



**Politecnico di Bari  
I<sup>a</sup> Facoltà di Ingegneria**

**Corso di Laurea in Ingegneria Informatica**

**Corso di Fisica Tecnica  
A.A. 2002/2003**

# **Elementi di elaborazione numerica dei segnali audio**

**Dott. Ing. Francesco Martellotta**

# 1

## La conversione analogico-digitale

### 1.1 Introduzione

Un *segnale* è una funzione o una grandezza che contiene informazione, relativamente allo stato o al comportamento di un sistema fisico. Da un punto di vista matematico un segnale è rappresentato da una funzione di una o più variabili indipendenti. Nel caso in esame il *segnale audio* è la variazione di pressione (ovvero la pressione acustica) prodotta da una data sorgente e misurata in un dato punto in funzione del *tempo*.

Un segnale reale è definito in un insieme temporale *continuo* (rappresentato cioè mediante una funzione continua), così come il suo valore istantaneo può variare in un insieme continuo (cioè per incrementi infinitamente piccoli), pertanto lo si definisce *tempo-continuo* e ad *ampiezza continua*, o, più sinteticamente, *segnale analogico*. Tale denominazione deriva dal fatto che spesso i segnali che si elaborano sono relativi a grandezze fisiche in relazione fra loro, che assumono variazioni *analoghe* nei confronti della variabile indipendente.

~~~~~  
Ad esempio il segnale elettrico prodotto da un microfono riproduce  
in termini di variazione di una grandezza elettrica le variazioni della  
pressione acustica agente sul diaframma.  
~~~~~

L'analisi computerizzata di un segnale audio analogico richiede che questo venga *discretizzato*, cioè che venga rappresentato da un insieme *discreto* di valori numerici. In particolare, il segnale analogico deve essere sottoposto a due operazioni di discretizzazione: la prima legata alla rappresentazione della variabile indipendente (il tempo); la seconda legata alla rappresentazione dei valori istantanei assunti dal segnale. Si passa così da un segnale tempo-continuo ad ampiezza continua, ad un segnale *tempo-discreto* ad *ampiezza discreta*, o, più sinteticamente, *segnale numerico* (o *digitale*). La conversione viene effettuata tramite un *converti-*

*tore analogico-digitale (A/D)* in cui vengono eseguite le due operazioni di discretizzazione, denominate rispettivamente *campionamento* e *quantizzazione*.

Il convertitore A/D viene inserito in un circuito schematicizzato in Figura 1. La pressione acustica viene convertita in una debole tensione elettrica dal microfono, che svolge la funzione di trasduttore (cioè di dispositivo che cambia la natura del segnale senza alterarne la forma). La tensione prodotta dal microfono, prima di poter essere elaborata ulteriormente, deve essere amplificata dall'amplificatore. Sia il microfono, sia l'amplificatore sono componenti analogici, infatti l'andamento della tensione prima e dopo l'amplificazione è analogo alla pressione acustica in ingresso a meno di una costante moltiplicativa. A questo punto entra in azione il convertitore A/D.

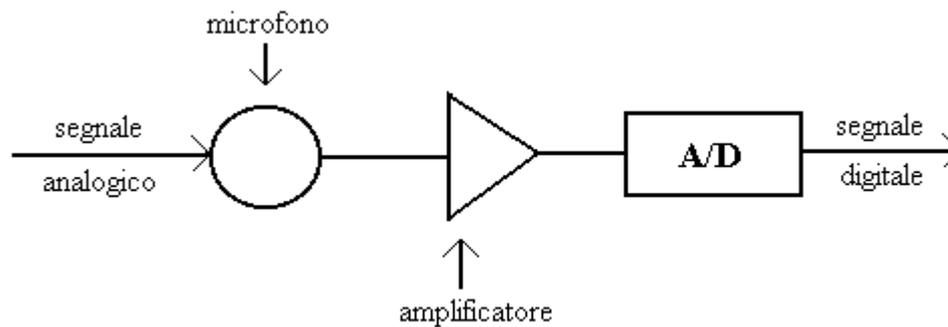


Figura 1 – Rappresentazione schematica di un circuito per la conversione di un segnale analogico in un segnale digitale

## 1.2 Il campionamento

Il *campionamento* consente di trasformare un segnale *tempo-continuo* in un segnale *tempo-discreto*, e consiste nel discretizzare una funzione continua  $x(t)$ , rappresentandola come una *sequenza* di numeri  $x[n]$ , dove  $n$  è una variabile intera compresa nel generico intervallo  $[N_1, N_2]$ .

Concretamente il campionamento del segnale tempo-continuo  $x(t)$  si realizza “estraendo” dal segnale stesso i valori che esso assume a istanti equispaziati, multipli di un intervallo  $T_C$ , detto *periodo di campionamento*. Pertanto il valore  $n$ -esimo della sequenza sarà pari al valore della funzione continua all'istante  $nT_C$ :

$$x[n] = x(nT_C). \quad (1)$$

Direttamente legata al periodo di campionamento è la *frequenza di campionamento*, definita come:

$$f_c = \frac{1}{T_C} \text{ (Hz)}, \quad (2)$$

che misura la frequenza con cui il campionatore estrae i valori della sequenza, ovvero il numero di campioni che vengono estratti ogni secondo.

