Campionamento e Sintesi 1

Andrea Agostini

11 maggio 2023

Indice

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

La pressione atmosferica standard al livello del mare è poco più di 100 000 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 µPa e 20 Pa.

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 µPa e 20 Pa). 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 343 m/s. Questo vuol dire che, approssimativamente, il suono in queste condizioni percorre 30 cm in 1 ms; 1 m in 3 ms; e 1 km in 3 s.

[1] Questo 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.

1.2 Ampiezza e rappresentazione nel dominio del tempo

Ci interessa individuare alcune grandezze che variano in maniera idealmente proporzionale[2] tra loro, per lo stesso fenomeno sonoro:

- lo spostamento dell'emittente s_e
- 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

$$s_e \propto s_r \propto \Delta p \propto V \tag{1.1}$$

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

$$A \propto s_e \propto s_r \propto \Delta p \propto V \tag{1.2}$$

[2]Nel 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.

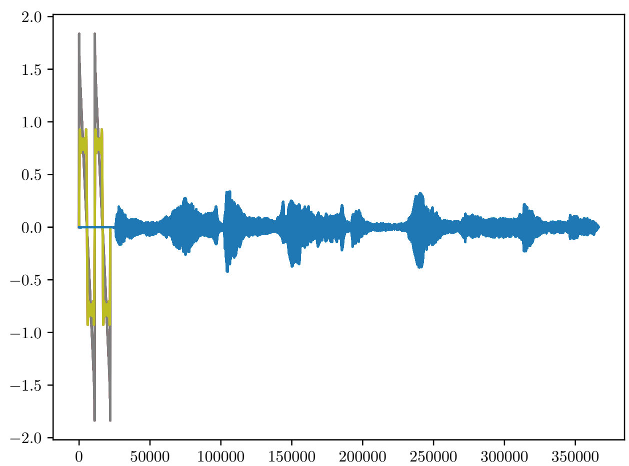

Figura 1.1: A PGF histogram from `matplotlib`.

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:

Ampiezza istantanea

L'*ampiezza istantanea* è l'ampiezza misurata a un istante specifico del fenomeno sonoro. È un valore reale[3] *segnato*[4] : 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 (cioè, dal punto di vista del ricevente, una depressione).

Le ragioni per la scelta dell'istante a cui compiere la misurazione possono essere varie, ma in generale l'ampiezza istantanea viene misurata ripetutamente, a intervalli di tempo regolari, nelle operazioni di campionamento del segnale: se le misure sono abbastanza frequenti, la successione delle ampiezze istantanee può costituire una buona approssimazione numerica dell'andamento del segnale. Di fatto, virtualmente tutte le tecniche di campionamento digitale del segnale audio sono basate su questo principio.

Di per sé, l'ampiezza istantanea 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).

Ampiezza di picco

L'*ampiezza di picco* è riferita non a un istante ma a un intervallo di tempo, ed è l'ampiezza di valore assoluto massimo entro tale intervallo. 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: è, a tutti gli effetti, l'ampiezza istantanea rilevata all'istante in cui questa è massima in valore assoluto.

Un caso tipico in cui è importante conoscere l'ampiezza di picco è la regolazione del guadagno in un sistema di registrazione: si producono segnali

---

[3] I numeri reali, appartenenti all'insieme $\mathbb{R}$, sono tutti i numeri rappresentabili su una retta. Per chi non ha mai incontrato i numeri complessi, appartenenti all'insieme $\mathbb{C}$ e rappresentabili su un piano, l'insieme $\mathbb{R}$ è, intuitivamente, l'insieme di *tutti i numeri*. Purtroppo o per fortuna, la matematica moderna è un po' più complicata di così. Dovremo accennare altre volte a $\mathbb{C}$, perché i numeri complessi sono molto utili per rappresentare certi aspetti dei segnali audio.

[4] Cioè, potenzialmente negativo.

forti e si regola il livello di premplificazione in maniera che l'ampiezza di picco ricevuta dal registratore (o, in un sistema digitale, dal convertitore) non ecceda il livello massimo che questo può trattare. In questo caso, come in molti altri, non è interessante sapere in quale direzione è stata rilevata la massima ampiezza: per questa ragione, viene normalmente considerato il valore assoluto dell'ampiezza di picco.

L'ampiezza di picco può essere raggiunta più volte nell'intervallo di tempo considerato, anche con segni diversi: per esempio, una sinusoide correttamente simmetrica rispetto allo 0 raggiunge la sua ampiezza di picco due volte per ogni ciclo, una volta dal lato positivo e una dal lato negativo.

Ampiezza picco-picco

L'*ampiezza picco-picco* è il valore assoluto della differenza tra la massima e la minima ampiezza entro un dato intervallo di tempo. In questo caso, "massima" e "minima" vanno intese in senso numerico, considerando il segno, e non in valore assoluto: cioè, le ampiezze negative saranno tutte considerate più piccole delle ampiezze positive (e quindi, per esempio, un'ampiezza di -0.5 è considerata più piccola di un'ampiezza di 1; e un'ampiezza di 0.1 è considerata più grande di una di -0.5).

