

Trattamento e Codifica di Dati Multimediali

Ilaria Battiston

Anno scolastico 2018-2019

Contents

1	Multimedia	3
2	Multimedia processing	4
3	Segnali	5
4	Sequenze	9
5	Analisi di Fourier	9
6	Campionamento	18
7	Convoluzione	19
8	Quantizzazione	20
9	Sistemi lineari	23
10	Equazioni alle differenze	29
11	Risposta all'impulso	31
12	Trasformata zeta	33
13	Analisi dei sistemi	37
14	Sistemi LTI e trasformata zeta	38
15	Filtri con poli e zeri	41
16	Immagini	45
17	Audio	49
18	Video	53
19	DCT wavelets	56
20	Tecniche di compressione	60
21	Valutazione della qualità di compressione	66
22	Compressione JPEG	68
23	Compressione video	71
24	Formati immagine	75
25	Compressione audio	76

1 Multimedia

Multimedia: utilizzo di diversi mezzi che concorrono insieme, solitamente in maniera interattiva, per trasferire informazione. I media hanno caratteristiche e standard differenti per essere catturati, immagazzinati, manipolati e trasmessi.

I segnali devono essere trattati tenendo conto del **sistema percettivo**. La tecnologia multimediale cerca di simulare il sistema percettivo umano, minimizzando il quantitativo di dati da processare per ottenere un'informazione.

Il trattamento dei media si basa sull'analisi multimodale: essa permette di trarre conclusioni su contenuti di diversi tipi e di individuare stimoli legati a segnali. Per questi scopi devono essere introdotte misure oggettive e soggettive.

Un altro aspetto del multimodale è l'analisi da segnali che provengono da diverse fonti, come i dati fisiologici (ottenuti tramite sensori).

1.1 Segnali

I segnali possono essere classificati in base a **dominio** e **codominio**.

Dominio:

- $D = R$, segnale a tempo (spazio) continuo $x(t)$, $t \in R$ (possono assumere tutti i possibili valori reali);
- $D = K$, segnale a tempo discreto, $x(t)$, $t \in K$ con K numerabile, $K \in \{\dots, t_{-1}, t_0, t_1, \dots\}$. Il numero di valori t può essere comunque infinito.

Codominio:

- $C = R$, segnale continuo nelle ampiezze;
- $C = K$, segnale discreto nelle ampiezze con K numerabile e tipicamente finito $\{x_1, x_2, \dots, x_n\}$.

La variabile dipendente è definita sul dominio, quella indipendente sul codominio. I segnali possono essere reali o complessi (parte reale e immaginaria oppure modulo e fase, strettamente in relazione tra loro).

Le ampiezze tipicamente sono un insieme finito di valori, quindi al crescere di t (variabile indipendente, infiniti valori) il **range** di $s(t)$ (variabile dipendente) sarà limitato, anche se composto da infiniti valori. Il **passo** (distanza tra un campione e il successivo) dev'essere costante.

La quantizzazione è uno step intermedio della trasformazione tra segnale analogico e digitale, cioè la divisione del tempo (spazio) in passi e l'approssimazione nel dominio. La definizione del passo deve preservare la qualità del segnale, e va stabilita una finestra temporale da considerare.

Le fasi della rappresentazione da analogica a digitale sono **campionamento**, **quantizzazione** e **codifica**.

Bisogna tenere conto di alcuni aspetti delicati, come i limiti di memoria, banda e tempi di processing: questi sono fattori importanti per determinare passo e finestra temporale. I valori cambiano anche a seconda del campo di analisi, per ottenere un'adeguata quantità di informazioni.

2 Multimedia processing

Per un efficace trattamento dei segnali multimediali è necessario minimizzare il quantitativo di dati processati, individuando solo quelli strettamente indispensabili. Oltre alla conversione e la compressione ci sono fasi come la memorizzazione e la trasmissione.

L'output è generalmente diverso dall'input: il segnale analogico x_a passa attraverso un campionario, da cui esce come x_n (non più analogico). L'output $x_q(n)$ contiene un insieme discreto di valori (sottoinsieme del dominio) che poi verrà trasformato in segnale quantizzato $x_q(n)$ e poi in bit.



Il valore di $x(n)$ è il valore della funzione analogica preso nT volte, dove T è il passo di campionamento. La frequenza è impossibile da ottenere senza la dimensione del passo, ed è legata a quanto velocemente varia il segnale.