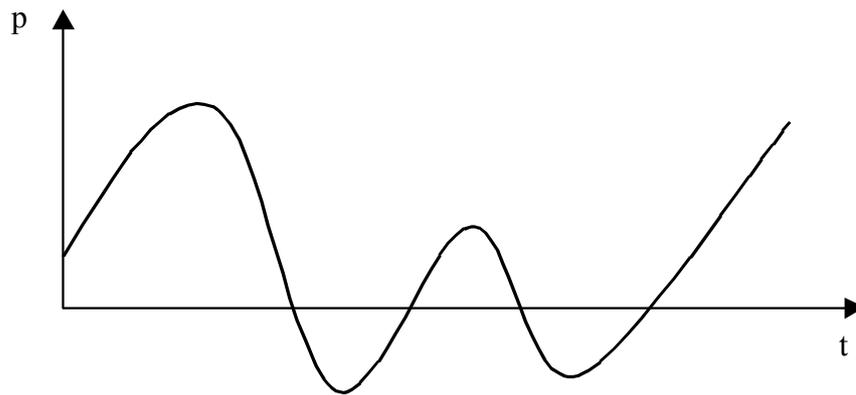


Figura 2 – Rappresentazione di un segnale nel piano analogico

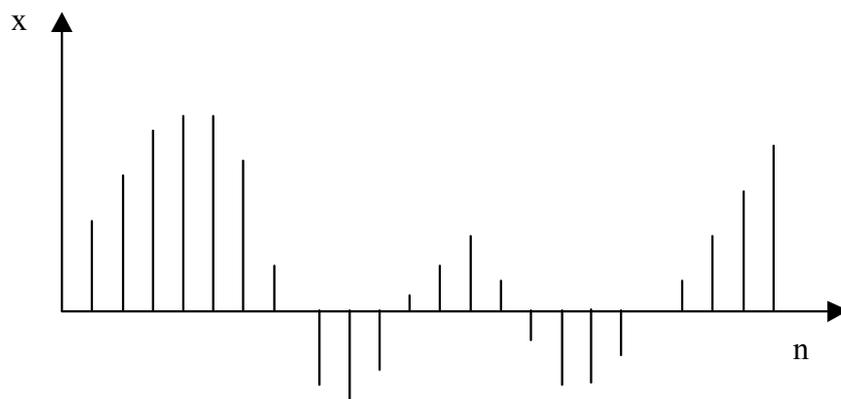


Figura 3 – Rappresentazione nel piano digitale dello stesso segnale di Figura 2

La Figura 2 mostra un generico segnale analogico rappresentativo della variazione della pressione acustica in funzione del tempo  $t$ . La Figura 3 rappresenta il corrispondente segnale *tempo-discreto* ottenuto dopo il campionamento. I valori della sequenza distano fra loro un tempo  $T$  pari al periodo di campionamento e la funzione  $x[n]$  risulta definita solo per valori interi di  $n$ , mentre *non è definita per valori non interi di  $n$* .

Intuitivamente, è evidente che la scelta della frequenza di campionamento di un segnale condiziona in misura significativa l'accuratezza con cui questo viene convertito e, di conseguenza, la possibilità di “ricostruire” (e riprodurre) il segnale analogico. La frequenza di campionamento risulta legata alla componente spettrale di più alta frequenza presente nel segnale dal *teorema del campionamento*.

### 1.2.1 Il teorema del campionamento di Nyquist-Shannon

Il *teorema del campionamento* stabilisce che un segnale tempo-continuo la cui componente spettrale di frequenza massima è  $f_{\max}$  può essere ricostruito senza errori a partire dal corrispondente segnale tempo-discreto solo se la frequenza di campionamento  $f_c$  è maggiore o al più uguale a  $2f_{\max}$ . Se questa condizione è soddisfatta allora è possibile ricostruire il segnale  $x(t)$  a partire dalla sequenza  $x[n]$  mediante la *formula di interpolazione di Nyquist-Shannon*:

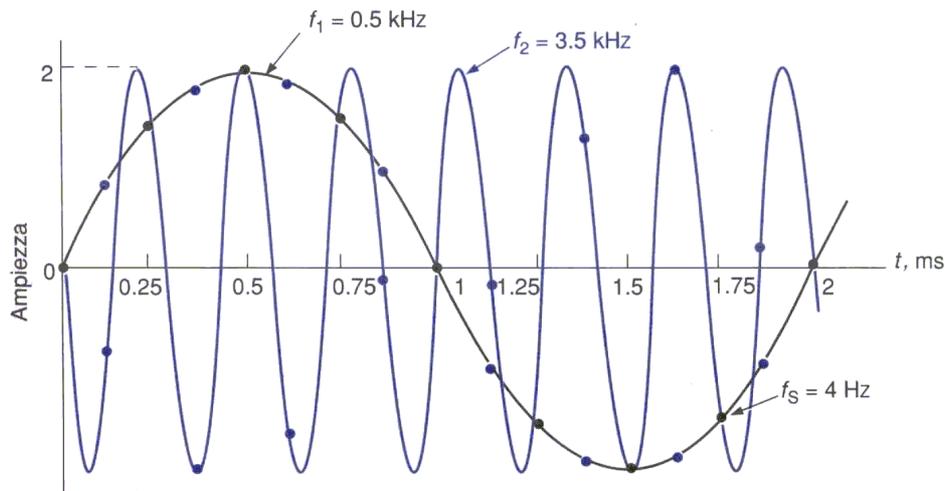


Figura 4 – Esempio di aliasing

$$x(t) = \sum_{k=-\infty}^{+\infty} x[k] \frac{\sin\left[\frac{\pi}{T}(t - kT)\right]}{\frac{\pi}{T}(t - kT)} \quad (3)$$

Come conseguenza del teorema del campionamento, se si campiona un suono ad una frequenza  $f_c$ , la componente spettrale di massima frequenza del suono non dovrà essere superiore alla frequenza  $f_c/2$ , di solito denominata *frequenza di Nyquist* ed indicata con  $f_N$ . Se nel suono sono presenti componenti spettrali di frequenza maggiore di  $f_N$  si dice che il suono viene *sottocampionato* e si verifica un fenomeno denominato *aliasing*. L'aliasing consiste nella comparsa di componenti spettrali spurie con frequenza  $f_{alias}$  dipendente dalla frequenza di Nyquist e dalla frequenza effettiva del suono  $f_{eff}$ , secondo la seguente relazione:

$$f_{alias} = f_N - (f_{eff} - f_N) = 2f_N - f_{eff} = f_c - f_{eff} \quad (4)$$

La Figura 4 mostra un esempio di aliasing conseguente ad un segnale costituito da un tono puro a 3.5 kHz che viene campionato a 4 kHz (cioè ogni 0.25 s). Si può osservare che andando ad interpolare fra i punti campionati (indicati in nero in figura) si ottiene un tono con una frequenza di 0.5 kHz, in accordo con la (4).

