

Campionamento e Sintesi 1

Andrea Agostini

30 aprile 2023

Capitolo 1

Pressione, ampiezza, intensità

1.1 Pressione sonora

Il suono è costituito da (o associato a) onde di pressione longitudinali che si propagano in un mezzo elastico (tipicamente l'aria).

L'onda longitudinale è tale perché lo spostamento di energia avviene lungo lo stesso asse dello spostamento di materia (come per esempio nel caso di una molla che viene colpita a un'estremità). Il caso opposto, che non riguarda il suono, è l'onda trasversale, in cui lo spostamento di energia avviene perpendicolarmente rispetto all'asse dello spostamento di materia: l'esempio classico è l'onda del mare, in cui l'energia si sposta orizzontalmente (dal mare aperto verso la riva) mentre la materia si sposta verticalmente (la singola molecola d'acqua si alza e si abbassa a causa del moto ondoso).

Il suono sarà in generale prodotto da un corpo emittente, che per semplicità immaginiamo puntiforme. Nel nostro modello, consideriamo anche la presenza di un recettore, pure puntiforme. Il corpo emittente oscilla in maniera più o meno regolare attorno a un suo punto di riposo. Così facendo, crea nell'aria un'alternanza di compressioni nella direzione del recettore e compressioni nella direzione opposta a quella del recettore (che possiamo considerare decompressioni nella direzione del recettore). Questi fronti di compressioni e decompressioni si propagano in forma di superfici sferiche.

La pressione si misura in pascal (Pa) e i suoi multipli e sottomultipli, come

- il millipascal (mPa: $\frac{1}{1000}$ Pa),

- il micropascal (μPa : $\frac{1}{1000000}\text{Pa}$),
- il kilopascal (kPa : 1000Pa),
- il bar (bar : $100\,000\text{Pa}$).

La pressione atmosferica standard al livello del mare è poco più di $100\,000\text{Pa}$. Il fenomeno sonoro provoca piccole variazioni nella pressione atmosferica: le variazioni tipiche dei suoni udibili avvengono in un ambito compreso all'incirca tra $20\mu\text{Pa}$ e 20Pa .

Il recettore del fenomeno sonoro (ad esempio la membrana del timpano, o una membrana microfonica) è immerso nell'atmosfera, che esercita eguale pressione da entrambi i lati: di conseguenza, possiamo considerare in generale nulla la pressione atmosferica complessiva a cui la membrana è sottoposta.¹ Il fenomeno sonoro provoca invece l'applicazione di una pressione direzionata sul recettore, che ne sarà compresso verso l'interno o verso l'esterno: ha senso quindi dire che il recettore è sottoposto a una pressione sonora dell'ambito descritto sopra (tra $20\mu\text{Pa}$ e 20Pa). Per convenzione, parleremo di pressione positiva quando la pressione esterna è superiore alla pressione interna (e quindi la membrana del timpano sarà spinta verso l'interno del cranio); di pressione negativa nel caso opposto (in cui la membrana del timpano è "risucchiata" verso l'esterno del cranio).

L'onda sonora impiega tempo a propagarsi: al livello del mare, con aria secca e temperatura di 20°C , la velocità di propagazione è di circa 343m/s . Questo vuol dire che, approssimativamente, il suono in queste condizioni percorre 30cm in 1ms ; 1m in 3ms ; e 1km in 3s .

1.2 Ampiezza e rappresentazione nel dominio del tempo

Ci interessa individuare alcune grandezze che variano in maniera idealmente proporzionale² tra loro, per lo stesso fenomeno sonoro:

- lo spostamento dell'emittente s_e

¹Questo non è vero ad esempio nel caso di sbalzi d'altitudine o di immersioni subacquee: in questi casi infatti bisogna compensare le variazioni di pressione a cui il timpano è sottoposto.

²Nel mondo reale, questa proporzionalità è sempre approssimativa. Rispetto alle definizioni della fisica classica, e in un contesto ideale in cui non esistono fenomeni parassiti che disturbano l'emissione, la trasmissione e la ricezione del suono (come ad esempio rumori d'ambiente, correnti d'aria, superfici riflettenti, l'elasticità non perfetta dei materiali e così via), è spesso utile considerarla esatta per semplicità.

- lo spostamento del recettore s_r (se il mezzo elastico non induce distorsioni: in generale possiamo considerare che le distorsioni indotte dall'aria siano estremamente ridotte)
- le variazioni di pressione atmosferica Δp (in assenza di altri fenomeni perturbativi)
- la tensione elettrica nella catena elettroacustica V (assumendo che trasduttori, conduttori e amplificatori si comportino in maniera ideale)

Possiamo allora scrivere che

$$s_e \propto s_r \propto \Delta p \propto V \quad (1.1)$$

dove il simbolo \propto indica una relazione di proporzionalità diretta.

Poiché la percezione e il trattamento del fenomeno sonoro avvengono in generale in termini astratti rispetto alla materialità delle grandezze fisiche coinvolte, spesso utilizziamo il concetto di *ampiezza* del segnale sonoro. L'ampiezza è una misura adimensionale proporzionale alle grandezze elencate sopra; le rappresenta tutte, e non coincide con nessuna in particolare. Il suo ambito è arbitrario e convenzionale; in condizioni normali, si adotta talvolta un ambito compreso tra -1 e 1, ma ci sono moltissime eccezioni. Il concetto di ampiezza è cruciale nella rappresentazione digitale del suono. Quindi

$$A \propto s_e \propto s_r \propto \Delta p \propto V \quad (1.2)$$

Se si traccia l'andamento di una qualsiasi di queste grandezze rispetto al tempo si ottiene un grafico che rappresenta il fenomeno sonoro *nel dominio del tempo*. Il fatto che si tratti di grandezze tutte proporzionali tra loro fa sì che, cambiando l'unità di misura sull'asse verticale, il grafico mantenga la stessa forma. Possiamo quindi parlare di *forma d'onda* (in inglese *waveform*). Questo tipo di grafico è talvolta chiamato *audiogramma*.

INSERIRE QUI ESEMPI DI FORME D'ONDA

1.3 Misure dell'ampiezza

Intuitivamente, osservare quanto il tracciato del grafico si distanzia dall'asse orizzontale dà un'idea approssimativa di “quanto forte” è il suono, o della sua dinamica (non uso il termine “intensità”, che ha un significato specifico definito più avanti). È utile definire alcune maniere diverse di misurare l'ampiezza di un fenomeno sonoro:

1.3.1 Ampiezza istantanea

L'*ampiezza istantanea* è l'ampiezza misurata a un istante specifico del fenomeno sonoro. Di per sé, non ci dà un'informazione affidabile sul livello dinamico percepito, perché qualsiasi fenomeno sonoro è costituito dalla rapida alternanza di pressioni relativamente alte (cioè pressioni che, in una direzione o nell'altra, deviano in maniera importante dalla pressione atmosferica media) e pressioni relativamente basse (cioè pressioni prossime alla pressione atmosferica media). L'ampiezza istantanea è un valore *segnato*: convenzionalmente, è positiva se rappresenta una pressione nella direzione nella direzione dall'emittente al recettore; negativa se rappresenta una pressione nella direzione dal recettore al ricevente. Naturalmente, può avere senso considerare il valore assoluto dell'ampiezza istantanea.

