

# Principi per la memorizzazione digitale della musica

Agostino Dovier

Univ. di Udine, DIMI

## 1 Preliminari

Ogni suono/rumore *udibile* è generato da una *vibrazione*. Il corpo vibrante, per essere udito deve *oscillare* con una *frequenza* compresa tra 16/20 Hz e 16/20 kHz.<sup>1</sup> Per fissare le idee, i suoni intuitivamente definiti *bassi* (per esempio quelli prodotti da un contrabbasso) hanno frequenza molto bassa (16–200 Hz), mentre quelli intuitivamente definiti *alti* (per esempio quelli prodotti da un ottavino o da un cimbalo) hanno frequenza molto alta (più di 10 kHz).

Grazie ad un noto teorema dovuto a Fourier, è appurato che ogni suono/rumore può essere scomposto mediante una somma di oscillazioni puramente sinusoidali (si pensi al grafico della funzione trigonometrica *seno*).

In particolare, per i *suoni*, vi è una semplice relazione che lega le frequenze delle varie sinusoidi presenti.

Il criterio di *accordatura* degli strumenti è basato su una nota campione: si impone che il LA immediatamente sopra il DO centrale del pianoforte oscilli a 440 Hz;<sup>2</sup> il LA dell’ottava successiva viene accordato in modo da avere ‘lo stesso suono’ o, in termini fisici la frequenza *doppia*. Il LA dell’ottava ancora successiva avrà frequenza doppia del doppio di 440 Hz e così via. Simmetricamente, per quanto riguarda le ottave più basse, la frequenza va di volta in volta dimezzata.

Tra una ottava e l’altra bisogna trovar spazio per 12 semitoni.<sup>3</sup> Senza entrare in dettaglio sulle ragioni di perchè ciò avvenga e ricordando che quasi nessun musicista si serve di questa legge fisica (l’accordatura ad “orecchio” è spesso priva di errore) si vuole evidenziare che il *rapporto* tra la frequenza di una nota e la frequenza della nota immediatamente precedente (un semitono sotto) è *fisso* e vale  $\sqrt[12]{2}$ . Come verifica che ciò abbia senso si osservi che la frequenza della nota sita 12 semitoni sopra un’altra di frequenza  $f$  (la sua ottava) è

$$f \times \underbrace{\sqrt[12]{2} \times \dots \times \sqrt[12]{2}}_{12} = f \times 2.$$

<sup>1</sup> Un Hz (Hertz) sta ad indicare una vibrazione al secondo. Qui “k” indica 1000, non 1024 (=  $2^{10}$ ) come nella misura della memoria di un calcolatore.

<sup>2</sup> La regolazione avviene agendo sulla tensione delle corde vibranti; il lettore interessato può trovare informazioni dettagliate sulle relazioni che intercorrono tra corde (lunghezza, diametro, tensione) e frequenza generata dalla loro oscillazione in qualunque testo di FISICA.

<sup>3</sup> Ad esempio DO, DO $\sharp$ , RE, RE $\sharp$ , MI, FA, FA $\sharp$ , SOL, SOL $\sharp$ , LA, LA $\sharp$ , SI.

In base a tale regola risulta facile calcolare la frequenza del DO sito un'ottava sopra al DO centrale:

$$440\text{Hz} \times \sqrt[12]{2} \times \sqrt[12]{2} \times \sqrt[12]{2} = 523.2\text{Hz}$$

Dimezzando e duplicando si possono quindi ottenere le frequenze base di tutti i DO udibili:

ottava	frequenza (Hz)
0	16.3
1	32.7
2	65.4
3	130.8
4	261.6

ottava	frequenza (Hz)
5	523.2

ottava	frequenza (Hz)
6	1046.5
7	2093.0
8	4186.4
9	8372.0
10	16744.0

Quando un corpo vibra alla frequenza  $f$  in esso sono pure presenti vibrazioni alla frequenze  $2 \times f, 3 \times f, \dots$ . Tali vibrazioni producono i cosiddetti *suoni armonici*. La differenza di *timbro* che caratterizza i vari strumenti musicali è determinata dalla diversa ampiezza dei suoni armonici in ogni singola nota eseguita.<sup>4</sup>

Anche i rumori sottostanno alla legge di Fourier, tuttavia non vi è alcuna regolarità tra le varie sinusoidi presenti.

