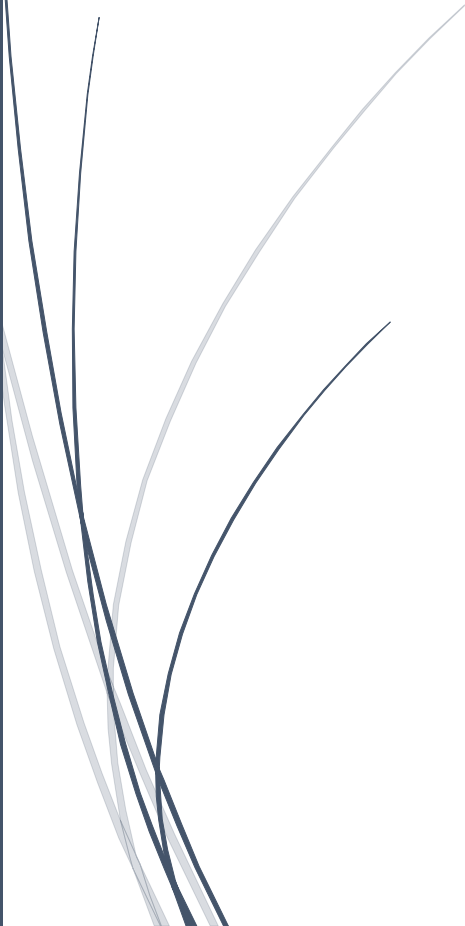


A dark blue vertical bar runs down the left side of the page. A blue arrow points to the right from this bar, containing the text '2018/2019'.

2018/2019

Approfondimento Informatica Musicale

Several thin, curved lines in dark blue and light grey originate from the bottom left and sweep upwards and to the right.

*SALVATORE CRISTALDI
EDOARDO CANNAVÒ*

Sommario

Lezione 1.....	4
Decibel	5
Lezione 2.....	7
Legge dell'inverso del quadrato:	7
Velocità del suono	7
Frequenza e Lunghezza d'onda	7
Riflessione.....	8
Diffrazione	8
Rifrazione.....	9
Assorbimento	9
Lezione 3.....	9
Teoria di Fourier	9
Spettrogramma	12
Scala Diatonica.....	12
LEZIONE 4	13
Attack:.....	13
Decay:	13
Sustain:	13
Release:	13
Spettro & Timbro.....	14
Rumore	14
Spettri di frequenza	14
Rumore bianco:	15
Rumore rosa:	15
Rumore Marrone:.....	16
Rumore Blu:	16
Rumore Viola:	16
Rumore Grigio:	17
Rumore Rosso:.....	17
Lezione 5-6-7	18
PSICOACUSTICA	18
COCLEA	18
Ossicini.....	19
Muscolo Stapedio	19
Soglie di Tolleranza del Rumore	20

Il suono - Percezione umana	21
Diagramma di Fletcher-Munson	21
Altezza di un Suono	22
Timbro percepito -Le formanti delle vocali	22
Localizzazione delle sorgenti sonore	23
Lezioni 8.....	25
Audio analogico:	25
Audio digitale:.....	26
Filtraggio:.....	27
Campionamento:.....	27
Quantizzazione:	28
Codifica:	28
Lezione 9.....	29
Quantizzazione	29
Esempio di Differenza tra quantizzazione a 4 bit e 3 bit	29
Quantizzazione Lineare	30
PCM	30
Bit Di Parità.....	31
Quantizzazione Non Lineare.....	31
Lezione 10 Digitalizzazione.....	33
Equalizzatori	33
Lezione 11.....	35
Filtro telefono:.....	35
Filtro midcut:	35
Filtro midpass:	35
Compressore:	37
Limitatore:	37
Espansore:	37
Limita Rumore (Noise gate):.....	38
Lezione 12.....	39
μ -law:.....	39
A-law:.....	40
PCM:	41
DPCM:.....	41
ADPCM:.....	42
Lezione 13.....	43

COMPRESSIONE	43
Entropia percettiva	43
Block coding.....	43
Sub-band coding.....	43
Transform coding.....	43
Hybrid coding.....	44
Schema generale di tipo percettivo	44
Lezione 14.....	46
MPEG	46
MPEG-1	46
MPEG-2.....	46
MPEG-4.....	47
MPEG-7	47
MPEG-21.....	47
LAYER 1	48
Layer 2	48
LAYER 3	49
Formati Audio Avanzati	49
PCM	50
WAV	50
MP3	50
WMA	50
FLAC	50
LEZIONE 15	51
MIDI Clock:.....	53
MIDI Time Code:.....	53
Evoluzione del MIDI.....	55

Lezione 1

Il **suono** (dal latino sonus) è la sensazione data dalla vibrazione di un corpo in oscillazione. Tale vibrazione, che si propaga nell'aria o in un altro mezzo elastico come un insieme di onde meccaniche longitudinali. Invece dall'altro canto l'audio è un segnale elettromagnetico che rappresenta e trasporta informazione sonora. L'audio è generato tra due dispositivi elettromagnetici e quindi esso permette la trasmissione del suono stesso.

Nel corso della storia il suono è stato possibile prima udirlo poi registrarlo e riprodurlo. Le prime apparecchiature atte a registrare e riprodurre suoni erano di tipo meccanico e non erano in grado di registrare la voce umana. La prima riproduzione automatica di musica può farsi risalire al IX secolo, quando i fratelli Banu Musa inventarono "il primo strumento musicale meccanico". Successivamente nel 1796, venne inventato da un orologiaio svizzero il carillon e tra il 1857 e il 1877 vennero inventati il fonautografo e il fonogramma. Dove il primo è stata la prima apparecchiatura conosciuta atta a registrare il suono, mentre il secondo fu un dispositivo in grado di registrare e riprodurre un suono.

Diamo adesso la definizione di onda e vediamo come essa si può classificare. Un'onda è una perturbazione che nasce da una sorgente e si propaga nel tempo e nello spazio, trasportando energia o quantità di moto senza comportare uno spostamento della materia. Essa è classificabile in base al:

- Mezzo di propagazione
- Rispetto alla direzione delle particelle
- Forma
- Periodicità.

Mezzo di propagazione:

Il mezzo può essere o meccanico o elettromagnetico e per questo l'onda si può ulteriormente classifica in onde meccanica o elettromagnetica. Un'onda meccanica è la propagazione di una perturbazione in un mezzo gassoso, liquido o solido con il trasporto di energia e quantità di moto e quindi interesserà le particelle di materia. Un'onda elettromagnetica è una perturbazione che interesserà la variazione del campo elettrico e magnetico. Essa di può propagare nel vuoto.

Direzione delle particelle:

L'onda può essere trasversale o longitudinale. La prima si ha quando le particelle si propagano in direzione perpendicolarmente alla direzione di propagazione. La seconda, invece, quando le particelle del mezzo in cui si propaga l'**onda** oscillano lungo la direzione di propagazione.

Forma:

La forma innanzitutto mi descrive il comportamento dell'ampiezza nel tempo. Un'onda può essere sinusoidale, dove il suo comportamento nel tempo è descritto dalla funzione seno o avere tutt'altra forma.

Periodicità:

Un'onda si dice periodica se si ripete uguale a se stessa, ovvero delle sequenze di oscillazioni, per intervalli di tempo di periodo T . Formalmente: una funzione si dice periodica di periodo T se $\exists T > 0 : \forall t \in \mathbb{R} \Rightarrow f(t) = f(t + T)$. Aperiodica se non ha nessuna regolarità nel tempo.

Data un'onda possiamo individuare delle componenti: ampiezza, frequenza, fase, fase iniziale, periodo, pulsazione, velocità e lunghezza d'onda.

L'**ampiezza**, chiamata anche intensità, è la massima variazione di una grandezza in un'**oscillazione periodica**. In parole povere descrivere il contenuto energetico dell'onda e caratterizza pure il volume

del suono cioè la forza stessa di un suono, determinato dalla pressione che l'onda sonora esercita sul timpano. Proprio questa "pressione" che si viene a creare può essere misurata in vari modi: SIL e SPL.

***Tabella dei DeciBel.**

Esempio di causa	Livello sonoro [dB]	Percezione umana
Minimo percepibile	0	Calma
Rumore in uno studio pro, fruscio delle foglie	20	
Bisbiglio, camera da letto silenziosa la notte	30	
Biblioteca	40	
Strada tranquilla, casa di giorno	50	Intrusione nella privacy
Conversazione normale	60	Telefono difficile da usare
Ristorante, uffici rumorosi, autostrada, aspirapolvere a 1 m	70	
Asciugacapelli, sveglia a 60 cm	80	Fastidioso
Camion pesante a 10 m	90	Molto fastidioso, iniziare a usare protezioni
Petardi, discoteca, chitarra elettrica al massimo (100W)	100	
Concerto rock, massimo sforzo vocale, rullante	110	Doloroso
Ciacson a 1 m	120	
Tuono	130	Danni all'udito istantanei
Sirena anti-aerea, jet al decollo a 30 m	140	
Razzo al decollo	180	

L'SPL (Sound Pressure Level) è una misura logaritmica della pressione sonora efficace di un'onda meccanica (sonora) rispetto ad una sorgente sonora di riferimento, viene misurata in dB.

SPL

Dove p_0 è la pressione sonora di riferimento pari a $25 \mu\text{Pa}$, che rappresenta la soglia minima di udibilità per un tono puro a 100 Hz.

A seguito di quanto detto è lecito chiedersi: perché usare la scala logaritmica e non il semplice Pascal ?

Il motivo è semplice: a causa dell'enorme range in cui lavoriamo. Infatti, possiamo passare da $25 \mu\text{Pa}$ fino a oltre 100 Pa, dove 130 Pa rappresenta la soglia del dolore fisico. A sinistra ne vediamo degli esempi.

Come già detto la soglia minima di udibilità è pari a $25 \mu\text{Pa}$ quella massima è 30 N/m^2 ovvero molto più grande! Per questo motivo abbiamo bisogno del Decibel.

Decibel

Ma cos'è il decibel ? Il **decibel** (simbolo dB) è la decima parte del bel (simbolo **B**): $10 \text{ dB} = 1 \text{ B}$ ed è un'unità di misura logaritmica del rapporto fra due grandezze omogenee e si usa per comprimere il range delle grandezze. È una grandezza relativa perché ha bisogno di grandezze omogenee, altre grandezze simili sono il Bel e il Neper. Esso ha delle proprietà:

- È adimensionale.
- Va specificata l'unità di misura di partenza.
- Sfrutta la proprietà dei logaritmi del prodotto.
- A un aumento di 10 dB corrisponde un aumento della grandezza originale di un fattore 10.

Inoltre, esso può essere usato per misurare grandezze (tensioni, potenze ecc.) direttamente in decibel, ovvero riferendo la grandezza alla sua unità di misura e sono chiamati **Decibel Assoluti**.

Ritornando al discorso precedente la pressione oltre ad essere misurata in SPL può essere misurata in SIL.

Il SIL o **intensità acustica** o **sonora** (Sound Intensity Level) è una grandezza fisica definita come il rapporto tra la potenza di un'onda sonora e l'area della superficie che da essa viene attraversata; oppure come l'energia che nell'unità di tempo attraversa l'unità di superficie posta in un punto perpendicolarmente alla direzione di propagazione del suono.

In questo caso la soglia minima di udibilità è 10^{-12} W/m^2 . Si esprime come:

$$= 10 \log_{10} \frac{I}{I_0} \text{ dB}$$

Frequenza: è il numero di oscillazione complete nell'unità di tempo e si misura in Hz.

Fase: la fase di una funzione periodica ad un certo istante temporale è la frazione di periodo trascorsa rispetto ad un tempo fissato.

Fase iniziale: rappresenta il periodo trascorso rispetto all'istante di tempo 0.

Periodo: è una grandezza fisica relativa alle onde definita come l'intervallo temporale corrispondente alla lunghezza d'onda oppure come il tempo necessario per compiere un'oscillazione completa. $T = 1/f$.

Pulsazione: è una grandezza che misura la velocità con cui viene effettuata un'oscillazione completa. Si misura in radianti/secondo. $\omega = 2\pi f = 2\pi/T$.

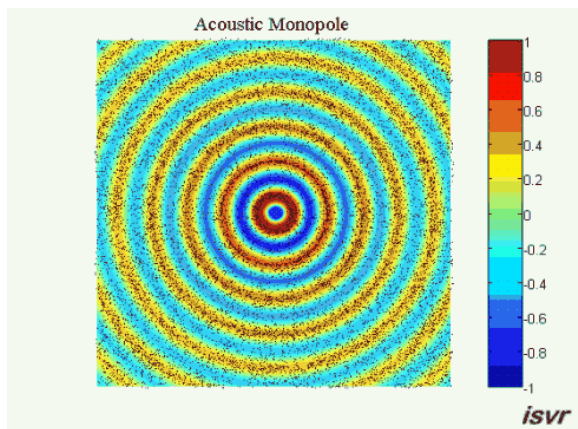
Lunghezza d'onda: la distanza percorsa tra due creste o fra due ventri della relativa forma d'onda, e viene comunemente indicata dalla lettera greca λ . Si misura in metri.

Velocità d'onda: è lo spazio percorso dalla perturbazione nel tempo e si misura metri su secondo e dipende dal mezzo in cui essa si propaga. $\lambda = vT = v/f$.



Lezione 2

Legge dell'inverso del quadrato:



Il caso del **monopolo acustico** riassume la legge dell'inverso del quadrato. Il monopolo, considerato come **sorgente puntiforme** (si definisce sorgente puntiforme quando quest'ultima ha dimensioni molto più piccole della lunghezza d'onda in gioco) emette **onde sferiche** nell'ambiente circostante in assenza di ostacoli. Il suono che si andrà ad irradiare sarà uniforme in tutte le direzioni e l'intensità sonora diminuirà sempre più all'aumentare della distanza.

La Legge dell'inverso del quadrato stabilisce che l'intensità del suono in campo libero è inversamente proporzionale al quadrato della distanza dalla sorgente. Se la distanza

raddoppia, l'intensità si riduce a 1/4; se triplica, l'intensità si riduce a 1/9; se quadruplica, l'intensità si riduce a 1/16.

Velocità del suono

La **velocità del suono** è la velocità con cui un suono si propaga in un certo ambiente, detto *mezzo*. La velocità del suono varia a seconda del mezzo (ad esempio, il suono si propaga più velocemente nell'acqua che non nell'aria), e varia anche al variare delle proprietà del mezzo, specialmente con la sua temperatura. Nell'aria, la velocità del suono a 0 °C è di 331 m/s (pari a 1.191,60 km/h) .

Approssimando possiamo ricavare una legge lineare $V(T) = (331,45 + (0,62 * T))$ m/s con T la temperatura misurata in °C).

La velocità di propagazione del suono è maggiore nei solidi che nei fluidi.

Frequenza e Lunghezza d'onda

La **frequenza** non è altro che il **numero di volte in cui il fenomeno ondulatorio si ripete nell'unità di tempo**, il suo inverso è il **tempo (periodo, misurato in secondi)** che occorre per compiere un'oscillazione completa.

$$T = \frac{1}{f};$$

Se volessimo definire la **lunghezza d'onda** potremmo dire che essa non è altro che il **periodo metrico dell'onda**. La **frequenza è l'inverso della lunghezza d'onda**, la sua unità di misura è l'Hertz (Hz). Queste grandezze sono legate tra loro dalla formula:

$$\lambda f = c ;$$

Dove:

λ = lunghezza d'onda

f = frequenza (Hertz)

c = **velocità del suono**

Frequenza e lunghezza d'onda, sono inversamente proporzionali tra loro. **All'aumentare della frequenza diminuisce la lunghezza d'onda** e viceversa.

Riflessione

E' il fenomeno acustico che si verifica quando l'onda sonora incontra un ostacolo e torna indietro (cambia quindi direzione).

Un'onda che incide su una superficie piana con un angolo di incidenza α (tra la normale alla superficie e la direzione di propagazione dell'onda) viene riflessa con un angolo di riflessione pari ad α (quindi gli angoli di incidenza e riflessione sono uguali). Le superfici concave vengono evitate in acustica in quanto tendono a concentrare il suono in un preciso punto (focalizzazione) creando distribuzioni sonore disomogenee. Vengono invece utilizzate per la costruzione di microfoni direzionali in quanto consentono di amplificare segnali anche molto deboli.

Viceversa le superfici convesse hanno la proprietà di diffondere il suono (dispersione) e dunque sono ampiamente usate per migliorare l'acustica degli ambienti. Quando un'onda che si riflette su una superficie convessa, il prolungamento dell'onda riflessa passa per il fuoco della superficie.

L'eco è un esempio di riflessione: onde sonore contro un ostacolo che vengono a loro volta nuovamente "percepite" dall'emettitore più o meno immutate e con un certo ritardo rispetto al suono diretto. Tale ritardo non dev'essere inferiore ad 1/10 di secondo. Al di sotto di tale valore non si può più parlare di eco ma di riverbero. Un tipico esempio di riverbero è quello prodotto in una stanza dalla riflessione di onde sonore sulle pareti perimetrali.

Si parla propriamente di eco quando le singole riflessioni dell'onda sonora sono percepite distintamente dall'ascoltatore.

In termini più generali, l'eco può essere definita come un'onda che viene riflessa da una discontinuità nel mezzo di propagazione, e che ritorna con una intensità e ritardo sufficiente per essere percepita. Può essere "utile" (come nei sonar) o "indesiderata" (come nei sistemi telefonici)

Ricapitolando: si ha riverbero quando l'onda incidente si confonde nell'orecchio dell'ascoltatore con l'onda riflessa, mentre si ha eco quando le due onde risultano distinte.

Diffrazione

Mutamento di direzione di un onda sonora a causa di un ostacolo (superarlo girandoci attorno). Se il pannello (ostacolo) è più piccolo della lunghezza d'onda, la diffrazione sarà significativa e l'onda si ricostituisce dietro l'ostacolo dopo aver perso un pò di energia; se invece il pannello è grande, l'onda viene in gran parte riflessa e si forma una zona d'ombra. Tutto dipende dalla frequenza in quanto suoni con una grande lunghezza d'onda (e dunque bassa frequenza) superano con facilità ostacoli con una dimensione minore della loro lunghezza d'onda. Questo è uno dei motivi per cui le prime frequenze che vengono attenuate sono quelle alte mentre quelle basse si propagano a distanze molto maggiori.

Rifrazione

Con tale termine si indica il fenomeno secondo il quale un'onda che attraversa due mezzi di diversa densità cambia direzione nel passaggio dall'uno all'altro. Tale comportamento è facilmente spiegabile se teniamo presente che il suono viaggia più velocemente in mezzi più densi.

Assorbimento

Puo` essere descritto come la conversione di energia acustica in energia termica da parte di una superficie. In altre parole, quando un suono viene a contatto con un ostacolo, gli trasferisce energia che viene dissipata sotto forma di calore. Il coefficiente di assorbimento esprime il rapporto tra energia assorbita e energia incidente.

Lezione 3

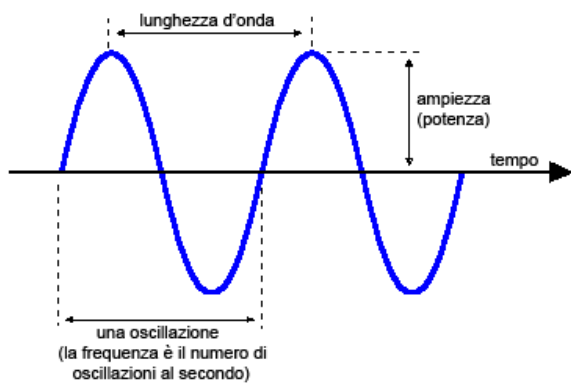
Teoria di Fourier

I suoni innanzitutto possono essere distinti in **periodici** e **non periodici**. Nei primi è possibile distinguere un ciclo di oscillazione che si ripete nel tempo mentre nei secondi (classificati generalmente come rumore) no. Se un suono è periodico sarà possibile associare ad esso un'altezza stabilita e pertanto avremo come percezione sonora un tono o una nota. Invece, nei suoni non periodici questo non è possibile poiché il cervello non riesce a contare i cicli di oscillazione. Nella musica entrambe le tipologie sono molto importanti in quanto i suoni periodici sono utilizzati per creare la melodia e l'armonia mentre quelli non periodici sono utilizzati principalmente per la ritmica.