1.3.2 Ampiezza di picco

L'*ampiezza di picco* è l'ampiezza di valore assoluto massimo entro un dato intervallo di tempo. Se l'intervallo è ragionevolmente ampio, può costituire una prima stima del livello dinamico percepito. L'ampiezza di picco è, in linea di principio, un valore segnato, per le stesse ragioni per cui lo è l'ampiezza istantanea. Nella pratica, spesso non è interessante sapere in quale direzione è stata esercitata la massima ampiezza rinvenuta: per questa ragione, viene normalmente considerato il valore assoluto dell'ampiezza di picco.

1.3.3 Ampiezza picco-picco

L'*ampiezza picco-picco* è la differenza tra la massima e la minima ampiezza entro un dato intervallo di tempo. Nel caso di una forma d'onda perfettamente centrata sullo 0 coincide con il doppio dell'ampiezza di picco. Se però la forma d'onda non è centrata sullo 0, o se la posizione dello 0 non è nota con esattezza (per esempio perché esiste una componente di tensione continua,³ l'ampiezza picco-picco fornisce l'unica possibile misura affidabile. L'ampiezza picco-picco, secondo la definizione data sopra, non è mai negativa.

³Una componente di tensione continua provoca una traslazione sull'asse verticale della rappresentazione nel dominio del tempo.

1.3.4 Ampiezza RMS

L'*ampiezza RMS* è una misura che tiene conto dell'andamento dell'ampiezza nel tempo. Di conseguenza, può rappresentare in maniera più accurata il livello dinamico percepito, che a parità di ampiezza di picco (o picco-picco) sarà maggiore se, su scale di durata di frazioni di secondo, il segnale rimane forte più a lungo. La definizione di ampiezza RMS per un segnale campionato in ampiezza a intervalli di tempo regolari, come i segnali digitali che trattiamo abitualmente, è

$$A_{\text{RMS}} = \sqrt{\frac{A_1^2 + A_2^2 + \dots + A_n^2}{n}} \quad (1.3)$$

dove n è il numero totale di campioni e $A_1 \dots A_n$ sono i valori di ampiezza istantanea dei singoli campioni entro la finestra temporale considerata. L'ampiezza RMS è quindi la radice quadrata della media dei quadrati delle ampiezze istantanee. Se consideriamo invece un segnale continuo nel tempo, la definizione di RMS deve fare ricorso al concetto di integrale:⁴

$$A_{\text{RMS}} = \sqrt{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} (A(t))^2 dt}$$

dove t indica il tempo, t_1 e t_2 sono rispettivamente il tempo iniziale e finale considerati e $A(t)$ è l'ampiezza nel tempo.

L'ampiezza RMS non può essere negativa, dal momento che non può esserlo il quadrato di un numero reale.

1.4 Intensità sonora

Dal punto di vista percettivo, più importante del concetto di pressione è quello di *intensità sonora*. Per definirla correttamente abbiamo bisogno di introdurre alcune nozioni:

1.4.1 Energia e lavoro

Energia e lavoro sono due grandezze strettamente correlate, rappresentate dalle stesse unità di misura come il Joule e la caloria. Il lavoro è definito come la forza che sposta un corpo moltiplicata per lo spostamento effettuato:

$$W = F \cdot s \quad (1.4)$$

⁴Un integrale può essere visto come la somma di infiniti campioni di durata infinitesimale: ci serve quindi a calcolare il valore totalizzato da un fenomeno continuo. Se non sei familiare con questo formalismo ignora tranquillamente i dettagli di questa formula.

L'energia può essere vista come la capacità di un corpo di svolgere un lavoro: si misura in termini del lavoro che in virtù di essa un corpo è in grado di svolgere. È una grandezza fisica fondamentale che obbedisce alla legge di conservazione: viene trasferita, non si crea e non si distrugge. In senso proprio, quindi, ciò che viene trasferito da un corpo vibrante alla membrana del nostro timpano è l'energia sonora, non la pressione sonora. In altri termini, il fatto che una pressione sonora si applichi al timpano è concettualmente una conseguenza del trasferimento di energia, e non viceversa.

L'energia è proporzionale al quadrato della pressione sonora e, di conseguenza, al quadrato dell'ampiezza e di tutti gli altri fenomeni direttamente proporzionali all'ampiezza:⁵

$$E \propto \Delta p^2 \propto \Delta A^2 \quad (1.5)$$

1.4.2 Potenza

La *potenza* (misurata in Watt e indicata qui dalla lettera P maiuscola, da non confondersi con la p minuscola con cui indichiamo la pressione) è una misura della quantità di lavoro svolto, o di energia dispiegata, per unità di tempo:

$$P = \frac{W}{t} \quad (1.6)$$

A parità di lavoro o di energia, la potenza è maggiore se il tempo è più breve. La stessa quantità di energia sonora dispiegata in un secondo o in un'ora darà luogo a fenomeni sonori di potenza diversa, maggiore nel primo caso, minore nel secondo. A parità di tempo, la potenza è proporzionale all'energia e quindi al quadrato dell'ampiezza:

$$P \propto E \propto \Delta p^2 \propto A^2 \quad (1.7)$$

1.4.3 Intensità

L'intensità sonora è una misura della potenza riferita all'area su cui questa è distribuita:

$$I = \frac{P}{a} \quad (1.8)$$

⁵La ragione di questa relazione discende dalle definizioni formali di pressione e di energia. Potresti chiederti: ma *perché* è così? La fisica dovrebbe rispondere che la scienza si occupa del come, non del perché che è invece appannaggio della filosofia; ma qui puoi trovare alcuni tentativi, più o meno fantasiosi, di spiegazioni intuitive: <https://languagelog.ldc.upenn.edu/n11/?p=6508>. E, a dire il vero, anche il concetto di *conseguenza*, impiegato in maniera un po' infingarda qualche riga sopra, sarebbe una questione filosofica più che scientifica.

L'intensità sonora non ha, in generale, una sua specifica unità di misura, ma è misurata in generale in Watt per metro quadrato ($\frac{W}{m^2}$ o $W \cdot m^{-2}$).

Possiamo assumere che il suono, in assenza di ostacoli, si propaghi su un fronte d'onda di forma corrispondente a una superficie sferica a partire dall'emittente, che consideriamo puntiforme. La superficie sferica naturalmente sarà sempre più ampia via via che il fronte d'onda si allontana dalla sorgente, secondo la formula

$$a = 4\pi r^2 \quad (1.9)$$

dove a è l'area della sfera e r il raggio, che in questo caso corrisponde alla distanza tra l'emittente e il fronte d'onda che stiamo considerando. Di conseguenza, possiamo dire che l'energia dispiegata dal fenomeno sonoro in un certo lasso di tempo si distribuisce su una superficie la cui area aumenta in maniera proporzionale al quadrato della distanza. In altri termini, posso pensare l'intensità sonora come la potenza portata dall'onda sonora su una certa superficie, in senso perpendicolare a essa, e a una certa distanza dall'emittente. E dunque, a parità di tutti gli altri parametri, l'intensità sonora è inversamente proporzionale al quadrato della distanza:

$$I \propto \frac{1}{r^2} \quad (1.10)$$

Questa relazione è chiamata *legge dell'inverso del quadrato* e si applica a un vasto insieme di altri fenomeni, tra cui l'intensità luminosa e l'attrazione gravitazionale.

