

Istituto Musicale L.Boccherini
Lucca

Master MAI 2015/2016
IMT - complesso San Michele
Lucca

Elementi di Acustica e metodologie di
sintesi del suono digitale

materiale del corso

Leonello Tarabella
Laboratorio Segnali ed Immagini
ISTI-CNR, Pisa

estatti da:

Musica Informatica, filosofia, storia e tecnologia della computer music -
Leonello Tarabella, Apogeo/Maggioli, 2014, isbn 978-88-916-0505-4

Argomenti del Corso

1	Acustica	7
1.1	L'aria	9
1.2	Sinusoide	12
1.3	Frequenza e periodo	15
1.4	Logaritmo	16
1.5	Il suono	20
1.6	Altezza e scale musicali	25
1.7	Intensità	40
1.8	Timbro	46
2	Analisi di segnali sonori	49
2.1	Pi greco e senoide	50
2.2	Segnale e spettro	57
2.3	Analisi di Fourier	62
2.3.1	Calcolo delle ampiezze delle armoniche	64
2.5	Spettri dinamici e involuppo	76
3	Analogico e digitale	79
3.1	Segnale analogico	80
3.2	Digitalizzazione del segnale analogico	81
3.3	Campionamento	82
3.4	Quantizzazione	94
3.5	Codifica	103
3.6	Convertitori	105
3.6.1	Convertitore digitale/analogico DAC	105
3.6.2	Convertitore analogico/digitale ADC	110
3.7	Considerazioni	111

4	Sintesi di segnali sonori.	113
4.2	Sintesi di segnali.	120
4.3	Oscillatore digitale.	121
4.4	Involuppo.	128
4.5	Sintesi armonica additiva.	130
4.6	Modulazione di Frequenza (FM)	134
4.6.1	Timbriche mediante FM.	143
4.6.2	FM modificata.	147
4.7	Distorsione non lineare (wave-shaping)	148
4.8	Sintesi sottrattiva.	153
4.9.4	Algoritmo di Karplus-Strong.	173
4.10	Sintesi per modelli fisici.	176
6	MIDI Musical Instrument Digital Interface.	223
6.1	L'interfaccia.	225
6.2	Il linguaggio.	228
6.2.1	Messaggi di canale (channel message)	229
6.2.2	Messaggi di sistema (system message)	231
6.4	Standard MIDI File e General MIDI.	234
7	Linguaggi musicali.	239
7.1	Linguaggi testuali.	242
7.1.1	CSound.	251
7.2	Linguaggi visuali.	268
7.2.1	MAX/MSP e Pure Data	268
7.4	Composizione algoritmica.	278
8	Interazione.	293
8.1	Dispositivi di input e interfacce gestuali.	301
8.2	Mapping.	304
8.3	Esecuzione e gestualità.	305

Acustica



- 1.1 L'aria
- 1.2 Sinusoide
- 1.3 Frequenza e periodo
- 1.4 Logaritmo
- 1.5 Il suono
- 1.6 Altezza e scale musicali
- 1.7 Intensità
- 1.8 Timbro



L'acustica è la parte della Fisica che si occupa del **suono** e ha un duplice aspetto: quello *oggettivo*, misurabile matematicamente nelle grandezze fisiche che lo caratterizzano, e quello *percettivo*, relativo alla sensibilità dell'apparato uditivo umano. Il primo aspetto è settore di indagine della *fisica acustica*, mentre il secondo è settore di indagine della *psico-acustica*, che mette in relazione il fenomeno fisico con le nostre reazioni sensoriali.

Il suono è il risultato dello stimolo indotto nell'orecchio dalla vibrazione di particelle d'aria causata da oggetti in movimento. Tutto ciò che si muove provoca uno spostamento delle particelle d'aria circostanti, che a loro volta provocano lo spostamento di quelle a loro vicine e così via, fino a ch  non viene incontrato qualche altro corpo solido, come il timpano dell'orecchio che viene messo in movimento *per simpatia* (*syn-p thos*, sento insieme) in ragione della sua conformazione e della distanza dall'oggetto sorgente.

Lo spostamento di particelle d'aria implica la creazione momentanea di micro-zone di maggiore e minore densit , cio  di pressione, in relazione a quella d'ambiente, dovuta rispettivamente all'accumulo ed alla rarefazione delle particelle d'aria stesse. In assenza di perturbazioni della pressione, cio  che "percepriamo"   il *silenzio*.

Fondamentalmente, ciò che caratterizza un suono da un punto di vista percettivo è l'**altezza** (dal grave all'acuto), legata alla *velocità di vibrazione* dell'oggetto sorgente, e l'**intensità** (dal *piano* al *forte*) legata all'ampiezza e quindi all'*energia* dell'oscillazione stessa. Altezza ed intensità sono quantità misurabili in termini matematici e quindi dominio di interesse della Fisica acustica.

La Psico-acustica coinvolge invece la sfera del soggettivo e cioè il complesso di sensazioni che noi proviamo in presenza di determinati stimoli sonori. Per cui, ad esempio, mentre è possibile dire esattamente qual è l'altezza di un suono (in musica, dal grave dei contrabbassi all'alto dei violini e degli ottavini) non è invece possibile definire quantitativamente, e cioè in termini matematici, il **timbro** di un suono, al quale perciò ci si riferisce per analogia agli altri sensi: vista, gusto, tatto.

E così un suono può essere definito morbido, brillante, piatto, pastoso, metallico, ecc., con caratterizzazioni del tutto soggettive e decisamente sfuggenti ad una classificazione di tipo fisico-matematico. Esistono tuttavia modelli e metodologie, che in seguito esamineremo, usate per caratterizzare il timbro di un suono.

Grande importanza assume anche “il dove” e “il come” il messaggio sonoro viene generato e percepito: l'ambiente infatti trasforma e caratterizza il suono. Allo stesso tempo l'ambiente caratterizza se stesso attraverso il suono: i passi in un bosco tappezzato dalle foglie d'autunno “suonano” in maniera del tutto diversa dai passi all'interno di un grande spazio chiuso come quello di una chiesa.

È per questo motivo che la fruizione di musica è fortemente dipendente dall'ambiente in cui avviene l'evento ed è per questo che, quando si può, si sceglie il luogo più adatto in funzione della musica da eseguire e, viceversa, si sceglie una certa tipologia di musica in funzione di un certo luogo prestabilito.

Ad esempio: nella musica operistica i cantanti spaziano con la loro voce per un raggio d'azione di una ventina di metri in un ambiente poco riverberante come è di fatto il *Teatro all'italiana*, che garantisce l'intelligibilità delle parole; mentre per canti corali si preferisce un ambiente riverberante che assicura la fusione delle molte voci con l'ambiente e che, tra l'altro, fa perdere la percezione della direzione del suono.

Nella realtà in cui siamo immersi, le sorgenti sonore sono costituite dai rumori provocati dai fenomeni atmosferici, da oggetti in movimento, dal verso degli animali, dalla voce umana e dal suono degli strumenti musicali; in generale, comunque, la distinzione tra *suono* e *rumore* non riguarda l'orecchio, ma piuttosto l'interpretazione che il cervello fa degli stimoli acustici provenienti dall'orecchio. Questo è pertanto di interesse della psico-acustica.

1.1 L'aria

Si è detto che il suono è il risultato della variazione di pressione dell'aria provocata da qualcuno o da qualcosa. Per l'appunto, l'*aria*. L'aria è l'elemento fondamentale e vitale alla nostra esistenza bio-fisica, al punto che i greci facevano coincidere nella parola *anemòs* il significato di respiro e di anima; e nelle filosofie orientali il respiro è il *prana* che è sinonimo di *energia vitale*.

L'aria è il sottile strato gassoso di uno spessore di poco più di venti/trenta chilometri che avvolge il nostro pianeta, e più propriamente detto *atmosfera*: se la Terra fosse una sfera di un metro di diametro l'atmosfera avrebbe lo spessore di un foglio di carta. L'aria ci tiene compatti, ci fa respirare, ci dà la sensazione del caldo e del freddo, trasporta gli odori, trasporta il polline e una grande varietà di semi aerei, consente di volare agli insetti, agli uccelli ed agli aerei, uniforma e sfuma la luce modellando le forme degli oggetti, e contribuisce alla percezione della distanza delle cose, come teorizzato ed usato da Leonardo da Vinci pittore con la *prospettiva aerea*, tant'è che in una giornata particolarmente limpida dopo la pioggia anche le montagne "sembrano più vicine".

E l'aria fa esistere il suono.

Si immagini ora di approntare un congegno costituito da una membrana tesa all'interno di un supporto rigido, e di applicare in maniera solidale al centro della membrana un'asse collegato ad una leva, come illustrato nella Figura 1.1.

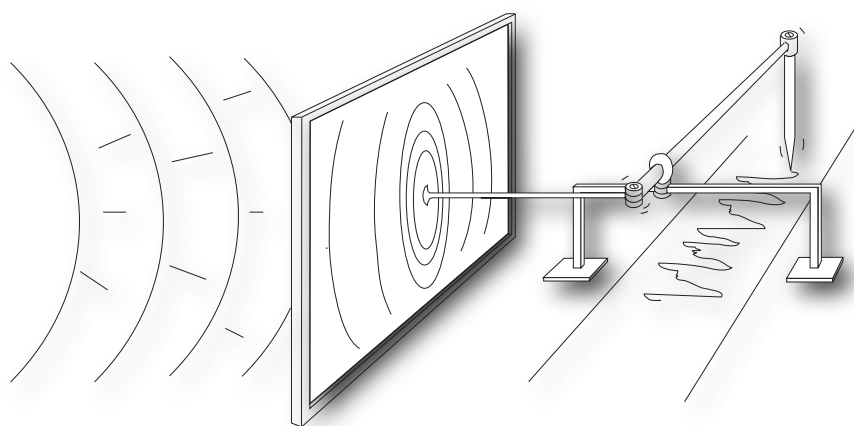


Figura 1.1 Visualizzazione del suono

La leva è incernierata in modo tale che piccoli spostamenti da un capo si traducano all'altro con spostamenti molto ampi e se a questo secondo capo viene applicata una penna che scrive sopra un nastro di carta scorrevole, come succede ad esempio in un sismografo o in un cardiografo, è allora possibile in un certo senso “vedere” quello che l'orecchio sente.

Posto dunque tale congegno in prossimità di una sorgente sonora, ci si deve aspettare che la membrana cominci ad oscillare in “simpatia” con la sorgente sonora stessa e che il grafico risultante ne sia la rappresentazione. È questo il principio di funzionamento del *fono-autografo* che Eduard-Léon Scott de Martinville (1817-1879) con geniale intuizione inventò, costruì e brevettò nel 1857 a Parigi.

Charles Cros (poeta, inventore e scrittore francese, 1842-1888) basandosi sui risultati di Scott, realizzò una macchina per la riproduzione del suono; e sulla base di queste esperienze, Thomas Alva Edison (1847-1931) realizzò, brevettò nel 1878 (US Pat. No. 200521) e commercializzò il *fonografo*.

I primi fonografi di Edison usavano come supporto di registrazione un cilindro e sfruttavano un movimento verticale dello stilo (esiste in rete un fiorente mercato di cilindri di Edison).

Ma fu Emile Berliner (1851-1929) che, sullo stesso principio dell'incisione di Scott-Cros-Edison, inventò il *disco*, che offriva caratteristiche decisamente vantaggiose dal punto di vista commerciale, perché poteva essere stampato in serie, anche se il cilindro garantiva uniformità di qualità per tutta la sua durata.

Se realizzare il marchingegno di Figura 1.1 risulterebbe piuttosto laborioso, più facile è invece realizzare il *telefono-a-bicchieri-di-carta-e-spago*, esperienza di gioco che più o meno tutti abbiamo fatto durante l'infanzia. Cosa ci sta sotto va però ben oltre il semplice gioco e sta alla base dei principi che hanno dato inizio alla storia della registrazione e della trasmissione della voce e della musica iniziata con il fono-autografo di Scott de Martinville.

Si tratta infatti di trasmettere le vibrazioni del fondo del bicchiere di carta di chi parla al fondo del bicchiere di chi ascolta attraverso una cordicella tesa e dritta fissata ai due bicchieri. Di fatto, le due membrane dei fondi dei bicchieri e la cordella formano un sistema unico che si muove in maniera solidale: questo significa che il filo trasmette meccanicamente l'*analogo* delle vibrazioni del fondo-microfono al fondo-altoparlante all'altro capo del filo.

I microfoni veri usati nel telefono e nei normali impianti di amplificazione sono invece realizzati con membrane di metallo o materiale plastico che in prossimità di campi elettrici o elettromagnetici, *trasducono* (e cioè convertono) la vibrazione meccanica in variazione di corrente elettrica, che può quindi essere trasmessa o memorizzata per essere ascoltata a distanza attraverso un

altoparlante. Quello che viene trasmesso è esattamente la “versione elettrica” di ciò che succede meccanicamente sulla cordicella del telefono di carta.

La storia della registrazione sonora è cominciata esattamente così, e tutto ciò che è venuto dopo è stato caratterizzato da un continuo miglioramento degli elementi in gioco, eventualmente sostituendo e/o integrando parti meccaniche con parti elettromeccaniche e/o elettroniche.

Mentre la tecnologia al contorno andava evolvendo, parallelamente veniva creata e messa a punto la teorizzazione analitico/matematica del fenomeno sonoro in sé e, più in generale, della *teoria del segnale*, dove per **segnale** si intende la *variazione nel tempo di una grandezza fisica*, che nel nostro caso è la pressione dell'aria.

Ciò che interessa dal punto di vista percettivo è perciò la *variazione della pressione dell'aria* e non quello della pressione effettiva nell'ambiente che, essendo costante, non induce alcuna vibrazione sulla membrana del timpano dell'orecchio e non produce perciò alcun fenomeno uditivo.

La curva delle variazioni di pressione dell'aria in funzione del tempo disegnata dal congegno di Figura 1.1, rappresenta la **forma d'onda** relativa ad una determinata eccitazione.

Perciò una forma d'onda relativa ad un fenomeno sonoro può essere descritta e rappresentata da una *funzione del tempo*, che rappresenta in tutto e per tutto la variazione continua di pressione dell'aria.

In altri termini si dice che il *segnale* è una *funzione continua del tempo* che rappresenta l'*analogo* della variazione della pressione: da qui deriva il termine *analogico* (Capitolo 3) che è quindi, più che sinonimo, termine strettamente legato al concetto di *continuità*, come è il naturale trascorre del tempo e di tutti i fenomeni naturali.

Il microfono è il tipico dispositivo elettronico analogico in grado di convertire una variazione di pressione in una variazione di tensione. Esso viene detto per l'appunto, *trasduttore*. In maniera duale, l'altoparlante è il trasduttore analogico che converte una variazione continua di tensione elettrica in una variazione di pressione dell'aria e quindi in suono.

I prossimi paragrafi riguardano tre concetti matematici, la **sinusoide**, la **frequenza** ed il **logaritmo**, necessari alla trattazione di tutti gli argomenti del libro ed in particolare alla definizione del suono (§1.5) nelle caratteristiche fondamentali di **altezza**, **intensità** e **timbro**.

1.2 Sinusoide

Costruite un semplice pendolo utilizzando una cordicella di 30-40 centimetri ed un oggetto qualunque appeso ad un'estremità; tenete l'altra estremità tra le dita di una mano e, avendo cura di tenere la cordicella tesa, spostate di lato con l'altra mano l'oggetto e rilasciatelo: il pendolo comincia ad oscillare... La velocità, che è massima (nei due sensi) nella parte bassa centrale, diminuisce in maniera graduale in prossimità degli estremi destro e sinistro, fino a fermarsi per un brevissimo istante per cambiare direzione.

Questo modo continuo e graduale di trasformazione è un particolare "andamento sinuoso" che risulta molto evidente se si accorcia la cordicella, aumentandone così la velocità di oscillazione; a circa 10 centimetri, l'oscillazione diventa quasi una pulsazione che ha grande eleganza ritmica e senso musicale. Non a caso nella musica jazz si parla di *swing*.

La figura seguente mostra un pendolo costituito da un secchiello pieno di sabbia con il fondo forato, appeso ad un cavo, che oscilla sopra un nastro che scorre a velocità costante:

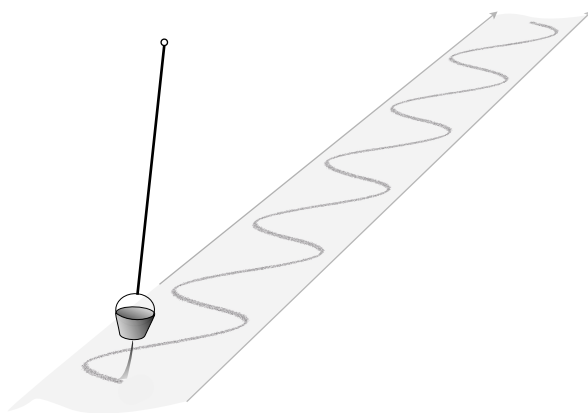


Figura 1.2 Pendolo e sinusoide

L'andamento oscillante e continuo del secchiello che pendola viene visualizzato dalla scia tracciata dalla sabbia che ha la forma di una **sinusoide**.

Anche il metronomo meccanico vecchio stile è una sorta di pendolo "rovesciato" e dall'esperienza si sa che spostando il peso scorrevole sulla barretta d'acciaio verso l'alto si hanno oscillazioni di durata maggiore di quando lo si sposta verso il basso. È importante perciò cercare la relazione che lega la lunghezza del pendolo alla durata di un'oscillazione completa.

Dalla Fisica si sa che la durata dell'oscillazione, per piccole oscillazioni, dipende soltanto dalla lunghezza del pendolo. Tale durata viene definita *periodo*, viene indicata con T ed è data con buona approssimazione dalla formula

$$T = 2\sqrt{L}$$

dove L è la lunghezza del pendolo. Se L viene espressa in metri, il valore di T risulta espresso in secondi, per cui per avere ad esempio un'oscillazione completa di 1 secondo, il pendolo deve essere lungo 0,25 metri e cioè 25 centimetri.

È possibile costruire la sinusoide in modo geometrico con righello e compasso a partire da un cerchio: se si divide il cerchio in parti uguali, ad esempio in 16 parti, come le fette di una torta, si ha che ogni fetta è di $22,5^\circ$ perché $22,5^\circ \times 16 = 360^\circ$.

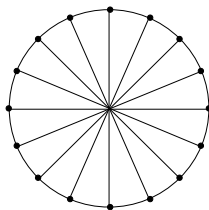


Figura 1.3 Divisione del cerchio in 16 parti uguali

Se si fa “rotolare” il cerchio per un intero giro, si ottiene un segmento lungo $\pi = 3,1415\dots$ volte la lunghezza del diametro del cerchio e quindi 2π volte la lunghezza del raggio.



Figura 1.4 Angoli in radianti

Succede inoltre che i 16 punti del cerchio lasciano sul segmento 16 tacche distanziate l'una dall'altra $1/16$ (un sedicesimo) di circonferenza: il segmento tra una tacca e l'altra misura perciò $2\pi/16$. E naturalmente la lunghezza “dritta” sul segmento orizzontale è uguale alla lunghezza “curva” sul cerchio. È questo il modo alternativo di misura gli angoli detto in **radianti** invece che in *gradi*. Perciò 360° corrisponde a 2π radianti, 180° a π radianti e 90° a $\pi/2$ radianti e così via.

Riportando il segmento all'altezza del centro del cerchio, facendo coincidere il primo punto del segmento a sinistra con il punto più a destra del cerchio e spostando ogni tacca su o giù fino a raggiungere l'altezza del punto corrispondente al punto sul cerchio che l'ha generata, si ottiene questa figura:

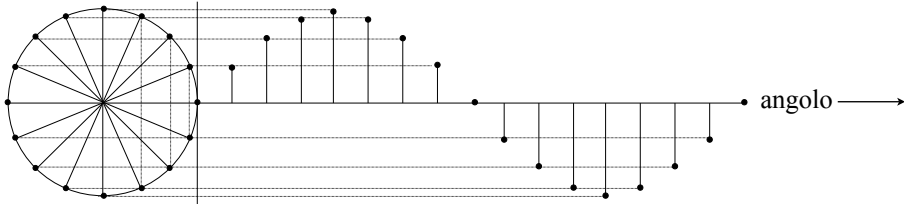


Figura 1.5 Costruzione geometrica della sinusoide

Già si vede che i punti in successione “disegnano” una *curva sinusoidale*, che diventa sempre più evidente se immaginiamo di aumentare il numero delle divisioni sul cerchio (e quindi dei punti che vanno a formare la sinusoide) fino ad arrivare ad un’infinità di punti, o più semplicemente ad un solo punto che si muove con continuità sul cerchio che genera una curva continua sinusoidale.

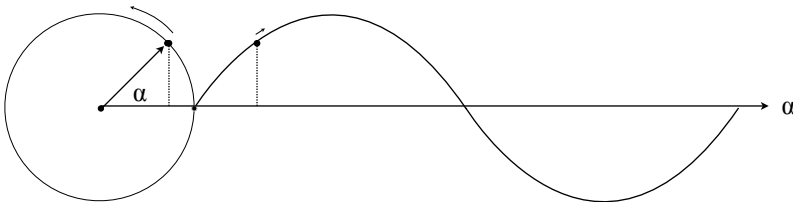


Figura 1.6 Cerchio e sinusoide

La cosa interessante del processo di divisione che aumenta sempre di più all’infinito è che tutto ciò sta alla base di quello che vedremo nel Capitolo 3 relativo all’*analogico* ed al *digitale*.

C’è inoltre da rilevare che, a parte l’oscillazione del pendolo, *in Natura non esistono fenomeni che oscillano in maniera puramente sinusoidale con velocità tali da provocare suono*. Tuttavia nello studio dell’Acustica si fa continuamente riferimento alla *sinusoide* come strumento matematico per l’analisi del suono e, nel nostro caso, per la creazione di suoni per sintesi elettronica.

Nel Paragrafo 2.1 verrà ripreso l’argomento della sinusoide esponendone più in dettaglio le caratteristiche analitiche. Per il momento è sufficiente tenere presente che ciò che caratterizza un suono è, in prima istanza, la sua *altezza* (grave, medio, acuto...), corrispondente alla velocità di oscillazione (*frequenza*) del mezzo meccanico che produce quel suono.

1.3 Frequenza e periodo

Nel linguaggio corrente la parola *frequenza* richiama il concetto di presenza di qualcuno nel solito luogo a scadenze fisse, come ad esempio un alunno a scuola: ‘...l'alunno frequenta molto, frequenta poco...’; oppure si dice che un fenomeno è *poco* o *molto* frequente quando accade poche o molte volte nel tempo. Il termine *frequenza* non si riferisce ad una proprietà caratteristica del fenomeno acustico (o di quello elettromagnetico delle trasmissioni radio e della luce) ma più semplicemente indica il *ripetersi a periodi regolari di tempo* di un fenomeno.

Il termine *frequenza* (di per sé astratto) si applica a fenomeni caratterizzati da accadimenti che si ripetono ad intervalli di tempo regolari e cioè *periodicamente*: il **periodo** è allora l'intervallo di tempo che trascorre tra un accadimento ed il successivo, che nel caso del pendolo o di una corda che vibra è dato dal riposizionamento su uno stesso punto. E anche se si dice *la frequenza del pendolo è...* la dizione corretta è *il pendolo oscilla con una frequenza di...* e cioè: *il pendolo compie un certo numero di oscillazioni nell'unità di tempo prescelta...* dove l'unità prescelta è il *secondo*. La definizione di frequenza è dunque

frequenza = numero di oscillazioni al secondo

L'unità di misura della frequenza è l'Hertz (dal nome del fisico tedesco Heinrich Rudolf Hertz) e si abbrevia con Hz. Dire che un suono è a 100 Hz vuol dire che l'oggetto vibrante (ad esempio, una corda) che provoca quel suono, oscilla fisicamente 100 volte al secondo, il che vuole dire che una sola oscillazione *dura* 1/100 (un centesimo) di secondo e quindi il *periodo di oscillazione* è di 0,01 secondi e la sua frequenza di oscillazione è 100 Hz.

Il *periodo* viene indicato con T e corrisponde all'inverso della frequenza, per cui $T=1/F$ e, viceversa, la frequenza è l'inverso del periodo, per cui $F=1/T$.

Esistono nel linguaggio parlato espressioni come *essere sulla stessa lunghezza d'onda*, *essere sulla stessa frequenza* e *essere in sintonia* per dire in maniera suggestiva *avere le stesse idee*, *essere d'accordo* anche se è evidente che noi umani non abbiamo niente a che fare con “frequenze di trasmissione e ricezione”. *Essere in sintonia* è espressione pre-tecnologica (*sintonia*, come *simpatia*, è parola greca) che nelle relazioni umane significa proprio *essere d'accordo*. Con l'invenzione della Radio il termine *sintonia* venne scelto per indicare la *selezione*, da parte dell'apparecchio ricevente, di una stazione trasmittente in modo da essere sulla stessa lunghezza d'onda, sulla stessa frequenza. Succede ora che in un percorso a ritroso, la tecnologia offre al linguaggio delle umane relazioni la sua terminologia fatta di frequenze e lunghezze d'onda.

1.4 Logaritmo

Se avete 20 anni di età, vuol dire che fino ad oggi avete vissuto circa 7500 giorni; se avete 26 anni avete vissuto circa 10.000 giorni. Sapete che è impossibile ricordarli tutti, ma siete comunque perfettamente coscienti di averli vissuti tutti, uno dopo l'altro, ed avete la precisa consapevolezza di tutta la vostra vita vissuta fino ad oggi.

Se a partire dalle scuole elementari ogni sera prima di addormentarvi avete annotato su un foglio le cose essenziali accadute e fatte durante la giornata, avreste ora un diario di qualche migliaio di pagine.

Ma se vi si chiedesse di scrivere *ora*, a memoria, il diario della vostra vita a ritroso, sareste in grado di riempire un foglio intero per quello che riguarda oggi, un foglio per ieri e avanti ieri, un foglio per il resto della settimana, un foglio per la settimana precedente, un foglio per il resto del mese, un foglio per il mese precedente, e così via. E cioè: su un singolo foglio riportereste le cose fatte ed accadute relative ad un numero di giorni, a ritroso, sempre più grande.

L'aumento progressivo del numero-di-giorni-per-pagina si riscontra negli stessi termini anche quando dal diario personale si passa al "diario collettivo" costituito dalla Storia: basta consultare un libro di testo scolastico per constatare che, grosso modo, lo stesso numero di pagine (una ventina o una cinquantina) riportano le cose relative agli *anni* del XX secolo, ai *decenni* del XIX, XVIII e XVII secolo, ai *secoli* del Rinascimento, del Medio Evo e dell'Impero Romano, ai *millenni* della storia dei Greci, degli Egiziani, degli Assiro/Babilonesi, dei Sumeri..., alle *decine di millenni* dell'Età del Ferro, del Bronzo, del Neolitico... Mesolitico... Paleolitico ...

Per cose accadute di recente usiamo il giorno come unità di misura dicendo *3-4-5 giorni fa*, ma oltre i 7-8 giorni diciamo *la scorsa settimana*, *2-3 settimane fa...* e poi *il mese scorso*; e non diciamo mai *145 giorni fa* ma *3-4 mesi fa* e poi *l'anno scorso* e così via; e per una cosa accaduta 10 anni fa non diciamo *3600 giorni fa* o *520 settimane fa* o *120 mesi fa*. Più è lontano il momento o il periodo di tempo rammentato, più grande è l'unità di misura utilizzata: giorno, settimana, mese, anno, decina d'anni, secolo, millennio... Le diverse unità di misura sono perciò una sorta di "contenitore" grosso modo della stessa grandezza che raccoglie una quantità di ricordi sempre decrescente a fronte di quantità di giorni ed eventi sempre crescente.

Secondo una delle più accreditate favole relative al *gioco degli scacchi*, si dice che a Sissa (Sussa Ibn Dahir al-Hindi) venne offerto dal Re, come ricompensa per l'invenzione di quel meraviglioso gioco, di poter esaudire un qualunque suo desiderio: Sissa chiese una quantità di grano pari al raddoppio di chicchi di grano a partire da 1 per la prima casella della scacchiera, per tutte le restanti

caselle fino alla 64-esima, e cioè: $1, 1 \times 2 = 2, 2 \times 2 = 4, 2 \times 2 \times 2 = 8, 2 \times 2 \times 2 \times 2 = 16, \dots$ che in matematica viene messo in forma “compatta” con la notazione $2^0, 2^1, 2^2, 2^3, 2^4, \dots$ dove il piccolo numero posto in alto a destra è detto *esponente* del numero *base* 2 e indica quante volte la base (in questo caso, 2) deve essere moltiplicata per se stessa. Questa notazione descrive come *elevare la base a potenza* per cui, ad esempio, $2 \times 2 \times 2 = 2^3$ si legge *2 elevato a 3*.

Quindi, a fronte di un *incremento lineare* (0, 1, 2, 3, 4, 5, 6, 7, 8...) dell’esponente, abbiamo un *incremento esponenziale* del valore risultante: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32786, 65536, 131072, 262144,

Procedendo col meccanismo del raddoppio ad ogni successiva casella, il numero di chicchi di grano cresce così rapidamente che alla 64-esima casella, tale numero è pari a circa 18-miliardi-di-miliardi, corrispondenti all’incirca a 1000 anni dell’attuale produzione mondiale di grano, costituita da così tanti chicchi che, se stesi in maniera compatta e uniforme per terra, coprirebbero la superficie dell’intero pianeta Terra, montagne, laghi, mari e oceani compresi.

Dante, nel XXVIII canto del Paradiso, per indicare lo strabiliante numero di Angeli che là sfavillavano come scintille, dice: *L’incendio suo seguiva ogni scintilla - ed eran tante, che ’l numer loro - più che ’l doppiar delli scacchi s’immilla* (vv.91-93).

Scelta perciò una *base* arbitraria (nell’esempio appena visto, 2), l’esponente che si dà alla base per ottenere quel numero, viene definito *logaritmo*; perciò se $2^4 = 16$ allora il logaritmo di 16 in base 2 è 4 e si scrive $\log_2(16)=4$; per cui $\log_2(1024) = 10$ e $\log_2(18.446.744.073.709.551.616) = 64$.

Se si sceglie come *base* il valore 10, allora l’esponente indica l’ordine di grandezza: $10^1=10, 10^2=100, 10^3=1000, \dots 10^6=1.000.000, \dots$ e questo è il modo che di solito usiamo per riferirci alle distanze: diciamo 1 metro, 10 metri, 100 metri, 1 chilometro, 2, 3, 4...10 e poi saltiamo a 20, 30...90, 100, 200, 300...1000, 2000, 3000...

Allora, $\log_{10}(100)=2, \log_{10}(1000)=3, \log_{10}(1.000.000)=6$.

La *base* che di solito si usa nei calcoli è il numero $e=2,718281828\dots$ detto numero di Eulero, ed il logaritmo e l’esponenziale che ne derivano sono detti *naturali* perché molti fenomeni della Natura (meccanici, elettrici, statistici) si comportano con andamento esponenziale e/o logaritmico in base e .

Trattare quantità numeriche che crescono in modo esponenziale ci viene difficile; ma considerandole logaritmiche, le ridimensioniamo alla nostra portata cognitiva.

La misura “lineare” del tempo e delle distanze della realtà vera è soverchiante e, per poterla gestire, in una sorta di rivincita psicologica ed operativa, la riduciamo (senza privarla tuttavia del contenuto) a logaritmo.

Il nostro pensiero è “logaritmico” per economicità cognitiva. Focalizzare l’attenzione sul fatto che abbiamo vissuto migliaia di giorni, che le azioni che compiamo quotidianamente, da quelle intime e personali a quelle logistiche e relazionali, le abbiamo fatte migliaia di volte un giorno dopo l’altro, con la consapevolezza che dovrebbe esistere in noi un ricordo lineare di “un noi” replicato tante volte quanti i giorni della nostra vita vissuta fino ad oggi, dà quasi un senso di smarrimento, di vertigine e di angoscia che svanisce in un catartico e rasserenante “ricordo logaritmico”. Il logaritmo perciò, prima ancora di essere uno strumento matematico, è un espediente psicologico per “ridimensionare” quantità numeriche (come quelle relative allo spazio e al tempo) che crescono vertiginosamente in maniera esponenziale.

La cosa interessante è che l’esponente che si dà alla base può essere un qualunque valore anche “non intero” per cui, ad esempio:

se $2^3=8$ e $2^4=16$ si ha che $2^{3,57}=9,5136\dots$, $2^{3,69}=12,906\dots$ e così via;

se $\log_2(8)=3$ e $\log_2(16)=4$, allora $\log_2(11)=3,459\dots$, $\log_2(15)=3,9\dots$

Se si usa una x per indicare un *qualunque valore* cui elevare una base b e con y il valore risultante, si ha un’espressione, più formalmente chiamata *funzione*, che indica *con continuità* l’andamento della *curva esponenziale* $y=b^x$ e negli stessi termini la funzione $y=\log_b(x)$ rappresenta la *curva logaritmo*.

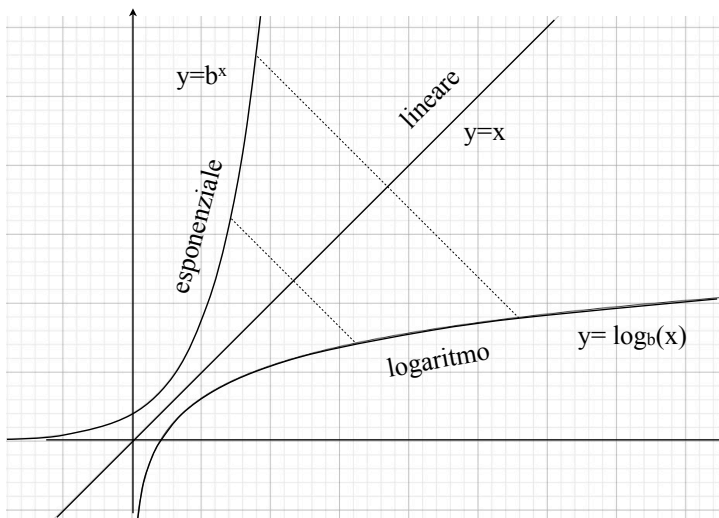


Figura 1.7 Curve lineare, esponenziale e logaritmica

Curva esponenziale e curva logaritmica sono perfettamente simmetriche rispetto ad una retta a 45 gradi, che si esprime con la *funzione lineare* $y=x$ in base alla quale ad ogni valore di x corrisponde lo stesso valore di y ; esponenziale e logaritmo sono funzioni l'una l'inversa dell'altra per cui

$$y = b^{\log_b(x)} = x \quad \text{e} \quad y = \log_b(b^x) = x$$

L'eventuale diario scrupolosamente scritto giorno dopo giorno mette in relazione uno-a-uno fogli e giorni secondo un andamento *lineare*. Ma il diario scritto a posteriori mette in relazione fogli e giorni in maniera sempre crescente in modo *esponenziale* e fogli e quantità di ricordi riportati (rispetto al foglio precedente) in modo *logaritmico*.

Il logaritmo è stato introdotto da Napier (Nepero) nel 1617 come osservazione dell'andamento degli interessi del denaro ma risultò da subito essere uno straordinario strumento matematico per eseguire velocemente operazioni complesse come la moltiplicazione, l'elevamento al quadrato, l'estrazione di radice... Una delle tante proprietà del logaritmo è quella per la quale la somma dei logaritmi di due numeri è uguale al logaritmo della moltiplicazione dei numeri stessi e cioè: $\log(a)+\log(b)=\log(a \times b)$ per cui sommare due lunghezze con due righelli logaritmici vuol dire moltiplicare i due valori corrispondenti. Con questo principio sono stati costruiti nel tempo *regoli calcolatori* (vedi in internet) di varie dimensioni con i quali eseguire in maniera *analogica* operazioni complesse con una precisione di 2 o 3 cifre, sufficienti nella maggior parte delle situazioni pratiche. Molte generazioni di ingegneri hanno costruito palazzi, grattacieli, ponti, dighe, ferrovie, automobili, apparecchiature elettroniche... eseguendo i calcoli di progetto con il regolo calcolatore. Dopo più di 100.000 giorni di gloriosa esistenza trascorsi linearmente (ma si fa prima a dire in maniera logaritmica "dopo più di tre secoli") il regolo calcolatore è andato prima in soffitta e poi nel dimenticatoio.

Per quello che qui ci riguarda, il logaritmo e l'esponenziale sono gli strumenti matematici che si usano per la comprensione e la misura analitica delle due grandezze fondamentali del suono: *altezza* ed *intensità*. Infatti essi mettono in relazione precisa il fenomeno fisico, oggetto di studio dell'Acustica, con l'aspetto percettivo del nostro apparato uditivo, oggetto di studio della Psico-acustica, e quindi della Musica.

Suonando in progressione i La su una tastiera di pianoforte, a partire dal più basso a quello più alto, si ha la sensazione di una graduale ed uniforme variazione dell'altezza del suono, che è ancor più evidente quando si suonano tutte le note che stanno tra due La successivi.

Tuttavia, se si misurano le frequenze delle note ci si accorge che la distanza in frequenza tra due note successive cresce progressivamente con *andamento esponenziale*. Ma la sensazione psico-acustica è quella di una progressione graduale e uniformemente *lineare*, perché così risponde il nostro apparato uditivo, secondo meccanismi psico-fisiologici di interpretazione rigorosamente *logaritmici* predisposti dalla Natura.

Riprenderemo l'argomento nel Paragrafo 1.6 dedicato alle scale musicali.

Il nostro apparato uditivo risponde in maniera logaritmica anche per quello che riguarda l'intensità sonora per cui, per esempio, a fronte della sensazione psico-acustica del raddoppio di volume sonoro, è necessaria una quantità quadrupla di energia fisica esercitata con maggiore pressione sui tasti del pianoforte o con maggiore insufflazione su uno strumento a fiato.

Sarà questo l'argomento del Paragrafo 1.7 dedicato all'intensità sonora e alla sua misura in *decibel*.

1.5 Il suono

Il pendolo visto nel Paragrafo 1.2, una volta messo in moto, rallenta gradualmente la sua corsa perché "cozza" continuamente contro le particelle dell'aria in cui è immerso e perciò, per un naturale processo di trasformazione di energia, provoca un effetto percepibile come leggerissimo soffio.

Ma se l'oggetto oscillante è una lama di acciaio o una corda tesa, allora il numero di oscillazioni al secondo è in questo caso decisamente superiore (50 e oltre) a quello del pendolo ed accade ugualmente che la lama e/o la corda investano e spostino le particelle dell'aria circostante, creando zone di accumulo e di rarefazione delle particelle stesse.

Se siete in mezzo alla folla e date uno spintone alle persone a voi vicine, si crea un accalcamento di persone nella direzione della vostra spinta e un vuoto dietro di voi; ma subito dopo le persone, si ridistribuiscono in maniera omogenea nello spazio.

È quello che succede alle particelle dell'aria, soltanto che la lama e la corda danno alle particelle dell'aria ripetuti spintoni che trasmettono accumulo e rarefazione a quelle vicine ed a quelle vicine delle vicine e così via. E soprattutto lo fanno in tutte le direzioni.

Per comprendere meglio il fenomeno nel suo complesso, figuratevi di essere nei pressi di una laghetto di pesca sportiva un tantino insolito e surreale, in una bella giornata senza vento, dove tutti i pescatori sono interessati soltanto a godersi la piacevole situazione, piuttosto che alla pesca vera e propria.

Supponete ora di lanciare un sasso in un punto qualunque della superficie dell'acqua che, in assenza di vento, è del tutto piatta: come si sa, si formano dei cerchi concentrici al punto di immersione del sasso che si espandono fino ad arrivare ai sugheri delle lenze dei pescatori e poi si frantumano e si disperdono tra le erbe e le canne ai bordi del laghetto. Facciamo una serie di osservazioni.

1) Il sasso che entra in acqua, per una frazione di secondo spinge verso il basso la porzione d'acqua investita creando una "buca" momentanea che subito viene "ricoperta" per tornare a livellare la superficie del laghetto. Ma questa spinta, a differenza del caso dello spintone nella folla, si trasmette alle particelle d'acqua "circostanti" e lo fa "in tutte le direzioni". E proprio per questo la perturbazione che si crea ha una forma circolare.

2) Un improbabile osservatore immerso nel laghetto a qualche metro di distanza da dove è caduto il sasso con una maschera subacquea in posizione tale che il filo dell'acqua corra orizzontalmente dividendo a metà la visione sopra e sott'acqua, vedrebbe passare la perturbazione sul vetro della maschera osservando il passaggio di un'onda sinuosa. Se nel campo visivo della maschera ci fosse una fogliolina o il sughero della lenza di un pescatore, il nostro subacqueo osserverebbe la fogliolina/sughero alzarsi ed abbassarsi senza spostarsi lateralmente in nessuna direzione.

Anche la *hola* allo stadio funziona esattamente così: lì, non è l'energia potenziale del pendolo o della lama/corda o del sasso, ma l'energia dell'entusiasmo agonistico che qualcuno in certo momento decide di sprigionare e che tutti assecondano alzandosi e rimettendosi seduti con perfetto tempismo a dar vita alla *hola*, che di per sé non esiste, eppure è ben riconoscibile dal lato opposto dello stadio. Anche in questo caso, come in quello della fogliolina, nessuno si è mosso dal suo posto ma, semplicemente, al momento giusto si è alzato e poi si è abbassato.

3) I pescatori del laghetto, anche se non hanno visto il sasso cadere in acqua e non hanno sentito il tonfo, osservano comunque il movimento sinuoso del sughero della propria lenza. Cosa ancor più importante è che ciascun pescatore fa questa osservazione in momenti diversi l'uno dall'altro a seconda della distanza dal punto dove è caduto il sasso. Sassi grossi provocano onde alte, sassi piccoli provocano onde basse ma, cosa ancora molto importante da sottolineare, tutte con la stessa velocità di propagazione.

La velocità di propagazione dell'onda dipende soltanto dalla densità dell'acqua. Se il liquido fosse olio, che ha una densità minore dell'acqua (infatti galleggia sull'acqua), la velocità di propagazione sarebbe minore. Se fosse mercurio, che a temperatura ambiente è liquido ed ha una densità decisamente più alta dell'acqua, la velocità di propagazione sarebbe più alta.

Quella che fino ad ora abbiamo chiamato *perturbazione dell'aria*, e cioè accumulo e rarefazione delle particelle d'aria, come succede nell'esempio dello spintone tra la folla, si definisce più formalmente come *variazione della pressione dell'aria* rispetto alla pressione atmosferica di riferimento. Quanto descritto per il laghetto e per il Teatro può essere visualizzato con il seguente diagramma:

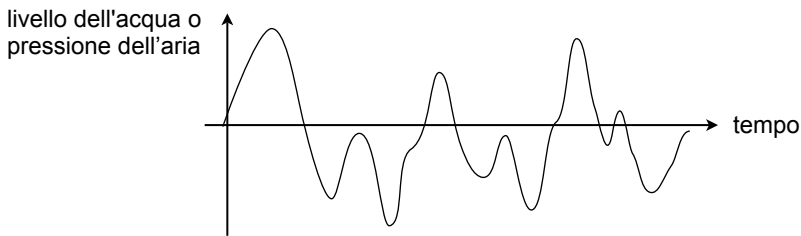


Figura 1.9 Variazione della pressione dell'aria e del livello dell'acqua in un punto

Attenzione: il grafico di Figura 1.9 non è la descrizione di quello che si vede in due dimensioni sulla superficie del laghetto o che si immagina accadere nelle tre dimensioni dello spazio del Teatro, ma quello che accade *nel tempo in un solo punto di osservazione*. Nella Figura 1.1 la membrana, che nell'esempio in due dimensioni del laghetto corrisponde al sughero delle lenze ed in quello in tre dimensioni del Teatro corrisponde al timpano dell'orecchio dello spettatore, rileva quello che accade in un solo punto e lascia la traccia di ciò che accade in quel punto nel tempo con continuità.

Nel primo caso, la traccia lasciata corrisponde alla *variazione nel tempo* del livello dell'acqua rispetto a quello di quiete, mentre nel secondo caso corrisponde alla *variazione nel tempo* della pressione dell'aria rispetto allo stato di quiete delle particelle dell'aria e quindi di silenzio.

L'asse orizzontale, detto *asse delle ascisse*, riporta il trascorrere del tempo, mentre l'asse verticale, detto *asse delle ordinate* dice, in corrispondenza di ogni istante del tempo, quanto vale la pressione.

L'orecchio, stimolato da tali onde, è sensibile comunque soltanto a quelle vibrazioni la cui frequenza è mediamente compresa tra 16 Hz e 16000 Hz; quest'intervallo frequenziale, definito *campo di udibilità* umano, è quindi quello significativo a livello percettivo. Al di là dei limiti inferiore e superiore di tale intervallo frequenziale si hanno rispettivamente gli **infrasuoni** e gli **ultrasuoni**, così definiti proprio in relazione a noi umani e non hanno significato se non quello del *silenzio*. Silenzio, tuttavia, non significa necessariamente "assenza di suono": sappiamo che molti animali sentono e reagiscono a fischi, che noi con percepiamo perché al di sopra della nostra soglia frequenziale di udibilità.

1.6 Altezza e scale musicali

Al variare della frequenza di un segnale sonoro varia anche la relativa percezione di altezza; ma una variazione di 100 Hz, ad esempio, tra 800 Hz e 900 Hz, non è così netta come tra 200 Hz e 300 Hz, perché a livello percettivo, sulla base della fisiologia dell'orecchio umano, la relazione che intercorre tra la frequenza di un suono e l'altezza percepita a livello psico-acustico è di tipo logaritmico. Quindi, per poter percepire la stessa variazione rilevata, per esempio, tra un suono a 200 Hz e uno a 400 Hz dove si ha un raddoppio della frequenza, è necessario passare da 400 Hz a 800 Hz.

In Natura i suoni esistono senza soluzione di continuità dal grave all'acuto, come succede ad esempio con una sirena d'allarme, mentre in ambito musicale le note costituiscono un "insieme discreto" delle frequenze udibili. La continuità frequenziale che il nostro apparato uditivo è in grado di percepire quello che in musica viene definita come *glissato* che si può ottenere facilmente nel violino e negli altri strumenti della famiglia a corda con manico liscio.

Per visualizzare le due diverse modalità di *continuo* e *discreto* basta pensare ad uno scivolo ed una scalinata (§3.4, Figura 3.15): con entrambe si può salire e scendere a livelli diversi, ma sulla scalinata ci sono soltanto livelli predefiniti, senza livelli intermedi. E non a caso in musica si parla di *scala*.

Perciò, se ad una nota corrisponde una ben precisa frequenza, non è detto che ad una frequenza corrisponda una nota. Il problema è allora come scegliere, nelle infinite possibilità che il continuo offre, i gradini della scala, e cioè le singole note.

Sin dall'antichità si è sempre cercato di definire un modello formale per rappresentare la suddivisione dell'ottava e scegliere le note per formare le scale su cui modulare melodie. Si tratta di una lunga storia che affonda le sue radici nella cultura greca classica di cui, purtroppo, per quello che riguarda la Musica, i pochi frammenti che ci sono pervenuti non sono sufficienti per una corretta ricostruzione. Anche per gli strumenti musicali possiamo fare affidamento soltanto su dipinti che raffigurano musicisti che suonano l'aulos, la cetra e il flauto di pan, dai quali è possibile soltanto una sommaria ricostruzione del reale "sound" della musica greca antica.

Fortunatamente però esiste una, per così dire, "memoria digitale" sulla natura della musica greca: si tratta dell'analisi aritmetico/geometrica che Pitagora di Samo (575-495 a.C.) fece sull'esistente cultura musicale dell'epoca, pervenuta a noi attraverso gli scritti di epoche successive da parte di Gaudenzio, Nicomano e Boezio.

La domanda che sta alla base dell'analisi di Pitagora è la seguente: la scala musicale in uso da secoli nel mondo greco è di origine culturale oppure

“risiede” già in Natura? È così perché “dettata dalla Natura” come ad esempio il senso del ritmo è connaturato con il battito del cuore, il passo e la corsa, oppure esistono precise cause e motivazioni storiche e culturali che l’hanno così definita? La risposta sta tra le due considerazioni in misura di una poco definibile percentuale a favore dell’una o dell’altra.

Sicuramente è innata la sensazione dell’*ottava*, rappresentata da un intervallo che abbraccia le altezze comprese tra un valore frequenziale f e il suo doppio $2f$ come ad esempio la differenza frequenziale che esiste tra un Do ed il Do successivo (si chiama ottava perché corrisponde alla distanza tra il primo e l’ottavo grado della scala musicale della nostra cultura occidentale). Pitagora fa questa considerazione confrontando le lunghezze (l’una la metà dell’altra) di una corda e di una flauto di canna.

Poi ci sono le *armoniche naturali*, dette anche *suoni armonici*, date dai multipli della frequenza-base (ottenibili per esempio con diverse modalità di soffio nell’imboccatura del flauto) che sono, per loro natura, consonanti tra loro.

Detta f la frequenza di riferimento, i multipli sono dati dal doppio $2f$, dal triplo $3f$, dal quadruplo $4f$... e così via. La frequenza di riferimento f viene detta anche la **fondamentale** e la numerazione delle armoniche coincide con il fattore moltiplicativo, per cui $2f$ è la seconda armonica, $3f$ la terza... e così via.

Si noti che la seconda armonica coincide con l’ottava (in senso musicale) della fondamentale e l’ottava dell’ottava coincide con la quarta armonica.

Il fatto che le note della scala cadano in qualche modo sulle armoniche naturali e cioè sui multipli della fondamentale, ha come diretta conseguenza di avere suoni “consonanti” tra loro, vale a dire, nel senso etimologico del termine, che “con-suonano” e si amalgamano bene insieme. In breve, sono “naturalmente” intonate fra loro... e le ragioni sono molteplici. Consideriamo allora il seguente grafico che riporta i multipli della fondamentale, le armoniche e le ottave. Per comodità descrittiva, fissiamo arbitrariamente la fondamentale alla frequenza di 100Hz.

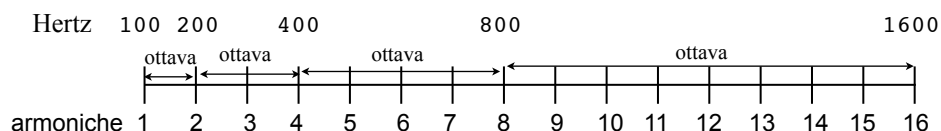


Figura 1.10 Ottave e armoniche naturali

Si osserva che le armoniche si distribuiscono sull'asse delle frequenze in maniera lineare - com'è giusto che sia - e che le ottave della fondamentale si trovano a distanze esponenziali - com'è giusto che sia - in base all'effetto psico-acustico che ciò ha sul nostro apparato auditivo.

Considerando 100 Hz come la frequenza della *nota-base*, e cioè il primo grado della scala, le note-base delle ottave superiori si trovano a 200, 400, 800 e 1600 Hz. All'interno di ogni ottava si trova perciò un numero sempre maggiore di armoniche: nella seconda ottava 200÷400 Hz troviamo soltanto l'armonica a 300 Hz; nella terza ottava 400÷800 Hz troviamo le armoniche 500, 600 e 700. Nella quarta ottava compresa tra 800 e 1600 Hz troviamo ben 7 armoniche.