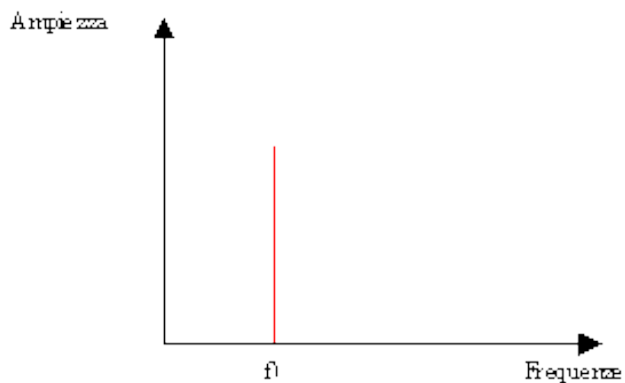
La **frequenza** di un suono è il numero di vibrazioni complete che la sorgente compie in un secondo o alternativamente il numero di compressioni e rarefazioni che subisce una particella d'aria in un secondo. E' la principale responsabile dell'altezza di un suono, l'altezza di un suono è il parametro legato alla sensazione di gravità acutezza che si percepisce di un suono. Non tutti i suoni hanno un'altezza definita, l'altezza è una caratteristica che risulta dalla periodicità di un segnale cioè dal fatto che il segnale ripeta lo stesso andamento per un po' di tempo

Differenza tra un suono puro e suono complesso:

- Un **suono puro** detto anche tono è costituito da una sola frequenza e quindi descritto da un'onda sinusoidale semplice, l'andamento del segnale è molto arrotondato, il periodo è composto da una singola compressione ed una singola rarefazione ben definite, l'ascolto non è particolarmente interessante. I suoni puri, infatti, possono essere prodotti quasi esclusivamente in laboratorio.

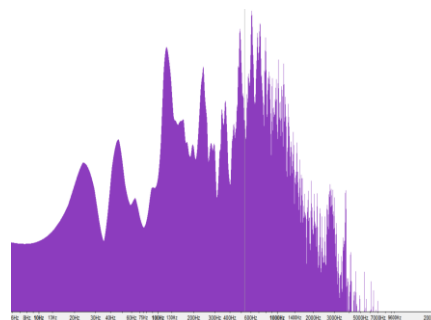
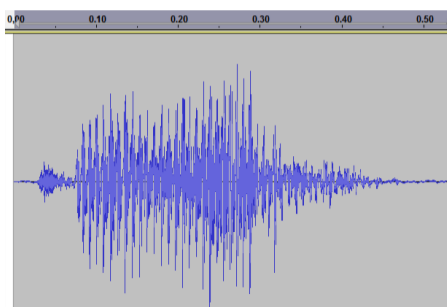


Esempio di un Tono puro.



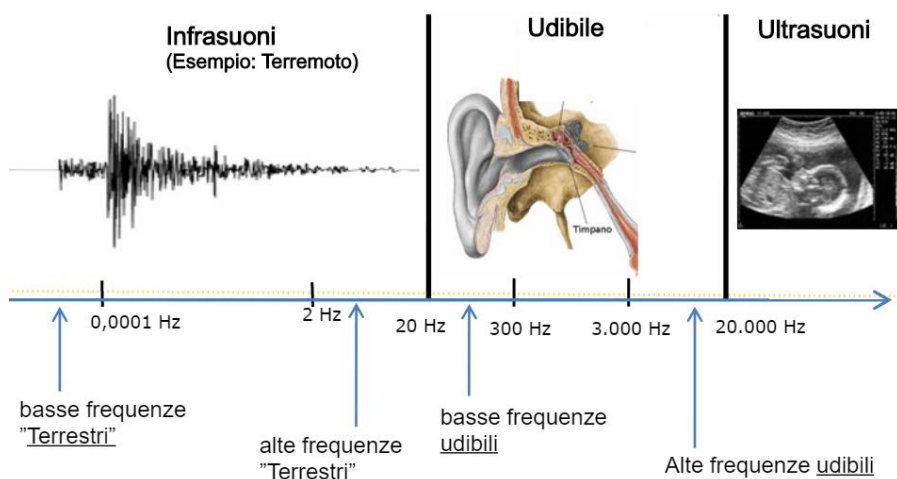
Spettro di un Tono puro.

-Un **suono complesso** consiste invece in più frequenze sommate in un'onda dall'andamento articolato, in un singolo periodo possono essere comprese più alternanze di compressioni e rarefazioni intermedie, l'ascolto rivela il timbro caratteristico di una sorgente e dell'ambiente circostante. In generale i suoni in natura sono di tipo complesso e lo specifico andamento deriva dal metodo di produzione del suono da parte della sorgente, la tipica eccezione è il diapason che riesce a produrre un suono sinusoidale quasi puro.



I suoni sono segnali che hanno frequenze comprese all'incirca tra i 20 e i 20000 Hertz, questi limiti derivano direttamente dal nostro apparato uditivo.

Gli **infrasuoni** sono caratterizzati da vibrazioni che hanno una frequenza inferiore ai 20Hz, mentre gli **ultrasuoni** sono caratterizzati da vibrazioni con frequenza superiore a 20 KHz.



In natura non esistono dei segnali realmente periodici, ma i suoni musicali, prodotti cioè da sorgenti che sono strumenti musicali tradizionali, hanno delle fasi di periodicità significative, e per essi ha senso parlare della sensazione di altezza. Il parametro percettivo che meglio corrisponde alla forma d'onda è il timbro. Il **Timbro** è quella caratteristica del suono per la quale, a parità di frequenza, è possibile distinguere due suoni prodotti da sorgenti diverse come ad es. il suono di uno strumento da quello di un altro (vedi un flauto e una chitarra). Il timbro dipende dal contenuto spettrale del suono costituito dalla somma di componenti sinusoidali che si può studiare attraverso l'analisi di Fourier.

Due elementi contribuiscono alla ricchezza delle forme d'onda complesse (e quindi al Timbro):

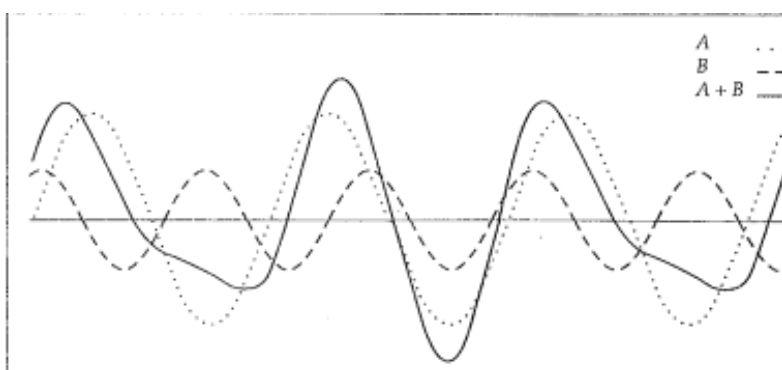
nel dominio della frequenza le componenti spettrali;

nel dominio del tempo, i transitori con cui ogni segnale reale deve fare i conti in quanto occorre un tempo finito perché inizi o si estingua una vibrazione.

Ci concentriamo in primo luogo sullo studio delle componenti spettrali che richiede l'analisi di Fourier. L'analisi di Fourier è un metodo analitico che ci permette di evidenziare in un segnale molti aspetti relativi alla natura del suono che non sono riconducibili direttamente alla rappresentazione nel dominio del tempo.

Chiariamo le differenze tra il dominio del tempo e dominio della frequenza. Una rappresentazione nel dominio del tempo descrive l'andamento di una grandezza, nel nostro caso l'intensità sonora. Ovvero considerando un punto della curva, esso rappresenterà l'intensità del suono in un preciso istante di tempo corrispondente in ascissa. Analogamente, quindi, se una rappresentazione è nel dominio della frequenza, la curva che descrive un segnale associa l'intensità alle frequenze specifiche. Nel 1800 il matematico francese Fourier dimostrò che *qualsiasi segnale complesso poteva essere descritto come una somma di segnali sinusoidali semplici*. È l'autore del metodo matematico noto come **Trasformata di Fourier** che permette di individuare le componenti di frequenza di un segnale. L'idea principale è che un segnale complesso è costituito da più segnali semplici; quindi, quando si menzionano le frequenze di un segnale complesso, ci si riferisce alle frequenze di questi segnali semplici. Requisito del segnale complesso è la *periodicità*, ossia un andamento, anche solo parziale e limitato, che si ripete a intervalli di tempo regolari. La sovrapposizione delle onde implica che, istante di tempo, delle onde semplici vengono sommate per ottenere l'ampiezza totale del segnale. Infine, il fatto che le frequenze componenti siano armoniche della frequenza fondamentale significa che tali frequenze sono multipli interi della frequenza fondamentale, cioè se la frequenza fondamentale è f , le armoniche di f sono $2f, 3f$, e così via. Non è detto che siano presenti tutte le armoniche ma tutte le frequenze presenti sono multipli della frequenza fondamentale.

Figura 1.14
Somma di sinusoidi:
l'onda a tratto solido
è data dalla somma
delle due onde
tratteggiate e punteggiate

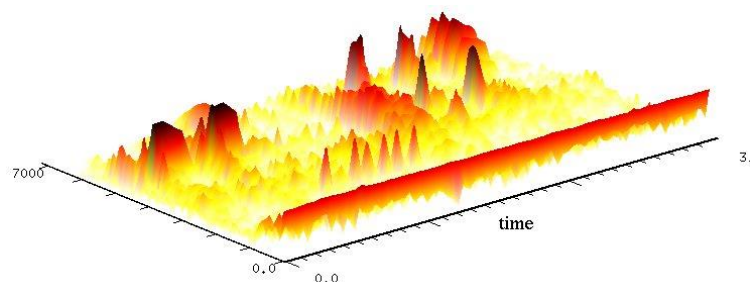


Spettrogramma

Uno **spettrogramma** o Sonogramma è la rappresentazione grafica dell'intensità di un suono in funzione del tempo e della frequenza.

E' possibile rappresentare la funzione $i(t,f)$ come una superficie (su un diagramma cartesiano con assi t, f ed i), di solito per gli spettrogrammi si usa un'altra rappresentazione grafica, nella quale:

- sull'asse delle ascisse è riportato il tempo in scala lineare;
- sull'asse delle ordinate è riportata la frequenza in scala lineare o logaritmica;
- a ciascun punto di data ascissa e data ordinata è assegnata una tonalità di grigio, o un colore, rappresentante l'intensità del suono in un dato istante di tempo e a una data frequenza; la relazione fra l'intensità del suono e la scala di grigi o di colori può essere lineare o logaritmica

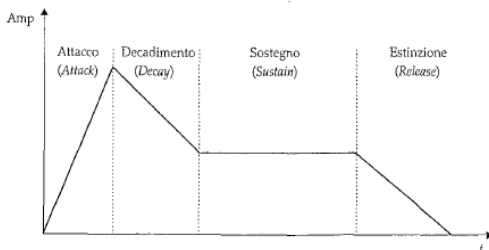


Scala Diatonica

Qualsiasi composizione musicale, vocale o strumentale, si costruisce dopo aver scelto il carattere espressivo che deve avere utilizzato suoni con particolari relazioni tra loro. Tale serie di suoni è la scala diatonica. La scala musicale è detta diatonica, quando è formata dalla successione di 8 suoni consecutivi, di cui il primo e l'ottavo hanno lo stesso nome ma diversa altezza.

LEZIONE 4

Si consideri questo esempio: la maggior parte delle canzoni segue un certo andamento dove si ha un inizio, dove la canzone stessa cresce poi vi è un corpo centrale e infine un fine dove la canzone (di solito) finisce pian piano, come quando un chitarrista esegue una nota, percepiamo un primo impatto sonoro, poi piano piano la nota si estingue.



Ora attraverso questa analogia possiamo descrivere il tempo di vita di un generico suono che descrive un certo andamento detto **inviluppo**. L'inviluppo è quindi l'andamento dell'ampiezza di un suono dal momento in cui viene generato a quando si estingue, cioè la sua evoluzione stessa. L'inviluppo si divide in 4 transitori o fasi: attack, decay, sustain, release.

Attack:

È il tempo che impiega il volume per passare da zero al suo valore massimo. In un organo l'attacco è immediato in quanto la nota viene suonata al massimo volume non appena si preme un tasto. Invece altri suoni possono essere caratterizzati da un attacco più graduale, come ad esempio un flauto. In un pianoforte può durare circa 1/100 di secondo. È ovvio che ogni suono ha una fase di attacco, in quanto ogni sistema fisico vibrante risponde con un tempo caratteristico: il tempo necessario all'instaurarsi delle onde stazionarie, o all'affermarsi di un particolare modo di vibrazione del sistema.

Decay:

Rappresenta il tempo che il suono impiega a passare dal volume massimo raggiunto durante la fase di attack al volume di sustain. Cioè il tempo che serve per raggiungere un'ampiezza costante. Esso è presente in quei strumenti (es. tromba) in cui il suono scatta solo se un determinato parametro fisico (ad esempio pressione del soffio) supera una certa soglia.

Sustain:

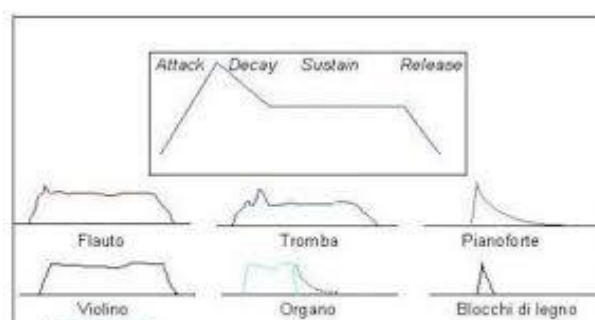
Qui l'ampiezza si mantiene costante, mentre l'esecutore continua a fornire energia. Questa fase non esiste per gli strumenti a percussione.

Release:

In questa fase il volume diminuisce fino al valore 0. Impostando il rilascio si definisce la rapidità con cui avviene questa diminuzione. È la fase che inizia nel momento in cui l'esecutore smette di fornire energia allo strumento e il suono decade più o meno rapidamente. Tale fase può essere anche molto lunga negli strumenti a percussione (si pensi alle note base di un pianoforte, o al suono di un gong), mentre di solito è breve negli archi e nei fiati. Ovviamente tutti i suoni hanno un rilascio. In un organo la fase del rilascio è istantanea: non appena si rilascia il tasto dell'organo, la nota cessa; invece nell'emulazione del suono di una chitarra questa fase è lenta.

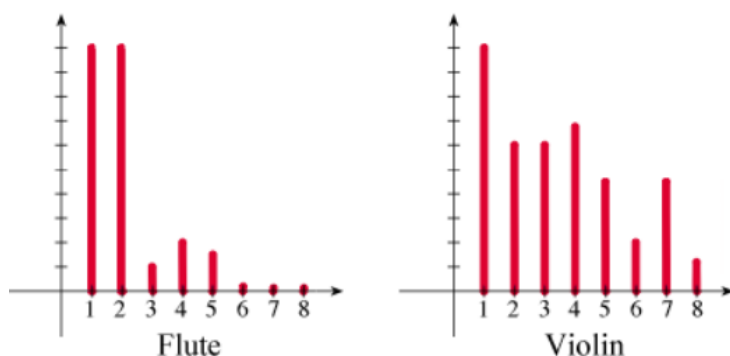
Ogni strumento possiede un inviluppo ADSR caratteristico.

Esempi: ➔



Spettro & Timbro

Il timbro è la qualità percepita di un suono che ci permette di distinguere due suoni che hanno la stessa altezza e la stessa intensità. In parole più semplici il timbro è la qualità del suono che ci permette di distinguere la voce di un violino da quella di un flauto, quando i due strumenti stiano emettendo una stessa nota. Per esempio, il timbro ci permette di distinguere l'uno dall'altro i diversi strumenti che compongono un'orchestra anche questi suonano insieme. In pratica quando uno strumento emette una nota di una determinata frequenza, esso, a causa dei vincoli imposti dalla "geometria" delle parti oscillanti degli strumenti musicali, genera, insieme alla nota fondamentale, più note tutte di frequenza multipla intera della fondamentale stessa (armoniche). In sostanza lo spettro di un suono ne caratterizza un timbro. Tale caratterizzazione è data dal numero e dal contributo delle varie frequenze dello spettro.



Come si può vedere, le **note musicali** sono segnali periodici non sinusoidali composti da più armoniche:

- l'**armonica fondamentale** (quella a frequenza più bassa) determina l'**altezza** percepita della nota;
- le **altre armoniche** invece determinano il **timbro**, cioè permettono di distinguere strumenti diversi.

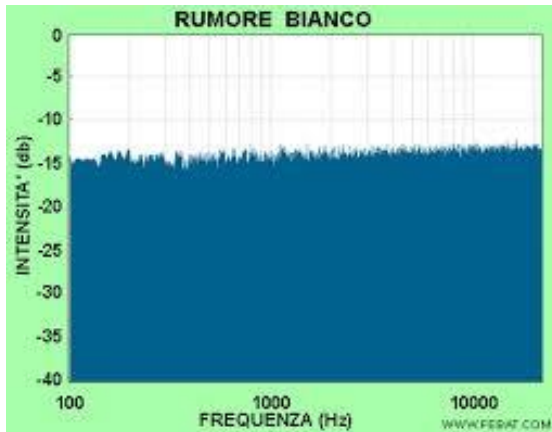
Rumore

Nella nostra vita quotidiana suoni, rumori e colori sono componenti essenziali della nostra vita. Tra di essi, tra suono e rumore, c'è una profonda differenza: nel primo si hanno oscillazioni regolari e rapide, mentre nel rumore tale oscillazioni sono irregolari ed imprevedibili. Esso è percepito come un segnale di disturbo rispetto all'informazione trasmessa in un sistema. Nella maggior parte dei casi essendo classificati come "disturbi" si cerca di attenuarli il più possibile. Ma ve ne sono alcuni che, caratterizzati da una componente aleatoria, vengono studiati. I principali rumori prendono il nome dei colori dello spettro visibile.

Spettri di frequenza

Gli spettri di frequenza dei rumori sono molto simili alle **radiazioni elettromagnetiche** assorbite dall'occhio umano per vedere i colori. Per capire ancora meglio: come il bianco è la somma di tutti i colori, il rumore bianco è la somma di tutte le frequenze udibili dall'orecchio umano. Essi sono classificabili in rumori: bianco, rosa, marrone, blu, viola, grigio, rosso.

Rumore bianco:



La sua ampiezza è costante e non è periodico nel tempo. È un rumore che in natura non esiste ma è creato artificialmente da alcune apparecchiature. È chiamato *bianco* per analogia con il fatto che una radiazione elettromagnetica di simile spettro all'interno della banda della luce visibile apparirebbe all'occhio umano come luce bianca.

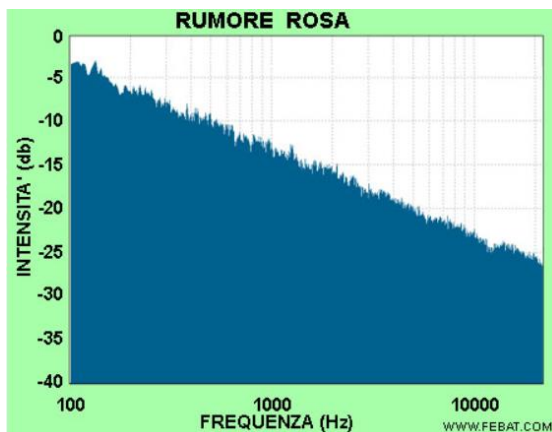
Il suo utilizzo è molto vario poiché viene percepito dall'orecchio come un qualcosa di piacevole e rilassante. Molte persone utilizzano il rumore del phon

per concentrarsi o per distendersi. Matematicamente parlando si può parlare di un vettore casuale \mathbf{w} ed è rumore bianco se e solo se il suo vettore medio e la sua matrice di autocorrelazione sono rispettivamente:

$$\begin{aligned}\mu_w &= \mathbb{E}\{\mathbf{w}\} = 0 \\ R_{ww} &= \mathbb{E}\{\mathbf{w}\mathbf{w}^T\} = \sigma^2 \mathbf{I}\end{aligned}$$

Si ha cioè un vettore medio nullo e una matrice di autocorrelazione che è un multiplo della matrice identità. Quando la matrice di autocorrelazione è un multiplo della matrice identità si dice che la correlazione è sferica.