D'altra parte, l'intensità sonora è direttamente proporzionale alla potenza, all'energia e al quadrato della pressione e dell'ampiezza:

$$I \propto P \propto E \propto \Delta p^2 \propto A^2 \quad (1.11)$$

Di conseguenza,

$$\frac{1}{r^2} \propto \Delta p^2 \propto A^2 \quad (1.12)$$

e quindi

$$\frac{1}{r} \propto \Delta p \propto A \quad (1.13)$$

È importante osservare che queste grandezze non variano con la legge dell'inverso del quadrato, ma con la *legge dell'inverso della distanza*.

1.5 Riepilogo e classificazione delle grandezze

Abbiamo quindi definito due categorie di grandezze fisiche nelle quali ci imatteremo ancora, e sulle quali non dev'esserci confusione:

Quantità proporzionali alla potenza (*power quantities*)

La prima categoria include l'energia, il lavoro, la potenza e l'intensità sonora. È importante ricordare che da un punto di vista concettuale la quantità fondamentale tra queste è l'energia, che gode del principio di conservazione. L'energia, il lavoro e la potenza non variano con la distanza: la quantità di energia (e quindi di lavoro e di potenza) in gioco è quella trasferita o trasferibile dall'emittente ai corpi circostanti, e rispetto a essa la distanza non è un parametro rilevante. La distanza entra in gioco quando si considera l'intensità sonora, che segue la *legge dell'inverso del quadrato*.

Quantità proporzionali alla radice quadrata della potenza (*root-power quantities*)

La seconda categoria include la pressione sonora e la tensione elettrica. La potenza è proporzionale al quadrato di queste grandezze: di conseguenza, queste saranno proporzionali alla radice quadrata della potenza, da cui il nome:

$$W \propto \Delta p^2 \implies \Delta p \propto \sqrt{W} \quad (1.14)$$

Si tratta di grandezze a cui non si applica principio di conservazione e che variano in maniera proporzionale alla distanza. Concettualmente, possiamo considerare i fenomeni rappresentati da queste grandezze come conseguenze del trasferimento di energia. D'altra parte la pressione sonora è il fenomeno che viene direttamente misurato da una membrana di microfono o dal nostro timpano.

L'ampiezza, in senso proprio, non è una grandezza fisica perché non descrive un fenomeno specifico; possiamo però considerarla un'astrazione delle *root-power quantities*, utile quando vogliamo considerare il fenomeno sonoro in maniera indipendente dal fenomeno fisico che l'ha prodotto; questo è particolarmente rilevante nel momento in cui trattiamo il suono dal punto di vista elettroacustico e, di conseguenza, numerico.

1.6 Decibel

Le pressioni (o, dovremmo dire, i Δp) che possiamo percepire come fenomeni sonori — abbastanza forti da essere udite, non così forti da danneggiare le nostre orecchie — si situano approssimativamente in un ambito compreso tra i 20 μPa e i 20 Pa. La nostra scala percettiva della dinamica, però, è tutt'altro che lineare rispetto all'ambito di pressione così misurato; la distanza percettiva tra un suono A di pressione 100 μPa e un suono B di 200 μPa sarà

decisamente diversa rispetto a quella tra un suono C di pressione 1000 μPa e un suono D di 1100 μPa , nonostante la differenza tra A e B sia la stessa che tra C e D, e cioè di 100 μPa .

Invece, la distanza percettiva tra un suono A di pressione 100 μPa e un suono B di 200 μPa sarà decisamente diversa rispetto a quella tra un suono C di pressione 1000 μPa e un suono D di 2000 μPa : l'osservazione rilevante è che il rapporto di pressione tra A e B è lo stesso che tra C e D, e cioè $\frac{1}{2}$.

Il fatto che tendiamo a graduare molti fenomeni secondo i rapporti moltiplicativi tra le grandezze fisiche coinvolte piuttosto che secondo le loro differenze è talvolta considerata una legge fondamentale della percezione umana, chiamata *legge di Fechner*. Guardando questa relazione moltiplicativa da un punto di vista leggermente diverso, possiamo dire che la nostra scala percettiva è tendenzialmente *logaritmica* rispetto alla grandezza fisica osservata, dal momento che lo stesso rapporto moltiplicativo corrisponde alla stessa differenza tra i logaritmi. Tornando all'ultimo esempio,

$$\begin{aligned} \log_{10} 100 &= 2 \\ \log_{10} 200 &\approx 2.301 \\ \log_{10} 1000 &= 3 \\ \log_{10} 2000 &\approx 3.301 \\ \log_{10} 200 - \log_{10} 100 &= \log_{10} 2000 - \log_{10} 1000 \approx 0.301 \end{aligned} \tag{1.15}$$

e anche, per la definizione stessa di logaritmo,

$$\log_{10} \frac{200}{100} = \log_{10} \frac{2000}{1000} = \log_{10} 2 \approx 0.301 \tag{1.16}$$

e, già che ci siamo,

$$\log_{10} \frac{100}{200} = \log_{10} \frac{1000}{2000} = \log_{10} \frac{1}{2} \approx -0.301 \tag{1.17}$$

Da questa constatazione discende la definizione di bel (B), un'unità di misura adimensionale che, in una delle sue possibili definizioni, esprime il logaritmo decimale del rapporto tra due intensità sonore:

$$L_B = \log_{10} \frac{I}{I_0} \tag{1.18}$$

dove I è l'intensità sonora che vogliamo misurare rispetto a un'intensità di riferimento I_0 .

Molto più spesso del bel, in realtà, è usato il decibel, corrispondente alla decima parte del bel:

$$L_{dB} = 10 \cdot \log_{10} \frac{I}{I_0} \quad (1.19)$$

Formule analoghe possono essere usate per esprimere la relazione tra due livelli di energia sonora o di potenza sonora, dal momento che si tratta di grandezze proporzionali tra loro:

$$L_{dB} = 10 \cdot \log_{10} \frac{E}{E_0} \quad (1.20)$$

$$L_{dB} = 10 \cdot \log_{10} \frac{P}{P_0} \quad (1.21)$$

Una proprietà fondamentale di questa definizione di decibel è la sua coerenza con un'altra definizione equivalente, ma riferita al rapporto tra due pressioni sonore:

$$L_{dB} = 20 \cdot \log_{10} \frac{\delta p}{\delta p_0} \quad (1.22)$$

L'equivalenza tra queste due definizioni discende dalla relazione quadratica tra pressione e intensità sonora e dalla proprietà fondamentale dei logaritmi, che permette di “tramutare” un esponente dentro il logaritmo in un coefficiente fuori dal logaritmo:

$$\log x^a = a \cdot \log x \quad (1.23)$$

e quindi:

$$\begin{aligned} L_{dB} &= 10 \cdot \log_{10} \frac{I}{I_0} \\ &= 10 \cdot \log_{10} \frac{\Delta p^2}{\Delta p_0^2} \\ &= 10 \cdot \log_{10} \left(\frac{\Delta p}{\Delta p_0} \right)^2 \\ &= 2 \cdot 10 \log_{10} \frac{\Delta p}{\Delta p_0} \\ &= 20 \log_{10} \frac{\Delta p}{\Delta p_0} \end{aligned} \quad (1.24)$$