Definiamo allora *ottava-base* quella compresa tra 100 Hz e 200 Hz sulla quale costruire la scala musicale seguendo il metodo di Pitagora (Figura 1.11a).

La terza armonica, quella a 300 Hz, si trova nella seconda ottava e perciò per riportarla nell'ottava-base è necessario dividerla per 2, ottenendo 150 Hz.

Questo *intervallo* corrisponde, come vedremo, al *quinto grado* della scala e perciò si chiama *intervallo di quinta* con la fondamentale e si indica con V.

Quindi "terza armonica diviso 2" dice che la V è data da $3/2$ del I grado

$$V = 3/2 I$$

Anzi, dice di più: dice che moltiplicando per $3/2$ la frequenza di un qualsiasi suono, si ottiene un nuovo suono alla distanza intervallare di una *quinta*. Allora, moltiplicando per $3/2$ la V si ottiene $9/4$, che corrisponde al II grado appartenente all'ottava superiore all'ottava-base. Perciò la II dell'ottava-base si ottiene dividendo per 2 il rapporto $9/4$ ottenuto con la V della V e cioè:

$$II = (3/2) \times (3/2) \times (1/2) I = 9/8 I$$

La figura seguente mostra quanto fin qui detto indicando con le lettere dell'alfabeto la successione dei passaggi:

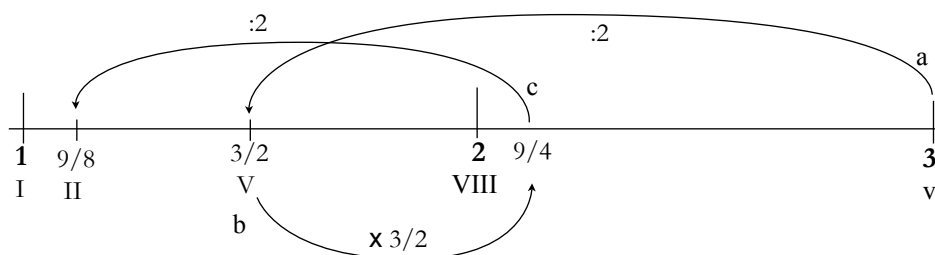


Figura 1.11a Metodo pitagorico per la costruzione della scala musicale

Il metodo pitagorico basato sulle quinte è dunque il seguente:

- 1) il nuovo grado si ottiene moltiplicando per $3/2$ l'ultimo trovato e
- 2) se cade nell'ottava-base (se cioè il rapporto ottenuto è <2) allora è quello;
- 3) se cade oltre l'ottava-base (se cioè il rapporto ottenuto è >2) allora si divide per 2 (cioè si moltiplica per $1/2$).

La tabella seguente riporta i valori così ottenuti nell'ordine per i gradi V, II, VI, III e VII ed espressi infine in relazione al I grado.

V	=	$3/2$ I		
II	=	$3/2$ V ($\times 1/2$)	= $(3/2) \times (3/2) \times (1/2)$ I	= $9/8$ I
VI	=	$3/2$ II	= $(3/2) \times (9/8)$ I	= $27/16$ I
III	=	$3/2$ VI ($\times 1/2$)	= $(3/2) \times (27/16) \times (1/2)$ I	= $81/64$ I
VII	=	$3/2$ III	= $(3/2) \times (81/64)$ I	= $243/128$ I

Nella figura seguente viene riportato il procedimento delle quinte indicando con le lettere minuscole dell'alfabeto la successione dei passi, dove una freccia in avanti moltiplica per $3/2$ il rapporto di partenza e una freccia all'indietro lo divide per 2.

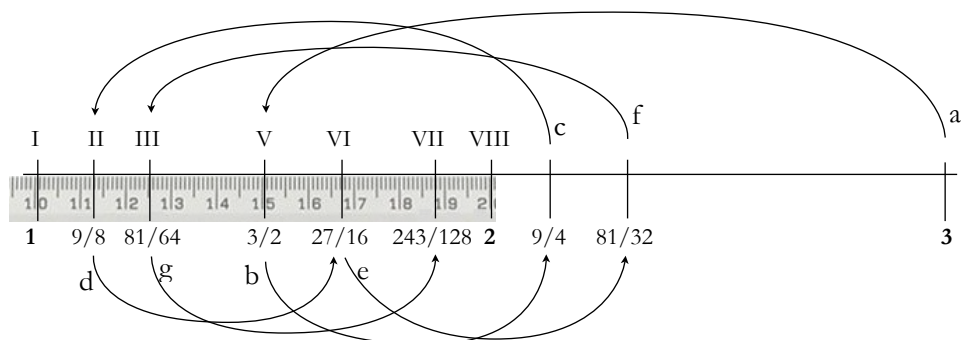


Figura 1.11b Metodo pitagorico per la costruzione della scala musicale

Alla scala manca soltanto il IV grado. Purtroppo con il sistema pitagorico delle quinte si intravede una strada molto lunga per arrivare ad un ipotetico “VII grado bemolle” (peraltro di moderna concezione) che moltiplicato per $3/2$ darebbe il IV grado. Infatti, continuando a moltiplicare per $3/2$ a partire

dal VII grado (243/128) in poi, i valori al numeratore ed al denominatore diventano ben presto numeri a 4, 5 e 6 cifre, per ottenere finalmente 177147/131072 pari a 1,35152... che dovrebbe corrispondere al IV grado.

Dovrebbe, perché in effetti moltiplicando 1,35152... per 3/2 si ottiene 2,027... che, anche se di poco, è superiore a 2 e non uguale 2 come dovrebbe essere il valore aspettato per l'VIII grado. Lo scarto di 0,027 è il cosiddetto *comma pitagorico*, anche se è poco probabile che sia stato lo stesso Pitagora ad definirlo così e soprattutto ad ottenere questo risultato in assenza delle attuali modalità di calcolo aritmetico in un contesto storico/culturale in cui la musica era per altro basata sulle *scale modali* a sette note: *ionica, dorica, frigia, lidia, misolidia, eolia e locria*.

Pitagora, perciò, trova il IV grado al di fuori della progressione delle quinte, imponendo “a ritroso” quel rapporto che moltiplicato per 3/2 dà 2 (e cioè l'VIII grado): questo rapporto è 4/3 perché $(4/3) \times (3/2) = 2$.

La scala risulta essere la giustapposizione di due semi-scale uguali tra loro di quattro note (I-II-III-IV e V-VI-VII-VIII) ciascuna formata da due intervalli “grandi” detti *tono* ed un intervallo “piccolo” detto *mezzotono*. Ciascuna delle due semi-scale *tono-tono-mezzotono* simmetriche e consonanti, costituisce il *tetracordo*.

Oltre ad essere consonante dal punto di vista aritmetico/acustico, la scala di sette note sembra essere “cònsona” al pensiero greco: il tetracordo *tono-tono-mezzotono* (che sta alla base anche della moderna modulazione armonica II-V-I ben nota ai jazzisti) dà la sensazione di una conclusione di un discorso sugli intervalli III-IV e VII-VIII; e questo sembra essere molto attinente al pensiero logico-deduttivo del sillogismo di cui è permeato il pensiero greco e la nostra cultura occidentale. Una controprova può essere trovata nel fatto che la scala pentatonica orientale, dove prevale un senso di sospensione e di un non definito punto di partenza e/o di arrivo, è più cònsona al pensiero *tao* e *zen*, che è tutt'altro che logico-deduttivo.

La scala musicale pitagorica completa è dunque:

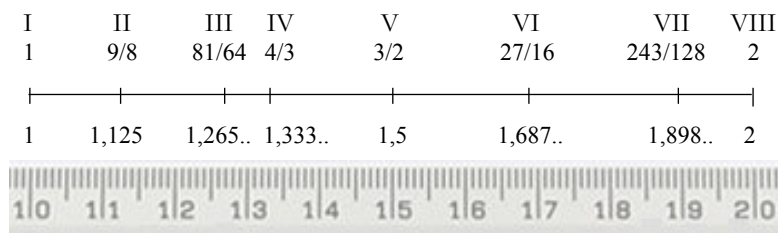


Figura 1.12 Scala musicale pitagorica

La cosa interessante che risulta applicando il metodo pitagorico delle quinte, è che il rapporto tra i toni interi (I-II, II-III, IV-V, V-VI e VI-VII) è sempre $9/8=1,125$

$$\begin{aligned} \text{II/I} &= (9/8)/1 && = 9/8 = 1,125 \\ \text{III/II} &= (81/64) / (9/8) && = 9/8 = 1,125 \\ \text{V/IV} &= (3/2) / (4/3) && = 9/8 = 1,125 \\ \text{VI/V} &= (27/16) / (3/2) && = 9/8 = 1,125 \\ \text{VII/VI} &= (243/128) / (27/16) && = 9/8 = 1,125 \end{aligned}$$

e per i “mezzo-tono” (III-IV e VII-VIII) il rapporto è $256/243 \approx 1,0535$

$$\begin{aligned} \text{IV/III} &= (4/3) / (81/64) = 256/243 = 1,0534979... \approx 1,0535 \\ \text{VIII/VII} &= 2 / (243/128) = 256/243 = 1,0534979... \approx 1,0535 \end{aligned}$$

Purtroppo, però, quello che abbiamo definito “mezzo-tono” non è affatto la metà del tono: infatti moltiplicando per se stesso il rapporto che rappresenta il “mezzo-tono” dovremmo ottenere il rapporto che rappresenta il tono intero; invece $(256/243) \times (256/243)$ non è uguale a $9/8$. In forma esplicita:

$$1,0535 \times 1,0535 = 1,0985... \text{ che è un valore minore di } 1,125$$

Questo risultato aritmetico ha determinato conseguenze molto importanti nell’evoluzione della musica sia dal punto di vista della teoria musicale in termini di “alfabeto, grammatica e sintassi” sia da quello pratico nella strutturazione meccanica degli strumenti musicali.

Ora si può immaginare che nel corso dei secoli i musicisti, ciascuno a proprio modo e in base alle possibilità che lo strumento usato consentiva, abbiano esplorato nelle proprie composizioni l’uso di note che non fossero soltanto quelle della scala pitagorica, cercandole “a orecchio” proprio là “in mezzo” ai toni interi I-II, II-III, IV-V, V-VI e VI-VII.

Ma quando si eseguono i dovuti calcoli usando il rapporto $256/243$ del “mezzo-tono” pitagorico IV/III e VIII/VII per trovare il “presunto semitono” in mezzo agli intervalli interi, succede che moltiplicandolo col precedente e dividendolo dal successivo, si ottengono due risultati diversi tra loro.

Per esempio, tra il secondo ed il terzo grado si ha:

$$\begin{aligned} \text{dal II in avanti (\#): } & (9/8) \times (256/243) = 32/27 = 1,185... \\ \text{dal III indietro (b): } & (81/64)/(256/243) = 19683/16384 = 1,2013... \end{aligned}$$

Infatti, moltiplicando per se stessa 12 volte la radice dodicesima di 2 si ottiene 2 ($r^{12}=2$). Quindi, presa una nota a caso come il Do dell'esempio appena visto, il Do dell'ottava superiore è dato dal valore frequenziale doppio di quello di partenza, come doveva essere.

Con il temperamento equabile, adottato dal XVIII secolo in poi, le note *diesizzate* coincidono con le successive note *bemollizzate*. Il *do#* quindi diventa uguale al *reb*, il *mi#* al *fa naturale*, ecc.. In sede di accordo internazionale è stata fissata un'altezza standard che definisce il La centrale a 440 Hz, dalla quale è possibile ricavare la frequenza di tutte le altre note.

Il termine *temperamento* è stato mutuato dal Disegno (*temperare le matite*) e dalla Pittura dove *temperare* vuol dire *predisporre all'uso la tavolozza dei colori*, che nel caso della musica diventa la "tavolozza delle note".

Per misurare piccole differenze frequenziali, è stato poi definito il *cent* come *centesima parte del semitono* e quindi *1200a parte dell'ottava*:

$$\text{cent} = \sqrt[100]{1,059463\dots} = \sqrt[1200]{2} = 1,000578\dots$$

La seguente tabella mette a confronto le due scale in termini di *rapporto* frequenziale con il primo grado e *distanza* in numero di cent dal primo grado:

Grado	Scala pitagorica		Scala equabile		differenza in cent
	rapporto	cent	rapporto	cent	
I	1	0	1		0
I# - II^b			1,05954..	100	
II	1,125	204	1,1224...	200	4
II# - III^b			1,1892....	300	
III	1,2656	408	1,2599....	400	8
IV	1,3333	498	1,3348....	500	-2
IV# - V^b			1,4142...	600	
V	1,5	702	1,4983..	700	2
V# - VI^b			1,58740...	800	
VI	1,6875	906	1,6818...	900	6
VI# - VII^b			1,78179...	1000	
VII	1,8984	1110	1,8877....	1100	10
VIII	2	1200	2,000	1200	0

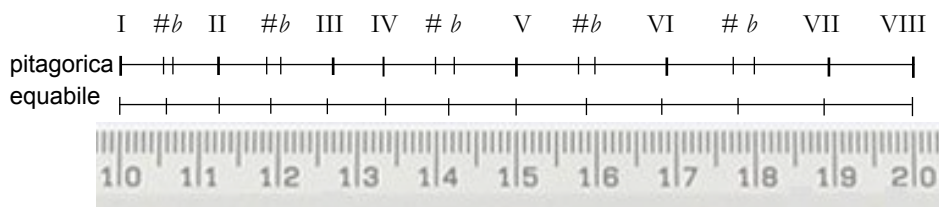


Figura 1.15 Confronto tra scala pitagorica e scala equabile

L'equidistanza in frequenza dei suoni della scala temperata risponde contemporaneamente a più requisiti, per cui diventata lo standard di riferimento della Musica, con più di 3 secoli di storia e molti generi e stili diversi tra loro.

Va innanzitutto osservato che la progressione basata sul rapporto irrazionale della *radice dodicesima di 2* = 1,059463... individua con sorprendente precisione (scarti di pochi cent, psico-acusticamente insignificanti) le note naturali derivanti dai rapporti interi, garantendo così anche la "compatibilità" della scala temperata con la scala naturale di Pitagora e quindi con la tradizione musicale.

Allo stesso tempo, con assoluta novità, la progressione è coerente con la fisiologia dell'orecchio e la naturale sensazione di uniforme gradualità dei semitoni temperati all'interno di un'ottava, già fisiologicamente esistente nel salto esponenziale di ottava.

E finalmente 1 tono intero è fatto di 2 semitoni!

La Figura 1.16 illustra l'andamento esponenziale delle frequenze rispetto all'andamento lineare del numero di semitono.

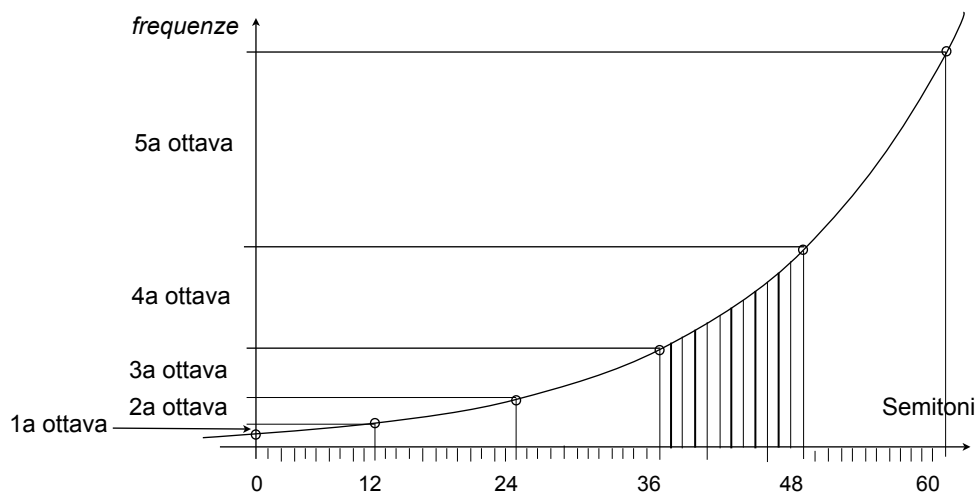


Figura 1.16 Andamento esponenziale delle note della scala temperata

Una volta definita, la scala temperata introduce il *concetto di astrazione* relativo alla combinazione orizzontale (melodia) e verticale (armonia) delle note tant'è che, in Musica, piuttosto che di “nomi” si preferisce parlare di *distanze intervallari*: invece di dire *Do-Mi* si dice *intervallo di terza* (4 semitoni); invece di dire *Re-La* si dice *intervallo di quinta* (7 semitoni); e così via.

Un'intera composizione può essere “spostata” verso il basso o verso l'alto semplicemente cambiando una nota di riferimento avendo cura di mantenere inalterate tutte le distanze intervallari.

È quello che si chiama *cambio di tonalità* che, insieme al meccanismo di *modulazione*, sta alla base della moderna Armonia a partire dal *Clavicembalo ben Temperato* (1722) con cui Bach dimostrò e consacrò definitivamente l'efficienza e la superiorità di questo nuovo sistema di riferimento musicale rispetto a quello (o quelli) del passato.

L'andamento esponenziale/logaritmico dei valori di frequenza rispetto a quello lineare della successione di semitoni temperati, è diretta conseguenza della formula iterativa $f_{n+1} = f_n \times r$ ed è riscontrabile grossolanamente nella sagoma del pianoforte a coda, nel profilo della sequenza delle canne d'organo e del flauto di pan, e nella tastiera del manico di una chitarra.

1.7 Intensità

L'*intensità acustica* riguarda ciò che comunemente viene chiamato il *volume* di un suono dal *piano* al *forte*: si dice infatti ‘...alzare e abbassare il volume...’. L'*intensità acustica* è una grandezza della Fisica acustica, mentre il volume sonoro è la corrispondente reazione percettiva del nostro apparato uditivo.

L'*intensità acustica* si misura in *decibel* ma... il decibel *non* è l'unità di misura dell'*intensità acustica*.

Se qualcuno vi dicesse che ‘...ieri, in discoteca 0 (zero!) decibel mi hanno spaccato i timpani...’ rispondereste che ‘..110-120 decibel.. spaccano i timpani!’. In realtà avreste ragione entrambi, anche se formalmente avreste usato, come si vedrà nel seguito, in modo non del tutto corretto la parola *decibel*.

L'*intensità acustica* ha a che fare con i concetti di *energia*, *forza*, *potenza* e *pressione*.

Nella Fisica classica, in special modo nella Meccanica, l'energia viene definita come la *capacità di un sistema di generare una forza per compiere un'azione*. Ora, poiché una qualunque azione viene svolta da una *forza* che è espressione di energia, la definizione appena vista, è di fatto “circolare” ...e non porta a niente. In realtà tutte le “cose” della Fisica vengono definite in modo circolare: la Fisica è infatti il settore di indagine che analizza i *fenomeni* (*ciò che appare*) per comprenderne il *comportamento* e li formalizza attraverso formule matematiche, senza però poter arrivare al *perché* ed all'*essenza* delle “cose”, perché... “perché non ci è dato!” L'Energia *esiste* di per sé in varie forme dal momento della Creazione/BigBang e non si crea né si distrugge ma, semplicemente, si *trasforma*. Quindi, termini di uso corrente come “produzione” e “sorgente” di energia in riferimento, ad esempio, alle centrali elettriche, non sono corretti. Corretto è dire, caso mai: *produzione di energia elettrica dalla trasformazione di energia eolica, idrica, fotovoltaica, ecc.* che a loro volta sono date dalla trasformazione di energia termica e luminosa proveniente dal Sole ...che a loro volta sono date dalla trasformazione di energia nucleare (esistente dalla Creazione/BigBang.....).

Il suono è una forma di energia che risulta dalla trasformazione di energia meccanica trasmessa all'*aria*: per gli strumenti a fiato si tratta della trasformazione diretta dell'energia proveniente dal sistema respiratorio. Per gli strumenti a corda si tratta di un doppio passaggio di energia: dagli arti alle corde, dalle corde all'aria.

E poiché il suono è un fenomeno che si manifesta nel tempo, è corretto parlare di *flusso di energia* che esprime la *quantità di energia trasformata nell'unità di tempo*: minore è il tempo, maggiore è la velocità (e viceversa) con cui l'energia viene trasformata. Tale velocità è legata alla *potenza* del dispositivo che trasforma l'energia: un motociclo che da fermo raggiunge una velocità di 100Km/h in 10 secondi è più potente di uno che, a parità di utilizzo della stessa quantità di energia (benzina), la raggiunge in 25 secondi. La *potenza* si misura in *watt* dal nome dell'ingegnere scozzese James Watt (1736-1819).

Poiché il suono si manifesta come energia trasmessa all'aria che si espande in modo sferico, è possibile pensarla come *forza esercitata su una superficie* come succede per l'appunto sul timpano dell'orecchio, e viene definita *pressione* che nel nostro caso è *pressione sonora* o *pressione acustica*.

La *pressione sonora* in realtà è data dalla *variazione della pressione atmosferica* dovuta al passaggio di un'onda sonora perché, come si sa, il suono è provocato da un oggetto che, oscillando, crea zone di *compressione* e *de-compressione* che si espandono a sfera.

La quantità di particelle d'aria spostate, e quindi la pressione sonora, è proporzionale alla superficie dell'oggetto oscillante ed all'ampiezza dell'oscillazione stessa: una lama di metallo investe una quantità di particelle d'aria maggiore di quelle di una corda di violino o di chitarra così come un ventaglio sposta una quantità d'aria maggiore rispetto ad una mano agitata!

Il *flusso di energia* di un'onda sonora relativa alla superficie su cui viene esercitata la pressione sonora, e cioè la quantità di energia che attraversa o impatta una superficie di area stabilita in un intervallo di tempo stabilito, viene definita *intensità acustica* ed è data perciò dal rapporto tra *potenza* e *superficie*:

$$\text{intensità acustica} = \text{Potenza/Superficie}$$

L'intensità acustica si misura in *watt su metro quadrato* (W/m^2) ed è responsabile della sensazione psico-acustica di maggiore o minore volume sonoro nel senso comune del termine. L'intensità acustica minima percepibile dal sistema uditivo di un soggetto adulto mediamente sano, è di 10^{-12} Watt/m^2 (all'incirca quello di una zanzara in avvicinamento notturno) ed è chiamata *soglia di udibilità*; l'intensità acustica massima è di $1 \text{ Watt}/\text{m}^2$ (all'incirca quella provocata dall'esplosione di ordigno) ed è chiamata *soglia del dolore*.

10^{-12} vuol dire 0,000.000.000.001 e cioè un milionesimo di un milionesimo. Per avere un'idea precisa di cosa sia il rapporto tra 10^{-12} e 1 (che è lo stesso che c'è tra 1 e 1.000.000.000.000) basti pensare al rapporto che c'è tra 1 millimetro ed un milione di milioni di millimetri e cioè 1.000.000 Km che equivalgono a tre volte la distanza Terra-Luna, come illustrato nella Figura 1.20.

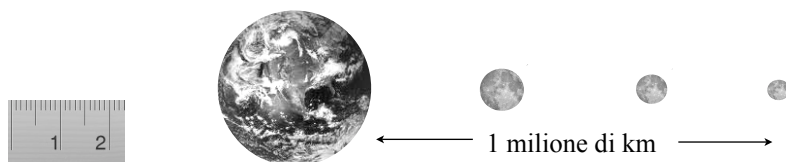


Figura 1.20 1 millimetro e 1 milione di chilometri

Ora, siccome è significativo parlare di millimetri e centimetri fino a quando si parla, ad esempio, di misure domestiche come mobili, mattonelle, infissi... non lo è più quando si parla di distanze tra una città e l'altra: non ha senso dire che la distanza tra Roma e Milano è 453.984.754 millimetri!!...

Per la verità, a qualunque ambito ci si riferisca di solito sono sufficienti 2 cifre (al massimo 3) di precisione: se un turista vi chiede la distanza per una strada, un monumento, o la stazione ferroviaria della vostra città, direte 700 metri, 1 Km e 2; non direte mai (anche ammesso che lo sappiate con precisione) 752 metri. La distanza tra Roma e Milano è 450 Km e non 453 Km. Due cifre sono sufficienti. La distanza tra due capitali europee è migliaia, centinaia di chilometri: Roma-Parigi 1.500 Km, Madrid-Londra 2.300 Km. E così per il tempo. Se si parla di oggi si dirà: 1 ora e 1/4; se si parla dei giorni passati si dirà 10-12 giorni fa, senza menzionare le ore; e poi 3-4 settimane fa; e poi 2 mesi fa...

Per questo motivo, rappresentare in maniera lineare (§1.4) un intervallo di valori che va da un millimetro ad una lunghezza pari a 3 volte la distanza Terra-Luna non è significativo né pratico. Quello che conta è allora l'*ordine di grandezza* e l'ordine di grandezza corrisponde all'esponente del 10 per cui 0 ($10^0=1$) vuol dire millimetri, 1 ($10^1=10$) vuol dire centimetri, 2 ($10^2=100$) vuol dire decimetri, 3 ($10^3=1000$) vuol dire metri, 6 vuol dire chilometri, ...e così via.

Ma l'esponente che si dà alla base, che nel nostro caso è 10 ed esprime l'ordine di grandezza, è per definizione il *logaritmo*. Proprio per questo si preferisce una misura logaritmica che consente in maniera pratica e significativa una *rappresentazione compressa* di un grande intervallo di valori in cui, oltretutto, sono significative solo 2 cifre di precisione.

L'espedito matematico di usare il logaritmo in base 10 del valore numerico anziché usare il valore numerico stesso risultò piuttosto geniale ed efficace ed in un certo qual modo pareva "spreco" se utilizzato nel solo caso dell'intensità acustica. Perché in effetti ci si accorse subito che la misura logaritmica poteva essere generalizzata ed applicabile a tutte quelle grandezze fisiche che presentano una grande escursione di valori.

Il primo caso si presentò negli anni '20 del secolo scorso, nella misura dell'attenuazione della potenza del segnale sulle linee telefoniche dei laboratori Bell e fu all'inizio chiamato *Transmission Unit* e ribattezzato *bel* in onore di Alexander Graham Bell dopo la sua morte. Perciò fu deciso di applicare il logaritmo non tanto al valore della potenza da misurare ma piuttosto al *rapporto* tra il *valore-da-misurare* P ed un *valore-di-riferimento* P_r prescelto. In questo modo il valore ottenuto è un numero puro e perciò non riferito ad alcuna grandezza fisica.

$$bel = \log_{10} \left(\frac{P}{P_r} \right)$$

Il valore di riferimento può essere la soglia di udibilità $P_u=10^{-12}$ Watt/m² oppure la soglia del dolore $P_d=1$ Watt/m². Se si sceglie $P_r=P_u$ allora l'intensità acustica relativa, ad esempio, traffico stradale medio che è dell'ordine di grandezza di 10^{-5} Watt/m², misurata in *bel* è:

$$\log_{10} \left(\frac{10^{-5}}{10^{-12}} \right) = \log_{10} (10^7) = 7bel$$

La seguente tabella mostra i valori di livello di intensità espressi in *bel* relativi ad ambienti e sorgenti di rumore.

Sorgente	watt/m ²	log ₁₀ (watt/m ²) = bel
Aereo a reazione al decollo (soglia del dolore)	1,0 = 10 ⁰	12
Discoteca	0,1 = 10 ⁻¹	11
Martello pneumatico	0,01 = 10 ⁻²	10
Stazione ferroviaria	0,001 = 10 ⁻³	9
Stadio	0,0001 = 10 ⁻⁴	8
Traffico stradale	0,00001 = 10 ⁻⁵	7
Sala da concerto	0,000001 = 10 ⁻⁶	6
Mensa sociale	0,0000001 = 10 ⁻⁷	5
Pranzo a casa: forchette contro piatti	0,00000001 = 10 ⁻⁸	4
Conversazione	0,000000001 = 10 ⁻⁹	3
Biblioteca	0,0000000001 = 10 ⁻¹⁰	2
Bisbiglio	0,000000000001 = 10 ⁻¹¹	1
Zanzara (soglia di udibilità)	0,0000000000001 = 10 ⁻¹²	0

Dalla tabella si vede che l'intensità acustica di una sala da concerto è di circa 10^{-6} Watt/m² corrispondenti a 6 bel; ma dal punto di vista psico-acustico il nostro apparato uditivo è in grado di percepire variazioni di livello molto più raffinate dei due livelli 6 e 7 bel. Per cui in realtà i livelli intermedi espressi con una cifra dopo la virgola come 6,1-6,2 ... 6,9 bel, sono necessari ed in pratica, ancora una volta, 2 cifre sono significative e sufficienti.

Allora, invece di usare il *bel* nell'intervallo $0\div 12$ con una cifra dopo la virgola, si preferisce moltiplicare per 10 il bel ed avere valori che vanno da 0 a 120 e oltre. Da qui il *decibel*: deci-bel = 10 bel = $10 \log_{10} (P/P_0)$.

La definizione formale e corretta di *decibel* è dunque

10 volte il logaritmo in base 10 del rapporto tra due potenze

e si abbrevia con il simbolo *dB*.

Un aumento di volume sonoro percepito come raddoppio, corrisponde a circa +3dB; infatti, ponendo $p_2=2p_1$ si ha:

$$10 \log_{10} (p_2/p_1) = 10 \log_{10} (2p_1/p_1) = 10 \log_{10} (2) = 10 \times 0,3010299... \approx 3dB.$$

Siccome il rapporto si può fare anche tra il valore da misurare e il valore della soglia del dolore $P_d=1\text{watt}/\text{m}^2$, allora misurando il valore massimo $P_{\text{max}}=P_d$ si ha: $10 \log_{10} (P_{\text{max}}/P_d) = 10 \times \log_{10} (1) = 0 \text{ dB}$e da qui la spiegazione della frase con cui abbiamo cominciato il paragrafo relativa alla discoteca!

Considerando la curva logaritmica riportata nel grafico seguente, possiamo notare che nel primo caso (riferimento= P_u) i valori stanno sulla parte positiva della curva e vanno da 0dB (min) a 120÷130dB (max); nel secondo caso (riferimento= P_d) i valori vanno da $-\infty\text{dB}(\text{min})$ a 0dB(max) e stanno sulla parte negativa della curva logaritmica. In questo secondo caso, nella pratica viene considerato il valore di -60dB come soglia minima corrispondente al “silenzio”.

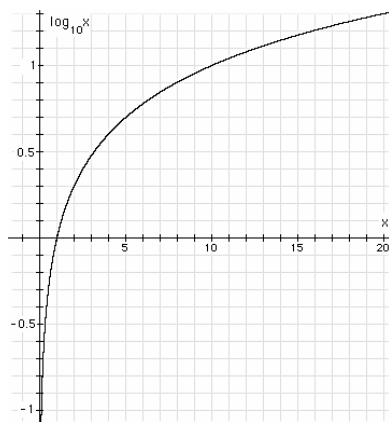


Figura 1.21 Andamento della funzione logaritmica

1.8 Timbro

A differenza di quanto si è visto per l'altezza e l'intensità, il **timbro** del suono non può essere espresso con un valore numerico rispetto ad un'unità di misura: il timbro è infatti un *parametro* che, in stretta relazione alla fonte di emissione del suono, definisce il particolare *colore* in grado di qualificarlo come appartenente a sorgenti diverse. Il nostro sistema uditivo è molto efficiente nel riconoscere il timbro dei vari strumenti musicali, della voce delle persone e dei rumori della natura e degli oggetti percossi, spostati, ecc.; ma non esiste un modo per definire il timbro in maniera analitica.

Proprio per questo, di solito, i timbri vengono caratterizzati per associazione ai vari strumenti musicali o al tipo di sorgente di emissione sonora: si parla infatti del timbro del pianoforte, della chitarra ed anche, in maniera più qualitativa, di timbro metallico, di timbro dolce, cupo, e così via. Nella voce cantata, ad esempio, le diverse timbriche sono espresse sia dai vari registri del canto (basso, tenore, contralto e soprano) sia dai tipi di voce caratteristiche del linguaggio parlato: voce afona, profonda, argentina, ecc.

Timbri diversi comunque non caratterizzano sempre sorgenti diverse: anche uno stesso strumento musicale può cambiare il proprio timbro se suonato con modalità diverse. Basti pensare al suono della tromba con la sordina.

Si tratta sempre e comunque di considerazioni qualitative del tutto informali e lontane dal concetto di “misura” visto nei casi dell'altezza e dell'intensità. Per arrivare ad una definizione di timbro il più possibile formale è necessario introdurre il concetto di **forma d'onda** e di **spettro**, che saranno trattati in modo più preciso nel prossimo capitolo.

Per introdurre il concetto di forma d'onda e di spettro si può fare un'analogia, con le dovute differenze, con ciò che accade nella generazione dei colori sullo schermo del computer, delle fotocamere digitali e dei telefonini.

Se si guarda molto da vicino lo schermo di una di queste apparecchiature (meglio se con una lente di ingrandimento) si vedrà una griglia molto fitta di quadratini che, come le tessere di un mosaico, formano la figura nella sua completezza vista dalla distanza minima di una decina di centimetri. Ciascuna di quelle tessere, dette *pixel* come forma contratta di *picture-element*, è in realtà formata da 3 elementi colorati dei cosiddetti colori fondamentali: il rosso, il verde ed il blu.

Nella terminologia corrente i tre colori vengono abbreviati con le loro iniziali maiuscole RGB dei nomi in inglese dei tre colori: Red, Green, Blue.

L'intensità di ciascuno dei tre colori può variare, secondo una tradizione tecnologica ormai consolidata, da un minimo di 0 ad un massimo di 255.

Così, ad esempio, con la combinazione di valori (R=255, G=0, B=0) si ottiene il rosso, con la combinazione di valori (R=0, G=255, B=0) si ottiene il verde e con la combinazione di valori (R=0, G=0, B=255) si ottiene il blu. E questo è di per sé evidente. Ma se si pone (R=255, G=255, B=0) si ottiene il giallo e con la combinazione (R=255, G=0, B=255) si ottiene il viola e con la combinazione (R=255, G=125, B=0) si ottiene l'arancione.

Una volta trovato un colore nuovo, con lievi variazioni dei singoli valori, si possono trovare tutte le sfumature e le gradazioni di colori volute. Utilizzando lo strumento **Colori** che potete attivare da un qualunque programma di scrittura o di disegno del vostro computer, potete verificare quanto qui esposto.

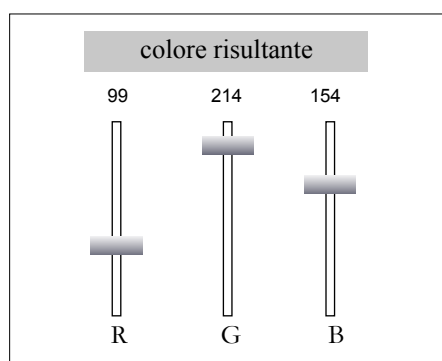


Figura 1.23 Colore RGB

Ciascuno dei colori fondamentali RGB può assumere 256 valori da 0 a 255 con gradazione del colore corrispondente che va dal nero al massimo di brillantezza del colore stesso. In questo modo si possono ottenere $256 \times 256 \times 256 = 16.777.216$ colori diversi (questo è il motivo per cui si dice che la gamma dei colori dello schermo del computer è a *milioni di colori*) che coprono con sufficiente realismo la gamma dei colori del visibile e cioè dei colori dell'arcobaleno.

Quindi, ogni colore o sfumatura di colore può essere rappresentata con una terna di valori numerici che consente di passare da una descrizione *qualitativa* ad una *quantitativa*. Per cui, ad esempio, il colore *tenue-violetto-che-tende-al-lilla* può essere espresso in forma sicuramente poco poetica ma decisamente molto precisa con: R222-G191-B244.

La terna dei valori RGB costituisce, in prima approssimazione, lo *spettro* di quel colore.

Un altro esempio si può fare considerando l'etichetta incollata sulle bottiglie di acqua minerale, dove è sempre presente un riquadro con la scritta analisi chimico-fisica e una lista di sali minerali come il sodio, il potassio, il calcio ecc. con le quantità percentuali presenti, compreso il *pH*, che è la misura dell'acidità.

Ebbene, ogni marca di acqua minerale presenta i propri valori e va da sé che al gusto a livello percettivo corrisponde la combinazione dei valori percentuali dei sali minerali e del *pH*. L'insieme dei valori delle percentuali dei sali minerali ed il *pH* costituiscono, ancora in prima approssimazione, lo *spettro* di quella particolare acqua minerale.

È la parola *analisi* su cui è importante porre l'attenzione, perché è questo l'argomento del prossimo capitolo: *analisi* è parola greca che significa *scomposizione* ed è davvero appropriata perché, come negli esempi del colore e dell'acqua minerale, è dall'operazione di *analisi* che risultano i valori numerici delle *componenti* che, nell'insieme, costituiscono lo *spettro*. Lo spettro è quindi la rappresentazione numerica (negli esempi visti) del colore percepito a livello sensoriale visivo o dell'acqua percepita a livello sensoriale del gusto.

Per il timbro di un suono, con le dovute differenze, succede la stessa cosa: la rappresentazione grafica proveniente dal marchingegno della Figura 1.1 si chiama (in prima approssimazione) *forma d'onda* e dall'analisi della forma d'onda risultano le sue componenti che sono di solito tantissime, ma fortunatamente tutte della stessa natura e cioè tutte di forma sinusoidale.

La forma d'onda del segnale e lo spettro altro non sono che due diversi modi di rappresentare lo stesso fenomeno: infatti, se il segnale rappresenta il suono come variazione di pressione al trascorrere del tempo, detto anche *dominio del tempo*, lo spettro rappresenta il suono come insieme dei valori delle componenti sinusoidali elementari nello spazio delle frequenze, detto anche *dominio delle frequenze*.

Questa dualità di rappresentazione del suono come segnale e spettro, trova corrispondenza nella fisiologia dell'orecchio. Attraverso meccanismi interpretativi del nostro cervello, la simultaneità delle singole vibrazioni ciliari della coclea vengono da noi riconosciute come timbro; e questo è dominio di indagine e di conoscenza della Neuro-Fisiologia.

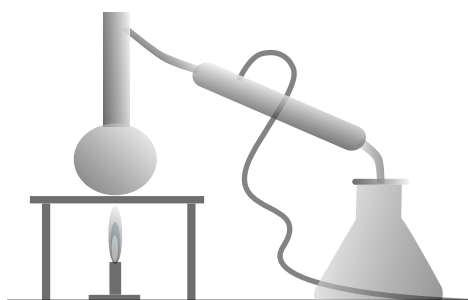
I domini di nostro interesse sono invece quello del *tempo* e quello delle *frequenze* che consentono di legare i concetti di frequenza, intensità e timbro per una più formale ed approfondita indagine analitica e definizione del suono.

L'analisi del suono è l'argomento del prossimo capitolo.

Analisi di segnali sonori



- 2.1 Pi greco e sinusoida
- 2.2 Segnale e spettro
- 2.3 Analisi di Fourier
 - 2.3.1 Calcolo delle ampiezze delle armoniche
- 2.4 Onda quadra
- 2.5 Spettri dinamici e involuppo



L'espressione *timbro-di-un-suono* è per la verità un po' ambigua ed in certo senso, tautologica: è come dire la *personalità-di-una-persona*! È chiaro che i due concetti si sovrappongono: una persona è, coincide con, il suo carattere, i suoi modi di fare, i gusti, le preferenze, le attitudini, il mestiere, ecc. E se qualcuna di queste componenti dovesse modificarsi per motivi, ad esempio, traumatici, di quella persona si direbbe: *...non è più lui/lei...*

Allo stesso modo, un particolare suono si identifica con il "suo" timbro e perciò nel seguito si userà la parola *timbro* e *suono* in maniera del tutto equivalente.

A proposito delle “dovute differenze” più volte menzionate a riguardo delle analogie tra timbro/colore e timbro/sapore-dell’acqua, c’è da dire che:

- 1) così come un colore può essere visto come “amalgama” (ossia sovrapposizione graduata dei tre colori fondamentali RGB) ed il sapore dell’acqua può essere considerato come “miscela” di un certo numero di sali minerali diversi ed in percentuali diverse, allo stesso modo un timbro può essere considerato come l’insieme di tantissime componenti elementari, tutte della stessa natura e tutte di forma sinusoidale;
- 2) mentre il *fattore tempo* è del tutto estraneo all’esistenza del colore e del sapore, cruciale è invece all’esistenza del suono: *il suono* come fenomeno fisico in sé è dato dalla variazione nel tempo della pressione dell’aria; anzi, *ogni suono* dal più semplice al più complesso (musicale, vocale, ecc.) è caratterizzato da una sua specifica ed individuale vita interna nella quale è possibile ravvisare diversi momenti e fasi in cui *quel suono* nasce, si evolve e si estingue.

Non è possibile schematizzare con pochi valori parametrici un particolare suono come si è visto invece con il colore “tenue-violetto-che-tende-al-lilla”; ma c’è in ogni caso l’esigenza di una metodologia che dia la possibilità di descrivere analiticamente un suono. Tale metodologia esiste e fa ricorso, come primo passo, alla sua rappresentazione grafico/matematica e cioè al *segnale e/o forma d’onda*. Poiché si è detto al punto 1) che il timbro e la corrispondente forma d’onda derivano dall’insieme delle componenti elementari sinusoidali, è necessario ora formulare con più precisione i concetti di *sinusoide* e di *frequenza* perché “strumenti matematici” fondamentali alla trattazione dell’Acustica.

2.1 Pi greco e sinusoide

Uno degli argomenti di interesse della geometria e della matematica è sempre stato quello relativo al rapporto esistente tra le misure degli elementi delle figure geometriche principali come ad esempio il rapporto tra il *lato* e la *diagonale* del *quadrato* ed il rapporto tra il *lato* e l’*altezza* del *triangolo equilatero*. Ma quello che fra tutte ha suscitato nel tempo più interesse, è senza dubbio il rapporto esistente tra le misure del *diametro* e della *circonferenza* del *cerchio*.

Come già si era visto nella Figura 1.4, se si fa rotolare un cerchio di diametro D , quando il cerchio ha compiuto un giro completo, la distanza percorsa è pari a 3 volte la misura del diametro più un “pezzetto” che, in maniera operativamente approssimativa, diciamo valere $0,14D$; per cui la

circonferenza del cerchio “misura” 3,14 volte il diametro del cerchio, come illustrato in Figura 2.1

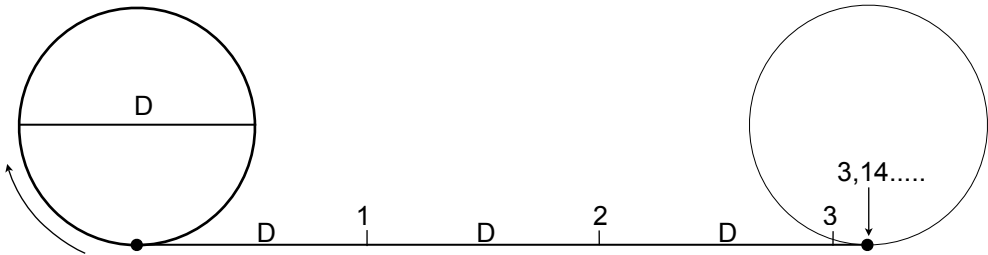


Figura 2.1 Misure del diametro e della circonferenza del cerchio

anche se, come si sa, il *vero rapporto* tra circonferenza e diametro è un valore con un numero infinito di cifre dopo la virgola: tale valore viene detto **pi greco** ed espresso con la lettera greca π che è la prima della parola *περίμετρος* (perimetro).

Intorno al calcolo della vera misura di quel “pezzetto”, e quindi di π , c’è praticamente tutta la storia della Matematica che ha visto impegnati studiosi di tutte le epoche e luoghi (greci, egizi, babilonesi, ellenisti, arabi, cinesi di due-tre millenni fa) e, più recentemente, di matematici come Newton, Eulero, ecc.

Oltre al fatto che il valore di π può essere ottenuto misurando fisicamente oggetti circolari con l’uso di una cordella graduata e flessibile come è ad esempio il metro da sarta (con il risultato di ottenere pochissime cifre decimali), sono stati proposti nel tempo numerosi metodi basati (vedi Wikipedia) su procedimenti di calcolo che danno la possibilità di trovare un numero di cifre decimali fino ad un limite prestabilito come per esempio:

$$\pi = 3,1415926535897932384626433832795028841971693993751058209749\dots$$

π è così affascinante [1] da aver stimolato una collettività di appassionati a creare poesie, racconti e romanzi dove le parole sono scelte, in sequenza, con un numero di lettere corrispondenti alle cifre di π per recitarle a memoria come fa il protagonista di “Vita di Pi”; i matematici, in base al modo anglosassone di scrivere le date del calendario, hanno dichiarato il 14 di marzo il π -day (3/14).

π è onnipresente nelle varie discipline della matematica (trigonometria, statistica, analisi, ecc.) e della fisica (cinematica, meccanica, ottica, elettromagnetismo, ecc.) e, naturalmente, nell’acustica.

Poiché $\pi = \text{circonferenza} / \text{diametro}$ ed il diametro D è 2 volte il raggio R , la lunghezza della circonferenza c è uguale a $D\pi = 2R\pi$ o anche $c = 2\pi R$.

Come mostrato in Figura 2.2, l'angolo α che il raggio forma con l'asse orizzontale x può essere misurato in *radiani* (rivedi Figura 1.4) come rapporto tra la misura dell'arco di circonferenza s e la misura del raggio R : $\alpha = s/R$.

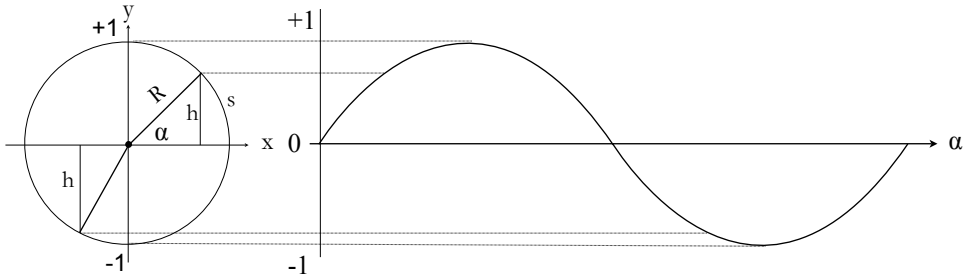


Figura 2.2 Seno dell'angolo in radianti e sinusoidi

La distanza h dall'asse orizzontale x di un punto nelle diverse posizioni sulla circonferenza di un cerchio di raggio R , traccia una *sinusoide*. Il rapporto tra la misura di h e quella del raggio R viene definito *seno dell'angolo α* e alla *funzione* matematica $y = \text{sen}(\alpha) = h/R$ corrisponde, graficamente, la *sinusoide* che perciò varia tra i valori -1 e $+1$.

La sinusoidi ha una "sorella gemella" chiamata **co-sinusoidi** che viene generata da un punto che si muove ad una distanza angolare di $\pi/2$ radianti rispetto al punto che genera la sinusoidi:

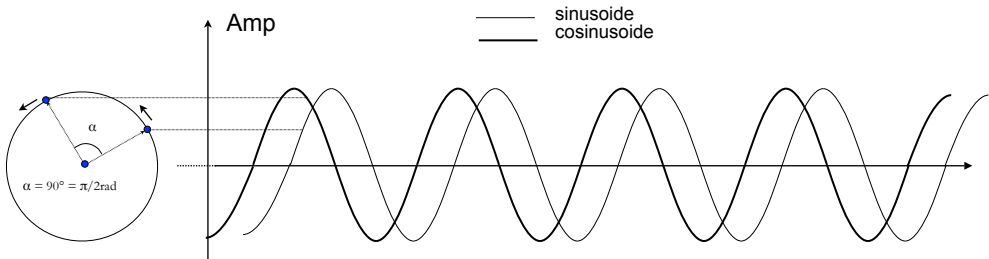


Figura 2.3 Sinusoidi e cosinusoidi

Poiché la sinusoidi e la cosinusoidi sono identiche nell'andamento e nella forma, ma una è indietro (o in avanti) rispetto all'altra, è possibile definire la *cosinusoidi in funzione della sinusoidi* nel seguente modo:

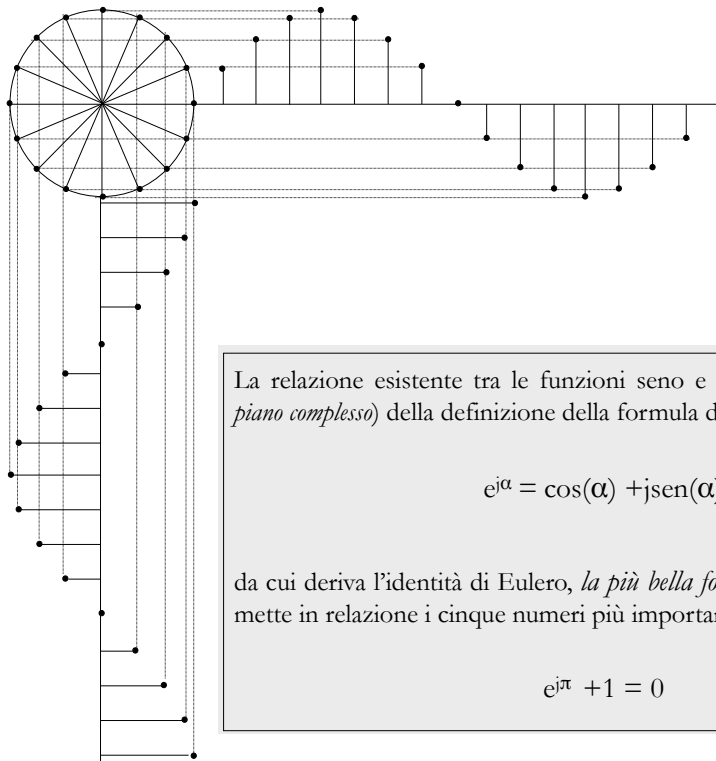
$$\cos(\alpha) = \text{sen}(\alpha + \pi/2)$$

In realtà le “sorelle” della sinusoida sono infinite perché se si mette un qualunque altro valore al posto di $\pi/2$, tradizionalmente indicato con la lettera greca φ (phi), si ottiene ancora una sinusoida, ma *traslata* di un angolo φ rispetto a quella originale:

$$y(\alpha) = \text{sen}(\alpha + \varphi)$$

La traslazione è detta **fase** e può assumere soltanto valori compresi tra $-\pi$ e $+\pi$. Questa osservazione relativa alla fase, tornerà particolarmente utile nel Capitolo 4 (in particolare in §4.6) dedicato ai *modelli di sintesi sonora*.

Per completezza formale si può infine osservare che, se la sinusoida è data dalla *distanza dall'asse orizzontale* del punto che si muove sulla circonferenza di raggio unitario, la cosinusoida è data dalla *distanza dall'asse verticale* di quel punto.