Rumore rosa:

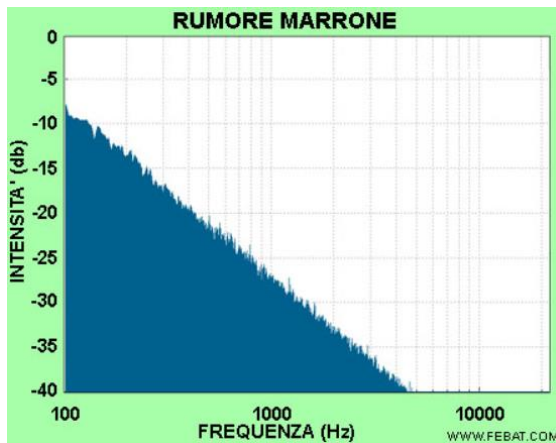


È un rumore particolare in cui le componenti a bassa frequenza hanno potenza maggiore, a differenza del rumore bianco in cui la potenza è uguale per qualsiasi frequenza. Esso è caratterizzata da una distribuzione spettrale di potenza inversamente proporzionale.

È impiegato nei grandi uffici o nelle sale d'attesa per favorire la concentrazione e il rilassamento. Questo tipo di rumore è strutturato in modo tale da compensare la sensibilità dell'orecchio umano alle varie frequenze, e viene utilizzato per l'equalizzazione del suono in ambito professionale. In questo segnale di rumore

l'intensità diminuisce di 3 dB per ottava.

Rumore Marrone:

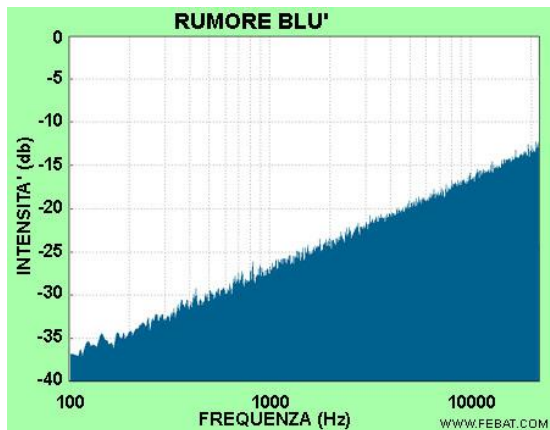


Il rumore marrone o rumore browniano è il rumore prodotto dal moto browniano. L'energia di questo rumore risiede nelle **basse frequenze**, molto abbondanti rispetto al rumore bianco e a quella rosa. Il suono appare simile a quello del tuono. Oppure il fragore delle cascate d'acqua.

Come il rumore rosa lo spettro presenta una relazione inversamente proporzionale. L'intensità si riduce – 6dB per ottava.

Lo spettro di potenza del rumore marrone è dato da $S(\omega) = \frac{S_0}{\omega^2}$.

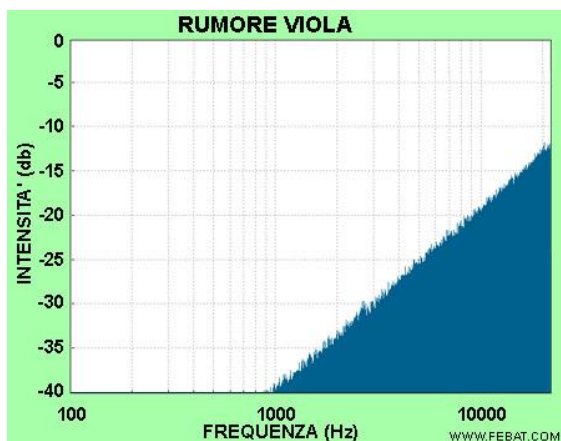
Rumore Blu:



La sua ampiezza non è costante e presenta una prevalenza di **frequenze alte**. Ha una **dipendenza direttamente proporzionale con incremento di 3 dB per ottava**.

Presenta un rumore stridulo e artificiale.

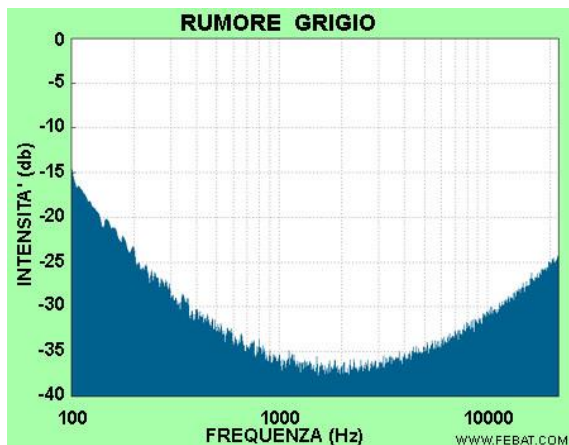
Rumore Viola:



Ha una crescita proporzionale all'aumentare della frequenza con incremento di 6 dB a ottava. Esso è considerato quello più fastidioso e produce un sibilo molto stridente.

Il rumore viola è un segnale adatto al trattamento degli acufeni (disturbi uditivi). Il suono prodotto ricorda quello di un getto di vapore.

Rumore Grigio:



Il **rumore grigio** è rumore casuale sottoposto ad una curva di filtro pari alla sensibilità psicoacustica (curve isofoniche) in un determinato intervallo di frequenze, dando all'ascoltatore la percezione che sia altrettanto forte a tutte le frequenze.

Viene utilizzato per equalizzare i segnali audio in modo tale che tutte le frequenze vengano percepite allo stesso volume da parte di un ascoltatore umano.

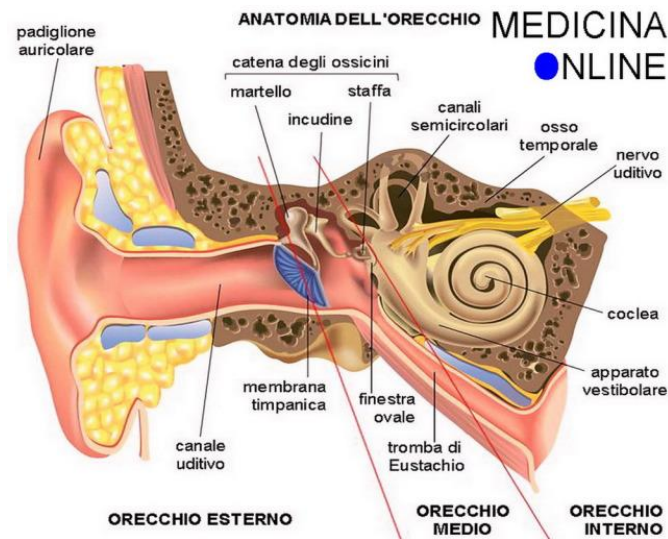
Rumore Rosso:

Il **rumore rosso** presenta un'ulteriore accentuazione della presenza di basse frequenze. Esso suona come un rimbombo molto basso (un treno che passa nel metrò, il rumore di un motore che fa vibrare, per risonanza, le pareti di una stanza) e si colloca alla soglia in inferiore delle frequenze udibili. È utilizzato negli effetti acustici musicali da film.

Lezione 5-6-7

PSICOACUSTICA

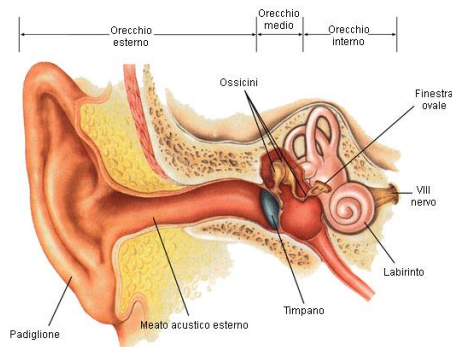
Il sistema uditivo umano ha una struttura complessa e svolge funzioni notevolmente avanzate. Non solo è in grado di elaborare un ampio insieme di stimoli, ma può identificare precisamente l'altezza o il timbro di un suono, o la direzione da cui esso proviene. Molte funzioni del sistema uditivo vengono svolte dall'organo che chiamiamo orecchio, ma grande enfasi di recente viene attribuita alla elaborazione che ha luogo nel sistema nervoso centrale. L'orecchio è spesso diviso in tre parti principali: orecchio esterno, orecchio medio ed orecchio interno.



L'orecchio esterno è formato dalla pinna esterna e dal canale uditivo (meato). L'orecchio medio inizia con la membrana del timpano, alla quale è attaccato il primo dei tre ossicini (chiamati martello, incudine e staffa) che compongono questo stadio. Il compito di questi ultimi è quello di amplificare il moto del timpano (essi formano un sistema di leve) e di trasferirlo ad un'altra membrana, la finestra ovale. Con la finestra ovale inizia l'orecchio interno, formato principalmente dalla coclea. La coclea contiene i meccanismi per trasformare le variazioni di pressione in corrispondenza del timpano in impulsi nervosi che vengono interpretati dal cervello come suono.

COCLEA

La coclea ci aiuta a percepire l'altezza (pitch) di un suono puro. Questo organo è costituito da una cavità divisa in due da una membrana (membrana basilare) e contenente un liquido incompressibile (perilinfo) che passa da una parte all'altra della membrana attraverso un'apertura. Le vibrazioni trasmesse dall'orecchio medio alla finestra ovale si trasmettono al fluido del dotto cocleare, che a sua volta provoca un moto della membrana basilare simile a quello di una bandiera. Lungo la membrana basilare, circa 30000 recettori nervosi (cellule ciliate) convertono il moto della membrana in segnali che sono a loro volta trasmessi ai neuroni del nervo acustico.



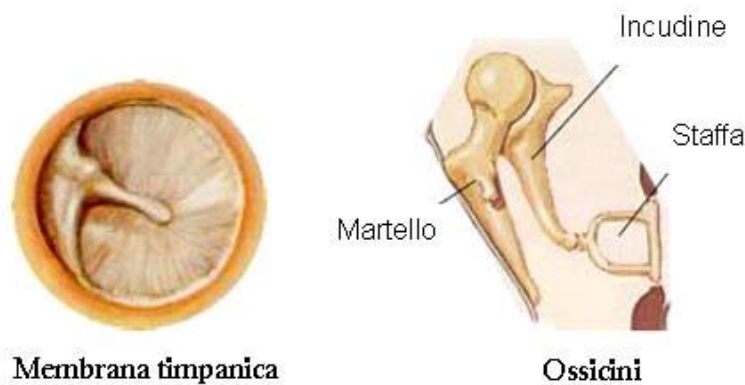
Ossicini

Gli **ossicini** sono, in ordine dal timpano all'orecchio interno (dall'esterno all'interno del cranio), il martello, l'incudine e la staffa.

- Il **martello** si articola con l'incudine ed è attaccato alla membrana timpanica che trasmette il moto alla catena degli ossicini.
- L'**incudine** è collegata ad entrambi gli altri ossicini.
- La **staffa** è articolata con l'incudine ed è collegata alla membrana della *finestra ovale*, l'apertura ellittica od ovale esistente fra l'orecchio medio ed il vestibolo dell'orecchio interno.

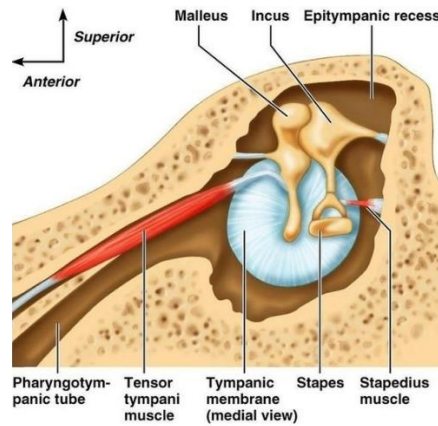
Quando le onde sonore fanno vibrare la membrana timpanica (timpano), questa mette in movimento il manico del martello, a cui è solidamente legata; la testa del martello conseguentemente vibra a sua volta, trasmettendo il suono agli altri due ossicini. La vibrazione della base della staffa attraverso la finestra ovale mette in movimento la perilinfa del vestibolo e quindi della coclea, causando il movimento delle ciglia delle cellule sensorie qui presenti.

Gli ossicini consentono alle vibrazioni del timpano di trasmettersi senza riduzione di forza alla finestra ovale. Tuttavia, la portata dei movimenti degli ossicini è controllata (e regolata) da alcuni muscoli ad essi connessi (il *tensore timpanico* e lo *stapedio*).



Muscolo Stapedio

Lo **stapedio** è il più piccolo dei muscoli striati del corpo umano e ha un aspetto le cui fibre convergono su un tendine centrale. Con appena un millimetro di lunghezza, il suo scopo è quello di stabilizzare l'osso più piccolo del corpo: la staffa. Lo stapedio si trova nell'orecchio medio ed ha il compito di prevenire un eccessivo movimento della staffa, aiutando così il controllo dell'ampiezza delle onde sonore dall'ambiente esterno all'interno dell'orecchio. È paragonabile al ruolo del muscolo tensore del timpano, che affievolisce primariamente quei suoni associati con la masticazione.



Soglie di Tolleranza del Rumore



Riguardo alla tutela della salute e sicurezza nei luoghi di lavoro, "in base alle conoscenze attuali, non si considera a rischio l'esposizione al rumore fino a 80 decibel. Il riferimento normativo fondamentale in materia di prevenzione dai rischi da rumore è il **decreto legislativo 81/2008** che prescrive i "requisiti minimi per la protezione dei lavoratori contro i rischi per la salute e la sicurezza derivanti dall'esposizione al rumore durante il lavoro". In particolare fissa **"tre livelli di esposizione** (80, 85 e 87 decibel ossia il valore medio, ponderato in funzione del tempo, dei livelli di esposizione al rumore per una giornata lavorativa) e i corrispondenti adempimenti ai quali sono tenuti i datori di lavoro qualora vengano superati i livelli stessi.

Se il livello di esposizione è superiore agli **80 decibel**:

- "messa a disposizione dei DPI dell'udito, da parte del datore di lavoro;
- obbligo di formazione e informazione dei lavoratori in merito ai rischi provenienti dall'esposizione al rumore, alle misure adottate, ai DPI dell'udito, all'uso corretto delle attrezzature, al significato del ruolo del controllo sanitario e della valutazione del rumore".

Se il livello di esposizione è superiore agli **85 decibel**:

- "obbligatorietà dell'utilizzo dei DPI dell'udito;
- obbligo di formazione e informazione dei lavoratori in merito ai rischi provenienti dall'esposizione al rumore.
- controllo sanitario.

Se il livello di esposizione è superiore agli **87 decibel**:

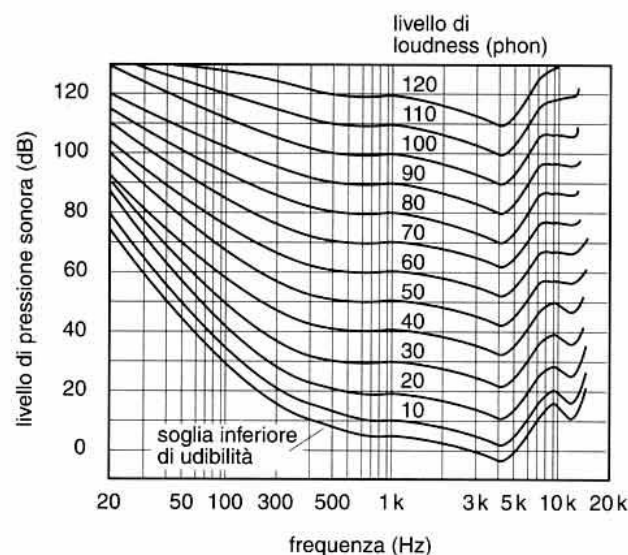
- "adozione immediata di misure atte a riportare l'esposizione al di sotto dei valori limite di esposizione (Dispositivi di Protezione e/o interventi su attrezzature, strutture o ambienti);
- individuazione delle cause dell'esposizione eccessiva;
- modifica delle misure di protezione e di prevenzione per evitare che la situazione si ripeta".

Il suono - Percezione umana

Le grandezze percettive con cui di solito si descrive un suono sono **volume**, **altezza** e **timbro**. Il **volume** di suono è il correlato percettivo diretto dell'intensità sonora. La **frequenza** è il principale responsabile della sensazione di altezza di un suono; per i suoni complessi l'altezza percepita dipende direttamente dalla frequenza fondamentale del suono, ma esistono casi in cui tale frequenza non è presente nello spettro, e viene ricostruita completamente dalla nostra percezione (si parla di frequenze fantasma). Infine, il **timbro** ha come correlato fisico diretto lo spettro o la forma d'onda del suono nel dominio della frequenza o del tempo.

Diagramma di Fletcher-Munson

L'intensità sonora è una grandezza fisica che, oggettivamente, misura il flusso di energia trasportata dall'onda sonora. Tale grandezza non descrive correttamente però l'intensità percepita (in inglese **loudness**) in quanto essa dipende in modo decisivo dalla frequenza del suono, e in misura minore, anche dal timbro del suono stesso. L'intensità percepita presenta quindi un legame complesso con l'intensità sonora. Per descriverla adeguatamente si è pensato di ricorrere alla rappresentazione mediante curve isofoniche che riportano, al variare della frequenza, il luogo geometrico dei punti per i quali l'intensità percepita è costante. Esse vengono ottenute chiedendo all'ascoltatore, mentre ascolta suoni di diversa frequenza, di regolare la manopola del volume in modo da percepirli con la stessa intensità. Questo compito viene rappresentato dall'audiogramma normale di **Fletcher-Munson**, è il frutto di studi compiuti su un gran numero di individui aventi un udito normale e privo di difetti. È limitato inferiormente da una curva che corrisponde alla **soglia di udibilità** e superiormente da tre tipi di curve dette: soglia del disturbo, soglia del dolore e soglia del danno uditivo che naturalmente non è determinabile sperimentalmente. La zona tra la soglia di udibilità e quella del dolore è detta **area di udibilità normale**. Per tracciare l'audiogramma sono stati effettuati dei confronti tra una serie di toni puri, sparsi in tutta la banda delle frequenze, e un tono puro di riferimento con frequenza di 1000 Hz.



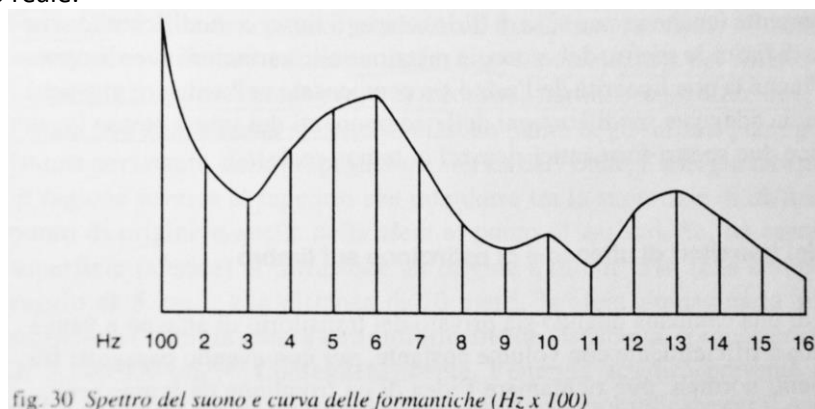
Ogni curva fornisce un valore diverso del livello di intensità soggettiva, la cui unità di misura è il phon. Le isofoniche sono graficamente abbastanza simili tra loro ma non sono mai sovrapponibili e la loro forma mostra la forte non linearità che caratterizza l'udito umano. Dall'audiogramma si nota come l'orecchio umano sia più sensibile alle medie frequenze fra qualche centinaio e qualche migliaio di Hz.

