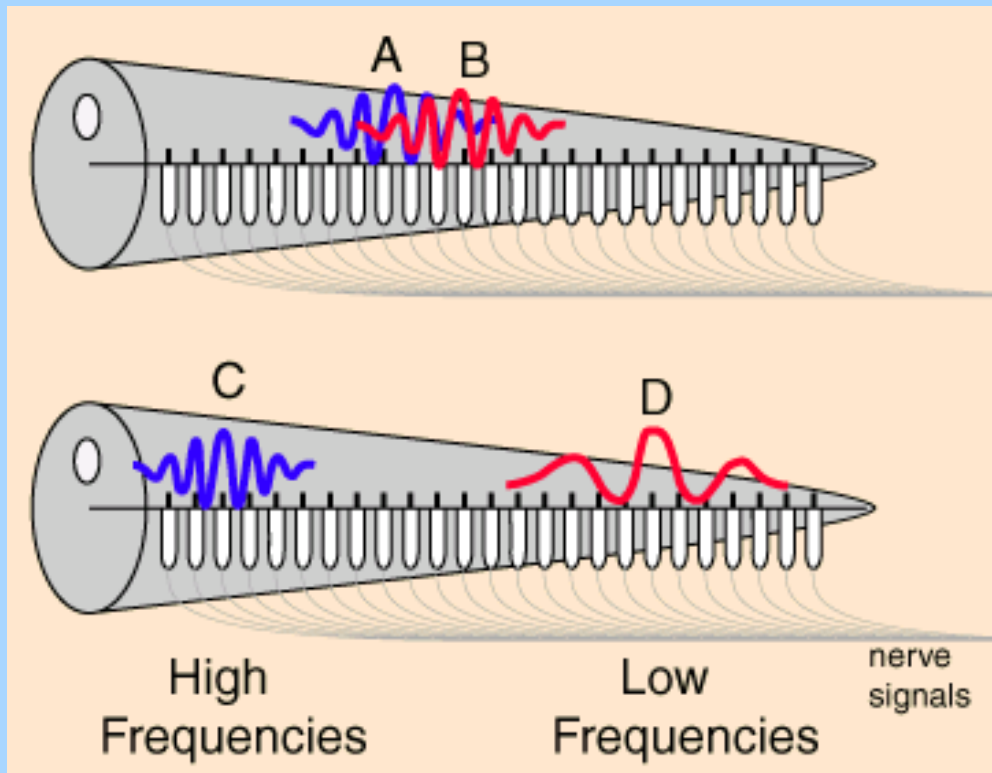
Dynamic Level	Phons	Sones
fff	100	64
...	90	32
f	80	16
...	70	8
p	60	4
...	50	2
ppp	40	1

Filtri di pesatura



Profili standard di filtro usati per rendere la misura strumentale dell'intensità vicina al comportamento dell'orecchio umano medio. I tre profili differenti sono per abbinare l'orecchio alle intensità di suono differenti (curva A,B,C). In genere quella più usata è la A. Si parla allora di dBA

Più suoni contemporaneamente



Quando un suono si somma ad un altro suono, l'aumento nella sonorità percepita dipende dal relativo rapporto di frequenza al primo suono. La visione di questo processo può essere ottenuta dalla **teoria posizionale** della percezione. Se il secondo suono ampiamente è separato nell'altezza dal primo, quindi non competono per gli stessi terminali nervosi sulla membrana basilare dell'orecchio interno. L'aggiunta del secondo suono di sonorità uguale rende circa due volte più intenso il suono totale. Ma se i due suoni sono vicini in frequenza, all'interno di una **banda critica**, allora gli effetti di saturazione nell'organo di Corti sono tali che la sonorità totale percepita è soltanto un po' più grande di uno o l'altro suono da solo.

JND (Just Noticeable Difference) o soglia percettiva di ampiezza

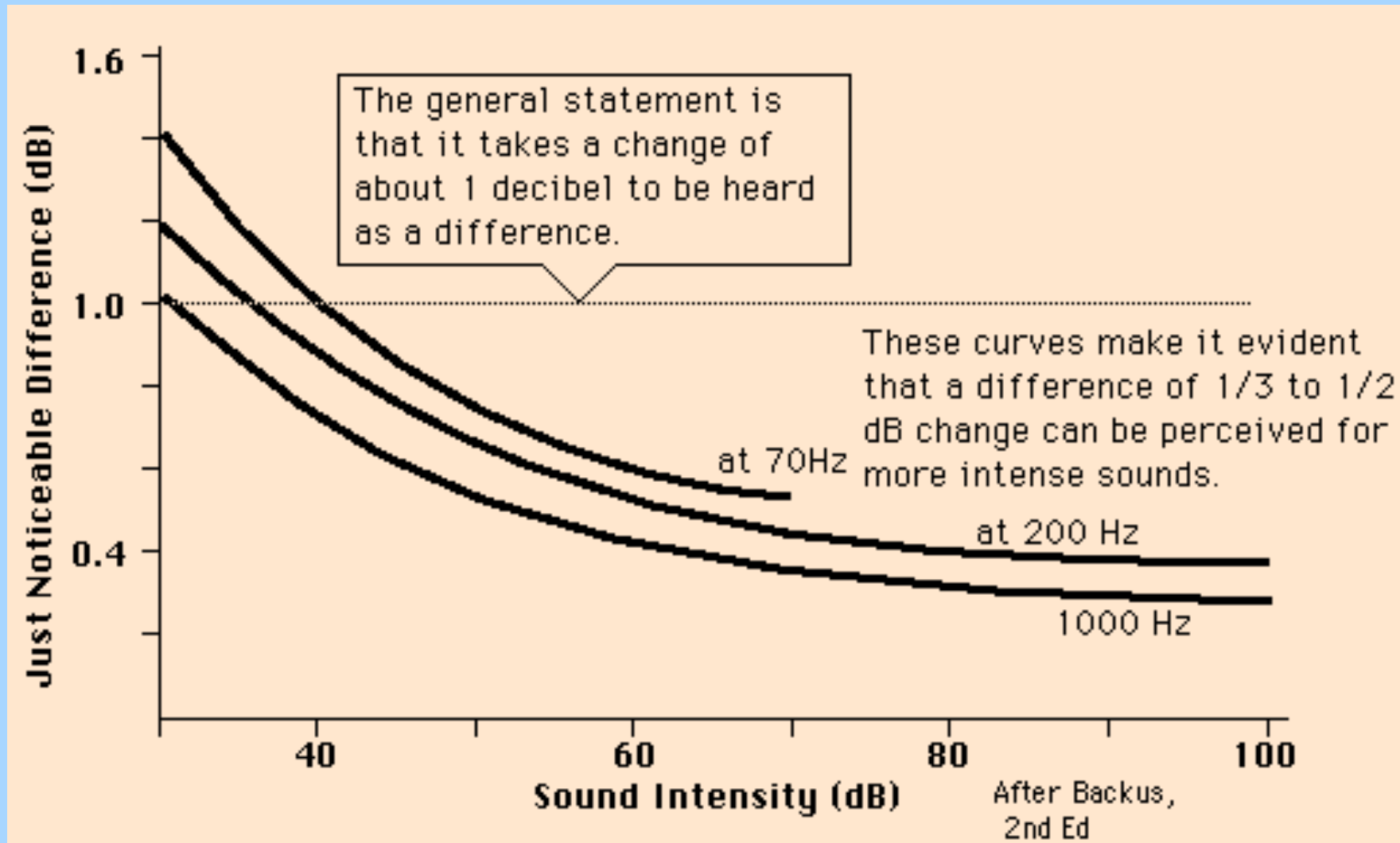
Un riferimento generale utile è che la soglia percettiva (JND) d'intensità del suono per l'orecchio umano è di circa 1 decibel.

JND = 1 decibel

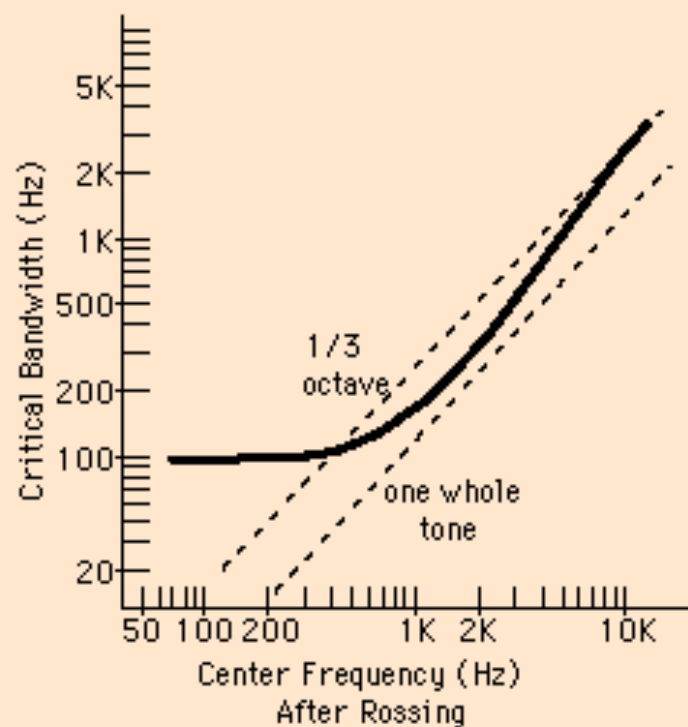
Infatti, l'uso del fattore di 10 nella definizione del decibel deve generare un'unità che è circa il minimo cambiamento rilevabile nell'intensità del suono.

Vi sono alcune variazioni al proposito: il jnd è circa 1 dB per i suoni deboli intorno ai dB 30-40 alle basse e medie frequenze. Può ridursi a 1/3 - 1/2 di decibel per i suoni forti.