Per evitare problemi di aliasing ci sono, quindi, due possibilità: aumentare la frequenza di campionamento fino a soddisfare la condizione  $f_c > 2f_{max}$ , oppure limitare la larghezza di banda  $B$  del segnale da campionare in modo che sia soddisfatta la condizione  $B = f_N$ . Poiché la composizione spettrale di un suono non è, in genere, nota a priori, la prima strada non è tecnicamente praticabile. La seconda, invece, consente di adottare una frequenza di campionamento fissa limitando la larghezza di banda del segnale mediante l'interposizione di un *filtro anti-aliasing* fra l'amplificatore ed il convertitore A/D dello schema di Figura 1.

Un filtro *anti-aliasing* è un filtro che lascia passare tutte le frequenze minori della frequenza di Nyquist mentre impedisce il passaggio di quelle maggiori che sarebbero causa di aliasing. Tecnicamente un filtro di questo tipo viene denomina-

to *filtro passa-basso* con frequenza di taglio pari alla  $f_N$ . In un filtro passa-basso *ideale* al di sotto della frequenza di taglio il segnale resta immutato (banda passante), mentre al di sopra della frequenza di taglio il segnale viene completamente azzerato (banda oscura). In un filtro *reale* (v. Figura 5) fra le due bande vi è una *banda di transizione* in corrispondenza della quale si verifica la caduta e, inoltre, sia nella banda passante sia nella banda oscura la risposta del filtro non è perfettamente piatta ma presenta delle fluttuazioni.

*La frequenza di campionamento di 44.1 kHz, derivata dalla tecnica di registrazione dei CD musicali e oggi comunemente adottata dalle più comuni schede audio per PC, si basa sul fatto che la sensibilità dell'orecchio umano arriva fino a 20 kHz e quindi le frequenze superiori, non essendo percepite, non sono di interesse ai fini della riproduzione. Questo significa che si potrebbe campionare a 40 kHz avendo però cura di filtrare il segnale in ingresso in quanto eventuali componenti aventi frequenza maggiore di 20 kHz pur non percepite dall'orecchio darebbero luogo a frequenze spurie nel campo udibile per effetto dell'aliasing. Pertanto è di fondamentale importanza impiegare sempre un filtro anti-aliasing. L'utilizzo della frequenza di campionamento di 44.1 kHz consente poi di limitare le contaminazioni dovute alla banda di transizione del filtro<sup>1</sup>.*

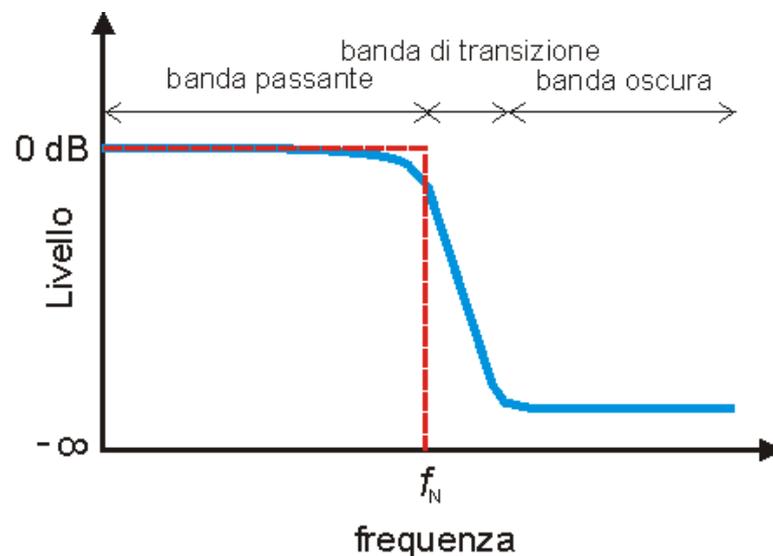


Figura 5 – Schematizzazione della risposta di un filtro passa-basso con frequenza di taglio  $f_N$ . (---) Filtro ideale; (—) Filtro reale

<sup>1</sup> La frequenza di campionamento di 44.1 kHz trae la sua origine dalla registrazione digitale dei segnali audio utilizzando un videoregistratore a nastro. Nel sistema TV europeo sono disponibili 25 quadri al secondo, ciascuno di 625 linee, e ciascuna linea può registrare tre parole digitali per canale, poiché una parte delle linee è utilizzata dai codici per la correzione degli errori restano 588 linee per quadro e quindi la frequenza di campionamento risultante è data da  $f_C = 3 \cdot 588 \cdot 25 = 44100$  Hz.

## 1.3 La quantizzazione

La seconda operazione di discretizzazione necessaria per la conversione analogico-digitale è la *quantizzazione*, in cui il valore istantaneo che il segnale assume –in genere un numero reale con infinite cifre decimali– viene approssimato con un numero finito di cifre decimali dipendenti dal numero di bit disponibili per allocare il valore. Il fatto di utilizzare registri di lunghezza finita, ovvero un numero finito di cifre sulle quali effettuare i calcoli, può essere causa di errori di *troncamento* (se le cifre meno significative vengano scartate), o di *arrotondamento* (se l'ultima cifra significativa viene arrotondata in base al valore delle cifre scartate). Questi errori possono essere considerati in termini di non linearità del sistema o, dato che questo tipo di analisi può risultare piuttosto complessa, più semplicemente come un “rumore” che si sovrappone al segnale.

Da quanto detto, si evince che quanto maggiore è il numero  $b$  di bit utilizzati dal quantizzatore, tanto più piccolo sarà l'errore commesso e, di conseguenza, il “rumore” presente nel segnale digitale.