La relazione esistente tra le funzioni seno e coseno sta alla base (nel piano complesso) della definizione della formula di Eulero

$$e^{j\alpha} = \cos(\alpha) + j\text{sen}(\alpha)$$

da cui deriva l'identità di Eulero, *la più bella formula matematica* perché mette in relazione i cinque numeri più importanti:

$$e^{j\pi} + 1 = 0$$

Figura 2.4 Costruzione geometrica della sinusoida e della cosinusoida

Vediamo ora la stessa cosa in modo dinamico considerando un punto P sulla circonferenza che viene “trascinato” dal raggio del cerchio che si muove “come fa il raggio luminoso sullo schermo del radar”. Stabiliamo inoltre, convenzionalmente, una posizione iniziale 0 scelta “ad ore 3 dell’orologio” e che la rotazione del raggio avvenga in *sensu anti-orario*.

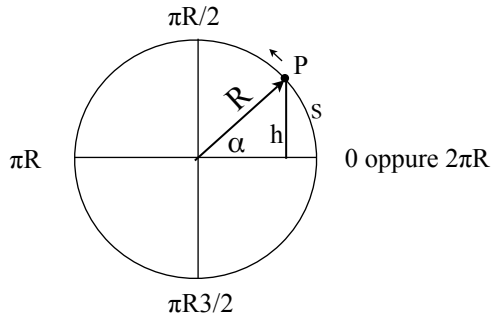


Figura 2.5 Movimento e posizioni di un punto sulla circonferenza

Quando raggiunge “ore 12”, il punto ha percorso una distanza di 1/4 di circonferenza $2\pi R$, che espressa in termini di raggio R, è pari a $2\pi R/4 = \pi R/2$; quando il punto raggiunge “ore 9” la distanza percorsa è pari alla metà della circonferenza, per cui $2\pi R/2 = \pi R$; quando raggiunge “ore 6” ha percorso 3/4 di circonferenza, e cioè $2\pi R \times 3/4 = \pi R 3/2$ e quando raggiunge il punto di partenza, ha percorso la distanza di $2\pi R$.

Facciamo questa osservazione: dire che un’automobile corre alla velocità di 100 chilometri all’ora, vuol dire che in 1 ora quell’automobile percorre una traiettoria di 100 chilometri di lunghezza. La *velocità* è data perciò dal rapporto tra lo *spazio* s percorso ed il *tempo* t impiegato per percorrerla secondo la relazione $v = s/t$ tant’è vero che, a parità di spazio percorso, se corre più velocemente, l’automobile impiega un tempo minore e se corre ad una velocità più bassa, impiega più tempo. Si può dire la stessa cosa in altri termini: lo spazio percorso è dato dalla velocità v del veicolo moltiplicato il tempo t trascorso, secondo la relazione $s = v \times t$ dall’ovvio significato.

Le due relazioni $v = s/t$ e $s = v \times t$ valgono anche quando la traiettoria percorsa dall’automobile è una pista circolare. Si tratta di un caso particolare in cui il veicolo percorre “linearmente” una traiettoria curva la cui lunghezza s è esprimibile in funzione del raggio del cerchio, per cui $s = \text{circonferenza} = 2\pi R$.

Inoltre, quando il veicolo percorre molte volte la pista circolare con velocità costante, è possibile analizzarne il movimento misurando il tempo T impiegato a percorrere un giro: il tempo T viene detto **periodo**.

Ad esempio: se il veicolo corre su una pista circolare di 80 metri di raggio con una velocità di 90 Km/h che corrisponde a 25 metri/secondo, si ha che

$s=c=2\pi \times 80 \approx 502$ metri, e poiché $t=s/v$ (in base alla $s=v \times t$)
 il periodo di percorrenza di un giro è $T=502/25 \approx 20$ secondi.

Perciò, nel caso particolare di un movimento circolare a velocità costante, poiché lo spazio s è dato dalla lunghezza della circonferenza $2\pi R$ ed il tempo t è dato dal periodo T impiegato a percorrere un giro intero, l'espressione $v=s/t$ diventa $v=2\pi R/T$.

È allora possibile conoscere in ogni momento la distanza percorsa sulla circonferenza *in funzione del tempo* trascorso, utilizzando la relazione $s=v \times t$ che con $v=2\pi R/T$, diventa

$$s = \frac{2\pi R}{T}t \text{ e siccome (Figura 2.2) } \alpha = \frac{s}{R} \text{ allora } \alpha = \frac{2\pi R}{RT}t \text{ e cioè } \alpha = \frac{2\pi}{T}t$$

e poiché $y=\text{sen}(\alpha)$ allora
$$y(t) = \text{sen}\left(\frac{2\pi}{T}t\right) \quad (*)$$

È importante sottolineare che mentre la $y=\text{sen}(\alpha)$ (in relazione a quanto illustrato nella Figura 2.2) si riferisce alla pura formulazione matematico/geometrica (statica), la (*) si riferisce ad un fenomeno fisico che si svolge nel tempo (dinamico), come illustrato nella Figura 2.6.

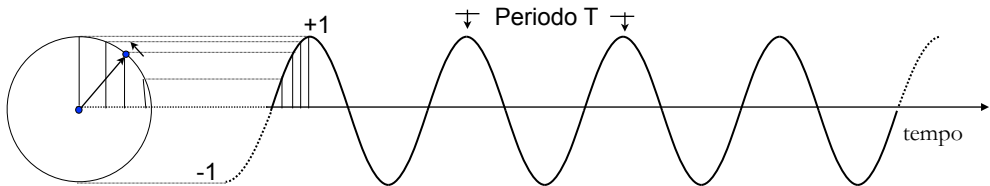


Figura 2.6 Periodo della sinusoide

Ricordando infine che $F=1/T$ allora

la $y(t) = \text{sen}\left(\frac{2\pi}{T}t\right)$ può essere scritta come
$$y(t) = \text{sen}(2\pi Ft)$$

Per esempio, se il periodo di accadimento di un evento è 0,5 secondi, allora la frequenza dell'evento è di 2 volte-al-secondo: $T=0,5 \quad F=1/T=1/0,5=2\text{Hz}$; se $F=50 \text{ Hz}$ allora $T=1/F=1/50=0,02$ secondi e cioè 2 centesimi di secondo.

Il grafico di Figura 2.7 riporta una porzione di tempo della durata di un centesimo di secondo (0,01 sec.) ed un'oscillazione sinusoidale a frequenza $F=400$ Hz, tenendo presente che 400 oscillazioni al secondo corrispondono a 4 oscillazioni ogni centesimo di secondo, e che il periodo di ciascuna oscillazione è $T=1/400=0,0025$ secondi =2,5 millisecondi.

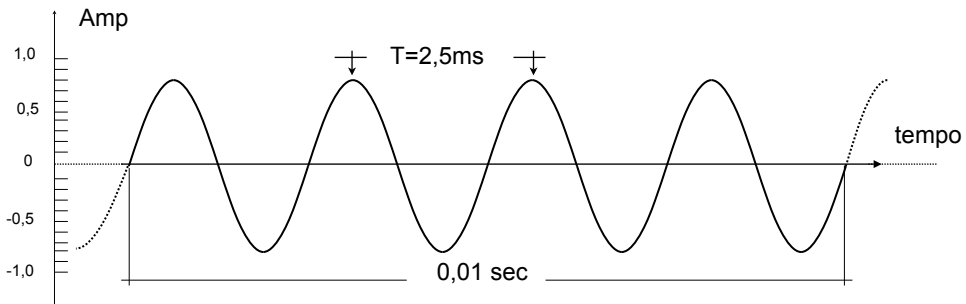


Figura 2.7 Sinusoide a 400 Hz

Prima di proseguire è importante sottolineare il fatto che non esistono meccanismi che oscillano in maniera puramente sinusoidale: una corda di chitarra, di violino o di pianoforte, oscilla con una frequenza precisa che ne determina percettivamente l'altezza (*pitch*), ma la sua variazione di posizione rispetto ad una posizione centrale è molto complessa e ben lontana dall'essere descritta da una semplice forma sinusoidale.

La **forma d'onda** che risulta graficando nel tempo tale posizione come in Figura 1.1, viene detto **segnale** e dall'analisi del segnale si può ottenere una rappresentazione in termini di componenti sinusoidali elementari.

Per semplificare la trattazione dell'analisi di un suono, ignoreremo il concetto di fase tenendo comunque presente che è vero che una sinusoide è completamente descritta dai 3 parametri *frequenza*, *ampiezza* e *fase*, ma è vero anche che la fase, per il nostro apparato uditivo non è significativa dal punto di vista timbrico.

Lo è invece, come vedremo nel Paragrafo 5.5, per il riconoscimento della direzione del suono.

La fase è comunque necessaria se si vuole ricostruire l'esatta forma d'onda: in relazione alla Figura 2.12 del prossimo paragrafo, la forma d'onda che risulta dalla somma punto per punto delle 3 sinusoidi, sarebbe diversa se anche una sola delle sinusoidi fosse traslata in avanti o indietro, e cioè se avesse un valore di fase diverso. Ma lo spettro delle ampiezze, e quindi l'effetto psicoacustico, e cioè il timbro del suono corrispondente, non subirebbe alcuna variazione.

2.2 Segnale e spettro

Relativamente all'acustica, dunque, una sinusoida è individuata dai due parametri **periodo** e **ampiezza**, analogamente a come un rettangolo è individuato dalle due dimensioni *base* e *altezza*: la base corrisponde al periodo e l'altezza corrisponde al doppio dell'ampiezza come mostrato in Figura 2.8:

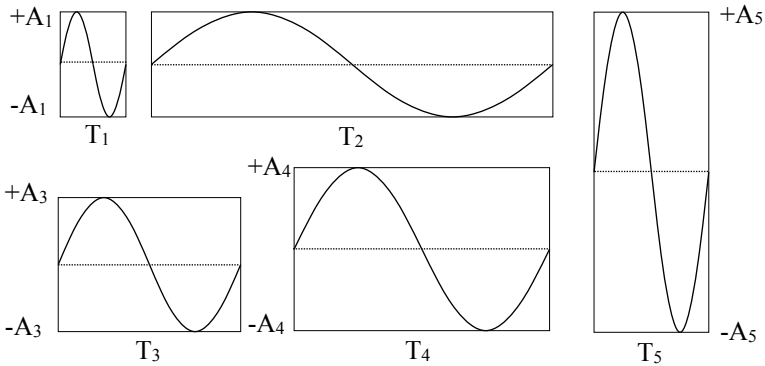


Figura 2.8 Sinusoidi varie

Il grafico di Figura 2.9 riporta, a mo' di esempio, 3 sinusoidi di ampiezza e periodo (e quindi, frequenza) diverse tra loro: $A_1=1$ e le altre considerate scalate proporzionalmente ad A_1

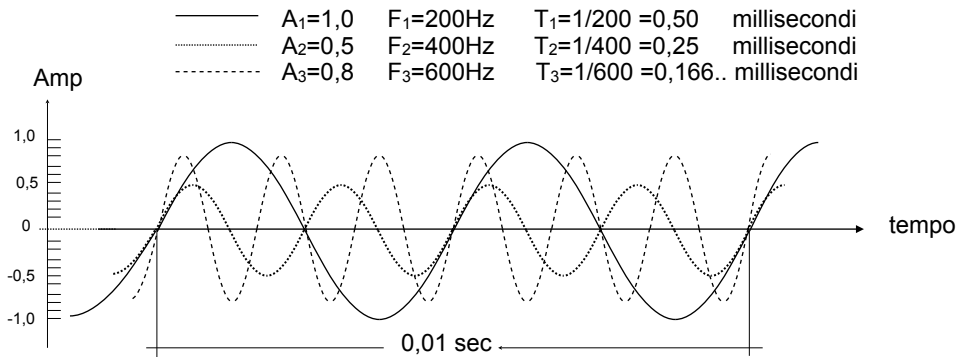


Figura 2.9 Sinusoidi contemporanee sovrapposte

Fino a che il numero di sinusoidi è basso, si riesce ad individuarle distintamente sul grafico; ma quando il numero sale fino a 10, 20 e oltre, il grafico diventa illeggibile e di nessuna pratica utilità.

Ma dal momento che una sinusoidale è descritta dai soli valori di ampiezza e frequenza, riportando il valore della *frequenza* sull'asse orizzontale e l'*ampiezza* sull'asse verticale, è possibile rappresentare con chiarezza e senza perdere informazione, anche un alto numero di sinusoidi.

Questa diversa rappresentazione è detta **spettro delle ampiezze** nel dominio delle frequenze.

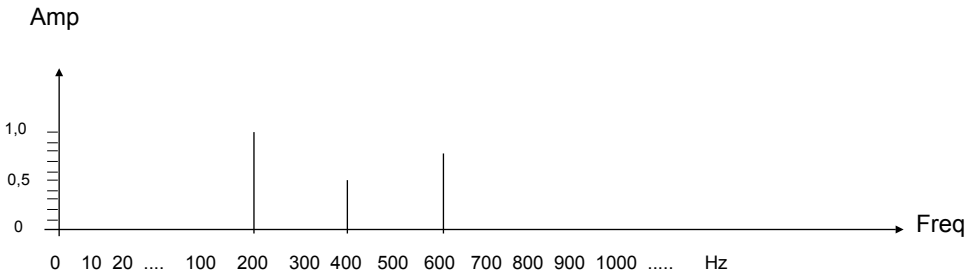


Figura 2.10 Spettro delle ampiezze

In pratica lo spettro delle ampiezze riporta una linea verticale di altezza proporzionale all'*ampiezza* di quella oscillazione sinusoidale se sull'asse del tempo c'è una sinusoidale a quella frequenza.

Tuttavia, ciò che il timpano dell'orecchio rileva non sono le tre sorgenti sonore sinusoidali distinte, ma l'effetto della loro sovrapposizione. Infatti il timpano è una membrana come la pelle di un piccolo tamburo che subisce una deformazione provocata dalla variazione micrometrica e velocissima della pressione dell'aria generata da sorgenti sonore circostanti.

Si tratta di una deformazione concavo/convessa tale per cui in ogni istante il timpano si trova in una ed una sola posizione "in fuori" o "in dentro" con piccoli spostamenti di millesimi di millimetro. La pressione totale esercitata sul timpano è quella data dalla sovrapposizione dei contributi di ogni singola sorgente sonora della *realtà acustica* che ci circonda e quindi la deformazione concavo/convessa è data, istante per istante, dalla sovrapposizione della pressione/depressione dell'aria che le singole sorgenti creano.

Il marchingegno di Figura 1.1 è quindi un'allegoria della meccanica dell'orecchio: la membrana corrisponde al timpano e il sistema di leve corrisponde agli *ossicini staffa-incudine-martello* che trasmettono le vibrazioni del timpano all'orecchio interno. Il movimento del timpano provocato dalla variazione di pressione dell'aria, è quello tracciato dalla penna sul nastro scorrevole, vale a dire che in ogni istante il timpano si trova in una ed una sola ben precisa posizione *in-fuori* o *in-dentro*.

La Figura 2.11 mostra l'andamento della pressione dell'aria che investe il timpano relativo ad un brano di musica della durata di 3 millisecondi dove stanno suonando tutti insieme un pianoforte, una batteria, una un contrabbasso, un sassofono ed una voce.

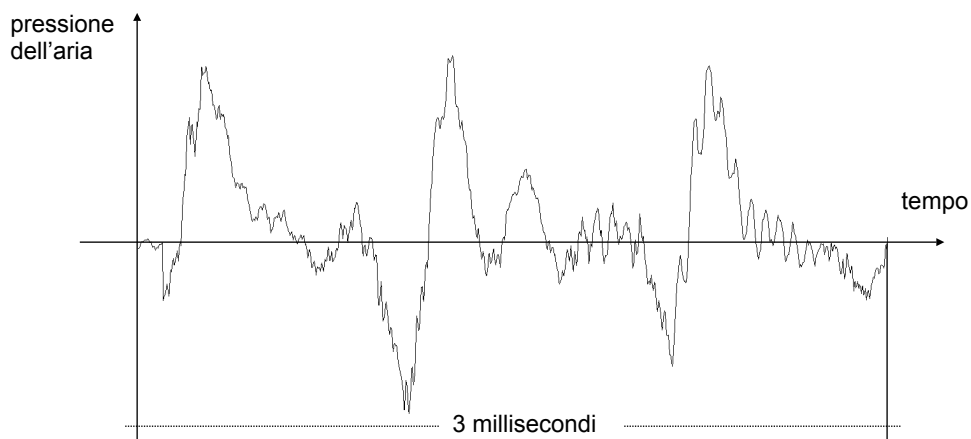


Figura 2.11 Segnale musicale

A livello percettivo noi siamo in grado di riconoscere distintamente le diverse sorgenti sonore come strumenti musicali, voci, rumori, ecc., ma a livello puramente acustico, il timpano rileva una sola variazione di pressione che è “l’amalgama”, la “sovrapposizione”, la *sommatoria* di tutte le variazioni di pressione da loro provocate.

La variazione della pressione dell'aria viene rappresentata con ciò che più volte abbiamo definito *segnale* (per l'esattezza, *segnale monodimensionale*) che sul disco di vinile (Figura 3.1) viene inciso come solco “congelato” in modo puramente meccanico in *analogia* con la variazione della pressione dell'aria provocata dai musicisti e “catturata” dal microfono.

Il segnale mono-dimensionale contiene l'informazione acustica relativa a tutti gli strumenti, alla voce, ai rumori, ecc. così come è nella realtà.

Vediamo allora cosa succede nel caso semplice di 3 sinusoidi a frequenza ed ampiezza diverse tra loro e contemporanee, supponendo che a ciascuna sinusoide corrisponda una distinta ed ipotetica “sorgente sonora”: si tratta di fare la somma algebrica punto per punto dei valori delle tre sinusoidi come illustrato nella Figura 2.12.

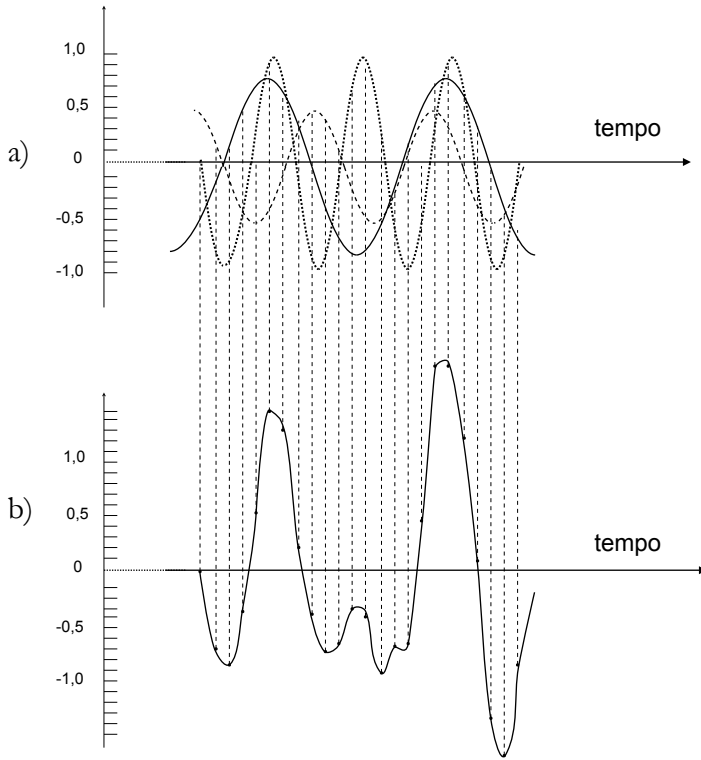


Figura 2.12 Somma di sinusoidi

La situazione che si ha di solito è perciò quella relativa ad un *segnale* nel *dominio del tempo* e il corrispondente *spettro delle ampiezze* nel *dominio delle frequenze*, anche se è bene osservare che in questo caso lo spettro è conosciuto in anticipo in quanto il segnale è stato “costruito” dalla sovrapposizione di tre segnali sinusoidali di cui si conosce frequenza ed ampiezza.

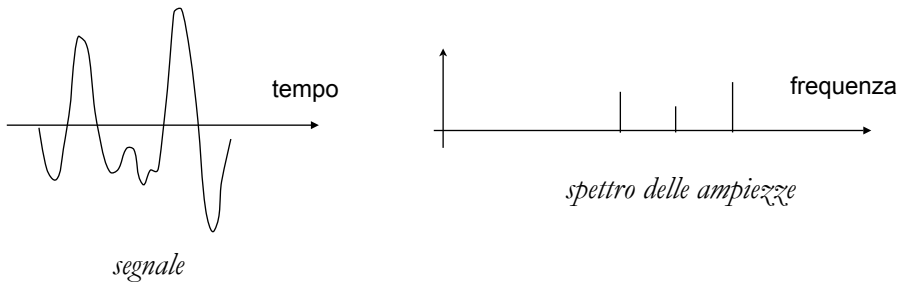


Figura 2.13 Segnale e spettro delle ampiezze

Nella realtà, invece, il *segnale* è la rappresentazione di ciò che accade veramente alla pressione dell'aria, come nel caso di Figura 2.11; perciò lo *spettro delle ampiezze* che risulta da un'operazione di analisi condotta sul segnale stesso, che è un'astrazione, una pura invenzione matematica.

Come nella *Statistica*: quando si dice che *il 20% degli alunni di una scuola ha gli occhi azzurri*, si è detto una cosa del tutto astratta che risulta dall'analisi di fatti reali oggettivi; perché la cosa vera è che Giuseppina, Pierpaola, Mariateresa, Giorgio... hanno gli occhi azzurri ed il giorno in cui Giuseppina, Giorgio, Mariateresa.... usciranno da quella scuola, loro continueranno ad avere gli occhi azzurri, ma la media statistica...

Analizzare le cose, in generale, vuol dire *separare* gli uni dagli altri gli elementi che compongono le cose. E separare vuole dire *mettere ordine*. E mettere ordine costa molto di più che fare disordine mischiando le cose. E non è detto che l'energia di per sé sia sufficiente per fare ordine: occorre disporre o mettere a punto una strategia organizzativa e poi metterla in opera.

Per esempio: quando al bar chiedete un cappuccino, il barista mischia il caffè con il latte e così facendo crea *disordine*. L'eventuale richiesta "della procedura inversa" e cioè di separare il latte dal caffè a partire dal cappuccino, è... abbastanza poco probabile che venga soddisfatta anche ammesso di avere a disposizione tanto tempo, tanto denaro e tanti baristi disposti a farlo.

In Fisica, per l'esattezza nella Termodinamica, la misura dell'ordine e del disordine si chiama *entropia* ed è stato proprio lo studio della propagazione del calore nei corpi solidi che ha portato alla metodologia di analisi sonora nota come *Analisi di Fourier* e poi *Trasformata di Fourier*.

Riassumendo:

- 1) *analizzare* significa separare le componenti che costituiscono un fenomeno o un amalgama complesso;
- 2) per fare l'analisi di un fenomeno o un amalgama complesso occorre una valida metodologia, seguendo la quale si è in grado di estrarre il valore delle componenti;
- 3) si trova ciò che si cerca! il risultato dell'analisi è la misura del "trovato ciò che si è cercato" e se la misura=0, quella cosa che si cercava... non c'è.

Un suono, anche un singolo suono come quello di un campanello, di un toc-toc sulla porta, di un flauto ecc., è il risultato dell'accumulo di tanti elementi che vibrano in ragione della fisica che ci sta dietro, ciascuno indipendentemente dagli altri. Siamo noi che, percettivamente, di fronte al fenomeno, organizziamo, classifichiamo e riconosciamo tutto ciò come *suono*.

2.3 Analisi di Fourier

Jean Baptiste Joseph Fourier visse in epoca napoleonica (1765-1830) ed ebbe stretti rapporti con Napoleone stesso. J.B.J. Fourier (che non deve essere confuso con il filosofo suo contemporaneo Charles Fourier) fu fisico matematico e politico [2][3].

Durante la campagna d'Egitto di cui fecero parte moltissimi scienziati francesi dell'epoca, oltre ad essere stato nominato Governatore del Basso Egitto (1800-1812) e ad aver partecipato alla spedizione archeologica nella quale fu trovata la Stele di Rosetta che nelle mani dell'amico Champollion avrebbe rivelato il segreto dei geroglifici, Fourier si occupò del fenomeno fisico della *trasmissione del calore*.

Fourier, che a vent'anni si lamentava di non essere ancora entrato nel firmamento degli immortali che avevano dato il loro contributo alla Scienza, con il suo lavoro sulla termodinamica, lui ignaro, sarebbe diventato un onnipresente e luminoso astro nel settore dell'acustica e dell'elettronica.

Nel suo libro *Teoria Analitica del Calore* del 1822, egli introdusse il concetto di *serie geometrica* come strumento, per l'appunto, di analisi del fenomeno termodinamico ed espone come *teorema*.

Il *Teorema di Fourier*, scolpito come epigrafe sulla sua tomba nel Cimitero di Montmartre a Parigi (dove sono anche le tombe di Ampère, Foucault, Adolph Sax, Truffaut, Berlioz...), nella sostanza dice che:

*un segnale periodico può essere scomposto nella somma di infiniti termini
di funzioni trigonometriche con frequenze multiple della fondamentale.*

È bene ribadire il fatto che Fourier fece questa considerazione senza pensare al suono, perché l'idea del suono come variazione di pressione dell'aria e della sua rappresentazione come segnale (Figura 1.1) è nata molto tempo dopo la sua morte con le esperienze di Scott de Martinville, Charles Cros e di Edison (§1.1).

Ed è stato negli anni '20-'30 del 1900, che il Teorema di Fourier con l'elettronica e l'analisi del suono si è rivelato di fondamentale importanza.

Il teorema di Fourier si esprime matematicamente con la *Serie di Fourier* cioè come *somma di infinite sinusoidi di frequenza doppia, tripla, quadrupla, ecc., ciascuna di ampiezza e fase opportune per ricostruire correttamente il segnale di partenza*.

Ma prima di proseguire è bene chiarire cosa si intende con il termine *periodico* che compare nella formulazione del teorema.

Un suono è provocato, ormai si sa, da un'oscillazione che si ripete molte volte al secondo; il che significa che la durata di una sola oscillazione è molto breve. È chiaro che per avere la sensazione del suono è necessario che all'orecchio arrivino molte oscillazioni per una durata per lo meno di qualche decimo di secondo.

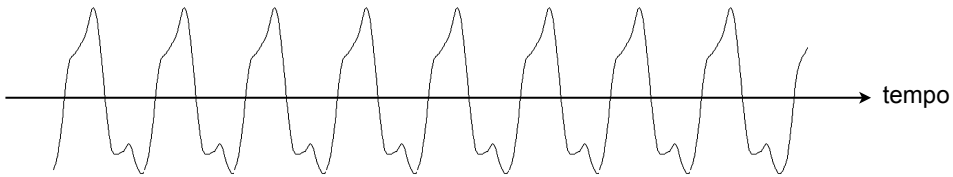


Figura 2.14 Ripetizioni di oscillazioni *identiche*

Una sola oscillazione che ha un periodo della durata media di 1 centesimo o di 1 millesimo di secondo, viene avvertito come un semplice “*fs*” o un “*tkc*” qualunque sia la forma dell'onda. Ma contando sul fatto che tutte le ripetizioni dell'onda che provocano la sensazione del suono sono uguali, l'analisi si fa su una sola ripetizione, su un solo periodo. Tuttavia, nella realtà non esistono suoni in cui le ripetizioni sono l'una uguale alla precedente e/o alla successiva: la Figura 2.14 mostra una situazione del tutto fittizia e teorica. La Figura 2.15 si riferisce invece al suono vero della nota LA₄=440Hz del pianoforte qualche centesimo di secondo subito dopo l'attacco e quindi in una zona sufficientemente stabile dal punto di vista timbrico e frequenziale.

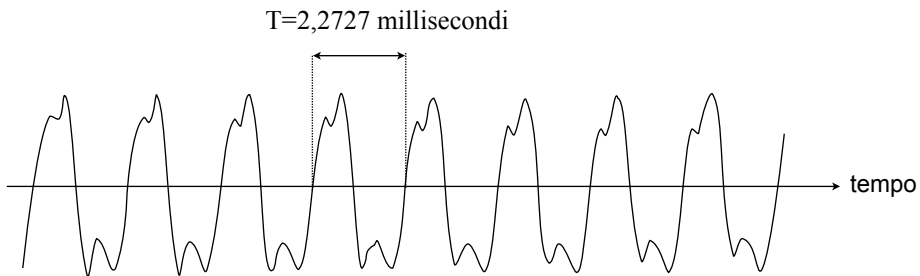


Figura 2.15 Ripetizioni di oscillazioni *“somiglianti”*

Si può osservare che, nonostante il segnale si riferisca ad una “zona di stabilità” del suono, le ripetizioni della forma d'onda della durata di $1/440=2,2727$ millisecondi, presentano piccole differenze. Questo fatto ha delle conseguenze sullo spettro e porta a dire che l'analisi del suono è completa quando viene fatta su ciascuna delle forme d'onda della successione con il risultato di avere *spettri dinamici* (§2.5, Figura 2.36).

2.3.1 Calcolo delle ampiezze delle armoniche

Il Teorema e le formule delle pagine precedenti dicono che è possibile costruire una forma d'onda complessa a partire dai valori delle ampiezze delle armoniche $A_1, A_2, A_3, \dots, A_n$ e, viceversa, che è possibile trovare i valori delle ampiezze delle armoniche $A_1, A_2, A_3, \dots, A_n$ che compongono una forma d'onda: in questo caso, quante più se ne cerca e se ne trova, tanto migliore è il risultato di una eventuale ricostruzione del segnale mediante serie di Fourier.

L'analisi di Fourier di un segnale somiglia un po' all'analisi chimica di un composto dove per conoscere la quantità di un elemento X (se presente) si immette nel composto una quantità nota N di un elemento che ha la proprietà di reagire con l'elemento X e poi si misura l'effetto della reazione di X con N.

Nell'analisi di Fourier si sa in partenza che il segnale è "composto" di sinusoidi a frequenza diversa: si tratta perciò di misurare l'ampiezza di molte componenti sinusoidali scelte alle frequenze multiple della fondamentale. Qui l'elemento noto e l'elemento cercato sono la sinusoide stessa presa alla frequenza voluta, e la "reazione" con l'intero "composto" costituito dal segnale è un'operazione matematica. L'operazione consiste (vedi Figure 2.16-2.17) nel *combinare* il segnale (e cioè la forma d'onda) con una sinusoide a frequenza nota e nel calcolare l'*area della superficie* che risulta da tale *combinazione*: il risultato del calcolo *rappresenta il contributo* della componente sinusoidale a quella frequenza.

"Combinare" segnale e sinusoide vuole dire, come verrà illustrato operativamente nella Figura 2.17, *moltiplicare punto a punto le due curve*. Questa procedura viene replicata con le componenti sinusoidali multiple (e cioè con le armoniche) del periodo di segnale analizzato. Vediamo allora nel dettaglio la procedura da seguire per calcolare il valore della fondamentale.

Predisponiamo innanzitutto una griglia di riferimento (Figura 2.16 [i]) sulla quale riportare un frammento di segnale corrispondente ad un periodo T di una *forma d'onda* come potrebbe essere, ad esempio, quella in evidenza nella Figura 2.15 relativa al segnale della nota La440 del pianoforte: la griglia riporta *in orizzontale* la durata del periodo T con 8 sotto-divisioni utilizzate per trovare (mediante moltiplicazione punto a punto delle due curve) la curva sotto la quale calcolare l'area; e *in verticale* una scala di riferimento del tutto indicativa tra -1 e +1 con suddivisione in passi di 0,25 di ausilio alla procedura di calcolo.

Sulla griglia viene posta una forma d'onda sinusoidale di eguale periodo ad ampiezza massima di riferimento uguale ad 1 (Figura 2.16 [ii]); il risultato della sovrapposizione (Figura 2.16 [iii]) viene utilizzata per generare la *curva prodotto* delle due forme d'onda (Figura 2.17) per eseguire, finalmente, il calcolo dell'area della superficie.

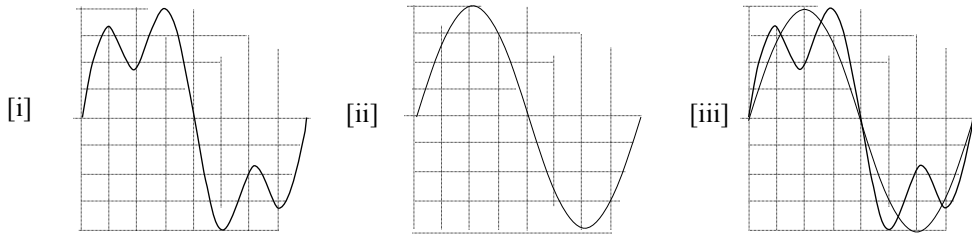


Figura 2.16 Sovrapposizione di sinusoide a forma d'onda

La curva risultante dal prodotto delle due curve, viene trovata per punti:

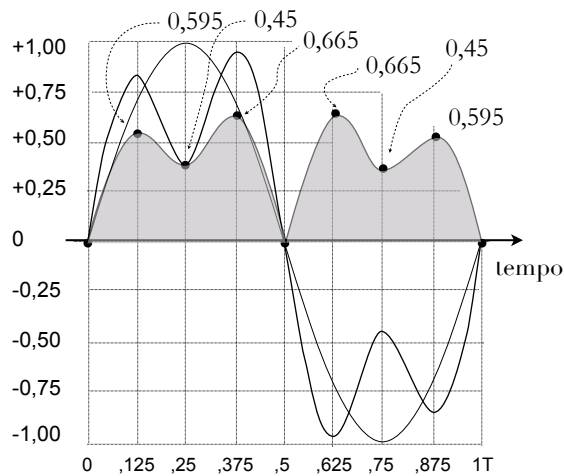


Figura 2.17 Curva risultante dal prodotto della forma d'onda con la sinusoide

- Il punto *zero* a 0 vale $0=0 \times 0$;
- il punto *uno* a $0,125 T$ vale $+0,595 = +0,70 \text{ (su sinusoide)} \times 0,85 \text{ (su forma d'onda)}$
- il punto *due* a $0,25 T$ vale $+0,45 = 1 \times 0,45$
- il punto *tre* a $0,375 T$ vale $+0,665 = +0,70 \times 0,95$
- il punto *quattro* a $0,5 T$ vale $0 = 0 \times 0$
- il punto *cinque* a $0,625 T$ vale $+0,665 = -0,70 \times -0,95$
- il punto *sei* a $0,750 T$ vale $+0,45 = -1 \times -0,45$
- il punto *sette* a $0,875 T$ vale $+0,595 = -0,70 \times -0,85$
- l'ultimo punto a $1T$ vale $0=0 \times 0$

L'area della *superficie in grigio sotto la curva risultante*, rappresenta il contributo della sinusoide a frequenza fondamentale.

Ma come si fa a calcolare l'area di una superficie delimitata da una linea curva qualsiasi?

La cosa che sappiamo fare per certo è calcolare l'area di un rettangolo come *base x altezza*, per cui l'unico modo per ottenere l'area di una superficie curva è quello di sommare l'area di tanti rettangoli sovrapposti ad essa (Figura 2.18) ottenendo, però, un risultato approssimativo: per ogni rettangolo la *base* (uguale per tutti) è quella stabilita nella quadrettatura del grafico e corrisponde al passo di divisione orizzontale, tradizionalmente indicato con Δt (*delta t*), che nel nostro caso vale $0,125T$; l'*altezza* di ciascun rettangolo è data, invece, dal valore trovato, per ogni punto, come prodotto del valore della forma d'onda con quello della senoide in quel punto.

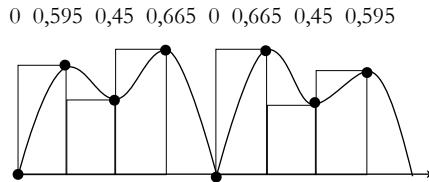


Figura 2.18 Calcolo dell'area sotto la curva (prima armonica o fondamentale)

Quindi, con $\Delta t=0,125$ l'area totale è data da

$$\begin{aligned}
 & 0 \times 0,125 + 0,595 \times 0,125 + 0,45 \times 0,125 + 0,665 \times 0,125 + 0 \times 0,125 + 0,665 \times 0,125 + 0,45 \times 0,125 + 0,595 \times 0,125 = \\
 & 0 + 0,074375 + 0,05625 + 0,08312 + 0 + 0,08312 + 0,05625 + 0,074375 = \\
 & = 0,4275 \text{ che rappresenta il contributo dell'ampiezza della prima armonica.}
 \end{aligned}$$

Il valore trovato è decisamente approssimato, ma è possibile ottenere un risultato più preciso utilizzando una suddivisione orizzontale (asse del tempo) più raffinata, come ad esempio con 16 suddivisioni e cioè con un $\Delta t=0,0625T$.

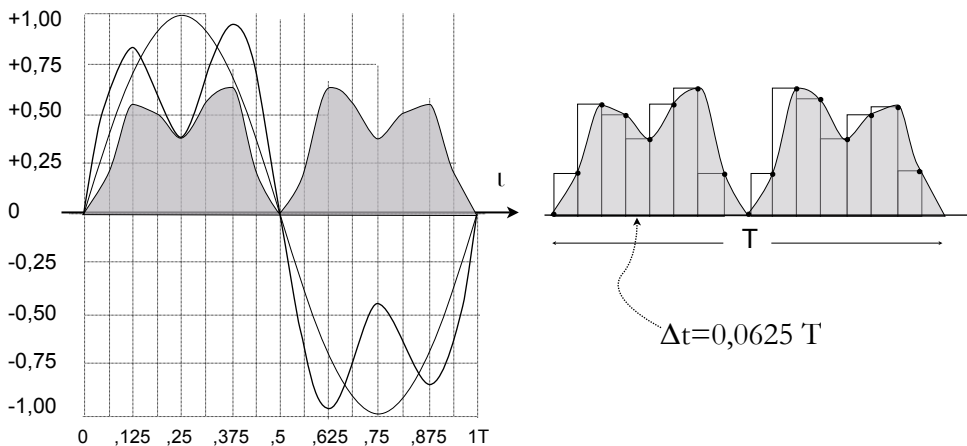


Figura 2.19 Calcolo più preciso dell'area sotto la curva (prima armonica)

Quindi, maggiore è il numero di suddivisioni, migliore è il risultato ottenuto.

Per trovare il valore della *seconda armonica* si sovrappone alla forma d'onda una senoide di periodo $T/2$ e cioè a frequenza doppia del segnale, nello stesso spazio temporale del periodo T come illustrato in Figura 2.20.

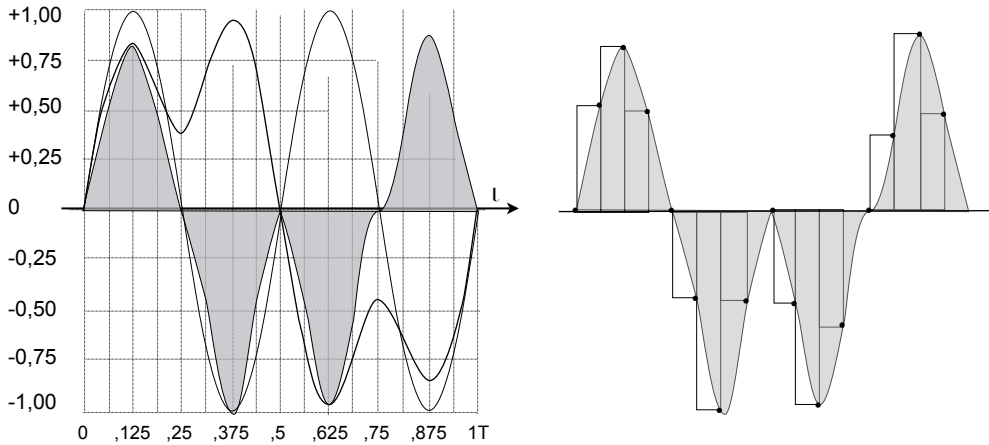


Figura 2.20 Calcolo della seconda armonica

Come prima, si trova la curva per punti e poi si calcola l'area della superficie tra curva e asse delle del tempo. In questo caso la curva trovata individua parti nella zona positiva e altre parti nella zona negativa: questo significa che il risultato è dato dalla *somma algebrica* e cioè dalla somma e differenza delle aree nelle parte positiva e nella parte negativa.

Per trovare il valore della *terza armonica* si sovrappone alla forma d'onda una senoide di periodo $T/3$ (e cioè a frequenza tripla) e poi si calcola l'area della superficie tra la curva trovata e l'asse del tempo.

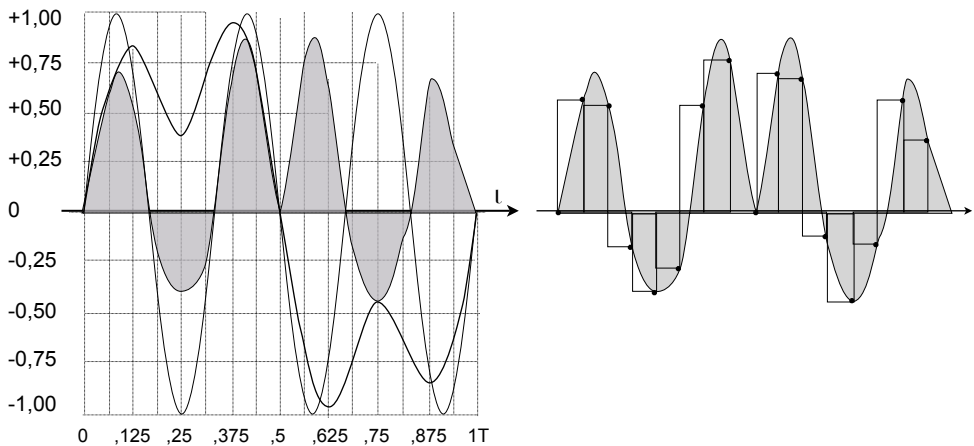


Figura 2.21 Calcolo della terza armonica

In riferimento al calcolo della prima armonica (Figura 2.17), se i valori dei punti trovati dal prodotto dei valori della forma d'onda con quelli della sinusoidale, vengono messi in una lista, possiamo indicare tutti quei valori con V_i intendendo con i l'*indice* che individua la sua posizione nella lista e che può assumere un valore che va da 0 a 7:

V_i	0	0,595	0,45	0,665	0	0,665	0,45	0,595
i	0	1	2	3	4	5	6	7

così facendo, il calcolo dell'area può essere espresso in forma compatta come

$$area = \sum_{i=0}^{i=7} V_i \Delta t$$

e si legge: *sommatoria per i che va da 0 a 7 di V_i moltiplicato Δt .*

Ogni valore V_i è il risultato del prodotto del valore della *forma d'onda* al punto i con il valore della *sinusoidale* sempre al punto i ; allora, se si indica con $s(i)$ il valore della forma d'onda originaria all'istante i -esimo e con $sen(x_i)$ il valore della sinusoidale sempre all'istante i -esimo, tenendo presente che

$$x_i = \frac{2\pi}{T} i \Delta t$$

si ha che $V_i = s(i) \times sen(x_i)$ per cui la formula precedente diventa

$$area = \sum_{i=0}^{i=7} s(i) sen(x_i) \Delta t$$

Quando si è eseguito il calcolo dell'area (Figura 2.18) si è detto che essa *rappresenta il contributo* e non *il valore* della prima armonica; in effetti il valore trovato è dato dall'area equivalente di un rettangolo che ha per *base il periodo T* e per *altezza il valore dell'armonica*. Per cui l'effettivo valore dell'armonica è $A_1 = area/T$ e cioè

$$A_1 = \frac{1}{T} \sum_{i=0}^{i=7} s(i) sen(x_i) \Delta t$$

Quindi, per trovare il valore di una generica *armonica n* , la formula che esprime la procedura esposta è la seguente:

$$A_n = \frac{1}{T} \sum_{i=0}^{i=7} s(i) sen(nx_i) \Delta t$$

Quanto abbiamo visto a commento della Figura 2.18, e cioè l'inizio di un processo che porta ad una precisione di calcolo sempre maggiore, nella matematica del *calcolo infinitesimale* viene espresso con: “...un numero di divisioni che tende all'infinito e Δt che tende a 0...” oppure più semplicemente, “*passando al limite...*” per ottenere il vero valore. Anche la simbologia matematica cambia, ed il simbolo di sommatoria \sum “spigoloso” come i rettangoli, viene sostituito dal simbolo \int chiamato *integrale* più sinuoso e continuo proprio a significare un calcolo dell'area fatto nella sua completezza (e cioè integrale!)

$$A_n = \frac{1}{T} \int_0^T s(t) \text{sen}(2\pi nFt) dt$$

È questo il tipo di formula che si trova (in maniera più complessa e completa) nei libri di matematica e di fisica dove viene trattata la Trasformata di Fourier.

Il meccanismo che *tende sempre più al raffinato* e quindi ad una maggior precisione del risultato, sta alla base del *calcolo infinitesimale* proposto verso il 1680 da Leibniz e Newton (c'è tutta una storia ed una diatriba in proposito) ed è parte fondamentale della matematica messa a punto da Eulero (1707-1783), Gauss (1777-1855) e altri grandi matematici.

Ci sono tre osservazioni conclusive che è importante fare.

- 1) La formula appena vista è utilizzabile soltanto quando la forma del segnale $s(t)$ è nota in termini di funzione matematica come ad esempio l'*onda quadra*, l'*onda triangolare* e l'*onda a dente di sega*. Nel caso di segnali reali, e cioè a livello operativo, il modo di ottenere il valore dell'ampiezza di un'armonica e di conseguenza lo spettro del segnale, è quello appena visto.
- 2) Concettualmente corretta, la formula e la procedura seguita nell'esempio, non è tuttavia completa per operare su segnali che non presentano particolari simmetrie come avviene nella realtà: il Teorema menziona infatti le *funzioni trigonometriche* intendendo con questo *seni e coseni* oppure *seni (o coseni) con fase* da cui risulta esistere anche lo *spettro delle fasi* oltre quello delle ampiezze, come si vedrà nel Paragrafo 5.5 e seguenti.
- 3) Il procedimento seguito per il calcolo integrale è lo stesso che sta alla base di quello di conversione dall'analogico al digitale, argomento del Capitolo 3.

2.5 Spettri dinamici e inviluppo

L'applicazione del Teorema di Fourier ad un segnale periodico relativo ad un suono come si è fatto fin qui, non esaurisce l'indagine sul timbro: infatti il teorema si riferisce ad un segnale che *si presuppone* si ripeta periodicamente sempre uguale a se stesso come nella Figura 2.15. Ma nella realtà la situazione è invece più simile a quella della Figura 2.16 dove le oscillazioni differiscono nella forma le une dalle altre, anche se tutte sono della stessa durata (e cioè dello stesso periodo) che ne determina l'altezza frequenziale a livello percettivo.

Si deve notare inoltre che la situazione di Figura 2.16 è abbastanza particolare perché riferita ad un solo suono di un solo strumento e in un preciso momento: il La440 di un pianoforte dopo l'attacco. Tutta la procedura di analisi che abbiamo condotto, conta sul fatto che già si sa su cosa si sta operando, mentre in generale invece, come si vedrà nel Paragrafo 4.9.3, l'analisi viene condotta su una porzione di segnale di lunghezza predefinita che si riferisce ad una situazione musicale complessa come quella dell'esempio di Figura 2.11.

Supponiamo comunque di avere eseguito il calcolo per trovare i valori delle prime 6 armoniche di un periodo estratto dal segnale di Figura 2.16 con la procedura esposta nel Paragrafo 2.3.1 e di avere ottenuto questo risultato:

$$A_1=0,74; \quad A_2=0,23; \quad A_3=0,33; \quad A_4=0,55; \quad A_5=0,23; \quad A_6=0,55$$

e di rappresentarlo in forma di spettro come mostrato nella Figura 2.35.

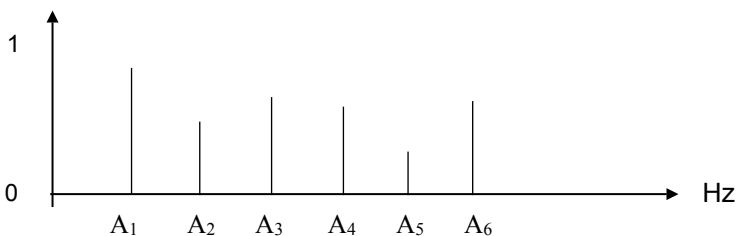


Figura 2.35 Spettro

Questo grafico è una schematizzazione del tutto indicativa e non riporta un riferimento ad una precisa scala: l'obiettivo è infatti ora quello di mettere in evidenza l'aspetto *dinamico* dello spettro che rappresenta l'evoluzione timbrica di un suono.

Eseguendo infatti tutta la procedura del calcolo delle ampiezze delle armoniche su periodi successivi di un segnale, si ottengono spettri in cui si riscontra una variazione del valore di ogni singola armonica. Questo significa che all'evoluzione della forma d'onda corrisponde l'*evoluzione dello spettro* come mostrato in Figura 2.36.

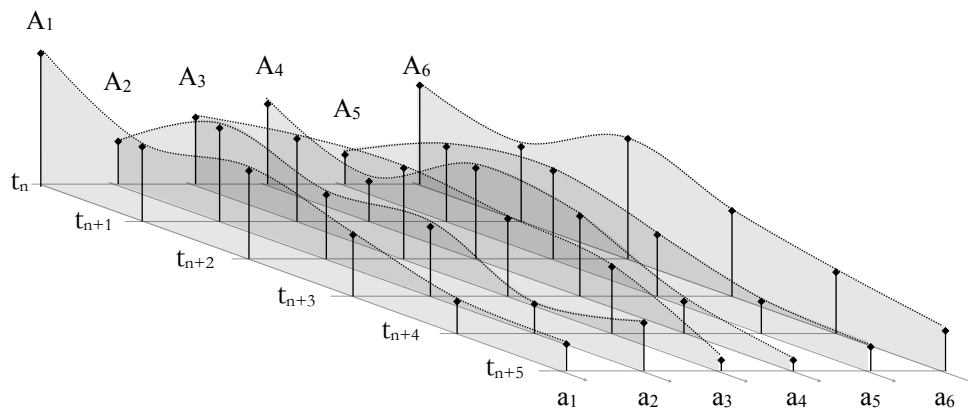


Figura 2.36 Spettri dinamici

La variazione nel tempo del valore di un'armonica si chiama **involuppo**. Nei suoni reali, in special modo nei suoni emessi da strumenti musicali, l'intensità totale e quella delle componenti non è costante per tutta la sua durata: in generale, nell'evoluzione di un suono dal momento dell'emissione allo spegnimento, si possono individuare, in special modo riferito all'andamento del *volume globale*, almeno tre fasi principali: *attacco*, *sostegno*, *rilascio*.

È evidente comunque che la suddivisione in tre fasi è del tutto approssimativa della realtà: esistono infatti casi in cui la schematizzazione non risulta del tutto adatta come ad esempio per il suono prodotto da una campana o da un tamburo che, in modi diversi, non presentano una vera e propria fase di sostegno. La terminologia introdotta è tuttavia utile a livello analitico ed è ormai entrata nell'uso comune.