Altezza di un Suono

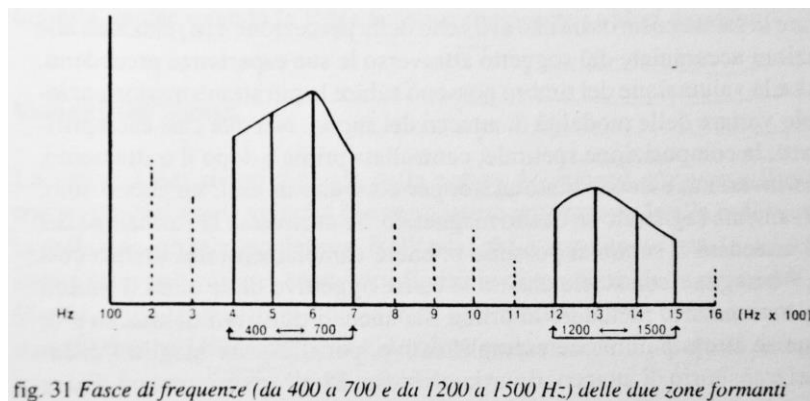
L'altezza è la frequenza fondamentale di una nota musicale o suono che viene percepita, ed è una delle caratteristiche principali di un suono. L'altezza indica se un suono è acuto piuttosto che grave e dipende dalla frequenza dell'onda sonora che lo ha generato. In particolare: più la frequenza di un'onda sonora è elevata e più il suono ci sembrerà acuto, mentre più è bassa la frequenza e più il suono ci apparirà grave. Nonostante la frequenza fondamentale reale possa essere determinata con una misura fisica, essa può differire dall'altezza percepita per via degli ipertoni (sono le componenti di un suono complesso dotate di una frequenza superiore a quella della fondamentale) e degli armonici naturali del suono (Gli **armonici naturali** sono una successione di suoni le cui frequenze sono multipli di una nota di base, chiamata fondamentale.). Il sistema di percezione uditiva umano può avere anche difficoltà a distinguere differenze di altezza fra le note, in alcune circostanze.

Timbro percepito -Le formanti delle vocali

Il **timbro**, è la qualità che, a parità di frequenza, distingue un suono da un altro. Il timbro dipende dalla forma dell'onda sonora, determinata dalla sovrapposizione delle onde sinusoidali caratterizzate dai suoni fondamentali e dai loro armonici. Dal punto di vista della produzione del suono, il timbro è determinato dalla natura (forma e composizione) della sorgente del suono e dalla maniera in cui questa viene posta in oscillazione. Mentre le altezze sonore, le durate e le intensità possono essere "quantificate" e ordinate lungo una [scala](#) perché grandezze fisiche oggettive e misurabili, ciò non avviene per il timbro che è una grandezza multidimensionale. Oggi si tende ad accertare nello spettro del suono la presenza di fasce di frequenze di maggior ampiezza, che sono messe in risalto da fenomeni di risonanza propri, ma ben definiti, della sorgente sonora e che prendono il nome molto significativo di **formanti**: la loro considerazione ha dato vita alla teoria formantica. Secondo la teoria formantica del timbro quest'ultimo dipende in modo prevalente dalla presenza più o meno cospicua delle zone formanti e dalla larghezza della fascia di frequenze abbracciata da ciascuna di esse. In questa figura è data un'esemplificazione delle zone formanti riguardanti un caso reale:



La curva segnata sopra le componenti dello spettro del suono descrive le zone formanti inerenti al caso, le quali interessano due fasce di frequenze che possono essere così delimitate (nella figura sotto).



Ciò significa che nei suoni emessi da quella sorgente sonora si ha un'esaltazione più o meno accentuata di tutte le componenti che vengono a trovarsi nell'ambito delle due zone formantiche.

Il timbro è influenzato anche dall'intensità del suono, in quanto una variazione apprezzabile di questa può esaltare o attenuare in modo non uniforme, rispetto alle caratteristiche uditive, i vari gruppi delle componenti. Anche l'altezza influisce sul timbro: lo spostamento di una stessa situazione armonica in una regione o in un'altra del campo di udibilità, può infatti alterare l'effetto in funzione della diversa sensibilità dell'orecchio al variare della frequenza.

Mascheramento link

(<https://fmilotta.github.io/teaching/computermusic/Projects/ComputerMusic-Project-05c-2018-IT.pdf>)

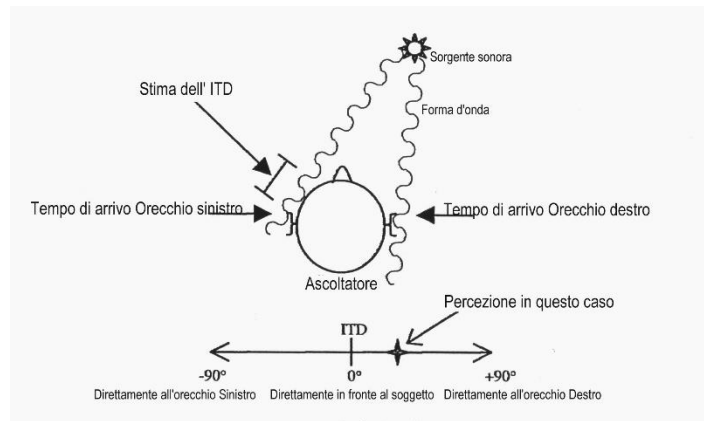
Localizzazione delle sorgenti sonore

Il meccanismo fondamentale si basa sul fatto (un po' come nella visione) di possedere due orecchie e non una sola. Tramite l'ascolto binaurale, cioè con entrambe le orecchie, il nostro sistema percettivo è in grado di confrontare le caratteristiche fisiche del suono che perviene alle due orecchie e di ricavare, da tale confronto, informazioni sulla posizione della sorgente che l'ha generato.

Per chiarire la strategia che il nostro sistema percettivo utilizza per individuare la direzione di provenienza di un suono, immaginiamo di avere un altoparlante che genera un suono e un ascoltatore che ruota la testa in modo da avere il suo orecchio destro più vicino all'altoparlante di quello sinistro.

In tale situazione si verificano due effetti:

- essendo l'orecchio destro più vicino del sinistro esso raccoglierà, tramite i padiglioni auricolari, il suono in anticipo. La differenza dei tempi di arrivo del suono viene chiamata **ITD** (acronimo dell'inglese Interaural Time Difference). Una stima di tale grandezza si ottiene dividendo il maggior percorso che deve compiere il suono per arrivare all'orecchio più lontano (la "larghezza della testa") per la velocità del suono in aria. Ruotando la testa si può fare in modo che l'ITD si annulli (o comunque scenda al di sotto del minimo valore rilevabile).
- L'orecchio più lontano si trova nella "zona d'ombra" creata dalla testa e riceve il suono con intensità minore. Tale differenza in intensità viene chiamata **IID** (acronimo dell'inglese Interaural Intensity Difference). Elaborando l'IID il sistema uditivo riceve ulteriori informazioni sulla direzione di provenienza del suono. Il più piccolo valore di IID che il nostro sistema uditivo può apprezzare è dell'ordine di 1 dB.



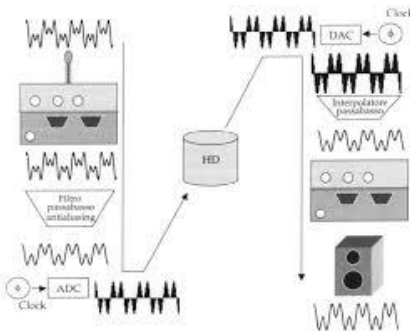
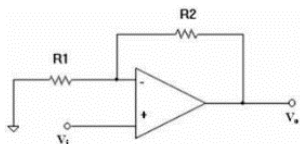
Lezioni 8

Come dalle lezioni precedenti, abbiamo detto che l'audio è un flusso informativo, avendo il suono, come caratteristica, una natura temporale. Esso si può dividere in *analogico* e *digitale*. Ma non solo l'audio ma anche l'informazione elettronica ovvero i segnali stessi. In generale "analogico" vuol dire che è tempo continuo e a valore continui, invece "digitale" vuol dire che è a tempo discreto (periodo finito) e a valori discreti (dati da numeri).

Audio analogico:

L'audio è detto analogico quando la sua rappresentazione consiste di forme d'onda continue. In questo tipo di audio si ha la variazione naturale delle ampiezze, ma a tale variazione viene associata una corrispondente variazione dei valori di tensioni. Ci sorge un dubbio, *come si ottiene a partire da un suono ?*

Abbiamo bisogno in primo luogo di un **trasduttore** è un dispositivo di qualsiasi genere destinato a convertire una grandezza fisica in un'altra. In sostanza esso converte le onde di pressione generate dal suono in grandezze elettriche. Successivamente un circuito elettronico chiamato **preamplificatore** riceve in entrata segnali analogici di potenza molto bassa e li rimette in uscita amplificati a sufficienza per poter essere accettati dagli ingressi di altri dispositivi elettronici come gli amplificatori operazionali. Poi verranno amplificati da un **amplificatore**, in figura è mostrato un amplificatore operazionale non invertente. Alla fine, verranno registrati e copiati su supporti fisici come il disco in vinile o audiocassetta.



Il disco in vinile è un supporto per la memorizzazione analogica dei segnali sonori. Ufficialmente introdotto nel 1948, è stato l'evoluzione del disco in gomma lacca da 78 giri. Esattamente come il suo antenato, il disco in vinile presenta un solco a spirale (inciso a partire del bordo esterno) in cui è presente la traccia analogica del suono registrato.

L'audiocassetta è un supporto fonografico a *nastro magnetico*. È composta da due bobine rinchiusi in un contenitore di plastica che raccoglie il nastro su cui si registra l'audio e da cui si

riproduce per mezzo della cosiddetta *piastra*. Quindi sfrutta la variazione dell'intensità dei campi magnetici.

Dopo queste operazioni verrà letto, preamplificato, amplificato e infine riprodotto.

Nella trasduzione di solito si verifica un'alterazione dei parametri originali. Questa alterazione si traduce in distorsione. Tale distorsione si propagherà fino alla riproduzione del suono. Essa è misurabile attraverso un indice di qualità del sistema detto SNR (Signal Noise Ratio), chiamato anche rapporto segnale-rumore è una grandezza numerica che mette in relazione la potenza del segnale utile rispetto a quella del rumore in un qualsiasi sistema di acquisizione.

$$SNR = \frac{P_{segnale}}{P_{rumore}} \quad \text{con } 0 \leq SNR < \infty \quad \text{L'SNR è misurabile anche in decibel} \quad \rightarrow \quad SNR_{db} = 20 \log_{10} \frac{S}{N}$$

Questo parametro deve essere il più alto possibile.

L'audio analogico in generale ha dei pro e contro:

Pro:

- La registrazione analogica su nastro è migliore di quella digitale, perché l'onda sonora, sotto forma di corrente alternata, viene registrata sul nastro in maniera continua.
- Rappresentazione di molte frequenze.
- Apparecchiature poco sofisticate come il disco in vinile e l'audiocassetta.

Contro:

- Le particelle metalliche che ricoprono la superficie del nastro sono piuttosto pigre e non si spostano dal loro stato di quiete a meno che il campo magnetico della testina di registrazione non sia abbastanza forte da polarizzarle. Quindi i suoni più deboli di intensità non vengono incisi correttamente, così come le alte frequenze;
- I transienti sono avvenimenti molto rapidi, dinamici e ricchi di alte frequenze, qualità che il nastro non riesce a soddisfare in pieno;
- Il rumore di fondo è una caratteristica intrinseca del nastro, alla quale si è cercato di ovviare con filtri di pre-enfasi e circuiti per la riduzione del rumore, che aggiungono a loro volta ulteriore distorsione;
- Deterioramento del materiale che avviene ad ogni passaggio e rischio di autocancellazione del materiale.
- Dipendenza del supporto fisico.

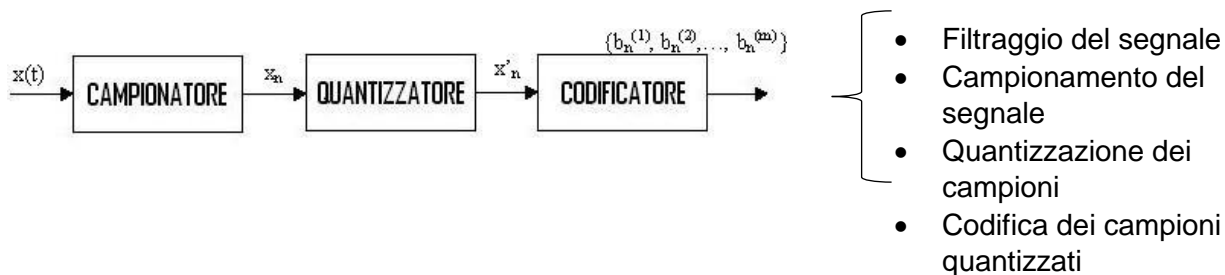
Audio digitale:

L'audio digitale viene trattato dagli elaboratori a differenza del primo.

Questo audio, a differenza dell'analogico, non imita più la curva d'ampiezza ma attraverso dei campioni, determinati dal teorema del campionamento chiamato pure teorema di Nyquist- Shannon, permette di assegnare dei valori che rappresentano, in quel momento, la curva d'ampiezza.

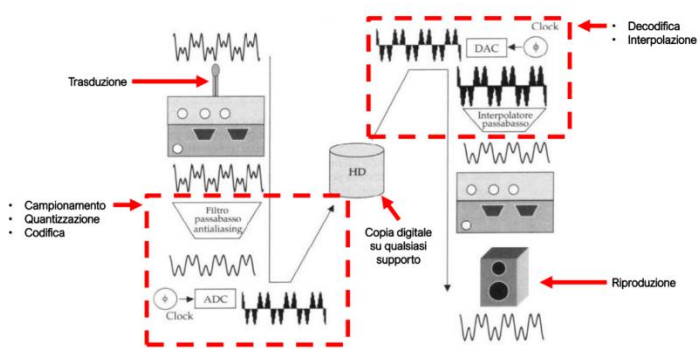
Anche qui come ottenere un audio digitale a partire da in suono ?

Come prima, abbiamo bisogno di un trasduttore, e successivamente di un ADC ovvero un convertitore analogico-digitale. Questo dispositivo permette di effettuare una conversione da valori e tempi continui a valori e tempi discreti attraverso 4 fasi:



Infine, viene rappresentato in un formato specifico e archiviato in una memoria di massa, come ad esempio un CD-ROM.

Invece, come ottenere un suono a partire da un audio digitale ? (Ovvero il percorso inverso).



In primo luogo, viene interpretato il formato di memorizzazione e il suo output viene inviato a un DAC ovvero un convertitore digitale-analogico. Esso trasforma il segnale a valori e tempo discreti in a valori e tempo continui, quindi si ha un flusso audio digitale in input (da una memoria o trasferito attraverso una rete) in un segnale analogico da inviare in output ad una periferica audio (ad es. un set di altoparlanti

o cuffie audio) per essere riprodotto/trasdotta in un segnale sonoro. Alla fine, le variazioni della grandezza elettrica vengono interpretate da un dispositivo (ad esempio casse audio), per produrre una vibrazione, ossia la sorgente dell'onda sonora.

Rivediamo la **conversione analogico-digitale**.

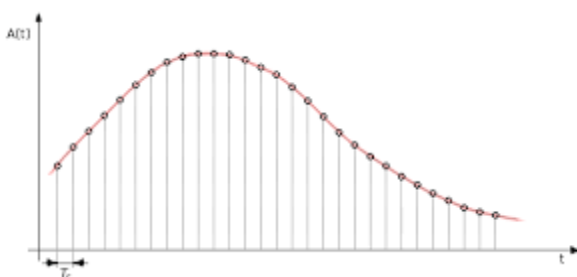
Filtraggio:

Si utilizza un filtro anti-alias che è un filtro analogico utilizzato prima del campionamento di un segnale, al fine di restringere la banda del segnale stesso per soddisfare approssimativamente il teorema del campionamento di Nyquist-Shannon.

Campionamento:

In questa fase si trasforma il segnale dal tempo al tempo discreto. In sostanza il campionamento consiste nell'andare a "sentire" (misurare, registrare) il valore del segnale analogico in diversi istanti di tempo o dello spazio. Il teorema che stabilisce quale sia la frequenza minima di campionamento con una determinata caratterizzazione in frequenza (trasformata di Fourier) affinché il segnale analogico possa essere ricostruito a valle a partire da quello discreto in input è il **teorema del campionamento** o teorema di Shannon-Nyquist che enuncia:

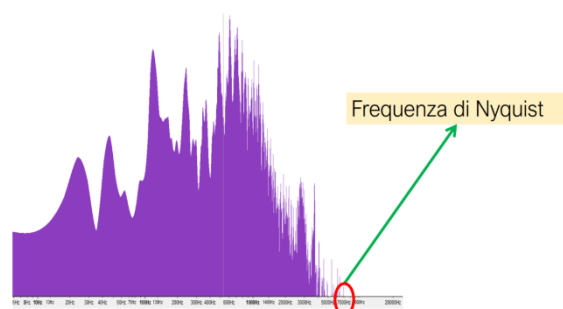
Al fine di poter ricostruire fedelmente un segnale campionato, è necessario che il tasso di campionamento sia almeno il doppio della frequenza di Nyquist, quindi: $f_c > 2f_N$.



Dove f_m è la massima frequenza dello spettro del segnale da campionare, dato per certo che sia limitato in banda e periodico. Detta pure **frequenza di Nyquist**.

Dove f_c è la frequenza con cui si prendono i campioni prende il nome di tasso di campionamento.

A quanto detto precedentemente bisogna aggiungere che per conoscere la frequenza di Nyquist



basta utilizzare la Serie di Fourier per ottenere lo spettro. Invece per segnali non periodici si può agire utilizzando la Trasformata di Fourier. Mentre, per segnali a banda non limitata (spettro non limitato), si è costretti ad usare un filtro passa basso per riportarsi ad un segnale a banda limitata.

È auspicabile rimuovere tutte le frequenze indesiderate al fine di ricostruire fedelmente il

segnale originale. Qualora ci fossero si utilizza un passabasso che elimina tutte le frequenze superiori a quelle di Nyquist.

Qualora si campionasse al di sotto della f_N , si verrebbe a creare il fenomeno dell'*aliasing*. L'*aliasing*, o distorsione da campionamento lento, da sottocampionamento, è il fenomeno per il quale due segnali analogici diversi possono diventare indistinguibili una volta campionati, cioè si causa la sovrapposizione dei segnali. Esso si verifica se si campiona al di sotto di $2 f_N$.

Dopo aver annunciato il teorema, possiamo intuire che più prendiamo alta la frequenza di campionamento meglio è, ma in realtà non conviene farlo perché si avrebbe sia un costo che uno spreco maggiore.

Allora, qual è la giusta frequenza di campionamento ? La giusta frequenza corrisponde al più del doppio della massima frequenza percepibile dagli umani che è di 20kHz. Quindi si campiona a 44,1 kHz.

Quantizzazione:

È un processo di discretizzazione nel dominio temporale (ascissa del riferimento cartesiano) definito campionamento. Cioè si passa da valori continui a valori discreti.

Codifica:

Assegnazione di segni o simboli convenzionalmente designati per rappresentare un'informazione, che nel nostro caso sono i campioni.

L'audio multicanale l'audio presentante più flussi informativi elettronici ognuno dei quali rappresentante un diverso flusso informativo sonoro destinato ad essere riprodotto da un diverso diffusore acustico, per esempio l'altoparlante, posizionato nell'ambiente d'ascolto diversamente rispetto a tutti gli altri e secondo regole prestabilite. Essi possono essere combinati per migliorare l'esperienza e la sensazione sonora.

L'audio multicanale può essere suddiviso in mono, se è presente un canale, stereo, se ve ne sono due.

Coloro che sono i padri dell'informazione sono Claude Elwood Shannon e Harry Theodor Nyquist.

- https://it.wikipedia.org/wiki/Claude_Shannon.
- https://it.wikipedia.org/wiki/Harry_Nyquist

Lezione 9

Quantizzazione

Un segnale analogico può essere visto come una tensione elettrica i cui valori variano in un certo periodo di tempo con continuità.