Se  $b$  è il numero di bit a disposizione, il numero di possibili livelli di quantizzazione è pari a  $2^b$ , allocabili in innumerevoli modi, ciascuno dei quali corrisponde a schemi di quantizzazione caratterizzati da vantaggi e svantaggi. Le due principali classi di schemi di quantizzazione sono quella *uniforme* (o *lineare*) e quella *non uniforme* (o *non lineare*), a seconda che i passi di quantizzazione siano di ampiezza uniforme o non uniforme. La scelta dell'uno o dell'altro tipo dipende da numerosi fattori, quali la conoscenza della distribuzione delle ampiezze del segnale, la possibilità di ridurre il numero di bit, la velocità e la complessità dei processi di elaborazione successiva, la precisione di calcolo, la dispersione in frequenza dell'errore di quantizzazione. Un'altra distinzione, fissata la lunghezza  $b$  del registro utilizzato in fase di conversione, riguarda la rappresentazione numerica che può essere a *virgola fissa* (i campioni vengono cioè convertiti in numeri interi), o a *virgola mobile*.

### 1.3.1 La quantizzazione uniforme

Nel campo audio per i sistemi di registrazione e riproduzione, sia professionali sia basati su PC, il sistema utilizzato è quello a *quantizzazione uniforme con aritmetica a virgola fissa a 16 bit*. Questo sistema di codifica viene comunemente denominato PCM (“Pulse-code modulation”) (cfr. Rif. [4]).

Con il sistema PCM, una volta fissato il passo di quantizzazione  $\Delta$ , la massima ampiezza che può essere convertita, nel caso di un segnale temporale a media nulla, è pari a  $\Delta \cdot 2^{b-1}$ . Quindi, assumendo  $\Delta$  pari a uno e tenendo conto dello zero, l'intervallo di dei possibili valori che può assumere il segnale quantizzato varia da  $-32767$  a  $+32767$ .

L'errore di quantizzazione è compreso nel campo  $\pm\Delta/2$  e può essere considerato come un segnale reale con uno spettro non limitato in banda e di densità di potenza spettrale uniforme (quindi assimilabile a rumore bianco). Questo significa

che il numero di bit utilizzati dal quantizzatore condiziona il rapporto segnale rumore (S/R) del segnale digitale. Infatti è possibile dimostrare (cfr. Rif. [2], pp. 206-209) che il rapporto S/R è pari a:

$$6b - 1,25 \text{ dB.} \quad (5)$$

In base a tale relazione il rapporto S/R ottenibile impiegando il sistema PCM è pari a 94,5 dB. Tuttavia, qualora il segnale in ingresso è molto basso il quantizzatore si comporta come un dispositivo non lineare che “tosa” il segnale, generando distorsione quando il segnale viene riconvertito in forma analogica. Per minimizzare questo inconveniente vi sono diverse soluzioni, fra cui l’aumento del numero di bit, il *sovracampionamento* (che aumenta il rapporto S/R a parità di bit usati), o l’uso di una quantizzazione non uniforme (in modo da aumentare il grado di dettaglio per piccole ampiezze, v. par. successivo). Un altro sistema alternativo, che conduce ad un pratico miglioramento della qualità del rumore percepito in riproduzione, a scapito di un modesto incremento del rumore di fondo, consiste nel sommare al segnale di ingresso del quantizzatore un rumore (“dither”) con un livello RMS pari a circa  $\Delta/3$ . In questo modo le distorsioni appaiono coperte dal rumore bianco aggiunto che risulta più gradevole all’ascolto, mentre il peggioramento nel rapporto S/R risulta comunque accettabile.

### 1.3.2 La quantizzazione non uniforme\*

La quantizzazione non uniforme viene solitamente utilizzata nelle applicazioni telefoniche dove il segnale trasmesso (il parlato) è caratterizzato da una distribuzione delle ampiezze tipicamente simile a quella di Figura 6, in cui le ampiezze estreme hanno una bassissima frequenza al contrario di quelle intermedie. In un sistema a quantizzazione uniforme viene assegnato lo stesso numero di livelli di quantizzazione sia alle piccole sia alle grandi ampiezze, ma queste ultime sono utilizzate solo raramente, per cui il sistema non è efficiente in quanto alcuni bit risultano “sprecati”.

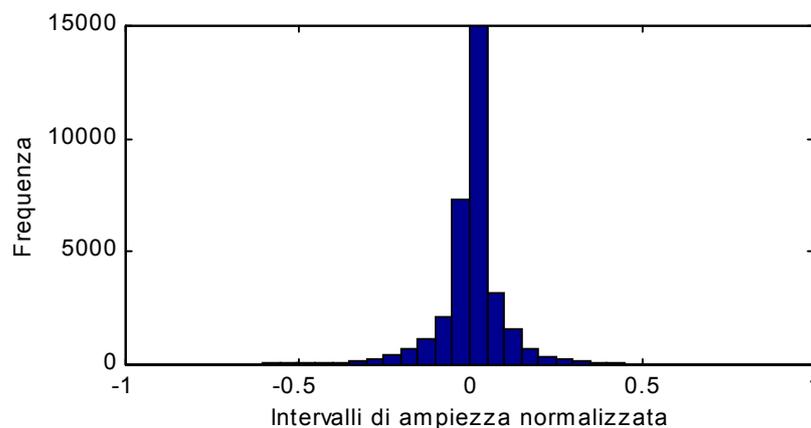


Figura 6 – Distribuzione delle ampiezze normalizzate (da -1 a +1) per un segnale vocale

\* Questo paragrafo ha carattere “informativo”, non costituisce argomento di esame ma costituisce un complemento per avere una visione più ampia della problematica. Ulteriori informazioni possono essere trovate nei Riff. [5],[6].