Variazioni nel JND



Bande critiche



Center Freq (Hz)	Critical bandwidth (Hz)
100	90
200	90
500	110
1000	150
2000	280
5000	700
10000	1200

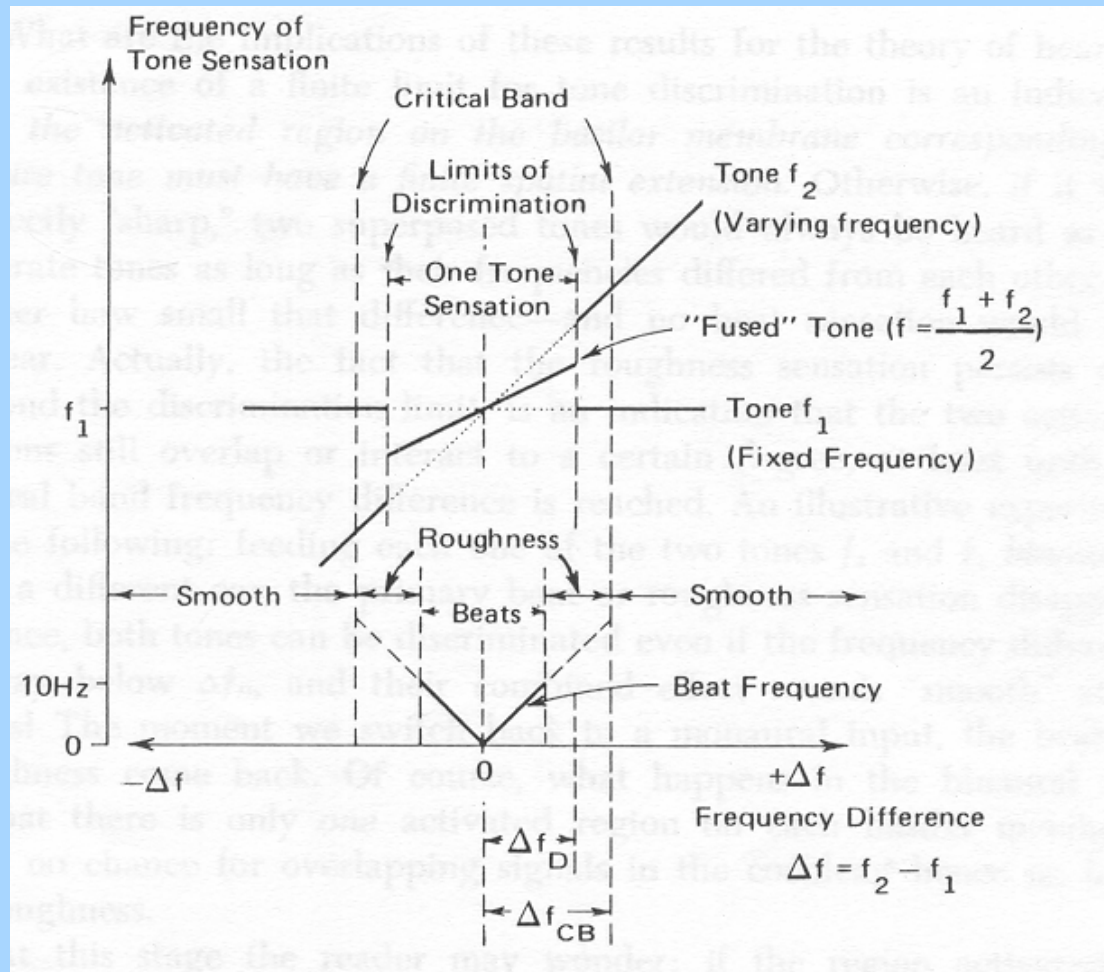
Rossing 2nd Ed p74

Per le frequenze gravi la **banda critica** è larga circa 90 hertz. Per le più alte frequenze, è fra un tono intero e 1/3 di ottava corrispondente circa ad un terza minore.

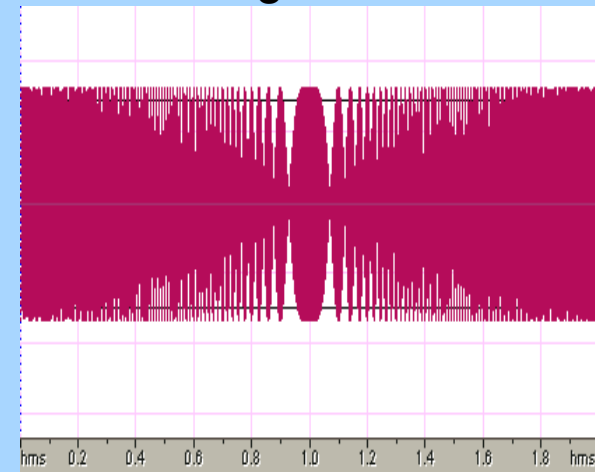
Banda Critica e Battimenti

Questo schema è la rappresentazione della sensazione uditiva prodotta dalla sovrapposizione di due suoni puri (sinusoidali) di frequenza quasi uguale

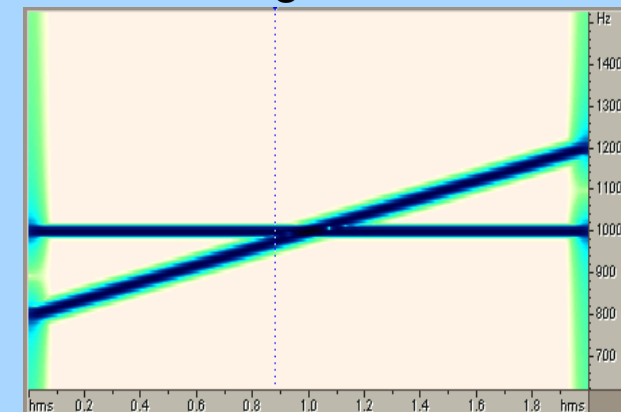
f_1 e $f_2 = f_1 + \Delta f$



oscillogramma

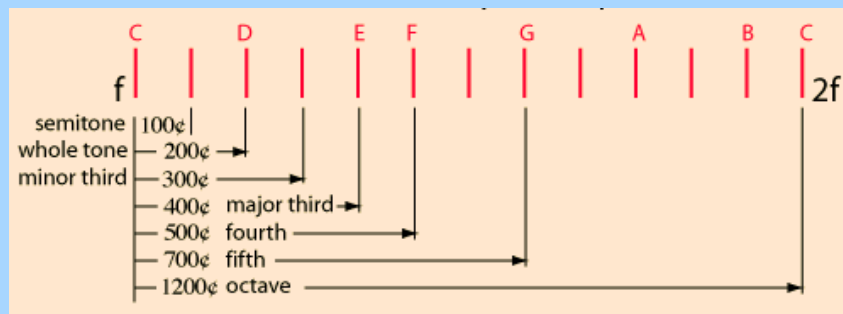


sonogramma



JND in Frequenza (soglia percettiva di frequenza)

La JND di altezza si deve esprimere come un rapporto o intervallo musicale dall'orecchio umano tenda a rispondere ugualmente ad uguali rapporti di frequenza. È conveniente esprimere la soglia frequenziale in **cents** (1 Cent = 1/100 di semitono temperato). Il valore usuale de JND frequenziale è di cinque centes. Uno dei vantaggi della notazione in cents sta nel fatto che esprime lo stesso intervallo musicale, senza distinzioni della zona di frequenza.



Divisione d'ottava nel temperamento equabile

Calcolando gli intervalli con i Cents

La relazione principale è data dalla seguente formula che esprime il rapporto d'ottava in cents in relazione alle frequenze:

$$\frac{f_2}{f_1} = 2^{\frac{\mathcal{C}}{1200}}$$

Ciò è conveniente per calcolare la frequenza corrispondente ad un certo numero di cents. Per calcolare il numero di cents a partire da due frequenze, la relazione sopra si deve invertire. Applicando il logaritmo naturale ad ambo i membri si ottiene:

$$\ln\left[\frac{f_2}{f_1}\right] = \frac{\mathcal{C}}{1200} \ln 2 \quad \text{or} \quad \mathcal{C} = 1200 \frac{\ln\left[\frac{f_2}{f_1}\right]}{\ln 2}$$

Vantaggi nell'uso dei Cents

Esaminando il semitono da A a Bb in punti diversi nell'estensione del pianoforte testimonia il fatto che se espresso in cents, ogni semitono temperato è lo stesso. Espresso in Hz differenzi, ogni semitono è diverso. Il valore di intervallo nei cents esprime il rapporto delle frequenze, che è lo stesso per ogni semitono temperato.

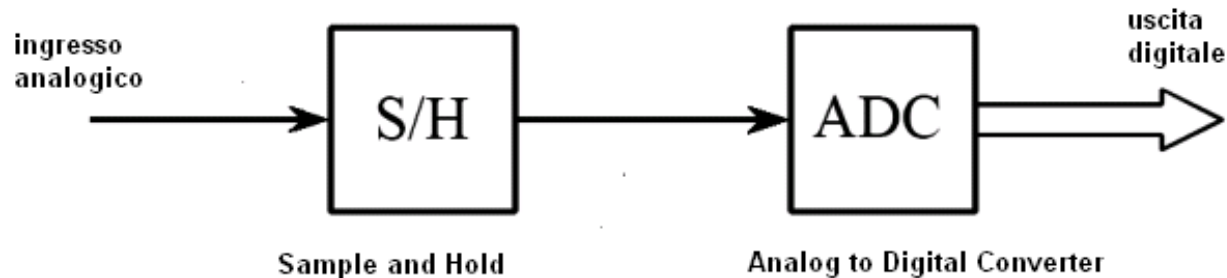
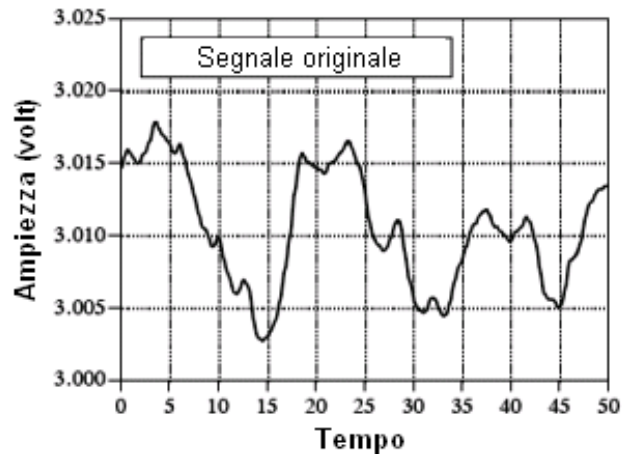
B_{b0}	29.135 Hz	B_{b4}	466.16 Hz	B_{b7}	3729.3 Hz
A_0	27.5 Hz	A_4	440.0 Hz	A_7	3520 Hz
Δf in Hz	1.635 Hz	26.16 Hz	209.3 Hz		
Interval	100¢	100¢	100¢		
JND = 5¢	→ 0.08 Hz	5¢	→ 1.27 Hz	5¢	→ 10 Hz

Si noti come il JND è sempre minore degli intervalli di semitono e che uno spostamento di 5 cents sposta meno di 1/10 di Hz alle note gravi e circa 10Hz nella zona più acuta della tastiera.

CAPITOLO 2

- Il segnale digitale e i suoi supporti
- Analisi, sintesi ed elaborazione del suono
- Linguaggi per l'analisi, la sintesi e l'elaborazione del suono

Campionamento..