Formule analoghe a quella che definisce il decibel in riferimento alla pressione sonora possono essere usate considerando altre grandezze a essa proporzionali, come l'ampiezza e la tensione elettrica:

$$L_{dB} = 20 \cdot \log_{10} \frac{A}{A_0} \quad (1.25)$$

$$L_{dB} = 20 \cdot \log_{10} \frac{V}{V_0} \quad (1.26)$$

Quale che sia il valore di riferimento di pressione sonora, intensità, potenza, voltaggio, ampiezza o altro, ci sono alcune considerazioni generali che possono essere fatte:

- A un suono al livello di riferimento corrisponde sempre un livello di 0 dB: se per esempio I è uguale a I_0 , e poiché il logaritmo di 1 in qualsiasi base è 0, allora

$$L_{dB} = 10 \cdot \log_{10} \frac{I}{I_0} = 10 \cdot \log_{10} 1 = 10 \cdot 0 = 0 \quad (1.27)$$

e lo stesso vale se invece dell'intensità si considera la pressione acustica, o l'ampiezza, o qualsiasi altra delle grandezze che abbiamo citato.

- A suoni più forti del valore di riferimento (cioè con ampiezza, pressione, intensità ecc. superiori a esso) corrispondono livelli in decibel positivi; a suoni più deboli del valore di riferimento corrispondono livelli in decibel negativi. Infatti nel primo caso il rapporto tra valore misurato e valore di riferimento sarà maggiore di 1, e quindi il suo logaritmo maggiore di 0; nel secondo caso il rapporto sarà maggiore di 0 e minore di 1, e quindi il suo logaritmo minore di 0.
- Il valore di riferimento non può essere il silenzio perfetto (corrispondente a un valore di intensità sonora, pressione sonora, ampiezza ecc. pari a 0), perché in questo caso metteremmo uno 0 sotto la linea di frazione e la divisione per 0 ha risultato indefinito.
- In senso proprio, neppure il valore misurato può essere il silenzio perfetto, perché questo vorrebbe dire calcolare il logaritmo di 0 che pure non è definito. Con un leggero abuso di terminologia, si usa però spesso scrivere che il silenzio perfetto corrisponde a $-\infty$ dB. In senso proprio, il silenzio perfetto esiste solo nel contesto astratto delle rappresentazioni numeriche.
- In ogni caso, anche quando il valore misurato o il valore di riferimento derivano da misurazioni con segno come la misurazione dell'ampiezza istantanea, nel calcolo del livello in dB si considera sempre il loro valore assoluto, perché il logaritmo dei numeri negativi non è definito.

La scala dei decibel, come dicevamo, rappresenta una modellizzazione semplice della maniera in cui la percezione umana gradua l'intensità del

fenomeno sonoro. In maniera estremamente approssimativa, ma non completamente sbagliata, possiamo considerare che una differenza di 1, o 10, o 50 dB tra due suoni altrimenti identici abbia sempre lo stesso “peso” percettivo. Questo è particolarmente importante in elettroacustica e informatica musicale, contesti nei quali è semplice controllare in maniera precisa l’ampiezza di un suono senza modificare gli altri suoi parametri.

1.7 Livelli di riferimento convenzionali per i decibel

Il decibel è una misura relativa, che però può essere impiegata in maniera assoluta se viene fissato un valore di riferimento per A_0 , Δp_0 , V_0 , A_0 e così via. Esistono alcuni valori di riferimento convenzionali che danno luogo a scale assolute utili in contesti diversi. Queste comprendono, tra le altre:

- dB_{SPL} , dove SPL sta per *sound pressure level*: misura la pressione acustica rispetto a una P_0 di 20 μPa , corrispondenti approssimativamente alla soglia di udibilità umana. Quando nel linguaggio comune si dice “in quel locale la musica era a 90 dB” si sottintende che si sta parlando di dB_{SPL} . Tutti i suoni udibili avranno livelli non negativi in dB_{SPL} . La pressione sonora di 20 Pa che viene spesso indicata come la massima che l’orecchio umano può tollerare senza subire danni permanenti corrisponderà quindi a un livello sonoro di 120 dB_{SPL} :

$$L_{dB} = 20 \cdot \log_{10} \frac{20\text{Pa}}{20\mu\text{Pa}} = 20 \cdot \log_{10} 10^6 = 20 \cdot 6 = 120 \quad (1.28)$$

- dB_{FS} , dove FS sta per *full scale*: misura l’ampiezza rispetto al massimo livello rappresentabile in un sistema audio digitale a rappresentazione intera o in virgola fissa.⁶ Tutti i suoni correttamente rappresentabili in un tale sistema avranno livelli non positivi in dB_{FS} .

Esistono numerose altre scale di riferimento usate in elettroacustica e psicoacustica, ma non le approfondiremo qui.

⁶Maggiori dettagli su queste rappresentazioni saranno dati più avanti. Per il momento, possiamo considerare come minimo che si tratta delle rappresentazioni usate dalle interfacce audio e dalla maggior parte dei formati di file audio.

1.8 Piccola aritmetica dei decibel

Succede spesso di dover convertire in decibel un rapporto di ampiezze, o viceversa. Questo può essere facilmente fatto con Max tramite gli oggetti *atodb* e *dbtoa*, ma è utile avere in mente almeno alcune identità semplici:

- Se un dato rapporto di ampiezza corrisponde a un guadagno pari a L_{dB} , allora l'inverso di tale rapporto corrisponderà a un guadagno pari a $-L_{dB}$: se

$$20 \cdot \log_{10} \frac{A}{A_0} = L_{dB} \quad (1.29)$$

allora

$$\begin{aligned} 20 \cdot \log_{10} \frac{A_0}{A} &= \\ 20 \cdot \log_{10} \frac{1}{\frac{A}{A_0}} &= \\ 20 \cdot \log_{10} \left(\frac{A}{A_0} \right)^{-1} &= \\ -1 \cdot 20 \cdot \log_{10} \frac{A}{A_0} &= \\ -1 \cdot L_{dB} &= \\ -L_{dB} & \end{aligned} \quad (1.30)$$

- Raddoppiare l'ampiezza corrisponde a un guadagno di circa 6 dB, dal momento che $\log_{10} 2 \approx 0.301$ e quindi $20 \cdot \log_{10} 2 \approx 6.02$. Poiché uno scarto di 0.02 dB è nella maggior parte dei casi del tutto irrilevante, spesso si considera l'identità come esatta anziché approssimata.
- Dimezzare l'ampiezza corrisponde a un guadagno di circa -6 dB.
- Moltiplicare per 10 l'ampiezza corrisponde a un guadagno di esattamente 20 dB, poiché $\log_{10} 10 = 1$ e quindi $20 \cdot \log_{10} 10 = 20$.
- Dividere per 10 l'ampiezza corrisponde a un guadagno di esattamente -20 dB.