Poiché nelle trasmissioni telefoniche l'interesse principale è quello di minimizzare il flusso di dati, per rendere più efficiente la quantizzazione e ridurre il numero di bit necessari si utilizza la tecnica del *companding* (*compression-expanding*) su cui si basano le codifiche *A-law* e  *$\mu$ -law*. Questi sistemi di quantizzazione espandono i piccoli valori e comprimono quelli grandi, cioè utilizzano un maggior numero di livelli di quantizzazione per le ampiezze di segnale più basse (con passi di quantizzazione più piccoli), ed un minor numero di intervalli di quantizzazione per le ampiezze più elevate (con passi di quantizzazione più grandi). La Figura 7 mostra il rapporto fra l'ampiezza (normalizzata) del segnale in ingresso (rappresentata sull'asse delle ascisse) e gli intervalli dei livelli di quantizzazione (normalizzati) assegnati in uscita (sull'asse delle ordinate). Si può osservare, ad esempio, che ai segnali di ampiezza compresa fra  $-0.2$  e  $+0.2$  è assegnato il 60% dei livelli di quantizzazione disponibili.

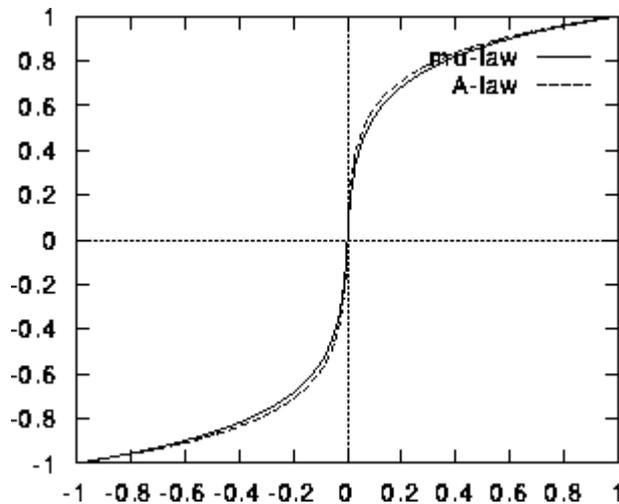


Figura 7 – Relazione fra ampiezza (normalizzata) del segnale in ingresso (asse x) e livelli di quantizzazione (normalizzati) corrispondenti.

# 2

## L'analisi in frequenza dei segnali digitali

### 2.1 Il teorema di Fourier per segnali tempo-continui

Il *Teorema di Fourier* stabilisce che un qualsiasi suono periodico è sempre rappresentabile come sovrapposizione di un opportuno numero  $m$  (eventualmente infinito) di onde sinusoidali di data ampiezza  $A_i$ , frequenza  $f_i$  e fase  $\varphi_i$ :

$$x(t) = \sum_{i=1}^{m \rightarrow \infty} A_i \sin(\omega_i t + \varphi_i), \text{ con } \omega_i = 2\pi f_i. \quad (6)$$

Il Teorema di Fourier è lo strumento matematico che ci consente di passare dal dominio del tempo a quello delle frequenze: infatti, la distribuzione delle ampiezze  $A_i$  associate alle rispettive frequenze che compongono il segnale fornisce la composizione spettrale dei suoni complessi.

Nel caso in cui un suono è *aperiodico* il segnale può ancora essere scomposto in una *serie di Fourier*, ma nella forma integrale:

$$x(t) = \int_{-\infty}^{+\infty} X(s) e^{i2\pi s t} ds, \quad (7)$$

dove i termini  $X(s)$  si ottengono applicando la *trasformata di Fourier* al segnale  $x(t)$ :

$$X(s) = \int_{-\infty}^{+\infty} x(t) e^{-i2\pi s t} dt, \quad (8)$$

e forniscono la *densità di potenza spettrale* relativa alla frequenza  $s$  in un intorno  $ds$  della stessa. Lo spettro di un suono aperiodico è pertanto *continuo*.

Poiché l'applicazione della trasformata di Fourier alla Eq. (8) produce come risultato la Eq. (7), si parla di *trasformata inversa* o *antitrasformata*, e ne consegue che una volta noto il segnale è possibile, utilizzando il medesimo strumento matematico, passare dal dominio del tempo a quello delle frequenze e viceversa.

## 2.2 La trasformata di Fourier discreta (DFT)

Per i sistemi tempo-discreti si è visto che il segnale  $x(t)$  è rappresentato mediante una sequenza  $x[n]$ . Se la sequenza è periodica, ovvero se  $x[n] = x[n + kn]$ , con  $k$  intero, allora è possibile rappresentarla come somma di sequenze esponenziali complesse con frequenze multiple intere della frequenza fondamentale della sequenza periodica  $2\pi/N$ ,  $N$  essendo il numero dei campioni che contenuti nel singolo periodo della sequenza. Si scompone così  $x[n]$  nella *serie discreta di Fourier* (DFS):

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{i\frac{2\pi}{N}nk}, \quad \text{con } \frac{2\pi k}{N} = \omega_k \quad (9)$$

Poiché l'esponenziale complesso  $e^{i2\pi nk/N}$  è periodico in  $k$  con periodo  $N$ , esistono solo  $N$  esponenziali complessi distinti il cui periodo è un sottomultiplo intero del periodo fondamentale. Gli  $N$  termini  $X(k)$  sono dati dalla *trasformata discreta di Fourier* (DFT):

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-i\frac{2\pi}{N}nk} \quad (10)$$

La rappresentazione in serie discreta di Fourier, valida per sequenze periodiche, può essere opportunamente applicata anche a sequenze di durata finita, trasformando la sequenza finita di lunghezza  $N$  in una sequenza periodica di periodo  $N$  il cui andamento nel periodo sia identico alla sequenza di durata finita. La DFS e la DFT risultanti dovranno però essere definite solamente nei rispettivi campi di esistenza:

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-i\frac{2\pi}{N}nk}, \quad \text{per } 0 \leq k \leq N-1 \quad (11)$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{i\frac{2\pi}{N}nk}, \quad \text{per } 0 \leq n \leq N-1 \quad (12)$$

mentre  $X(k) = 0$  al di fuori dell'intervallo  $0 \leq k \leq N-1$ , e  $x[n] = 0$  al di fuori dell'intervallo  $0 \leq n \leq N-1$ .