Intuitivamente, l'ampiezza picco-picco calcola la distanza tra i valori estremi di ampiezza istantanea che il segnale tocca entro l'intervallo temporale. Nel caso di una forma d'onda perfettamente centrata sullo 0, essa coincide con il doppio del valore assoluto dell'ampiezza di picco. Per esempio, il segnale nell'intervallo considerato potrebbe avere ampiezza massima a 0.7 e minima a -0.7: entrambi i valori possono essere considerati ampiezza di picco, visto che il loro valore assoluto 0.7 è identico. In questo caso, l'ampiezza picco-picco, cioè il valore assoluto della differenza tra massimo e minimo è $0.7 - (-0.7) = 0.7 + 0.7 = 1.4$.

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),[5] l'ampiezza picco-picco fornisce l'unica possibile misura affidabile. Ad esempio, se il segnale descritto sopra si trovasse traslato di -0.5 sull'asse verticale, il suo massimo sarebbe $0.7 - 0.5 = 0.2$ e il minimo $-0.7 - 0.5 = -1.2$. Questo però non modificherebbe l'ampiezza picco-picco,

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

che sarebbe $0.2 - (-1.2) = 1.4$. L'ampiezza picco-picco, secondo la definizione data sopra, non è mai negativa. Inoltre, in un segnale bipolare (cioè, un segnale che attraversa la linea dello 0) è sempre maggiore dell'ampiezza di picco.

Ampiezza RMS

L'*ampiezza RMS* (da *root-mean-square*: vedremo tra poco perché) è una misura presa rispetto a un intervallo di tempo, e tiene conto dell'andamento dell'ampiezza nell'intervallo. 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. L'ampiezza RMS non è mai maggiore dell'ampiezza di picco, e in generale — salvo i casi di forme d'onda rettangolari e di segnali costanti — è minore di essa.

La definizione di ampiezza RMS per un segnale campionato in ampiezza a intervalli di tempo regolari, come i segnali digitali che trattiamo abitualmente, è

$$A_{\mathrm{RMS}} = \sqrt{\frac{A_1^2 + A_2^2 + \ldots + A_n^2}{n}} \tag{1.3}$$

dove n è il numero totale di campioni e A_1 ... 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, da cui il nome.

Se consideriamo invece un segnale continuo nel tempo, com'è il caso per segnali la definizione di RMS deve fare ricorso al concetto di integrale:[6]

$$A_{\mathrm{RMS}} = \sqrt{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} [A(t)]^2 \mathrm{d}t}$$

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.

[6] Un 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'ampiezza RMS non è mai negativa.[7]

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:

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:

$$W = F \cdot s \tag{1.4}$$

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.[8] 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

[7] Volendo essere pignoli dovremmo dire che ogni numero reale positivo ha due radici quadrate, una positiva e una negativa: ma per il calcolo della RMS consideriamo solo quella positiva. D'altra parte, solo numeri reali non negativi hanno una radice quadrata reale: ma la formula della RMS mette sotto radice una somma di quadrati, e il quadrato di qualsiasi numero reale è non negativo. Tutto questo per dire che la RMS è ben definita.

[8] Questa è una semplificazione: nelle reazioni nucleari, la materia viene trasformata in energia; ed è teoricamente possibile trasformare energia in materia. Nella meccanica classica, però, l'energia è sempre conservata.

proporzionali all'ampiezza:[9]

$$E \propto \Delta p^2 \propto \Delta A^2 \tag{1.5}$$

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:

$$P = \frac{W}{t} \tag{1.6}$$

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 (perché il lavoro è concentrato in un tempo più breve), minore nel secondo. A parità di tempo, la potenza è proporzionale all'energia e quindi al quadrato dell'ampiezza:

$$P \propto E \propto \Delta p^2 \propto A^2 \tag{1.7}$$

Intensità

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

$$I = \frac{P}{a} \tag{1.8}$$

L'intensità sonora non ha una sua specifica unità di misura, ma è misurata 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

[9] La 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/nll/?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.

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

$$a = 4\pi r^2 \tag{1.9}$$

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:

$$I \propto \frac{1}{r^2} \tag{1.10}$$

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:

$$I \propto P \propto E \propto \Delta p^2 \propto A^2 \tag{1.11}$$

Di conseguenza,

$$\frac{1}{r^2} \propto \Delta p^2 \propto A^2 \tag{1.12}$$

e quindi

$$\frac{1}{r} \propto \Delta p \propto A \tag{1.13}$$

Questo è un risultato molto importante: pressione, ampiezza e spostamento sono inversamente proporzionali alla distanza, non al suo quadrato. Si dice quindi che seguono la *legge dell'inverso della distanza*.

1.5 Riepilogo e classificazione delle grandezze

Abbiamo quindi definito due categorie di grandezze fisiche nelle quali ci imbatteremo 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:

$$W \propto \Delta p^2 \implies \Delta p \propto \sqrt{W} \tag{1.14}$$

Si tratta di grandezze a cui non si applica principio di conservazione e che variano in maniera inversamente 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 Intermezzo: logaritmi — un ripasso

Il prossimo concetto che introdurremo richiede uno strumento matematico che dovresti già conoscere, ma che potrebbe meritare un veloce ripasso: il *logaritmo*.

Il logaritmo è definito come l'operazione inversa rispetto all'elevamento a potenza: cioè, se $b^x = y$ allora $\log_b y = x$. Chiamiamo b la *base* del logaritmo.