Considerando che, in virtù della proprietà fondamentale dei logaritmi, moltiplicazioni dell'ampiezza corrispondono sempre a somme algebriche di decibel ed elevamenti a potenza dell'ampiezza corrispondono a moltiplicazioni di decibel, le identità riportate sopra possono essere utili per fare alcuni calcoli in maniera semplice. Per esempio:

- Un guadagno di 100, cioè $10 \cdot 10$, corrisponde a $20 + 20 = 40$ dB.
- Un guadagno di 20, cioè $2 \cdot 10$, corrisponde a circa $6 + 20 = 26$ dB.
- Un'attenuazione di 1000, cioè un guadagno di $\frac{1}{1000}$ o 10^{-3} corrisponde a $20 \cdot -3 = -60$ dB.
- 8 dB, cioè $20 - 2 \cdot 6$ dB, corrispondono a un guadagno di circa $10 \cdot 2^{-2} = 10 \cdot \frac{1}{4} = 2.5$
- 10 dB, cioè $20 \cdot \frac{1}{2}$ dB, corrispondono a un guadagno di $10^{\frac{1}{2}} = \sqrt{10} \approx 3.16$

Queste operazioni corrispondono, tra l'altro, al collegamento in serie di amplificatori o attenuatori, come avviene normalmente nel percorso del segnale lungo una catena elettroacustica o all'interno di un mixer. Se per esempio lo stadio di guadagno di un mixer è regolato a 40 dB, il fader del corrispondente canale a -10 dB e il livello di uscita a -20 dB (assumendo che non ci siano altre trasformazioni del segnale, come equalizzazioni o altro) sapremo che il segnale in uscita dal mixer sarà $40 - 10 - 20 = 10$ dB più forte rispetto a quello in entrata, e quindi che il voltaggio in uscita sarà circa pari al voltaggio in entrata moltiplicato per 3.16.

Capitolo 2

Frequenza

2.1 Frequenza e sue unità di misura

La nostra rappresentazione cognitiva del suono non avviene nei termini di rapide oscillazioni di pressione dell'aria, ma piuttosto in termini di successioni ed evoluzioni nel tempo di altezze e timbri. Per fare ciò, individuiamo nell'andamento della pressione sonora strutture anche grossolanamente regolari nel tempo e le analizziamo dal punto di vista frequenziale: cioè, in sostanza, valutiamo quante volte tali strutture si ripetono, anche in maniera estremamente approssimativa, in una data unità temporale. Da tale valutazione siamo in grado di trarre informazioni sulle qualità e sull'attività dell'emittente. A complemento di ciò, siamo dotati di un apparato fonatorio in grado di emettere vibrazioni regolari di pressione acustica in maniera estremamente controllata: dalla relazione tra emissione e percezione di tali vibrazioni nasce la parola; la musica, la sua origine, il suo ruolo evolutivo sono fatti più misteriosi e sfuggenti, ma sono basati almeno parzialmente sugli stessi meccanismi.

Introduciamo quindi il concetto di frequenza e la sua misurazione: dato un fenomeno che varia in maniera regolare, possiamo individuare il suo periodo, cioè l'intervallo temporale che intercorre tra due sue ripetizioni. Chiamiamo questo periodo Δt e lo misuriamo tramite unità di misura di tempo: millisecondi, secondi, minuti, ore, giorni...

La frequenza indica quante volte la variazione regolare avviene in una data unità di tempo: per esempio, quanti periodi ci sono in un secondo, in un minuto, in un'ora. Possiamo quindi definirla come l'inverso del periodo:

$$f = \frac{1}{\Delta t} = \Delta t^{-1} \quad (2.1)$$

e quindi

$$\Delta t = \frac{1}{f} = f^{-1} \quad (2.2)$$

Possiamo misurare la frequenza ad esempio in hertz (Hz, cicli per secondo), kilohertz (kHz, migliaia di Hertz o cicli per millisecondo), pulsazioni al minuto (BPM, cicli per minuto). Nota che le formule scritte sopra si possono applicare in maniera letterale solo se le unità del tempo e della frequenza corrispondono; se così non è, sarà necessaria una conversione: per esempio,

$$f_{Hz} = \frac{1}{\Delta t_s} \quad (2.3)$$

e

$$f_{kHz} = \frac{1}{\Delta t_{ms}} \quad (2.4)$$

ma

$$f_{kHz} = 1000 \frac{1}{\Delta t_s} = \frac{1000}{\Delta t_s} \quad (2.5)$$

e

$$f_{BPM} = 60 \frac{1}{\Delta t_s} = \frac{60}{\Delta t_s} \quad (2.6)$$

Tutto quanto detto fin qui riguarda in senso stretto la *frequenza fondamentale* di un segnale, che è appunto l'inverso della durata di una delle sue ripetizioni nell'unità di tempo.

Sappiamo però che qualsiasi segnale può essere scomposto in una somma potenzialmente infinita di segnali sinusoidali. Prima di tutto allora abbiamo bisogno di una definizione di senoide.

2.2 Seni e coseni

Esistono funzioni matematiche dette periodiche, che ripetono all'infinito un comportamento sempre uguale e che ci riguardano molto da vicino, visto che sono quelle per le quali ha senso definire una frequenza.

Le funzioni periodiche sono molte — un'infinità, letteralmente — ma ce ne sono due che hanno alcune proprietà in virtù delle quali esse costituiscono il mattone elementare a partire dal quale costruiremo gran parte del nostro discorso sul suono. Queste funzioni sono il *seno* e il *coseno*, che rappresentano rispettivamente la proiezione sull'asse verticale e orizzontale di un punto della circonferenza goniometrica:¹ se prendo un raggio della circonferenza

¹La circonferenza goniometrica (*unit circle* in inglese) è una circonferenza tracciata nel piano cartesiano, con centro corrispondente all'origine degli assi e raggio pari a 1.

che forma un angolo α in senso antiorario rispetto all'asse delle ascisse, il seno di α è la proiezione sull'asse verticale del punto in cui questo raggio tocca la circonferenza; e il coseno di α è la proiezione dello stesso punto sull'asse orizzontale.

Vediamo ora alcuni fatti fondamentali relativi a seno e coseno.

2.2.1 Radianti

L'argomento di una funzione seno o coseno è un angolo che viene normalmente espresso in radianti, dove il radiante è una misura adimensionale che rappresenta il rapporto tra la lunghezza di un arco di circonferenza definito da un dato angolo e il raggio della stessa circonferenza. Consideriamo l'arco definito da un angolo di 360° : in questo caso, l'arco corrisponde all'intera circonferenza, la cui lunghezza c possiamo calcolare a partire dal raggio r come

$$c = 2\pi r \quad (2.7)$$

e quindi l'angolo in radianti corrispondente a 360° sarà

$$360^\circ = \frac{2\pi r}{r} = 2\pi \quad (2.8)$$

Altri misure di angoli che incontreremo spesso sono

$$90^\circ = \frac{\pi}{2} \quad (2.9)$$

e

$$180^\circ = \pi \quad (2.10)$$

Nota che un angolo espresso in radianti non porta un'unità di misura.