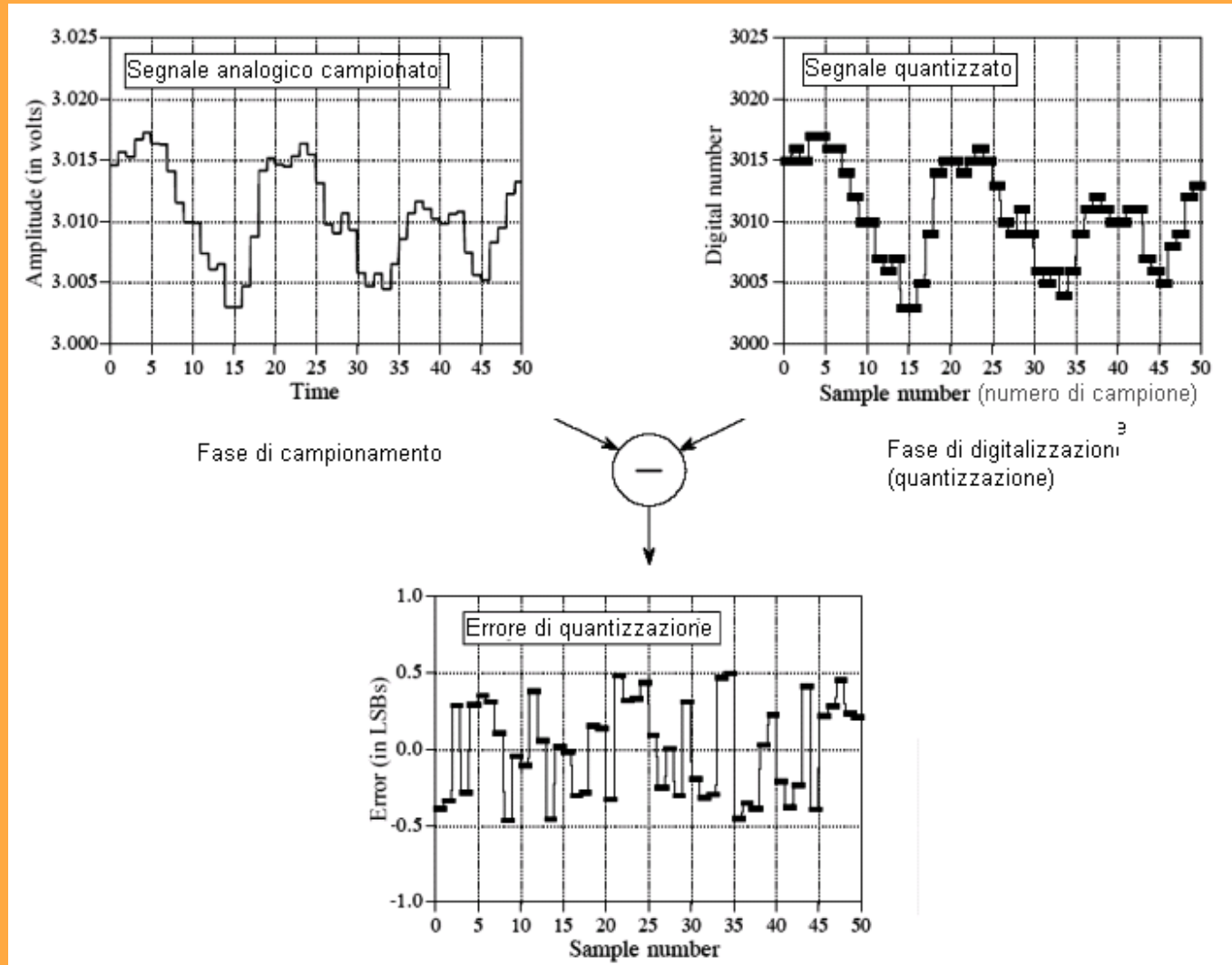


Nello schema il processo di conversione analogico/digitale. Sono evidenziate le due fasi costitutive del processo:

- **campionamento** (discretizzazione temporale)
- **quantizzazione** (discretizzazione di ampiezza)

In ogni processo di conversione viene commesso in misura variabile un certo numero di errori numerici poiché l'ampiezza è rappresentata da un numero finito di valori numerici. In ogni caso l'errore può essere mantenuto sufficientemente basso.

..Campionamento



Nella fase di campionamento vengono prelevati valori d'ampiezza ad una velocità definita **FREQUENZA DI CAMPIONAMENTO** (*Sampling Rate*) espressa in Hz.

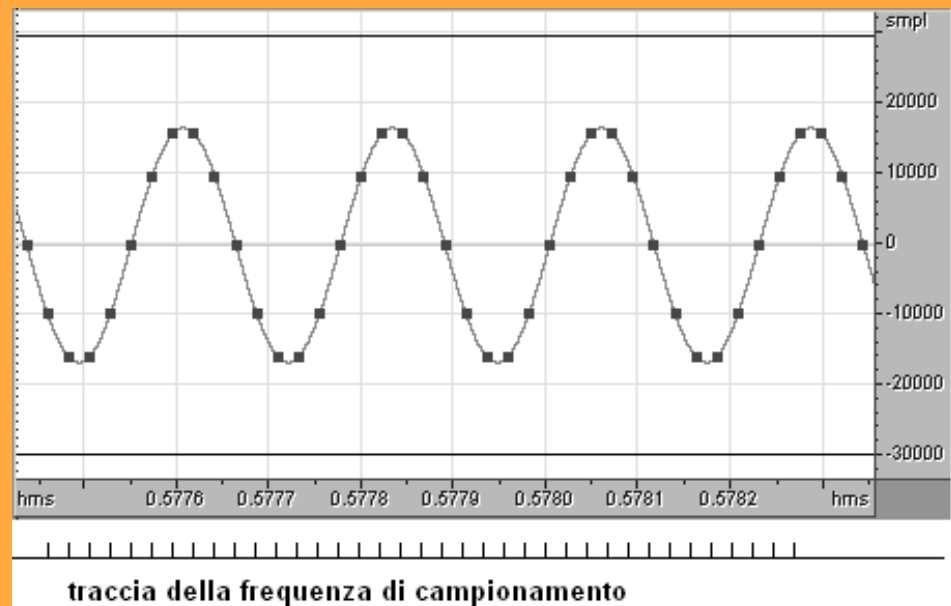
Teorema di campionamento

Il Teorema del Campionamento (detto anche di Shannon) può essere sinteticamente espresso nel modo seguente:

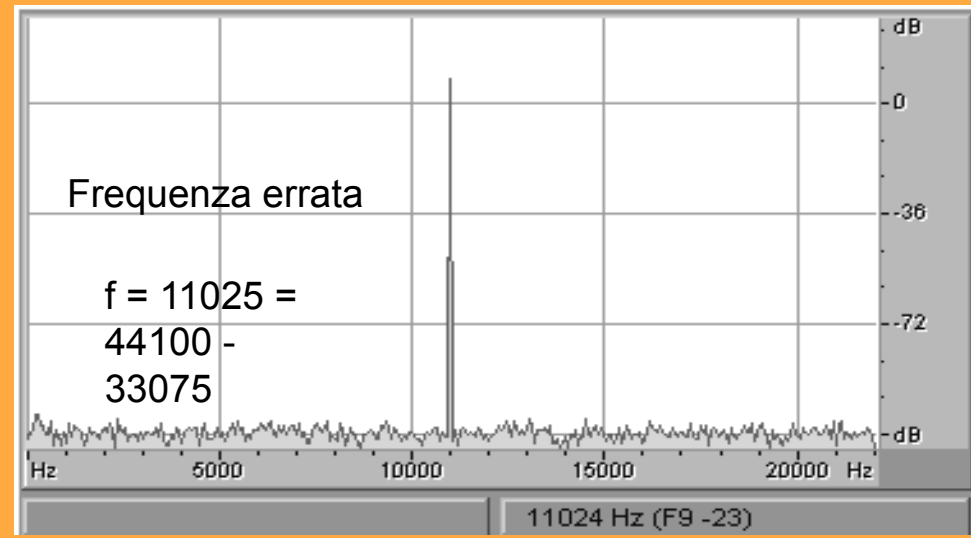
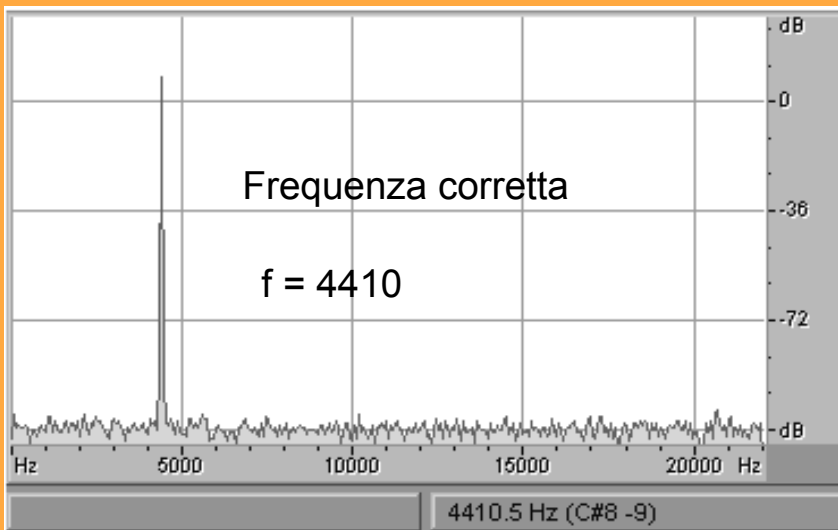
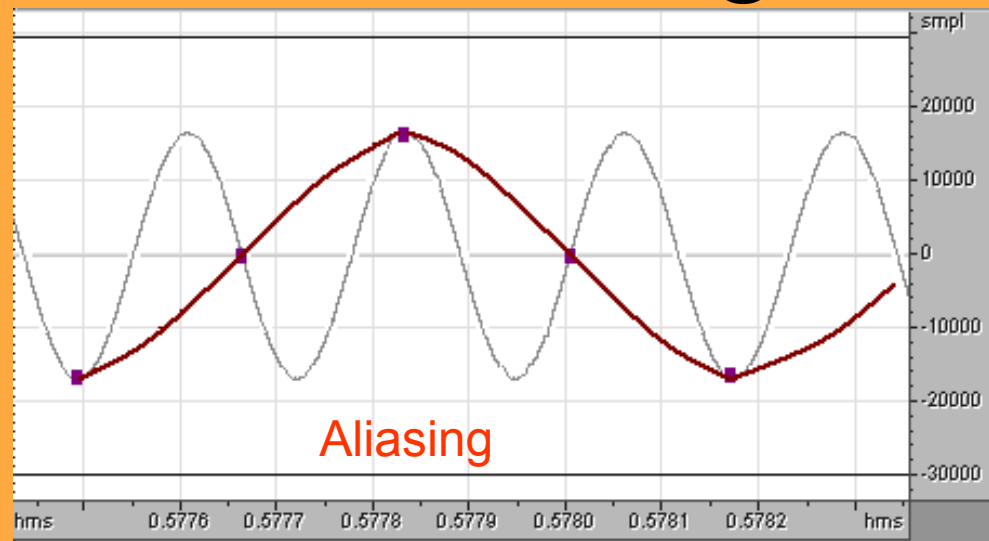
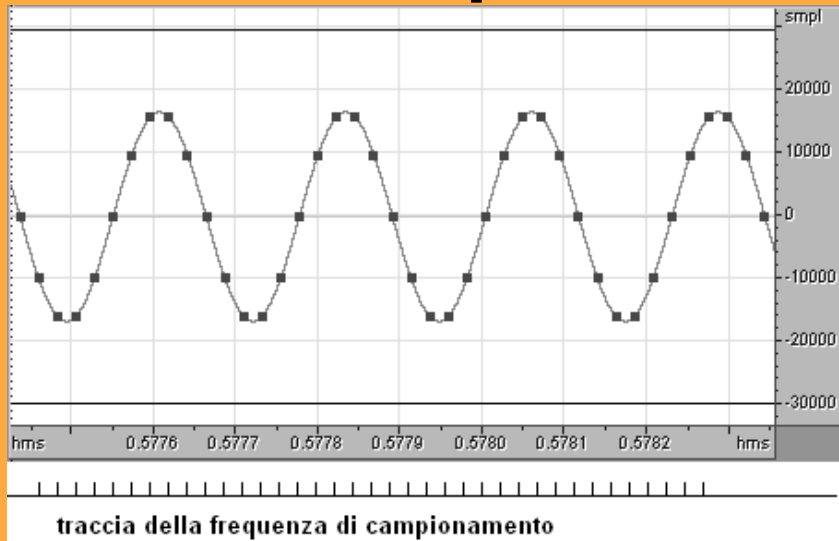
$$SR \geq 2 F_{\max}$$