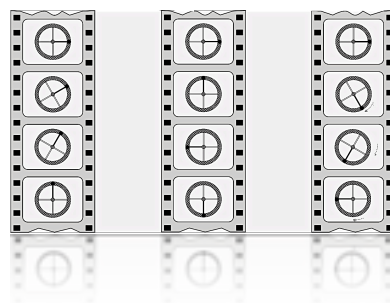
Tutto insieme, involuppo delle armoniche e involuppo del volume globale, contribuisce alla rappresentazione del timbro. Si riprenderà l'argomento nel Capitolo 4 dedicato alla *sintesi sonora* (§4.4) dove l'involuppo diventa anche e soprattutto "strumento di modellazione" dei suoni sintetici.

Riferimenti

- [1] *Le fascinant nombre π* - Jean-Paul Delahaye, Bibliothèque pour la science, diffusion BELIN, 8,rue Fèron, 75006 Paris, 1998
- [2] it.wikipedia.org/wiki/Jean_Baptiste_Joseph_Fourier
- [3] www.dm.unipi.it/mat_dia_med/Fourier.pdf
- [4] *su Google immagini* > schema elettrico onda quadra
- [5] it.wikipedia.org/wiki/Onda_quadra
- [6] www.falstad.com/fourier/

Analógico e digitale

- 3.1 Segnale analógico
- 3.2 Digitalizzazione del segnale analógico
- 3.3 Campionamento
- 3.4 Quantizzazione
- 3.5 Codifica
- 3.6 Convertitori
 - 3.6.1 Convertitore digitale/analógico DAC
 - 3.6.2 Convertitore analógico/digitale ADC
- 3.7 Considerazioni



Come abbiamo visto nel Paragrafo 1.1, il termine *analógico* viene usato come sinonimo di *continuo*. Va rilevato tuttavia che prima dell'arrivo della tecnologia digitale, questo termine era del tutto inesistente: le apparecchiature ed i supporti di registrazione, di trasmissione e di riproduzione sonora e visiva erano semplicemente *elettroniche*.

Del resto, possiamo osservare che anche prima dell'introduzione del treno a trazione elettrica, il *treno a vapore* era semplicemente *il treno*; e che prima dell'invenzione e dell'introduzione del motore a reazione, l'*aereo ad elica* era semplicemente l'*aereo*.

Perciò, *a-vapore*, *ad-elica* e *analógico*, sono dizioni coniate per distinguere una nuova tecnologia che all'inizio ha affiancato e poi ha gradualmente sostituito quella già esistente.

Digitale è una di quelle parole apparentemente di origine anglosassone, ma in realtà di origine latina. Come *media* e *bus*: *media* è il plurale del neutro *medium* che vuol dire *mezzo* nel senso di *centrale* (dito medio) e di *attrezzo* o *strumento*; *bus*, che nell'accezione corrente significa mezzo di trasporto pubblico, viene in realtà da *omni-bus* che vuol dire *per-tutti* e quindi *pubblico* per distinguerlo (quando fu introdotto a Parigi e a Londra) dalle carrozze private; quindi *bus* altro non è che la desinenza latina dell'ablativo plurale di *omnis*.

Digitale viene da *digitus*, che in latino ha il doppio significato di *numero* e di *dito* e quindi, in senso lato, di *contare*. Per cui l'espressione *digitare un numero* (per esempio il PIN) è una ridondanza, anzi, quasi una tautologia. L'espressione è però più valida se le si dà il significato di *comporre (sul telefono o sul computer) un valore numerico premendo i tasti con le dita*. Nel seguito useremo le parole *digitale* e *numerico* in maniera del tutto equivalente e come più suona meglio nel contesto.

3.1 Segnale analogico

La Figura 1.1 descrive quello che succede in maniera meccanica nel processo di cattura del fenomeno fisico/acustico (e cioè della variazione di pressione dell'aria) e della sua trasformazione in rappresentazione grafica che abbiamo detto chiamarsi *segnale*. La tecnologia del disco di vinile mostra chiaramente quella che è l'essenza dell'analogico e del continuo, come mostra la Figura 3.1 in questo ingrandimento.

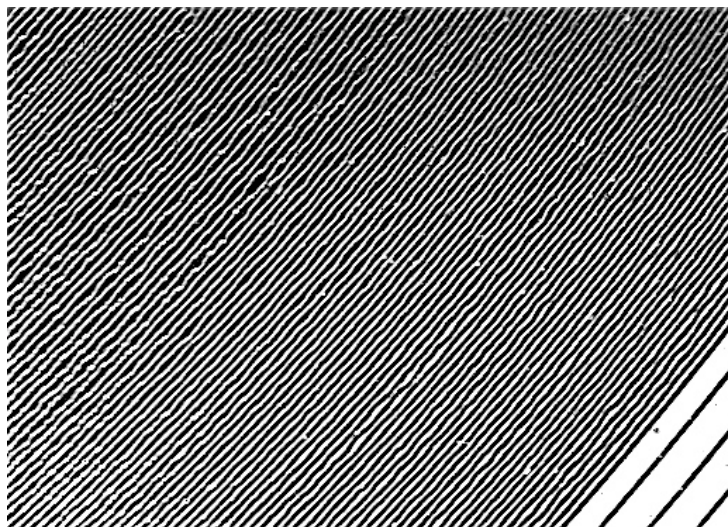


Figura 3.1 Solco su disco di vinile

La puntina di lettura del giradischi rileva in maniera meccanica le oscillazioni fissate nel solco; la puntina oscilla all'interno di un campo elettro-magnetico costituito da un solenoide che genera un segnale elettrico utilizzato (amplificato) dall'altoparlante per produrre suono.

Essendo tutto continuo (fenomeno acustico, segnale micro-fonico, solco che memorizza meccanicamente le oscillazioni continue e lettura) anche la formalizzazione della teoria dei segnali e del loro trattamento fa riferimento alla matematica del calcolo infinitesimale degli *integrali* e delle *derivate*.

Ma con la realizzazione degli elaboratori elettronici e più in generale con lo sviluppo della tecnologia digitale, si è sviluppata anche una nuova formalizzazione del fenomeno acustico e delle modalità operative di generazione, di trattamento e di riproduzione dei segnali, dovute al diverso modo di rappresentazione.

Con l'uso dell'elaboratore, il segnale relativo al fenomeno acustico viene rappresentato in forma di successione di valori numerici e perciò le tecniche di analisi, di sintesi, di elaborazione, di registrazione e di riproduzione diventano algoritmi informatici e la teoria matematica di supporto fa riferimento a successioni numeriche e ad operatori come sommatorie ed equazioni alle differenze piuttosto che ad integrali e derivate.

3.2 Digitalizzazione del segnale analogico

Dunque, *digitale* e/o *numerico* diventano etichette che caratterizzano una tecnologia rispetto ad un'altra e cioè la tecnologia informatica del discreto, rispetto alla tecnologia elettronica analogica del continuo (nastri a cassette, dischi di vinile, ecc.). Vedremo più avanti il significato di *discreto*.

Ma se è vero che tutta l'elaborazione, la memorizzazione e la trasmissione del segnale viene ora eseguita in termini numerici attraverso algoritmi eseguiti sul computer o sistemi digitali specializzati, resta comunque il fatto che *noi* umani continuiamo ad essere indubbiamente "analogici" per cui i due componenti di ingresso ed uscita audio di un sistema di registrazione, di amplificazione e di trasmissione radio (il microfono e l'altoparlante) sono e restano dispositivi elettronici analogici.

Si tratta quindi di operare una conversione dell'analogico al digitale quando si vuole introdurre un segnale audio (proveniente per esempio da un microfono) nel computer e di operare una conversione dal digitale all'analogico quando si vuole fare uscire un segnale audio dal computer verso un altoparlante.

La figura seguente mostra i dispositivi essenziali per eseguire le due operazioni: la *conversione analogico-digitale* (A-to-D oppure A/D) e la *conversione digitale-analogico* (D-to-A oppure D/A).

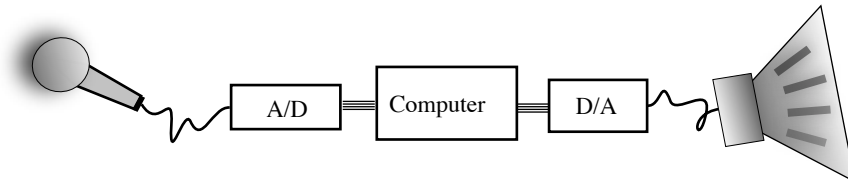


Figura 3.2 Conversione A-to-D e D-to-A di segnale audio

Nell'insieme, il processo di conversione dall'analogico al digitale si chiama **digitalizzazione del segnale** ed è proprio al momento della conversione che si stabiliscono i parametri relativi al grado di approssimazione e quindi alla qualità con cui il segnale reale viene messo in rappresentazione numerica. Per la conversione dal digitale all'analogico (D/A oppure D-to-A) si parla di *rendering* e cioè di come le informazioni numeriche vengono utilizzate per generare *segnale analogico* e quindi *suono* attraverso gli altoparlanti. Si tratta allora di vedere cosa voglia dire operativamente *digitalizzare un segnale analogico* e cosa comporti tale operazione in termini di vantaggi e svantaggi.

Il processo di digitalizzazione del segnale analogico consiste principalmente di tre operazioni distinte: il **campionamento**, la **quantizzazione** e la **codifica**.

Il *campionamento* è l'operazione con cui ad intervalli di tempo regolari prestabiliti, viene prelevato il valore istantaneo del segnale.

La *quantizzazione* è l'operazione di approssimazione ad un valore intero di una scala predefinita, del valore reale del campione.

La *codifica* è l'operazione con cui i valori quantizzati vengono rappresentati in forma numerica, di solito espressi in forma binaria.

3.3 Campionamento

Nella dizione corrente un *campione* è una piccola porzione di sostanza prelevata da un contesto più ampio per eseguire una serie di analisi chimico/fisiche, come ad esempio si fa per il sangue, l'acqua di falda, il terreno, ecc. In questi casi l'operazione di campionamento si fa una tantum e la si ripete solo se esistono ragionevoli motivazioni che lo richiedono. Nel nostro caso, il

campionamento (in inglese *sampling*) prevede invece una lunga serie di brevi prelievi, detti per l'appunto *campioni*, dal naturale e continuo fluire del fenomeno acustico, prelievi eseguiti a successivi e regolari intervalli di tempo.

La distanza temporale che intercorre tra un prelievo e il successivo è detta **periodo di campionamento** e quindi il numero di campioni prelevati nell'unità di tempo, che di solito è 1 secondo, è detta **frequenza di campionamento**.

Il processo di campionamento in sé non è una novità vera e propria introdotta con la tecnologia informatica per l'audio digitale: il Cinema (fine '800) e la TV (metà '900) funzionano proprio sul principio del campionamento.

È noto che il movimento delle immagini di un film è il risultato della proiezione di una sequenza di 24 fotografie al secondo relative ad un'immagine reale in movimento.

Oggi, nell'era della fotografia digitale, tutte le funzioni meccaniche vengono simulate a programma, compreso, con un pizzico di nostalgia, il "click" acustico dello scatto vero e proprio, del tutto estraneo al contesto puramente elettronico!

Va ricordato che con una terminologia ormai non più in uso, le fotografie venivano chiamate anche *istantanee*, proprio perché si tratta di catturare e di congelare un istante della realtà continua: quanto dura l'istante durante il quale avviene lo scatto dell'otturatore e la cattura dell'immagine sulla pellicola dipende essenzialmente dalla tipologia del soggetto e dalle condizioni di luce (combinazione tra tempo e diaframma ben nota ai fotografi), ed è comunque dell'ordine di qualche millisecondo.

Dunque, la ripresa cinematografica tradizionale consiste di fatto nello scatto di 24 fotografie istantanee al secondo distanziate l'una dall'altra per l'appunto $1/24$ di secondo e cioè 41,66... millisecondi, con una durata di scatto dell'ordine di qualche millisecondo (diciamo 3,66 millisecondi).

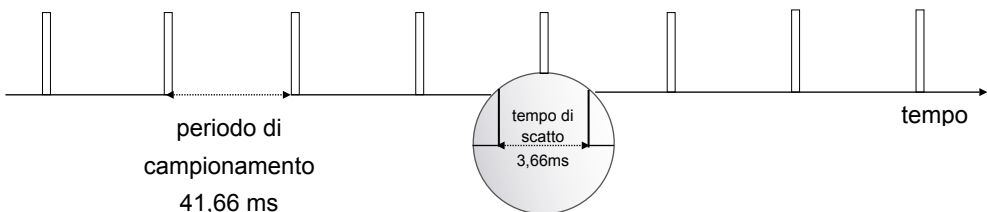


Figura 3.3 Periodo di campionamento e tempo di scatto

Durante i circa 38msec ($38=41,66-3,66$) in cui i meccanismi della cinepresa provvedono alla chiusura dell'obiettivo tramite l'otturatore, all'avanzamento della pellicola ed alla preparazione per lo scatto successivo, nel rettangolo di ripresa può succedere di tutto, senza essere però catturato e memorizzato: una bottiglia lanciata in campo tra i giocatori di una partita di calcio da un focoso tifoso... se ciò accadesse proprio in quei 38 millisecondi in cui l'otturatore della cinepresa è chiuso, agli effetti della ripresa, e quindi della proiezione, è come se nulla fosse accaduto.

Nelle trasmissioni televisive dove si commentano le partite di calcio ed in particolare quando si contestano, ad esempio, situazioni di fuori-gioco oppure azioni tipo il fallo-di-mano ecc., si esaminava la ripresa alla *moviola*, che consentiva di esaminare la scena al rallenty e fotogramma per fotogramma.

La moviola, (dall'inglese *movie=cinema*) era la macchina meccanica usata per il montaggio dei film. Tuttavia anche la moviola non sempre dava un indiscutibile responso, perché non era detto che l'azione fallosa fosse accaduta durante lo scatto del fotogramma ma, proprio perché veloce, era molto probabile ($38/41,66 \approx 90\%$) che fosse avvenuta durante i 38 millisecondi in cui l'otturatore era chiuso.

Con le tecniche di ripresa ormai a sola tecnologia digitale sia per la televisione sia per il cinema, si è passati a tassi di ripresa dell'ordine di $100\div 150$ fotogrammi al secondo per cui, nello stesso frammento di tempo, c'è ora una maggiore quantità di informazione ed anche la riproposta al rallenty e molto più fluida ...anche se, per definizione, non continua!

La scelta della frequenza di campionamento è dunque cruciale, perché si tratta di non perdere informazioni relative alla realtà, in vista della successiva proiezione ma anche per soddisfare la necessità pratica di trovare un giusto compromesso con la quantità di informazioni in relazione al supporto di memorizzazione.

Quando ancora si usava la pellicola, sul proiettore di una sala cinematografica la bobina di un tempo-di-film della durata di 45/50 minuti, aveva un diametro di circa un metro e la velocità di scorrimento davanti all'obiettivo, con 24 fotogrammi di altezza ciascuno di circa 26 millimetri, era di oltre mezzo metro al secondo.

Un campionamento a 100 fotogrammi/secondo avrebbe implicato una velocità di scorrimento di $2\div 2,5$ m/sec e quindi bobine di $3\div 4$ metri di diametro... insomma, si sarebbero creati un sacco di complicazioni meccaniche, economiche e gestionali. Problemi, questi, ormai superati dalla conversione al digitale sia nella ripresa sia nella proiezione in sala, tant'è che alcuni film oggi utilizzano da 50 a 100 frame/sec.

Comunque, con una velocità di ripresa e di proiezione di 24 o 25 fotogrammi al secondo, com'è ancora nello standard cinematografico e in quello televisivo, qualunque movimento che avviene con velocità sufficientemente bassa com'è nella gestualità delle persone, viene ricostruito in maniera fedele alla realtà. Ma fenomeni che accadono con velocità superiori, come tipicamente succede per gli oggetti rotanti, non sempre vengono correttamente ricostruiti durante la proiezione. Si pensi ad esempio alle ruote delle automobili con i cerchi a raggi o comunque con una figura ripetuta più volte sul cerchio: succede spesso di vedere al cinema o in TV le ruote girare in senso contrario rispetto al senso di marcia del mezzo.

Consideriamo una ruota con 4 raggi e del diametro di 30 cm e quindi una circonferenza di 30π cm = 94 cm circa. Nella figura seguente consideriamo inoltre che il senso di rotazione sia anti-orario e che la sequenza delle immagini sia dall'alto verso il basso.

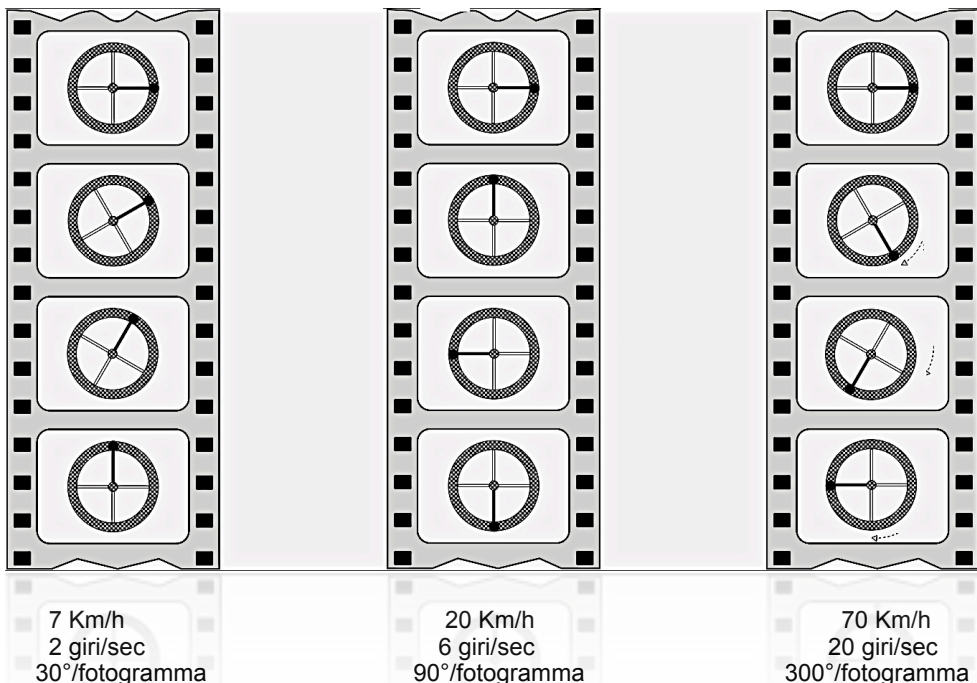


Figura 3.4 Movimento reale ed apparente

Quando inizialmente il veicolo si muove presumibilmente ad una velocità di circa 7 Km/h pari a $0,96 \times 2 = 1,92$ metri al secondo, la ruota compie 2 giri completi ogni secondo, pari ad un angolo di $2 \times 2\pi$ radianti. Poiché i fotogrammi di ripresa sono 24 al secondo, ciascun raggio della ruota viene ripreso ad incrementi di angolo di $4\pi/24 = \pi/6$ radianti ($=30^\circ$) ed in proiezione il movimento viene ricostruito correttamente.

Quando il veicolo raggiunge una velocità di ≈ 20 Km/h pari a 5,76 m/sec e cioè $6 \times 0,96$ m/sec., la ruota compie 6 giri completi al secondo pari ad un angolo di $6 \times 2\pi$ radianti. Questa volta i raggi vengono ripresi ad incrementi di $6 \times 2\pi/24 = \pi/2$ radianti ($=90^\circ$) e siccome i raggi sono quattro, le immagini successive sono uguali tra loro.

Il risultato apparente è di una ruota ferma di un'auto in corsa. In realtà si vede che qualcosa non torna proprio perché l'immagine della ruota è confusa e questa confusione è data dalla continua sovrapposizione di immagini diverse come senz'altro diversi sono i raggi nei particolari.

Quando il veicolo raggiunge una velocità di ≈ 70 Km/h pari a 19,2 m/s e cioè $20 \times 0,96$ m/s, la ruota compie ogni secondo 20 giri completi pari ad un angolo di $20 \times 2\pi$ radianti. Questa volta i raggi vengono ripresi ad incrementi di $20 \times 2\pi/24 = 10\pi/6$ radianti ($=300^\circ$).

Ma $10\pi/6 = 5\pi/3 = -\pi/3$ e cioè -60° e siccome sul fotogramma non c'è nessuna informazione riguardo al fatto che la ruota abbia fatto "quasi" un giro completo in senso antiorario o "solo" una piccola porzione di giro in senso orario, l'occhio che lo vede "preferisce" la più semplice tra le due situazioni.

Questo vuol dire che la successione dei fotogrammi riproposti viene interpretata come un *movimento apparente* che risulta essere contrario a quello originale. Per l'esattezza, la frequenza di questa rotazione apparente è data da $-\pi/3 = -8\pi/24 = -4 \times 2\pi/24$ che corrisponde a -4 giri al secondo e cioè una frequenza di -4Hz, dove il segno - significa *senso di rotazione contrario a quello originario*.

Perciò, come dice il **teorema di Shannon/Nyquist**, se si campiona un segnale che contiene frequenze superiori alla *metà della frequenza di campionamento* (e cioè se si *sotto-campiona*) si genera un movimento periodico *apparente* diverso da quello originale.

Il fenomeno è conosciuto come **aliasing** o **foldover**: *aliasing* è parola latina (*alias=altro, diverso*: si pensi alla parola *alibi*) "decli/coniugata" all'inglese e dice in maniera generica che il *sotto-campionamento* produce una cosa diversa dalla realtà. *Fold-over* (ripiegare intorno) è più precisa perché dice *come* e *dove* viene generata "l'altra cosa".

Come abbiamo visto nel Capitolo 2 (§2.1, Figure 2.6 e 2.7) il *numero di giri al secondo* corrisponde alla *frequenza*. Consideriamo allora un asse orizzontale sul quale segnare le frequenze relative al campionamento dei fotogrammi alle diverse velocità di rotazione della ruota in corrispondenza delle diverse velocità lineari del veicolo:

$$\begin{aligned} F_c &= 24\text{Hz} & F_c/2 &= 12\text{Hz} \\ 7\text{Km/h} &\longrightarrow 2\text{Hz} \\ 20\text{Km/h} &\longrightarrow 6\text{Hz} \\ 70\text{Km/h} &\longrightarrow 20\text{Hz} \longrightarrow -4\text{Hz} \end{aligned}$$

La Figura 3.5 mostra come le rotazioni a frequenze che stanno al di sotto di $F_c/2$ (nel nostro caso = 12Hz) come 2 e 6 Hz, vengono ricostruite correttamente, mentre quelle a frequenze che stanno al di sopra di $F_c/2$ generano rotazioni la cui frequenza risulta essere riflessa come in uno specchio (o ripiegata=*folded*) proprio intorno alla frequenza $F_c/2$ (detta **frequenza di Nyquist, F_n**) con senso di rotazione contrario a quello originario.

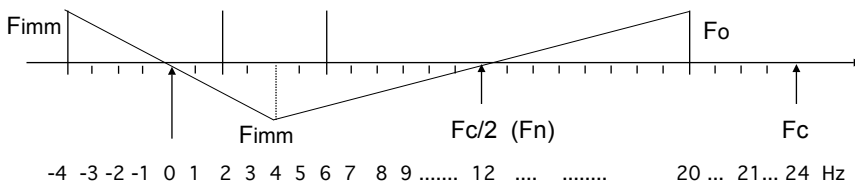


Figura 3.5 Foldover o aliasing

Ed è proprio perché la F_n si comporta come uno specchio che si parla di *frequenza immagine*. Perciò, detta F_o la frequenza originale e F_{imm} la frequenza immagine, si ha

$$F_{imm} = F_n - (F_o - F_n) = 2F_n - F_o = F_o - F_c$$

dove F_c rappresenta la frequenza di campionamento e F_o la frequenza del segnale originale. Abbiamo visto infatti che la rotazione di 20 giri/sec che corrisponde a 20 Hz, genera una frequenza immagine di (20-24) Hz = -4 Hz.

Il risultato negativo indica proprio il fatto che il senso di rotazione è contrario rispetto all'originale e viene indicata sul grafico con un segmento tratteggiato e nella parte inferiore del grafico stesso. La rotazione contraria si traduce in uno *sfasamento* di π radianti = 180° .

Ma cosa succede se la velocità del veicolo aumenta al punto che la ruota gira con una frequenza che supera anche la frequenza di campionamento?

Dalla formula $F_{\text{imm}} = F_c - F_o$ risulta che la frequenza immagine cade, in questo caso, nell'intervallo tra 0 ed $F_c/2$ ma il senso di rotazione torna ad essere quello originale perché il segno del risultato è positivo. Ad esempio:

$$90 \text{ Km/h} \rightarrow 26 \text{ Hz} \quad \text{da cui} \quad F_{\text{imm}} = (26 - 24) \text{ Hz} = 2 \text{ Hz}$$

Si può verificare infatti che una rotazione pari a 26 Hz corrisponde, sulla pellicola, ad una rotazione di 2 Hz: infatti due successivi fotogrammi che distano $1/24$ di secondo, riportano una rotazione di $26 \times 2\pi/24$ radianti, e cioè

$$\begin{aligned} (2+24) \times 2\pi/24 &= \\ 4\pi/24 + 2\pi &= \\ \pi/6 + \text{un giro intero} &=> 30^\circ \end{aligned}$$

Riassumendo:

- 1) tra 0 e $F_c/2$ lo spettro viene riproposto correttamente;
- 2) tra $F_c/2$ ed F_c lo spettro viene replicato sullo spettro originale ma con senso di rotazione invertito e cioè sfasato di $180^\circ = \pi$ radianti;
- 3) lo spettro oltre F_c viene replicato su quello originale.

A tutto quanto detto fin qui, fa riferimento il **teorema di Shannon** in base al quale la *frequenza di campionamento* e cioè il numero di campioni prelevati nell'unità di tempo, deve essere fissata almeno al doppio della massima frequenza *che interessa* contenuta nel segnale originale. E più praticamente si può dire anche: fissata una frequenza di campionamento F_c , la frequenza $F_c/2$ è quella al di sotto della quale non si verifica *foldover/aliasing*.

Ci si deve aspettare dunque, come vedremo meglio nel seguito, che succeda così anche per gli spettri nelle regioni frequenziali sempre più alte. Ma ancora un'osservazione prima di tornare sull'audio digitale.

Alla luce del Sole, la realtà tutta si muove in maniera continua e quindi non vediamo mai le ruote delle automobili "girare all'indietro" rispetto al senso di marcia. Ma alla luce delle lampade elettriche, sì! Di notte, in grandi spazi come tipicamente sono i caselli autostradali illuminati da alti lampioni, quando le automobili si avvicinano rallentando ai box per pagare il pedaggio, è possibile osservare il fenomeno dell'*aliasing/foldover* "dal vivo".

Il fatto è che la luce dei lampioni è alimentata con corrente alternata come quella domestica che pulsa 100 volte al secondo (le due semionde sinusoidali positiva e negativa dei 50 Hz delle corrente di rete). L'occhio non si accorge di questa pulsazione a causa della persistenza dell'immagine sulla *retina* di circa 1/16 di secondo, per cui la luce appare continua. Ma non appena c'è un elemento rotante che si avvicina alla frequenza di 100 Hz, ecco che compare l'effetto.

È poi vero che l'effetto compare anche in pieno giorno alla luce del Sole quando la ruote vengono viste al di là di una cancellata o dei piloncini verticali posti a distanze regolari che sorreggono il guard-rail lungo le autostrade.

Si tratta comunque di un effetto artificiale che si chiama *effetto stroboscopio* che viene usato per calibrare la giusta velocità di rotazione dei giradischi di vinile, mediante tacche opportunamente distanziate in funzione del pulsare della frequenza di rete: quando l'immagine apparente è ferma allora la velocità di rotazione del piatto è quella giusta. L'effetto stroboscopio è il risultato dell'*interferenza* di due fenomeni periodici a frequenze diverse.

Quello che abbiamo visto nel caso della ripresa televisiva e/o cinematografica (sia a pellicola sia digitale) vale anche nel processo di campionamento di un segnale sonoro. È chiaro che ora si parla di frequenze molto più alte perché l'apparato uditivo umano è in grado di percepire suoni la cui frequenza è compresa nell'intervallo che va da un minimo di 16 Hz ad un massimo di 16.000 Hz.

Tuttavia, mentre nel cinema/televisione ogni campione è costituito da una figura e quindi da una superficie (due dimensioni, base e altezza) rappresentata da alcuni milioni di singoli punti, nel caso del segnale audio un campione è costituito da un singolo valore numerico. Per questo motivo si dice anche che il segnale sonoro è *mono-dimensionale*.

Quindi, per rispettare il teorema di Shannon, che nel settore dell'audio è conosciuto come **teorema di Nyquist**, la frequenza di campionamento deve essere posta *almeno* a 32 KHz.

Vediamo allora quali sono gli effetti del campionamento e del sotto-campionamento per un segnale sonoro analogico. Per semplicità di esposizione prendiamo in considerazione una sinusoida a 100 Hz ed una frequenza di campionamento di 800 Hz.

Poiché il *periodo* è l'inverso della *frequenza* ($T=1/F$) un periodo completo della sinusoida dura $1/100 \text{ Hz} = 0,01$ secondi e il periodo di campionamento, che indichiamo con Δt , dura $1/800 \text{ Hz} = 0,00125$ sec. Questo vuol dire che un ciclo intero di sinusoida viene "fotografato" dal processo di campionamento a distanze temporali di $\Delta t = 0,00125 \text{ sec.}$, per un totale di 8 volte ad ogni ciclo.

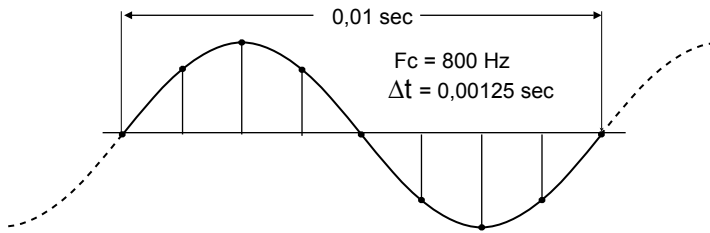


Figura 3.6 Sinusoide campionata

Il risultato del campionamento è una successione di valori che ricostruiscono correttamente la sinusoide, anche se “correttamente” è un termine eufemistico dal momento che 8 punti così distanziati e “sgranati” non riproducono di certo una sinusoide.

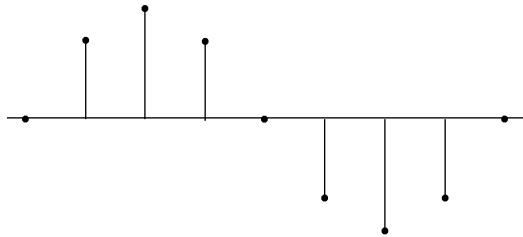


Figura 3.7 Risultato del campionamento

Va notato tuttavia che in una situazione reale ed operativa, come è ad esempio la frequenza di campionamento di 44.100 Hz dello standard CD, 8 campioni riproducono la frequenza di 5.512 Hz, che è già una frequenza molto alta. Ma questa è l'essenza del campionamento!

Cosa del tutto diversa è quella che accade se la frequenza sta al di sopra di $F_c/2$ che nel nostro caso è 400 Hz: un suono sinusoidale a frequenza di 600Hz si manifesta con 6 oscillazioni nello stesso spazio temporale di 0,01 secondi.

$F_c/2$ è la **frequenza di Nyquist**.

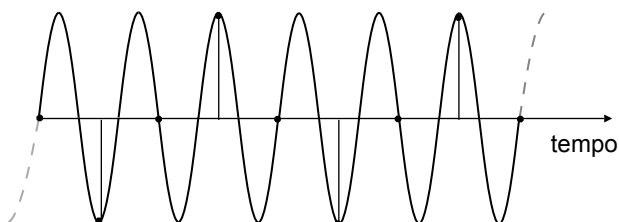


Figura 3.8 Sinusoide sotto-campionata

Come prima, nello spazio temporale di 0,01sec. il processo di campionamento a 800 Hz prende 8 valori e siccome trova quello che c'è in quel momento, con la successione di campioni si ricostruisce una sinusoide a 200Hz (due oscillazioni di 0,01 secondi) e invertita rispetto all'asse del tempo.

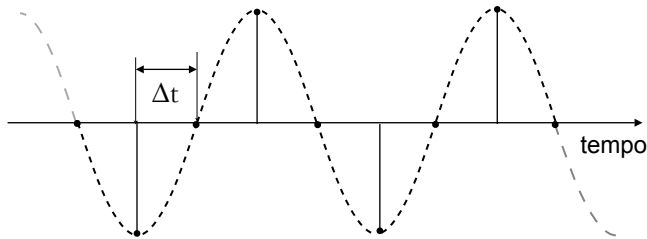


Figura 3.9 Risultato del sotto-campionamento

È questa la frequenza immagine data da $F_{imm} = F_s - F_c$ che nel nostro caso vale $(600-800) \text{ Hz} = -200\text{Hz}$.

La figura 3.10 riassume sull'asse delle frequenze quanto accade contemporaneamente sull'asse del tempo nelle figure 3.6, 3.7, 3.8 e 3.9, dove F_s è la frequenza del segnale originario ed F_c è la frequenza di campionamento.

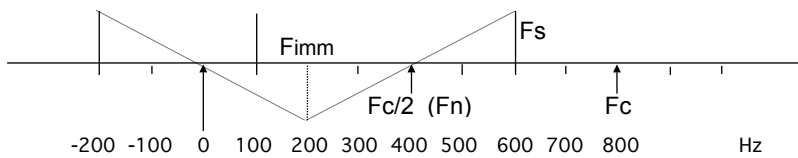


Figura 3.10 Risultato del campionamento e del sotto-campionamento

In una situazione reale come quella della figura seguente che rappresenta 40 millisecondi circa di musica, è necessaria una frequenza di campionamento molto alta per memorizzare e ricostruire poi tutta l'informazione acustica ivi contenuta.

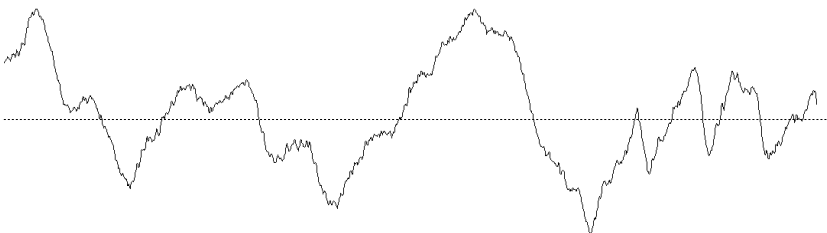


Figura 3.11 Segnale musicale

Con una frequenza di campionamento di 800 Hz, che equivale ad un periodo T pari a $\Delta t=0,00125\text{sec.}$ come quello visto nell'esempio, si prenderebbero una quarantina di valori del segnale che sarebbero del tutto insufficienti alla ricostruzione: tutti i dettagli delle "increspature" dovute ai suoni di frequenza più alta, responsabili del colore timbrico del suono, andrebbero perduti.

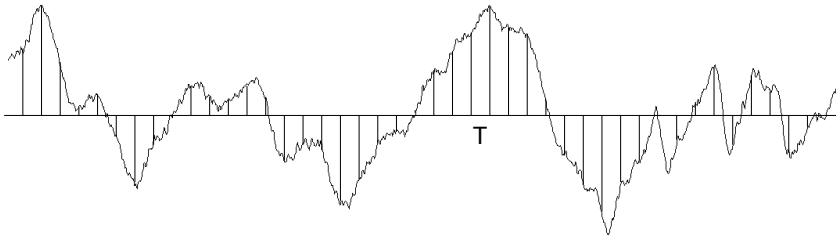


Figura 3.12 Sotto-campionamento del segnale musicale

Inoltre, gli effetti del *foldover* sarebbero disastrosi! Accadrebbe che tutte le frequenze comprese tra $F_c/2$ e F_c sarebbero riflesse e sovrapposte a quelle comprese tra 0 e $F_c/2$ (*frequenza di Nyquist*) come pure quelle comprese tra F_c e $3F_c/2$ (sovrapposte ma non riflesse).

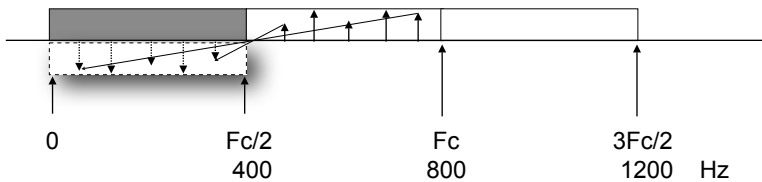


Figura 3.13 Effetto del sotto-campionamento

Ci sono due modi per evitare tutto ciò e di solito vengono usati entrambi:

- 1) aumentare la frequenza di campionamento;
- 2) *filtrare*, vale a dire, eliminare le frequenze che stanno al di sopra di $F_c/2$.

Nella Figura 3.12, aumentare la frequenza di campionamento significa diminuire il valore di Δt e cioè infittire i momenti di prelievo, il che alla fin fine vuol dire prelevare maggiori dettagli nel segnale, come si può riscontrare anche visivamente.

Di *filtri* si dirà ampiamente nel Capitolo 5. Per il momento è sufficiente pensare ad un filtro come ad un dispositivo che agisce sul segnale secondo certi criteri e lo modifica in modo tale da alterarne il contenuto spettrale.

Il concetto di filtro coincide con quello che si ha a livello colloquiale e cioè di *scelta e/o selezione* operata in base a prefissati criteri su una pluralità di “oggetti”. Per esempio: un colino, un setaccio, noi stessi nei rapporti interpersonali, ecc.

Nell'elettro-acustica i filtri si caratterizzano in base alla modalità di intervento sullo spettro (riscontrato a posteriori). Nel caso in questione, si tratta di un *filtro passa-basso* che elimina da un segnale audio le frequenze più alte; e proprio perché questo filtro ha il compito di prevenire l'effetto di un eventuale sotto-campionamento, viene detto anche **filtro anti-aliasing**.

Nel processo di conversione A-to-D il filtro anti-aliasing è un dispositivo elettronico che viene posto tra sorgente del segnale (ad esempio, microfono) e il convertitore analogico-digitale come mostrato nella Figura 3.14.

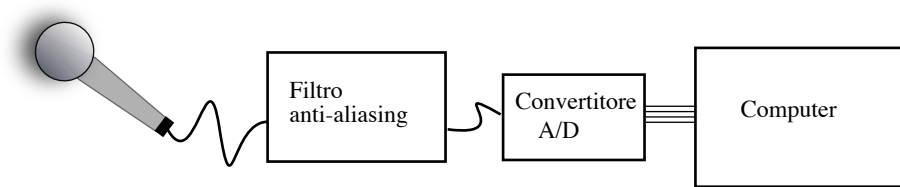


Figura 3.14 Filtro anti-aliasing nel processo di conversione A/D

Ciò sostanzialmente è fattibile proprio perché il segnale audio è un segnale mono-dimensionale e cioè relativo ad “una sola” quantità fisica che varia nel tempo: pressione dell’aria trasdotta in tensione elettrica, come quella che esce da un microfono o da un jack di chitarra.

Ma nella ripresa cinematografica, dove il “segnale” è bi-dimensionale perché costituito da immagini, non è possibile filtrare elementi ruotanti e quindi non è possibile eliminare l'effetto dell'aliasing se non alzando notevolmente la frequenza di campionamento che, come abbiamo già visto, comporterebbe appesantimenti tecnologici e pratici poco giustificabili.

Anzi, come si vede spesso in alcune pubblicità di automobili, si fa di necessità virtù utilizzando proprio l'effetto-stroboscopio dell'aliasing in modo artistico.

Ma quando le velocità di rotazione sono troppo alte come succede per le pale delle eliche degli aerei non c'è proprio possibilità di eliminare l'effetto del fold-over/aliasing.

3.4 Quantizzazione

La **quantizzazione** è l'operazione con la quale una grandezza continua viene espressa con un numero finito di valori. *Quantizzare* viene dal latino *quantum* e si riferisce a s/oggetti non divisibili come ad esempio le persone: un gruppo è fatto da 4 o da 5 persone... mai da 4,35!

Quantizzare corrisponde perciò ad *approssimare valori che variano nel continuo con valori interi*. Si pensi ad uno scivolo ed una gradinata:

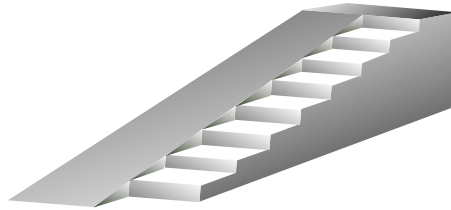


Figura 3.15 Scivolo e gradinata

sullo scivolo si può essere a qualunque livello con continuità tra il punto più basso ed il punto più alto; ma sulla scalinata si è su un gradino o si è su quello precedente o su quello successivo. Sullo scivolo c'è il continuo mentre sulla scalinata c'è il *discreto* (*discretum*=distaccato, separato).

“Approssimare” viene da *ad-proximum* e cioè *verso il più vicino* perché *proximo*=più vicino (e non a caso *Proxima Centauri* è la stella *più vicina* al Sole).

Ad esempio: il valore intero più vicino a 45,32 è 45 e il valore intero più vicino a 45,84 è 46:

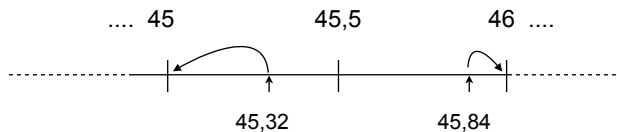


Figura 3.16 Approssimazione di un valore reale a numero intero

Nel linguaggio di tutti i giorni si tende a dare al termine *prossimo* il significato di *successivo* come ad esempio: la “prossima puntata”, la “prossima volta”, la “prossima fermata del pullman”; infatti se si è sull'autobus e l'autobus è appena partito dalla fermata corrente, con “la prossima fermata” si intende ovviamente la successiva.

Ma in effetti la “prossima” in senso etimologico (*proxima*), e cioè “la più vicina” in termini di distanza, è quella appena lasciata. Per essere corretti bisognerebbe dire perciò la “fermata successiva”, la “puntata successiva”, la “volta successiva”; ma essendo il fattore temporale (e non quello spaziale) quello determinante, nella dizione corrente *prossimo* equivale a *successivo*.

In Matematica prevale invece il fattore spaziale (topologico, per essere precisi) e quindi l'operazione di approssimazione viene eseguita alla lettera con questo algoritmo:

se il valore decimale (le cifre dopo la virgola) è minore di 0,5
allora si prende il valore intero più basso eliminando i valori dopo la virgola;
altrimenti (cioè se il valore dopo la virgola è maggiore o uguale a 0,5)
 si prende il valore intero successivo.

Si dice anche *approssimare per difetto* (caso *allora*) o *approssimare per eccesso* (caso *altrimenti*).

Più elegantemente, si ottiene lo stesso risultato *aggiungendo sempre* il valore 0,5 al valore reale in questione ed *eliminando sempre* le cifre decimali dopo la virgola che risultano, senza ricorrere al *se-allora-altrimenti*.

Così ad esempio:

$$45,32+0,5 = 45,82 \quad \text{dà} \quad 45$$

$$45,84+0,5 = 46,34 \quad \text{dà} \quad 46$$

È chiaro che approssimando valori reali a valori interi, e cioè quantizzando, si commette un errore che viene per l'appunto detto *errore di quantizzazione* ed è misurabile in termini di differenza tra il valore reale ed il valore numerico intero utilizzato, che non è mai comunque (per definizione) più grande di 0,5 *in valore assoluto* e cioè sempre compreso tra -0,5 e +0,5.

I segnali sonori come quelli rappresentati nella Figure 3.11 e 3.12 variano tra un valore minimo ed un valore massimo che dipende da una scala di riferimento prefissata. Si tratta allora di scegliere i termini della quantizzazione e cioè un congruo numero di livelli verso i quali approssimare i valori reali della forma d'onda in esame.

Nel caso di una quantizzazione a 2 soli livelli, tutti i valori che stanno tra 0 e 0,4999... vengono approssimati al valore 0 mentre tutti i valori che stanno tra 0,5 e 1 vengono approssimati ad 1.

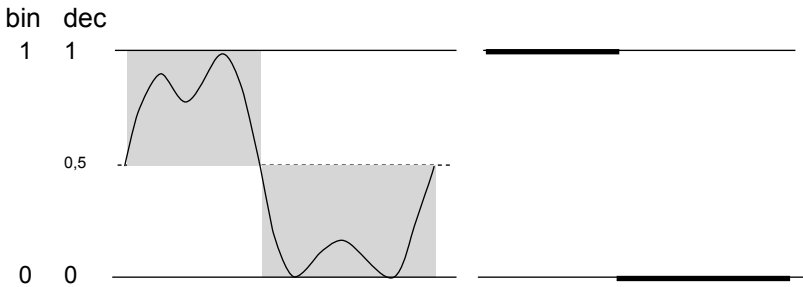


Figura 3.17 Quantizzazione a 2 livelli

Se i livelli sono 4 che vanno dal minimo di 0 al massimo di 3

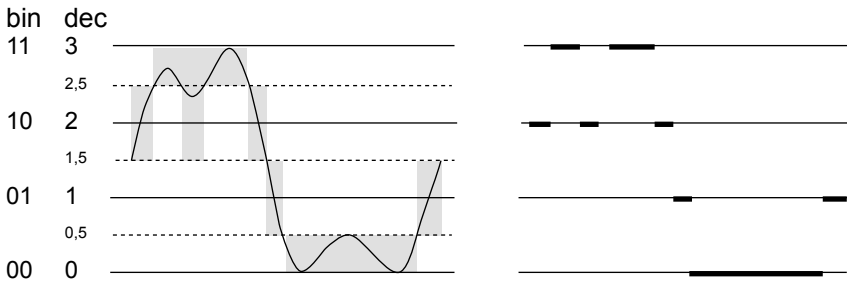


Figura 3.18 Quantizzazione a 4 livelli

- i valori che stanno tra 0 e 0,4999... vengono quantizzati a 0
- i valori che stanno tra 0,5 e 1,4999... vengono quantizzati a 1
- i valori che stanno tra 1,5 e 2,4999... vengono quantizzati a 2
- i valori che stanno tra 2,5 e 3 vengono quantizzati a 3

Se i livelli sono 8 che vanno dal minimo di 0 al massimo di 7

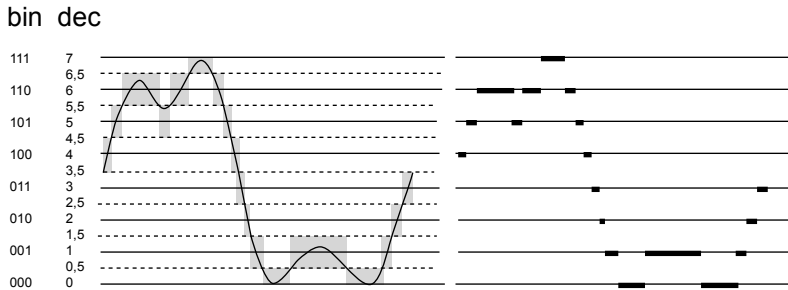


Figura 3.19 Quantizzazione a 8 livelli

i valori che stanno tra 0 e 0,4999... vengono quantizzati a 0
 i valori che stanno tra 0,5 e 1,4999... vengono quantizzati a 1
 i valori che stanno tra 1,5 e 2,4999... vengono quantizzati a 2
 i valori che stanno tra 2,5 e 3,4999... vengono quantizzati a 3
 i valori che stanno tra 3,5 e 4,4999... vengono quantizzati a 4
 i valori che stanno tra 4,5 e 5,4999... vengono quantizzati a 5
 i valori che stanno tra 5,5 e 6,4999... vengono quantizzati a 6
 i valori che stanno tra 6,5 e 7 vengono quantizzati a 7

Poiché il modo di aumentare il numero dei livelli di quantizzazione *nlq* lasciando invariata l'ampiezza massima consiste nel raddoppiare il numero dei precedenti livelli dimezzando così lo spazio tra un livello e l'altro, il *Nlq* usato per la codifica ha corrispondenza col numero di bit usato per rappresentare il numero binario che esprime i livelli.

Ad esempio: per 2 livelli di quantizzazione è sufficiente 1 bit ($2^1=2$, Figura 3.17); per 4 livelli sono necessari 2 bit ($2^2=4$, Figura 3.18); per 8 livelli sono necessari 3 bit ($2^3=8$, Figura 3.19); ...e così via.

E poiché il valore numerico intero cui viene approssimato il valore reale viene di solito espresso per l'appunto in *notazione binaria*, l'errore di quantizzazione dipende dalla quantità di cifre binarie utilizzate.

L'errore di quantizzazione viene indicato con

$$\text{err}_q = | V_r - V_q |$$

dove V_r è il valore effettivo del segnale analogico e V_q è il valore quantizzato e le due stanghette verticali stanno ad indicare il *valore assoluto* e cioè la misura senza segno della distanza tra i due valori.

Come abbiamo già detto, l'errore di quantizzazione è sempre compreso tra $-0,5$ e $+0,5$. Infatti:

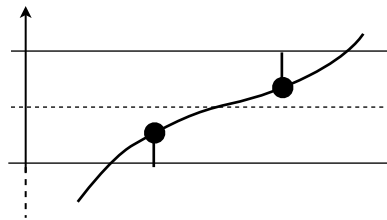


Figura 3.21 Errore di quantizzazione

Perciò il grafico risultante dagli scarti tra valori reali e quantizzati è costituito da valori casuali tra $-0,5$ e $+0,5$:

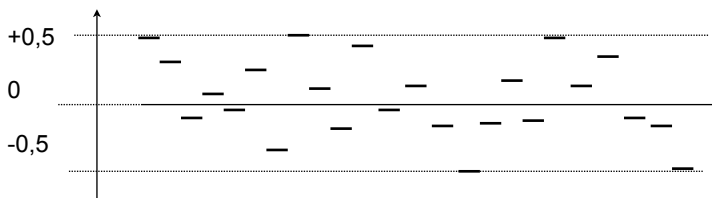


Figura 3.22 Scarti tra valori reali e quantizzati

L'errore di quantizzazione si traduce acusticamente in *rumore di quantizzazione* e si manifesta come “soffio” o “fruscio”. Di fatto, l'alterazione dovuta alla quantizzazione dei campioni espressi con valori reali “disturba e sporca” il segnale vero e proprio.

Per questo motivo il grado di bontà dell'approssimazione si esprime come rapporto tra la *potenza del segnale reale* (P_{segn}) e la *potenza del rumore di quantizzazione* (P_{rum}) e cioè **rapporto segnale/rumore** indicato con SNR (*Signal to Noise Ratio*) o anche S/N.

Se il **numero dei livelli di quantizzazione** Nlq è tale per cui il valore di n utilizzato nella relazione $Nlq=2^n$ è almeno 7 o 8, il rapporto tra la potenza del segnale e la potenza del rumore è dato con buona approssimazione da

$$SNR = \frac{P_{\text{segnale}}}{P_{\text{rumore}}} \approx 2^{2n}$$

dove il simbolo \approx vuol dire per l'appunto “uguale con buona approssimazione”. Questa formula porta ad una serie di considerazioni e chiarimenti.

Per definizione, un segnale è dato dalla variazione di una grandezza fisica (pressione dell'aria, voltaggio elettrico, ecc.) che veicola informazione. Tale variazione è “rilevata a posteriori” (ed è questa la cosa importante!) e quindi **non** descrivibile con una funzione matematica; non descrivibile, ma pur sempre analizzabile!