2.2.2 Relazione tra seni e coseni

Il seno e il coseno hanno esattamente lo stesso andamento, traslato di un quarto di cerchio (cioè 90°). È facile immaginare il perché: un cerchio è sempre “fatto nello stesso modo” comunque io lo ruoti, e ruotando tutta la figura di 90° quello che prima era l'asse dell'ordinata si ritrova ora sull'asse dell'ascissa. Quindi

$$\cos \alpha = \sin\left(\alpha + \frac{\pi}{2}\right) \quad (2.11)$$

2.2.3 Codominio

Il codominio del seno e del coseno — cioè, l'ambito di valori y che una funzione seno o coseno $y = \sin x$ o $y = \cos x$ può assumere — è compreso tra -1 e 1. Anche questo è facile da vedere, per via del fatto che la circonferenza goniometrica ha raggio 1 e quindi tutti i suoi punti sono compresi tra -1 e 1 sia sull'ascissa che sull'ordinata. Quindi, per qualsiasi angolo x

$$-1 \leq \sin x \leq 1 \quad (2.12)$$

e

$$-1 \leq \cos x \leq 1 \quad (2.13)$$

Di conseguenza, l'ampiezza picco-picco delle funzioni seno e coseno è 2.

2.2.4 Dominio e periodicità

Il dominio del seno e del coseno è infinito: angoli più grandi di un giro (cioè di 2π) sono possibili e ben definiti, e i punti che questi definiscono sulla circonferenza corrispondono a punti definiti da angoli più piccoli del giro. Possiamo dire che, continuando a far ruotare il raggio dopo che ha compiuto un giro, questo continua a ripercorrere sempre lo stesso percorso. Possiamo estendere il ragionamento ad angoli negativi, che possono essere visti come angoli calcolati facendo ruotare il raggio in senso orario anziché antiorario.

Questa considerazione in definitiva spiega perché seno e coseno sono funzioni periodiche, che si ripetono uguali a sé stesse a ogni giro di circonferenza, cioè ogni 2π radianti. Quindi

$$\forall n \in \mathbb{Z}, \sin(x + 2\pi n) = \sin x \quad (2.14)$$

e

$$\forall n \in \mathbb{Z}, \cos(x + 2\pi n) = \cos x \quad (2.15)$$

Per questa ragione, di solito si preferisce utilizzare come argomenti del seno e del coseno valori compresi tra 0 e 2π o, talvolta, tra $-\pi$ e π , a meno che non ci siano ragioni forti per fare altrimenti.

2.3 Sinusoidi

Si può generalizzare la nozione di seno e coseno in quella più astratta di sinusoidi. Una sinusoidi è una qualsiasi funzione che abbia l'andamento di un seno o di un coseno, ma che abbia un periodo potenzialmente diverso da

2π ; che sia potenzialmente traslata lungo l'asse orizzontale; e la cui ampiezza picco-picco sia potenzialmente diversa da 2, fermo restando che il codominio dev'essere simmetrico rispetto allo 0.

2.3.1 Frequenza

Il termine della frequenza controlla la periodicità della funzione. Ad esempio, possiamo definire una funzione sinusoidale con frequenza 1 (cioè che si ripete esattamente una volta per ogni incremento unitario sull'asse x) come

$$s(x) = \sin(2\pi x) \quad (2.16)$$

Più in generale, possiamo definire una funzione sinusoidale con frequenza f come

$$s(x) = \sin(2\pi f x) \quad (2.17)$$

Poiché, nel caso specifico del discorso sul suono, x rappresenta generalmente il tempo, spesso scriviamo

$$s(t) = \sin(2\pi f t) \quad (2.18)$$

Alcuni testi, come il Puckette, preferiscono parlare di velocità angolare anziché di frequenza, introducendo un termine

$$\omega = 2\pi f \quad (2.19)$$

per cui

$$f = \frac{\omega}{2\pi} \quad (2.20)$$

ω rappresenta non il periodo, cioè i cicli per unità di tempo, ma i radianti per unità di tempo. Questo rende più semplice ma meno tangibile l'equazione della sinusoide, che diventa

$$s(t) = \sin(\omega t) \quad (2.21)$$

La funzione sinusoidale è considerata, citando Curtis Roads, *sub specie æternitatis*: la sua espressione non contiene un termine che rappresenti una sua durata nel tempo. Dal momento che il suo dominio non è limitato, una sinusoide è un fenomeno che non ha un inizio e non ha una fine.

2.3.2 Ampiezza

L'ampiezza di una senoide può essere espressa tramite un coefficiente che moltiplica il seno:

$$s(t) = k \cdot \sin(2\pi ft) \quad (2.22)$$

Il valore assoluto di questo coefficiente rappresenta l'ampiezza di picco della senoide; il doppio del valore assoluto, $2k$, ne rappresenta l'ampiezza picco-picco. Se k è negativo, la senoide si ritrova rispecchiata rispetto all'asse orizzontale. Se k è 0 stiamo rappresentando il silenzio, e il termine della frequenza è ininfluente e perde significato (come pure quello della fase che vedremo tra poco).

2.3.3 Fase

Aggiungendo all'argomento un termine fisso ϕ , detto *fase*, si può traslare l'intera funzione sinusoidale sull'asse orizzontale:

$$s(t) = k \cdot \sin(2\pi ft + \phi) \quad (2.23)$$

In generale, il termine di fase è considerato compreso tra 0 e 2π o tra $-\pi$ e π : qualsiasi fase fuori dall'ambito di riferimento è equivalente a una fase entro esso, dal momento che la senoide ha un periodo di 2π . Il rovesciamento rispetto all'asse orizzontale prodotto da un coefficiente k negativo è equivalente a una rotazione di fase di π :

$$k \cdot \sin(2\pi ft + \phi) = -k \cdot \sin(2\pi ft + \phi + \pi) = -k \cdot \sin(2\pi ft + \phi - \pi) \quad (2.24)$$

Inoltre, una rotazione di fase di 90° corrisponde alla trasformazione della funzione seno in una funzione coseno:

$$k \cdot \sin(2\pi ft + \phi) = k \cdot \cos(2\pi ft + \phi - \frac{\pi}{2}) \quad (2.25)$$

2.3.4 Somme di sinusoidi alla stessa frequenza

Sommando due sinusoidi alla stessa frequenza si ottiene un'altra senoide, generalmente di ampiezza e fase diversa ma con la stessa frequenza:

$$k \cdot \sin(2\pi ft + \phi_1) + k_2 \cdot \sin(2\pi ft + \phi_2) = k \cdot \sin(2\pi ft + \phi) \quad (2.26)$$

Inoltre, una senoide con qualsiasi fase può essere espressa come la somma di una senoide seno e una senoide coseno con fase nulla (cioè, di una

sinusoide seno con fase nulla e una con termine di fase ruotato di π) alla stessa frequenza della sinusoide originale:

$$k \cdot \sin(2\pi ft + \phi) = k_1 \cdot \sin(2\pi ft) + k_2 \cdot \cos(2\pi ft) \quad (2.27)$$

Se vuoi approfondire queste due formule, ti consiglio di guardare qui: <https://www.dsprelated.com/showarticle/635.php>.

2.4 Percezione della frequenza, spazio delle altezze

La frequenza fondamentale è il parametro essenziale legato alla percezione d'altezza di un suono. In linea di principio, quindi, il concetto di altezza può essere applicato soltanto a suoni periodici. Nella pratica, tuttavia, perché la percezione umana possa attribuire un'altezza a un suono, non serve che questo sia esattamente periodico — e, naturalmente, non serve che sia illimitato nel tempo: è sufficiente solo una periodicità molto approssimativa ed estesa a un tempo relativamente breve (in generale, qualcosa attorno al decimo di secondo — di più per suoni gravi, talvolta meno per suoni acuti).