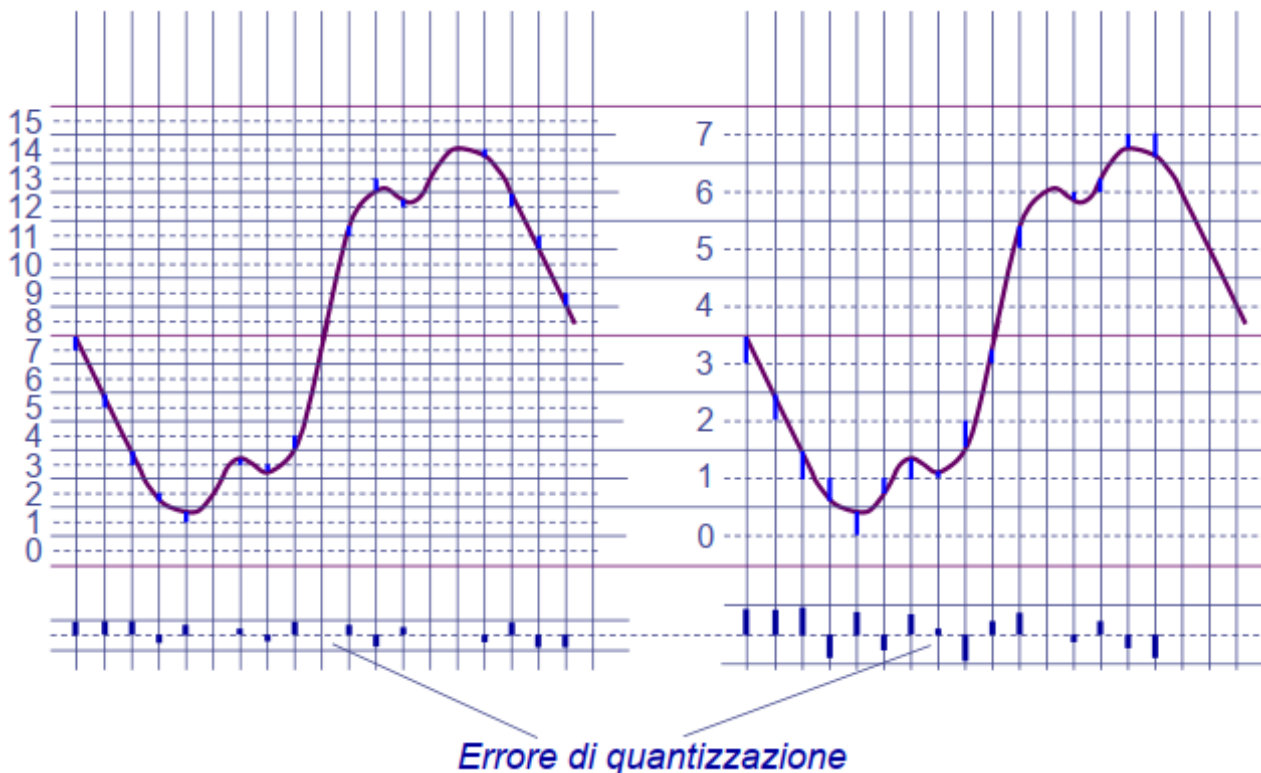
L'informazione di tipo digitale è invece caratterizzata da sequenze di valori binari di una certa lunghezza (parole binarie);

i valori binari vengono detti "bit", e le parole binarie costituite da n bit (ovvero di lunghezza n) possono assumere 2^n configurazioni diverse; ad esempio se $n=2$ le parole binarie possono essere 2^2 : 00, 01, 10, 11.

Nel segnale analogico il limite in precisione è determinato dalla percentuale di rumore presente. Nell'informazione digitale invece il limite è strettamente legato al numero di bit utilizzato per rappresentare i campioni del segnale: con due bit, si possono rappresentare quattro valori; con tre bit, otto valori; con quattro bit, sedici valori; e così via.

La **quantizzazione** è appunto l'operazione di assegnare valori discreti per descrivere un segnale continuo; più bit vengono usati, tanto più è accurata la quantizzazione effettuata. Questo perché nella quantizzazione il rapporto segnale/rumore è proporzionale al numero di bit adoperati con una riduzione del "rumore di quantizzazione" in misura di 6 dB per ogni bit adoperato. Più sono i gradini, minore sarà l'errore di quantizzazione (o rumore)

Esempio di Differenza tra quantizzazione a 4 bit e 3 bit



Nella conversione analogico-digitale, la differenza tra il segnale analogico reale e il valore digitale quantizzato dello stesso viene chiamata **errore di quantizzazione** o **distorsione di quantizzazione**. Questo errore è dovuto all'arrotondamento o al troncamento del segnale digitalizzato. Il segnale errore è

talvolta considerato come un segnale casuale aggiuntivo chiamato **rumore di quantizzazione** a causa del suo comportamento stocastico (aleatorio).

Il **rumore di quantizzazione** è il modello usato per descrivere l'errore di quantizzazione nella conversione analogico-digitale nei sistemi di telecomunicazioni e nell'elaborazione numerica dei segnali. Si tratta un errore di arrotondamento tra la tensione analogica d'ingresso e il valore digitale all'uscita, di tipo non lineare e dipendente dal segnale in ingresso.

Esistono diversi modelli del rumore di quantizzazione. In un convertitore analogico-digitale ideale l'errore di quantizzazione è uniformemente distribuito tra $-\frac{1}{2}LSB$ e $\frac{1}{2}LSB$ (**LSB** indica il bit meno significativo del quantizzatore) mentre il segnale ha una distribuzione uniforme lungo tutti i livelli di quantizzazione. In queste condizioni, il rapporto segnale/rumore di quantizzazione (SQNR) può essere calcolato come:

$$SQNR = 20 \log_{10}(2^Q) \cong 6.02 \cdot Q \text{ dB}$$

dove Q è il numero di bit usati nella quantizzazione.

Quando il segnale che si presenta all'ingresso è un'onda sinusoidale' la distribuzione del segnale non è più uniforme e la corrispondente equazione risulta:

$$SQNR \approx 1,761 + 6.02 \cdot Q \text{ dB}$$

Il **dithering**, nella elaborazione numerica di segnali, è una forma di rumore con una opportuna distribuzione, che viene volontariamente aggiunto ai campioni con l'obiettivo di minimizzare la distorsione introdotta dal troncamento nel caso in cui si riquantizzino i campioni stessi. Il dithering viene usato abitualmente nell'elaborazione di segnali video e audio campionati e quantizzati.

Quantizzazione Lineare

La dimensione degli intervalli di decisione in cui si suddivide l'insieme dei possibili valori è sempre la medesima. Ovunque cada (all'interno della dinamica) il valore quantizzato, l'accuratezza dell'operazione sarà sempre la stessa. Un esempio di quantizzazione uniforme è la PCM.

PCM

La **Pulse-Code Modulation** (letteralmente "modulazione a codice di impulsi", abbreviato PCM), in elettronica e nelle telecomunicazioni, è un metodo di rappresentazione digitale di un segnale analogico. Il metodo utilizza un campionamento dell'ascissa del segnale a intervalli regolari; i valori letti vengono poi quantizzati in ordinata ed infine digitalizzati (in genere codificati in forma binaria). La PCM è ampiamente utilizzata nei sistemi di telefonia, ma si basano su questo principio anche molti standard video. Poiché la PCM pura richiede un **bitrate** molto elevato, gli standard video di consumo come DVD o DVR sono basati su sue varianti che fanno uso di tecniche di compressione. Molto frequentemente, la codifica PCM viene impiegata per facilitare le trasmissioni digitali in forma seriale.

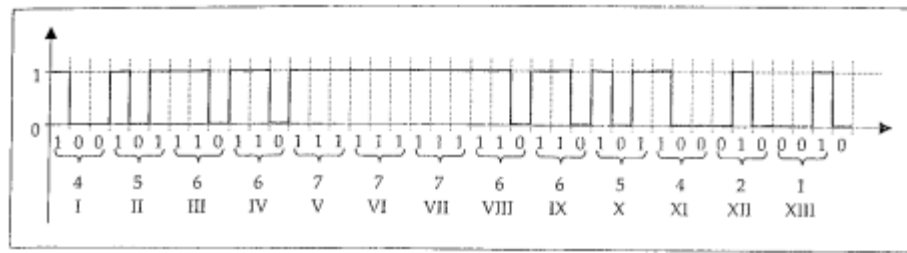


Figura 2.10 Modulsione del codice di impulso (PCM)

Bit Di Parità

La codifica PCM è usata in tutti i settori dell'archiviazione e della trasmissione digitale dei dati. Per irrobustire la codifica si aggiungono extra che sono utilizzati per il controllo di eventuali errori nella codifica. Un esempio molto semplice di controllo degli errori è un bit di parità: in questo caso si aggiunge un bit ad una sequenza fissata di bit (una o più parole binarie). Il significato del bit extra può essere il seguente:

- Se nella sequenza il numero di bit a 1 è pari, allora il Bit verrà apposto a 0;
- Se nella sequenza il numero di bit a 1 è dispari, allora il bit verrà a posto a 1;

Nel caso in cui un Bit per errore subisce un Flip (varia da 0 a 1 o viceversa), il bit di parità permette di scoprirlo.

Ad esempio, supponiamo che ci siano 5 bit a 1 nella sequenza. Il bit di parità viene posto a 1. Se uno qualsiasi dei bit cambia polarizzazione (sia da 1 a 0 che da 0 a 1), allora si ha un numero pari di 1 e si ha un'inconsistenza tra la sequenza di Bit e il Bit di parità. (si noti che è anche possibile che l'errore sia nel bit di parità).

Quantizzazione Non Lineare

Una soluzione alternativa per evitare la degradazione dei segnali deboli è data da una spaziatura non uniforme delle regioni di quantizzazione, una quantizzazione non lineare. Benché il numero delle regioni continui a rimanere lo stesso (in quanto dipende dal numero di bit), la dimensione delle regioni di quantizzazione diminuisce per le regioni delle ampiezze deboli, e aumenta di conseguenza per le regioni delle ampiezze forti. Quindi, l'errore di quantizzazione decresce per le ampiezze deboli e aumenta per le ampiezze forti: cioè, il guadagno in qualità alle ampiezze deboli è compensato da una perdita di qualità alle ampiezze forti. Per la quantizzazione non lineare, il rapporto SQNR può assumere forme diverse, in quanto l'ampiezza dell'errore varia con l'ampiezza del segnale il con il tipo di spaziatura scelto.

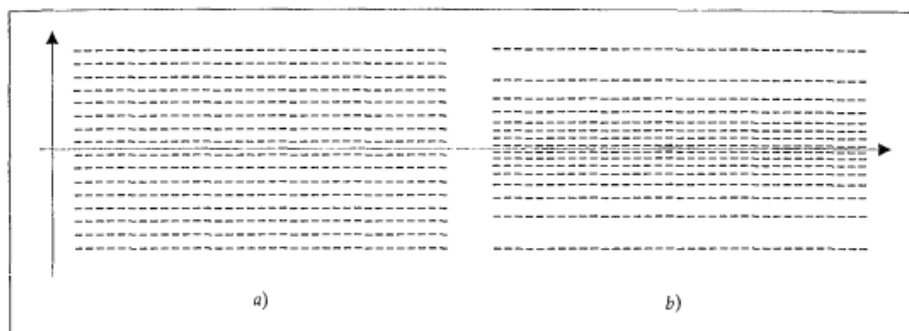


Figura 2.11 Quantizzazione lineare (a) e non lineare (b)

La quantizzazione non lineare si comporta in ambito digitale come una compansion: L'idea è essenzialmente quella di ridurre il rumore alle ampiezze deboli, ammettendo invece che aumenti alle ampiezze forti, dove comunque viene mascherato. Questo risultato viene raggiunto mediante una spaziatura non uniforme delle regioni di quantizzazione: le regioni vicino alle ampiezza nulla vengono quantizzate in modo molto più fine delle regioni e ampiezza elevate, un esempio è la quantizzazione logaritmica che si ritrova molto spesso in alcuni formati audio, la caratteristica fondamentale è che la dimensione delle regioni di quantizzazione crescono con l'ampiezza del segnale secondo una curva logaritmica.

Lezione 10 Digitalizzazione

Equalizzatori

Con **equalizzazione**, nel campo del trattamento dei segnali audio, si indica l'operazione di livellamento delle varie parti dello spettro di frequenze che compongono un segnale audio.

L'equalizzazione è una tecnica di trattamento dei segnali audio e rappresenta il procedimento di filtraggio a cui è sottoposto un segnale audio per variarne il contenuto timbrico.

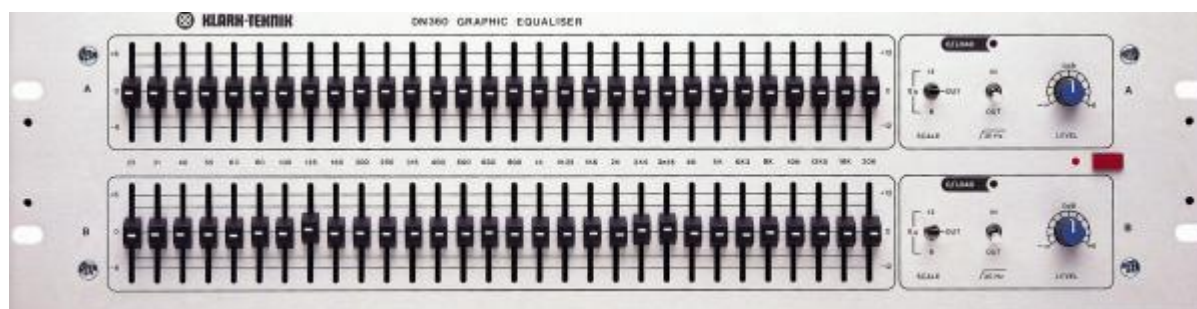
I campi di applicazione dell'equalizzazione sono molteplici:

- correzione timbrica durante la ripresa di suoni
- correzione durante la fase di post-produzione della timbrica di segnali registrati
- eliminazione / riduzione di fruscii, rumori ricorrenti
- creazione di suoni nuovi, manipolati o simulati (simulazione di una telefonata in uno spettacolo)

A seconda della tipologia di progettazione, gli equalizzatori si dividono essenzialmente in **equalizzatori grafici** ed **equalizzatori parametrici**.

Gli **Equalizzatori Grafici** dividono l'intero spettro musicale in più porzioni e l'utente può regolare per mezzo dei regolatori a cursore il livello di quella determinata porzione, che corrisponderà ad esempio ai bassi, oppure medio-bassi, ecc ecc.

L'equalizzatore grafico di solito si trova come hardware esterno e non è altro che una catena di vari filtri peaking con frequenza e bandwidth fissa in cui possiamo scegliere solo il guadagno o l'attenuazione. È composto nella maggior parte dei casi da 31 o 15 bande fisse ma possiamo trovare alcune eccezioni come l'API 560, classico equalizzatore da studio a 10 bande, ricreato in digitale anche da Waves. (sono più **user friendly**)



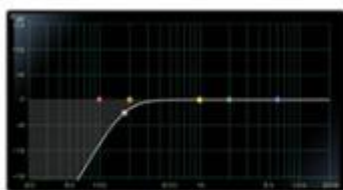
Gli **Equalizzatori Parametrici** sono più complessi, ma permettono di effettuare correzioni evitando particolari danni al segnale, essendo utilizzati particolarmente nelle riprese dei suoni in spettacoli dal vivo, dove gli errori e le pesanti correzioni timbriche possono essere dannosi.

Un equalizzatore parametrico è in media dotato nei mixer analogici di 4 filtri invece di 7, 15 o ancora 31 come nei tipi grafici, ma ognuno di questi filtri può essere regolato nei suoi parametri, da cui il nome.

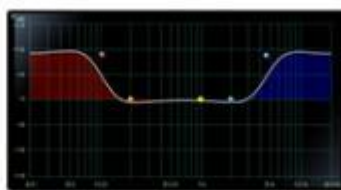
L'equalizzatore parametrico è lo strumento più potente con cui possiamo lavorare ed è quello presente nella maggior parte dei mixer hardware e dei plugin di equalizzazione.

È composto da diversi tipi di filtri che si dividono in:

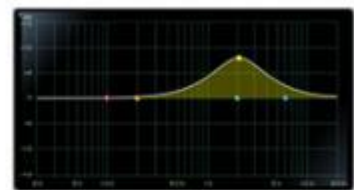
- **Filtri Passa Alto (HPF) e Passa Basso (LPF):** sono filtri che tagliano da una certa frequenza in giù (nel caso del filtro passa alto) o in su (filtro passa basso). L'attenuazione di ciò che viene tagliato dipende dalla pendenza del filtro, che di solito può essere 6, 12, 18 o 24 dB per ogni ottava. Nel mondo analogico di solito abbiamo la sola selezione della frequenza ma in digitale spesso abbiamo anche la possibilità di scegliere la pendenza e a volte l'enfatizzazione nel punto di taglio (resonance).
- **Filtri Shelving:** utilizzati per alte e basse frequenze, enfatizzano o attenuano dalla frequenza selezionata in giù (Low Shelving) o in su (High Shelving). Hanno quindi solitamente 2 parametri: frequenza selezionata e dB di guadagno/attenuazione. A volte è presente un terzo parametro che controlla la pendenza nel punto di guadagno o attenuazione.
- **Filtri Parametrici o Peaking Filter:** sono filtri che enfatizzano o attenuano la frequenza selezionata. Quanto le frequenze intorno a quella selezionata verranno attenuate o enfatizzate dipenderà dal parametro di bandwidth o Q: più alto sarà il valore di Q più sarà stretta la campana di guadagno e attenuazione. I parametri in questo caso saranno 3: frequenza, guadagno/attenuazione e larghezza della campana. Se uno di questi 3 parametri manca, il filtro viene chiamato **semiparametrico**.



Filtro Passa Alto (HPF)



High and Low Shelving



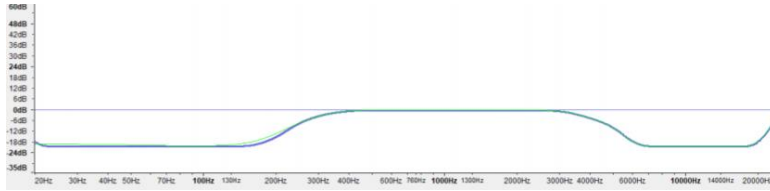
Peaking Filter

Alcuni filtri parametrici hanno un valore di **Q** fisso al variare del guadagno o dell'attenuazione (Costant Q), altri invece hanno larghezza di banda variabile (Proportional Q) che si stringe all'aumentare del guadagno o attenuazione. Questo in modo da aumentare la precisione in caso di tagli o guadagni su specifiche frequenze.

Certi filtri hanno delle campane non simmetriche in caso di taglio o attenuazione, come il peaking filter del **Renaissance Equalizer di Waves**, dove la campana si stringe in caso di taglio, in modo da rimuovere frequenze precise, e si allarga in caso di guadagno in modo da non essere troppo risonante su una specifica frequenza.

Lezione 11

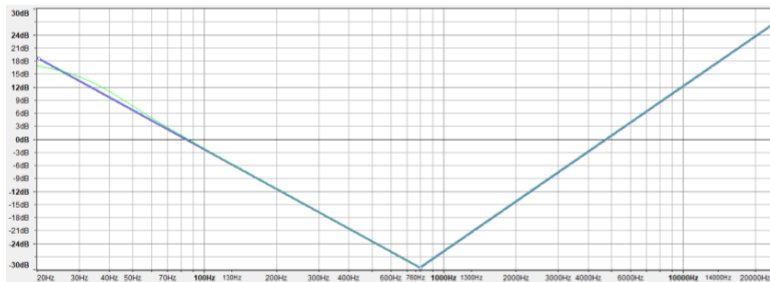
In frequenza vi sono moltissime tipologie di filtri, tra questi vi sono il filtro *telefono*, filtro *midcut*, filtro *midpass*.



300Hz a 4000Hz.

Filtro telefono:

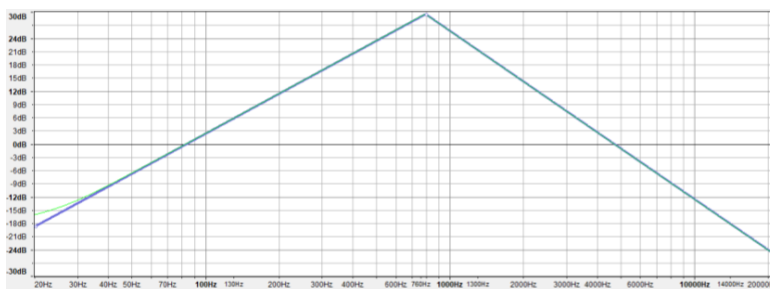
Come si evince, esso attenua sia le alte che le basse frequenze e in sostanza attenua tutte le frequenze diverse dall'intervallo che va da



questo filtro si ottiene un suono un po' vuoto e metallico.