I logaritmi hanno alcune proprietà molto utili, che sono facilmente dimostrabili ma che mi limiterò a enunciare:

- Il logaritmo di 1 in qualsiasi base è 0, perché qualsiasi numero elevato a potenza 0 dà 1.

- Il logaritmo di qualsiasi numero maggiore di 1 è positivo in qualsiasi base; il logaritmo di qualsiasi numero maggiore di 0 e minore di 1 è negativo in qualsiasi base.

- Il logaritmo di 0 e di numeri negativi non esiste.

- La somma di due logaritmi di uguale base è uguale al logaritmo (nella stessa base) del prodotto: $\log_b x + \log_b y = \log_b(xy)$. Questa identità rende tra l'altro evidente il fatto che il logaritmo non è una funzione lineare.

- È possibile trasformare un esponente dentro il logaritmo in un prodotto fuori dal logaritmo: $\log_b x^y = y \log_b x$. Queste due ultime affermazioni possono essere intese intuitivamente pensando che il logaritmo ci permette di "scendere di livello" con le operazioni aritmetiche: trasforma l'elevamento a potenza in moltiplicazione e la moltiplicazione in addizione. Per lungo tempo, questa semplificazione dei calcoli è stata la prima ragione per usare i logaritmi.

- I logaritmi in tutte le basi sono direttamente proporzionali tra loro: cioè, è possibile passare da una base all'altra applicando al logaritmo un coefficiente — più specificatamente, $\log_b x = \frac{\log_c b}{\log_c a}$, dove c è una base scelta arbitrariamente. Questa formula, detta *formula del cambio di base*, ci dice tra l'altro che se conosciamo i logaritmi in una base qualsiasi possiamo calcolare logaritmi in qualsiasi altra base.

- Il logaritmo ci dà una misura dei rapporti tra numeri: la differenza dei logaritmi di coppie di numeri in rapporto 2:1 sarà sempre la stessa (il valore esatto della differenza dipende dalla base del logaritmo). Quindi $\frac{x}{y} = \frac{z}{w} \implies \log_b x - \log_b y = \log_b z - \log_b w$. Per esempio, in base 10 un rapporto di 2:1 corrisponderà a una differenza tra i logaritmi di circa 0.301 (incontreremo di nuovo molto presto questo valore).

- Il logaritmo del reciproco di un numero è il negativo del logaritmo di quel numero: $log_b\frac{1}{x} = -log_b x$.

- Il logaritmo in base 10 ci dà una misura della "lunghezza tipografica" di un numero intero: $log_{10}10 = 1$, $log_{10}100 = 2$, $log_{10}1000 = 3$ e così via; per argomenti intermedi avremo valori di logaritmo intermedi (ma, ovviamente, non *linearmente* intermedi: $log_{10}55 \approx 1,74$, che è molto più vicino a $log_{10}100$ che a $log_{10}10$ anche se 55 è la media esatta tra 10 e 100). Generalizzando ulteriormente, il logaritmo in base b ci dà una misura della "lunghezza tipografica" di un numero rappresentato in base di numerazione b.

1.7 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 analoga 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:

$$L_{dB} = 10 \cdot log_{10}\frac{I}{I_0} \tag{1.19}$$

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:

$$L_{dB} = 10 \cdot log_{10}\frac{E}{E_0} \tag{1.20}$$

$$L_{dB} = 10 \cdot log_{10} \frac{P}{P_0} \tag{1.21}$$

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:

$$L_{dB} = 20 \cdot log_{10} \frac{\Delta p}{\Delta p_0} \tag{1.22}$$

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

$$logx^a = a \cdot logx \tag{1.23}$$

e quindi:

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

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:

$$L_{dB} = 20 \cdot log_{10} \frac{A}{A_0} \tag{1.25}$$

$$L_{dB} = 20 \cdot log_{10} \frac{V}{V_0} \tag{1.26}$$

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

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

  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. Ad ogni modo, il silenzio perfetto non esiste nel mondo fisico, ma 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.

1.8 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, W_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 $\mathrm{dB_{SPL}}$:

$$L_{dB} = 20 \cdot log_{10} \frac{20\text{Pa}}{20\text{µPa}} = 20 \cdot log_{10} 10^6 = 20 \cdot 6 = 120 \tag{1.28}$$

- 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.[10] 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.

[10]Maggiori 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.

1.9 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

$$20 \cdot log_{10} \frac{A}{A_0} = L_{dB} \tag{1.29}$$

  allora

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

- 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 delle proprietà dei logaritmi che abbiamo visto sopra, 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:

$$f = \frac{1}{\Delta t} = \Delta t^{-1} \tag{2.1}$$

e quindi

$$\Delta t = \frac{1}{f} = f^{-1} \tag{2.2}$$

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,

$$f_{Hz} = \frac{1}{\Delta t_s} \tag{2.3}$$

e

$$f_{kHz} = \frac{1}{\Delta t_{ms}} \tag{2.4}$$

ma

$$f_{kHz} = 1000\frac{1}{\Delta t_s} = \frac{1000}{\Delta t_s} \tag{2.5}$$

e

$$f_{BPM} = 60\frac{1}{\Delta t_s} = \frac{60}{\Delta t_s} \tag{2.6}$$

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 sinusoide.