Un tipo di analisi che è possibile eseguire su un segnale, consiste nel calcolo della *potenza*; più correttamente, della *potenza media*. Questo calcolo è decisamente più semplice dell'analisi di Fourier (§2.3) e dà informazione su ciò che comunemente viene detto *volume* sonoro.

La figura seguente riporta due segnali campionati e quantizzati di due registrazioni eseguite con lo stesso livello di ingresso, relative a due diverse intensità (potenze) esercitate sullo strumento:



Figura 3.23 Segnali audio con lo stesso livello di ingresso

Il valore del livello della potenza media di una porzione di segnale costituito da n campioni si calcola con la *radice quadrata della media dei campioni elevati al quadrato* (in inglese **Root Mean Square**, RMS in sigla).

Dalla Fisica si sa infatti che la *relazione* esistente tra la *potenza* di una sorgente sonora misurata in Watt ed il *segnale* elettrico trasdotto dal microfono misurato in Volt, è di tipo *quadratico*; questo vuol dire che, per esempio, al *raddoppio* dell'ampiezza del segnale rilevato, corrisponde una sorgente sonora che l'ha generato, di potenza *quadrupla*.

SNR rapporto segnale/rumore		
n	scala lineare: 2^{2n}	scala logaritmica: dB
6	4096	36
7	16384	42
8	65536	48
9	262144	54
10	1048576	60
11	4194304	66
12	16777216	72
13	67108864	78
14	268435456	84
15	1073741824	90
16	4294967296	96

Figura 3.24 Rapporto segnale rumore SNR

L'intervallo volumetrico tra i due livelli minimo e massimo si chiama *gamma dinamica* o più semplicemente *dinamica*: la dinamica dipende dal numero di bit di quantizzazione e viene misurata anch'essa in deciBel come rapporto tra la minima e la massima potenza sonora registrabile.

È curioso osservare come il termine *dinamica* venga usato nello stesso ambito ed in un certo senso con lo stesso significato ma con due... “intenzioni” diverse a seconda del punto di vista dal quale si considera la cosa: nel dominio della fisica acustica, la dinamica è una misura ben precisa espressa con un valore in deciBel che nel caso particolare del digitale corrisponde a quello dell'SNR; in campo musicale, invece, la dinamica si riferisce a ben più sfumate e poco definibili modalità espressive operate da un musicista sul suo strumento mediante il controllo e la modulazione di intensità sonora.

Ma al di là della misura dell'SNR, della dinamica e del modo di calcolarle, c'è tutta una serie di considerazioni relative ad aspetti pratici, piuttosto che ad elementi di fisica acustica, e risultati matematici veri propri, che faremo nel Paragrafo 3.7 dopo aver esposto l'ultima fase della digitalizzazione del segnale (la codifica) e i dispositivi di conversione analogico-digitale e digitale-analogico.

3.6.1 Convertitore digitale/analogico DAC

Le pile a stilo di tipo AA in commercio che sono usate normalmente per alimentare le apparecchiature elettriche ed elettroniche, sono tutte a 1,5 volt; la dimensione (ministilo, stilo, torcetta, ecc.) ha a che fare, in linea di massima, con la durata nel tempo prima della scarica totale.



Figura 3.28 Pila stilo tipo AA

Di solito, per funzionare, le apparecchiature hanno bisogno di tensioni di alimentazione superiori a 1,5 volt, per cui si utilizzano più pile disposte *in serie* e cioè con il polo + di una in contatto con il polo - della successiva.

In questo modo la tensione totale ai capi della serie è data dalla somma delle singole tensioni, vale a dire che per un numero N di pile in serie, la tensione ai capi della serie è pari a $N \times 1,5$ volt.

Dipendentemente dalle dimensioni dell'apparecchiatura e dalla forma dell'alloggiamento delle pile, queste possono essere disposte una in coda all'altra oppure l'una parallelamente all'altra ma con senso invertito; in questo secondo caso è una barretta di metallo che realizza il contatto tra il polo + di una e il polo - dell'altra. In entrambi i casi, comunque, il collegamento elettrico è *in serie*.

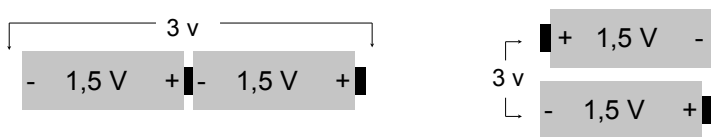


Figura 3.29 Pile in serie

In una torcia elettrica, dove di solito vengono inserite almeno 3 pile per un totale di 4,5 volt, il circuito elettrico per alimentare la lampada attraverso i fili di collegamento, è quello mostrato in Figura 3.30:

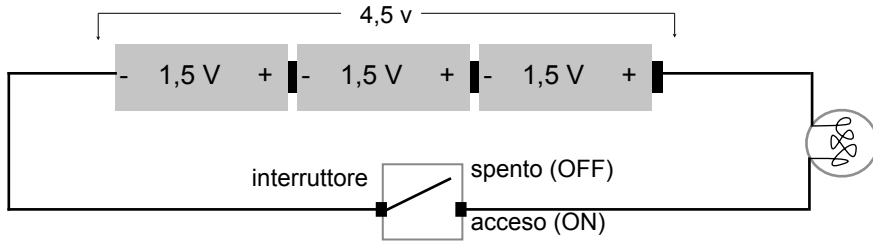


Figura 3.30 Schema elettrico di una torcia

Del circuito fa parte anche l'*interruttore* che è il dispositivo elettro-meccanico per controllare l'accensione della lampada.

Esiste un altro dispositivo elettro-meccanico che svolge una funzione più complessa della semplice interruzione di un circuito: si tratta del *deviatore* che è in grado per l'appunto di deviare il flusso della corrente elettrica su due terminali 0 e 1, e quindi su due diversi fili di collegamento:

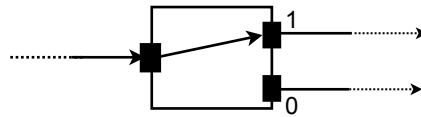


Figura 3.31 Deviatore

Il deviatore è il dispositivo essenziale al funzionamento del convertitore digitale/analogico.

Per semplificarne l'esposizione, si pensi ora a particolari pile (peraltro inesistenti sul mercato) che erogano 1 volt di tensione invece che a quelle reali che erogano 1,5 volt: in questo modo una serie di pile genera una tensione di 2 volt, tre pile 3 volt, quattro pile 4 volt, e così via.

Si consideri allora il seguente schema che realizza un convertitore DAC a 2 bit:

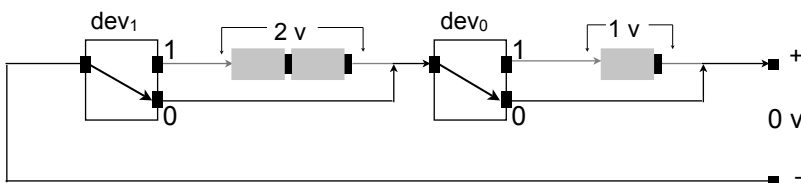
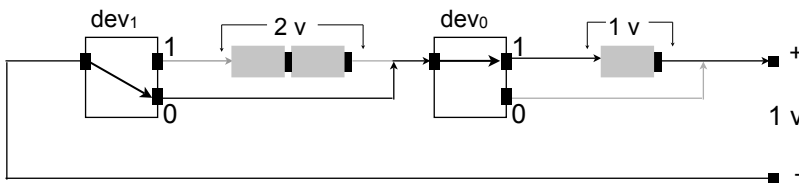


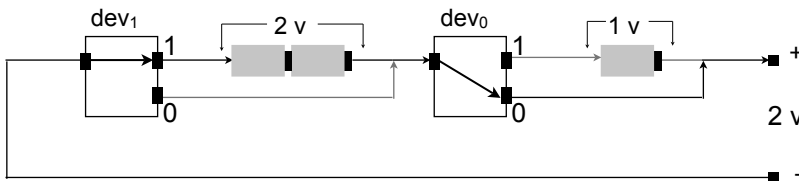
Figura 3.32 Convertitore DAC a 2 bit in configurazione 00

Un deviatore in posizione 0 in pratica esclude dal circuito la pila (o la serie di pile) successive: nel caso della Figura 3.32 entrambi i deviatori sono in posizione 0 e quindi ai terminali + e - del circuito non c'è nessuna tensione, ovvero 0 volt; partendo infatti dal terminale - verso il deviatore 1 fino al terminale + si osserva che dal percorso sono escluse tutte le pile.

Nel successiva situazione (Figura 3.33) il deviatore 0 è sulla posizione 1 e quindi il circuito include la pila di 1 volt: la tensione riscontrata ai terminali è perciò di 1 volt.

**Figura 3.33** Convertitore DAC a 2 bit in configurazione 01

Nel successiva situazione (Figura 3.34) il deviatore 0 è sulla posizione 0 ed il deviatore 1 è sulla posizione 1 che include nel circuito la serie di pile che dà 2 volt: la tensione riscontrata ai terminali è perciò di 2 volt.

**Figura 3.34** Convertitore DAC a 2 bit in configurazione 10

Infine, la situazione (Figura 3.35) in cui entrambi i deviatori sono nella posizione 1, mette in serie tutte e tre le pile per un totale di 3 volt riscontrati ai terminali + e - del circuito.

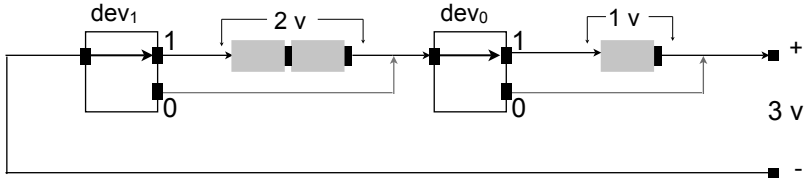


Figura 3.35 Convertitore DAC a 2 bit in configurazione 11

In pratica la tensione ai terminali è determinata dalla combinazione dei due deviatori secondo la seguente tabella nella quale è facile ravvisare la numerazione binaria e la corrispondente decimale.

dev1	dev0	volt
0	0	0
0	1	1
1	0	2
1	1	3

Lo schema di Figura 3.36 mostra un convertitore DAC a 3 bit in configurazione 101 che dà ai terminali del circuito una tensione di 5 volt.

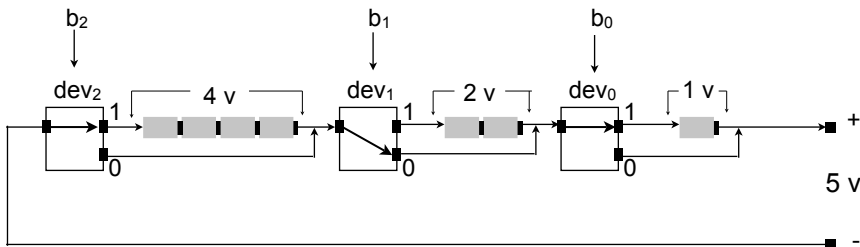


Figura 3.36 Convertitore DAC a 3 bit in configurazione 101

Il lettore verifichi come le 8 combinazioni dei 3 deviatori controllati dai valori binari 0/1 dei bit corrispondenti al numero binario (b2 b1 b0) in ingresso diano in uscita al DAC i corrispondenti otto valori di tensione da 0 a 7 volt secondo seguente la tabella:

b2	b1	b0	volt
----	----	----	------

0	0	0	0
0	0	1	1
0	1	0	2
0	1	1	3
1	0	0	4
1	0	1	5
1	1	0	6
1	1	1	7

Questo è il principio di funzionamento del DAC. Ma i convertitori DAC reali sono realizzati in tutt'altro modo per comprendere il funzionamento dei quali, sono necessarie conoscenze di elettrotecnica ed elettronica che vanno oltre gli scopi di questo libro.

In generale, valgono comunque le seguenti considerazioni.

Il numero di bit utilizzati nella conversione, determina il numero di livelli di tensione che il DAC può produrre, e costituisce una delle principali caratteristiche ad esso associate: la **risoluzione**. Un DAC con 8 bit di risoluzione può produrre $2^8=256$ livelli diversi di tensione e un DAC a 16 bit, $2^{16}=65536$ livelli. E cioè, come si è detto più volte, all'aumentare della risoluzione aumenta la precisione e quindi la qualità del segnale.

Un'altra caratteristica importante associata al DAC è il **tempo di conversione** (*settling time*) ossia il tempo minimo necessario per la stabilizzazione del livello di tensione elettrica che rappresenta l'uscita e perciò l'intervallo di tempo che è necessario lasciare trascorrere prima della presentazione del successivo valore numerico da convertire.

3.6.2 Convertitore analogico/digitale ADC

La tecnica usata per effettuare la conversione è basata sul *confronto* tra il valore "istantaneo" del segnale analogico in ingresso ed un altro valore analogico generato da un DAC interno all'ADC stesso.

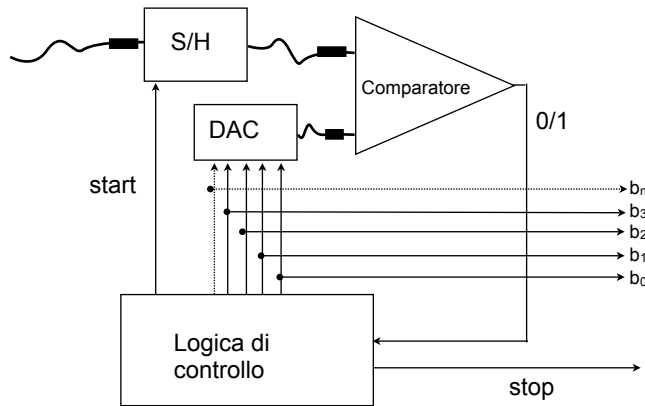


Figura 3.37 Schema funzionale del Convertitore ADC

Gli elementi fondamentali dell'ADC sono 4:

- S/H (*Sample and Hold* = *cattura e trattiene*)
- DAC (interno all'ADC)
- Comparatore
- Logica di controllo.

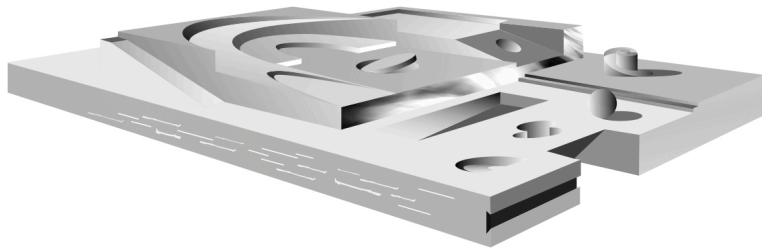
Quando la logica di controllo emette il *comando di start*, il S/H cattura il valore analogico presente all'ingresso; il DAC genera un valore analogico in base al valore numerico generato dalla logica di controllo secondo una certa strategia; il Comparatore esegue il confronto tra il valore analogico trattenuto nell'S/H ed il valore generato dal DAC e risponde in modo affermativo *solo quando* i due valori sono uguali o, meglio, quando differiscono di un valore minore di un intervallo di quantizzazione: in questo caso la logica di controllo segnala all'esterno, attraverso il *comando di stop*, che $b_0b_1b_2b_3..b_n$ è il valore convertito.

Il tempo di conversione dell'ADC dipende soprattutto dall'efficienza dell'algoritmo di ricerca del valore da confrontare con quello in ingresso che generalmente si basa sulla ben nota *ricerca binaria* o *logaritmica*.

Sintesi di segnali sonori

4

- 4.1 Max Mathews
- 4.2 Sintesi di segnali
- 4.3 Oscillatore digitale
- 4.4 Involuppo
- 4.5 Sintesi armonica additiva
- 4.6 Modulazione di Frequenza (FM)
 - 4.6.1 Timbriche mediante FM
 - 4.6.2 FM modificata
- 4.7 Distorsione non lineare (wave-shaping)
- 4.8 Sintesi sottrattiva
- 4.9 Altre tecniche di sintesi
 - 4.9.1 Sintesi per campionamento
 - 4.9.2 Sintesi granulare
 - 4.9.3 Re-sintesi
 - 4.9.4 Algoritmo di Karplus-Strong
- 4.10 Sintesi per modelli fisici



Una delle prime apparecchiature, se non addirittura la prima apparecchiatura, di sintesi sonora elettronica è stato il *Theremin* realizzato da Leon Theremin negli anni '20 del secolo scorso. Questo strumento, che oggi conosce una stagione di revival notevole, è basato su un circuito elettronico oscillante in grado di generare un segnale elettrico analogico essenzialmente di forma sinusoidale con ampiezza e frequenza variabili, controllate con le mani che agiscono a distanza sul campo elettrico di due antenne. Il segnale elettrico generato dall'oscillatore viene poi direttamente amplificato e trasdotto in suono da un altoparlante.

In seguito furono realizzati sintetizzatori analogici più complessi basati su più oscillatori con forme diverse come onde quadre, triangolari e a dente di sega, facilmente realizzabili con elementi della componentistica elettronica analogica di base come resistenze, condensatori, induttanze, tubi elettronici e transistori. Le architetture dei sintetizzatori analogici erano basate su *moduli*, ciascuno dei quali in grado di svolgere una ben precisa funzione, come ad esempio quella di generare un segnale sinusoidale con frequenza e ampiezza variabile oppure filtrare, e cioè modificare, un segnale complesso (rumore bianco, onda quadra, onda triangolare, ecc.) in ingresso e restituirlo in uscita con lo spettro opportunamente variato, come si fa con i controlli di tono dei bassi/medi/alti di un normale amplificatore o di un mixer.

La filosofia di utilizzo di un sintetizzatore analogico consiste principalmente nel concatenare vari moduli collegandone le uscite agli ingressi di altri moduli per mezzo di cavi audio ed intervenendo poi sui controlli manuali di ciascun modulo della catena. Senza entrare nei dettagli del funzionamento dei moduli di generazione e di elaborazione di segnali elettro-acustici, basta qui sottolineare due cose essenziali: 1) la funzionalità svolta da un modulo coincide con la struttura elettronica del modulo stesso; 2) un modulo è in grado di svolgere solo e soltanto quella funzione.

In effetti, la coincidenza della funzionalità con l'architettura stessa di un'apparecchiatura è una caratteristica tipica di tutta l'elettronica analogica: un apparecchio radio, un televisore, un lettore/registratore di cassette audio o video a nastro, un amplificatore, una telecamera di sorveglianza, un mixer, ecc., sono apparecchiature che svolgono una e quella sola funzione per la quale la circuiteria elettronica è stata progettata e realizzata.

Va sottolineato inoltre il fatto, solo apparentemente banale ma invece di sostanziale importanza, che è sufficiente agire sull'interruttore di alimentazione dell'apparecchio per metterlo istantaneamente in funzione: c'è da figurarsi infatti che la sola accensione metta in moto la circolazione degli elettroni nei vari componenti elettronici e... tutto funziona!

Il computer, che in italiano è *calcolatore* (che peraltro è proprio la traduzione della parola inglese *computer*) è nato come macchina di calcolo matematico, ma ben presto si è evoluto come macchina in grado di svolgere più e differenziate funzioni.

Il computer è una macchina *general purpose* e cioè "a scopo generico" che di per sé è in grado soltanto di eseguire calcoli e di spostare informazioni binarie da una parte all'altra della memoria; diventa di volta in volta macchina *special purpose* come un editore di testi, editore di grafici e di fotografie, stazione di posta elettronica, navigatore di rete, macchina di gestione contabile, ecc., in base al programma software installato ed attivato.

La parte elettronica, e cioè l'*hardware*, è sempre la stessa ma è il *software* che *anima* il computer per cui accendere-il-computer di per sé non vuol dire niente, perché è necessario sapere e decidere in quale veste usarlo scegliendo una specifica funzionalità attraverso l'attivazione di un opportuna applicazione.

Tutta un'altra storia rispetto alle apparecchiature analogiche, che sono di fatto macchine *special purpose* dove, per così dire, "il software coincide con l'hardware". In questi termini, in riferimento ai moduli di un sintetizzatore analogico, il software coincide con l'hardware che ne implementa le funzionalità. La lenta ed inarrestabile trasformazione dagli anni '60 agli anni '90 di un mondo tecnologico analogico in quello digitale, ha coinvolto a livello planetario migliaia e migliaia di ingegneri, informatici, fisici, matematici ed ovviamente il Mercato. Ma per quello che ci riguarda, e cioè il passaggio dal sintetizzatore analogico a quello digitale, esiste un preciso momento, un luogo ed un personaggio chiave che vale la pena qui ricordare.

4.2 Sintesi di segnali

Scopo della sintesi è quello di generare artificialmente un segnale sonoro con caratteristiche timbriche desiderate attraverso determinate strategie operative dette, nel caso della tecnologia digitale, *tecniche (o modelli) di sintesi digitale*.

Un modello di sintesi digitale consiste in un programma informatico che produce una sequenza di valori numerici corrispondenti ai valori campionati del segnale audio.

Con l'uso di computer sufficientemente potenti in relazione alla complessità del segnale, la sequenza di campioni generata viene inviata direttamente al DAC, ottenendo suono in tempo reale; altrimenti i valori calcolati vengono accumulati sull'hard-disk ed utilizzati in un secondo momento, ottenendo così suono in tempo differito.

La realizzazione e la scelta di una tecnica di sintesi dipendono dalla sua flessibilità e complessità algoritmica, ossia dalla

- 1) quantità di risorsa di calcolo richiesta;
- 2) varietà delle timbriche ottenibili;
- 3) semplicità d'uso e numero dei parametri di controllo.

Esistono due principali categorie di modelli di sintesi: i modelli di **sintesi additiva** ed i modelli di **sintesi sottrattiva**.

Nelle sintesi additive il segnale sonoro è dato dal risultato della sovrapposizione nel tempo di molti segnali elementari, come è ad esempio quello della sinusoide; nel dominio delle frequenze questa operazione corrisponde all'introduzione di righe spettrali. Le sintesi additive trattate in questo capitolo sono: la **sintesi armonica**, la sintesi per **modulazione di frequenza** e per **distorsione-non-lineare** (o **wave-shaping**).

La sintesi sottrattiva consiste nell'azione combinata della generazione di un segnale che copre gran parte o tutto lo spettro dell'udibile (come il **rumore bianco** che tratteremo in seguito) e dell'attenuazione e/o eliminazione di determinate zone dello spettro.

Gli elementi principali della sintesi sono l'**oscillatore digitale**, il **generatore di inviluppo** ed i **filtri**. L'oscillatore e l'inviluppo sono argomento dei prossimi paragrafi, mentre ai filtri, trattandosi di un argomento articolato e complesso, è dedicato l'intero Capitolo 5.

4.3 Oscillatore digitale

Uno degli elementi fondamentali della musica elettronica analogica era l'oscillatore sinusoidale realizzato per mezzo di una circuiteria elettronica; si trattava quindi di un oggetto fisico che generava direttamente un segnale elettrico analogico. Un **oscillatore digitale** è invece un *oggetto software*, e cioè un programma informatico che calcola i valori numerici relativi all'oscillazione sinusoidale e li trasmette, con una predefinita frequenza di campionamento, ad un convertitore DAC attraverso il quale si produce il segnale analogico, corrispondente a quello generato direttamente da un oscillatore realizzato con circuiteria elettronica analogica.

Poiché l'obiettivo è quello di poter disporre di molti oscillatori per creare timbriche complesse e più "strumenti" contemporanei per eseguire musica polifonica, è necessario eseguire i calcoli relativi ad un singolo oscillatore nel minor tempo possibile. Una soluzione da sempre adottata da fisici, astronomi e ingegneri, soprattutto quando non esistevano le calcolatrici ed i calcolatori elettronici, è stata quella di pre-calcolare una volta per tutte il valore delle funzioni principali come quelle trigonometriche e logaritmiche, e raccoglierle in opportune tabelle. Del resto, anche la "tabellina pitagorica" è un semplice esempio di pre-calcolo delle moltiplicazioni delle unità da "consultare" all'occorrenza.

Una tabella da cui prelevare valori pre-calcolati, viene definita **look-up table**. Supponiamo allora di avere pre-calcolato i valori della funzione trigonometrica *seno* per un intero ciclo suddiviso in 48 parti e cioè con un passo del valore di $2\pi/48=0,13$ radianti, come mostrato in Figura 4.1

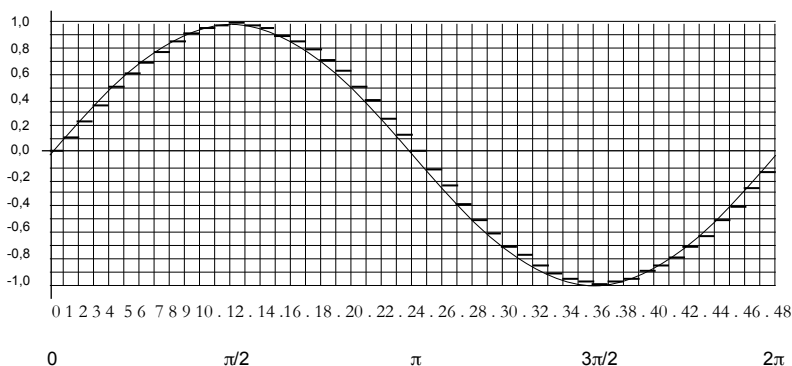


Figura 4.1 Calcolo di un ciclo di senoide

e di averli messi in sequenza in una tabella come mostrato in Figura 4.2a:

Tab	0,00	0,16	0,27	0,34	0,47	0,69	0,76	0,82	0,89	0,92	0,97	...	-0,27	-0,16
i	0	1	2	3	4	5	6	7	8	9	10	...	46	47

Figura 4.2a Tabella relativa al calcolo di un ciclo di senoide

Come si era già visto nel Paragrafo 2.3.1 (pag. 68), è possibile individuare i valori della tabella utilizzando la notazione Tab_i dove i può assumere valori da 0 a 47 per cui, ad esempio, quando $i=3$, Tab_i vale 0,34; quando $i=46$, Tab_i vale -0,27; ecc.

In questo esempio si utilizza una tabella di 48 posizioni (da 0 a 47) in modo che nello spazio della pagina del libro si possa seguire visivamente la descrizione dell'oscillatore digitale per il quale si ipotizza anche una frequenza di campionamento commisurata alla dimensione della tabella, ossia di 48 Hz.

È vero che l'esempio ha poco senso dal punto di vista acustico/musicale, perché con questi valori si ottengono soltanto frequenze bassissime; ma è vero anche che basta moltiplicare per 1000 sia la lunghezza della tabella sia la frequenza di campionamento affinché il tutto abbia senso a livello operativo.

Supponiamo allora di prelevare i valori dalla tabella *in maniera temporizzata, in successione e ciclicamente*: in maniera temporizzata vuol dire ad intervalli di tempo regolari $T=1/F_c$ dove F_c è una predefinita frequenza di campionamento; in successione vuol dire che i valori della tabella Tab_i vengono prelevati con valori crescenti dell'indice i ; "ciclicamente" vuol dire che quando $i=47$, il successivo valore di i è 0. La tabella Tab_i di Figura 4.2a va pensata infatti con struttura "circolare" come in Figura 4.2b.

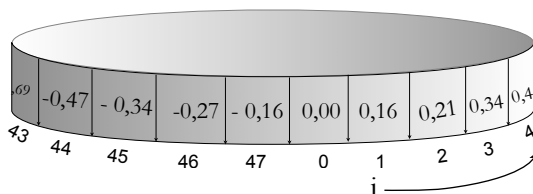


Figura 4.2b Tabella "circolare" relativa al calcolo di un ciclo di senoide

Dal momento che abbiamo ipotizzato una frequenza di campionamento $F_c=48\text{Hz}$, ogni singolo campione viene prelevato dalla tabella e inviato al DAC ogni $T= 1/48$ di secondo, e perciò i 48 valori dell'intera tabella vengono emessi in 1 secondo.

E poiché la scansione dell'intera tabella avviene in 1 secondo, all'uscita del DAC si osserva un'oscillazione sinusoidale con frequenza di 1Hz.

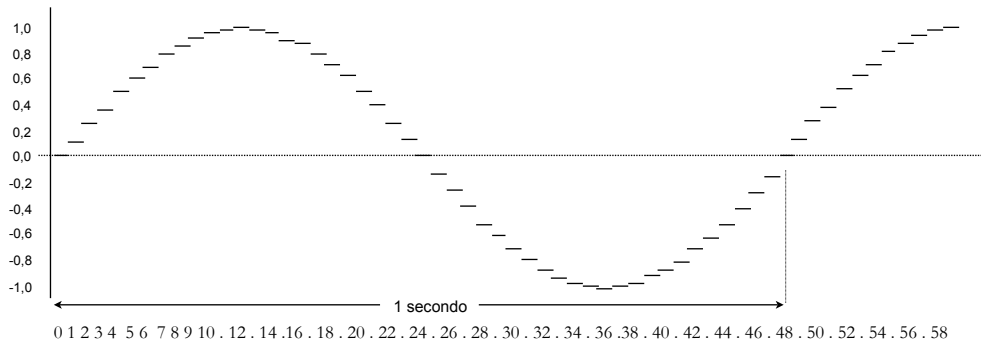


Figura 4.3 Generazione di una sinusoida a frequenza 1 Hz

Si osserva perciò che la frequenza f dell'oscillazione è data dal rapporto fra la frequenza di campionamento F_c e la lunghezza della tabella L_t

$$f = \frac{F_c}{L_t} \quad (4.1)$$

Per variare la frequenza dell'oscillazione, la formula suggerisce due soluzioni: variare la lunghezza di tabella tenendo F_c fissa, oppure variare la frequenza di campionamento tenendo L_t fissa. Ma, anche se formalmente corrette, le due soluzioni non sono utilizzabili a livello operativo: nel primo caso sarebbe necessario un grandissimo numero di tabelle e il secondo caso è impraticabile perché la F_c non può che essere una sola, e cioè quella di riferimento.

La soluzione al problema è altrove.

Il meccanismo dell'oscillatore digitale si basa essenzialmente sull'incremento dell'indice i per mezzo dell'operazione $i_n = i_{cor} + 1$ in base alla quale, ad ogni istante di campionamento e di prelievo del valore dalla tabella con Tab_i , il nuovo valore di i_n è dato dal valore corrente di $i_{cor} + 1$, vale a dire che la scansione della tabella è fatta in successione su tutte le posizioni:

0 1 2 3 4 ... 44 45 46 47 0 1 2 3 4 5...

Ma se l'operazione di incremento dell'indice viene fatto con $i_n=i_{cor}+2$, la scansione della tabella avviene con passo 2, il che corrisponde (partendo con $i=0$) a prendere soltanto i valori 0 2 4 6 8 10... ..44 46 0 2 4 6 8... ..40 42 46 0 2 4... ottenendo questa volta una sinusoide con frequenza di 2 Hz .

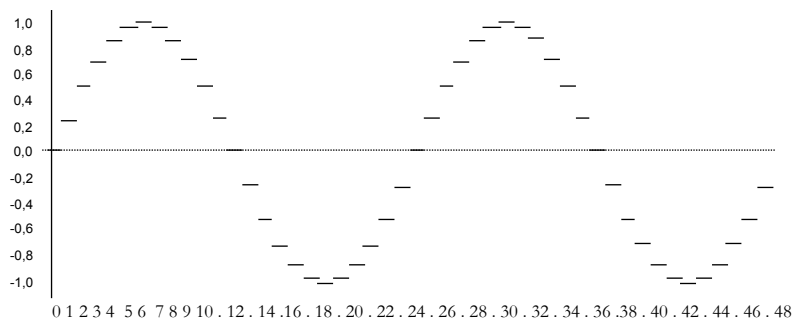


Figura 4.4 Generazione di una sinusoide a frequenza 2 Hz

Infatti con la stessa frequenza di campionamento F_c , e cioè con lo stesso numero di campioni emessi nello stesso tempo di 1 secondo, vengono generati 2 cicli completi di forma sinusoidale. È chiaro che con un incremento di 3 si ottiene una frequenza di 3 Hz e con un incremento di 4 si ottiene una frequenza di 4 Hz e così via, per cui la frequenza dell'oscillazione è data dal rapporto tra frequenza di campionamento F_c e lunghezza della tabella L_t , moltiplicato il passo p corrispondente all'incremento:

$$f = p \frac{F_c}{L_t} \quad (4.2)$$

che indica la diretta proporzionalità tra l'incremento applicato e la frequenza ottenuta. E se si sceglie di fissare F_c uguale a L_t , la formula si semplifica in

$$f = p \quad (4.3)$$

Ma cosa succede se si vuole generare un suono con frequenza espressa da un valore non intero? Ad esempio: sappiamo che la nota LA, secondo lo standard, ha una frequenza di 440Hz per cui, con una tabella di 48.000 valori ed una frequenza di campionamento di 48KHz, applicando la (4.2) o la (4.3), si deve usare un passo di scansione $p=440$.

Ma in base alle considerazioni fatte nel Capitolo 2, il DO temperato tre semitoni sopra il LA ha una frequenza di 523,25 Hz e quindi, con le stesse condizioni, dovremmo usare un passo di 523,25.

L'oscillatore digitale viene di solito rappresentato in forma schematica come in Figura 4.6.

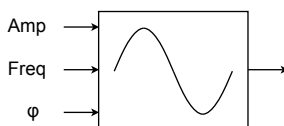


Figura 4.6 Oscillatore digitale

Dal punto di vista acustico/musicale è interessante, poi, la *variazione nel tempo dell'ampiezza* dell'oscillazione che viene controllata con ciò che nel Paragrafo 2.5 abbiamo definito *inviluppo* e che è argomento del prossimo paragrafo.

4.4 Inviluppo

Il termine **inviluppo** deriva dalla forma dell'andamento della linea che raccorda i picchi dell'oscillazione relativa ad un suono durante la sua evoluzione: infatti tale andamento "avviluppa" la forma d'onda stessa. In inglese il termine è *envelope* che significa "busta da lettere", per cui la forma dell'andamento figurativamente "imbusta" e modella la forma d'onda del suono.

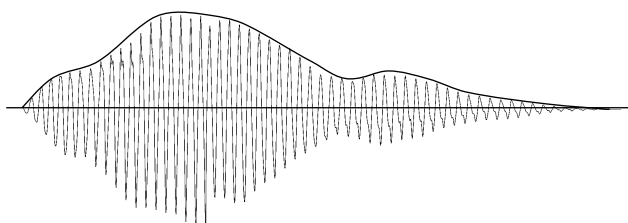


Figura 4.7 Forma d'onda inviluppata

L'inviluppo rappresenta l'evoluzione del volume totale di un suono dal momento dell'inizio dell'emissione fino allo spegnimento, nella quale è possibile individuare (§2.5) tre fasi principali (**attacco**, **sostegno**, **rilascio**). L'inviluppo (e più precisamente l'*inviluppo volumetrico*) è il parametro che più di altri caratterizza un suono: un suono percussivo (tamburo, piatto, campana,

ecc.) ha un attacco brevissimo, non ha sostegno ed ha un rilascio lungo; il suono di uno strumento a fiato ha un attacco lungo, ed un rilascio breve.

Nella sintesi sonora esiste un meccanismo (o funzione) chiamato *generatore di inviluppo* che si può realizzare, al pari dell'oscillatore digitale, con una *look-up table* e cioè con una tabella contenente i valori relativi al profilo dell'inviluppo voluto. In questo caso la tabella viene scandita una sola volta durante la generazione di un suono e la sua attivazione coincide con l'inizio della scansione. In inglese si usa una terminologia da armi da fuoco: l'attivazione della scansione viene detta *trig*, *trigger* (sparo, grilletto della pistola) e la scansione *one-shot*, un colpo solo.

Il generatore di inviluppo, più in generale, viene usato per modellare il contributo di tutti gli elementi (ciascuno indipendentemente dall'altro) che caratterizzano un suono: l'ampiezza di un'armonica (§2.3.1), la quantità di rumore (§4.8), la frequenza di taglio di un filtro (§5.3), ecc. È possibile definire un profilo di inviluppo "per punti" come successione di coppie numeriche *tempo-valore* o direttamente disegnati sullo schermo del computer con il mouse con andamenti tra punto e punto diversi tra loro: lineari, logaritmici, esponenziali, ecc., a seconda del linguaggio musicale usato (Capitolo 7).

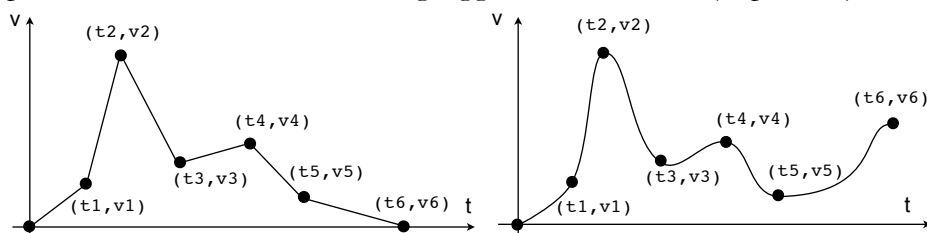


Figura 4.8a Definizione libera di inviluppi

Esiste poi una forma di inviluppo schematica standard nota come ADSR, dalle iniziali dei nomi delle 4 fasi che lo caratterizzano: **attacco**, **decadimento**, **sostegno**, **rilascio** (*attack*, *decay*, *sustain*, *release*), come mostrato in Figura 4.8b.

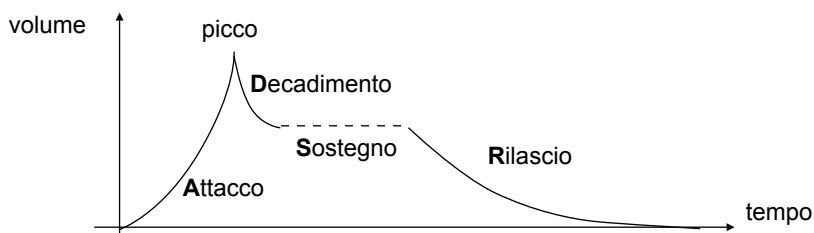


Figura 4.8b Inviluppo ADSR

Per definire un inviluppo ADSR occorrono e bastano 5 valori:

- 1) la *velocità* di Attacco; 2) il *volume* di Picco; 3) la *velocità* di Decadimento;
- 4) il *volume* di Sostegno; 5) la *velocità* di Rilascio.

Una volta definiti il livello di picco e il livello di sostegno, le *velocità* di attacco (0-picco), di decadimento (picco-sostegno) e di rilascio (sostegno-0) definiscono implicitamente anche le *durate* di queste tre fasi. In particolare, la *fase di sostegno* può avere durata variabile per coprire la casistica delle figure musicali. L' ADSR viene schematizzato come in Figura 4.9:

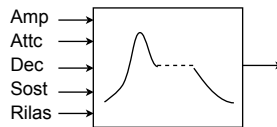


Figura 4.9 Generatore d'involuppo ADSR

Un generatore di involuppo ADSR (Figura 4.9) è più adatto alla definizione di algoritmi di sintesi nell'ambito della musica con note musicali da pentagramma; generatori di involuppi relativi ad involuppi definiti liberamente (Figura 4.8a) sono invece più adatti per composizioni musicali dove anche l'evoluzione del timbro è parte della composizione stessa, come succede nella musica elettronica classica.

4.5 Sintesi armonica additiva

Dal punto di vista teorico la sintesi armonica additiva è data dalla sommatoria di un certo numero di componenti armoniche sinusoidali secondo il seguente schema:

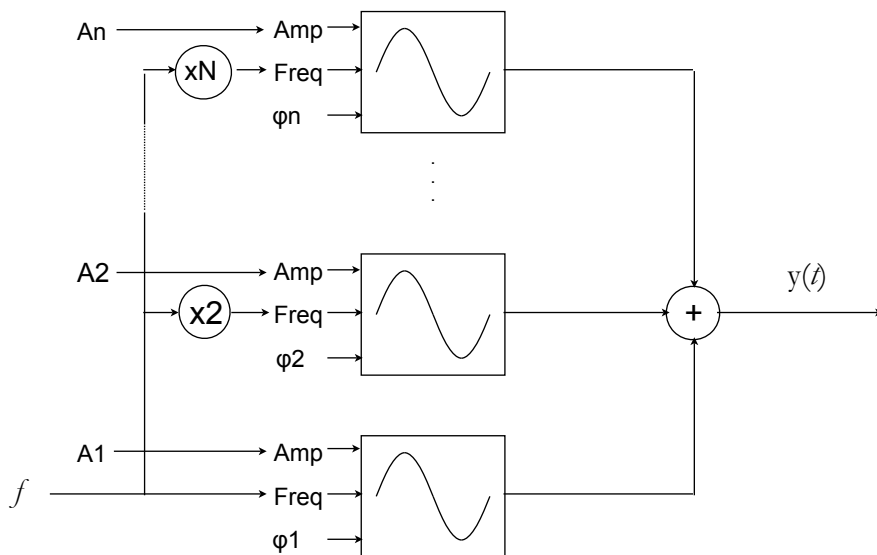


Figura 4.10 Modello della sintesi armonica additiva

che analiticamente si esprime secondo la relazione:

$$y(t) = \sum_{i=1}^{i=n} A_i \cdot \sin(2\pi i f t + \varphi_i)$$

dove $y(t)$ è il segnale sintetizzato, A_i è il valore dell'ampiezza della generica armonica avente fase φ_i e la frequenza multipla della frequenza fondamentale f .

Questa tecnica di sintesi viene spesso detta *Sintesi Armonica di Fourier* per la somiglianza con il noto sviluppo in serie (§2.3); la dizione, che è anche un omaggio a Fourier, è plausibile se riferita soprattutto alle prime esperienze di musica elettronica analogica ed alle tastiere elettroniche degli anni '60 e '70 anche di produzione italiana, come Farfisa ed Elka, dove la modalità di generazione del segnale si basava nella sostanza, vista l'epoca, sull'implementazione elettronica dello schema e della formula appena viste.

Su alcune di queste tastiere era presente un gruppo di potenziometri a slitta, in numero limitato ma significativo, che permettevano di dosare il contributo delle singole armoniche in maniera diretta e molto intuitiva.

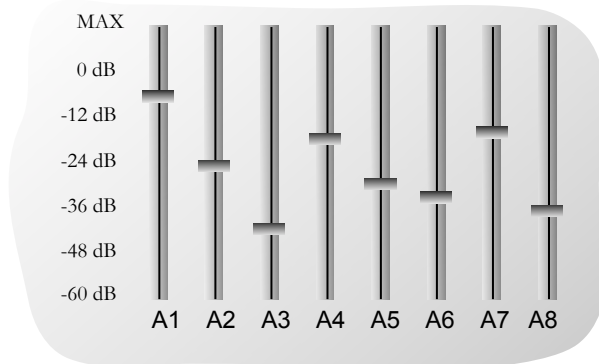


Figura 4.11 Controllo delle armoniche su tastiere analogiche anni '60-'70

Il musicista poteva dunque creare il proprio timbro agendo direttamente sui potenziometri a cursore e se si fosse fatta l'analisi di Fourier sulla forma d'onda risultante visualizzata su un oscilloscopio, la procedura matematica avrebbe prodotto lo spettro delle armoniche con l'esatto riscontro nella disposizione degli stessi potenziometri.

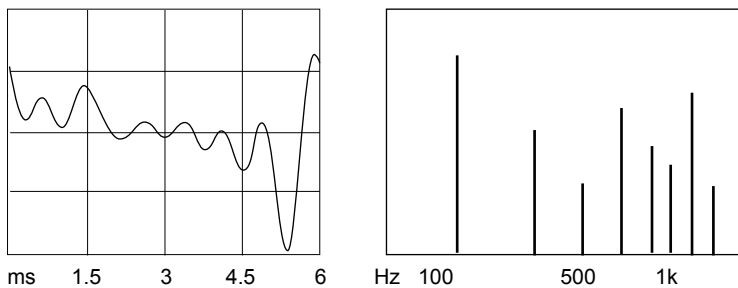


Figura 4.12 Forma d'onda e spettro

Si trattava di sonorità del tutto nuove ed accattivanti che hanno segnato quell'epoca (una canzone fra tante: *A Wither Shade of Pale*) e che tuttavia erano connotate dalla loro "staticità" timbrica e dall'assenza di controllo dinamico di attacco e rilascio, dato che i tasti della tastiera musicale erano semplici interruttori ON/OFF.

La dizione “sintesi armonica di Fourier” è invece decisamente impropria nel caso specifico della sintesi digitale perché permette di introdurre, mediante programmazione software, varianti sostanziali per ottenere maggiore ricchezza timbrica e controllo espressivo dinamico.

Anche se per “sintesi additiva” si intende una categoria di modelli di sintesi, nel caso del digitale per *sintesi additiva* si intende la procedura descritto nel seguito; la dizione è pertinente anche perché gli altri modelli di sintesi additiva (*modulazione di frequenza, distorsione-non-lineare*) hanno il loro nome che ben le caratterizza.

Il modello della sintesi additiva è rappresentato in Figura 4.13.

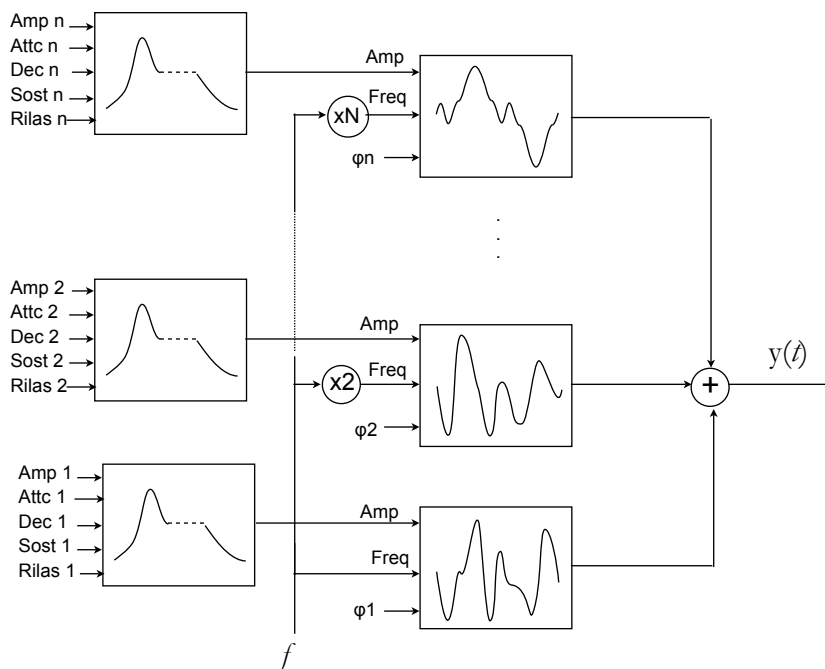


Figura 4.13 Modello della sintesi additiva

Le varianti rispetto al modello della sintesi armonica illustrata in Figura 4.12, sono notevoli e riguardano l'ampiezza, la frequenza e la tabella di riferimento dei singoli oscillatori.

Nel dettaglio:

- 1) l'ampiezza di ogni singola componente può variare in funzione del tempo sotto il controllo di un generatore di inviluppo ADSR per il quale è possibile stabilire: ampiezza massima, tempo di attacco e di decadimento, volume di sostegno e tempo di rilascio;
- 2) ciascun oscillatore può avere un fattore moltiplicativo arbitrario rispetto alla frequenza fondamentale;
- 3) la tabella di riferimento è definibile nei più svariati dei modi.

Utilizzando tale modello, che costituisce il modello più generale di riferimento per la sintesi sonora digitale, possono essere create le timbriche più disparate intervenendo sui numerosi valori parametrici.

L'interesse per questo modello di sintesi negli anni ha avuto “andamento carsico” in relazione all'effettiva usabilità, ed infatti negli ultimi tempi, proprio grazie alla potenza di calcolo a disposizione con gli attuali computer, si sta assistendo ad un rinnovato interesse sull'argomento. Si veda ad esempio [6].

Tuttavia, all'epoca in cui questo modello fu definito, la potenza di calcolo dei computer non era sufficiente per poter ottenere suono in tempo reale.

Vennero fatte molte simulazioni con il linguaggio Music V di Max Mathews, fino a generare sonorità anche di strumenti musicali della tradizione; ma il non poter “suonare” direttamente la potenziale cornucopia di timbriche che il modello consentiva era causa di non poca frustrazione per i musicisti dell'epoca dell'area elettronica.

Ma fu proprio quel senso di strisciante frustrazione che spinse molti ricercatori ad esplorare modalità più “economiche” in termini di risorse di calcolo, ma efficienti sia per la controllabilità, sia per la varietà timbrica ottenibile.

I modelli di sintesi più significativi risalenti agli inizi degli anni '70 sono la **modulazione-di-frequenza**, o **FM**, che ebbe anche un grandissimo successo commerciale, e la sintesi per **distorsione-non-lineare**, o **wave-shaping**, che purtroppo non ha avuto quel successo commerciale che si sarebbe meritato in relazione alla genialità ed all'originalità della metodologia utilizzata.

4.6 Modulazione di Frequenza (FM)

Come si sa (§2.1) un'onda sinusoidale è definita dai tre valori parametrici relativi alla *frequenza*, all'*ampiezza* ed alla *fase*. Mentre la frequenza e l'ampiezza hanno significato specifico e peculiare, la fase ha significato solo se si mettono in relazione tra loro più onde sinusoidali; inoltre, come abbiamo già visto, la fase non ha alcuna rilevanza psico-acustica, perché “sfasare” significa alla fin fine “spostare nel tempo” e “tutto insieme” il risultato dell'oscillazione.

È un po' la stessa differenza che c'è nel “fare una passeggiata solitaria” e “camminare in un corteo”: solo in quest'ultimo caso ha senso dire che si è “avanti” o “indietro” rispetto a qualcun altro.

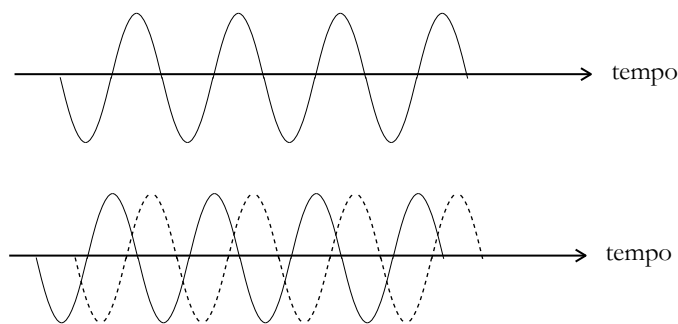


Figura 4.14 Onde sinusoidali sfasate

Ma la fase acquista significato e valore, anche nel caso di una sola oscillazione, se è *variabile nel tempo*. In questo caso l'oscillazione totale che varia nel tempo non è più una semplice senoide ma una forma “modellata” dal modo in cui varia la fase. Si dice per l'appunto che la fase *modula* la frequenza di base in quella che è la *modulazione di frequenza*.