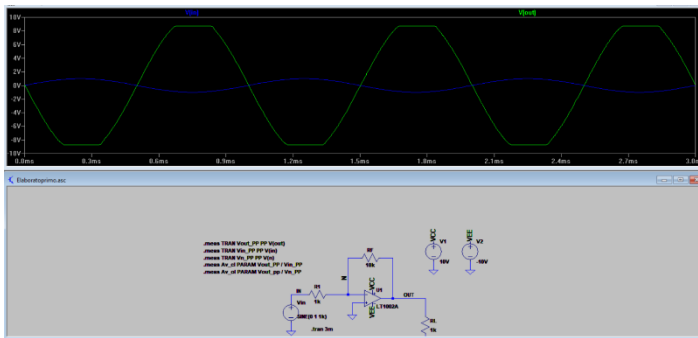
Filtro midcut:

Il filtro midcut o a "V", dove per midcut si intende "taglio dei medi". Banalmente si intuisce che ha una forma a V e amplifica le basse e alte frequenze. Esso non annulla le frequenze medie, ma le attenua mentre alcune le annulla. Applicando



Filtro midpass:

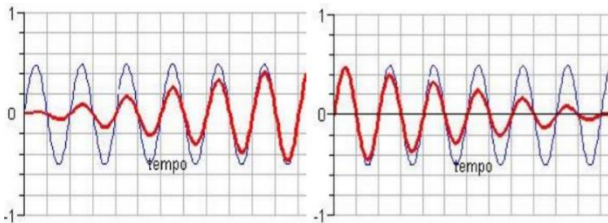
Speculare del midcut, esso attenua le basse e alte frequenze e amplifica quelle centrali.



A volte può capitare quando un amplificatore è spinto a produrre un segnale in uscita più elevato della tensione di alimentazione, amplificherà il segnale solo fino alla massima capacità e a quel punto il segnale non può essere amplificato ulteriormente. Graficando la tensione di un segnale in questo caso, si verificherà il fenomeno del **clipping**. In questo caso il segnale saturerà fino a un massimo data

dalla tensione d'ingresso.

In questa figura possiamo vedere un esempio di amplificatore invertente in cui è evidente il fenomeno del clipping, attraverso la simulazione su LTSPICE.



Di solito in una traccia audio sono presenti i fenomeni di *fade-in* & *fade-out*. Essi sono rispettivamente l'aumento graduale di un segnale audio e la dissolvenza, ossia la riduzione graduale di un segnale audio.

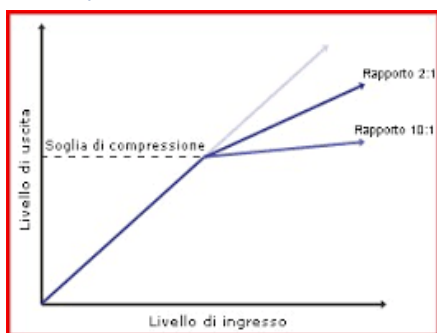
Adesso diamo la definizione di **equalizzatore**: L'equalizzatore è uno strumento per bilanciare le frequenze di un segnale.

Esso stesso agisce sia sul range dinamico sia sull'amplificazione al fine di "sistemare" le frequenze.

Con l'equalizzatore abbiamo detto che possiamo agire sull'amplificazione, allora possiamo effettuare il processo di **Normalizzazione**. Tramite questo processo si può effettuare la massima amplificazione possibile senza introdurre distorsione digitale ovvero il clipping.

Sul range dinamico, che è la differenza in dB tra il massimo volume che può sostenere il componente senza distorsioni e il rumore di fondo che esso produce, sono possibili delle operazioni, rispettivamente: Compressore, Limitatore, Espansore, Limita Rumore.

Compressore:



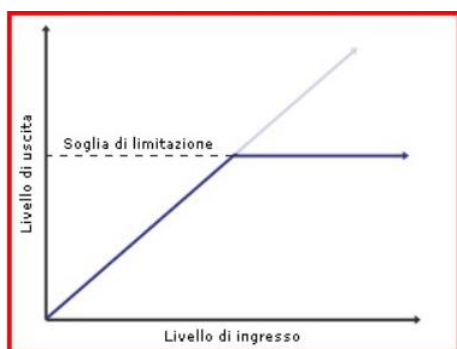
La compressione può essere descritta, in termini semplici, come un'operazione di controllo automatico del volume, cioè una riduzione della gamma dinamica di un segnale. I suoni forti che superano una certa soglia vengono ridotti di volume, mentre i suoni deboli non vengono modificati sotto la soglia o sono aumentati sotto essa. Il compressore modificherà tutti i valori a disposizione, successivamente sarà applicata una *Normalizzazione*.

L'aumento risultante del livello medio può essere utile per l'audio riprodotto in un ambiente rumoroso, come in un'auto o nel parlato, per rendere la voce distante più rumorosa di quella ravvicinata.

Esso, tipicamente, consta di 6 parametri: soglia, ratio, attacco, rilascio, knee, Make-Up Gain.

Vedesi: <https://tecnologiamicale.wordpress.com/2013/07/14/il-mastering-volume-dinamica-e-processor-di-dinamica/>.

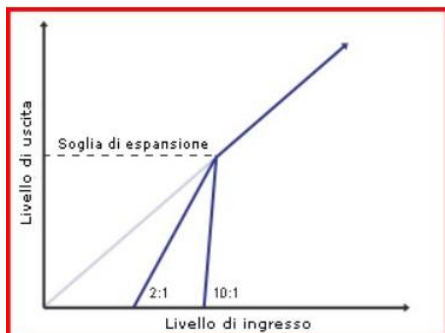
Limitatore:



Molto spesso si ha a che fare con un segnale che è abbastanza costante come livello ma che occasionalmente presenta rapidi picchi di modulazione che portano il sistema a saturare o a distorcere. Esempi di questi picchi di modulazione sono i colpi di arma da fuoco o i suoni generati dai timpani in un'orchestra. Per correggere tali fluttuazioni di segnale è necessario un limitatore, cioè un apparato che opera come un normale amplificatore per segnali al di sotto di un determinato livello prefissato, ma che diventa un compressore per segnali che superano tale livello. Quindi tutti i valori di ampiezza superiori ad una soglia vengono ridotti al

valore della soglia stessa e a differenza del compressore non modifica tutti i valori, ma solo quelli oltre soglia. Può essere anticipato da un'amplificazione o normalizzazione.

Espansore:

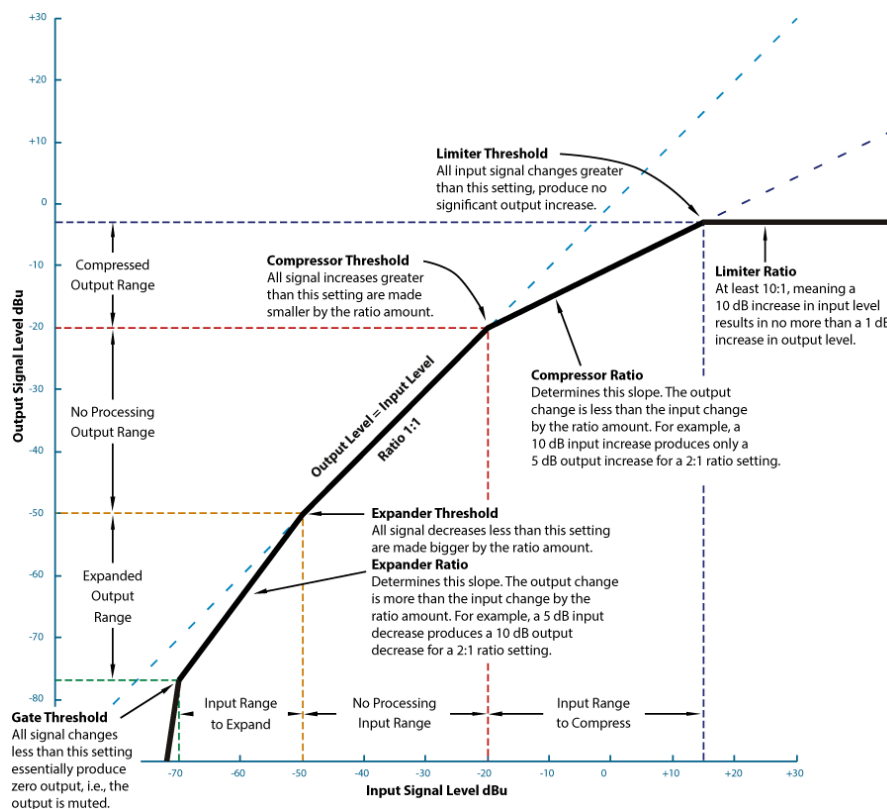


Gli espansori svolgono l'operazione inversa dei compressori; essi rendono più tenui segnali già deboli e più forti segnali già forti. La pendenza delle rette è sempre maggiore di quella a 45° di un amplificatore lineare. Aumenterà la gamma dinamica del segnale quindi aumenta le ampiezze al di sopra della soglia e diminuirà quelle negative sotto la soglia. Intuiamo che modificherà tutti i valori.

Limita Rumore (Noise gate):

Esso ha l'effetto opposto del Limitatore, infatti tutti i valori in ampiezza inferiori ad una soglia vengono aumentati al valore della soglia. Esso modificherà solo i valori sotto la soglia. I limita rumore attenuano i segnali di una quantità fissa, nota come range. Un limita rumore consente al segnale principale di passare solo quando è al di sopra di una soglia impostata: in questo caso il gate è 'open'. Se il segnale scende al di sotto della soglia, nessun segnale può passare (o il segnale è sostanzialmente attenuato): il gate è 'chiuso'. Un noise gate viene utilizzato quando il livello del "segnale" è sopra il livello del "rumore" indesiderato.

RIASSUNTO DELLE OPERAZIONI SUL RANGE DINAMICO.

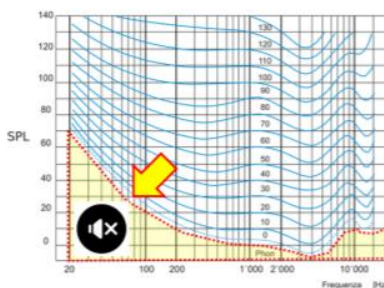


Lezione 12

Quante volte ci è capitato di dover “zippare” qualcosa. Magari per rendere più piccoli i dati da inviare e soprattutto ridurre lo spazio! Invece di mandare mega e megabyte di roba basterebbe solo qualche kilobyte. Premesso ciò, questo discorso si può traslare anche nel mondo audio. Allora chiediamoci: *Perché dovremmo comprimere ?*

A metà anni 80 ci fu un rapido successo dei CD audio che portò alla luce il successo dell'audio digitale. Ma essi richiedevano grandi tassi di trasferimento dati. Così nacque l'esigenza di iniziare a comprimere. Vantaggi:

1. **Riduzione di tempi e costi di trasmissione.**
2. **Riduzione dello spazio di memoria occupato.**



Nel comprimere una traccia audio poter effettuare una compressione del silenzio, per far ciò possiamo usare la tecnica lossy. Questa tecnica permette la compressione dati che porta alla perdita di parte dell'informazione originale durante la fase di compressione/decompressione dei dati che la rappresentano. Dal diagramma di Fletcher-Munson notiamo che alcuni segnali non sono udibili per cui attraverso una soglia posso essere interpretati come silenzio, questo perché il vero e proprio silenzio si avrebbe a 0 dB.

Riprendendo quanto detto nella prima parte, possiamo definire come “Size”, il numero di bit necessari per rappresentare il segnale senza comprimere. Naturalmente lo vorremmo il più piccolo possibile e per far ciò utilizziamo le tecniche di compressione.

$$\text{Size} = f_c \times N \times D \times C$$

durata

Dove f_c è la frequenza di campionamento, N la profondità in bit, D la del flusso audio e C il numero di canali.

Inoltre, il **bit rate** o **bps** è la velocità di trasmissione dei bit su un canale di trasmissione in un dato intervallo di tempo.

$$\text{Bps} = f_c \times N \times C$$

I metodi della compressione permettono di abbassare il bit rate preservando una buona qualità.

Vediamo altri algoritmi per la compressione.

μ -law:

L'algoritmo μ -law è un algoritmo di compressione, utilizzato principalmente nei sistemi di telecomunicazione digitale PCM a 8 bit nel Nord America e in Giappone. È una delle due versioni dello standard G.711. Comunque, la codifica μ -law, grazie a una quantizzazione logaritmica, offre una codifica di (bit-per-campione) /8 a 1. Quindi essa permette di raggiungere una gamma dinamica pari a quella ottenuta da una quantizzazione lineare di 13-14 bit. Infatti, la quantizzazione logaritmica non opera una spaziatura uniforme sulla gamma dinamica, come quella lineare, ma i campioni di ampiezza debole, ovvero quelli intorno allo zero, vengono quantizzati con maggiore accuratezza.

In sostanza questi algoritmi di compressione riducono la gamma dinamica di un segnale audio. Nei sistemi analogici, ciò può causare un aumento del rapporto segnale-rumore (SNR) raggiunto durante la trasmissione mentre nel dominio digitale, può ridurre l'errore di quantizzazione.

Aggiungiamo che la quantizzazione non lineare assicura maggior precisione alle ampiezze più basse con soli 8 bit.

Vediamo la sua espressione per la codifica.

$$Y = \begin{cases} 128 + \frac{127}{\ln(1+\mu)} \times \ln(1+\mu|x|) & x \geq 0 \\ 127 - \frac{127}{\ln(1+\mu)} \times \ln(1+\mu|x|) & x < 0 \end{cases}$$

Dove $\mu = 255$, e x è il valore del segnale in input normalizzato tra -1 e +1.
L'obiettivo di questa formula è condurre i valori dall'intervallo $[-32768, 32767]$ all'intervallo $[0, 255]$.
 Y appartiene all'intervallo $[0, 255]$.

Analizziamola in dettaglio.

Il logaritmo naturale di questa somma va da 0 a 5.545, mentre il rapporto $127/\ln(1+\mu)$ è una costante e vale 22,9. Di conseguenza il secondo addendo va da 0 a 127.

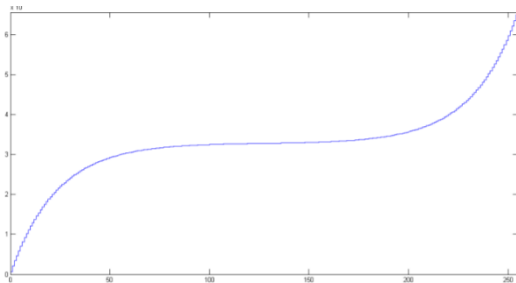
Questa formula comprime i campioni a 16 bit in modo non lineare su campioni da 8 bit.

Se volessimo decodificare ?

$$x = \begin{cases} \frac{\exp\left(\frac{128+Y}{127} \times \ln(1+\mu)\right) - 1}{\mu} & Y \geq 128 \\ \frac{\exp\left(\frac{127-Y}{127} \times \ln(1+\mu)\right) - 1}{\mu} & Y < 128 \end{cases}$$

Applichiamo questa formula, per tornare ai valori a 16 bit.
Se facessimo un esempio numerico noteremmo che i valori ottenuti dalla decompressione non siano esattamente quelli codificati. Infatti, questa è

una tecnica **Lossy**.



Possiamo notare che gli estremi siano quantizzati in maniera meno precisa rispetto a quelli centrali.

Sull'asse sono presenti a 16 bit. Sull'asse orizzontale si trovano i corrispondenti interi a 8 bit ottenuti dalla codifica.

A-law:

L'algoritmo A-law è un algoritmo di compressione utilizzato nei sistemi di comunicazione digitale PCM a 8 bit europei per ottimizzare, ossia modificare, la gamma dinamica di un segnale analogico per la digitalizzazione. È una delle due versioni dello standard G.711 di ITU-T. La codifica A-law è in uso in Europa. Grazie alla quantizzazione non lineare, permette di ottenere con soli 8 bit la stessa qualità (es: SQNR) che si otterrebbe con una quantizzazione lineare a 13 bit.

$$Y = \text{sign}(X) \begin{cases} \frac{A|X|}{1+\ln A} & |X| < \frac{1}{A} \\ \frac{1+\ln A|X|}{1+\ln A} & \frac{1}{A} < |X| \leq 1 \end{cases}$$

X è un valore normalizzato tra -1 e 1, mentre A è un fattore pari a 87.7, mentre Y è codificato tra -1 e 1.

Mentre per decodificare ?

Sia Y il valore codificato di ampiezza tra -1 e 1 e A un fattore pari a 87.7.

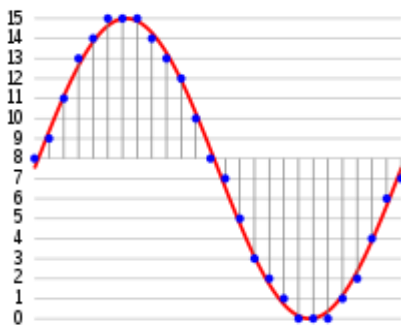
$$X = \text{sign}(Y) \begin{cases} \frac{|Y|(1+\ln A)}{A} & |Y| < \frac{1}{1+\ln A} \\ \frac{e^{|Y|(1+\ln A)-1}}{A} & \frac{1}{1+\ln A} < |Y| \leq 1 \end{cases} \quad \text{Valgono le stesse considerazioni per la } \mu\text{-law.}$$

Esistono delle tabelle di conversione per passare da codeword di 14 a 8 bit (μ -law) e da 13 a 8 bit (A -law).

Abbiamo appena parlato delle due codifiche μ -law e A -law. Esse sono utilizzate per i servizi di telefonia digitale su linee ISDN. L'idea da utilizzare è che il segnale sia sufficientemente compresso da poter essere inviato su tali linee. Queste linee consentono un traffico di 64 kbps codificato in tempo reale. Il tasso di campionamento nelle applicazioni di telefonia è di 8000 camp/sec, che funziona perfettamente nelle frequenze di 0-4 KHz, dove si propaga la voce umana.

Esistono altre codifiche come: PCM, DPCM, ADPCM.

PCM:



La codifica PCM o Pulse Code Modulation utilizza un campionamento dell'ascissa del segnale a intervalli regolari; i valori letti vengono poi quantizzati in ordinata ed infine digitalizzati (in genere codificati in forma binaria). La PCM è ampiamente utilizzata nei sistemi di telefonia, ma si basano su questo principio anche molti standard video, come l'ITU-R BT.601. Poiché la PCM pura richiede un bitrate molto elevato, gli standard video di consumo come DVD o DVR sono basati su sue varianti che fanno uso di tecniche di compressione. La lunghezza delle parole binarie dipende dai bit di quantizzazioni

utilizzati. Se uso 4 bit si possono assumere solo valori tra 0 e 15.

DPCM:

La codifica DPCM o Differential Pulse Code Modulation è una codifica di segnale che utilizza la linea di base della modulazione PCM ma aggiunge alcune funzionalità basate sulla predizione dei campioni del segnale. L'input può essere sia analogico o digitale. Si può usare in due modi.

La prima opzione è prendere i valori di due campioni consecutivi, se sono analogici si quantizzano, e calcolarne la differenza. L'output sarà tale differenza.

La seconda opzione consiste non più di prendere la differenza dei campioni ma di prendere la differenza relativa all'output di un modello locale del processo di decodifica. La differenza può essere anche quantizzata. Per i segnali audio questo tipo di codifica riduce il numero di bit richiesti per campione di circa il 25% rispetto al PCM. La DPCM viene calcolata attraverso l'espressione:

$$\sum_{i=1}^N C_i * [x(k-i)T_c] \quad \text{Dove } T_c \text{ è il tempo di campionamento.}$$

ADPCM:

La codifica ADPCM o Adaptive Differential Pulse Code Modulation è una variante del DPCM che rende arbitraria la dimensione dei passi di campionamento, così da consentire ulteriori riduzioni della larghezza di banda richiesta per un dato rapporto segnale-rumore. Quindi oltre a codificare le differenze, tale codifica, utilizza un meccanismo di predizione che, unito a un algoritmo di riquantizzazione, permette la sistemazione, ovvero si adatta alle differenze da codificare. Per la predizione del campione successivo si somma ± 1 al campione predetto attuale e si sceglierà la predizione più vicina al valore effettivo.