## 2 Campionatura

La tecnica della campionatura viene utilizzata per la registrazione digitale di un brano musicale. Viene usata nei *compact disc*, nei *DAT*, e, grazie al calo del costo della memoria di massa dei calcolatori, nelle sale di incisione, per memorizzare su HD.

Il principio utilizzato è molto semplice; considerate, ad esempio, le misure della temperatura corporea in un ospedale. Esse vengono 'registrate' ad intervalli di tempo regolari e riportate in un diagramma; alla fine i punti ottenuti sono uniti insieme per ottenere un'approssimazione (è infatti precisa solo negli istanti in cui viene misurata) dell'andamento nel tempo della temperatura del paziente. Se viene raddoppiato il numero di misurazioni al giorno (la frequenza) allora il diagramma diviene una migliore approssimazione dell'andamento reale.

La campionatura di un segnale sonoro avviene esattamente in questo modo: la sorgente sonora viene captata mediante un microfono e tradotta in un segnale elettrico al fine di poter essere più facilmente misurata. Ad intervalli di tempo regolari, viene rilevato il valore del segnale ottenuto. Vi sono due problemi:

<sup>4</sup> A questo proposito, è facile verificare numericamente, ad esempio, che se  $f = 532.2$  (frequenza del DO centrale), allora  $2 \times f, 4 \times f, 8 \times f, \dots$  sono le frequenze dei successivi DO. Inoltre  $3 \times f, 6 \times f, \dots$  sono frequenze relative alla nota SOL, mentre  $5 \times f, 10 \times f, \dots$  sono relative alla nota MI. Dunque una singola nota 'reale' nasconde al suo interno un accordo maggiore.

- quanto dev'essere piccolo l'intervallo di tempo per garantire una memorizzazione sensata (ovvero quale dev'essere la frequenza di campionamento)?
- Come convertire la misurazione (tipicamente un valore in Volt) in un dato in grado di essere memorizzato in un calcolatore?

Per quanto riguarda il primo problema, si è sperimentalmente accertato che, per garantire la fedeltà richiesta dagli audiofili, la frequenza di campionamento dev'essere almeno doppia della massima frequenza del segnale da registrare (nel nostro caso 20 kHz). I compact disc campionano alla frequenza di 44.1 kHz, mentre i più moderni *DAT* (registrazione digitale su cassetta) funzionano con la frequenza di 47 kHz. Ben più basse le frequenze di campionamento utilizzate dai dispositivi atti a registrare solo la voce umana (ad es. segreterie telefoniche): 8–10 kHz sono più che sufficienti.

Per il secondo problema, si assuma esistere un valore massimo per il volume prodotto dagli esecutori; a tale valore corrisponderà un segnale oscillante di ampiezza massima (sia positivo che negativo). Si assuma inoltre che il dispositivo di registrazione digitale disponga di due bytes (16 bit) per codificare il suono. Con 16 bit, in complemento a 2, si possono rappresentare i valori interi da  $-2^{15} = -32768$  a  $2^{15} - 1 = 32767$ . Al valore massimo di cui sopra sarà assegnato il valore  $-32767$  o  $32767$  a seconda che la lettura sia effettuata sulla cresta della semionda negativa o positiva rispettivamente. Al valore di tensione nullo (assenza di musica, ma anche istante in cui il segnale passa da un segno al segno opposto) si fa corrispondere lo 0. Gli altri valori binari sono utilizzati per codificare i valori di tensione intermedi; tuttavia ciò non avviene nel modo più intuitivo, ovvero con una connessione lineare tra valore in tensione e valore codificato, ma utilizzando una codifica *logaritmica* del valore. La descrizione formale di ciò va però al di là dello scopo di queste note.

Per codificare (con una registrazione 'mono') 6 minuti (= 360s) di musica con la campionatura a 44.1 kHz usando una lettura a 16 bit (= 2 bytes), sono dunque necessari

$$360 \times 44100 \times 2 = 31752000 \approx 32MB,$$

indipendentemente da numero e tipo di strumenti contemporaneamente utilizzati per produrre la musica.