Pertanto, impiegando la DFT è possibile ottenere lo spettro della sequenza, ovvero la densità di potenza spettrale, definita su  $N$  righe spettrali equispaziate che individuano tutte le frequenze fino alla frequenza di campionamento  $f_C$ . Tuttavia, in base al teorema del campionamento, si è visto che sono significative solo le frequenze delimitate superiormente dalla frequenza di Nyquist e quindi, essendo  $f_N = f_C/2$ , devono essere prese in considerazione solo le prime  $N/2$  righe. La più piccola frequenza analizzata sarà data da  $f_C/N$ , valore che corrisponde anche al *passo* delle  $N$  righe spettrali, inversamente proporzionale alla *risoluzione in frequenza* dello spettro. Questo tipo di rappresentazione per passi di ampiezza costante attribuisce la stessa importanza tanto alle basse che alle alte frequenze, in contrasto con il comportamento del nostro orecchio che ha una risoluzione più alta

alle basse frequenze (dove le bande critiche hanno una larghezza di 100 Hz) e progressivamente più bassa alle alte frequenze (dove le bande critiche tendono a divenire sempre più ampie, con la conseguenza che il nostro orecchio risulta meno sensibile ai “dettagli” dello spettro).

### 2.2.1 La trasformata di Fourier veloce

La fortuna dell'analisi spettrale dei segnali digitali è in larga misura legata alla definizione, avvenuta nel 1965 ad opera di Tukey e Cooley, di un algoritmo per eseguire la trasformata di Fourier in modo altamente efficiente. Questo algoritmo, noto sinteticamente come FFT (*fast Fourier transform*) consente di calcolare i termini  $X(k)$  impiegando solo  $N \log_2 N$  operazioni, contro le  $N^2$  richieste dall'algoritmo normale. La condizione affinché l'algoritmo FFT possa essere applicato è che  $N$  deve essere una potenza di 2.

L'idea di base è quella di utilizzare un approccio ricorsivo: calcolare prima la trasformata dei valori della sequenza con indice pari ( $x[0], x[2], \dots, x[N-2]$ ), poi quelli con indice dispari ( $x[1], x[3], \dots, x[N-1]$ ), e infine combinare i risultati dei due per ottenere la trasformata di Fourier dell'intera sequenza. Una trattazione più approfondita del funzionamento dell'algoritmo può essere trovata in ([7],[8],[9]).

## 2.3 Impiego pratico della DFT

L'impiego pratico della trasformata discreta di Fourier per l'analisi spettrale dei segnali digitali richiede, in alcuni casi, l'impiego di accorgimenti indispensabili per ottenere risultati corretti.

Un tipico caso è quello di un segnale in cui *il primo e l'ultimo campione sono diversi*. Poiché l'applicazione della DFT richiede che il segnale venga reso periodico, si otterrebbe un andamento simile a quello riportato in Figura 8, in cui compare una discontinuità nel segnale. Dal punto di vista dell'analisi spettrale la discontinuità si traduce in un “gradino” (v. Figura 9) che abbraccia tutte le frequenze oltre a quelle originarie.

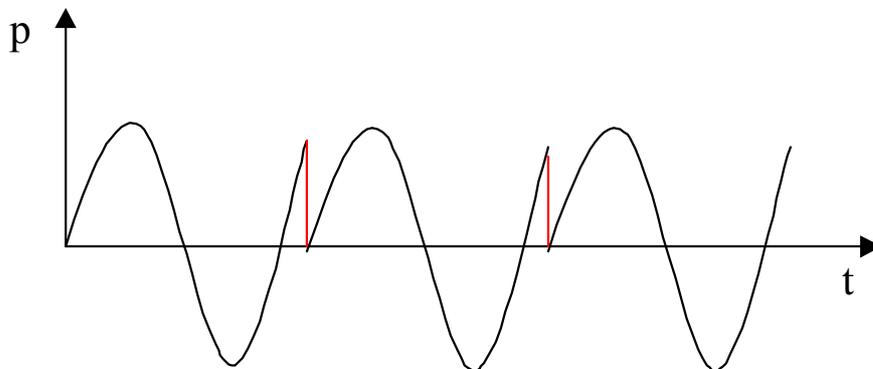


Figura 8 – Rappresentazione nel piano dei tempi di un segnale (costituito da una porzione non intera di una sinusoide) resa periodica per poter calcolare la DFT.

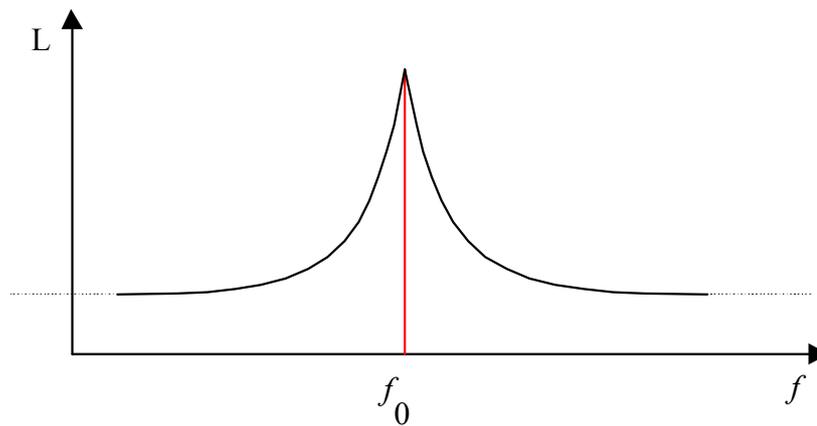


Figura 9 – Spettro relativo al segnale riprodotto in Figura 8. Si osserva che la densità di potenza spettrale assume valori non su tutte le frequenze dello spettro in aggiunta alla frequenza  $f_0$  della sinusoide originaria