La relazione analitica che descrive un'onda sinusoidale con fase prefissata è come sappiamo $y(t) = A \sin(2\pi ft + \varphi)$ e quella che descrive la modulazione di frequenza è

$$y(t) = A \sin(2\pi ft + \varphi(t)) \quad (4.6)$$

Considerando una generica forma d'onda che varia in funzione del tempo come è $\varphi(t)$ ed un'oscillazione sinusoidale a frequenza fissa, detta **portante** (*carrier*), si ha graficamente la seguente situazione:

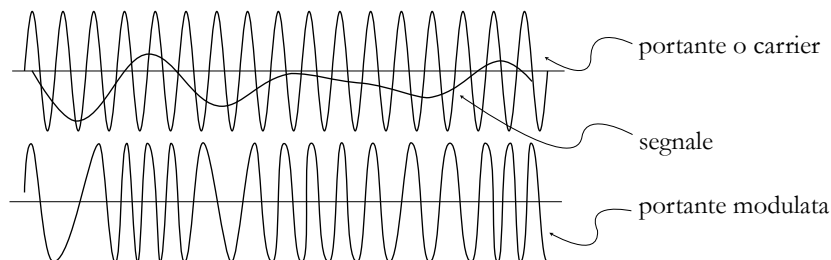


Figura 4.15 Segnale, portante e portante modulata

La Modulazione di Frequenza o *Frequency Modulation* (da cui FM) è in origine la tecnica usata per le trasmissioni radio. Brevettata e messa in opera già negli anni '30, fu adottata dalla RAI a partire dagli anni '50 ed è oggi lo standard utilizzato da tutte le Radio locali a partire dagli anni '70. La modulazione di frequenza ha molti vantaggi rispetto alla precedente Modulazione di Ampiezza (AM), peraltro ancora in uso: non risente dei disturbi causati dall'atmosfera terrestre e consente una banda audio completa (Stereo-HiFi). Secondo lo standard internazionale, l'intervallo di frequenze su cui le varie stazioni radiofoniche trasmettono, va da 87,5 MHz a 108,0 MHz per cui, ad esempio RAI3 trasmette su 94,5 MHz e Isoradio su 103,3 MHz.

Queste sono le frequenze dell'onda sinusoidale *portante* che corrisponde alla frequenza di sintonia di una specifica Stazione Radio, mentre il segnale modulante che costituisce il segnale audio vero e proprio (voce/musica) è ovviamente nella gamma tra 30 Hz e 15.000 Hz. Succede allora che lo spettro del segnale modulato in frequenza si manifesta con una riga in corrispondenza della frequenza portante di sintonia (per esempio 100 MHz) e due bande laterali della larghezza di 15KHz relative al segnale audio, come illustrato in Figura 4.16.

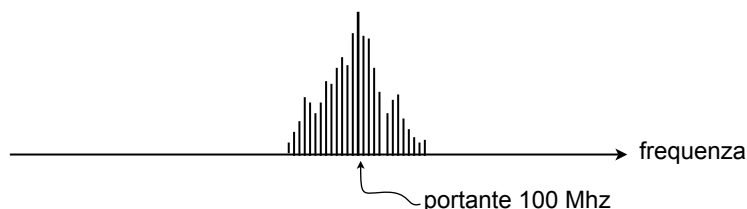


Figura 4.16 Portante e bande laterali nella Modulazione di Frequenza

L'idea di uno spettro ricco di componenti generate da una formula così semplice apparve a qualcuno come una gustosa leccornia nella ricerca di soluzioni economiche in termini di risorse di calcolo per generare timbriche complesse al costo computazione di due soli oscillatori digitali.

Quel qualcuno era John Chowning, giovane musicista diplomato nel 1959 alla Wittenberg University, Ohio, che (come già si faceva con i sintetizzatori analogici) *sostituì* nella formula (4.6) la funzione modulante relativa al segnale audio con una *semplice sinusoidale* utilizzando *entrambe* le frequenze, quella della portante e della modulante, nella *banda audio*.

Sulla FM-audio Chowning condusse un'accurata ed approfondita ricerca analitica e pratica con l'uso del linguaggio Music V, grazie anche alla collaborazione dello stesso Max Mathews.

Una volta verificato che la soluzione era straordinariamente economica, straordinariamente efficace e straordinariamente semplice ed agile da controllare, Max non ebbe dubbi a dire a John "Patent it!", "Brevettala!".

Chowning pubblicò i risultati della sua ricerca nel 1973 in [7] che è, tra l'altro, uno degli articoli più citati nella bibliografia della letteratura relativa alla Computer Music.

Il brevetto dell'FM digitale nello spettro audio (U.S. Patent 4.018.121 - 1975) fu condotto all'interno della Stanford University, California e fu rilasciato nel 1977.

La Stanford University concesse la licenza d'uso della tecnologia FM-audio alla Yamaha Corporation, che nel 1983 mise sul mercato la DX7, la prima tastiera a sintesi digitale basata sulla FM e la prima ad incorporare l'interfaccia MIDI (Capitolo 6).

Il brevetto della sintesi FM-audio ha portato alla Stanford University oltre 20 milioni di dollari, parte dei quali utilizzati per il finanziamento delle ricerche e per la realizzazione alla Stanford University del CCRMA, Center for Computer Research in Music and Acoustics, di cui Chowning è stato Direttore fino alla data del suo pensionamento.

Le sonorità della sintesi FM hanno definito lo stile della pop music degli anni '80, come la sintesi armonica aveva definito quello degli anni '60-'70.

Nella forma più semplice, lo schema della modulazione di frequenza è quello mostrato nelle Figura 4.17:

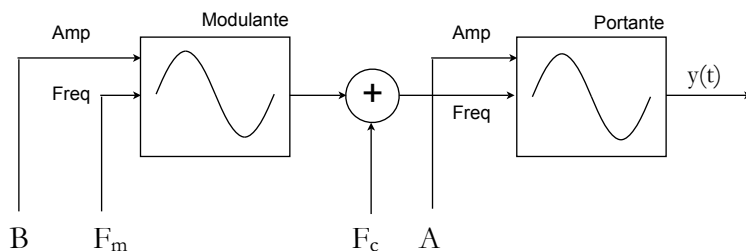


Figura 4.17 Modello della Modulazione di Frequenza

cui corrisponde la formula $y(t) = A \sin(2\pi F_c t + B \sin(2\pi F_m t))$.

Succede allora che il valore in ingresso all'oscillatore della portante è dato dal risultato della somma della quantità costante F_c con quello variabile in maniera sinusoidale proveniente dall'uscita dell'oscillatore della modulante, per cui all'uscita dell'oscillatore portante si riscontra un segnale $y(t)$ con andamento a "mantice di fisarmonica" come illustrato in Figura 4.18.

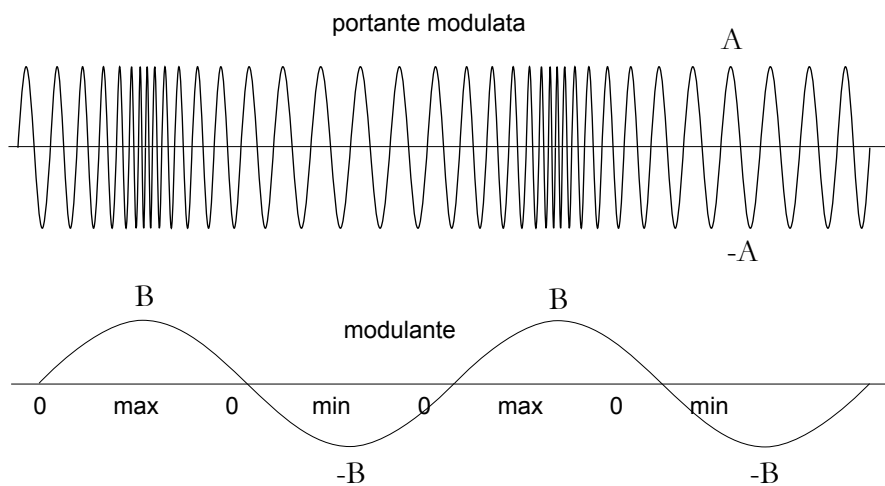


Figura 4.18 Modulazione di Frequenza

Nei punti dove la modulante vale 0, la frequenza della portante è quella originale F_c ; nei punti dove la modulante raggiunge il massimo B , la frequenza della portante è massima e nei punti dove raggiunge il minimo $-B$, la frequenza della portante è minima.

L'ampiezza della modulante determina perciò la *profondità* della variazione, detta *deviazione di picco* della portante; la frequenza della modulante determina invece la frequenza con cui avviene la deviazione di picco.

Come si era già osservato (Figura 4.16), lo spettro di una sinusoide modulata in frequenza è del tipo illustrato in Figura 4.19 dove le righe spettrali compaiono a destra e a sinistra della portante alle frequenze $F_c \pm kF_m$ con $k=1,2,3,4,5....$

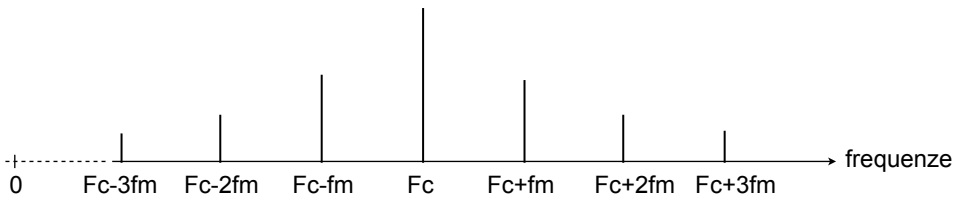


Figura 4.19 Spettro della Modulazione di Frequenza

A questo risultato si arriva in maniera analitica ricorrendo anche a quello che viene detto *sviluppo in serie di Bessel* per cui

$$y(t) = A \sin(2\pi F_c t + B \sin(2\pi F_m t)) \quad \text{diventa}$$

$$\begin{aligned}
 y(t) = & J_0(I) \text{sen}(\omega_c t) \\
 & + J_1(I) \text{sen}((\omega_c + \omega_m)t) + J_1(I) \text{sen}((\omega_c - \omega_m)t) \\
 & + J_2(I) \text{sen}((\omega_c + 2\omega_m)t) + J_2(I) \text{sen}((\omega_c - 2\omega_m)t) \\
 & + J_3(I) \text{sen}((\omega_c + 3\omega_m)t) + J_3(I) \text{sen}((\omega_c - 3\omega_m)t) \\
 & + \dots \\
 & + J_N(I) \text{sen}((\omega_c + N\omega_m)t) + J_N(I) \text{sen}((\omega_c - N\omega_m)t)
 \end{aligned}$$

Lo sviluppo in serie mostra la presenza della portante F_c ridimensionata da un fattore $J_0(I)$ e la presenza di altre frequenze distanziate dalla portante in ragione dei multipli della frequenza modulante con ampiezze stabilite dai coefficienti $J_1(I), J_2(I) \dots J_N(I)$ dati dalle funzioni di *Bessel del primo ordine*.

Sviluppo in serie e funzioni di Bessel (matematico e astronomo tedesco, 1784-1846) sono strumenti matematici ben noti ai fisici, astronomi ed agli ingegneri delle telecomunicazioni. Come si fa di solito a questo punto della trattazione della FM-audio, le funzioni di Bessel vengono introdotte tout-court ed utilizzate per consultazione grafica come si fa con la mappa della rete ferroviaria nazionale e dell'orario dei treni, senza chiedersi chi, quando e come ha progettato e realizzato la rete e stabilito l'orario.

Poiché l'obiettivo è quello di avere una timbrica che rimanga invariata su tutta la gamma delle altezze, come quando si imposta un timbro su una tastiera elettronica, è necessario modificare il modello di Figura 4.17 e la formula corrispondente, in modo da dare il valore alla frequenza e all'ampiezza dell'oscillatore modulante.

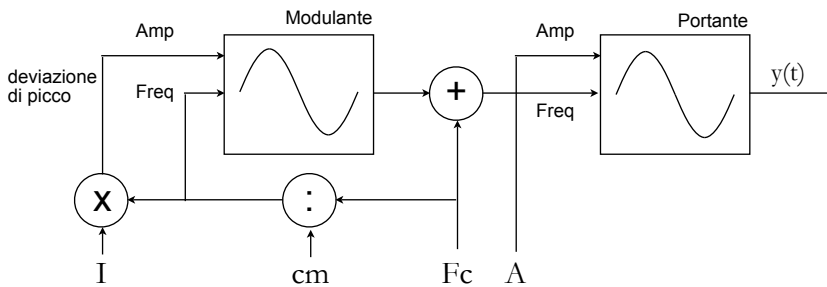


Figura 4.20 FM con Indice di modulazione e rapporto portante/modulante

La frequenza della modulante F_m viene impostata indirettamente attraverso un valore che esprime il “rapporto” con F_c e cioè $F_{carrier}/F_{modulant}$ indicato con cm : se ad esempio $cm=4$, ad una portante $F_c=600$ Hz corrisponde una $F_m=150$ Hz e le frequenze laterali sono a (450,750), (300,900) e (150,1050).

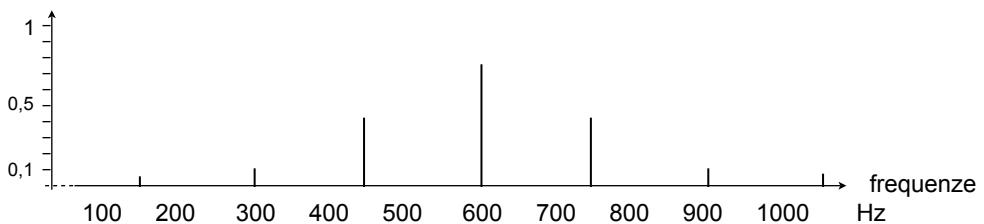


Figura 4.21 Spettro di FM con $F_c=600$ Hz e $cm=4$

Con lo stesso rapporto $cm=4$, ad $F_c=400$ Hz corrisponde una $F_m=100$ Hz, con frequenze laterali (300,500), (200,600), (100,700)....

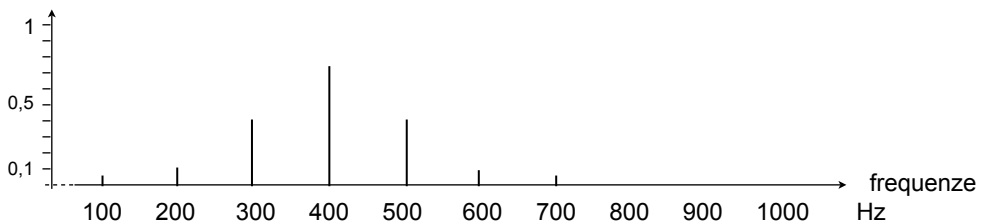


Figura 4.22 Spettro di FM con $F_c=400$ Hz e $cm=4$

L'altra modifica riguarda l'ampiezza della modulante.

Nello schema di Figura 4.17 e nella corrispondente formula, l'ampiezza della modulante è stabilita con il parametro B , e siccome la deviazione di picco della portante dipende dall'ampiezza della modulante, sempre nell'ottica di mantenere proporzionalmente invariato lo spettro alle diverse frequenze di pitch, è necessario poter scalare anche l'ampiezza B della modulante in funzione del pitch.

Il parametro che controlla l'ampiezza B diventa allora la *deviazione di picco* D_p che dipende dalla frequenza della portante F_c , dal rapporto $cm = F_c/F_m$ e da un *indice di modulazione* I , che è un parametro delle già annunciate funzioni di Bessel, secondo le relazioni

$$cm = F_c/F_m \quad (4.7.1)$$

$$F_m = F_c/cm \quad (4.7.2)$$

$$D_p = I \times F_m = I \times F_c/cm \quad (4.7.3)$$

La formula relativa allo schema di figura 4.19 è allora

$$y(t) = A \sin(2\pi F_c t + D_p \sin(2\pi F_m t)) \quad (4.8)$$

per cui, riassumendo:

-a) i parametri di controllo della sintesi FM sono 4:

- 1) l'ampiezza A della portante
- 2) la frequenza F_c della portante
- 3) il rapporto cm portante/modulante
- 4) l'indice di modulazione I

-b) cm stabilisce *dove* appaiono le righe spettrali

-c) I stabilisce *quanto* valgono le ampiezze delle righe spettrali.

Le funzioni di Bessel sono quelle rappresentate in Figura 4.23 che ne riporta soltanto le prime 5 (da J_0 a J_4) delle infinite possibili.

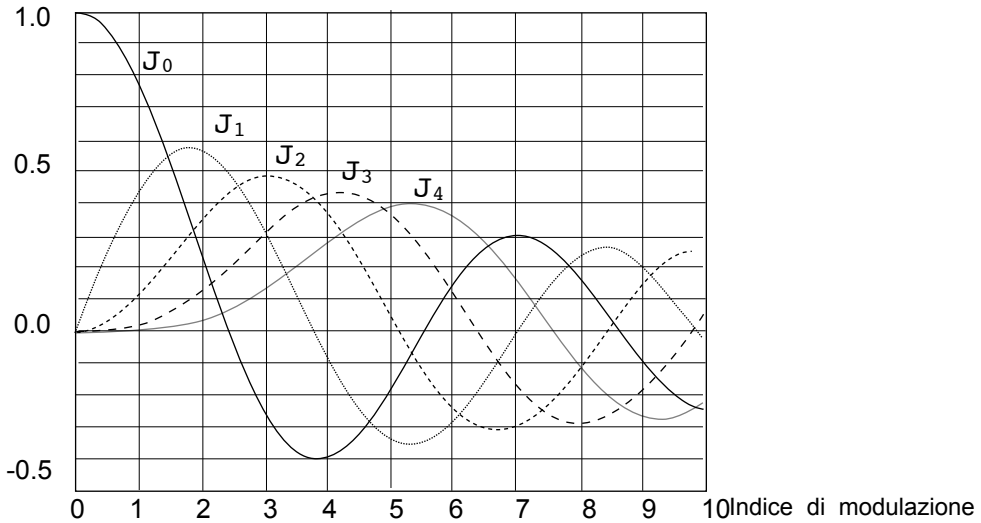


Figura 4.23 Funzioni di Bessel del primo ordine

Possiamo ora conoscere i valori delle ampiezze delle componenti spettrali leggendo graficamente i valori delle funzioni di Bessel in corrispondenza dell'indice di modulazione I .

Per $I=0$, la prima funzione di Bessel J_0 vale 1 e tutte le altre valgono 0; questo corrisponde al fatto che quando $I=0$ anche l'uscita dell'oscillatore modulante vale 0 e quindi all'uscita dell'oscillatore portante c'è la sola sinusoide di frequenza F_c .

Per $I=1$, $J_0=0,77$ $J_1=0,44$ $J_2=0,11$ $J_3=0,02$ $J_4=0,0$ che produce lo spettro di Figura 4.21 con $cm=4$ e $F_c=600$ e quello di Figura 4.22 con $cm=4$ e $F_c=400$.

Per $I=2$, si ha: $J_0=0,22$ $J_1=0,58$ $J_2=0,35$ $J_3=0,13$ $J_4=0,03$.

Per $I=3$ la componente J_0 assume un valore negativo e così anche le altre funzioni all'aumentare dell'indice di modulazione; il significato fisico/acustico di una componente negativa è quello di una componente dello stesso valore positivo sfasato di π radianti, pari a 180° , per cui le funzioni di Bessel devono essere re-interpretate con valori sempre positivi come mostrato in Figura 4.24:

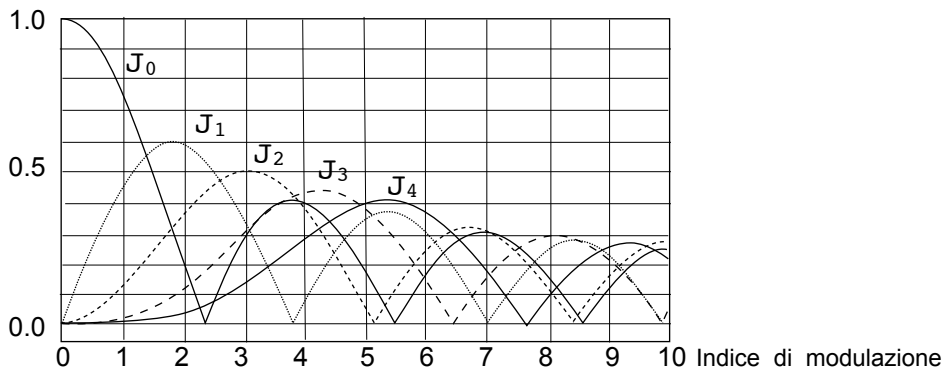


Figura 4.24 Funzioni di Bessel re-interpretate

L'indice di modulazione può variare in maniera *continua* da 0 a 10 ...ed oltre e lo si può far variare in maniera *dinamica* sotto il controllo di un involuppo per il quale si deve stabilire un massimo da raggiungere durante la sua l'evoluzione.

Lo schema completo ed operativo per la sintesi a Modulazione di Frequenza è allora quello in Figura 4.25:

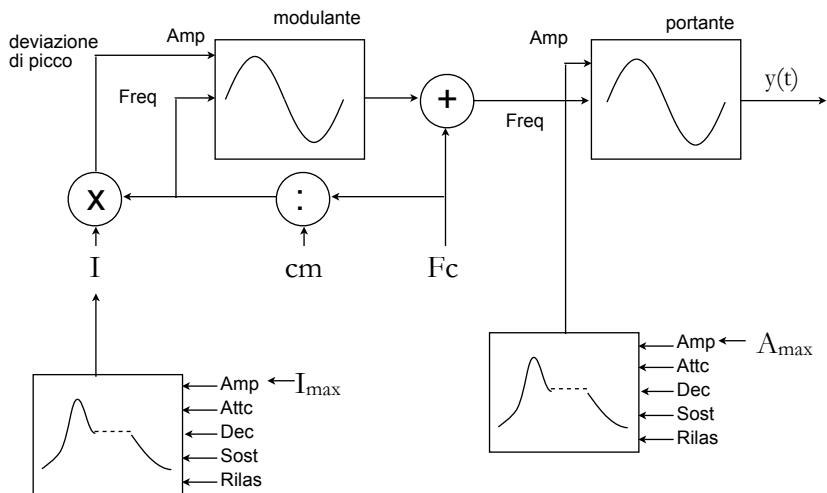


Figura 4.25 Modello completo della Modulazione di Frequenza

Nel prossimo paragrafo vengono date indicazioni di massima per ottenere diverse tipologie di timbriche al variare del rapporto portante/modulante cm e dell'indice di modulazione I .

4.8 Sintesi sottrattiva

Questa tecnica di sintesi è basata sull'eliminazione parziale o totale di alcune componenti o di intere zone frequenziali di un segnale sorgente caratterizzato da uno spettro molto ricco, con l'uso di appositi dispositivi, detti *filtri*.

In maniera figurata la sintesi sottrattiva ricorda l'azione di uno scultore che partendo da un blocco di marmo informe "sottrae" materiale per definire la forma desiderata. In questo caso il "blocco-di-marmo-informe" è di solito il segnale detto *rumore bianco*.

In Acustica viene definito "rumore" una sonorità per la quale non è definita né è percepibile un'altezza precisa (pitch) come invece è per le note musicali, e coincide con il significato usuale che si ha di rumore: il fruscio delle foglie, una porta che sbatte, il mare in tempesta, un applauso a fine concerto, il vento, ecc. L'aggettivo "bianco" viene utilizzato per analogia al concetto di "luce bianca".

Si è già detto che spesso vengono usate stesse terminologie nell'Acustica e nell'Ottica, come nel caso di "frequenza" e "ottava", anche se i due fenomeni hanno valenze e "principi di funzionamento" del tutto diversi: il suono è dato dalla diretta percezione di un agente che mette in azione le particelle dell'aria, mentre la visione delle cose del mondo circostante è data dall'effetto di "rimbalzo" verso i nostri occhi di una o più sorgenti di *fotoni*.

In maniera diretta percepiamo solo lucciole, lanterne, lampade, il Sole e le stelle perché - anche se è banale dirlo - noi vediamo le cose, le piante, la Luna, gli animali, noi stessi e gli altri, soltanto se illuminate e cioè se investite da fotoni emessi da una qualche sorgente luminosa come è il Sole.

Il Sole emette radiazioni elettro-magnetiche su un grandissimo intervallo di frequenze, dai GigaHz delle onde radio alle centinaia di milioni di milioni di milioni di Hertz (10^{20} Hz) dei raggi X e dei raggi gamma.

I nostri occhi sono sensibili alle radiazioni che hanno frequenza nell'intervallo 430÷750 TeraHertz (dove 1 Tera=1.000.000.000.000= 10^{12}) che sono, per quello che ci riguarda, le radiazioni luminose... la "luce".

Le radiazioni luminose arrivano a noi in un totale mélange frequenziale e si separano a ventaglio solo in rare occasioni sotto forma di arcobaleno o molto più banalmente quando una porta a vetri di un negozio viene investita dal Sole con un'angolazione tale da proiettare una striscia colorata chiamata *spettro solare* che va con continuità dal rosso al giallo al verde all'azzurro al blu al viola.

Si tratta del ben noto "effetto prisma" che può essere considerato un'allegoria dei concetti di segnale e spettro, filo conduttore di questi capitoli.

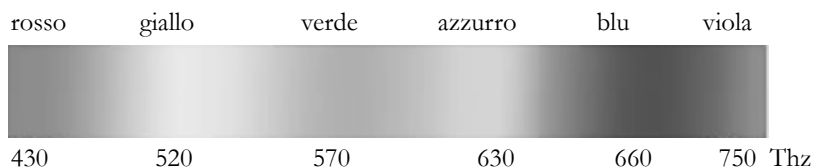


Figura 4.36 Spettro della Luce

Dunque, un oggetto appare bianco perché *riflette tutto* lo spettro solare mischiato e un oggetto appare nero perché *assorbe tutto* lo spettro solare e non riflette niente; un oggetto appare colorato perché ha la proprietà di assorbire tutte le componenti dello spettro della luce bianca a meno di una zona ristretta che viene riflessa.

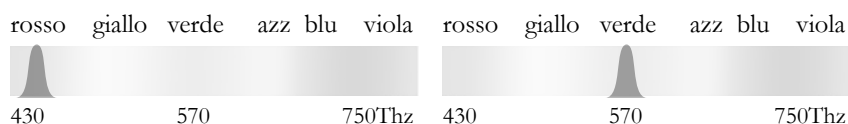


Figura 4.37 Colore come filtro

I petali dei papaveri appaiono rossi perché riflettono soltanto le radiazioni dello spettro relative al rosso ed assorbono tutte le altre; le foglie appaiono verdi perché riflettono soltanto le radiazioni dello spettro relative al verde ed assorbono tutte le altre.

Quindi, quello che noi chiamiamo “colore-delle-cose” è ciò che risulta dall’azione “sottrattiva” e cioè di mascheramento della gran parte dello spettro del visibile della luce bianca. In termini più propri, la maschera si chiama “filtro” ed il concetto coincide con quello che si ha nel linguaggio comune di filtro: scelta selettiva secondo certi criteri da un miscuglio di elementi dello stesso genere ma con caratteristiche diverse (dimensioni, peso, frequenza, ecc.).

Per analogia, dunque, il “rumore bianco” è dato dal miscuglio casuale di suoni su tutta la gamma dell’udibile ed è ragionevole affermare che l’effetto acustico del rumore bianco sia più o meno quello del vento perché nel vento le particelle dell’aria oscillano in maniera caotica su tutte le frequenze.

Il rumore bianco ha perciò una rappresentazione spettrale come quella mostra in Figura 4.38.

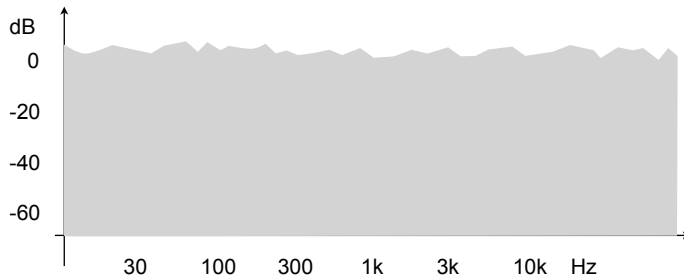


Figura 4.38 Spettro del rumore bianco

Nella musica elettronica analogica il rumore bianco è stato, insieme alla sinusoide, uno degli elementi base di generazione sonora, per lo più associato alla modulazione ed alla trasformazione attraverso l'uso dei filtri; in quest'ambito, generatore di rumore e filtri sono realizzati mediante opportune circuiterie elettroniche atte allo scopo.

Nella musica digitale il generatore di rumore è invece un algoritmo che genera numeri casuali secondo strategie ben note in letteratura e comunque esistente come funzione di base in un qualunque linguaggio di programmazione, mentre i filtri sono ben precise computazioni matematiche alla definizione delle quali si arriva dopo un percorso piuttosto complesso che sarà l'argomento del prossimo capitolo.

L'azione di un filtro, detta **risposta in frequenza** o più semplicemente *risposta*, è quella di prendere un segnale $x(t)$ in ingresso cui corrisponde lo spettro $X(f)$, di elaborarlo tramite eliminazione parziale o totale di alcune zone dello spettro e di restituirlo in uscita come segnale $y(t)$ cui corrisponde lo spettro $Y(f)$.



Figura 4.39 Azione di un filtro

- Esistono 4 tipi principali di filtri:
- 1) **passa-basso** (*low-pass*);
 - 2) **passa-alto** (*high-pass*);
 - 3) **passa-banda** (*band-pass*);
 - 4) **reiezione di banda** (*band-reject*).

Ai filtri passa-basso e passa-alto è associato un unico parametro di controllo detto **frequenza di taglio**:

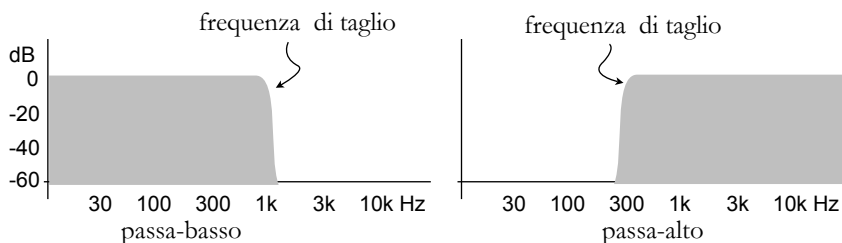


Figura 4.40 Risposta in frequenza dei filtri passa-basso e passa-alto

Il filtro *passa-basso* lascia passare le componenti frequenziali al di sotto della frequenza di taglio scelta e sopprime quelle al di sopra; il filtro *passa-alto* lascia passare le componenti frequenziali al di sopra della frequenza di taglio scelta e sopprime quelle al di sotto.

I parametri di controllo dei filtri passa-banda e reiezione-di-banda sono il **centro-banda** e la **larghezza-di-banda**:

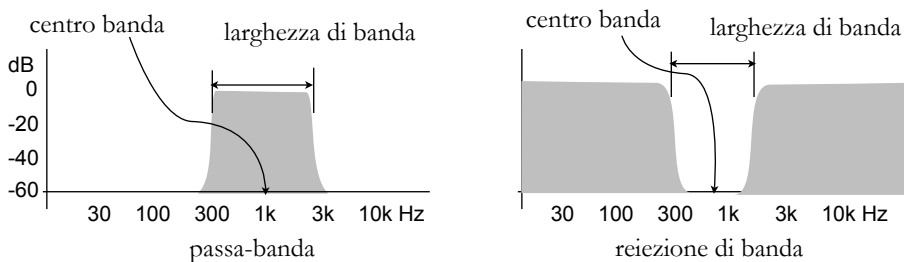


Figura 4.41 Risposta in frequenza dei filtri passa-banda e reiezione-di-banda

Il filtro *passa-banda* lascia passare le componenti dell'intervallo definito dal *centro-banda* e dalla *larghezza-di-banda* e sopprime tutte le altre; il filtro *reiezione-di-banda* sopprime le componenti dell'intervallo definito dal *centro-banda* e dalla *larghezza-di-banda* e fa passare tutte le altre. I parametri di controllo *centro-banda* e *larghezza-di-banda* sono indipendenti tra loro per cui è possibile scegliere e definire con molta precisione la zona dello spettro da eliminare.

Di filtri digitali si parlerà in maniera più esauriente nel prossimo capitolo.

I generatori di segnali ad ampio spettro sono detti anche **sorgenti di eccitazione** e sono: il già menzionato **rumore bianco**, il **rumore rosa**, il **generatore di impulsi**, l'**onda triangolare** e il **dente di sega**.

Il *rumore bianco* è caratterizzato da uno spettro di ampiezza costante su tutta la gamma dell'udibile.

Il *rumore rosa* è dato dal rumore bianco elaborato da un sistema di filtri che è caratterizzato da una sonorità di "soffio" più morbida di quella secca e tagliente del rumore bianco, ed ha caratteristiche spettrali meglio specificate nel prossimo capitolo (Figura 5.21).

L'*impulso* è un segnale periodico *asimmetrico* che si ottiene variando la percentuale delle durata tra la parte positiva e quella negativa dell'*onda quadra*: tale percentuale rispetto ad un intero periodo è detta *duty cycle*: (con *duty cycle*=50%, l'impulso corrisponde all'onda quadra).

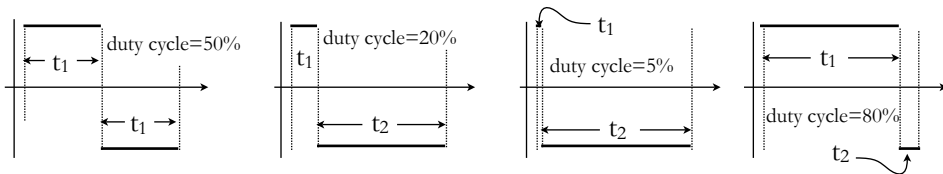


Figura 4.42 Onda quadra e impulso con duty cycle diversi

Anche l'*onda triangolare* (Figura 4.43) è un segnale periodico simmetrico.

Il *dente di sega* con rampa ascendente o discendente è un segnale periodico *asimmetrico* che si ottiene spostando il picco di un'onda triangolare rispettivamente a destra o a sinistra.

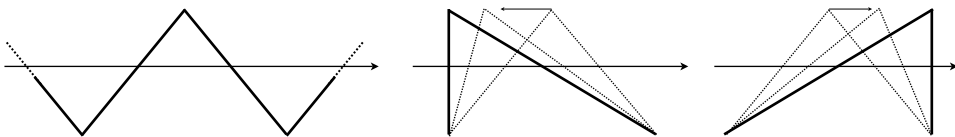


Figura 4.43 Onda triangolare e dente di sega

È possibile infine usare come fonte di eccitazione un qualsiasi campione sonoro di qualsiasi natura catturato da vivo; ed è comunque possibile fare una qualsiasi miscela additiva di più fonti di eccitazione. In ogni caso si fa sempre riferimento a queste sorgenti di eccitazione perché la computazione algoritmica (o la circuiteria elettronica se si è in ambito analogico) per generare questi segnali è decisamente semplice a fronte di una ricchezza di contenuto spettrale.

Nella sintesi sottrattiva il segnale generato da una sorgente di eccitazione passa attraverso un insieme di filtri, più propriamente detto *banco di filtri*, scelti tra le quattro tipologie appena viste, che ne modella le caratteristiche spettrali e quindi timbriche.

Affinché il segnale finale abbia spettro dinamico, le caratteristiche dei filtri (frequenza di taglio, larghezza di banda, ecc.) sono controllate da uno o più generatori di involuppo a seconda del tipo di filtro.

Esistono sostanzialmente due configurazioni di banco di filtri: in *serie* (o in *cascata*) ed in *parallelo*, come mostrato nella Figura 4.44:

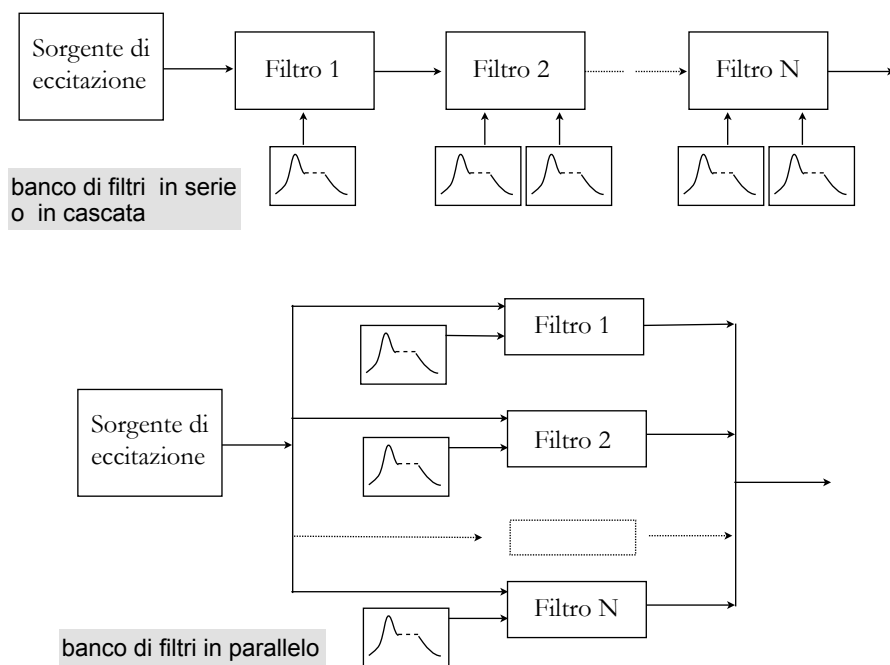


Figura 4.44 Banchi di filtri in serie e in parallelo

È poi possibile assemblare banchi di filtri più complessi riconducibili comunque sempre alla forma generale di "parallelo-di-serie" come illustrato in Figura 4.45:

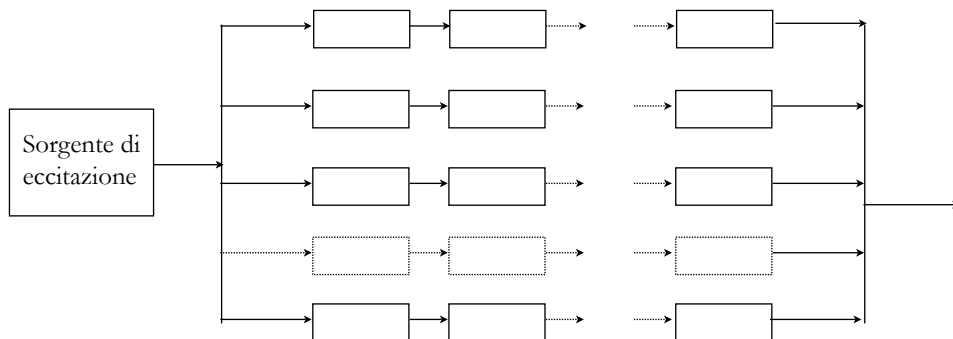


Figura 4.45 Banco di filtri in configurazione parallelo-di-serie

Ciascuno dei filtri del banco, può essere uno dei quattro tra i tipi descritti: passa-alto, passa-basso, passa-banda e reiezione-di-banda; a seconda del tipo, il filtro viene controllato dinamicamente da uno o da due generatori di involuppo come mostrato nei seguenti schemi di riferimento:

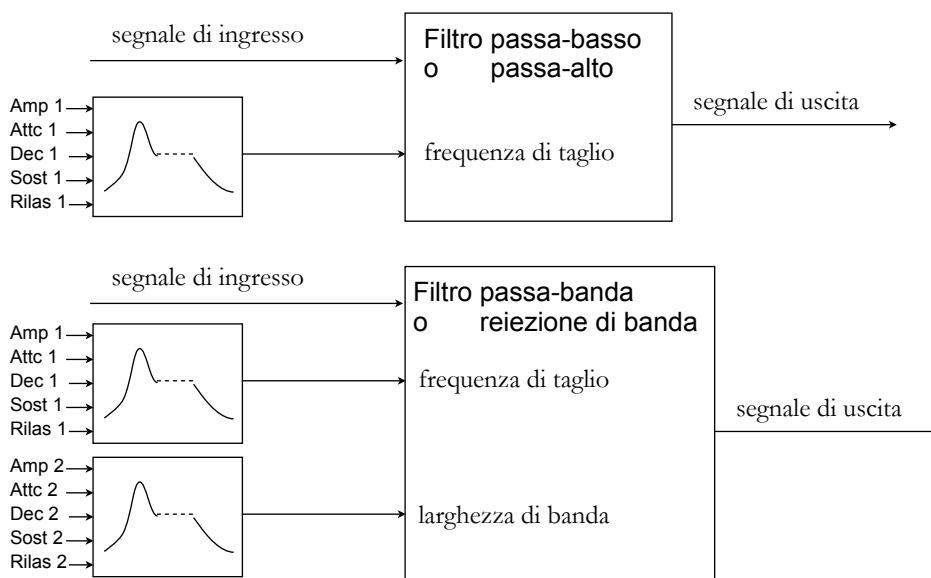


Figura 4.46 Controllo dinamico di un filtro

La sintesi sottrattiva è dunque una tecnica che dà grande libertà d'azione, sia nella definizione di una particolare configurazione eccitatore-filtri, sia per la presenza di così tanti valori parametrici di controllo.

Un'applicazione significativa di questa tecnica è data dalla *sintesi vocale*, perché gli elementi in gioco appena visti simulano e corrispondono abbastanza bene agli "elementi" funzionali dell'apparato fonatorio umano:

- il generatore di impulsi corrisponde all'azione della *glottide* e cioè delle corde vocali che generano suoni di altezza frequenziale riconoscibile;
- particolari posizioni di lingua e denti emettono rumore bianco e/o rosa quando si pronuncia la "s" di "salato" o "sc" di "sciocco";
- la gola, il naso ed in special modo la bocca con il movimento mascellare, creano un vera e propria cavità risonante dinamica che attenua e/o enfatizza differenti zone dello spettro come fanno i filtri.

Lo spettro della voce umana, in particolar modo nell'emissione delle vocali, ha un andamento a "colline-e-valli" più propriamente detti *formanti*:

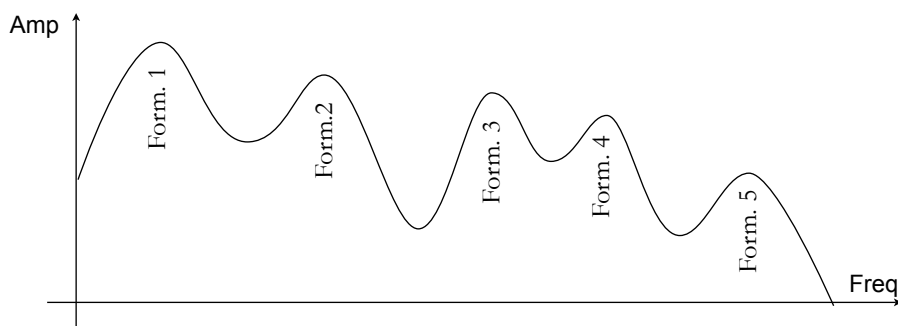


Figura 4.47 Spettro della voce umana a *formanti*

Questo risultato si raggiunge utilizzando, in prima approssimazione, sorgenti di eccitazione ad impulsi e a rumore bianco/rosa elaborate da un banco filtri passa-banda in parallelo (Figura 4.45).

Vale comunque la pena ricordare che esistono in letteratura modi alternativi per generare formanti, tra cui il più noto è il FOF, *Function d'Onde Formantique* sviluppato negli anni '80 da Xavier Rodet all'IRCAM di Parigi ed il VOSIM sviluppato da Tempellars e Kaegi negli anni '70 all'Istituto di Sonologia di Utrecht (Olanda).

4.10 Sintesi per modelli fisici

Nel primo capitolo si è visto (§1.2, Figura 1.2) che l'andamento della proiezione a terra dell'oscillazione del pendolo, visualizzata dalla sabbia che scende dal secchiello forato, è sinusoidale; anche un oggetto appeso ad una molla oscilla con moto sinusoidale:

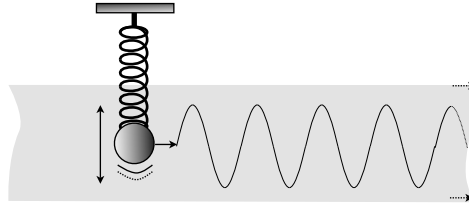


Figura 4.57 Oscillazione verticale sinusoidale

Sinusoidale è anche il moto di un oggetto che scivola su un piano connesso orizzontalmente ad una molla.

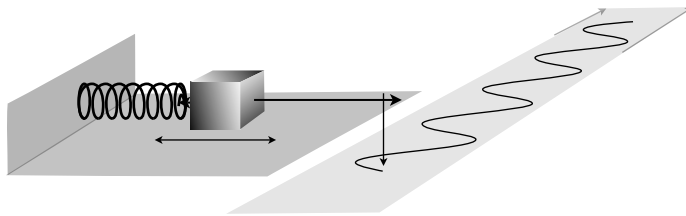


Figura 4.58 Oscillazione orizzontale sinusoidale

Tuttavia, come sappiamo dall'esperienza, l'energia introdotta inizialmente con la compressione o l'estensione della molla o lo spostamento laterale del pendolo si dissipa gradualmente, per cui la forma dell'andamento in tutte e tre le situazioni è nella realtà quella di una sinusoide smorzata:

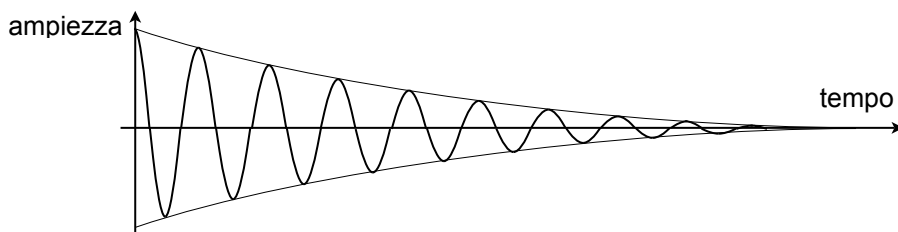


Figura 4.59 Oscillazione sinusoidale smorzata

Con gli elementi “oscillatore” ed “inviluppo” utilizzati nella descrizione delle tecniche di sintesi come la FM e la Wave-shaping, possiamo realizzare un generatore di oscillazione smorzata in questo modo:

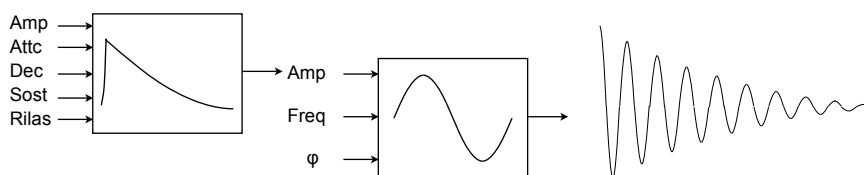


Figura 4.60 Generatore di oscillazione sinusoidale smorzata

Intervenendo sui parametri che controllano la frequenza dell'oscillatore, l'ampiezza massima, l'attacco ed il tempo di rilascio dell'inviluppo è possibile ottenere l'andamento tipico dei tre sistemi meccanici considerati. Tuttavia, si deve osservare che tra i valori parametrici relativi a frequenza-ampiezza-rilascio e la struttura meccanica di uno qualunque dei tre sistemi non c'è alcuna relazione; pertanto non è possibile risalire ad uno dei tre sistemi meccanici esaminando soltanto il grafico dell'oscillazione smorzata.

Vale infatti la seguente considerazione: tutti i modelli di sintesi esaminati fin qui seguono la prassi della riproduzione del segnale relativo al timbro di uno strumento a partire dall'analisi del segnale sonoro opportunamente digitalizzato e, cosa più importante, a prescindere dalla struttura fisica di quello strumento. Infatti, le tecniche di sintesi esaminate nei precedenti paragrafi sono *modelli matematici* implementati mediante algoritmi che generano segnale sotto il controllo di parametri che non hanno niente a che fare con le caratteristiche meccanico/fisico/acustiche di un eventuale strumento musicale reale di riferimento. Questi modelli vengono detti anche *modelli del ricevente*.

Un *modello fisico*, invece, focalizza l'attenzione sul comportamento meccanico/acustico di uno strumento musicale reale attraverso l'esame delle caratteristiche fisiche dei singoli elementi e delle loro interdipendenze all'interno dell'architettura dello strumento stesso. I modelli fisici vengo detti anche *modelli della sorgente*.

Naturalmente anche un modello fisico è di fatto una formulazione matematica che ne descrive il comportamento dinamico; tale formulazione presuppone nozioni di fisica e di matematica, oggetto di percorsi di studio superiori che implicano competenze ed interessi specialistici. Per questo motivo non ci addentreremo su questo terreno, ma faremo una serie di considerazioni sufficientemente esaurienti per comprendere la filosofia di lavoro e di utilizzo dei modelli fisici per la sintesi.

Dalla Fisica si sa che la frequenza di oscillazione del pendolo dipende dall'*accelerazione di gravità* e dalla *lunghezza* del pendolo stesso (più precisamente, la lunghezza è quella tra il fulcro e il *centro di massa* che dipende dalla *massa* appesa). L'*ampiezza* dell'oscillazione (iniziale) dipende dall'*elongazione iniziale*, e lo smorzamento dipende dall'*attrito* che si esercita sul fulcro. Simulare la meccanica del pendolo attraverso formulazioni matematiche significa allora avere un modello fisico i cui parametri di controllo sono: *lunghezza*, *massa*, *accelerazione di gravità*, *elongazione iniziale* e *attrito*.

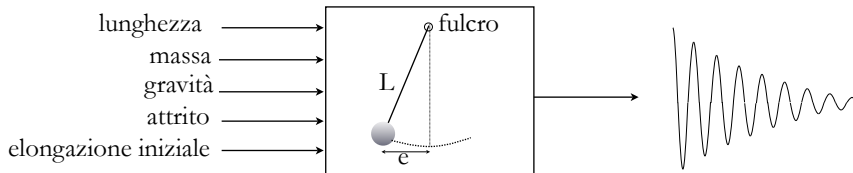


Figura 4.61 Modello fisico del pendolo

Nel caso del meccanismo di Figura 4.62 gli elementi in gioco sono: la *massa*, l'*elasticità* della molla, la *gravità*, l'*elongazione iniziale* e l'*attrito* insito nella meccanica della molla stessa.

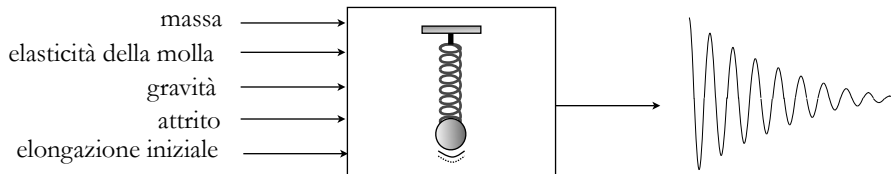


Figura 4.62 Modello fisico massa-molla

Nel caso del meccanismo di Figura 4.63 gli elementi in gioco sono: la *massa*, l'*elasticità* della molla, l'*elongazione iniziale* e l'*attrito* tra massa e piano di slittamento.