In caso di differenze ridotte c'è un vantaggio effettivo nel codificare differenze piuttosto che le codeword stesse, casomai si possono utilizzare delle Look-Up Table (LUT). Queste sono generalmente array, usate per sostituire operazioni di calcolo a runtime con una più semplice operazione di consultazione.

Lezione 13

COMPRESSIONE .

Entropia percettiva

Combinando le nozioni di mascheramento psicoacustico e i principi di quantizzazione, è possibile definire una misura dell'informazione percettivamente rilevante in un segnale audio. Uno studio di Johnston (dei Bell Labs) ha fissato una misura del limite teorico alla comprimibilità di un segnale, se si vuole ottenere una codifica trasparente." Tale limite è di 2.1 bit/campione. Quindi, una compressione con un bit-rate di 64kbps (con frequenza di campionamento a 44.1 kHz) non può essere una codifica trasparente, in quanto ogni campione verrà codificato con $64.000/44.100 = 1,45$ bit/ campione. La stima dell'entropia percettiva è stata condotta su un'ampia varietà di materiale sonoro con qualità CD con un algoritmo che ricalca a grandi linee lo standard MPEG. In realtà, occorre precisare che l'obiettivo della compressione non include sempre la trasparenza del segnale codificato. Spesso le necessità imposte dalla trasmissione via rete di un segnale o l'inserimento dell'audio in un prodotto multimediale su CD-ROM superano le richieste di trasparenza. Perciò i parametri di compressione e l'allocazione dei bit dipendono dal bit-rate ricercato, che può garantire o meno la trasparenza del segnale compresso. Se si riesce a produrre il bit-rate necessario a rappresentare il mascheramento completo del rumore e della distorsione, lo schema di compressione produrrà una codifica percettivamente trasparente, cioè il segnale decodificato sarà indistinguibile dal segnale originale.

Block coding

La quantizzazione non uniforme può essere vista molto facilmente come una codifica di tipo virgola mobile. La codifica si divide in due parti, la mantissa e l'esponente. La mantissa è il valore effettivo del campione, mentre l'esponente dice di quante posizioni a sinistra deve essere spostata la mantissa (lasciando dei bit a zero) per avere il valore assoluto corretto su una scala PCM lineare. L'esponente è quindi un fattore di scala, un po' come per la tabella dei passi di quantizzazione nella codifica IMA-ADPCM (in quel caso, la mantissa è l'output del Quantizer, e l'esponente è l'indice sulla tabella dei passi di quantizzazione). Poiché i segnali audio reali non variano molto tra un campione e il successivo, si può assumere che lo stesso esponente valga per un buon numero di campioni. Un risparmio consistente si ottiene quindi mantenendo lo stesso esponente per tutto un blocco di campioni, in modo da memorizzarlo solo una volta: all'interno del blocco, si sceglie il campione che necessita del maggior numero di bit di esponente, e una volta determinato tale numero di bit, questo viene mantenuto per tutto il blocco. Questo approccio è detto codifica per blocchi, o all'inglese block coding.

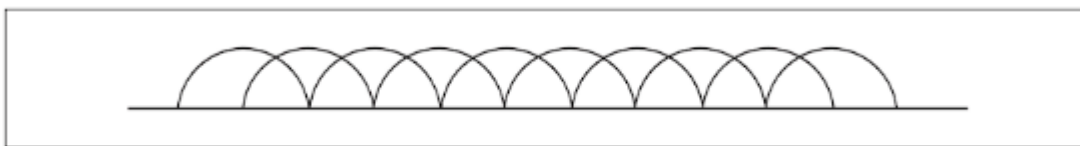
Sub-band coding

Nella codifica a sottobande la rappresentazione è realizzata applicando il segnale all'ingresso di un banco di filtri passabanda (filtri di analisi), e codificando separatamente le componenti spettrali ottenute. Tali componenti, opportunamente compresse, sono applicate in fase di decodifica ad un banco di filtri che ricostruiscono il segnale (filtri di sintesi).

Transform coding

Il segnale audio si può rappresentare sia nel dominio del tempo che nel dominio della frequenza. Sebbene siamo in genere abituati a pensare il segnale audio nel dominio del tempo, si può notare come un segnale audio cambi molto lentamente nel dominio della frequenza o dominio trasformato (dal momento che la

rappresentazione nel dominio della frequenza si ottiene da un'operazione di trasformata a partire dalla rappresentazione nel dominio del tempo). Il dominio della frequenza è inoltre vantaggioso per discriminare tra bande in cui si hanno comportamenti diversi dal punto di vista della gamma dinamica e del mascheramento: infatti, se la gamma dinamica è limitata in modo assoluto o per effetto del mascheramento è sufficiente un numero di bit inferiore rispetto al segnale originale per ottenere lo stesso rapporto segnale/rumore (SNR). Il problema con le trasformate è la presenza dei transitori, di durata variabile a seconda dei segnali, che rende difficoltosa l'accuratezza della discretizzazione nel tempo. La soluzione è di "finestrare" il segnale in segmenti molto brevi e quindi applicare la trasformata a ogni segmento di campioni (o blocco). Poiché la divisione netta in segmenti potrebbe introdurre artifici nei punti estremi della segmentazione (vedi il problema dei pre-echi), di solito si fa in modo di sovrapporre i segmenti in modo che ciascun campione appartenga a due finestre, anche se con ruoli differenti in quanto nella finestra più a sinistra si troverà nella metà finale e nella finestra più a destra nella metà iniziale. Come abbiamo accennato nel Capitolo 2, nel calcolo della RFT, la trasformata discreta di Fourier, la Soluzione in frequenza, cioè la dimensione del periodo di campionamento sulle frequenze, è uguale al numero di campioni della finestra. Nelle finestre sovrapposte si ha quindi il doppio dei campioni necessari; inoltre richiede un grande sforzo di computazione. Questi problemi sono stati risolti con l'uso della DCT, trasformata discreta del coseno (Discrete Cosine Transform), in alcuni casi, come nel caso di MP3. La DCT è una forma della Trasformata Discreta di Fourier (DFT), ottenuta mediante un raddoppiamento dei campioni audio a sua volta prodotto da una concatenazione speculare degli stessi campioni.



Hybrid coding

La rappresentazione a sottobande può essere integrata con la rappresentazione basata su trasformata. In tale approccio, tipicamente, il segnale è preliminarmente rappresentato in sottobande e successivamente ciascuna sottobanda è trasformata e compressa. Questo tipo di rappresentazione è adottata nello standard MPEG-1 Layer III, Also Known As (AKA) MP3.

Schema generale di tipo percettivo

Il segnale di input $X[n]$ viene segmentato in frame (fotogrammi, segmenti, finestre, blocchi) quasi-stazionari, che durano da 2 a 50 millisecondi (analisi nel tempo). Per "segmenti quasi stazionari" si intende che, essendo il segmento molto breve, al suo interno il segnale non presenta in pratica un inviluppo di ampiezza, cioè non contiene transitori. La porzione di segnale all'interno di ciascun blocco viene quindi analizzata nel dominio della frequenza (analisi nella frequenza). Questa analisi viene condotta o da una trasformata unica (transform coding) o da un banco di filtri che suddivide il segnale in sotto-bande di frequenza (sub-band coding). I meccanismi di compressione di tipo percettivo oscillano tra una codifica per sotto-bande e una codifica per trasformata, con molti casi di concomitanza delle due codifiche; in ogni caso, il segnale verrà rappresentato nel dominio della frequenza. Allo stesso tempo, il segnale viene anche sottoposto a un'analisi psicoacustica, che può eventualmente beneficiare della divisione in sotto-bande. Questa analisi beneficia di una trasformata molto più definita dell'analisi precedente. L'analisi psicoacustica opera una misura della distorsione percettiva, producendo le soglie di mascheramento che permettono di determinare il massimo

ammontare di distorsione armonica che si può introdurre nel segnale durante la fase di riquantizzazione, senza provocare delle alterazioni udibili nel segnale ricostruito. In particolare, il modello psicoacustico calcola il rapporto SMR (Signal-to-Mask Ratio), cioè il rapporto tra l'ampiezza del segnale e la soglia di mascheramento. I rapporti SMR, calcolati per blocco di frequenze, sono utilizzati per decidere quanti bit vengono allocati per la riquantizzazione del segnale in ogni banda, in modo da minimizzare il rumore di quantizzazione. La fase di quantizzazione e codifica usa i bit allocati per stabilire le dimensioni delle regioni di quantizzazione (fattori di scala sull'esponente) e codificare il segnale. Le ridondanze ancora presenti nel segnale vengono rimosse mediante schemi di compressione semplici di tipo lossy (come DPCM e ADPCM) e lossless basati sulla ridondanza statistica. Infine, la codifica elaborata viene formattata per produrre il bitstream: in questa fase si altera l'ordine delle informazioni per limitare gli errori di trasmissione e la loro propagazione, e si introducono eventuali informazioni aggiuntive. Poiché il modello psicoacustico di controllo della distorsione è adattativo sul segnale, la maggior parte degli algoritmi di compressione è a tasso variabile di bit. La decodifica "scompatta" il bitstream, ricostruisce i valori quantizzati nel dominio delle frequenze, e infine trasforma i campioni di frequenza in un segnale audio

Lezione 14

MPEG



Il **Moving Pictures Experts Group (MPEG)** ha creato una serie di standard per la rappresentazione codificata di immagini in movimento, la loro compressione, trasmissione su reti digitali e archiviazione su media quali CD e DVD, principalmente con l'obiettivo di occupare meno banda rispetto al formato analogico tradizionale (oltre agli altri vantaggi dati dalla digitalizzazione). I vari standard MPEG non si occupano solo della codifica audio/video vera e propria, ma anche dell'organizzazione delle informazioni ausiliarie per la sincronizzazione di diversi contenuti, la gestione di stream multipli quali diverse tracce audio, le EPG (guide elettroniche), ecc..

Una caratteristica peculiare di questi standard è quella di *non definire il processo di codifica ma specificare quello di decodifica*. Questo ha garantito una maggior flessibilità nelle implementazioni senza compromettere l'interoperabilità. Infatti in questo modo ciascun costruttore che voglia realizzare un encoder MPEG ha la libertà di applicare i meccanismi definiti dallo standard in vari modi, purchè lo stream prodotto rispetti la sintassi ufficiale e risulti quindi decodificabile da un decoder standard.

MPEG-1

Lo standard MPEG-1, formalizzato in ambito ISO all'inizio degli anni '90, si è rapidamente diffuso grazie ad alcune circostanze favorevoli. In primo luogo è uno standard aperto, realizzabile da chiunque a partire dal documento di definizione dello standard, acquistabile dall'ISO. In secondo luogo esso è largamente documentato grazie alla disponibilità di software pubblico sviluppato durante i lavori di standardizzazione al fine di verificare e esplicitare le funzionalità della codifica e decodifica. In terzo luogo, esso prevede uno sbilanciamento della complessità a carico del codificatore, mentre il decodificatore è il più quanto possibile semplificato, così da abilitare la produzione di riproduttori (player) a basso costo. Infine, la standardizzazione è intervenuta tempestivamente rispetto all'avvento della digitalizzazione dei contenuti multimediali, anticipando e scoraggiando il proliferare di soluzioni proprietarie di interoperabilità necessariamente limitate. Lo standard di codifica audio consta di tre schemi di codifica (Layers) di complessità ed efficienza di compressione crescente, che sono qui di seguito sinteticamente descritti. In tutti i casi, i dati codificati sono organizzati in trame (frame) decodificabili indipendentemente una dalle altre, costituite rispettivamente dalla rappresentazione codificata di 384 campioni per il Layer-I e di 1152 per i Layer-II e III. Osserviamo che tali trame sono di lunghezza in bit variabile con il bit-rate. Ad esempio, per audio Layer-III MPEG audio campionato a 44.1KHz ogni trama rappresenta un intervallo di 26.12msec. Per un bit-rate 384 kbits/sec la trama corrisponde a 1.25 KBytes. Il bit-rate dell'audio codificato con MPEG-1 varia da un minimo di 32 Kb/s per segnale mono ad un massimo che dipende dalla complessità della codifica adottata e dalla qualità desiderata. In Tabella sono riportati i bit-rate medi tipicamente richiesti dai diversi Layer per la codifica percettivamente trasparente, ovvero caratterizzata da artefatti di compressione non percepibili, di segnale audio stereo; la qualità del segnale decodificato è pertanto quella di un CD audio.

MPEG-2

Lo standard MPEG-2 Audio prevede tanti algoritmi di codifica compatibili con quelli definiti da MPEG-1 (MPEG-2 Forward/Backward Compatible), quanto algoritmi di codifica avanzati (MPEG-2 Advanced Audio Coding). Tali algoritmi sono analoghi a quelli definiti in MPEG-1, ma con alcuni raffinamenti, principalmente riferibili a:

- aumento della risoluzione in frequenza: MPEG-2 AAC opera una codifica a trasformata MDCT che conduce ad una risoluzione massima in frequenza di 1024 componenti spettrali, che sono raggruppate tipicamente in 43 bande cui vengono assegnati passi di quantizzazione (scale factors) differenti;
- maggiore flessibilità di adattamento della finestra temporale alle caratteristiche del segnale, che può variare da 2048 a 256 campioni, con conseguente riduzione dei fenomeni di pre-eco;
- introduzione di tecniche predittive nel dominio del tempo e della frequenza; in particolare MPEG-2 AAC introduce una tecnica di adattatività temporale della quantizzazione che prende il nome di Temporal Noise Shaping [4.5].

MPEG-4

Rilasciato nel 1998 introduce significative novità, quale il concetto di oggetto (AVO, Audio Visual Object), che può definire sia flussi audio video (elementary stream simili a quelli dell'MPEG-1 e 2) che oggetti sintetici descritti con un linguaggio simile al VRML (Virtual Reality Modeling Language), ed anche quello di scena con cui si descrive il contesto di tali oggetti.

I bitrate per cui l'MPEG-4 è stato sviluppato vanno dai 64Kbps a oltre 52Mbps anche se le implementazioni attuali lo limitano e ottimizzano per bassi valori, al di sotto di 1 Mbps dove i precedenti mostrano i loro limiti e dove questo invece garantisce notevole qualità.

A differenza dell'MPEG-1 e 2, nell'MPEG-4 non è stato definito un formato di dati multiplexati come nelle sezioni System dei precedenti.

MPEG-7

(Multimedia Content Description Interface)

MPEG-7 è uno standard nato per codificare i contenuti multimediali; non è uno standard nato per codificare flussi audio o video come, ad esempio, MPEG-1 MPEG-2 o MPEG-4 ma per definire come sono organizzati i dati multimediali.

La codifica utilizza l'XML per memorizzare dei metadati che utilizzano a loro volta il timecode del filmato permettendo di sincronizzare i flussi multimediali con particolari eventi. Per esempio, permette di sincronizzare un filmato con i suoi sottotitoli o un video con il testo della canzone.

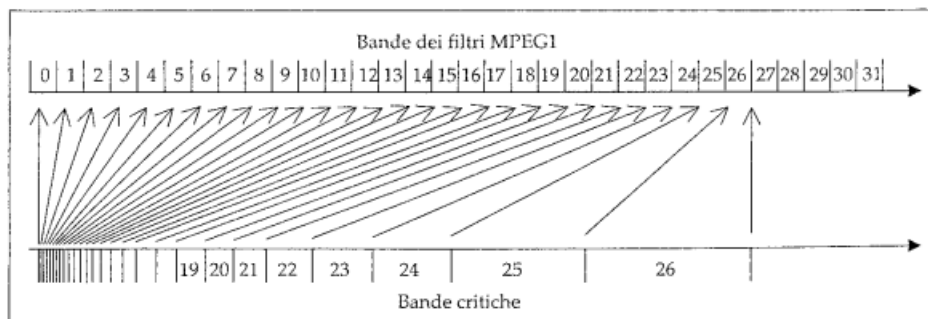
MPEG-21

MPEG-21 è nato per definire un framework per le applicazioni multimediali. È un progetto ancora nei primi stadi di vita, ricerca l'espansione e l'unificazione dei componenti dell'MPEG-4 e dell'MPEG-7 all'interno di un singolo framework che copra tutti gli aspetti. Lo standard comprende una protezione digitale del diritto d'autore, dei sistemi di pagamento, verifica e valutazione di qualità.

Una delle caratteristiche più importanti degli standard MPEG è il principio di minimizzare gli elementi normativi all'interno dello standard. In particolare, gli unici elementi che vengono imposti sono il formato dell'audio compresso e la struttura dell'algoritmo di decodifica. I moduli che compongono l'algoritmo di compressione MPEG audio sono lasciati quasi interamente alla creatività degli sviluppatori. Lo standard di compressione audio MPEG è diventato uno standard universale in vari settori, dall'elettronica di consumo alla

multimedialità. Lo standard MPEG1 (come pure MPEG2) lavora su tre livelli di compressione, detti layer. Il layer 1 è l'algoritmo di base; il layer 2 e 3 migliorano i moduli dello schema, producendo migliori compressioni al costo di incrementare la complessità sia dell'encoder che del decoder.

LAYER 1



L'analisi tempo-frequenza effettuata da layer 1 usa il banco di filtri che è alla base di tutti gli altri layer. Questo banco di filtri divide il segnale audio in 32 bande di frequenza della medesima larghezza. I filtri sono relativamente semplici e producono una buona risoluzione temporale e una ragionevole risoluzione in frequenza, se paragonata alle proprietà percettive del nostro sistema uditivo. La trasformata implementata dai filtri è la trasformata discreta del coseno (DCT), che è una variante della trasformata discreta di Fourier. La progettazione del Banco di filtri è un compromesso tra accuratezza ed efficienza. Come si nota in figura la larghezza uniforme delle bande risulta troppo ampia per le basse frequenze. Ad esempio, se il tasso di campionamento è di 48 kHz, le bande hanno una larghezza di 750 Hz, data da $24000/32$ (24.000 è la frequenza di Nyquist a 48 kHz). Quindi, l'allocatione dei bit non può essere modellata in modo fine sul mascheramento del rumore, che è molto influenzato dai limiti della banda critica. Di fatto alle basse frequenze, una banda del filtro include più bande critiche. Tuttavia, l'errore introdotto dal banco di filtri è di minima entità e non è udibile. I filtri adiacenti sul banco soffrono di sovrapposizione delle bande in modo significativo. Ciò avviene anche per le bande critiche nell'orecchio interno, ma nel caso del Banco di filtri una singola frequenza finisce per influenzare l'output di due filtri adiacenti in modo innaturale. Il banco di filtri produce 32 campioni di frequenza, un campione per banda, per ogni 32 campioni audio in input. A questo punto, per ciascuno delle 32 bande, si raggruppano assieme 12 campioni. Nel frattempo, il modello psicoacustico determina le soglie di mascheramento per ciascun blocco di 12 campioni. La fase di quantizzazione e codifica assegna a ogni blocco di 12 campioni un'allocatione di bit e un fattore di scala. I bit allocati per blocco vengono memorizzati perché il decoder sappia di quanti bit è fatto ogni campione. Ogni sottobanda può ricevere da 0 a 15 bit (0 bit si hanno ad esempio quando il mascheramento non fa udire nulla di una certa sottobanda). Il fattore di scala è un moltiplicatore che dimensiona i campioni di ogni sottobanda in modo opportuno rispetto all'intera gamma dinamica del quantizzatore. Il layer 1 formatta i 32 blocchi di 12 campioni (in tutto 384 campioni) in un frame. Oltre ai dati audio, il frame contiene un'intestazione, un codice di controllo degli errori di tipo CRC, ed eventualmente informazioni aggiuntive.

Layer 2

Il layer 2 apporta lievi miglioramenti alle prestazioni delle layer 1. I blocchi di campioni audio diventano più grandi e si impongono delle restrizioni sulla locazione di bit nel caso delle sottobande di media e alta frequenza. Nel frame, la rappresentazione dei bit allocati, fattore di scala e campioni quantizzati è resa più compatta. Questa compattazione permette di utilizzare più bit per i campioni.