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 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.

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

$$c = 2\pi r \tag{2.7}$$

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

$$360^\circ = \frac{2\pi r}{r} = 2\pi \tag{2.8}$$

Altri misure di angoli che incontreremo spesso sono

$$90^\circ = \frac{\pi}{2} \tag{2.9}$$

e

$$180^\circ = \pi \tag{2.10}$$

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

[1] La 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.

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

$$\cos\alpha = \sin(\alpha + \frac{\pi}{2}) \tag{2.11}$$

Codominio

Il codominio del seno e del coseno — cioè, l'ambito di valori y che una funzione seno o coseno $y = sinx$ o $y = cosx$ 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

$$-1 \leqslant \sin x \leqslant 1 \tag{2.12}$$

e

$$-1 \leqslant \cos x \leqslant 1 \tag{2.13}$$

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

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

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

e

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

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 sinusoide. Una sinusoide 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.

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

$$s(x) = \sin(2\pi x) \tag{2.16}$$

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

$$s(x) = \sin(2\pi f x) \tag{2.17}$$

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

$$s(t) = \sin(2\pi f t) \tag{2.18}$$

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

$$\omega = 2\pi f \tag{2.19}$$

per cui

$$f = \frac{\omega}{2\pi} \tag{2.20}$$

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

$$s(t) = \sin(\omega t) \tag{2.21}$$

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.

Ampiezza

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

$$s(t) = k \cdot \sin(2\pi f t) \tag{2.22}$$

Il valore assoluto di questo coefficiente rappresenta l'ampiezza di picco della sinusoide; il doppio del valore assoluto, $2k$, ne rappresenta l'ampiezza picco-picco. Se k è negativo, la sinusoide 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).

Fase

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

$$s(t) = k \cdot \sin(2\pi f t + \phi) \tag{2.23}$$

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 sinusoide 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, cioè di mezzo giro:

$$k \cdot \sin(2\pi f t + \phi) = -k \cdot \sin(2\pi f t + \phi + \pi) = -k \cdot \sin(2\pi f t + \phi - \pi) \tag{2.24}$$

Inoltre, una rotazione di fase di $\frac{\pi}{2}$ (un quarto di giro, o 90°) corrisponde alla trasformazione della funzione seno in una funzione coseno:

$$k \cdot \sin(2\pi f t + \phi) = k \cdot \cos(2\pi f t + \phi - \frac{\pi}{2}) \tag{2.25}$$

Somme di sinusoidi alla stessa frequenza

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

$$k_1 \cdot \sin(2\pi ft + \phi_1) + k_2 \cdot \sin(2\pi ft + \phi_2) = k \cdot \sin(2\pi ft + \phi) \tag{2.26}$$

Inoltre, una sinusoide con qualsiasi fase può essere espressa come la somma di una sinusoide seno e una sinusoide 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:

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

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, 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,[4] 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 $(\sqrt[12]{2})^2 = \sqrt[6]{2} \approx 1.122$, un valore leggermente più piccolo di $\frac{9}{8} = 1.125$; la terza maggiore, composta da quattro semitoni, ha rapporto pari a $(\sqrt[12]{2})^4 = \sqrt[3]{2} \approx 1.26$, un valore leggermente più grande di $\frac{5}{4} = 1.25$; l'ottava, coerentemente con la sua definizione data sopra, ha rapporto pari a $(\sqrt[12]{2})^{12} = 2$.

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

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

[4] Cioè, secondo il temperamento equabile, utilizzato in maniera quasi ubiqua nella maggior parte della musica prodotta a partire dall'inizio del XIX secolo.

2.5 Spettro di frequenza

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, ciascuna caratterizzata dalla sua frequenza, ampiezza e fase. 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. Dire che una componente non è presente in un segnale equivale a dire che la sua ampiezza è nulla.

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 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.

Segnali periodici, armoniche

Esiste 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 multiple della frequenza fondamentale del segnale.[5] 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,

[5] Questo equivale a dire che le componenti a frequenze diverse dai multipli della fondamentale hanno tutte ampiezza nulla.

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.

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 $s(t)$ su dominio infinito e restituisce un'altra equazione che descrive il segnale come una funzione della frequenza $F(f)$ su dominio infinito. Il codominio di questa funzione $F(f)$ è l'insieme dei numeri complessi $\mathbb{C}$. Questo vuol dire che il valore della funzione $F(f)$, corrispondente alla trasformata di Fourier del segnale originale $s(t)$, per una data frequenza f è un numero complesso; questo rappresenta in maniera compatta l'ampiezza e la fase della sinusoide a frequenza f che, sommata a tutte le altre sinusoidi con ampiezze e fasi ottenute per tutte le altre possibili frequenze, produrrà un segnale identico a quello originale. L'ampiezza di cui stiamo parlando qui è, di fatto, il termine di ampiezza dell'equazione sinusoidale presentata sopra, che abbiamo chiamato k; possiamo quindi considerarlo come l'ampiezza di picco non segnata della corrispondente sinusoide.

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 con le frequenze sull'asse orizzontale e le ampiezze 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.