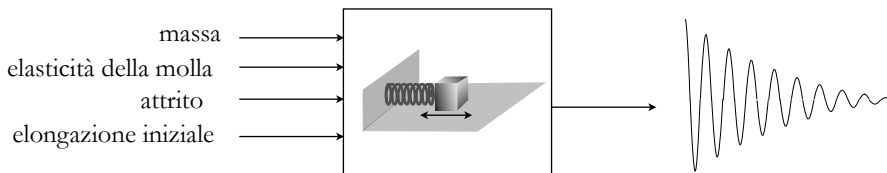


Figura 4.63 Modello fisico massa-molla e piano di slittamento

Una volta realizzato il modello fisico è possibile scegliere valori parametrici, anche molto distanti dalla situazione reale, con misure millimetriche della lunghezza del pendolo, masse minuscole o grandissime e accelerazioni di gravità diverse da quella terrestre ed ottenere in uscita il calcolo del segnale che viene utilizzato per generare suono. Per esempio: nei tre casi di modello fisico e nel caso di modello matematico di Figura 4.60, scelta una frequenza superiore a 1000 Hz, il suono risultante è quello di un “pling” di un campanellino.

Un metodo standard per approntare un modello fisico è quello di considerare un sistema fatto di masse e di molle come in Figura 4.64:

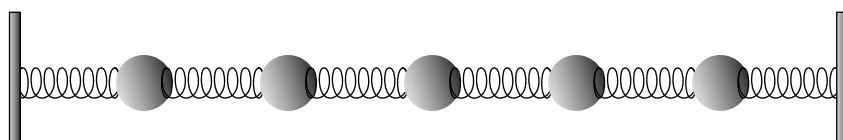


Figura 4.64 Modello fisico *masse e molle*

L'eccitazione anche di una sola massa provoca il movimento di tutte le altre in funzione dell'elasticità delle molle e delle masse stesse ed il segnale sonoro è, per esempio, quello corrispondente al moto di una massa scelta tra le altre.

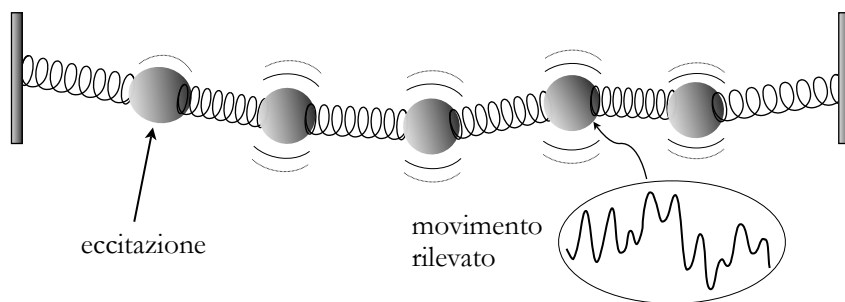


Figura 4.65 Modello fisico *masse e molle* eccitato

È chiaro che il movimento dell'intero sistema dipende da come e da dove viene esercitata l'eccitazione ed il suono risultante da dove si preleva il moto utilizzato come segnale. Il numero e le dimensioni delle masse e delle molle può essere qualsiasi ottenendo così sistemi con diversi gradi di *elasticità* (dato dalle molle) e *densità* (dato dalle masse) fino a simulare il comportamento fisico di una corda di lunghezza, diametro, tensione debitamente scelte e modi diversi di eccitazione: plectro, dita, archetto, ecc. Architetture più complesse bidimensionali possono simulare superfici vibranti come pelli di tamburo e casse risonanti come la cassa armonica di una chitarra, di un violino, ecc.

Altro discorso è quello relativo alla simulazione degli strumenti a fiato: qui si parla ovviamente di flussi di aria e di pressione e depressione formulati nel *Teorema di Bernulli* che è quello per cui “i camini tirano” e “gli aerei volano”, come recita il libro “*Le 5 equazioni che hanno cambiato il mondo*” [14]. In questo saggio vengono narrate le vite di Isaac Newton, Daniel Bernulli, Michael Faraday, Rudolf Clausius e Albert Einstein e le vicende che hanno portato alla formulazione delle equazioni che descrivono il comportamento dei principali fenomeni della Natura (gravità, fluido-dinamica, induzione elettro-magnetica, entropia e relatività) con la conoscenza delle quali è stato possibile realizzare tutte le meraviglie della tecnologia di cui facciamo quotidianamente uso.

Fondamentalmente, uno strumento musicale, considerato come un sistema meccanico/fisico, è schematizzabile con due componenti principali: l'eccitatore ed il risuonatore.

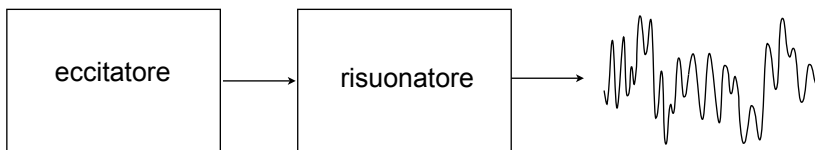


Figura 4.66 Approccio funzionale di un modello fisico

Negli strumenti a fiato l'eccitatore può essere il sistema ancia-bocchino (sassofono, clarinetto, ecc.) o labbra-bocchino (tromba, ottoni..) e il risuonatore il “tubo” cilindrico o conico dello strumento stesso. Caso diverso è quello del flauto dove è il flusso stesso dell'aria opportunamente canalizzato sull'imboccatura che mette in risonanza la colonna d'aria del “tubo-flauto”.

Negli strumenti a corda e percussione l'eccitatore è il martello di feltro del pianoforte, il plettro e le dita nella chitarra, l'archetto nel violino, viola, violoncello e contrabbasso ed il risuonatore è la corda e la cassa armonica; le bacchette ed i mallet sono gli eccitatori per i tamburi.

In base a questa distinzione, la simulazione tramite modello fisico consiste nell'esprimere con leggi opportune il funzionamento degli elementi essenziali dello strumento musicale e delle loro interazioni.

Il termine *simulatore* viene usato con lo stesso significato usato anche in altri ambiti (per esempio: simulatore di volo), dove programmi informatici sofisticati ripropongono in modo virtuale un sistema reale che reagisce in tempo reale ad informazioni date dall'uomo attraverso l'uso di sensori.

In definitiva, simulare uno strumento musicale con questo tipo di approccio significa assemblare idealmente in maniera opportuna gli elementi derivanti da una sua ipotetica suddivisione, rispettandone le relative leggi fisiche. Nel caso di simulazione di uno strumento reale è possibile arrivare ad ottenere un risultato sonoro il più possibile paragonabile a quello ottenuto con le tecniche di sintesi attuali. I vantaggi ottenibili dalle tecniche di sintesi basate sui modelli fisici sono molteplici:

- corrispondenza diretta tra i parametri a disposizione e la realtà costruttiva dello strumento;
- possibilità di gestire l'intervallo dei valori parametrici in gioco oltre i limiti imposti dalla realizzabilità fisica stessa;
- possibilità di realizzare “strumenti musicali” con caratteristiche ideali altrimenti non realizzabili in pratica, come ad esempio un’arpa eccitata dall’archetto di un violino, una tromba eccitata con ancia semplice, ecc.
- approccio più intuitivo nella realizzazione di uno strumento, dato che l’assemblaggio viene eseguito come se fossero a disposizione realmente i vari pezzi costituenti lo strumento stesso.

Esiste una vastissima letteratura intorno ai modelli fisici: è sufficiente una ricerca in rete con “physical modeling” per conoscere i risultati della ricerca di chi ha affrontato questo affascinante modo di sintesi audio: Perry Cook (Princeton), Julius Smith (Stanford), ecc.

Un modello fisico richiede una notevole mole di calcolo ed, al solito, per molto tempo non è stato possibile utilizzare in tempo reale questo approccio alla sintesi sonora. Nella metà degli anni '90 la Yamaha mise in commercio un sintetizzatore basato esclusivamente sul physical modeling che ebbe molto successo ed ancora utilizzato: si tratta del VL1, al quale furono affiancati altri modelli che includevano anche sensori di espressione associati alle varie forme di modulazione timbrica.

Infatti, accanto alla macchina in sé, grande importanza assume anche tutta una sensoristica, come il *breath-control* in grado di rilevare l'intensità del soffio dalla bocca e diverse tipologie di gestualità che, insieme alla tradizionale tastiera musicale, danno senso compiuto alle potenzialità espressive timbriche, in special modo in relazione alla corrispondenza diretta ed intuitiva tra azione esercitata e risultato sonoro, come accade con gli strumenti musicali fisici veri.

Riferimenti

- [1] www.seg.co.uk/telecomm/automat1.htm
- [2] en.wikipedia.org/wiki/Telephone_exchange
- [3] *La Scienza del Suono*, John R. Pierce, Zanichelli Editore, Bologna, 1988 (prima edizione).
- [4] *The Technology of Computer Music*, MIT Press, 1969.
www.scribd.com/doc/140253767/Max-Mathews-The-Technology-of-Computer-Music
- [5] it.wikipedia.org/wiki/Anno_bisestile
- [6] www.image-line.com/documents/morphine.html
- [7] The Synthesis of Complex Audio Spectra by Means of Frequency Modulation, J. Chowning, in *Journal of the Audio Engineering Society*, vol. 21, n° 7, 1973.
- [8] www.cim.mcgill.ca/~clark/nordmodularbook/fmalgorithms.htm
- [9] *Elements of Computer Music*, F. Richard Moore, Prentice-Hall, 1990.
- [10] www.csounds.com/journal/issue11/distortionSynthesis.html
- [11] Digital synthesis of Plucked-Strings and Drum Timbres, in *Computer Music Journal*, 1983, Mit Press;
- [12] compbio.soe.ucsc.edu/~karplus/papers/digitar.pdf
- [13] davidjaffesite.squarespace.comstorage/writings/Jaffe-Smith-Extensions-CMJ-1983.pdf
- [14] *Le 5 equazioni che hanno cambiato il mondo*, Michael Guillen, TEA Edizioni, Milano, 2005.

MIDI Musical Instrument Digital Interface



- 6.1 L'interfaccia
- 6.2 Il linguaggio
 - 6.2.1 Messaggi di canale (channel message)
 - 6.2.2 Messaggi di sistema (system message)
- 6.3 La sintassi MIDI
- 6.4 Standard MIDI File e General MIDI
- 6.5 Oltre il MIDI: ZIPI e OSC



Il MIDI è come il *Ginkgo biloba*: è un “fossile vivente”. In Biologia un fossile vivente è un organismo sviluppatosi in ere geologiche remote che non si è estinto ma che neppure si è evoluto mediante trasformazioni genetiche. Comparso nel Giurassico ed arrivato fino ai giorni nostri con pochissime modificazioni sostanziali, il ginkgo è una pianta che si distingue nei giardini pubblici o privati per un tipo di arborescenza molto rudimentale e soprattutto per la forma delle foglie con nervatura a ventaglio, decisamente primordiale rispetto a quella ramificata delle foglie di una qualsiasi altra pianta o albero che ha subito evoluzione.

Anche il MIDI presenta le caratteristiche tipiche dello stadio evolutivo della tecnologia digitale degli anni '80 e così è arrivato fino ad oggi senza subire evoluzione sia nella tipologia delle connessioni e dell'interfaccia, sia nelle caratteristiche della struttura della messaggistica del linguaggio.

Fino a tutti gli anni '70 la strumentazione musicale elettronica utilizzata nella musica leggera e rock era completamente analogica; ma con l'introduzione sul mercato di intere famiglie di microprocessori a basso costo e alte prestazioni, si intravide la possibilità di trasferire le metodologie operative tipicamente usate nella realizzazione di calcolatori general-purpose sulle apparecchiature musicali.

Si cominciò così a realizzare tastiere musicali per la generazione di opportuni codici binari che riportavano informazioni relative alla pressione dei tasti per controllare i generatori sonori installati all'interno dell'apparecchiatura stessa oppure, via trasmissione dati, apparecchiature esterne realizzate anche da case costruttrici diverse.

Nel 1981 fu presentata alla AES, Audio Engineering Society, una proposta contenente idee e suggerimenti relativi ad un'interfaccia seriale per collegare tra loro apparecchiature elettroniche usate in campo musicale. Questo schema di intenti siglato inizialmente con la dizione Universal Synthesizer Interface, servì da stimolo per tutti i maggiori costruttori di apparecchiature musicali ad uno sforzo comune rivolto alla definizione di uno standard di comunicazione.

Le specifiche di questo nuovo standard, che prese il nome di MIDI, Musical Instrument Digital Interface, furono presentate al pubblico agli inizi degli anni '80 e già nel 1983 apparvero sul mercato tastiere musicali e batterie elettroniche dotate di ingresso e di uscita MIDI. Il MIDI divenne ben presto uno standard riconosciuto in tutto il mondo dai costruttori di strumenti musicali ed intorno al 1983/1984 sorsero due comitati, l'americana MMA (MIDI Manufacturers Association) e il giapponese JMSC (Japanese MIDI Standards Committee), che avevano lo scopo di perfezionare il MIDI e di far rispettare le decisioni prese in proposito.

L'introduzione del MIDI provocò all'inizio una serie di malintesi tra i musicisti dell'area leggera e rock, per i quali fino a quel momento la parola "elettronica" riguardava esclusivamente tutto ciò che aveva a che fare con il segnale audio in bassa frequenza per cui, anche se ora il gergo si arricchiva delle due nuove parole "digitale" e "computerizzato", ciò non aiutava a capire che nei cavi di collegamento MIDI non passava segnale audio, ma dati!

L'informazione trasmessa sui cavi MIDI è infatti senz'altro più vicina alla notazione musicale del pentagramma piuttosto che alla rappresentazione digitale del segnale. La differenza è innanzitutto concettuale, ma riguarda anche il flusso della quantità di informazione. Per codificare e/o trasmettere un suono della durata di 1 solo secondo con caratteristiche di alta fedeltà e stereofonia, come succede ad esempio nello standard 16bit/44.100 Hz, c'è bisogno di $2 \times 2 \times 44.100 = 176.400$ bytes, mentre per codificare e trasmettere il comando di esecuzione di un suono di durata variabile tra 1 millisecondo e l'eternità, sono sufficienti appena 6 byte. I 176.400 byte al secondo sono

necessari indipendentemente dal fatto che il segnale si riferisca alle prime note di “Jingle Bell” oppure al “pieno” dell’inizio della Sagra della Primavera, mentre nel caso della corrispondente codifica MIDI la differenza in numero di byte tra le due situazioni è decisamente grandissima.

Quando un tasto viene premuto su una tastiera MIDI, la logica di controllo genera una sequenza di 3 byte che indicano nell’ordine: la segnalazione di evento “tasto premuto”, il numero del tasto premuto e la velocità con cui il tasto è stato premuto, che corrisponde al controllo dell’espressione. Nel momento in cui il dito si solleva dal tasto questo nuovo evento provoca la generazione di altri 3 byte che indicano nell’ordine: segnalazione di “tasto rilasciato”, numero del tasto rilasciato e velocità di rilascio. Il linguaggio MIDI è piuttosto complesso ed articolato e, oltre questa basilare codifica, copre molto bene la larghissima casistica delle informazioni musicali di espressione.

Con la sigla MIDI si intende indifferentemente sia “la” interfaccia sia “il” protocollo di comunicazione ed “il” linguaggio ed è per questo che a volte viene usato con l’articolo maschile e a volte con quello femminile.

6.1 L’interfaccia

La MIDI è un’interfaccia *seriale asincrona* caratterizzata da una velocità di trasmissione di 31.250 bit al secondo con modalità “start/stop” e quindi di circa 3KByte al secondo. Questa tipologia di interfaccia era caratteristica degli anni ’80, periodo in cui le periferiche del computer come stampanti e modem erano collegate con la *seriale standard RS-232* che consentiva velocità di trasmissione dati di alcune decine di migliaia di bit al secondo, ben lontane da quelle delle attuali interfacce *USB* e *FireWire* (sempre *seriali*) che consentono la trasmissione di decine e centinaia di megabit al secondo.

Ma ancora oggi la velocità del MIDI è quella originaria ed è proprio il numero 31.250 bit/sec che testimonia il livello di evoluzione tecnologica di quel periodo storico.

A cavallo fra gli anni ’70 ed ’80 fecero la loro comparsa sul mercato i primi personal computer, tra i quali di grandissimo successo furono il Commodore64 e l’AppleII. Il cuore di quei due computer era il microprocessore Motorola 6502 che aveva un’architettura a 8 bit ed una frequenza di clock di 1 MegaHz e cioè più di 1000 volte più lenta di quella degli attuali computer. Tutta la logica interna funzionava con il clock ad 1MHz ed i circuiti integrati di ingresso/uscita dati erano l’UART (Universal Asynchronous Receiver/Trasmitter) e l’ACIA (Asynchronous Communication Interface Adapter) preposte alla

conversione del contenuto di un byte della logica interna del 6502 in una sequenza di 8 bit in uscita ed una sequenza di 8 bit provenienti dalla linea seriale in ingresso, in un byte; l'UART e l'ACIA consentono di scegliere la velocità di trasmissione sulla base della frequenza di clock tramite divisioni binarie (2,4,8,16,32,64) e per il MIDI venne scelta la divisione per 32 da cui risulta la velocità di $1.000.000/32=31.250$ bit al secondo.

In sede di definizione dello standard MIDI, i progettisti scelsero il tipo di comunicazione seriale e cioè un giusto compromesso tra semplicità di collegamento tramite normali cavi multipolari ed efficienza funzionale. Le apparecchiature MIDI erano/sono caratterizzate dalla presenza delle porte che si presentano come connettori DIN a cinque poli:

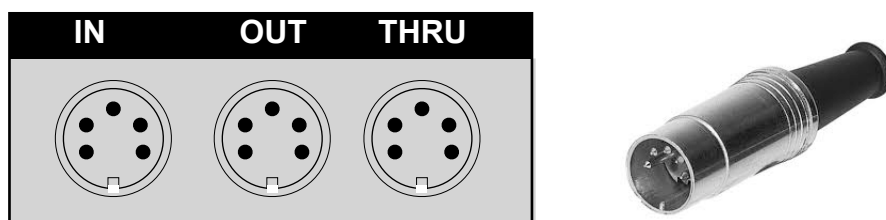


Figura 6.1 Connessioni e connettore MIDI

La porta IN riceve i messaggi inviati dall'esterno, la porta THRU ritrasmette all'esterno i messaggi ricevuti sulla porta IN e la porta OUT trasmette all'esterno solo i messaggi generati all'interno dello strumento.

Il modo più semplice di realizzare una rete MIDI è il collegamento *in cascata* consentito dalla stessa configurazione standard del terzetto di connettori MIDI IN-THRU-OUT presente su tutte le apparecchiature.

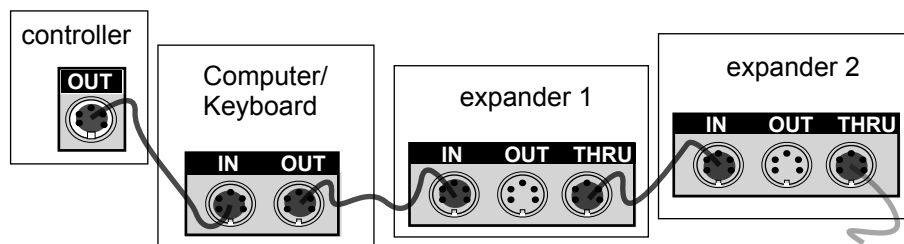


Figura 6.2 Rete MIDI standard

Situazioni più complesse in cui erano presenti molte apparecchiature MIDI venivano risolte con una configurazione *a stella* attraverso l'uso di un'apparecchiatura chiamata THRU-BOX che dispone di una sola porta IN e molte porte THRU; un'alternativa era data dalla configurazione *a matrice* attraverso l'uso di un'apparecchiatura chiamata PATCH-BAY che dispone di molte porte IN e molte porte OUT: in questo caso i collegamenti tra le apparecchiature erano configurabili via software senza dover intervenire fisicamente sui cavi di collegamento.

Per il Commodore64 e l'AppleII furono realizzate interfacce MIDI in forma di dispositivi esterni aggiuntivi che si inserivano in vari modi nella logica interna (tipicamente sul bus-dati).

Ma il computer che più degli altri contribuì alla diffusione dello standard MIDI con applicazioni musicali significative fu l'ATARI che disponeva, oltre alle normali alle porte di ingresso/uscita dati, dell'interfaccia MIDI direttamente cablata sulla logica interna e con i connettori IN-OUT montati sullo chassis.

È possibile suddividere la strumentazione MIDI in tre categorie principali: *sequencer*, *expander* e *controller*. I sequencer sono programmi operanti su computer atti alla preparazione di brani musicali in post-produzione in termini di sequenze di eventi che generano messaggi MIDI per il controllo degli expander, atti alla generazione sonora. L'evoluzione di queste applicazioni come CuBase, Pro-tools, Logic ed il popolare Garage-Band, consentono di trattare sulla stessa time-line sia segnale audio sia eventi MIDI altrimenti detti DAW, *Data and Audio Workstation*.

Gli expander, secondo la denominazione dell'epoca, sono i generatori di suono, come sintetizzatori e campionatori, e i processori-audio per trattare in tempo reale il suono con effetti di riverbero, flanger, chorus ecc.

All'epoca la periferica MIDI di sintesi e di elaborazione audio era una necessità ma oggi, viste le potenzialità di calcolo a disposizione, la sintesi e l'elaborazione del suono avvengono all'interno del computer stesso, mediante applicazioni denominate per l'appunto *virtual expander*.

I controller sono strumenti generatori di messaggi MIDI e coincidono in pratica con la strumentazione musicale classica opportunamente "midizzata": tastiere musicali, MIDI-sax, drum-pads, chitarre MIDI, pitch-to-MIDI-converter, ecc., e costituiscono cioè l'interfaccia uomo-mondoMIDI.

Va osservato che molti controller utilizzano oggi il protocollo ed il connettore USB (Universal Serial Bus) pur utilizzando la messaggistica del linguaggio MIDI originario, oggetto del prossimo paragrafo.

6.2 Il linguaggio

Qualunque sia la struttura di una rete MIDI, la comunicazione tra le varie apparecchiature viene fatta attraverso l'invio di *messaggi*: un *messaggio* è un pacchetto di byte costituito da un *comando* seguito da uno o più *dati*.

Un comando è un byte (byte=pacchetto di 8 bit) con il bit più significativo (MSB=il primo da sinistra) a 1, mentre un dato è un byte con il MSB a 0.

1xxxxyyy comando Odddddd dato

Quando un messaggio viene generato, esso è presente su tutta la rete ed arriva indistintamente a tutte le apparecchiature MIDI; affinché i messaggi abbiano effetto in maniera selettiva sulle singole apparecchiature ed all'interno di ogni singola apparecchiatura, si fa uso dei *canali logici* codificati nel byte-comando stesso, previa opportuna predisposizione dell'apparecchiatura con quelli che vengono definiti i *modi di ricezione* dell'apparecchiatura stessa.

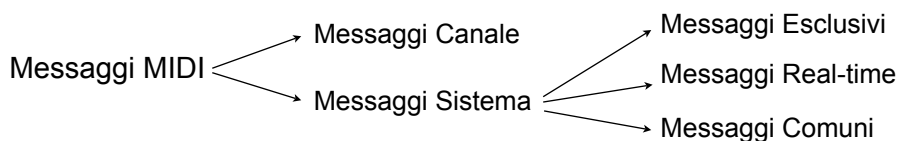
Un *byte-comando* contiene due tipi di informazione: il *tipo-del-comando* codificato sui tre bit *xxx* ed il numero del *canale logico* sul quale quel comando deve agire, codificato sui quattro bit *yyyy* del semibyte di destra.

Un semibyte viene espresso anche con una *cifra esadecimale* e cioè un valore che va da 0 a 15 codificato con \$0 \$1 \$2 \$3 \$4 \$5 \$6 \$7 \$8 \$9 \$A \$B \$C \$D \$E \$F cui corrispondono pacchetti di 4 *bit* nelle configurazioni 0000 0001 0010 0011 0100 0101 0110 0111 1000 1001 1010 1011 1100 1101 1110 1111

Poiché il MSB in un byte-comando è sempre 1, i comandi possibili sono 8 (1000=\$8, 1001=\$9, 1010=\$A, 1011=\$B, 1100=\$C, 1101=\$D, 1110=\$E, 1111=\$F) ed il numero dei canali possibili è 16 (0000÷1111 oppure \$0÷\$F). Mentre la codifica binaria (o esadecimale) va da 0 a 15, nella terminologia MIDI i canali sono numerati da 1 a 16, per cui \$0 è il canale 1, \$1 è il canale 2, \$F è il canale 16.

I valori del byte-dato vanno da 0 a 127 (00000000÷01111111) perché il MSB è sempre 0 e quindi le cifre binarie a disposizione sono 7, per cui $2^7-1=127$.

I messaggi che costituiscono il linguaggio MIDI vengono suddivisi in Messaggi di Canale e Messaggi di Sistema che a loro volta si suddividono in Messaggi Esclusivi, Messaggi Real-time e Messaggi Comuni.



6.2.1 Messaggi di canale (channel message)

Questi messaggi controllano le voci e riguardano per lo più le modalità esecutive di una nota. La parte comando di questi messaggi è un numero binario compreso tra 1000 e 1110 (che espresso in esadecimale è una delle possibili cifre \$8 \$9 \$A \$B \$C \$D \$E) seguito dal numero di canale “n”. Il semibyte di sinistra rappresenta il comando, mentre il semibyte di destra rappresenta il canale. Ogni messaggio ha poi la sua particolare configurazione:

Note Off	\$8n kk vv
Note On	\$9n kk vv
Polyphonic Key Pressure	\$An kk vv
Control Change	\$Bn cc vv
Program Change	\$Cn pp
Channel Pressure	\$Dn vv
Pitch Bend (Pitch Wheel)	\$En vv ww

I messaggi \$9n=NoteOn e \$8n=NoteOff riguardano rispettivamente l’attivazione e la disattivazione delle note. In entrambi i comandi “n” indica il canale logico su cui il comando ha effetto, “kk” indica il numero di tasto e “vv” indica la velocità di attacco se il comando è Note On e la velocità di rilascio, se il comando è NoteOff; il numero di tasto kk può assumere valori da 1 a 127 e quindi ben oltre i classici 88 tasti della tastiera di pianoforte.

Per esempio: $kk=57_{dec}=39_{esa}$ corrisponde al LA centrale; $kk=60_{dec}=3C_{esa}$ è il DO tre semitoni sopra. Tutte le altre note della tastiera si trovano incrementando o decrementando quei valori del corrispondente numero di semitoni.

Il messaggio NoteOn attiva l’emissione della nota “kk” iniziando dalla fase di attacco degli involucri predisposti per quel timbro ed il parametro “vv” influisce sulla pendenza. Gli involucri vengono poi scanditi fino al punto di rilascio e mantenuti in quella posizione fintantoché non arriva il messaggio di NoteOff. Poiché la velocità di rilascio nel NoteOff è tutto sommato poco significativa, è entrato in uso codificare questo messaggio per mezzo del messaggio NoteOn con $vv=0$. Questa soluzione porta ad una riduzione del traffico di informazioni sulla rete MIDI. Per convenzione infatti, se un messaggio segue ad un messaggio dello stesso tipo sullo stesso canale) non è necessario ripetere la parte comando del messaggio, ma è sufficiente la parte dato. Le due sequenze di messaggi sono equivalenti (numeri in esadecimale):

91 47 50 91 40 50 <=> 91 47 50 40 50

Il principio è esteso a tutti i comandi ma è particolarmente utile per i comandi After-Touch e Pitch-Bend, tant'è che per i comandi di canale si parla anche di *status byte* con il seguente significato: fino a che non arriva un nuovo comando (e cioè un byte con MSB=1) si intende che i byte-dato si riferiscono implicitamente all'ultimo comando; si dice anche che l'apparecchiatura è nello *status* determinato dall'ultimo comando.

Il Polyphonic Key Pressure (\$An kk vv) viene generato fra i messaggi di NoteOn e NoteOff e dà la possibilità di variare con continuità il valore volumetrico della nota dopo che è stata emessa, ma prima di essere rilasciata. Questo messaggio viene generato solo da tastiere equipaggiate con sensori particolari oltre a quelli ordinari che rilevano la velocità di pressione dei tasti.

Il Control Change (\$Bn cc vv) consente di comunicare cambiamenti di valori relativi a controlli di volume, percentuale di eco, posizione panoramica, ecc. Il valore "cc" esprime il numero del controllo, mentre "vv" esprime il valore del controllo. Esiste una tabella di riferimento ben precisa ma di solito i valori più utilizzati sono: \$07=volume, \$0B pan-pot, ecc.; i valori di "vv" vanno da 0 a 127 (\$00÷\$7F) e sono perciò in grado di controllare con raffinatezza il tipo di controllo.

Il Control Change viene usato anche per predisporre un'apparecchiatura ad una particolare modalità di ricezione:

\$Bn 7C 00 -> OmniOff

\$Bn 7D 00 -> OmniOn

\$Bn 7E 00 -> MonoOn

\$Bn 7F 00 -> PolyOn

Il comando Program Change (\$Cn pp) serve per comunicare un cambio di programma e cioè un nuovo timbro o una nuova configurazione orchestrale opportunamente predisposta e memorizzata sull'apparecchiatura. Il valore "pp" rappresenta il valore del preset voluto.

Il Channel Pressure (\$Dn vv) indica la pressione esercitata sui tasti dopo che i NoteOn sono stati generati, ma non ancora sono stati generati i relativi NoteOff; la pressione è specificata dal valore "vv".

Il Pitch Bend (\$En vv ww) comunica una variazione frequenziale nell'intervallo di +2/-2 semitoni intorno a quella del numero di tasto premuto espressa con molta raffinatezza per mezzo dei 14 bit dei due byte-dato pari ad un intervallo di valori tra +8192 e -8191.

6.2.2 Messaggi di sistema (system message)

I messaggi di sistema vengono riconosciuti indistintamente da tutti i moduli componenti una rete MIDI; la parte comando di questi messaggi è \$Fx dove “x” indica il tipo di messaggio-sistema con $x=0000\div 1111$ e non c’è perciò nessun riferimento ai canali; i messaggi di sistema si suddividono in messaggi *esclusivi*, messaggi *comuni* e messaggi *real-time*.

I messaggi esclusivi (System Exclusive Message) consentono alle case costruttrici di strumenti di definire sotto-linguaggi per l’appunto esclusivi per le proprie apparecchiature ed addirittura per ogni particolare apparecchiatura. La sintassi di questo messaggio è la seguente:

$$\text{\$F0 } \langle \text{ID} \rangle \langle \text{dato} \rangle \langle \text{dato} \rangle \langle \text{dato} \rangle \dots \text{\$F7}$$

dove \$F0 è l’identificatore di inizio messaggio esclusivo:

$\langle \text{ID} \rangle$ rappresenta il codice di identificazione della casa costruttrice che in sede di accordo MIDI è stato stabilito essere: \$43 per la Yamaha, \$41 per la Roland, \$04 per Moog e così via;

$\langle \text{dato} \rangle$ rappresenta il contenuto del messaggio come successione di dati (valori $0\div 127$);

\$F7 è l’identificatore di fine messaggio esclusivo.

I messaggi comuni (System Common Message) riguardano il brano in esecuzione (o in registrazione) e consentono di posizionarsi opportunamente all’interno del brano.

I messaggi real-time (System Real-Time Message) si occupano di mantenere il sincronismo tra i vari moduli, fungendo da metronomo di sistema.

I messaggi comuni e real-time (esplicitati nel dettaglio nel prossimo paragrafo) sono particolarmente utili nella codifica degli Standard Midi File perché consentono la “traduzione” in successione di eventi di brani musicali in quella che può essere considerata la traslitterazione di una partitura musicale polifonica che può essere utilizzata per la memorizzazione e/o la trasmissione in rete di brani musicali “midizzati”.

6.4 Standard MIDI File e General MIDI

Più o meno un secolo fa esisteva la modalità di “registrare” esecuzioni musicali eseguite su pianoforte per mezzo di *pianoforti a rullo* dove, non il suono, ma l'effetto meccanico dei tasti premuti veniva memorizzato in forma di perforazione di rulli di carta in corrispondenza dei tasti.

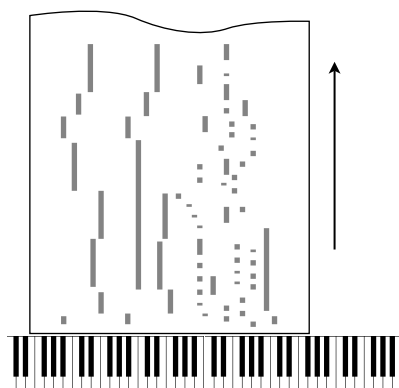


Figura 6.3 Piano-roll

Mandare in esecuzione oggi quei piano-roll significa qualcosa di “di più” che ascoltare l'audio di registrazioni su disco riprodotti da altoparlanti perché qui c'è la suggestione della vitalità del gesto e del tocco personale di quei formidabili pianisti e compositori come G. Mahler, E. Grieg, S. Rachmaninoff, C. Debussy, S. Joplin, S. Prokofiev, G. Gerswhin, J.R. Morton, ecc. che si possono “ora” ri-ascoltare “dal vivo” su pianoforti veri. Esiste in proposito tutto un mondo ancora vivace ed attivo interessante da scoprire [1][2].

L'informazione memorizzata su un rullo di carta di piano-roll è molto simile a quella dei messaggi midi NoteOn/NoteOff. Simile, non uguale, perché nel linguaggio MIDI non esiste il concetto di “tempo” e di “durata”. Ad esempio: il messaggio di Note-On riporta il “numero tasto” (*key-number*) corrispondente alla posizione di una nota sul pentagramma, ma non l'informazione della figura musicale minima, croma, ecc. La gestione del tempo viene delegata all'esecutore umano che agisce sui controller che generano i messaggi utilizzati dall'expander, o dal *virtual expander*, per generare suono nel momento stesso in cui ricevono il messaggio.

Il MIDI è infatti essenzialmente un linguaggio *real-time*, il che significa che quando i messaggi vengono generati, immediatamente l'evento corrispondente viene interpretato ed eseguito dal generatore di suoni con l'emissione di suono.

Memorizzare la successione di eventi così come vengono generati non sarebbe sufficiente per avere la rappresentazione statica “midizzata” di un brano musicale, come succede nel piano-roll, perché mancherebbe l’informazione “tempo”. Per ottenere l’equivalente del piano-roll si tratta allora di introdurre un’informazione aggiuntiva che quantifichi il ritardo tra un messaggio (anche complesso) ed il successivo. È questo il principio seguito per la definizione degli Standard MIDI File (SMF) intesi come formato di memorizzazione, di trasmissione in rete e di scambio tra diverse applicazioni.

Consideriamo ad esempio il seguente frammento musicale:

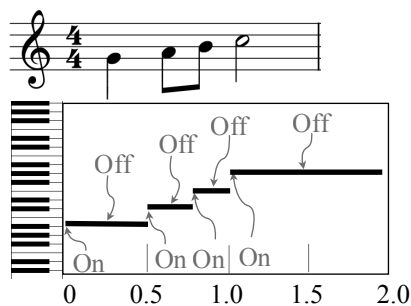


Figura 6.4 Corrispondenza pentagramma/piano-roll

e supponiamo un metronomo a 120 (battiti/minuto) in modo tale che la croma=0,25secondi, la semiminima=0,5secondi e la minima=1secondo. Assumiamo che valgano le seguente convenzioni:

[90 3A 30] messaggio MIDI data dalla tripla [NoteOn-keynum-velocity] dove

90_{esa} =NoteOn sul canale 1

80_{esa} =NoteOFF sul canale 1

37_{esa} = 55_{dec} = Sol3

39_{esa} = 57_{dec} = La3

3B_{esa} = 59_{dec} = Si3

3C_{esa} = 60_{dec} = Do4

40_{esa} = 64_{dec} = valore medio di velocity

0,xx attesa in frazioni di secondo tra un messaggio e l’altro.

Un possibile “MIDI file” relativo al frammento musicale di Figura 6.4 conterrebbe i seguenti dati:

```
[90 37 40] _0,25_ [80 37 00] _0,25_ [90 39 40] _0,125_ [80 39 00] _0,125_
[90 3B 40] _0,125_ [80 3B 00] _0,125_ [90 3C 40] _0,5_ [80 3C 00] _..._
```

È proprio questa la rappresentazione dinamica di un brano musicale in un MIDI file, descritta in termini di successione di messaggi NoteOn e NoteOff opportunamente temporizzati (*time-stamp*) da un'applicazione di esecuzione di MIDI file. Gli Standard MIDI File sono il risultato del secondo accordo nell'ambito della realizzazione di applicazioni per la produzione musicale computerizzata, deciso e coordinato dalla MMA, MIDI Manufacturers Association.

Un'altra importante convenzione stabilita dalla MMA riguarda la selezione dei timbri/strumenti sull'expander per avere la giusta esecuzione musicale. Un expander (si intende sempre anche un *virtual expander*) ha a disposizione una grande varietà di timbri per creare situazioni orchestrali differenti, ed è possibile fare assegnamenti canale-strumento, sia operando direttamente sul dispositivo, sia via MIDI mediante il messaggio ProgramChange, \$Cn pp che imposta il canale "n" a sintetizzare suoni con lo strumento "pp". Dal momento che "pp" può assumere valori tra 0 e 127, è possibile scegliere tra 128 strumenti diversi stabilito nello standard che viene definito *General MIDI* o GM.

Al breve frammento musicale di Figura 6.5, costituito da 2 voci, viene associata la corrispondente versione piano-roll; c'è da osservare infatti che i sequencer MIDI ed i DAW utilizzano questa rappresentazione perché rispondente visivamente alla successione ed alla sovrapposizione degli eventi del piano-roll meccanico vero e proprio. La tastiera di pianoforte posta in verticale serve soltanto da riferimento per le strisce che indicano le note la cui lunghezza descrive la loro durata, e non ha nessun riferimento al timbro di pianoforte; mediante il messaggio di Program Change è infatti possibile assegnare ad ogni traccia uno dei possibili 128 timbri stabiliti nel GM.

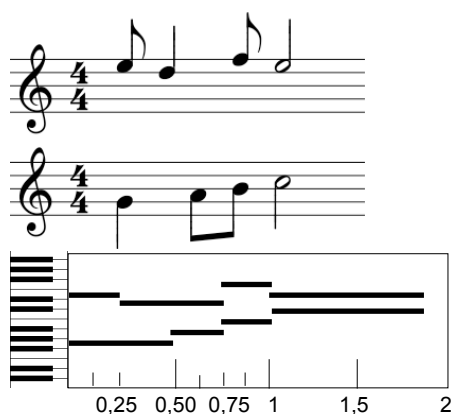


Figura 6.5 Corrispondenza pentagramma/piano-roll di due voci

In questo esempio la traccia 1 si riferisce al canale 1 con i messaggi \$90=NoteOn e \$80=NoteOff, mentre la traccia 2 si riferisce al canale 2 con i messaggi \$91=NoteOn e \$81=NoteOff.

In testa al MIDI file si mettono i messaggi di ProgramChange per predisporre l'expander agli strumenti 12.vibrafono per la traccia 1 e 57.tromba per la traccia 2, come risulta dalla carta del GM. Anche qui i valori sono espressi in esadecimale, perciò $12_{dec}=0C_{esa}$ e $57_{dec}=49_{esa}$ ecc.

Il "MIDI file" risulta ora essere:

```
[C0 0C][C1 49]      [90 37 40][91 44 40] _0,125_ [81 44 00] _0,125_ [80 37
00][91 42 40] _0,25_ [81 42 00][90 39 40] _0,125_ [80 39 00] _0,125_ [90
3B 40][91 45 40] _0,125_ [80 3B 00][81 45 00] _0,125_ [90 3C 40][91 44 40]
_0,5_ [80 3C 00][81 44 00] ..._
```

Un vero Standard MIDI file (SMF) ha una struttura molto più complessa di quella sopra riportata in quanto include anche messaggi che controllano i parametri di espressione e di dinamica attraverso l'uso di tutta la messaggistica MIDI; i due esempi servono solo a mettere in evidenza la logica di interlacciamento dei messaggi MIDI in successione lineare di eventi che invece, per sua natura, la musica polifonica "vorrebbe" in struttura parallela.

La struttura seriale dello SMF realizza il parallelismo della musica attraverso l'uso dei messaggi di canale che indirizzano i messaggi ai singoli strumenti stabiliti nello standard GM. La contemporaneità di più eventi, come succede negli accordi o più semplicemente nell'essere musicalmente "a tempo", viene risolta giustapponendo messaggi di NoteOn senza l'indicazione di attesa: va osservato tuttavia che c'è sempre un ritardo dell'ordine di 1 millisecondo tra più "eventi contemporanei" come ad esempio le note di un accordo: il ritardo è dovuto alla velocità di trasmissione dati di 3Kbyte/sec del MIDI che è comunque sotto la soglia di percezione psico-acustica.

Uno SMF viene preceduto da una testata iniziale o *header* (nell'esempio costituita solo dai messaggi ProgramChange per associare gli strumenti ai canali) che stabilisce sostanzialmente il numero di tracce ed il numero di strumenti secondo lo standard GM ed il tipo di codifica del messaggio relativo al ritardo: microsecondi, frazione di semiminima, frazioni di 24esimo o di 25esimo di secondo ecc.

Queste due ultime codifiche di misura del tempo sono state introdotte per gettare un ponte ad un altro standard, l'SMPTE, Society of Motion Picture and Television Engineers in uso nei sistemi di sonorizzazione cinematografica e televisiva, per dare una risposta operativa all'attività di post-produzione.

Linguaggi musicali



- 7.1 Linguaggi testuali
 - 7.1.1 CSound
 - 7.1.2 SuperCollider
 - 7.1.3 Chuck e STK
 - 7.1.4 pureCMusic
 - 7.1.5 Common Lisp Music
- 7.2 Linguaggi visuali
 - 7.2.1 MAX/MSP e Pure Data
 - 7.2.2 Kyma
- 7.3 Note conclusive
- 7.4 Composizione algoritmica



In alcuni settori di applicazione dell'informatica diversi da quello musicale, l'introduzione del calcolatore ha voluto dire per lo più facilitare, migliorare e ottimizzare pratiche e metodologie operative, senza però apportare (perché di fatto non richiesto) nuove possibilità e potenzialità espressive. Ad esempio: fino alla fine degli anni '70, giornali quotidiani e settimanali venivano composti utilizzando la *linotype*, che era una speciale macchina da scrivere in grado di generare fisicamente blocchetti di piombo riproducenti i caratteri premuti sulla tastiera ed assemblati sulle lastre di stampa. Solo a partire dagli anni '80 si è passati alla completa computerizzazione delle varie fasi della preparazione di un giornale, dove ciascun giornalista scrive il suo articolo al computer generando un file e l'impaginatore utilizza un sistema di video-composizione in grado di preparare le lastre finali per le rotative di stampa.

Tutto ciò migliorò sicuramente le condizioni di lavoro di coloro che operavano nelle redazioni dei giornali, tra l'altro liberati dall'incubo delle malattie polmonari dovute all'uso costante di piombo, ma il lettore non si accorse minimamente di tutto questo, perché l'innovazione tecnologica non aveva provocato cambiamenti nel modo di scrivere o nell'affrontare argomenti e soprattutto nelle modalità di fruire il giornale comprandolo all'edicola e leggendolo.

I personal computer danno oggi la possibilità di trattare parole e grafismi con modalità tali da risolvere tranquillamente anche tutti i problemi della scrittura musicale tradizionale su pentagramma.

L'introduzione dell'uso del computer in questo settore ha risolto il secolare problema della produzione di partiture, un tempo affidato alla figura del copista, tanto che attualmente è possibile trovare sul mercato prodotti molto sofisticati per l'editoria di musica pentagrammata che consentono di rappresentare anche le più inconsuete situazioni musicali, in termini di configurazioni ritmiche, simboli di espressione e controllo della dinamica.

Di solito questi editori danno anche la possibilità di produrre, per ogni simbolo scritto, il corrispondente effetto acustico attraverso la trasformazione in MIDI-file ed ottenere così la vera e propria esecuzione musicale.

Va comunque osservato che per sua storia e natura, il pentagramma è fortemente orientato a rappresentare la musica della cultura occidentale da eseguirsi con gli strumenti musicali basati sulla divisione dell'ottava in dodici semitoni temperati e con ben precise figurazioni ritmiche.

Viceversa, con l'uso dell'elettronica e in particolare dell'informatica, è stato possibile scendere al livello di dettaglio micro-strutturale del suono, livello al quale il concetto di nota musicale come quantità bi-dimensionale *altezza/durata* data da *posizione-sul-pentagramma/figura* perde il significato originario, per lasciare il posto al più generale concetto di *entità acustica* come quantità pluri-dimensionale di *frequenza, durata, timbro, volume, posizione-nello-spazio, effetto-speciale*, del tutto non rappresentabile su una partitura pentagrammata.

Nell'ambito della musica elettronica analogica degli anni '60-'70 erano state fatte esperienze dove la composizione musicale veniva in qualche modo graficata per essere documentata ed eventualmente riprodotta. Venivano proposti schematismi creati all'occasione, anche su fogli pentagrammati, nelle forme più inconsuete e senza una precisa simbologia di riferimento. E magari succedeva spesso che tali composizioni rimanessero al solo stadio di proposta e che la partitura assumesse valore artistico più come opera grafico-pittorica che musicale. E tutto questo perché nel nobile sforzo di superare i limiti geografici del territorio musicale tradizionale, ci si avventurava in un mondo totalmente privo di un qualche formalismo linguistico di riferimento.

La tecnologia informatica, dove invece il formalismo linguistico di riferimento esiste e si esplica nel concetto di *sintassi* che è alla base dei *linguaggi di programmazione*, consente di rivisitare le possibilità operative compositive del pentagramma e concludere che, quando si usano sistemi informatici, l'uso del pentagramma come linguaggio di composizione non porta alla innovazioni di rilievo nella forma e nei contenuti rispetto alle possibilità già presenti nell'attività compositiva, attuata secondo le modalità classiche: pianoforte, carta pentagrammata e penna.

Se si usa il computer con il pentagramma come unico elemento descrittivo di una composizione musicale, in generale si possono fare solo cose che in realtà si potrebbero fare anche senza l'uso del computer. È invece auspicabile poter allargare gli orizzonti creativi utilizzando le potenzialità operative che lo strumento informatico offre.

Ed è successo, infatti, che proprio le possibilità operative peculiari dei linguaggi di programmazione hanno attirato l'attenzione dei ricercatori del settore per utilizzarli come nuovi ambienti compositivi.

Questo capitolo è costituito dalla rassegna dei più rappresentativi linguaggi musicali che sono in uso ormai da molti anni.

Non si farà invece menzione dei *sequencer*, ossia di quegli applicativi basati sul sistema grafico del pentagramma e piano-roll, come CuBase e GarageBand, o dei *patcher* tipo Reaktor, con i quali è possibile configurare e rendere operativo un sintetizzatore virtuale “vecchio stile”.

I linguaggi di programmazione musicale si dividono in due categorie principali: *testuali* e *visuali* (o *grafici*). I linguaggi testuali consentono di descrivere algoritmi di sintesi e composizioni musicali attraverso linee di “testo scritto” in caratteri alfanumerici tipiche dei linguaggi di programmazione come il C, Java, Python ecc.

I linguaggi visuali consentono invece di descrivere algoritmi di sintesi e composizioni musicale attraverso il posizionamento su una finestra di lavoro con l'uso del mouse di elementi grafici che rappresentano oscillatori, generatori di inviluppo, filtri, cursori, manopole, ecc., e di farne gli opportuni collegamenti: il risultato finale è quello che viene chiamato un *patch*, il quale può essere attivato per produrre il risultato acustico/musicale desiderato.

In entrambi i casi si tratta di linguaggi concepiti e realizzati nelle loro prime versioni prototipali e commerciali molti anni fa: le versioni oggi disponibili sono il risultato dei continui aggiornamenti apportati via via negli anni, senza comunque introdurre novità sostanziali nella logica di funzionamento e nell'approccio operativo.

7.1 Linguaggi testuali

Il capostipite dei linguaggi musicali testuali è MUSIC V, realizzato negli anni '60 da Max Mathews (§4.1). L'erede diretto di MUSIC V è *C-Sound*, realizzato da Barry Vercoe al Massachusetts Institute of Technology, che è il primo linguaggio qui riportato (§7.1.1).

Il secondo ed il terzo linguaggio testuale qui riportati sono *SuperCollider* (§7.1.2) sviluppato a partire dagli anni '90 da James McCartney, e *Chuck* (§7.1.3) sviluppato da Ge Wang e Perry R. Cook alla Princeton University, che hanno le caratteristiche sintattiche dei linguaggi di programmazione standard.

Il quarto linguaggio è *pureCMusic* (§7.1.4), sviluppato all'interno dell'attività di ricerca sull'informatica musicale dell'ISTI-CNR di Pisa. Si tratta in questo caso di un linguaggio *embedded* e cioè ospite del linguaggio di programmazione general purpose C ed è di fatto una *framework*, come descritto in seguito.

Il quinto ed ultimo linguaggio è *Common Lisp Music* (§7.1.5), sviluppato da Bill Schottstaedt al CCRMA della Stanford University.

Per questi linguaggi si farà una descrizione delle funzionalità essenziali, rimandando il lettore ai manuali di riferimento per le specifiche e le modalità d'uso, avvertendolo comunque che per l'utilizzo di essi è necessaria una conoscenza dei concetti di base della programmazione algoritmica (qui di seguito sommariamente riportati) ed un minimo di esperienza esercitata con un qualunque linguaggio di programmazione.

Tutte le applicazioni che rendono operativo il computer, come un browser di navigazione internet (Safari, Chrome, Explorer), un WordProcessor, RealPlayer, PhotoShop, i video giochi e via dicendo, sono realizzate mediante l'uso dei linguaggi di programmazione come C++, Java, Python, ecc.

L'elemento principale di un linguaggio di programmazione, come abbiamo già visto nel Paragrafo 4.3 relativo all'oscillatore digitale, è l'*istruzione di assegnamento* che riflette in pratica il "mestiere" proprio del calcolatore e cioè: "fai qualcosa!" che è in generale un'operazione aritmetica del tipo

$$\text{alfa} = 3+4;$$

con il seguente significato operativo: il risultato dell'operazione 3+4 viene assegnato attraverso l'operatore = ad una *etichetta simbolica* "alfa" che contiene in questo caso il valore 7.

Nella terminologia della programmazione, l'etichetta simbolica viene detta *variabile* proprio perché può variare continuamente in valore numerico per effetto dell'esecuzione delle istruzioni del programma; ad esempio: dopo l'esecuzione dell'istruzione $\text{alfa}=10+\log_2(8)$; *alfa* contiene il valore 13.

Una variabile può essere anche una tabella e, come si è già visto nel Paragrafo 4.3 (Figura 4.2a), per individuare ogni elemento della tabella si fa uso di un *indice* come si fa con i numeri civici per individuare ogni singola abitazione di una strada. In questo caso l'etichetta è seguita da una coppia di parentesi quadre [] che racchiudono un numero o una variabile (semplice) che individua l'elemento.

Per esempio: se `beta` è la tabella

beta	15	0,10	5,21	3,34	8,5	0,5	-12	9,82	45,5	0,92	7,2
i	1	2	3	4	5	6	7	8	9	10	11

allora `beta[1]` vale 15; `beta[5]` vale 8,5; `beta[7]` vale -12; ecc., e dopo l'esecuzione di questo scorcio di programma

```
.....
ind=8;
val=beta[ind];
.....
```

`val` contiene il valore 9,82.