Nella nostra percezione dell'altezza è fondamentale il concetto di intervallo: l'intervallo è, semplificando il concetto, ciò che percepiamo come la “distanza” tra due altezze diverse. La riconoscibilità di una struttura melodica o armonica non è tanto data dalle altezze in termini assoluti, ma dalla configurazione (nel tempo, per la melodia; nella simultaneità, per l'armonia) dei suoi intervalli. L'intervallo, da parte sua, può essere definito come un particolare rapporto tra frequenze. Per ragioni storiche e teoriche molto lontane da ciò di cui ci occupiamo qui, certi rapporti semplici hanno nomi specifici: per esempio, un rapporto di frequenze 2:1 è chiamato ottava; un rapporto 3:2 è chiamato quinta; un rapporto 4:3 è chiamato quarta; 5:4 è chiamato terza maggiore; 6:5 terza minore; 9:8 seconda maggiore.

Il fatto che nominiamo le distanze tra le altezze in base ai loro rapporti implica che l'organizzazione dell'altezza rispetto alla scala delle frequenze sia di tipo logaritmico, in maniera non dissimile — ma in realtà molto più precisa nella nostra percezione — a quanto abbiamo visto a proposito delle ampiezze. Questo vuol dire, tra l'altro, che lo stesso intervallo corrisponderà a una differenza frequenziale molto più piccola nel grave che nell'acuto. Per esempio, la quinta a partire da una frequenza di 110 Hz si trova a 165 Hz,² e quindi la differenza tra le due frequenze coinvolte è di 55 Hz; d'altra parte,

²Questo corrisponde all'intervallo la-mi in chiave di basso.

la quinta a partire da una frequenza di 440 Hz si trova a 660 Hz,³ con una differenza frequenziale di 220 Hz.

Possiamo quindi dare una preliminare e parziale definizione di *spazio delle altezze*, in relazione logaritmica con lo spazio delle frequenze. Per approfondire tale relazione bisognerebbe addentrarsi nella teoria dei temperamenti e dei sistemi di accordatura. Per il momento, riporto senza dimostrarli solo alcuni punti basilari che stanno alla base della teoria musicale e dell'organologia occidentale moderna:

- Gli unici intervalli che, nella più diffusa prassi moderna, mantengono il rapporto di frequenza semplice enunciato sopra (che deriva da proprietà acustiche fondamentali del suono e che è alla base di molti temperamenti antichi e non solo) sono l'ottava, con rapporto 2:1, e i suoi multipli.
- L'ottava è divisa in 12 semitoni, tutti definiti dallo stesso rapporto frequenziale (sono i cosiddetti *semitoni temperati*). Tale rapporto frequenziale è pari a $\sqrt[12]{2}$.
- Tutti gli altri intervalli vengono approssimati elevando a potenza il rapporto del semitono. Per esempio, la seconda maggiore, composta da due semitoni, ha rapporto pari a $(\sqrt[12]{2})^2 = \sqrt[6]{2}$; la quinta, composta da sette semitoni, ha rapporto pari a $(\sqrt[12]{2})^7$; l'ottava, coerentemente con la sua definizione data sopra, ha rapporto pari a $(\sqrt[12]{2})^{12} = 2$.

2.5 Spettro di frequenza

2.5.1 Teorema di Fourier e rappresentazione nel dominio del tempo

Il *teorema di Fourier* ci dice che qualsiasi segnale può essere rappresentato come una somma di componenti sinusoidali. Questo vuol dire, in senso proprio, che sommando algebricamente tra loro sinusoidi (potenzialmente in numero infinito) con specifiche frequenze, ampiezze e fasi è possibile ottenere qualsiasi segnale, di qualsiasi complessità. Chiamiamo queste sinusoidi le *componenti*, o *parziali*, del segnale originale.

Ci sono alcune ricadute importanti di questa affermazione. In primo luogo, il teorema di Fourier non parla della durata delle componenti: l'espressione matematica di una sinusoide assume che questa sia infinita nel

³Questo corrisponde all'intervallo la-mi in chiave di violino.

tempo, dal momento che il suo dominio corrisponde all'intero insieme dei numeri reali. In secondo luogo, lo stesso teorema non dice che il segnale debba avere caratteristiche di staticità o periodicità nel tempo: qualsiasi fenomeno, in tutta la sua evoluzione temporale, può essere rappresentato in termini di somma di sinusoidi definite in termini di frequenza, ampiezza e fase. In generale, però, è possibile che questa rappresentazione richieda di considerare sinusoidi a frequenza estremamente bassa.

2.5.2 Segnali periodici, armoniche

Esiste poi un caso particolare del teorema di Fourier, che si applica a segnali periodici: questi possono essere rappresentati da una somma di sinusoidi le cui frequenze sono tutte multipli interi della frequenza fondamentale del segnale. L'ampiezza e la fase di ciascuna sinusoide determinano la forma d'onda del segnale originale. Tutte queste sinusoidi sono dette *armoniche* del segnale; diremo che le loro frequenze si trovano in rapporto armonico tra loro, e che il segnale nel suo complesso è armonico. È vero anche il contrario: qualsiasi somma di sinusoidi le cui frequenze siano tutte multiple di una frequenza data (cioè siano tutte in rapporto armonico tra loro) produrrà un segnale periodico, che avrà quest'ultima frequenza come fondamentale. Dunque, *i tutti e soli i segnali armonici sono periodici e tutti e soli i segnali periodici sono armonici.*

Tra l'altro, i rapporti interi tra le armoniche sono alla base dei rapporti frequenziali semplici corrispondenti agli intervalli. Inoltre, il fatto che l'armonicità di un fenomeno sonoro (qualità di estrema rilevanza e riconoscibilità percettiva) corrisponda a una sua strutturazione definita dai rapporti di frequenza delle sue parziali ha una relazione evidente con il fatto che la nostra percezione classifica le strutture di frequenza in base ai loro rapporti.

Possiamo considerare le armoniche come un caso particolare delle parziali di un segnale, e in generale non è sbagliato chiamarle comunque “parziali” o “componenti”, a meno che questo non generi confusione.

Convenzionalmente, a ciascuna armonica viene associato un indice che corrisponde al rapporto tra la sua frequenza e quella della fondamentale — cioè, all'intero che, moltiplicando la frequenza della fondamentale, produce la frequenza dell'armonica. Quindi la seconda armonica avrà frequenza doppia rispetto alla fondamentale; la terza armonica frequenza tripla; e così via. Secondo questa convenzione, la prima armonica coinciderà con la fondamentale.

Esiste un'altra convenzione adottata soprattutto da testi più vecchi per cui l'armonica di frequenza doppia rispetto alla fondamentale è chiamata

prima armonica; quella di frequenza tripla è chiamata seconda armonica; e così via. Da questa convenzione deriva quella, di uso frequente, per cui la fondamentale viene chiamata f_0 , anche in contesti in cui questa viene considerata la prima armonica.

2.5.3 Rappresentazione nel dominio della frequenza

La rappresentazione di un segnale in termini delle sinusoidi che lo compongono è detta *nel dominio della frequenza*, e si contrappone alla rappresentazione *nel dominio del tempo* che è quella per cui il segnale è rappresentato in termini delle variazioni di ampiezza nel tempo. In senso stretto, ogni segnale, qualunque sia la durata, può essere descritto in termini di una somma di infinite sinusoidi.