L'obiettivo è capire se esiste una soglia in grado di stabilire se le informazioni vengono perse in base alla dimensione del passo. Avere un'alta **frequenza di campionamento** significa avere una buona qualità ma un volume elevato dei dati, mentre una bassa frequenza produce fenomeni di aliasing (approssimazione a costante).

Un segnale qualsiasi è rappresentabile come integrali di infiniti termini (seni e coseni) con peso e ampiezza diversi. Tra essi bisogna preservare quello con frequenza massima, per evitare sovrapposizioni e cambiamenti. La scomposizione del segnale è effettuata tramite **analisi di Fourier**.

Il **campionamento** con frequenza massima consente inoltre di riconvertire il segnale digitale in analogico senza perdita di informazione, perché permette la conservazione delle frequenze.

La **quantizzazione**, invece, è un'operazione che comporta sempre perdite, quindi non è reversibile.

Se il numero di frequenze tende a infinito, com'è possibile individuare la massima? In questo caso

si introduce il **filtering**: l'eliminazione dei valori esclusi da una certa soglia (superiore o inferiore). Un altro scopo del **filtro anti-aliasing** è l'eliminazione delle frequenze troppo basse.

Un esempio di elaborazione numerica è il filtraggio, l'eliminazione delle frequenze fuori dal range accettabile (filtro passa-basso o alto). I sistemi sono combinazioni di più operatori lineari, quindi nonostante complessi sono scomponibili a causa delle loro proprietà.

3 Segnali

Un **segnale** rappresenta il comportamento di grandezze fisiche in funzione di una o più variabili indipendenti. Sono *monodimensionali* se rappresentati da una sola variabile, per esempio il suono (continuo). I dati EEG sono *multidimensionali* in variazione al tempo, agli elettrodi e ai soggetti.

Le immagini in bianco e nero sono segnali *bidimensionali* (coordinate spaziali) e monocanale (il grigio), mentre quelle a colori hanno 3 segnali dimensionali RGB. Il campionamento corrisponde al numero di pixel, e la quantizzazione è la profondità del colore (quanti bit per la codifica). Aumentando il numero di livelli, aumenta la capacità di rappresentare l'informazione.

Se la variabile indipendente continua viene discretizzata è stato effettuato un campionamento, in cui è necessario conoscere la distanza tra i campioni (digitali).

Il valore assunto dal segnale si definisce **ampiezza** (dipendente, codominio) mentre l'asse delle ascisse è il **dominio** (tempo o spazio). Si possono introdurre grandezze statistiche come media e varianza, indicate in modo diverso a seconda del tipo di segnale.

- Continuo:

$$- \mu = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) dt;$$

$$- \mu = \frac{1}{T_1 - T_0} \int_{T_0}^{T_1} x(t) dt;$$

- Discreto:

$$- \mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i.$$

Il segnale digitale ha solamente una casistica di μ perché non tende mai a infinito, e il livello dell'ampiezza è diverso. Se un segnale varia, ha una componente continua DC (*direct current*, contributo a frequenza 0 e valore medio), e componenti AC a corrente alternata che variano in base a come il segnale fluttua intorno al valore medio.

Una forma d'onda ripetuta ha escursioni costanti, descritte da una grandezza chiamata ampiezza picco-picco A_{pp} . Il periodo è arbitrario.

Deviazione standard e varianza forniscono informazioni aggiuntive su quanto lontano (e con quale potenza) il segnale fluttua dal valore medio. Alla varianza è fortemente legata la potenza del rumore: un'alta varianza implica un forte rumore.

$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu)^2$$

Se media e varianza di un segnale non cambiano nel tempo, esso è stazionario. Al contrario, se il segnale varia la media sarà diversa a seconda della finestra, e la media globale non dà informazioni.



La **periodicità** indica la ripetizione del segnale nel tempo, definito appunto in periodi. Non esistono segnali puramente periodici, ma si usano approssimazioni delle forme d'onda che assume il segnale. L'inversa del periodo è chiamata **frequenza fondamentale**: $f_0 = 1/T$.

3.1 Segnale sinusoidale

Un segnale sinusoidale è un seno o un coseno, monodimensionale in funzione del tempo o dello spazio.

$$A(T) = A_{med} + B \cdot \sin(2\pi ft + \varphi_0)$$

$$A_{med} = \frac{1}{T} \int_0^T A(t) dt \quad A_{pp} = A_{max} - A_{min} \quad B = A_{pp}/2$$

Parametri importanti in questi casi sono **frequenza** e **fase**.