dove F_{\max} = frequenza più acuta contenuta nel segnale da campionare. Ciò significa che nel caso del segnale audio, essendo l'estensione in frequenza da 20 a 20000 Hz, occorrerà una frequenza di campionamento (SR) maggiore di 40000 Hz. E' in pratica ciò che avviene nella realizzazione e riproduzione di CD audio per i quali la $SR = 44100$ (cioè un valore paria più del doppio della massima frequenza presente)

Dato un certo valore di sr , è possibile convertire segnali con frequenza massima $sr/2$. Tale frequenza limite viene chiamata **frequenza di Nyquist**



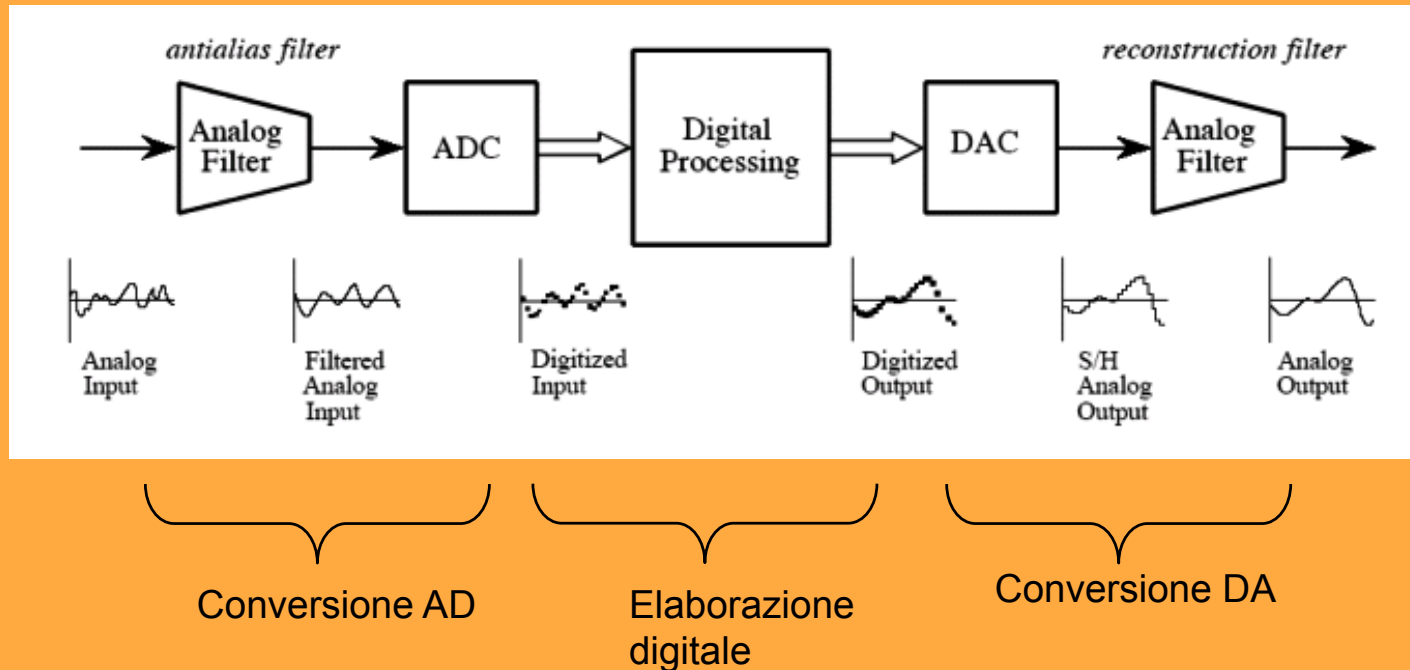
Campionamento e Aliasing



sr = 44100 f = 4410 10 samps/periodo

sr = 44100 f = 33075 1.33 samps/periodo

Conversione AD e DA

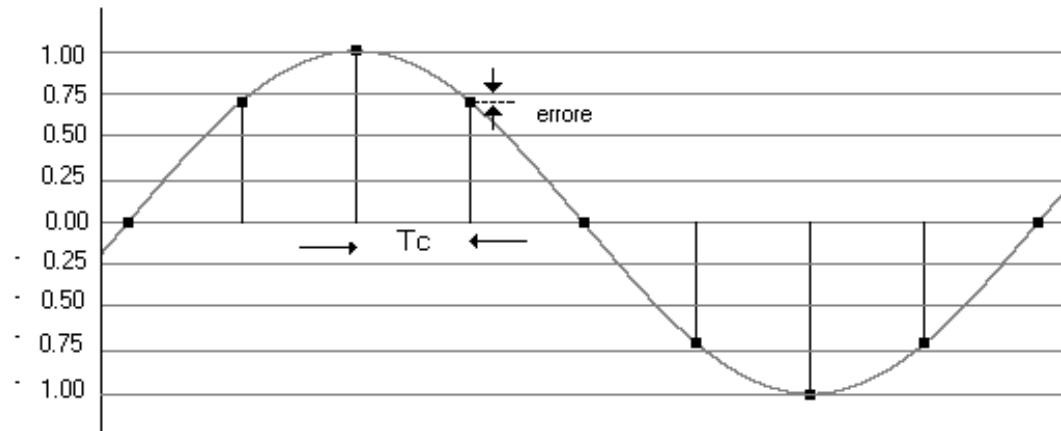


- Per garantire la conoscenza della massima frequenza in ingresso si antepone un **filtro anti alias**
- Per riportare il segnale numerico (digitale) in forma continua si pospone un **filtro di ricostruzione**

ADC = Analog to Digital Converter

DAC = Digital to Analog Converter

Quantizzazione



L'errore che sempre si commette quando si rappresenta l'ampiezza di un segnale campionato attraverso un numero finito di livelli, è detto **errore di quantizzazione**. L'effetto di tale errore produce un segnale indesiderato che prende il nome di **rumore di quantizzazione**. L'errore si può ridurre al minimo aumentando considerevolmente il numero di livelli.

Aritmetica dei segnali digitali

Nei sistemi di conversione, ogni livello viene codificato con una **parola binaria** (numero binario) costituita da M cifre binarie (**bit**) per un numero totale di livelli pari a 2^M . Ad esempio, impiegando una parola di 8 bit, si possono codificare $2^8 = 256$ diversi livelli. In questo caso, l'intera gamma dinamica del segnale d'ingresso è espressa su 256 valori, con una risoluzione d'ampiezza pari a $1/256$, cioè circa 0.0039. Ogni aumento di un bit nella codifica della parola binaria determina un raddoppio nel numero di livelli rappresentabili.

1 1 0 1 1 0 0 1

Numero binario a 8 bit (cifre). Ogni cifra ha un peso crescente da destra verso sinistra secondo la relazione:

2^N dove N = numero di cifre binarie

$$1 \times 2^7 + 1 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 217 \text{ (decimale)}$$

Ogni gruppo di 8 bit è indicato con il nome di BYTE

Qualità della quantizzazione

Per fornire un criterio di valutazione della qualità di un sistema di conversione è uso comune riferirsi alla cosiddetta *gamma dinamica*. Essa esprime il rapporto tra il segnale ed il rumore (**SNR o Signal-to-Noise Ratio**) quando il rumore è additivo e stazionario. Per questo parametro vale la relazione:

$$\text{SNR(dB)} = 6.02 N + 1.76$$

dove M è il numero di bit impiegati nella codifica dei livelli. Il fattore 6.02, deriva dal fatto che ogni raddoppio del numero di livelli provoca un raddoppio della massima ampiezza rappresentabile. (Risulta infatti che $20\log(2) = 6.02$ dB).

L'impiego di un maggior numero di livelli di quantizzazione consente di aumentare la dinamica totale e allo stesso tempo migliorare la risoluzione del sistema.

La dinamica di un'orchestra misurata all'interno di uno studio di registrazione può raggiungere valori prossimi a 120 dB; considerando che il livello di rumore di fondo può essere circa 20 dB, l'effettiva gamma dinamica può essere stimata circa **100 dB**. Da questo esempio si intuisce che la registrazione e la riproduzione digitale di segnali musicali necessita elevati valori di dinamica, ottenibili solamente impiegando convertitori il cui numero di bit consenta di raggiungere tali valori.