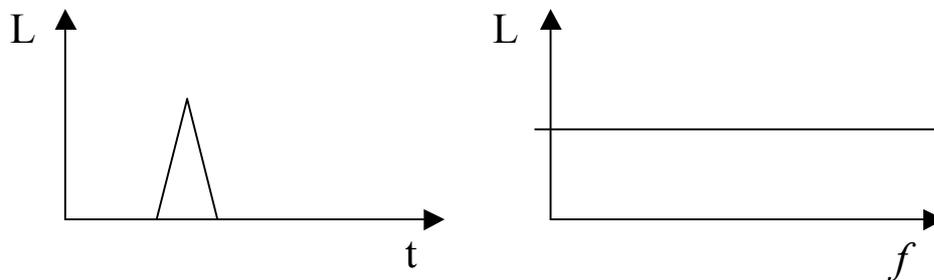


Figura 10 – Segnale impulsivo e spettro relativo

Il fenomeno dell'estensione dello spettro a tutte le frequenze è noto col nome di *leaking* (perdita), ed è dovuto al fatto che la discontinuità presente nel segnale ha uno spettro di tipo *bianco*. Questa è una proprietà di tutti i segnali che hanno un carattere *impulsivo*<sup>1</sup>, cioè segnali brevi ed intensi: quanto più breve è la loro durata, tanto maggiore è l'estensione del loro spettro. Per risolvere il problema del leaking è necessario che il segnale assuma alle estremità i medesimi valori e, preferibilmente, che si annulli. In questa maniera si elimina ogni discontinuità nel segnale reso periodico.

Per azzerare il segnale agli estremi la soluzione più semplice è quella di “finestrare” il segnale nel dominio del tempo, cioè moltiplicarlo per una funzione di pesatura che assuma valore nullo agli estremi. Esistono diverse di queste funzioni ciascuna caratterizzata da una sua forma (alcune sono raffigurate in Figura 11), ma la più utilizzata è la finestra di *Hanning* che è una funzione coseno traslata lungo l'asse delle ordinate e normalizzata ad uno.

Tuttavia, come si può facilmente intuire, l'uso della finestatura risolve un problema, quello del leaking, ma ne crea un altro, dal momento che il segnale finestrato *non è equivalente* a quello originario: i campioni più interni acquistano infatti un peso maggiore rispetto a quelli alle estremità. Per evitare anche questo problema, la procedura che si segue consiste nel suddividere la sequenza in “fette”

<sup>1</sup> Si definisce *impulso unitario* un segnale in cui  $x[0]=1$  e  $x[n]=0$  per ogni  $n > 0$ .

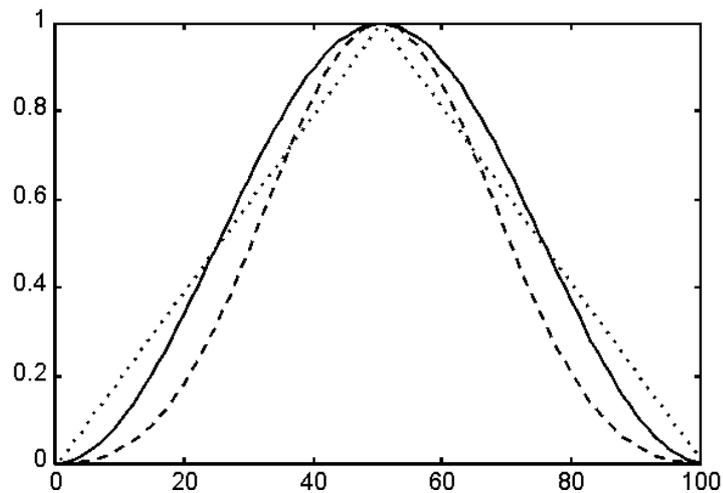


Figura 11 – Funzioni di pesatura: (—) Hanning; (- - -) Triangolare; (- · -) Blackman.

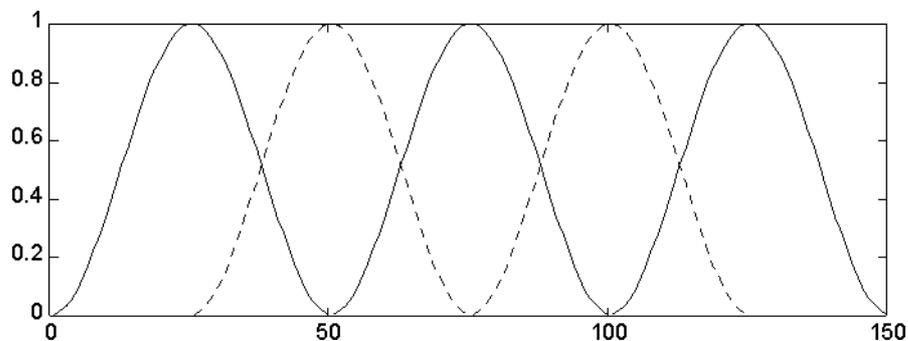


Figura 12 – Esempio di traslazione della finestra di Hanning con sovrapposizione del 50%

di ampiezza  $M$ , e, dopo aver applicato la finestrazione alla prima fetta  $a$  e aver calcolato lo spettro relativo, per tenere nel debito conto i campioni alle estremità si trasla la finestra in modo che sia a cavallo fra la prima e la seconda fetta: in questo modo si realizza una sovrapposizione che, per garantire l'equivalenza energetica fra segnale originario e quello pesato, deve essere (se si utilizza la Hanning<sup>1</sup>) del 50%. Ad ogni passo successivo si deve quindi traslare (o *shiftare*) la finestra di  $M/2$  campioni (v. Figura 12) e calcolare lo spettro risultante, fino a che si arriva alla fine della sequenza o, per essere precisi, al punto in cui la sequenza si ripete: questo per garantire che anche i campioni iniziali della sequenza abbiano lo stesso peso degli altri. Lo spettro finale si ottiene facendo la media di tutti gli spettri calcolati traslando la finestra prescelta.