La scelta dell'etichetta che simboleggia una variabile è del tutto arbitraria, ma è buona regola inventarsi un nome che richiami mnemonicamente il significato che ha quella variabile nel contesto del programma.

Le istruzioni di un programma vengono eseguite in sequenza secondo l'ordine con cui sono state scritte, ma è possibile alterare in vari modi il flusso del programma e "dirottare" l'esecuzione su differenti parti del programma stesso in base all'esito dell'interrogazione di una certa condizione oppure ripetere ciclicamente un blocco di istruzioni.

Si tratta di costrutti definiti *strutture di controllo*, presenti più o meno nella stessa forma in tutti i linguaggi di programmazione, che danno la possibilità di realizzare procedure algoritmiche molto complesse. Vediamo due esempi che ne illustrano le funzionalità.

Supponiamo di voler caricare una tabella di 12 elementi con i valori delle frequenze di un'ottava dal La centrale pari a 440 Hz, al La dell'ottava superiore: per fare questo, oltre alla stessa tabella che definiamo `tabfreq` occorre una variabile `ind` per individuare tutti gli elementi della tabella e una di riferimento per l'incremento esponenziale relativo al temperamento equabile visto nel Paragrafo 1.6 dedicato alle scale musicali.

La struttura di controllo utilizzata per eseguire questa operazione complessa è il *for*, che ha un suo modo di funzionare secondo una precisa sintassi.

Lo scorcio di programma che prepara la tabella è il seguente:

```
rad12=1,0594631; //radice dodicesima di 2
tabfreq[1]=440;
for (ind=2; ind<13; ind=ind+1)
{
    tabfreq[ind]=tabfreq[ind-1]*rad12;
}
```

La prima istruzione predispone la variabile (anche se in realtà non subirà variazioni durante l'esecuzione) `rad12` al valore 1,0594631 (Paragrafo 1.6, Figure 1.15-16); la doppia barra `//` consente di scrivere un commento sempre utile in fase di rilettura del programma.

La seconda istruzione carica il primo elemento della tabella, con indice 1, al valore 440 che è quello della frequenza del La centrale.

La terza istruzione fa uso della struttura di controllo *for* che esegue ciclicamente il blocco di istruzioni (qui una soltanto, ma di solito molte) racchiuso tra parentesi graffe `{ }`. Il *for* è seguito da tre specifiche separate da un `;` e racchiuse tra parentesi tonde `()` dove:

- la prima specifica `ind=2` inizializza la variabile `ind` al valore 2;
- la seconda specifica `ind<13` esegue un confronto tra `ind` e 13:
se il risultato del confronto è *vero* (e questo si verifica per tutti i valori 2,3,4,5,6,7,8,9,10,11 e 12 di `ind`) allora l'istruzione del blocco `{ }` viene eseguita, *altrimenti* il ciclo è concluso e il flusso del programma continua dall'istruzione successiva al blocco stesso.
- la terza specifica `ind=ind+1` provoca l'effetto di dare a `ind` il valore di `ind` corrente +1, e cioè: ad ogni ciclo `ind` viene incrementato di 1.

La prima volta del ciclo, `ind` vale 2 e perciò l'istruzione del blocco `{ }`

```
tabfreq[ind]=tabfreq[ind-1]*rad12;
```

assegna alla posizione 2 di `tabfreq` il valore della precedente posizione (e cioè la posizione 1 individuata da `[ind-1]`) moltiplicata per il valore di `rad12`, che produce il valore della frequenza del semitono successivo: La#/Sib.

Terminata l'ultima istruzione del blocco `{ }`, il *for* esegue l'incremento dell'indice (che ora vale 3) il quale viene confrontato con 13; siccome è *vero* che 3 è minore di 13, il blocco viene eseguito e questa volta la terza posizione di

`tabfreq` viene caricata con il valore della frequenza della posizione precedente relativa a *La#/Sib*, che moltiplicato per `rad12`, produce quella del *Si*...

...e così via, fino al momento in cui `ind` vale 12, la posizione 12 di `tabfreq` viene caricata con il valore della frequenza (880) del *La* dell'ottava sopra, `ind` viene incrementato di 1 e perciò vale 13; siccome il risultato della condizione `13<13` è *falso*, il ciclo *for* è concluso, e la tabella è pronta.

Esiste sostanzialmente un'altra struttura di controllo per l'esecuzione ciclica di blocchi: *while*, che fa ripetere l'esecuzione del blocco associato (con un numero di cicli non prefissato) in base al risultato del confronto di variabili che possono cambiare di valore all'interno del blocco:

```
while (condizione)
{
    ...
}
```

Il secondo esempio mostra come creare un intervallo di terza maggiore o di terza minore a seconda del valore *vero/falso* di una variabile `magmin`:

```
.....
nota1=440;
if(magmin) nota2=nota1*rad12*rad12*rad12*rad12;
else      nota2=nota1*rad12*rad12*rad12;
.....
```

Una volta predisposta la variabile `nota1` al valore 440 del *La*, se `magmin` è vera, allora la variabile `nota2` viene messa al valore di `nota1` moltiplicato 4 volte (tante quanti sono i semitoni di un intervallo di terza maggiore) per `rad12` ottenendo 503,965Hz=*Do#*; altrimenti per 3 volte (tanti quanti sono i semitoni di un intervallo di terza minore) e cioè 475,682Hz=*Do*.

Esiste un'ulteriore struttura di controllo, *case-of* detta anche *switch*, che consente di scegliere l'esecuzione di un numero un numero di blocchi superiore a 2, che porta il controllo dell'esecuzione al blocco di caso corrispondente al valore corrente della variabile `num`:

```
switch (num)
{
    case 1: {... blocco1} :break
    case 2: {... blocco2} :break
    ....
    case n: {... bloccoN} :break
}
```


Con queste essenziali nozioni di programmazione algoritmica, possiamo realizzare ora l'oscillatore digitale sinusoidale descritto nel Paragrafo 4.3 e lo facciamo scrivendo due "scorci" di programma che verranno usati più avanti in modo formale.

È necessario innanzitutto creare la tabella di "look-up" che contiene i valori di un intero ciclo di sinusoidi che per il caso operativo scegliamo essere di 44100 valori.

```

.....
float TabellaSeno[44100];
int i;
float deltaOmega=2π/44100.0;
for(i=0;i<44100;i++)
    TabellaSeno[i]=sin(i*deltaOmega);
.....

```

La prima istruzione definisce, attraverso la direttiva `float`, la tabella `TabellaSeno` di 44100 locazioni, indirizzabili con l'indice intero `i`, in grado di memorizzare in ciascuna di esse un valore reale (e cioè con cifre dopo la virgola).

L'istruzione successiva assegna a `deltaOmega` un valore corrispondente alla 44100-esima parte dell'angolo giro misurato in radianti (§1.2). Il costrutto iterativo `for` con l'indice `i` che scorre da 0 a 44099, assegna alle locazioni di `TabellaSeno` i valori di un intero ciclo di sinusoidi calcolati con l'uso della funzione `sin` predefinita ed preesistente nel linguaggio di programmazione.

Lo scorcio di programma che segue realizza il meccanismo vero e proprio dell'oscillatore digitale:

```

.....
valore = tabellaSeno[parte intera di indice];
indice = indice + frequenza;
se (indice >= 44100)
    allora indice=indice-44100;
.....

```

La prima istruzione preleva il valore dalla tabella in base alla parte intera del valore corrente di `indice`; la successiva istruzione incrementa il valore di `indice` in base alla frequenza dell'oscillazione voluta riportando, questa volta, anche la parte decimale (dopo la virgola); il costrutto `se (...)` `allora` assicura lo scorrimento circolare della tabella perché quando il valore di `indice` supera la lunghezza della tabella, viene riportato opportunamente indietro.

7.1.1 CSound

In CSound la definizione di un brano musicale si fa in termini di *orchestre* e di *partiture*: un'*orchestra* è un insieme di algoritmi di sintesi che definiscono *strumenti* per mezzo di elementi predefiniti nel linguaggio, come oscillatori, generatori di involuppo, filtri, ecc.; una *partitura* è una successione di istruzioni che invocano l'attivazione degli strumenti definiti nell'orchestra.

Gli elementi di base per costruire algoritmi di sintesi si chiamano *Unit Generators* (detti anche moduli o *opcodes*) a ciascuno dei quali, in base alla funzione svolta, corrisponde un'istruzione costituita da un nome simbolico e dalla successione di parametri di ingresso. Nel caso dell'oscillatore digitale, ad esempio, l'opcode è *oscil* e la forma sintattica che lo implementa è

```
uscita oscil ampiezza,frequenza,tabella
```

A questo corrisponde la rappresentazione grafica di riferimento che segue:

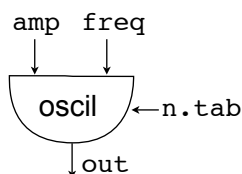


Figura 7.2 Rappresentazione grafica del modulo “oscil”

Ad ogni istante di campionamento, **oscil** produce un valore utilizzando i valori correnti di ampiezza, frequenza e numero di tabella, valore che viene assegnato alla variabile **uscita**. Dal momento che questo opcode funziona con il meccanismo dell'oscillatore digitale, è necessario prima definire la tabella di look-up da cui leggere i valori con passo proporzionale alla frequenza voluta ed in riferimento ad una prefissata frequenza di campionamento (§4.3). Per generare la tabella di look-up si fa uso di moduli di generazione detti *genN* che non vengono tuttavia invocati direttamente, ma si fa riferimento ad essi attraverso l'istruzione **f** in base alla seguente sintassi:

```
f num t lungh genN par1 par2 par3 ... dove
```

num è il valore di riferimento usato da **oscil**;
t è il tempo in secondi di attivazione della tabella;
lungh è la lunghezza desiderata della tabella;
genN è il valore che corrisponde alla tipologia di generazione della tabella;
parN sono i valori degli eventuali parametri necessari alla *genN*.

Ad esempio: `f 7 0 32768 10 1`

genera al tempo 0 una tabella di 32768 campioni relativi ad una sinusoide di ampiezza 1 e con numero di riferimento 7; è il numero 10 che richiama la **gen10** per generare una sinusoide ed il valore 1 ne determina l'ampiezza. Se a questo valore si fanno seguire altri valori come in quest'altro esempio

`f 3 0 32768 10 1 0.5 0.333 0.25`

allora la forma d'onda generata e memorizzata nella tabella (in questo caso etichettata con il numero 3) è quella risultante dalla somma dei primi 4 armonici con i valori specificati. Oltre alla modalità 10 (**gen10**) ne esistono altre che consentono di generare tabelle nei più svariati modi, incluso quello (**gen1**) di caricare un file sonoro di tipo wave, aiff, ecc.

Un altro opcode di fondamentale importanza è il generatore di involuppo **linen** (abbreviazione di *linear envelope*) che ha la seguente sintassi:

`uscita linen ampiezza, attacco, durata, rilascio`

cui corrisponde la rappresentazione grafica di riferimento

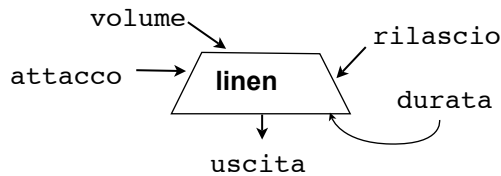


Figura 7.3 Rappresentazione grafica del modulo "linen"

Esso produce un valore variabile nel tempo, che viene assegnato ad **uscita**, utilizzando i valori correnti di ampiezza, attacco, durata e rilascio relativi ad una forma di involuppo elementare come quello in Figura 7.4:

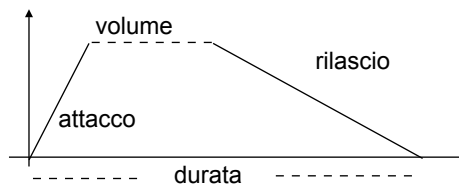


Figura 7.4 Forma elementare di involuppo

Per realizzare uno strumento, e cioè un algoritmo di sintesi, è necessario scrivere linee di istruzioni che determinano la presenza degli elementi nell'algoritmo stesso e dei collegamenti tra essi.

Il seguente semplice strumento costituito da un oscillatore involupato

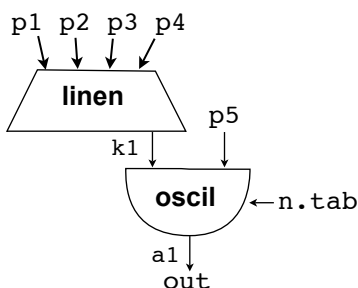


Figura 7.5 Esempio di strumento CSound

viene realizzato in CSound con le istruzioni:

```

          instr 7
k1      linen p4,p7,p3,p8
a1      oscil k1,p5,p6
          out   a1
          endin

```

La prima linea `instr` stabilisce l'inizio della definizione dello strumento seguito da un numero di riferimento (7 nell'esempio), che può essere un qualunque valore e che identifica quello strumento tra tutti gli altri che faranno parte dell'orchestra; la definizione dello strumento finisce con `endin`.

L'opcode `linen` è seguito da 4 parametri `pN` di cui si dirà tra poco; per il momento è importante notare che l'etichetta `k1` è presente sia nella parte "risultato" di `linen` sia nella parte "parametri" della successiva istruzione `oscil` il che stabilisce un collegamento tra l'uscita del generatore di involuppo e l'ingresso del controllo di volume dell'oscillatore. Il valore prodotto da `oscil` viene raccolto nella variabile `a1` che l'opcode `out` si incarica di mandare in uscita per la conversione dal digitale all'analogico, operazione che di fatto provoca l'emissione sonora dello strumento.

I parametri `pN` servono per il collegamento tra strumento e la seconda parte di cui è costituito un brano CSound e cioè con la partitura (lo *score*); lo *score* consiste di una serie di istruzioni (tante quante sono le note o più in generale i *suoni* del brano stesso) che hanno il seguente formato:

	i	7	0	2	10000	440	1	0.2	1.2
num d'ordine parametri		<i>p1</i>	<i>p2</i>	<i>p3</i>	<i>p4</i>	<i>p5</i>	<i>p6</i>	<i>p7</i>	<i>p8</i>
significato		<i>str</i>	<i>tempo</i>	<i>dur</i>	<i>amp</i>	<i>freq</i>	<i>n.tab</i>	<i>attacco</i>	<i>rilascio</i>

i è di fatto l'opcode che effettua la chiamata alla strumento (*instrument*) indicato in *p1* (in questo caso =7) per l'attivazione di un suono con caratteristiche definite dai valori parametrici che seguono. I parametri *p1* *p2* e *p3* sono riservati ed hanno sempre il significato di

- p1* numero di strumento dell'orchestra invocato
- p2* tempo di attivazione
- p3* durata dell'evento (nota o suono)

L'ordine degli altri parametri è libero, ma è necessario che venga rispettata la corrispondenza tra il numero del parametro presente nello strumento e la posizione dei valori parametrici presenti nelle istruzioni *instrument* dello score. L'opcode **f** che definisce le tabelle fa parte dello score per cui una partitura che genera tre suoni, ad esempio, con frequenze 110, 220 e 440 Hz di volume (rispettivamente) massimo, medio e basso e con durata di un secondo con lo strumento 7 definito in precedenza, è

```

f 3 0 32768 10 1 0.5 0.333 0.25
i 7 0 1 32000 110 3 0.2 0.8
i 7 1 1 10000 220 3 0.1 0.5
i 7 2 1 5000 440 3 0.4 0.8
  p1  p2  p3  p4  p5  p6  p7  p8
  str tempo dur amp freq n.tab attacco rilascio
end

```

Il parametro *p2* con il quale si definisce il tempo di attivazione è quello che consente anche l'esecuzione di più suoni contemporanei, come succede per gli accordi o più in generale per più parti musicali separate.

Nel caso di brani con note da pentagramma, è possibile utilizzare una speciale codifica simbolica (partendo dalla quale il linguaggio genera il corrispondente valore frequenziale) oppure utilizzare applicativi denominati *MIDI-to-CSound* facilmente reperibili in rete, che a partire da un MIDI file producono un file costituito da linee *instrument* contenenti già tutte le necessarie informazioni.

Un modo alternativo di usare CSound è quello di far generare le linee *instrument* per mezzo di programmi scritti in un linguaggio di programmazione qualsiasi come il C, il Basic, ed in particolar modo il Python, ormai il più diffuso e popolare. In questo modo è possibile svincolarsi dal tradizionale concetto di composizione musicale ed allargare i confini della creatività per mezzo della *composizione algoritmica*, cui faremo cenno nell'ultimo paragrafo di questo capitolo, utilizzando tutte le peculiarità operative delle tecniche di programmazione.

Nel corso degli anni, una numerosa comunità internazionale di utenti ha contribuito ad arricchire la già ricca libreria di funzionalità di base fatta di generatori, filtri, ecc. con i quali implementare tutte le tecniche di sintesi note (Capitolo 4) ed i più svariati effetti come riverbero, flanger, ecc.

La disponibilità di computer veloci dà oggi la possibilità di utilizzare CSound in tempo reale per la maggior parte di composizioni. E comunque, nel caso in cui il computer utilizzato non sia in grado di calcolare in tempo reale il segnale a causa della complessità ed il numero di instrument presenti nella composizione, Csound consente di memorizzare off-line il segnale su un file tipo .wave e di ascoltarlo in tempo differito alla fine dell'elaborazione.

Dal sito ufficiale di Csound [1] è possibile scaricare gratuitamente l'ultima versione del linguaggio funzionante per tutti i sistemi operativi (Windows, MAC-OS, Linux) ed utilizzarla liberamente sul proprio computer. Inclusa nel linguaggio c'è anche una guida ipertestuale consultabile in linea, relativa a tutti gli opcode raggruppati per famiglie con la descrizione dettagliata della funzione svolta e dell'uso dei parametri associati.

Alcune di queste versioni includono la possibilità di definire un'interfaccia grafica denominata *widget*, sulla quale posizionare cursori e manopole di vario genere, il cui effetto si riflette, mediante l'uso di un opportuno opcode di collegamento, sulla parametria dello strumento.

Fanno parte del pacchetto anche molti esempi e composizioni di vari autori che possono servire da esempi da cui prendere le mosse per fare le proprie composizioni. Particolari collegamenti rimandano alle figure di Max Mathews, Barry Vercoe e Richard Boulanger: quest'ultimo è un musicista che ha dedicato la sua vita alla composizione ed alla didattica.

A lui si deve un efficace tutorial (incluso anch'esso nel pacchetto) che, partendo da esempi elementari come quello descritto in questo paragrafo, porta gradualmente l'utente alla comprensione di situazioni compositive di complessità crescente. Boulanger è anche autore di un libro edito dalla MIT-Press dal titolo CSOUND [2], che è il libro di riferimento completo di questo linguaggio musicale.

Esiste anche un libro in italiano di (che non è la traduzione di CSound di Boulanger) ma che è altrettanto puntuale, didatticamente efficace e ricco di esempi e di documenti relativi ad esperienze compositive dei musicisti italiani Bianchini e Cipriani [3].

Va infine segnalato il nome di Gabriel Maldonado come il referente italiano di CSound: al sito [4] è possibile scaricare la versione per Windows che include molte funzioni di base per generare, assieme al suono, grafica sofisticata in realtime.

7.2 Linguaggi visuali

I linguaggi musicali descritti nei prossimi paragrafi sono MAX/MSP e PureData (§7.2.1) e Kyma (§7.2.2). Essi vengono definiti *linguaggi grafici* o *linguaggi visuali* poiché consentono di descrivere algoritmi di sintesi e composizioni musicali attraverso il posizionamento su una finestra di lavoro di elementi grafici, che rappresentano oscillatori, involucri, filtri, cursori, manopole, e di farne gli opportuni collegamenti. Kyma e MAX/MSP sono prodotti commerciali, mentre PureData, che di fatto è la versione “spartana” di MAX è scaricabile gratuitamente dalla rete come i linguaggi testuali descritti nei precedenti paragrafi.

7.2.1 MAX/MSP e Pure Data

Il nome di questo linguaggio musicale visuale è un omaggio a Max Mathews. La prima versione di MAX venne realizzata da Miller Puckette all'IRCAM di Parigi nel 1988 su macchine Macintosh ed era orientata esclusivamente alla gestione di eventi MIDI, mentre la generazione del suono era delegata ad expander MIDI esterni. In seguito Miller Puckette e David Zicarelli realizzarono versioni del linguaggio che proponevano un ambiente integrato per la gestione in tempo reale anche di segnali audio. Questa nuova versione venne chiamata MaxSignalProcessing da cui la sigla MAX/MSP.

Il concetto fondamentale in MAX è il *patch*, ovvero un insieme di oggetti elementari collegati tra loro che si scambiano messaggi: un oggetto riceve messaggi ai quali reagisce generando altri messaggi destinati ad altri oggetti. Il flusso dei messaggi, o *data-flow*, avviene attraverso linee di connessione. È possibile inoltre che un oggetto generi messaggi in maniera temporizzata o che riceva messaggi MIDI dall'esterno generati da un performer con l'uso di controller MIDI.

Quando MAX fu presentato ufficialmente la prima volta, ci fu un certo dibattito sollevato dai puristi dell'informatica, che ritenevano, a ragione, che le parole chiave del linguaggio “oggetto”, “messaggio”, “data-flow” e, come vedremo più avanti, “incapsulamento”, fossero usate impropriamente rispetto al loro significato formale. Tale dibattito fu riportato nell'articolo *Putting Max in Perspective* [11] che è ora consultabile in rete in [12]. L'articolo riporta anche l'intervento di risposta dello stesso Puckette che, in maniera tanto serafica quanto pratica ed operativa, risponde: *quando ho progettato MAX ho preso in considerazione prima di tutto l'affidabilità e poi la semplicità e l'espressività; tutto il resto riguarda soltanto l'informatica*. E bisogna ammettere che il successo commerciale e di popolarità che MAX ha ottenuto negli anni gli hanno dato ragione.

MAX è un linguaggio interpretato e sempre attivo, sia nella fase di editing che in quella operativa vera e propria. Infatti un patch può essere in modo *edit* o in modo *run*: in modo edit il mouse e la tastiera vengono usati per posizionare oggetti e stabilire collegamenti tra essi per creare il patch stesso; in modo run il mouse e la tastiera vengono usati per intervenire sui diversi controlli del patch, come pulsanti, interruttori, cursori, manopole, ecc.

La Figura 7.6 mostra un patch MAX che sintetizza due suoni sinusoidali controllabili in frequenza manualmente e separatamente.

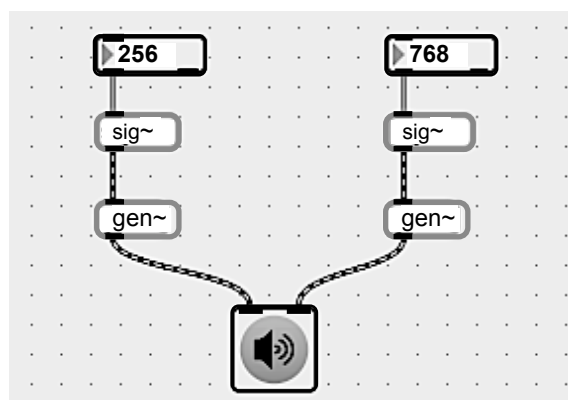


Figura 7.6 Patch MAX per la sintesi di due suoni sinusoidali

Gli oggetti MAX sono sostanzialmente figure rettangolari che in certi casi riportano all'interno una figura rappresentativa della funzione svolta, come ad esempio quella dell'altoparlante. I rettangoli/oggetti hanno sul bordo dei trattini in neretto che funzionano da "spinotto" di ingresso/uscita ai quali agganciare le linee di collegamento dati: quelle superiori sono per l'ingresso dati e quelle inferiori per l'uscita. Esistono due tipi di collegamento: dati/evento costituite da linee piene, e dati/segnale costituite da linee tratteggiate giallo/neri. Max è infatti un linguaggio visuale molto "colorato" e non tanto per ragioni estetiche quanto per ragioni pratiche: quando i patch sono densi di oggetti e di collegamenti, e questa è la norma, il colore diventa un fattore di fondamentale importanza per raccapezzarsi.

L'oggetto altoparlante ha due ingressi segnale per i canali audio di destra e di sinistra. Il segnale sinusoidale viene in questo caso generato dall'oggetto *gen~* che riceve in ingresso il valore della frequenza in termini di *passo-discansione*, come visto nella descrizione dell'oscillatore digitale (§4.3); la sintesi e l'elaborazione dei segnali viene infatti gestita da un insieme di oggetti identificati con il simbolo ~ (tilde) dall'ovvio significato iconico che gli oggetti

comunicano fra loro attraverso collegamenti di tipo segnale. Il valore del passo-di-scansione in ingresso a *gen~* viene calcolato dall'oggetto *sig~* che riceve in ingresso il valore della frequenza in Hz; quest'oggetto fa parte di quella categoria di oggetti che raccordano l'ambiente "eventi" con quello "segnali".

L'oggetto bordato in neretto e contrassegnato all'interno con un triangolino viene infatti controllato manualmente per mezzo del mouse: cliccando all'interno appare il cursore che consente di immettere un valore con la tastiera del computer; oppure cliccando sull'oggetto e tenendo premuto il mouse, si fa variare il contenuto numerico mediante scorrimento verticale del mouse, intervenendo sull'ordine di grandezza del numero con lo spostamento orizzontale del mouse.

Il patch mostrato in Figura 7.7 fa uso dell'algoritmo di Karplus-Strong visto nel Paragrafo 4.9.4; si deve notare che anche qui, come nel precedente patch, viene usato *gen~* perché in realtà quest'oggetto rappresenta un'intera classe di generatori all'interno della quale è possibile selezionare quello voluto.

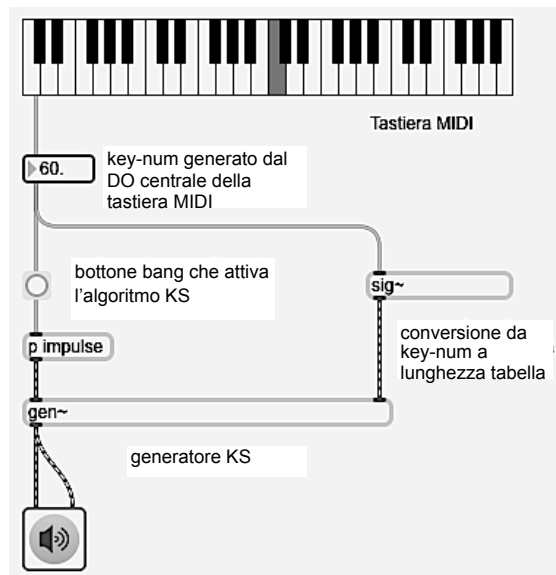


Figura 7.7 Patch MAX per la sintesi di due suoni sinusoidali

In questo caso l'uscita del segnale viene inviato ad entrambi i canali destro e sinistro dell'altoparlante. Karplus-Strong ha due ingressi: il primo, a sinistra, riceve il segnale di attivazione generato dall'oggetto *impulse* che a sua volta viene attivato dall'oggetto *bang* ogni volta che viene "premutato" un tasto sulla tastiera MIDI; ed il secondo, a destra, relativo alla frequenza data in termini di lunghezza di tabella. In particolare, quando con il mouse viene "premutato" un

tasto della tastiera MIDI, questa genera il messaggio MIDI corrispondente al numero di tasto (DO=60, in questo caso) che viene convertito dall'oggetto *sig~* da valore numero in segnale.

MAX dà anche la possibilità di inserire commenti che facilitano la rilettura del patch per modifiche ed estensioni successive. Un patch può essere incapsulato, secondo la terminologia di MAX, in un unico oggetto eventualmente replicabile in più esemplari per costruire patch più complessi che realizzano un "orchestra di strumenti". L'operazione di incapsulamento corrisponde in realtà, secondo una più formale terminologia informatica, a quella di raccolta in *sotto-programma* di una patch con funzionalità complessa.

MAX è un prodotto commerciale acquistabile via rete dalla Cycling'74 all'indirizzo [13] ed è comunque possibile scaricare gratuitamente una copia di MAX valida per un breve periodo di tempo, sufficiente per poter familiarizzare ed apprendere tutte le possibilità operative. Esiste anche un manuale di MAX in italiano che affronta l'argomento in maniera graduale e sistematica e riporta anche tanti esempi di patch e composizioni a complessità crescente, utilissimi sia in fase di apprendimento che come modelli per da cui prendere le mosse per le proprie composizioni [14].

Miller Puckette ha realizzato anche una versione alternativa al linguaggio, del tutto gratuita ma che offre più o meno tutte le potenzialità creative di MAX: si tratta di Pure Data (PD) che si presenta con una grafica più essenziale ma con la stessa logica di editing. La Figura 7.8 mostra il patch dell'algoritmo di sintesi FM presente nel manuale e nei tutorial che accompagnano il pacchetto scaricabile dalla rete [15].

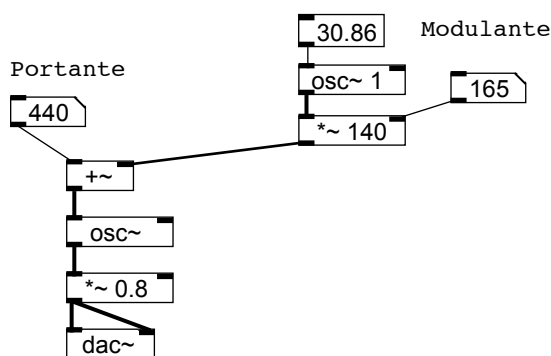


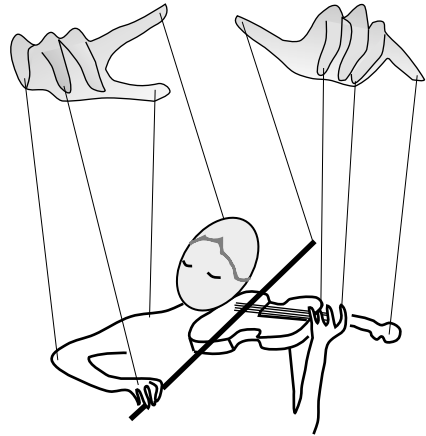
Figura 7.8 Patch PD di Modulazione di Frequenza

In PD tutti gli oggetti hanno la stessa forma "spartana" di semplici rettangoli: un'etichetta simbolica ne specifica la funzione. In particolare, gli oggetti "signal" sono caratterizzati dalla presenza di una *~*. Collegamenti "dati" vengono fatti con *linee sottili* mentre quelli "segnale" con *linee più larghe*.

Interazione



- 8.1 Dispositivi di input e interfacce gestuali
- 8.2 Mapping
- 8.3 Esecuzione e gestualità



Negli ultimi anni si è fatto un grande uso dei termini *multimedialità* ed *interattività* in seguito al diligente impiego della tecnologia informatica sulla miriade di attività umane che vanno dal lavoro, alle tele-comunicazioni, alla ricerca, al gioco. Il termine *multimediale* assume molteplici connotazioni e viene usato in relazione alla quasi totalità delle attività condotte con un qualsiasi apparato frutto della tecnologia informatica [1][2].

C'è da osservare innanzitutto che la parola *multi-media* è di origine latina, poi re-importata dal mondo anglosassone; e inoltre, che il successo di questo neologismo non è commisurato alla discrepanza tra etimologia e semantica del termine. Infatti, se si intende *medium* nel senso di *veicolo di contenuto artistico*, allora multi-media non è specifico della tecnologia digitale perché già il Cinema, l'Opera Lirica, il Teatro dell'Arte erano e sono forme di espressione "multimediali". E se si intende *medium* nel senso di *strumento o attrezzo*, dal momento che i diversi contenuti trattati (fotografie, filmati, testi, animazioni, parola e musica, ecc.) sono tutti rappresentati in formato digitale, e perciò omogenei tra loro, l'attrezzo, lo strumento, il medium che li genera e li assembla è *uno solo*: il computer. E allora sarebbe più appropriato il termine *mono-medialità*.

L'altra parola legata al mondo dell'informatica e, più in generale, alla tecnologia digitale è *interattività* che viene usata nel settore delle Arti Contemporanee per designare sia una *installazione interattiva* sia una *performance interattiva*.

Nell'installazione interattiva l'artista utilizza sensori elettronici di varia natura per coinvolgere le persone del pubblico, invitandole a scoprire gli effetti reattivi provocati dalla loro presenza attiva sugli elementi visivi e sonori che costituiscono l'installazione stessa: in questo caso l'*inter-azione* è parte dell'opera stessa.

Nella performance interattiva, invece, è l'artista che interagisce con il sistema informatico per eseguire la sua composizione musicale, o più in generale (per restare in linea con la dizione corrente), la sua composizione multi-mediale. In questo capitolo ci si occupa della *performance musicale interattiva*.

Nella musica della tradizione, di qualunque genere si tratti, suonare un brano di musica vuol dire, nell'essenza, generare suoni con appositi "attrezzi meccanici" controllati da una o più parti del corpo: bocca, braccia, mani e piedi. A questo proposito Marco Tiella, nel suo libro *L'officina di Orfeo* [3] dice che "...la musica, astratta per definizione, si materializza negli arnesi che traducono i testi musicali in suono."

Le azioni attraverso le quali i suoni vengono generati, vale a dire le modalità con le quali vengono scelti ed emessi e con le quali viene data loro espressione, possono avvenire in maniera integrata o distribuita a seconda dello strumento musicale in uso.

Ad esempio, nel pianoforte le dita eseguono in maniera integrata i compiti di selezionare una nota e di darle espressione; nella chitarra e negli strumenti ad arco i due compiti sono distribuiti rispettivamente tra la mano sinistra e la mano destra; negli ottoni (tromba, bassotuba, ...) sono essenzialmente le labbra ad avere il compito di selezionare le note e di controllarne l'espressione; negli strumenti ad ancia (clarinetto, sassofono, oboe...) le note sono principalmente stabilite attraverso combinazioni di chiavi aperta/chiusa operate dalle dita di entrambe le mani, mentre l'espressione viene controllata dai sistemi orale e respiratorio.

Uno strumento musicale tradizionale è perciò un "arnese compatto" che propone, nelle caratteristiche costruttive della forma e della meccanica, sia la possibilità di generare le note musicali stabilite dal temperamento equabile, sia l'adattamento ergonomico alla struttura del corpo umano per assicurarne la reale usabilità e il controllo espressivo. Molto probabilmente, una civiltà aliena potrebbe ricostruire la struttura e le funzionalità del corpo umano dall'analisi comparata della totalità degli strumenti musicali.

Una specifica azione effettuata su uno strumento musicale produce *sempre* lo stesso risultato sonoro, tant'è che imparare a suonare uno strumento vuol dire proprio imparare a coordinare le parti interessate del corpo per esercitare le corrette azioni desiderate su di esso.

Ciò vuol dire che esiste una relazione *uno-a-uno* tra le azioni effettuate ed i suoni ottenuti. Ad esempio, nel pianoforte la pressione di un tasto genera sempre la stessa nota con intensità proporzionale alla pressione del dito.

Anche una tastiera elettronica, che simula e ripropone le funzionalità di un pianoforte vero, si presenta come uno strumento “compatto”. Ma in realtà essa è un sistema composto da tre componenti principali di diversa natura, fisicamente separati l'uno dall'altro e collegati “in catena” con protocolli di comunicazione anche di natura diversa, come mostrato nella Figura 8.1:

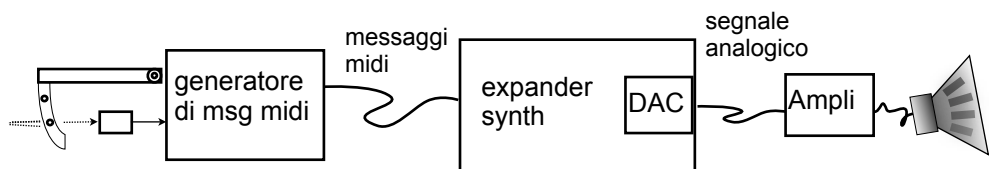


Figura 8.1 Dispositivi e protocolli di comunicazione in una tastiera elettronica

- 1) ad ogni tasto della tastiera è associato un dispositivo che rileva la modalità con cui viene premuto, tramite calcolo del tempo (e quindi della velocità) tra il passaggio tra due marker successivi, che viene utilizzata per generare un messaggio MIDI contenente l'informazione di *key-number* e *velocity*;
- 2) il messaggio viene trasferito tramite interfaccia MIDI ad un expander che effettua la sintesi per campionamento, per modulazione di frequenza o altre tecniche;
- 3) il segnale digitale generato viene convertito in analogico e mandato ad un impianto di amplificazione e di diffusione sonora.

Se nel pianoforte classico il percorso dell'*informazione* - relativa alla pressione del dito sul tasto che si riflette sul martello di feltro che percuote la corda che genera il suono - è breve, diretto e esclusivamente meccanico, in una tastiera elettronica tale percorso è molto più lungo e più complesso, perché realizzato con l'uso di dispositivi e messaggistiche diverse.

La divisione in tre elementi principali (Figura 8.1) su cui si basa l'architettura di una tastiera elettronica commerciale, fin qui definita "compatta", consente di realizzare un sistema "esploso" composto da tre dispositivi realmente separati ed opportunamente collegati tra loro: è infatti possibile assemblare sistemi elettronici di sintesi sonora utilizzando dispositivi come *tastiere-mute* che generano soltanto messaggi MIDI ed *expander/synth* di una qualunque casa costruttrice con modalità di sintesi diverse, come quelli prodotti in gran quantità in special modo negli anni '80 e '90.

Oggi, in alternativa agli expander, è possibile utilizzare un computer sul quale attivare applicazioni *virtual expander* che ripropongono "a software" le funzionalità di un sintetizzatore MIDI fisico con il vantaggio di poter utilizzare diverse tipologie di sintesi con la stessa configurazione di sistema come mostrato di Figura 8.2:

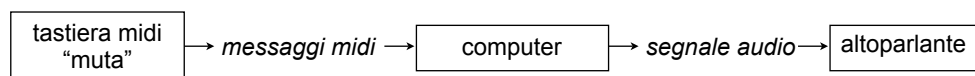


Figura 8.2 Sistema tastiera_muta/computer

In tutte e tre le situazioni (tastiera compatta, sistema tastiera-muta ed expander/synth, sistema tastiera-muta e computer) si hanno comunque sempre le stesse possibilità creative ed espressive (ancorché ampliate con timbriche ed effetti tipici della tecnologia elettronica ed informatica) di quelle offerte da un pianoforte reale: contesto melodico/armonico caratteristico delle dodici note del temperamento equabile e relazione uno-a-uno tra tasto e suono.

Lo schema di Figura 8.2 che mette in relazione, attraverso il ruolo attivo del computer, i messaggi generati dalla tastiera MIDI e il risultato musicale, viene perciò generalizzato con quello di Figura 8.3 che illustra l'essenza dello *strumentario informatico* usato nelle performance a controllo interattivo:

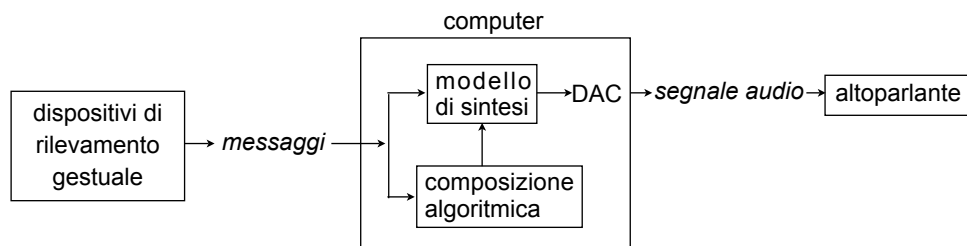


Figura 8.3 Lo strumentario musicale informatico

Le informazioni provenienti da un controllore (definito anche *dispositivo di rilevamento gestuale* o *sistema di rilevamento gestuale* del performer) vengono utilizzate in tempo reale dai programmi attivi sul computer per influenzare sia la composizione algoritmica al macro livello degli eventi, sia gli algoritmi di sintesi e di elaborazione sonora (micro livello), anche in modo indiretto attraverso la composizione algoritmica stessa.

È proprio in presenza di quello che è stato definito *ruolo attivo* del computer che si aprono possibilità creative ed espressive del tutto nuove. E del tutto nuove sono anche le modalità di intendere procedure, prassi e convenzioni consolidate nella musica della tradizione relative alla composizione, alla trascrizione su pentagramma, all'esecuzione, all'esecutore, ecc.

A questo proposito va osservato che ci sono per lo meno tre punti distinti da prendere in considerazione: 1) *dispositivi di input*, 2) *mapping* e 3) *meta-composizione*, che sono argomento dei prossimi paragrafi.

8.1 Dispositivi di input e interfacce gestuali

Una prima categoria di apparati atti all'interazione è quella costituita dai controller MIDI che diverse case costruttrici di strumenti musicali elettronici hanno via via realizzato ed immesso sul mercato. Prima fra tutte, la tastiera tipo pianoforte dotata di sensori che rilevano la velocità di pressione dei singoli tasti, e di solito equipaggiata con altri sensori per il bender e l'after-touch. In seguito sono state realizzate diverse tipologie di dispositivi differenti dalla tastiera in grado di generare tutta la messaggistica MIDI attraverso l'uso di pulsanti, manopole, cursori a slitta, sensori di percussione, touch-pad, ecc., denominati *controller midi*. È sufficiente digitare controller midi come parole chiave su un browser di ricerca in rete, per ottenere informazioni ed immagini su questi dispositivi di input.

Parallelamente a questi prodotti commerciali, sono stati realizzati negli anni '80 e '90 anche numerosi dispositivi e sistemi originali di input - spesso in modo prototipale ed in esemplari unici, grazie alla disponibilità sul mercato di una vasta gamma di componenti elettronici ed elettro-meccanici, che va sotto il nome di *sensoristica* - atti alla rilevazione di informazioni relative alla presenza ed ai movimenti del corpo umano.

Tali componenti sono stati utilizzati per progettare e realizzare dispositivi e sistemi originali denominati *interfacce gestuali*. Quella di inventare e realizzare nuovi dispositivi di interazione è stata infatti un'attività altrettanto perseguita di quella di realizzare modelli ed algoritmi di sintesi (Capitolo 4) e linguaggi musicali (Capitolo 7).

Le tecnologie sono quelle utilizzate nel settore del controllo automatico e robotico: raggi infrarossi, sonar, onde elettro-magnetiche, rilevatori di pressione ed elaborazione di immagini in tempo reale. Vengono qui riportati alcuni tra i molti dispositivi e sistemi realizzati negli anni nei vari centri di ricerca a livello internazionale, indicandone soltanto autori e caratteristiche principali, e rimandando il lettore ad una più approfondita analisi attraverso una ricerca in rete delle pubblicazioni apparse su riviste ed atti di congresso specializzati, ed immagini e video-clip che ne mostrano le funzionalità.

Primo fra tutti va ricordato il *Sequential Drum* (seguito poi dal *Radio Drum*) che Max Mathews realizzò assieme a Curtis Abbot, consistente in una tavola rettangolare percuotibile con bacchette di legno in grado di generare informazioni relative alla posizione x-y del punto ed all'intensità di battuta.

The Hands, realizzato dal Michael Waisvitz dello STEIM di Amsterdam, è un congegno antesignano del Data Glove, in grado di generare informazioni dal movimento delle dita e dalla rotazione nello spazio delle mani.

Un caso particolare è quello costituito dalla bacchetta del direttore d'orchestra, in inglese *baton*. Esistono per lo meno 4 differenti versioni realizzate indipendentemente l'una dall'altra in differenti centri di ricerca:

- *Radio Baton*, realizzata da Max Mathews e Robert Boie nel 1987, era in grado di misurare la posizione nello spazio di due bacchette contenenti trasmettitori radio rispetto ad una superficie ricevente.
- *Lightning System*, prodotto commercialmente da Donald Buchla, usava led a infrarossi per misurare la posizione della bacchetta, con output MIDI programmabile secondo i movimenti dell'esecutore.
- *Digital Baton*, realizzato da Theresa Marrin al MIT Media Lab, utilizzava accelerometri per individuare la posizione nello spazio e sensori di forza sull'impugnatura per controlli vari.
- *Light Baton*, realizzata da Paolo Carosi e Graziano Bertini presso il CNR di Pisa, usava una telecamera ed una scheda di digitalizzazione video per rilevare la posizione bidimensionale della punta della bacchetta da cui estrarre parametri relativi alla traiettoria e alla velocità di movimento.

Diversi sistemi si basano su sensori a luce infrarossa. *Palm Drive* è un controller realizzato presso il CNR di Pisa (1998-2004) che utilizza un insieme

di 24 sensori disposti su un supporto planare, in grado di misurare distanza ed inclinazione su due assi delle due mani dell'esecutore.

Recentemente è stato messo in commercio il dispositivo *Leap* in grado di ricostruire l'intera forma della mano in modo molto preciso e reattivo.

Altri sistemi sono basati sull'analisi in tempo reale di immagini provenienti da una telecamera ed opportunamente digitalizzate: *BigEye*, realizzato allo STEIM di Amsterdam e *EyesWeb*, del Dipartimento di Informatica Musicale dell'Università di Genova, sono due sistemi per l'estrazione di informazioni di controllo dai movimenti della figura umana.

In particolare, *EyesWeb* è una piattaforma gratuita e aperta costituita da potenti librerie di oggetti dedicati all'acquisizione e all'analisi di immagini in movimento di figure umane all'interno di uno spazio di ripresa da telecamera.

A livello commerciale va segnalato il dispositivo *Kinect* che la Microsoft ha prodotto, in origine, per il controllo della X-Box: questo dispositivo è in grado di ricostruire la sagoma "a fiammifero" di una persona (anzi, fino ad un massimo di 3 persone), per il quale esiste una libreria di funzioni scaricabile dalla rete, in grado di generare tutte le informazioni relative alla posizione dei singoli arti della persona (o delle persone) in movimento nello spazio d'azione tri-dimensionale di fronte al dispositivo.

Esiste poi la pratica ormai consolidata del "fai-da-te" rivolta alla realizzazione di interfacce gestuali originali e personali da utilizzare nelle più disparate situazioni creative di performance interattive, grazie alla disponibilità sul mercato di una vasta gamma di sensoristica a basso prezzo (sensori distanza, di temperatura, di movimento, di pressione, ecc.) e di dispositivi di controllo basati su elettronica digitale.

Tra tanti, il più popolare ed utilizzato è *ARDUINO* di produzione italiana che, come recita la pagina iniziale del sito [3], è una *piattaforma open-source per la prototipazione elettronica, basata su hardware e software flessibili e di semplice utilizzo*. Si tratta infatti di una scheda basata su un *microcontrollore* programmabile da computer cui è connesso via USB.

Arduino è equipaggiato di convertitori analogico/digitali ai quali è possibile collegare le diverse tipologie di sensori, e di convertitori D/A che consentono di controllare dispositivi esterni come LED, motori ecc. Una vastissima libreria di funzioni scritte in C consente la realizzazione di sistemi automatici complessi originali e personali.

Esiste poi un'altra categoria di dispositivi dedicati esclusivamente all'informatica musicale.

Un esempio è costituito da *Eobody SensorBox* [5] che è in grado ricevere in ingresso i segnali analogici di sensori di varia natura e di generare uno streaming di messaggi MIDI corrispondenti.

Questo dispositivo non è programmabile ed è più semplice da utilizzare: i vari sensori si connettono al SensorBox con spinotto jack tipo chitarra, e il SensorBox genera messaggi MIDI che vengono inviati al computer tramite interfaccia USB.

I messaggi possono così essere utilizzati da qualunque patch o composizione realizzata con i linguaggi musicali MAX, PD, SuperCollider, ecc.

8.2 Mapping

Tutta la Musica della tradizione si fonda sulla relazione uno-a-uno tra azioni esercitate sugli strumenti musicali e suono prodotto, che riflette sia la grafia del pentagramma sia l'alfabeto delle dodici note del temperamento equabile. Inoltre, proprio perché gli strumenti musicali sono attrezzi meccanici e proprio perché la generazione del suono dipende dalle caratteristiche fisiche dei materiali utilizzati per la loro realizzazione, le sfumature di colore timbrico e la pronuncia di note singole e di intere frasi musicali dipendono dalla diretta e meccanica interazione tra corpo umano e strumento.

Quando un musicista esegue un brano musicale leggendo lo spartito, oppure a memoria, oppure improvvisando, “esegue” simultaneamente per lo meno due *procedure*, che nel gergo informatico si dicono *processi con-correnti*: quella più propriamente musicale melodico/ritmico/armonico e quella più propriamente fisiologica che mette in funzione la muscolatura delle parti del corpo in contatto fisico con lo strumento.

Lo strumento è per l'appunto il mezzo con cui il/la musicista dà vita alla musica e, anche se può sembrare banale sottolinearlo, si osserva che i due processi, il fluire della musica e l'attività neurofisiologia, accadono in un “olismo zenico” “dentro” di lui/lei.

Qui la separazione tra musicista e strumento è netta. Ma nel “nuovo strumentario musicale”, come illustrato nella Figura 8.3, i dispositivi che costituiscono l'interfaccia tra il performer e il sistema informatico non vengono collegati direttamente ai generatori sonori elettronici: in mezzo c'è il computer che interpreta, in funzione di una opportuna programmazione, le informazioni provenienti dai sensori per generare segnale audio.

Il programma attivo sul computer, come succede nell'esempio di apertura di questo capitolo, legge i dati provenienti dai sensori e li utilizza per controllare gli algoritmi di generazione sonora, creando opportune corrispondenze tra azioni ed eventi musicali. Il termine *mapping* indica infatti la possibilità di configurare dinamicamente le modalità di collegamento tra i dati provenienti

dai controllori esterni e la generazione del suono vera e propria, come illustrato nell'esempio delle scalette semitonate.

Questo modo di mettere in relazione i gesti dell'esecutore con gli algoritmi che generano strutture sonore, si riflette anche e soprattutto sulle modalità compositive di un brano musicale, perché in questo contesto il performer non "suona uno strumento" ma piuttosto *controlla* una *meta-composizione* attraverso i dispositivi di input: il programma di mapping stesso che definisce il meccanismo algoritmico è la composizione musicale.

Una meta-composizione è una composizione musicale realizzata con le modalità della composizione algoritmica.

Come conseguenza della presenza del mapping, al termine meta-composizione si aggiunge quello di *meta-strumento* che evidenzia la relazione tra causa (il gesto) ed effetto (il risultato acustico-musicale) mediata dalla logica nell'algoritmo di mapping, di volta in volta implementata in modo arbitrario e creativo.

Il concetto di mapping rende indefiniti i confini tra strumento, composizione ed esecuzione, così netti invece nella situazione tradizionale. La semplice relazione uno-a-uno tra azioni sullo strumento e suono generato, diventa ora *uno-a-molti*, *molti-a-molti* se non addirittura, in modo poco formale ma suggestivo, *molti-ad-algoritmo*.

Non è facile formalizzare il concetto di mapping.

Anzi, a giudicare dai numerosi scritti esistenti in letteratura intorno all'argomento, sembra proprio che non sia questo l'obiettivo ma piuttosto quello di individuare espedienti, metodologie e strategie di pratica utilità per l'implementazione di algoritmi di mapping [6][7][8].