Gamma dinamica e numero di bit

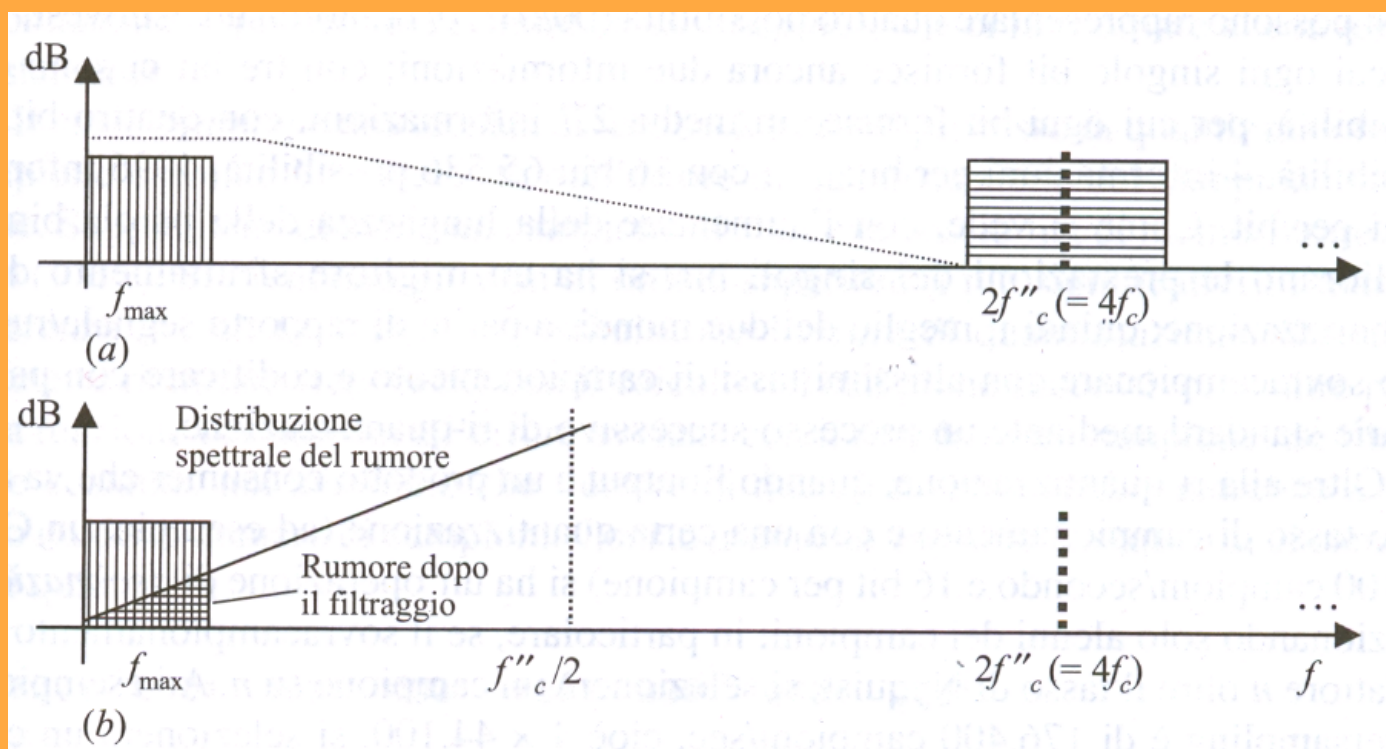
Gamma dinamica in funzione del numero di bit M nel convertitore

N (numero.bit)	SNR (dB)
8	49.92
9	55.94
10	61.96
11	67.98
12	74.00
13	80.02
14	86.04
15	92.06
16	98.08
17	104.10

Dalla tabella 1.3 si evince che, impiegando una codifica PCM uniforme, per garantire un'efficace gamma dinamica del sistema di conversione, occorre che il convertitore A/D impieghi una parola binaria di almeno 16 bit ed allo stato attuale ciò rappresenta una specifica standard nella codifica dei dati per il *compact disc*.

Sovracampionamento

Il sovracampionamento (*oversampling*) utilizza una sr molto maggiore di quella necessaria per soddisfare il teorema di Nyquist. Nella figura è mostrato lo spettro del segnale campionato a 4 volte (4x). Si noti (a) il filtro di ricostruzione low pass molto blando. Inoltre possono essere utilizzati convertitori istantanei (*flash*) senza S/H e in questo caso il rumore si distribuisce prevalentemente alle frequenze elevate, lontano dalla gamma audio



Noise Shaping e Codifiche Differenziali

Il **Noise Shaping** (modellazione di rumore) è una tecnica di campionamento in cui l'errore di campionamento del campione attuale è influenzato dall'errore del campione precedente. Questa tecnica sposta ancora più in alto lo spettro del rumore di quantizzazione.

Le tecniche di sovracampionamento prevedono anche l'impiego di conversioni di tipo differenziale (DPCM) che valutano le differenze dei valori piuttosto che il loro valore assoluto (PCM):

PCM Differenziale

DPCM (differenziale – bit multipli)

Δ Modulation (differenziale - 1 bit)

PCM standard

Σ -DPCM (integrazione – bit multipli)

$\Sigma - \Delta$ Modulation (integrazione – 1 bit) (Bitstream – Philips)

Questa ultima codifica si arriva a frequenze di sampling di 11 MHz.

Frequenze di Campionamento

Frequenza di campionamento	Utilizzo
8,000 Hz	telefono e comunicazioni senza fili walkie-talkie
11,025 Hz	un quarto della frequenza di campionamento degli audio CD; utilizzata per audio a bassa qualità PCM, MPEG e per analisi audio della banda passante per i subwoofer
22,050 Hz	La metà della frequenza di campionamento degli audio CD; utilizzata per audio a bassa qualità PCM, MPEG e per l'analisi dell'energia alle basse frequenze. Utilizzabile per digitalizzare i precedenti formati audio analogici del ventesimo secolo come ad esempio i 78 giri.
32,000 Hz	miniDV digital video camcorder , nastri video digitali con canali extra per le informazioni audio (es. DVCAM con 4 Canali per l'Audio), DAT , Digitales Satellitenradio tedesca, compressione audio digitale NICAM , utilizzata per la televisione analogica in alcuni stati. Radiomicrofoni di alta qualità.
44,100 Hz	audio CD , utilizzata spesso con la compressione audio MPEG-1 per la creazione di (VCD , SVCD , e MP3). Molte apparecchiature professionali utilizzano (o hanno presente come opzione) la frequenza di 44.1 kHz, inclusi i mixer, Equalizzatori, compressori, reverberi, crossover, registratori e radiomicrofoni professionali.
47,250 Hz	Frequenza del primo registratore commerciale PCM della Denon
48,000 Hz	Utilizzata per miniDV , digitale terrestre , DVD , e per la produzione di film .
50,000 Hz	Frequenza dei primi registratori commerciali degli ultimi anni 70 da 3M a Soundstream
50,400 Hz	Frequenza di campionamento utilizzata dal registratore audio digitale Mitsubishi X-80 .
88,200 Hz	produzione di un audio CD (multipla di 44,100 Hz).
96,000 Hz	DVD-Audio , alcune tracce DVD LPCM , tracce audio del formato BD-ROM (Disco Blu-ray) e HD DVD (High-Definition DVD).
176,400 Hz	produzione di un audio CD (multipla di 44,100 Hz).
192,000 Hz	DVD-Audio , alcune tracce DVD LPCM , tracce audio del formato BD-ROM (Disco Blu-ray) e HD DVD (High-Definition DVD). Registratori e sistemi di editing in Alta definizione
2,822,400 Hz	SACD , 1-bit sigma-delta modulation processo di campionamento conosciuto come Direct Stream Digital , sviluppato da Sony in collaborazione con Philips

Formato dei dati audio

- Ogni campione viene rappresentato con un numero binario composto di N bit
- Possono essere usati numeri interi senza segno
- Interi con segno (tecnica del complemento a 2)
- Con più canali generalmente i dati si susseguono in modo alternato:
campione 1 sx campione 1dx campione2 sx campione2 dx ecc..

Un' altra distinzione è quella tra la rappresentazione Little-Endian e Big-Endian.
Per esempio un campione a 16 bit (ad esempio il numero 9) può essere rappresentato alternativamente come la combinazione di due byte:

00001001 00000000 Little-Endian
00000000 00001001 Big-Endian

Scambiando la rappresentazione si commettono errori colossali (es. da 9 a 2304 nel caso si interpreti il primo dei due numeri come BE)

Principali formati

- Ogni struttura dati che rappresenta informazioni audio digitali è organizzata come files numerici, ciascuno dei quali è preceduto da un'intestazione (Header) che ne caratterizza il contenuto.
- Esistono molti formati audio. I principali sono elencati nella seguente tabella

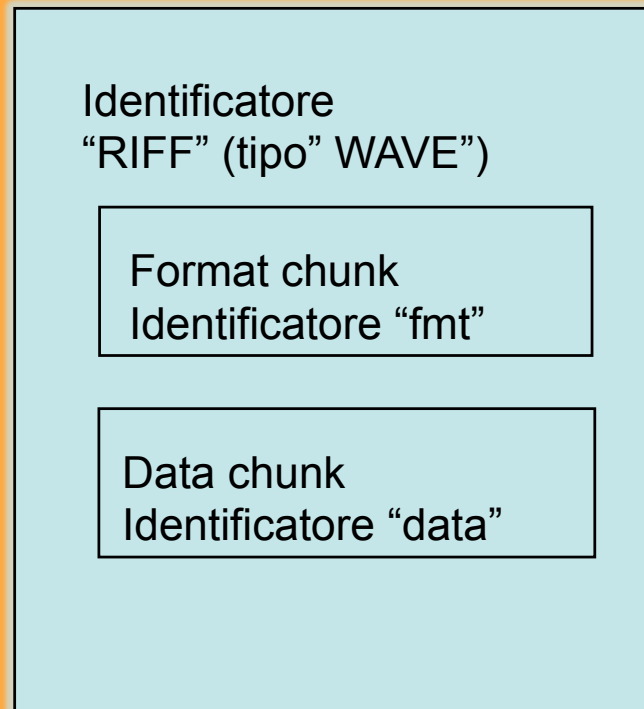
Tabella 3.1 Esempi di formati di file audio con intestazione

Estensione	Origine	Parametri
.au, .snd	NeXT, Sun	rate, #canali, codifica, info string
.aif(f), AIFF	Apple, SGI	rate, # canali, quantizz., molte info
.aif(f), AIFC	Apple, SGI	(AIFF con compressione)
.wav, WAVE	Microsoft	rate, #canali, quantizz., molte info
.voc	Creative	rate (8 bits/1 ch; silence deletion)
.iff, IFF/8SVX	Amiga	rate, # canali, info strumenti (8 bit)
.mp3, .mp4	MPEG	rate, #canali, qualità dei campioni
.sf	IRCAM	rate, #canali, codifica, info
.ra	Real Networks	rate, #canali, qualità dei campioni

WAVE, AIFF e formato RIFF

- Il formato **WAVE** (estensione .wav) è caratteristico dei dati audio in ambiente Microsoft, ma sono riconosciuti anche in ambiente Mac. Per l'ambiente Mac è più frequente l'uso di file **AIFF** (estensione .aif)
- I file di estensione .wav appartengono al formato **RIFF** (Resource Interchange File Format) .
- Il formato RIFF prevede la memorizzazione dei dati in sezioni logiche all'interno dei file. Tali sezioni sono dette "*chunks*".
- I chunks sono diversi e di vario tipo ma due di essi sono sempre obbligatoriamente presenti:
 1. il chunk di formato
 2. il chunk dati
- Chunks opzionali sono i *cue-points* o *marker chunks* per individuare punti precisi; i *playlist chunks* per l'ordine di riproduzione dei marker, gli *associated data list chunks* che contengono liste di nomi associati ai marker e i *sampler chunks* *instrument chunks* per la definizione dei parametri relativi ad un certo strumento digitale.