La frequenza si misura in Hertz, e rappresenta la rapidità con cui varia l'ampiezza in un intervallo temporale T . La pulsazione (intera variazione di ampiezza) è proporzionale alla frequenza, si ha che $\omega = 2\pi f$.

La fase segna l'alternarsi di positività o negatività del segnale, in particolare è significativa la fase iniziale φ_0 .

$$P(t) = |x(t)|^2 \quad \text{potenza istantanea}$$

$$E_x = \int_{-\infty}^{+\infty} P(x) dt \quad \text{energia: area di potenza istantanea}$$

Tanto più l'ampiezza si scosta dallo 0, più la potenza aumenta:

- Se $E_x < \infty$ il segnale ha energia finita;
- Quando $E_x = \infty$ si definisce la potenza media: $P_x = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt$.

Potenza media di un segnale periodico T : $P_x = \frac{1}{T} \int_T |x(t)|^2 dt$

Entrambi i valori hanno una componente continua (valore medio) calcolato come il limite dell'integrale o l'integrale (segnale periodico) di $x(t)$.

3.2 Decibel

Il **decibel** (dB) è un'unità di misura logaritmica (quindi non lineare), in cui lo scopo del logaritmo è visualizzare meglio grandi scale di valori e avvicinarsi alla percezione umana. La misura è quindi relativa e adimensionale.

$$decibel = dB \longrightarrow 10 \log_{10} \frac{P_1}{P_2} = 20 \log_{10} \frac{A_1}{A_2}$$

$$Bel \longrightarrow \log_{10} \frac{P_1}{P_2} \quad P = \text{potenza} \quad A = \text{ampiezza} \quad P \propto A^2$$

Le componenti della formula sono due pressioni, di cui il numeratore è la potenza del suono e il denominatore è la soglia minima di udibilità. La potenza è proporzionale al quadrato dell'ampiezza, quindi trasformando in scala logaritmica essa diventa un fattore moltiplicativo. $6dB$ rappresentano un raddoppio dell'ampiezza.

3.3 Trasformazioni di segnali

Operazioni molto comuni sono la traslazione, il cambio di scala e l'inversione temporale.

Se il segnale è reale, è possibile ritardarlo ma non anticiparlo.

- Ritardo: fissato un tempo t_0 , la traslazione trasforma il segnale $x(t)$ nel segnale $x(t - t_0)$;
- Anticipo: fissato un tempo t_0 , la traslazione trasforma il segnale $x(t)$ nel segnale $x(t + t_0)$;
- Cambio di scala: dato un numero reale $a > 0$, trasformazione del segnale $f(t)$ in $f(at)$:
 - Se $a > 1$ si ottiene una compressione lineare;
 - Se $a < 1$ si ottiene un allungamento lineare;
- Inversione: trasforma il segnale $f(t)$ nel segnale $f(-t)$.

Un segnale si dice pari se $f(t) = f(-t)$, dispari se $f(t) = -f(-t)$: il coseno è una funzione pari, il seno è dispari.

3.4 Segnali continui

Gradino: usato per selezionare la parte positiva dei segnali che tendono a $\pm\infty$.

Si definisce un **gradino unitario** $u(t)$:

$$u(t) = \begin{cases} 1 & \text{se } t \geq 0 \\ 0 & \text{se } t < 0 \end{cases}$$

Gradino traslato in t_0 : definito quando la finestra di osservazione è finita, centrata rispetto a 0. Se il segnale è fuori dall'intervallo assume valore 0:

$$u(t - t_0) = \begin{cases} 1 & \text{se } t \geq t_0 \\ 0 & \text{se } t < t_0 \end{cases}$$

Impulso rettangolare unitario $rect(t)$:

$$rect(t) = \begin{cases} 1 & \text{se } |t| \leq 1/2 \\ 0 & \text{se } |t| > 1/2 \end{cases}$$

Quest'ultima è una funzione che forma un rettangolo di area unitaria. Generalizzando, si ha un rettangolo di altezza A , base T e traslato in t_0 sostituendo nella formula unitaria t con $\frac{t-t_0}{T}$ e confrontandolo con $1/2$.

$$|t - t_0| \leq T/2 \begin{cases} \text{per } (t - t_0) > 0 & (t - t_0) \leq T/2 & t \leq t_0 + T/2 \\ \text{per } (t - t_0) < 0 & (t - t_0) \geq -T/2 & t \geq t_0 - T/2 \end{cases}$$

La moltiplicazione di un segnale per un rettangolo lo approssima con segmenti verticali o orizzontali a seconda dell'asse considerato. Media e varianza non sono le stesse rispetto alla funzione originale.