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. Da questo punto di vista, possiamo dire che nessun fenomeno sonoro che percepiamo è statico: in generale, attribuiremo un inizio e una fine a ciascun fenomeno sonoro; inoltre, le sue qualità ci appariranno come variabili nel tempo: una corda pizzicata, ad esempio, produce un suono che inizia forte e brillante e si estingue lentamente diventando progressivamente più flebile e scuro.[6] Se è vero che la trasformata di Fourier, nella sua formulazione originale e più astratta, tratta il segnale codificando in termini frequenziali il suo comportamento lungo qualsiasi scala temporale, è anche vero che la nostra percezione funziona in maniera molto diversa e, per avere una rappresentazione utile del fenomeno sonoro, è necessario tenerne conto.

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

[6] "Brillante" e "scuro" sono sinestesie che fanno riferimento rispettivamente alla maggiore o minore ampiezza delle componenti spettrali acute di un suono.

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 restituisce 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.[7] Anche la STFT restituisce numeri complessi, che rappresentano i termini di ampiezza non segnata e di fase delle sinusoidi che dobbiamo sommare per ottenere la porzione analizzata del segnale originale.

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 necessariamente 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 diver-

[7]In informatica musicale si parla spesso di *fast Fourier transform*, o *FFT*, che è un algoritmo per il calcolo della STFT. Un altro termine che si incontra spesso è *discrete Fourier transform*, o *DFT*, che fa riferimento esplicito al fatto che il calcolo viene svolto su una sequenza di campioni rilevati a intervalli temporali successivi. In senso proprio, la STFT è una DFT applicata a una porzione temporale breve (o "finestra") di un segnale.

so, 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.

Si può dire, in maniera intuitiva anche se formalmente non precisa, che lo spettro frequenziale restituito dalla STFT non è virtualmente mai esatto rispetto al segnale analizzato. Tuttavia, la STFT è perfettamente invertibile: le inesattezze che essa introduce vengono poi "metabolizzate" quando viene calcolata la trasformata inversa a breve termine, e contribuiscono alla ricostruzione del segnale originale. Inoltre, nonostante la sua apparente imprecisione, l'analisi del fenomeno sonoro restituita dalla STFT è abbastanza buona e prossima alla percezione da essere praticamente utile in una grande varietà di casi. La grande maggioranza degli analizzatori di spettro che incontriamo nei vari programmi audio che usiamo sono basati sulla STFT, e in generale funzionano piuttosto bene. Se tutto questo sembra complicato non preoccuparti: lo è, e ci torneremo nelle prossime annualità del corso, quando esploreremo come la STFT e la sua inversa possono essere usate per operare trasformazioni interessanti del suono.

La matematica della STFT non è banale, e non la approfondiremo oltre. Ti consiglio però questo sito `https://jackschaedler.github.io/circles-sines-signals/index.html` e il bellissimo libro per bambini superintelligenti "Who Is Fourier?" per avere due introduzioni informali ma rigorose sulla questione.

Rappresentazioni grafiche e spettrogrammi

Lo spettro di frequenze di un segnale può essere rappresentato graficamente. Incontriamo generalmente due tipi fondamentali di rappresentazioni grafiche. La prima è la rappresentazione dello spettro istantaneo, che è un grafico bidimensionale con la frequenza sull'asse orizzontale e l'ampiezza, generalmente non segnata, sull'asse verticale;[8] questa rappresentazione descrive lo spettro di un segnale in maniera indipendente dal tempo, assumendo quindi che questo possa estendersi indefinitamente nel futuro e nel passato; o anche lo spettro di un segnale considerato nel breve termine di

[8] In certi casi è utile rappresentare l'ampiezza come un valore segnato: il segno negativo, in questo caso, rappresenta un'inversione di fase di una componente rispetto a quelle con segno positivo.

una data finestra temporale, come per esempio lo spettro prodotto da una corda di chitarra un secondo dopo che è stata pizzicata.

La seconda rappresentazione è lo spettrogramma, che è un grafico tridimensionale che rappresenta una successione temporale di spettri a breve termine. Lo spettrogramma può essere molto utile per "leggere" il comportamento del suono nel tempo e, in generale, è una rappresentazione molto più eloquente e leggibile del grafico della forma d'onda, perché più prossima alla nostra percezione. Lo spettrogramma rappresenta il tempo sull'asse orizzontale, la frequenza sull'asse verticale e l'ampiezza (sempre non negativa) 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 "finestre" discrete: maggiore è la risoluzione sull'asse temporale e minore sarà quella sull'asse frequenziale, e viceversa.

I grafici del segnale nel dominio della frequenza possono mostrare la frequenza e l'ampiezza in maniera lineare o logaritmica (il tempo degli spettrogrammi invece è generalmente mostrato in maniera lineare).

Se la frequenza è rappresentata in maniera lineare, una stessa differenza frequenziale corrisponderà a una stessa distanza nel grafico: per esempio, una differenza di 1000 Hz potrebbe corrispondere a 1 cm sull'asse orizzontale. Lo stesso vale per l'ampiezza e le grandezze a essa linearmente correlate: per esempio, una differenza di 0.1 in ampiezza normalizzata, o una differenza di 0.5 V, potrebbero corrispondere a 1 cm sull'asse verticale.