Strutture dei formati audio



Byte offset

00



4 byte per i caratteri RIFF

04



4 byte per lunghezza rimanente file

08



dati audio



dati audio

Qui troviamo l'indicazione
specifica del formato WAVE
e il relativo contenuto di
campioni

Strutture del formato WAVE

→	dati audio	Byte offset	
		00	4 byte per i caratteri WAVE
	Format chunk	04	24 byte per il chunk del formato
	Data chunk	28	<i>Lunghezza variabile.....???</i>

Nel *format chunk* sono riportati tutti i descrittori della forma d'onda:

1. **Frequenza di campionamento**
2. **Nr. Canali** 1 = mono 2 = stereo o più canali organizzati in gruppi (*frames*)
3. **Codifica del segnale** (1 = PCM)
4. **Tasso di riproduzione** (campioni /sec)
5. **Media dei byte/sec**
6. **Allineamento dei frames**(Nr. Canali x Nr. Bi / 8) Es.: 1 x 16 / 8 = 2 (suono monofonico)

Struttura completa del formato WAVE

Struttura globale di un file WAVE

00	"RIFF"		
04	lunghezza di rDATA		
08	chunk	"WAVE"	
	rDATA		
12		chunk formato dati wFORMAT	"fmt"
16			Lunghezza di wDATA
20			Codifica
22			Canali
24			Tasso di riproduzione
26			Consumo medio di byte/sec
28			Allineamento
30		chunk dati wDATA	"data"
34			Lunghezza di dDATA
48			Dati audio
...			
...			

Supporti per l'audio digitale

I segnali numerici (digitali) possono essere:

1. Immagazzinati su un supporto fisico
2. Elaborati e riconvertiti quasi istantaneamente (real time processing)

Nel primo caso i supporti possono essere di varia natura e in primo luogo distinti per il tipo di supporto utilizzato. Si possono avere quindi:

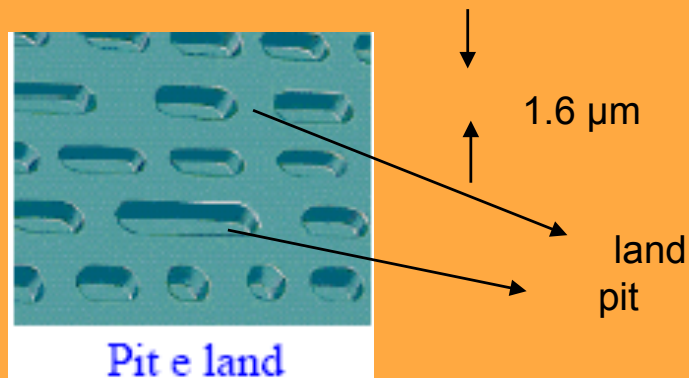
1. Supporti magnetici (Hard disk, floppy disk, nastro magnetico digitale)
2. Supporti ottici (CD – DVD – CD-ROM...)
3. Supporti a stato solido (Memorie RAM – ROM – FLASH – Memory Card)

CD

Il **CD** è ormai da tempo lo strumento fondamentale per la memorizzazione dell'audio digitale sia nella forma canonica, il **CD** originale che si compra nei negozi, che nelle altre varianti masterizzate (**CD-R Audio**, **CD-R con file WAV**, **CD-R con file MP3**, ecc.).

Il CD è stato sviluppato all'inizio degli anni '80 da Sony e Philips

Il CD è fisicamente un disco ottico di policarbonato di 12 cm di diametro. I bit sono impressi sotto forma di ***pit*** e ***land***. I ***pit*** sono le zone stampate nel substrato di plastica del disco (larghi circa **0.6 micron**) mentre i ***land*** sono le zone non stampate. Il substrato è posto vicino alla superficie superiore del disco ed è metallizzato al fine di riflettere il raggio laser. Durante la lettura ogni transizione ***pit-land*** e ***land-pit*** viene interpretata come un bit **1**, mentre le aree piane, che si trovano prima e dopo ogni transizione, sono qualificate come uno o più bit **0** consecutivi. I ***pit*** ed i ***land*** sono allineati in una traccia a spirale, che inizia vicino al diametro interno del disco e termina in prossimità del diametro esterno. I **CD** audio la cui durata è inferiore a quella massima, hanno un'area non utilizzata nella zona esterna del disco.



DATI : 2 canali (stereo) con $sr = 44100$

La lettura può avvenire secondo 2 modalità:

1. A velocità lineare costante (CLV) 150 kB/sec
- (2. A velocità angolare costante (CAV) (+ recenti)

Formato dati CD



Area Lead In

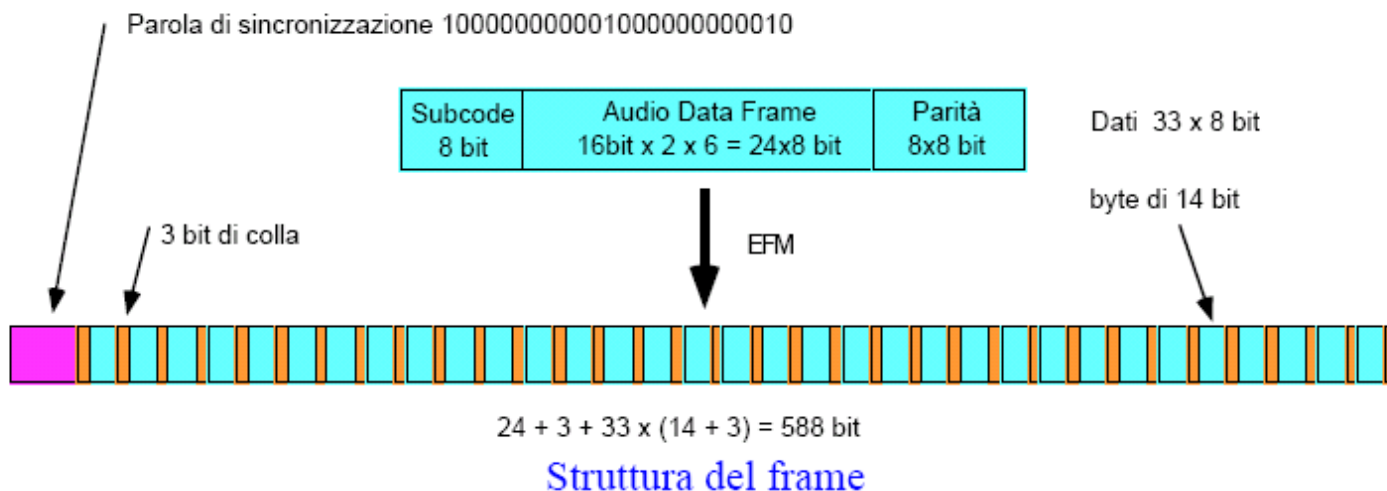
Area dati

Area Lead Out

- Con la tecnica CAV si è raggiunta progressivamente la velocità di **52x** il trasferimento base (**7.8 MB /sec** per una velocità di rotazione > 12000 rpm). Questa velocità è importante solo in fase di masterizzazione e nella lettura di CD-ROM

- Per ridurre al minimo la possibilità di errori in lettura dovuta a fattori elettro-meccanici i dati vengono codificati con una speciale tecnica denominata **Codifica Reed-Solomon – CIRC**

- La lettura avviene attraverso blocchi di 588 bit (**frame**) che hanno la struttura mostrata in figura



A queste informazioni si devono aggiungere le informazioni extra contenute nell'area Lead In / Out (N tracce, tempo assoluto, tempo di traccia)

Frames, Settori e Tracce di un CD audio

La struttura fisica dei dati di un CD audio è organizzata nel modo seguente:

- **FRAME** (588 bytes)
- **SETTORE** (98 frames)
- **TRACCIA** (costituita da un numero variabile di frames e che rappresenta un “brano”)

Alla frequenza di campionamento di **44100Hz** vi sono **44100 * 4** byte/sec di informazione buona. Dato che ogni frame contiene **24** byte la durata temporale di un frame è **1/7350** di secondo. Inoltre per completare un settore servono **98** frame e quindi un settore ha la durata di **1/75** di secondo, è questa l'unità temporale minima indirizzabile.