L'operazione attraverso la quale si ottiene la rappresentazione nel dominio della frequenza o *spettro* a partire da quella del dominio del tempo è la *trasformata di Fourier* (*Fourier transform*, o *FT*). Questo termine indica anche il risultato dell'operazione, per cui possiamo dire che la rappresentazione nel dominio della frequenza è la trasformata di Fourier della rappresentazione nel dominio del tempo.

La trasformata di Fourier è un'operazione invertibile: dalla rappresentazione nel dominio della frequenza possiamo ottenere quella nel dominio del tempo applicando alla prima la *trasformata inversa di Fourier*. In senso proprio, la trasformata di Fourier si svolge su una rappresentazione analitica del segnale: cioè, è un'operazione che prende un'equazione che descrive il segnale come una funzione del tempo su dominio infinito e restituisce un'altra equazione che descrive il segnale come una funzione della frequenza su dominio infinito.

Definiamo formalmente lo spettro come la distribuzione delle ampiezze e delle fasi delle componenti sinusoidali rispetto alla frequenza. In generale, la distribuzione a cui siamo più interessati è quella delle ampiezze, dal momento che la nostra percezione non distingue la fase delle componenti, se non in casi molto particolari. Questa distribuzione può essere rappresentata graficamente come un diagramma cartesiano che rappresenta le frequenze sull'asse orizzontale e le frequenze corrispondenti sull'asse verticale. Una singola sinusoide avrà allora una rappresentazione grafica costituita da un solo punto. La somma di due sinusoidi sarà rappresentata da una coppia di punti. Segnali puramente armonici saranno rappresentati come un insieme di punti a posizioni equidistanti sull'asse delle ascisse. Discuteremo più avanti alcune altre categorie fondamentali di configurazioni spettrali, e

le relazioni tra le loro rappresentazioni nel dominio della frequenza e del tempo.

2.5.4 Considerazioni percettive, STFT

La nostra percezione del suono è basata su una rappresentazione ibrida: da una parte abbiamo la capacità di distinguere timbri diversi indicando le loro caratteristiche spettrali (più scuro, più brillante, più nasale; o anche, per chi è allenato a distinguere e nominare questi dettagli, più ricco di armoniche pari, o dispari, o di alcune armoniche specifiche). Siamo anche in grado di distinguere frequenze e timbri diversi prodotti simultaneamente da corpi diversi (più note suonate simultaneamente da un pianoforte o una chitarra, più voci che parlano o cantano simultaneamente, la compresenza di suoni di origine differente). Alla base di questo tipo di percezione c'è una rappresentazione cognitiva del suono nel dominio della frequenza: il nostro orecchio interno contiene cellule acustiche sensibili a frequenze diverse, e queste cellule compiono un'operazione concettualmente simile a una trasformata di Fourier, salvo il fatto che l'ambito frequenziale a cui esse sono sensibili è limitato a una banda passante compresa approssimativamente (e ottimisticamente) tra i 20 Hz e i 20 kHz.

D'altra parte, la nostra percezione del suono avviene anche in termini di eventi che si susseguono nel tempo — fonemi, note, articolazioni interne del suono. Questo avviene in particolare per eventi la cui frequenza di produzione è più bassa di 20 Hz. Intuitivamente, la dimensione temporale del suono ci appare perfettamente contrapposta rispetto a quella frequenziale e indipendente da essa (potremmo dire *ortogonale*). Una partitura tradizionale ci offre una rappresentazione simbolica di questo paradigma di cognizione del suono: sull'asse orizzontale il tempo, su quello verticale la frequenza — o, nel caso di più suoni simultanei, le frequenze. Non deve però mai sfuggirci il fatto fondamentale per cui queste due dimensioni percettivamente contrapposte, il tempo e la frequenza, sono in realtà due maniere diverse di percepire fenomeni appartenenti alla stessa categoria, cioè variazioni della pressione dell'aria nel tempo.

Possiamo allora dire che la nostra percezione avviene secondo una rappresentazione frequenziale *a breve termine*: nel breve periodo, cioè per frequenze relativamente alte, si comporta come una rappresentazione nel dominio della frequenza; nel lungo periodo, cioè per frequenze relativamente basse, si comporta come una rappresentazione nel dominio del tempo.

Esiste una variante della trasformata di Fourier, chiamata *trasformata a breve termine di Fourier* (*short-term Fourier transform* o *STFT*), che re-

stituisce una scomposizione sinusoidale di un segnale limitato nel tempo (e, di conseguenza, anche nella frequenza). Non entriamo ora nei suoi dettagli: possiamo dire che, anche se dal punto di vista matematico il suo risultato è *esatto*, questo in generale non coincide però esattamente con la nostra percezione cognitiva del suono, che avviene secondo meccanismi molto più complessi della STFT. Tuttavia, spesso ne fornisce un'approssimazione utile alla descrizione del fenomeno sonoro, e di fatto i vari analizzatori spettrali che utilizziamo sono basati il più delle volte su questo strumento matematico.⁴

La STFT ha alcune limitazioni intrinseche. In primo luogo, il risultato che fornisce non è continuo, ma è costituito da “finestre” che discretizzano sia l'asse temporale che quello frequenziale; a una maggiore risoluzione temporale corrisponderà una peggiore risoluzione frequenziale, e viceversa. Inoltre, l'asse frequenziale è diviso in maniera lineare: ciascuna finestra frequenziale potrebbe avere, per esempio, una larghezza di 100 Hz. Poiché però la nostra percezione dell'altezza non è lineare ma approssimativamente logaritmica rispetto alla frequenza, questo vuol dire che i 100 Hz nella regione grave copriranno uno spazio percettivamente molto più ampio degli stessi 100 Hz nella regione acuta. Di conseguenza, dal punto di vista percettivo la STFT tende a dare risultati decisamente poco accurati rispetto alle regioni spettrali gravi, e inutilmente dettagliati rispetto a regioni spettrali estremamente acute. Esistono trasformate di tipo diverso, chiamate in generale *constant Q transform*, che organizzano lo spazio delle frequenze in maniera più prossima alla nostra percezione. Dal punto di vista matematico si tratta però di meccanismi molto più complessi, che vengono adottati decisamente più di rado rispetto alla STFT.

2.5.5 Spettrogrammi

Una rappresentazione grafica nel dominio del tempo dell'andamento spettrale a breve termine del suono può essere molto utile per “leggerne” il comportamento; in linea generale, è una rappresentazione molto più eloquente e leggibile di una rappresentazione interamente nel dominio del tempo, perché più prossima alla nostra percezione. Lo spettrogramma è allora tipicamente una rappresentazione tridimensionale, con il tempo sull'asse orizzontale, la frequenza sull'asse verticale e l'ampiezza sull'asse della profondità, che spesso è rappresentato tramite una scala di grigi o di colori. Il calcolo dello spettrogramma, che è spesso basato sulla STFT, non fornisce un risultato continuo, ma suddivide sia l'asse del tempo che quello della frequenza in

⁴In informatica musicale si parla spesso di *fast Fourier transform*, o *FFT*, che è un algoritmo per il calcolo della STFT.

“finestre” discrete: maggiore è la risoluzione sull’asse temporale e minore sarà quella sull’asse frequenziale, e viceversa.