Se invece la frequenza è rappresentata in maniera logaritmica, uguali rapporti di frequenza corrisponderanno a uguali distanze nel grafico: per esempio, un rapporto frequenziale di 2:1, ovvero un intervallo d'ottava, potrebbe corrispondere a 1 cm nel grafico; questo vuol dire che una rappresentazione frequenziale logaritmica è lineare nello spazio delle altezze, per cui intervalli uguali corrisponderanno a distanze grafiche uguali. Uno spettrogramma con rappresentazione logaritmica delle frequenze è un oggetto piuttosto simile a una partitura. Se l'ampiezza è rappresentata su una scala logaritmica, questo equivale a dire che le differenze in decibel sono mostrate linearmente: per esempio, un rapporto di ampiezze di 2:1, ovvero una differenza di circa 6 dB, potrebbe corrispondere a 1 cm.

2.6 Categorie di spettri

Possiamo suddividere gli spettri in alcune categorie differenti.

Spettri armonici

Come già detto, un segnale periodico nel dominio del tempo produce uno spettro armonico. Questo vuol dire che le frequenze di tutte le sue componenti sono multiple della frequenza fondamentale, corrispondente all'inverso del periodo del segnale.

In una rappresentazione grafica nel dominio della frequenza rappresentata linearmente, tutte le sue componenti sono equidistanti sull'asse delle ascisse, e la stessa distanza c'è tra la fondamentale e l'asse delle ordinate.

Il caso più semplice di spettro armonico è quello della sinusoide, che ha una sola componente di ampiezza non nulla, alla frequenza fondamentale.

È utile conoscere le conformazioni spettrali di alcuni altri segnali periodici:

- L'onda a dente di sega è composta da una somma di sinusoidi armoniche con termine di fase nullo. L'ampiezza di ciascuna armonica è inversamente proporzionale al suo ordine: se la fondamentale (armonica 1) ha ampiezza 1, allora la seconda armonica (di frequenza doppia) ha ampiezza $\frac{1}{2}$, la terza armonica (di frequenza tripla rispetto alla fondamentale) ha ampiezza $\frac{1}{3}$ e così via. In termini percettivi, questo produce un suono molto brillante: la decima armonica ha ampiezza $\frac{1}{10}$, corrispondente a -20 dB, rispetto alla fondamentale; la centesima armonica ha ampiezza $\frac{1}{100}$, corrispondente a -40 dB, il che vuol dire che nelle giuste condizioni può essere ancora chiaramente percepibile.

- L'onda quadra è composta da una somma di sinusoidi armoniche con termine di fase nullo. L'ampiezza di ciascuna armonica di ordine dispari è inversamente proporzionale al suo ordine; le armoniche di ordine pari hanno ampiezza nulla. Quindi, se la fondamentale (armonica 1) ha ampiezza 1, allora la terza armonica (di frequenza tripla) ha ampiezza $\frac{1}{3}$, la quinta armonica (di frequenza quintupla rispetto alla fondamentale) ha ampiezza $\frac{1}{5}$ e così via. Le armoniche di ordine 2, 4, 6 eccetera non sono presenti.

- Nell'onda triangolare, l'ampiezza di ciascuna armonica di ordine dispari è inversamente proporzionale al quadrato del suo ordine; le armoniche di ordine pari hanno ampiezza nulla. Quindi, se la fondamentale (armonica 1) ha ampiezza 1, allora la terza armonica (di frequenza tripla) ha ampiezza $\frac{1}{9}$, la quinta armonica (di frequenza quintupla rispetto alla fondamentale) ha ampiezza $\frac{1}{25}$ e così via. Le armoniche di ordine 2, 4, 6 eccetera non sono presenti. Le armoniche di ordine dispari hanno termini di fase alternati: se la fondamentale ha fase nulla, allora la terza armonica ha fase π (cioè fase opposta rispetto alla fondamentale), la quinta ha fase nulla, la settima fase π e così via.

- Il treno d'impulsi è composto da una somma di armoniche tutte alla stessa ampiezza, con termine di fase nullo. Questo produce un segnale che percettivamente è estremamente, innaturalmente brillante, dal momento che in linea generale gli spettri che troviamo nel mondo fisico tendono ad avere ampiezze digradanti al crescere della frequenza. Si può fare una considerazione interessante sul treno d'impulsi: se allontaniamo gli impulsi tra loro nel tempo, stiamo aumentando il loro periodo, quindi abbassando la frequenza fondamentale, quindi "avvicinando" tra loro le armoniche. Più gli impulsi sono radi, più lo spettro è "fitto" di armoniche di pari ampiezza. Se gli impulsi diventano infinitamente radi — cioè, se la frequenza diventa infinitesima, cioè se si considera un singolo impulso — allora lo spettro sarà costituito da armoniche infinitamente vicine. Questo caso particolare è il cosiddetto *delta di Dirac*, una funzione costituita da un singolo impulso non nullo di durata infinitesima, che contiene tutte le frequenze dello spettro, in un dominio infinito, in eguale quantità. Quando battiamo le mani per saggiare la risposta acustica di una stanza, o pizzichiamo una corda con un plettro, stiamo approssimando il comportamento di un delta di Dirac: idealmente, immettiamo nel sistema (che sia la stanza o la corda) ogni possibile frequenza attraverso un impulso estremamente breve, in maniera che il sistema risuoni nella maniera più ricca possibile producendo quella che viene chiamata la sua *risposta d'impulso*.