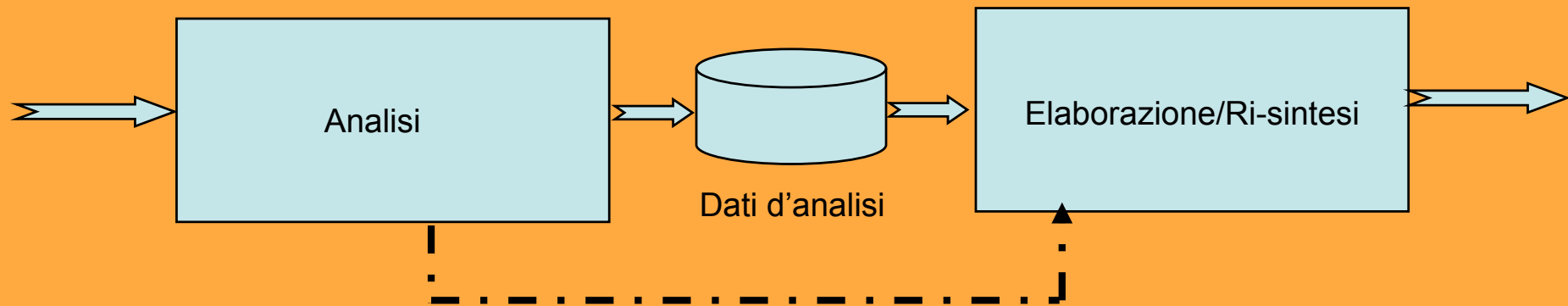
Le specifiche e i libri colorati

Oltre al CD audio sono stati sviluppati molti altri supporti, le cui specifiche individuali sono state riportate su una serie ben nota di libri con copertine di diverso colore secondo il seguente schema:

Pubblicazione	Anno	Denominazione supporto
Red Book	1982	CD-DA (audio)
Yellow Book	1985	CD-ROM (Read Only Memory)
Green Book	1987	CD-I (interactive)
Orange Book	1988	CD-Magneto Ottico – Photo CD
Orange Book II	1990	CD-R (Recordable – multisession)
Orange Book III	1997	CD-RW (ReWritable)
White Book	1993	Video-CD
Blue Book	1995	CD - Extra
Gold Book	1996	CD-HD (High Density)

Analisi del suono

- Il suono in forma digitalizzata consente di applicare su di esso un ragguardevole numero di **processi d'analisi** e di decomposizione dei dati registrati in *forma strutturata*.
- Il suono rappresentato in forma analitica può essere più facilmente elaborato e modificato



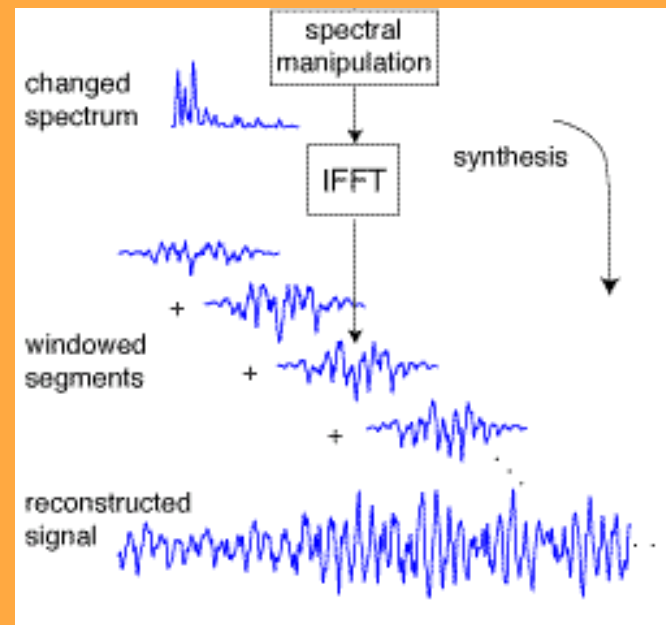
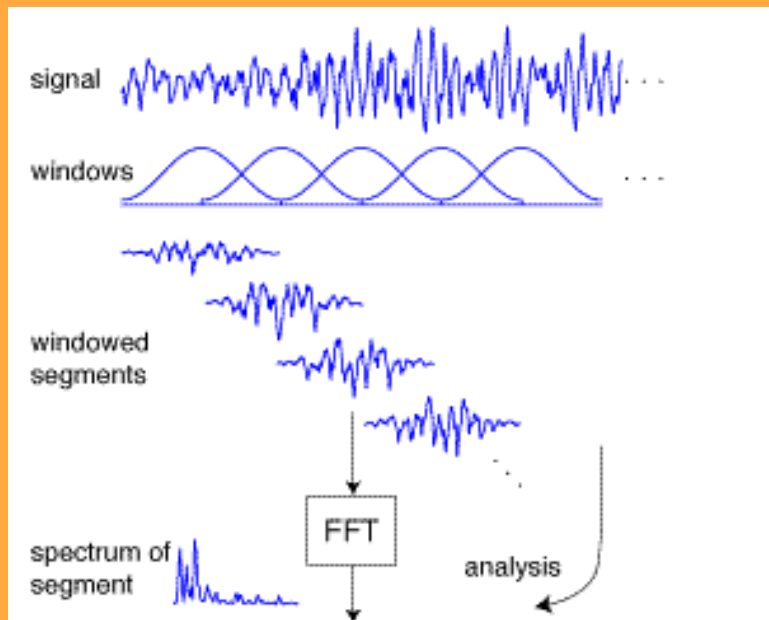
Il processo di analisi/sintesi può essere realizzato in tempo reale

Analisi in frequenza dei segnali campionati

- Come visto in precedenza, attraverso la Serie di Fourier (FS) è possibile decomporre e analizzare solo i segnali periodici. Oltre a ciò, il calcolo dei coefficienti della serie è possibile solo se si conosce l'espressione analitica del segnale da analizzare.
- Nel dominio digitale ogni segnale è rappresentato da una sequenza di numeri $x(nT)$ dove
 - n = campione n -esimo
 - T = periodo di campionamento ($1/sr$)In generale, poiché T è costante una sequenza viene indicata con la notazione $x(n)$.

Trasformata di Fourier Short-Time

- L'analisi di Fourier fa corrispondere un segnale variabile nel tempo in una rappresentazione statica del contenuto frequenziale in termini di componenti elementari.
- Per adattare tale tipo d'analisi al reale mondo dei segnali campionati, l'analisi viene condotta dinamicamente nel tempo. A tale proposito viene introdotta la "Short Time Fourier Transform" STFT



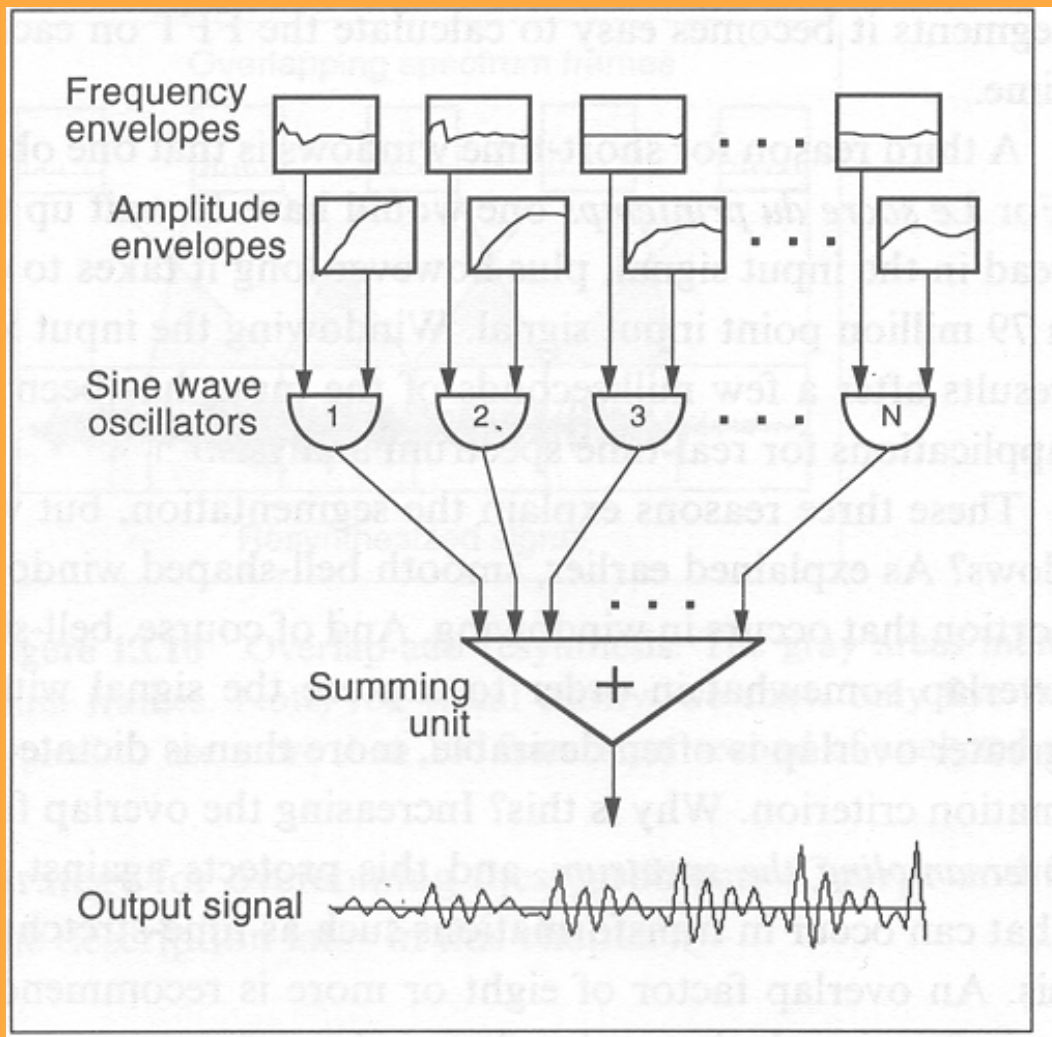
La finestatura consiste nel prelevamento di tratti brevi del segnale sfumati alle estremità per ridurre al minimo gli effetti di bordo.

Analisi STFT (il segnale viene finestrato)

Il segnale viene ri-sintetizzato

Metodo OA e SAR

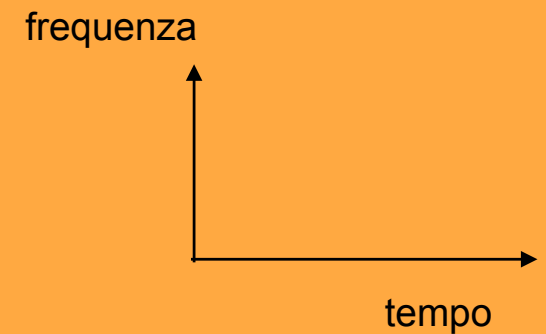
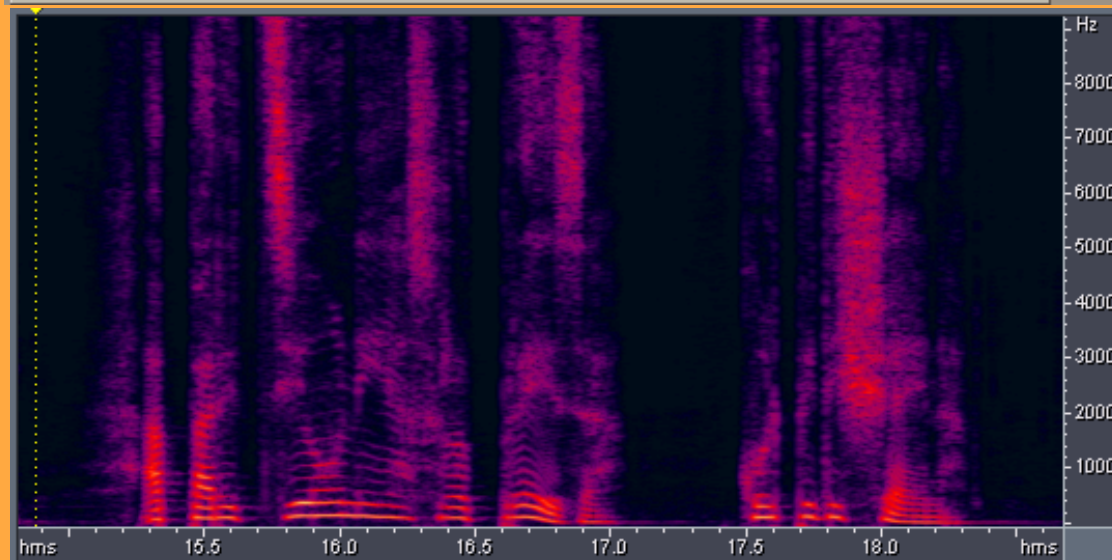
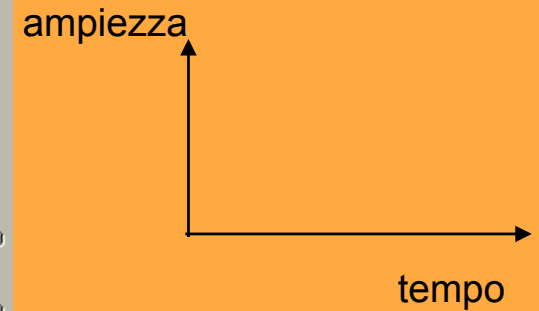
- La ri-sintesi può essere realizzata col metodo illustrato nella slide precedente cioè tramite **Overlap and Add** (Sovrapponi e Somma). Il metodo OA è computazionalmente efficiente.