I frame ora contengono 1152 campioni per canale audio (384 x 3), in quanto i dati riguardano tre blocchi di 12 campioni. L'allocazione dei bit è unica sui 3 gruppi da 12 campioni, ma i fattori di scala rimangono tre, uno per blocco come prima, in modo da evitare la distorsione. I fattori di scala si riducono a due o uno quando sono sufficientemente simili in valore per due blocchi o quando è certo che il mascheramento renda la distorsione non percepibile. In questi casi occorre aggiungere nel bitstream l'informazione che i fattori di scala sono condivisi e tra quali blocchi.

LAYER 3

Il layer 3 introduce molte caratteristiche nuove, che lo differenziano in modo sostanziale dagli altri due layer. In particolare viene introdotto un banco di filtri ibrido, che pone in cascata al banco usato nei layer 1 e 2 un modulo aggiuntivo basato su una trasformata discreta del coseno modificata. Tale modulo permette una più elevata risoluzione frequenziale, che produce una partizione dello spettro più simile alle bande critiche nelle basse frequenze. Ciascuna delle 32 sottobande può essere suddivisa 18 volte, con un numero massimo di 576 frequenze componenti. Esiste anche la possibilità di applicare una MDCT in soli 6 punti (invece che 18): in questi casi si perde in risoluzione in frequenza, ma si guadagna nel dominio del tempo per il trattamento dei transitori, e si riescono a evitare i pre-echi. In generale, l'algoritmo di compressione lavora con la MDCT a 18 punti, e richiama la MDCT a 6 punti solo quando si ha un innalzamento repentino della richiesta di bit da allocare per la quantizzazione. Le novità introdotte dal layer 3 sono la quantizzazione non uniforme, codifica di Huffman e bande raggruppate per uno stesso fattore di scala.

Formati Audio Avanzati



Esistono diverse categorie di file audio. In primis abbiamo i formati definiti come **non compressi**. La parola stessa definisce questa sezione di file. In questo caso, infatti, le onde sonore che sono state catturate vengono **convertite in formato digitale** senza alcuna ulteriore elaborazione. In cosa si traduce tutto questo? In file molto pesanti da salvare e che occupano parecchio spazio. Per creare dei file audio più facili da salvare e meno pesanti sono stati poi creati i formati con **compressione dati con perdita**. Questi sacrificano un po' di qualità a favore del poco spazio d'archiviazione richiesto.

Per rimediare anche a questo difetto, ovvero la perdita di qualità a favore della grandezza dei file, sono stati creati dei formati audio **compressi senza perdita di dati**. Ovviamente non saranno dei file piccoli come la

compressione con perdita però saranno comunque più piccoli dei non compressi. Va detto però che questo tipo di formati non sono molto usati se non in casi specifici.

PCM

PCM è il formato audio più comune **e usato sui CD e i DVD** che abbiamo in casa. È la conversione fedele dell'audio analogico all'interno di un file digitale. Non a caso stiamo parlando di un **formato audio non compresso**. Ricordiamo che i suoni analogici esistono come forme d'onda, e per convertire un'onda sonora in bit digitali il suono deve essere campionato e registrato a determinati intervalli.

WAV

WAV sta per Waveform Audio File Format (un tempo veniva etichettato anche come Audio per Windows, ma ora non più). È uno standard che è stato **sviluppato da Microsoft** e IBM nel 1991. Non è compresso ed è essenzialmente quello che si ottiene quando si estrae l'audio da un CD musicale con il computer. Occupa molto spazio (1.411 Kbps di musica stereo a 44.100 Hz/16 b), ma riproduce i suoni in maniera fedele.

MP3

Questo è sicuramente il formato audio più famoso. È stato rilasciato nel 1993 e negli anni ha incrementato il suo successo grazie **all'abbinata con i brani musicali**. Perché sono così usati? Semplice, si tratta di file con compressione con perdita di dati e i file generati sono talmente piccoli da garantire alle persone di creare una libreria musicale **senza la necessità di troppo spazio d'archiviazione**. E la qualità? È molto inferiore rispetto ad altri formati ma l'orecchio umano non sempre riesce a percepire in maniera significativa la differenza. MP3 inoltre è un **file universale**. Si può trovare su smartphone, computer, smart TV, tablet e moltissimi altri apparecchi elettronici.

WMA

WMA sta per Windows Media Audio. È stato rilasciato per la prima volta nel 1999 e ha attraversato diverse evoluzioni da allora, sempre mantenendo lo stesso nome. Come si può facilmente intuire è **un formato audio creato da Microsoft**. Fu pensato per sostituire gli MP3. Purtroppo sono pochissime piattaforme che lo supportano e perciò non è mai riuscito a mettere in difficoltà il predominio dell'MP3.

FLAC

FLAC sta per Free Codec Lossless Audio. È un formato compresso senza perdita di dati creato nel 2001. Un file FLAC può essere **compresso fino al 60% della sua dimensione** iniziale senza minimamente perdere in qualità del suono. Inoltre si tratta di un formato open-source. Per questo motivo è supportato nella maggior parte dei programmi e dispositivi. Si tratta della **migliore alternativa di qualità agli MP3**.

LEZIONE 15

Il MIDI (Musical Instrument Digital Interface) è utilizzato nella produzione di musica elettronica, con esso si indica il protocollo standard per l'interazione degli strumenti musicali elettronici, anche tramite un computer. Il MIDI sta ad indicare un linguaggio informatico, ossia una serie di specifiche che danno vita al protocollo, e un'interfaccia hardware, che consente il collegamento fisico tra dispositivi, cioè permette la comunicazione tra dispositivi hardware. Questo protocollo definisce:

- Protocollo (regole di protocollo)
- Interfaccia (hardware)
- Formato (Standard Midi File)

Esso è nato negli anni Ottanta. Il primo prototipo fu presentato da due progettisti della Sequential Circuit, Dave Smith e Chet Wood, che nel 1981 proposero le prime specifiche del MIDI in un documento pubblicato sotto il nome di The complete SCI MIDI. Esso era la risposta all'esigenza di far comunicare tra loro diversi strumenti musicali elettronici, tenendo conto delle caratteristiche di ognuno. L'immediata popolarità dello standard creò divisioni all'interno del comitato fondatore e queste divergenze furono risolte nel 1985, quando la IMA (International MIDI Association) pubblicò la versione 1.0 delle specifiche MIDI, istituendo di fatto lo standard del protocollo. Da lì nacque l'Atari 520ST che diventerà uno dei più richiesti.

In sostanza, il MIDI permette la necessità di far comunicare diversi hardware, per esempio sintetizzatori musicali, che hanno rappresentazioni diversi con diverse velocità di flusso dati. Quindi permette la sincronizzazione con dei segnali temporizzati e velocità di trasmissione fissate. Alla fine, esso permette:

- Una buona performance musicale con qualità e praticità del sistema, con basso costo.
- Codifica dati MIDI
- Trasferimento dell'ordine di kB.
- Decodifica e sintetizzazione.



*Un esempio di hardware che fa uso di MIDI è il **Sequencer**. È utilizzato nel campo musicale per la creazione e riproduzione di sequenze di segnali di controllo, che consentono di comandare uno strumento musicale elettronico. Consente l'esecuzione dei suoni e consente l'editing. Nel sequencer non viene memorizzato alcun segnale audio a parte quello di controllo; si può immaginare un sequencer come una "mano elettronica" automatica e programmabile che suona strumenti e regola pulsanti e potenziometri di sintetizzatori e processori audio.*

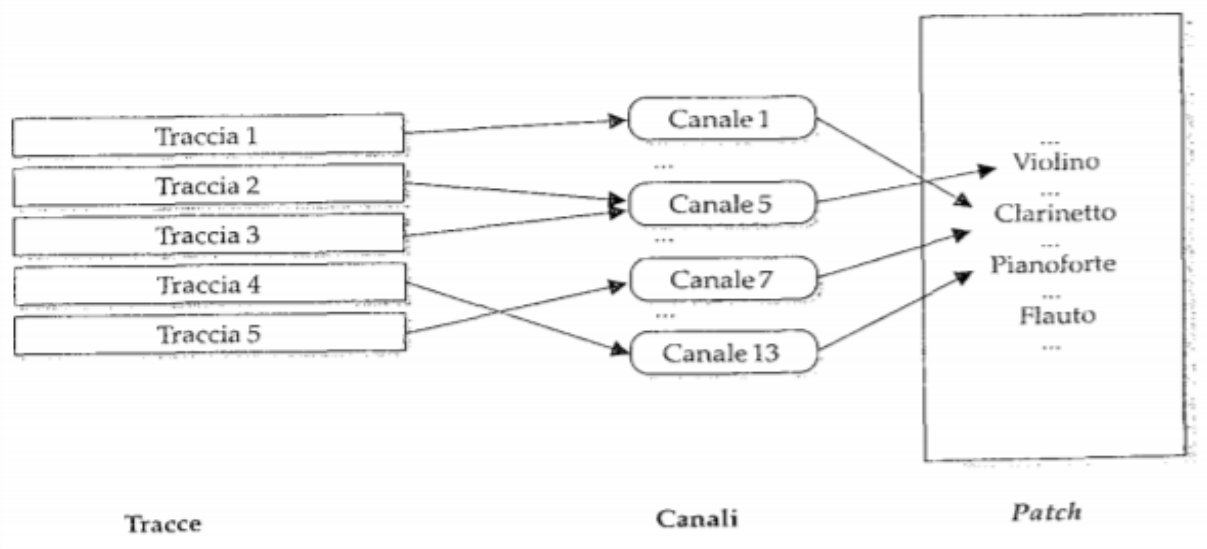
Sarà il MIDI stesso a fornire le istruzioni per le esecuzioni.

Il MIDI a livello software, cioè i tipi di informazioni da esso contenuti.

- **Canali:** Vi sono da 1 a 16 canali che permettono l'esecuzione di fino a 16 suoni diversi. Tutti i messaggi di uno stesso canale vengono associati allo stesso timbro.
- **Tracce:** Flusso strutturato e autonomo di messaggi MIDI, distinto logicamente da altre tracce. In pratica si assegna ad ogni traccia un canale con un rapporto 1 a N. Implementano il concetto di partiture e mixing, dove il mixing consente nel miscelare opportunamente fra loro uno o più suoni (anche brani).

- **Patch:** È il termine che indica il timbro prodotto da un generatore. I messaggi MIDI possono veicolare informazione relative a un massimo di 128 patch. Cioè possiamo selezionare 128 timbri diversi. Tale numero si può incrementare coi **banchi**. Poiché un banco contiene 128 patch con n banchi possiamo utilizzare n^{128} patch.

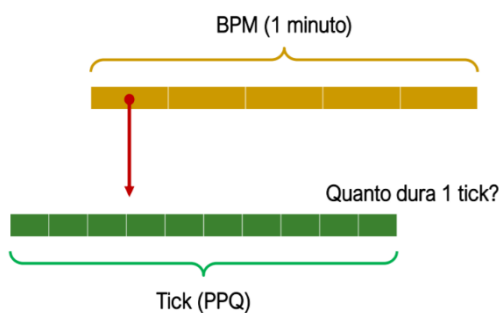
Riassunto:



Cosa succede quando premiamo un tasto di un sintetizzatore che fa corrispondere una nota ?

L'esecuzione di una nota viene codificata attraverso una serie di messaggi MIDI. La pressione sul tasto è rappresentata da due distinti messaggi: *Note-on*, *Note-off*. Il primo indica che abbiamo premuto il tasto, mentre col secondo che è avvenuto il rilascio. Entrambi sono caratterizzati da una *velocity*, cioè la velocità con cui il tasto di una tastiera si abbassa o si alza. L'altezza di una nota (*pitch*) è codificata assegnando un valore corrispondente al tasto premuto tra i 128 disponibili. Altre informazioni che riguardano una nota c'è l'*aftertouch* che indica eventuali variazioni di pressione del tasto mentre è premuto, dove si verifica il fenomeno del vibrato. Poi vi è il *controller* si possono codificare altre informazioni.

Informazione temporale:



Ai messaggi MIDI è necessario fornire anche l'informazione temporale che consenta a un sequencer di conoscere il momento esatto in cui reagire. Quindi ogni dispositivo MIDI è perciò dotato di un clock, un dispositivo interno di sincronizzazione, sulla base del quale ordina nel tempo i messaggi. L'unità di misura base sono i tick o PPQ (parti per quarto), la sua grandezza è detta *Division*, maggiore è la Division, maggiore sarà la risoluzione temporale possibile. Essa non è assoluta ma dipende dal

valore del quarto. Il numero dei quarti in un minuto è indicato BPM (pulsazioni per minuto), tipicamente compreso tra 40 e 240. I PPQ sono variabili tra 24 e 4096.

Queste specifiche riguardano singoli dispositivi, nel momento che si impiegano più dispositivi è necessario un'altra forma di sincronizzazione come la *MIDI Clock I* e *MIDI Time Code(MTC)*.

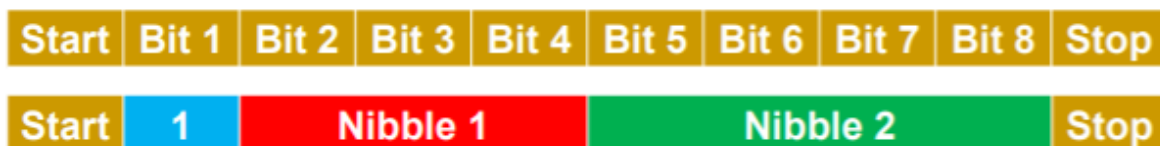
MIDI Clock: Device master con il suo clock interno produce un clock message ogni 24 PPQ, lo device slave userà il clock esterno: ovvero il clock message che arriva dal master, in modo tale da sincronizzarsi.

MIDI Time Code: Inserisce le informazioni standard del timecode SMPTE come una serie di messaggi MIDI. Ogni messaggio occupa un quarto di fotogramma. I messaggi sono trasmessi come sequenza di otto messaggi, per cui un valore completo di timecode è trasmesso ogni due fotogrammi. Quindi esso è un altro metodo di sincronizzazione usato per coordinare produzioni video ecc.

Vediamo ora come sono composti i messaggi MIDI.

MESSAGGI MIDI

I messaggi MIDI sono sequenze di parole di 10 bit, ma poiché il primo e l'ultimo indicano inizio e fine della parola possiamo considerare parole di 8 bit (1 byte).



Bit 1: Se il bit è posto = 1 si parla di *Status bytes*, sono numeri di otto cifre binarie in cui il bit più significativo è posto a 1. Questi byte servono a identificare il tipo di messaggio, ovvero lo scopo dei data byte che li seguono. Se il bit è = 0 si parla di *Data bytes*, sono numeri di otto cifre binarie in cui il bit più significativo è posto a 0. Per ogni Status byte deve sempre essere spedito il corretto numero di Data Bytes corrispondente. Essi conterranno i parametri dei messaggi.

A loro volta gli *Status Byte* possono essere di 2 tipi: *Messaggi di Canale*, *Messaggi di Sistema*.

Bit 2,3,4: Raggruppati nel Nibble 1

I Channel Message sono usati dai ricevitori in un sistema il cui numero di canali corrisponde al numero di canali codificato all'interno dello Status Byte e comporta delle operazioni come il cambio timbro e la riproduzione di una nota. I Channel Message sono identificati dai bit che vanno da 000 a 110.

I System message non sono codificati con i numeri di canale e sono rivolti a tutto il sistema. Sono usati per operazione di timing, sincronizzazione e specifiche. Sono identificati con i bit 111.

Bit 5,6,7,8: Raggruppati nel Nibble 2

A loro volta i Channel Message si dividono in *Voice Message* e *Mode Message*.

I Mode Message servono per definire la risposta degli strumenti ai Voice messages. I mode messages sono spediti sul basic channel di uno strumento. Ci dicono cosa deve suonare uno strumento.

I Voice Message descrivono gli eventi musicali: quale nota deve essere mandata in esecuzione, quale deve essere terminata, con quale volume, con quali modificazioni espressive. I voice messages sono spediti sui Voice channels. Vediamo Note-on e Note off.

- Se è Note-on e con Nibble 1 = 001 il Nibble 2 ha 4 bit e quindi 16 canali possibili, mentre il Data Byte ha 7 bit a disposizione ($Ni1+Ni2$) con 128 valori possibili per pitch e 128 valori possibili per velocity. Indica il suonare una nota
- Se è Note-off e con Nibble 1 = 000 indica di smettere di suonare una nota e ha velocity = 0. Si può ottimizzare sostituendo la coppia Note-on + Note-off con un solo Note-on seguito da due coppie di Data Byte.
- Se nel Nibble 1 = 101 il Data Byte ha 7 bit a disposizione ($Ni1+Ni2$) con 128 valori di pitch e 128 valori per variazione di pressione. Indica il tremolo. Il Nibble 2 ha 4 bit e quindi 16 canali possibili.
- Se nel Nibble 1 = 110 il Data Byte ha 7 bit a disposizione ($Ni1+Ni2$) con 128 valori di pitch e 128 valori per variazione di frequenza. Indica il vibrato. Il Nibble 2 ha 4 bit e quindi 16 canali possibili
- Se nel Nibble 1 = 100 il Data Byte ha 7 bit a disposizione ($Ni1+Ni2$) con 128 indici possibili. Il Nibble 2 ha 4 bit e quindi 16 canali possibili. Usato per il cambio timbro e per qualsiasi esigenza di indicizzazione.

Invece i System Message si dividono in Common Message, Real Time Message ed Exclusive Message.

I *Common Message* svolgono funzioni generali relative a tutto il sistema. Si pensi alle necessità di sincronizzazione richieste in un brano eseguito da un sintetizzatore. Qui abbiamo:

- **MIDI Time Code Quarter Frame**: è la categoria di messaggi impiegata dai dispositivi in grado di trasmettere/ricevere il MIDI Time Code. Sono necessari 8 messaggi trasmessi ogni quarto di frame.
- **Song Position Pointer**: una song è una sequenza di messaggi MIDI memorizzata in un dispositivo MIDI, esso indica il punto all'interno della song caricata, dove si deve posizionare il puntatore in modo tale da far partire la registrazione da quel punto.

- **Song Select**: consente di scegliere una song all'interno di quelle conservate in memoria dal dispositivo.
- **Tune Request**: Verifica l'accordatura dei dispositivi.

I *Real Time Message* si occupano del funzionamento sincronizzato dei diversi moduli di un sistema in tempo reale, esempio il coordinamento durante una registrazione che avvengano simultaneamente su più dispositivi. Qui abbiamo:

- **MIDI Clock**: garantiscono la sincronizzazione tra più dispositivi, non assoluto ma relativo.
- **Start/Continue/Stop**: questi tre messaggi coordinano il funzionamento di tutti i dispositivi che compongono il sistema MIDI.
- **Active Sensing**: viene inviato ogni 300 millisecondi, con la funzione di mantenere attiva la connessione tra master e slave.
- **System Reset**: è il messaggio che riporta tutti i dispositivi ai valori predefiniti.

Gli *Exclusive Message* sono messaggi attraverso i quali i costruttori sono in grado di specificare informazioni relative ai loro prodotti. Hanno 2 status byte posti all'inizio e fine della sequenza. Se non sono riconosciuti vengono ignorati.

Evoluzione del MIDI

Il General MIDI (GM) è uno standard più approfondito rispetto all' 1.0. Esso permette una migliore regolazione tra canali, patch e timbri. Invece General Standard (GS) e Standard XG definirono due standard interni che sopprimevano alle limitazioni del GM. Il GS, della Roland, codifica anche con l'impiego del Bank Select, il messaggio di scelta del banco, in cui si aggiungono altri controlli relativi a parametri timbrici e effetti audio. L'XG, della Yamaha, aumenta il numero dei preset e dei drumkit disponibili e prevede inoltre la scalabilità dei messaggi. In aggiunta a ciò c'era pure l'Xmidi (eXtended MIDI), una proposta d'estensione del MIDI, che consisteva nell'aumentare i canali da 16 a 324, la risoluzione dei valori (livello, velocità, ecc.) da 128 a 510 e quelli non lineari da 128 a 4374. Il tutto possibile attraverso una logica ternaria e non più binaria.