Segnali con la stessa distribuzione spettrale di ampiezze ma fasi diverse rispetto a quelle indicate sopra avranno, in generale, lo stesso "suono" ma forme d'onda diverse. Questa considerazione non vale solo per i tipi di segnali che abbiamo descritto, ma per qualsiasi altro segnale.

Osserviamo infine che sinusoide, onda quadra e onda triangolare condividono la caratteristica di essere *antisimmetriche*: la parte negativa del segnale, come rappresentato nel dominio del tempo, è speculare rispetto alla sua parte positiva. Rispetto al periodo della fondamentale, le sue armoniche dispari sono antisimmetriche, quelle pari no: per questa ragione i segnali antisimmetrici contengono solo armoniche di ordine dispari (questo è vero anche per la sinusoide, la cui unica componente è di ordine 1) e sono quindi detti essi stessi *dispari*.

Spettri inarmonici a componenti discrete

Esiste una categoria di spettri che non consideriamo armonici, e che sono costituiti da componenti sinusoidali individuali organizzate in maniera non regolare. Da un punto di vista puramente teorico bisogna fare una distinzione di una certa sottigliezza: se le frequenze delle componenti hanno rapporti razionali tra loro, in senso proprio il segnale sarà sempre periodico, con un periodo pari al minimo comune multiplo dei periodi delle componenti (e quindi con una frequenza pari al massimo comun divisore delle frequenze delle componenti). Per esempio, se un segnale è composto da una componente a 100 Hz, una a 212 Hz e una a 213.5 Hz ci sarà comunque una periodicità di 0.5 Hz. Solo rapporti di frequenza irrazionali (per esempio, tra $100\sqrt{2}$ Hz e 200π Hz) non definiscono un massimo comun divisore e quindi produrranno un segnale puramente inarmonico.

La nostra percezione di armonicità (e con essa la capacità di individuare percettivamente una frequenza fondamentale) tende però ad attenuarsi e a svanire in almeno due circostanze: quando la frequenza fondamentale scende al di sotto dei 20 o 30 Hz e quando il segnale non contiene armoniche di ordine basso. Nel primo esempio entrambe le condizioni sono vere: la fondamentale sarebbe estremamente grave, ampiamente al di fuori della nostra banda di percezione frequenziale, e le tre componenti ne costituirebbero rispettivamente la 200esima, 424esima e 427esima armonica. Nella pratica, quindi, quando una di queste due circostanze si avvera consideriamo inarmonico lo spettro del segnale. È facile riconoscere questo tipo di spettri guardando un loro grafico nel dominio della frequenza: se l'occhio non distingue l'equidistanza delle componenti, possiamo assumere che la qualità sonora dello spettro sarà di tipo inarmonico.

Esistono alcuni oggetti (come le campane e le piastre di metallo), oltre ad alcune specifiche tecniche di sintesi e trattamento sonora (come la modulazione ad anello, il *frequency shifting*, la sintesi per modulazione di frequenza) che tendono a produrre spettri ascrivibili a questa tipologia.

Spettri rumorosi (a componenti continue)

Non tutti gli spettri sono scomponibili in singole componenti sinusoidali a frequenze precisamente individuate: è infatti possibile che l'ampiezza sia non nulla per componenti a frequenze infinitamente vicine. Un caso che abbiamo già incontrato è il delta di Dirac. Questo equivale a dire che esistono regioni dello spazio delle frequenze in cui l'energia è distribuita in maniera continua. In termini di rappresentazione nel dominio del tempo, questo corrisponde a segnali caotici, il cui comportamento può essere descritto esclusivamente in maniera statistica. In termini di percezione sonora, questo produce suoni inarmonici che classifichiamo intuitivamente come "rumorosi": molte consonanti, gli strumenti a percussione ad altezza indeterminata come i piatti, il vento tra le foglie, una cascata.

Se in un suono a componenti continue emergono alcune bande di frequenza relativamente strette con ampiezze relativamente alte, possiamo attribuirgli un'intonazione: un esempio classico è il suono del flauto, soprattutto quando prodotto con molta "aria". In senso proprio, dovremmo dire che nessun suono del mondo reale ha componenti precisamente sinusoidali: a un'analisi abbastanza accurata, ciascuna delle cosiddette sinusoidi si rivelerebbe come una componente continua a banda molto stretta.

Tra i segnali a componenti continue, ricordiamo almeno:

- Il *rumore bianco*, il cui spettro contiene energia a tutte le frequenze, e la stessa quantità di energia per banda di frequenza: per esempio, la quantità di energia nella banda tra 100 e 200 Hz è uguale a quella nella banda tra 1100 e 1200 Hz. Questo rende il rumore bianco estremamente brillante e aspro. Nel dominio del tempo, il rumore bianco è modellizzabile come un segnale completamente casuale nel dominio del tempo.

- Il *rumore rosa*, il cui spettro contiene energia a tutte le frequenze, e la stessa quantità di energia per ottava (o, più generalmente, per intervallo nello spazio delle altezze): per esempio, la quantità di energia nella banda tra 100 e 200 Hz è uguale a quella nella banda tra 1000 e 2000 Hz.