In alternativa può essere utilizzato il metodo **Sinusoidal Additive Resynthesis** (Ri-sintesi Sinusoidale Additiva) che utilizza un banco di oscillatori, ciascuno controllato da inviluppi di ampiezza e frequenza.

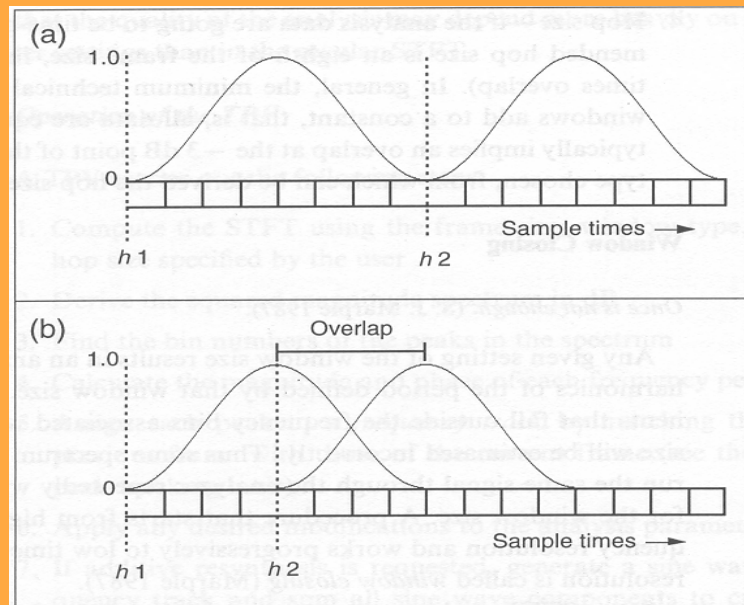
STFT e Sonogramma

Una delle più forme di rappresentazione grafica dei dati di analisi STFT è il **sonogramma** o spettrogramma. Viene originariamente utilizzato per l'analisi del segnale vocale ed oggi fa parte integrante di molti software di editing e analisi del suono digitale.



Phase Vocoder (PV)

- Una delle principali applicazioni della STFT è il Phase Vocoder. Un segnale audio viene decomposto in una sequenza di FRAMES di cui è possibile indicare la lunghezza.
- La tecnica del PV consente di analizzare e di risintetizzare il suono con un livello qualitativo che dipende dalla natura dei segnali stessi e dall'appropriata serie di parametri di analisi
- I parametri sono:
 1. *Frame size* (nr. di campioni analizzati in un intervallo temporale)
 2. *Window type* (morfologia della funzione per la finestrazione)
 3. *FFT size* (generalmente una potenza di 2)
 4. *Hop size* . *Overlap factor* (differenza temporale tra un frame e il successivo)

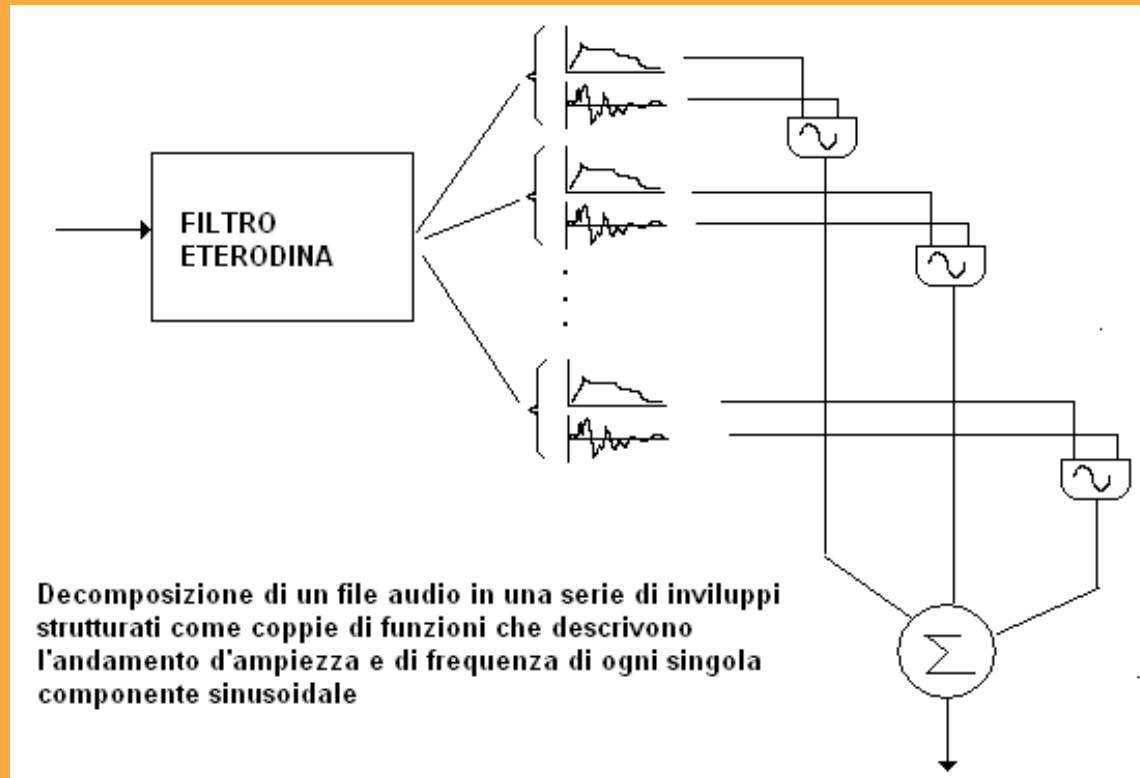


Variazione del parametro hop size per un frame di 8 campioni:

a) finestre non sovrapposte (hop = Frame size)

b) finestre parzialmente sovrapposte (hop < Frame size). In questo caso hop = 4 campioni

Analisi Eterodina

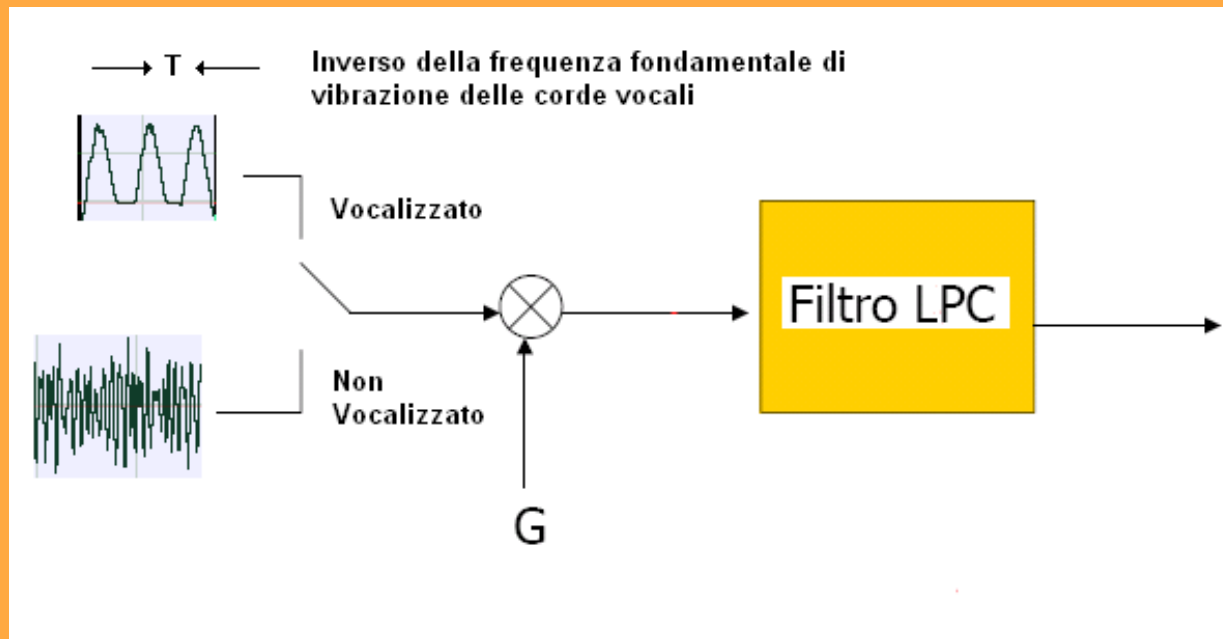


L'analisi è basata sull'estrazione di componenti armoniche (ampiezza e frequenza). Il suo utilizzo è adatto in particolare per i suoni periodici e "pseudoperiodici" tempo-varianti. Poco indicato per i suoni "inarmenici" e conforti componenti rumorose e di modulazione.

La fase di sintesi è realizzata anche in questo caso con un banco di oscillatori controllati in ampiezza e frequenza.

Linear Prediction (LPC)

La tecnica LPC deriva essenzialmente dagli studi per l'analisi e la sintesi vocale. Viene utilizzato un modello basato sul funzionamento del meccanismo di emissione vocale.



La sintesi è basata sulla modellazione dei due tipi di segnale (vocalizzato e non) attraverso un filtro che realizza la migliore simulazione delle risonanze del tratto vocale circa ogni 20 ms. Il termine "predizione" si riferisce al fatto che il campione attuale è in effetti una combinazione pesata di N campioni precedenti.