### 3 Memorizzazione nota per nota

Un secondo metodo per la memorizzazione della musica è quello utilizzato dai sistemi *MIDI*,<sup>5</sup> sistemi integranti calcolatore e strumenti musicali elettronici.

Con questa tecnica non si registra l'esecuzione, bensì si codifica la partitura. Analizziamo le proprietà di ogni nota necessarie per la sua memorizzazione.

<sup>5</sup> Acronimo per Musical Instrument-Device Interface.

**nome:** come mostrato nella sezione 1 con 10 ottave si riesce a coprire tutto lo spettro di frequenze *udibili*. In ogni ottava vi sono 12 note: con 120 valori si riesce a codificare il valore di ogni nota; si avrà cura di riservare un numero (ad esempio lo 0) per rappresentare la pausa: 7 bit sono pertanto sufficienti per memorizzare questo campo.

**valore:** quant'è il valore (ovvero la *durata*) possibile per ciascuna nota? Vanno rappresentate la semibreve (4/4), la minima (2/4), la semiminima (1/4), la croma (1/8), la semicroma (1/16), la biscroma (1/32), e la semibiscroma (1/64): 7 casi. Le stesse note possono essere seguite da un punto, da due punti, possono stare entro una terzina, entro una quintina o entro una settina. In tutto serviranno 42 possibilità: 6 bit sono sufficienti.

**intensità:** una nota può essere eseguita *ppp*, *pp*, *p*, *mp*, *mf*, *f*, *ff*, *fff*: 8 possibilità. 3 bit sono sufficienti.

**legatura:** una nota può essere o meno legata alla seguente: 1 bit.

Poco più di 2 byte sono dunque necessari per memorizzare le informazioni relative alla singola nota. Riducendo il numero di possibilità per il campo durata (ad esempio le note con puntino e doppio puntino possono essere descritte mediante legatura di valore, oppure si impediscono gruppetti rari quali le settine) si può facilmente arrivare a 2 byte.

L'intero brano avrà bisogno poi di alcune informazioni di controllo, quali

- tonalità del brano;
- intonazione delle singole tracce (ad esempio uno spartito per tromba sarà scritto un tono sopra) ed indirizzamento delle stesse ai diversi *canali* d'ingresso di un'apparecchiatura riprodotrice (sequencer);
- tempo d'esecuzione;
- informazioni varie quali ritornelli, commenti come testo sullo spartito, etc.

Queste ultime, tuttavia utilizzano una quantità di memoria trascurabile rispetto a quella necessaria per memorizzare i vari spartiti.

Si assuma di scrivere per intero un brano da 6 minuti, per un solo strumento e che il tempo d'esecuzione sia un 4/4 con 120 battiti al minuto (dunque 30 battute al minuto). Ciò implica che devono essere scritte (e memorizzate)  $30 \times 6 = 180$  battute. Supponiamo (è chiaramente un eccesso) che nello spartito di ogni strumento in ogni battuta vi siano mediamente 16 note, allora saranno necessari

$$180 \times 16 \times 2 = 5760B$$

più alcuni per il controllo generale. Supponendo che il brano di 6 minuti sia suonato da un'orchestra di 60 strumenti, con partiture tutte diverse tra loro (non ha più senso considerare 16 note in media a battuta: 8 è già un eccesso), saranno necessari

$$180 \times 8 \times 2 \times 60 = 172800B,$$

ovvero meno di  $200kB$ .

Con questo approccio si ha dunque un notevole risparmio in termini di memoria. I musicisti da *piano-bar* si sono impadroniti di queste moderne tecnologie preparandosi (o procurandosi) una vasta banca dati di brani (le voci impiegate sono di solito 4, con molte ripetizioni e note lunghe: pochissimo spazio è necessario). Il vero problema di questo approccio è il tempo necessario a memorizzare i vari spartiti. Prodotti in grado di leggere un spartito cartaceo (mediante uno *scanner*) e trasformarlo automaticamente in uno spartito *MIDI* non danno ancora l'affidabilità necessaria.