Questa procedura per il calcolo dello spettro permette anche di svincolare la risoluzione in frequenza dello spettro dal numero  $N$  dei campioni che compongono il segnale. Se si desidera una spaziatura in frequenza di  $\delta F$  e se la frequenza di campionamento è  $f_C$ , allora il numero di righe spettrali da calcolare sarà pari a:

<sup>1</sup> L'utilizzo preferenziale della finestra di Hanning è proprio legato al fatto che essa è l'unica fra le varie finestrate disponibili che garantisce, con una sovrapposizione del 50%, che tutti i campioni della sequenza contribuiscano in eguale misura allo spettro: in Figura 12 sommando le funzioni di pesatura di una finestra e di quella traslata si ottiene sempre un valore unitario.

$$R = \frac{f_c}{\delta F}, \quad (13)$$

di cui, per il teorema del campionamento, saranno significative solo le prime  $R/2$ .

A questo punto si avranno delle fette costituite da  $M$  campioni, sulle quali andare a calcolare uno spettro ad  $R$  punti. Per avere una adeguata risoluzione spettrale si avrà in genere  $R > M$ , per cui, dovendo eseguire la DFT sullo stesso numero di punti, si adotterà per entrambe il valore più alto. Inoltre, per poter applicare l'algoritmo FFT è anche conveniente che  $R$  sia una potenza di due. Le fette saranno quindi composte da  $R$  punti di cui i primi  $M$  saranno quelli effettivamente prelevati dalla sequenza e finestrati, mentre i restanti  $R - M$  saranno riempiti di zeri (*zero padding*).

## 2.4 Rappresentazioni grafiche dello spettro sonoro

Una volta calcolata la trasformata di Fourier per il segnale in esame è possibile rappresentare i risultati in diversi modi. La rappresentazione più semplice è quella che corrisponde alla definizione di *spettro* (v. Figura 13, sinistra) e riporta le frequenze in ascissa e la densità di potenza spettrale in ordinata. In genere tanto le frequenze che la densità spettrale vengono rappresentate su scala logaritmica per tenere conto della sensibilità dell'orecchio (lo spettro viene comunque calcolato in base alla spaziatura in frequenza fissata).

Il secondo tipo di rappresentazione è il *sonogramma* (o *spetrogramma*) (v. Figura 13, destra), che si basa sul calcolo dello spettro effettuato per "fette", così come illustrato nella sezione precedente. In questo caso anziché mediare i risultati dei vari calcoli, questi vengono associati ciascuno al relativo intervallo temporale a cui fanno riferimento, in modo da fornire una rappresentazione di come si modifica lo spettro nel tempo. In questo caso in ascissa si riportano i tempi, in ordinata le frequenze e la densità di potenza spettrale viene rappresentata mediante una opportuna scala cromatica.

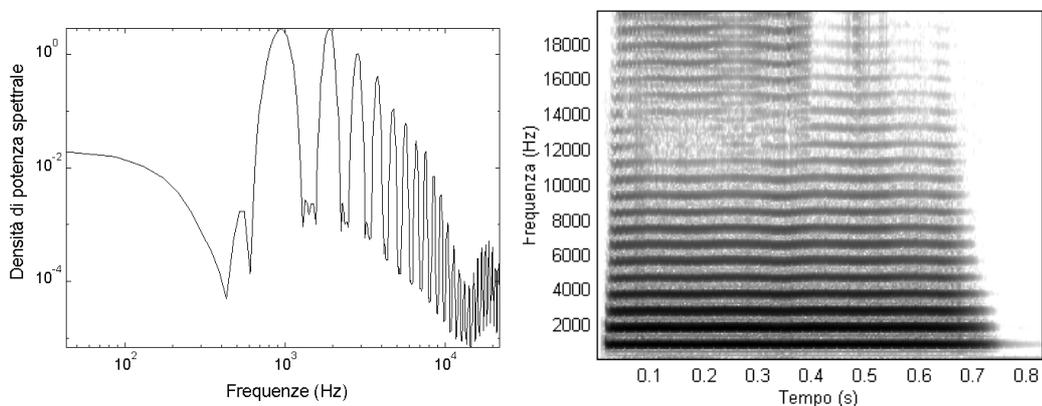


Figura 13 – Spettro relativo ad una nota ( $La_4$ ) emessa da una tromba. Sinistra: rappresentazione nel piano delle frequenze; destra: sonogramma (toni più scuri corrispondono a densità di potenza spettrale più elevate).

## Bibliografia

- [1] B. Gold, N. Morgan, *Speech and audio signal processing*, Wiley & Sons, New York, 2000.
- [2] S. Santoboni, *Elettroacustica*, Masson, Milano, 1996.

## Webliografia

- [3] Teorema del campionamento su Wikipedia:  
[www.wikipedia.org/wiki/Nyquist-Shannon\\_sampling\\_theorem](http://www.wikipedia.org/wiki/Nyquist-Shannon_sampling_theorem)
- [4] Pulse Code Modulation:  
[www.rasip.fer.hr/research/compress/algorithms/fund/pcm](http://www.rasip.fer.hr/research/compress/algorithms/fund/pcm)
- [5] Quantizzazione non lineare:  
[www.ee.ucla.edu/~dsplab/sqc/over.html](http://www.ee.ucla.edu/~dsplab/sqc/over.html)
- [6]  $\mu$ -law companding:  
[www-ccrma.stanford.edu/~jos/mdft/Mu\\_Law\\_Companding.html](http://www-ccrma.stanford.edu/~jos/mdft/Mu_Law_Companding.html)
- [7] L'algoritmo FFT su Wikipedia:  
[www.wikipedia.org/wiki/Fast\\_Fourier\\_Transform](http://www.wikipedia.org/wiki/Fast_Fourier_Transform)
- [8] L'algoritmo FFT (storia e teoria) su Mathworld:  
<http://mathworld.wolfram.com/FastFourierTransform.html>
- [9] L'algoritmo FFT passo per passo:  
[www.cs.cf.ac.uk/Dave/Vision\\_lecture/node20.html](http://www.cs.cf.ac.uk/Dave/Vision_lecture/node20.html)