Altre considerazioni

Per finire, aggiungiamo alcune considerazioni finali sul rapporto tra il segnale visto nel dominio del tempo e il suo spettro.

DC offset

Fin qui abbiamo sempre assunto che il segnale considerato nel dominio del tempo sia sempre esattamente centrato rispetto all'asse dello 0 (che può essere considerato la pressione d'ambiente, l'assenza di voltaggio, la posizione di riposo della singola particella d'aria o della membrana del timpano e del microfono). È possibile però che il segnale che consideriamo sia invece traslato rispetto allo 0: questo può accadere come risultato di un particolare tipo di azione meccanica (per esempio, l'arco del violino tende a "tirare" la corda in una sola direzione, e quindi produrrà un segnale decentrato), o di specifiche tecniche di sintesi e trattamento (come, ancora una volta, la sintesi per modulazione di frequenza), o per problemi elettrici nella catena elettroacustica (come l'immissione nel percorso del segnale di una componente di corrente continua, da cui il nome di *DC offset* che viene dato alla traslazione verticale della forma d'onda rispetto alla sua posizione normale). Questo tipo di traslazione può essere visto come la somma di un valore costante al segnale; ma un valore costante può essere ottenuto da una funzione sinusoidale con frequenza nulla e termine di fase e ampiezza non nulla: $k \cdot \sin(2\pi \cdot 0t + \phi) = k \cdot \sin\phi$ per qualsiasi t. Ha quindi senso considerare nella descrizione spettrale di un segnale anche una componente di frequenza 0 e fase di $\frac{\pi}{2}$, la cui ampiezza k corrisponderà alla traslazione sull'asse verticale della forma d'onda, dal momento che $\sin\frac{\pi}{2} = 1$.

Altre relazioni tra rappresentazione temporale e frequenziale

Esistono poi alcune caratteristiche della rappresentazione nel dominio del tempo che possono fornire indizi interessanti sullo spettro del segnale: naturalmente, la periodicità o aperiodicità può, almeno in casi semplici, essere stimata guardando la forma d'onda. Abbiamo inoltre già parlato della relazione tra assenza di armoniche pari e antisimmetricità. Un'altra osservazione utile è che la presenza di armoniche acute produce, in generale, una forma d'onda più "spigolosa" — e che, al converso, "spigoli" nel segnale produrranno spettri ricchi di armoniche acute, fino al caso limite di una discontinuità, un "salto verticale", nella forma d'onda, che teoricamente produce un delta di Dirac contenente ogni possibile frequenza a pari intensità (il famigerato "click").

Analisi della scena uditiva

In tutto questo, non abbiamo accennato, se non di sfuggita, al caso comunissimo in cui più corpi emettano vibrazioni acustiche simultaneamente, nella stessa scena uditiva — che siano le corde di un pianoforte o corpi senza nessun tipo di relazione materiale o causale. Da un punto di vista puramente astratto, matematico, potremmo dire che non c'è nulla da aggiungere: due note diverse suonate simultaneamente produrranno un segnale sonoro dato dalla somma algebrica delle loro forme d'onda nel tempo, e tutte le considerazioni fatte fin qui restano immutate. Quindi, tipicamente, le strutture spettrali armoniche di ciascuna delle due note si sommeranno, tipicamente producendo uno spettro che all'occhio apparirà come uno spettro inarmonico a componenti discrete.

L'esperienza però ci dice che la percezione che abbiamo di questo fenomeno è profondamente diversa: siamo perfettamente in grado di capire che ci sono più sorgenti sonore in gioco, ciascuna delle quali caratterizzata da uno spettro armonico; un ascoltatore allenato è capace di individuarne con sicurezza le relazioni di frequenza ("una terza maggiore"), ascriverle a specifiche categorie timbriche ("un violino e un flauto") e, se ha la fortuna di avere l'orecchio assoluto, discernere la fondamentale prodotta da ciascuna delle due ("un mi e un sol diesis in chiave di violino").

Ancora più evidente è la nostra capacità di selezionare, in una scena piena di suoni che consideriamo di disturbo, una sorgente sonora che ci interessa: dal momento che non viviamo in ambienti perfettamente silenziosi, è ciò che facciamo ogni volta che parliamo con qualcuno. Addirittura, c'è un fenomeno psicoacustico chiamato "effetto cocktail party" che mostra come siamo capaci di selezionare e condurre una conversazione che ci interessa in una stanza piena di altre conversazioni che potenzialmente si svolgono con un tono di voce più alto, come spesso accade in una festa. Il cervello umano è un riduttore di rumore straordinario.

Queste capacità riposano su meccanismi cognitivi estremamente complessi e dei quali abbiamo una comprensione molto limitata: questi meccanismi ci permettono di individuare e separare porzioni di spettro in base alle loro caratteristiche, in maniera raffinatissima, e sono fondamentali nel funzionamento cognitivo dell'udito e nella nostra capacità di utilizzare l'informazione acustica. Qui si aprirebbe il discorso estremamente ampio dell'*analisi della scena uditiva*, trattato in maniera approfondita nel testo classico di Albert S. Bregman intitolato appunto *Auditory Scene Analysis*.