Funzione **delta di Dirac** $\delta(t)$: distribuzione con rettangolo di base infinitesima e altezza infinita che abbia l'area unitaria, con la larghezza che tende a 0 e di conseguenza l'altezza che tende a infinito.

$$\int_{-\infty}^{\infty} \delta(t) dt = 1$$

Limite dell'impulso rettangolare di base Δ per $\Delta \rightarrow 0$, dove Δ è uno scalare sull'asse x :

$$\delta(t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \text{rect}\left(\frac{t}{\Delta}\right) \quad \delta(t - x) = 0 \quad \text{se } t \neq x$$

Viene introdotta per rappresentare fenomeni fisici di durata infinitesima (impulsi).

Il corrispondente di δ discreto nel dominio è la **delta di Kronecker** o impulso unitario, rappresentato con una freccia verticale di altezza (o peso) unitario. Questa funzione ha significative applicazioni nel trattamento dei segnali.

Si ha un impulso $A\delta(n - n_0)$ di ampiezza A e che occorre al tempo $n = n_0$.

Con $n - 2$, δ è in ritardo di 2 (se fosse + sarebbe anticipo).

Esistono anche l'analogo discreto della delta di Dirac e del gradino unitario:

$$\delta(n) = \begin{cases} 1 & \text{se } n = 0 \\ 0 & \text{se } n \neq 0 \end{cases} \quad u(n) = \begin{cases} 1 & \text{se } n \geq 0 \\ 0 & \text{se } n < 0 \end{cases}$$

$$x(n) \cdot \delta(n) = x(0)$$

$$f(n) \cdot \delta(n - n_0) = f(n_0)$$

$$\delta(n) = u(n) - u(n - 1)$$

Il gradino continuo, nel discreto diventa una successione di δ : $u(n) = \sum_{i=0}^{+\infty} \delta(n - i)$



4 Sequenze

Le sequenze $x(n)$ sono formate da segnali a tempo discreto. Se essi sono quantizzati in ampiezza, si parla di segnale digitale.

Sequenza **causale**: $x(n) : n > 0$ (con numeri positivi).

Sequenza **anticausale**: $x(n) : n < 0$ (con numeri negativi).

Sequenza **pari**: $x(n) = x(-n)$ (coseno).

Sequenza **dispari**: $x(n) = -x(-n) \rightarrow x(0) = 0$ (seno).

Sequenza **periodica**: $x(n) = x(n + T)$

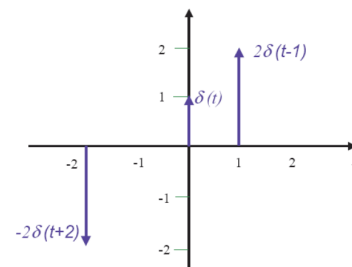
Sequenza **limitata**: $|x(n)| \leftarrow x_0 < \infty \quad \forall n$

Non è possibile tornare indietro nel tempo, quindi le sequenze considerate sono generalmente causali.

Nella figura, $\delta(n)$ rappresenta l'impulso unitario. Moltiplicando δ per un coefficiente k l'impulso avrà diversa lunghezza, o diversa direzione se k è negativo.

Un ritardo $t - t_0$ sposta l'impulso verso destra, un anticipo $t + t_0$ viceversa.

Se gli impulsi hanno un fattore moltiplicativo n , si definisce una funzione **rampa**: il segnale seguirà la diagonale.



5 Analisi di Fourier

L'analisi di Fourier decompone il segnale in costituenti sinusoidali di differenti frequenze. Il segnale non è più nel dominio tempo-spazio, ma delle **frequenze**: i dati sono gli stessi, cambia solo la rappresentazione.

Ogni funzione periodica e a quadrato sommabile può essere espressa come somma infinita e pesata di funzioni seno e coseno (combinazioni di funzioni armoniche).

Si ricorda che una sequenza **periodica** è $x(n) = x(n + T)$. Una funzione **armonica** è una funzione periodica del tipo:

$$y = A \sin(\varpi x + \varphi) \quad y = A \cos(\varpi x + \varphi)$$

Dove A è l'ampiezza, ϖ è la pulsazione, φ è la fase.

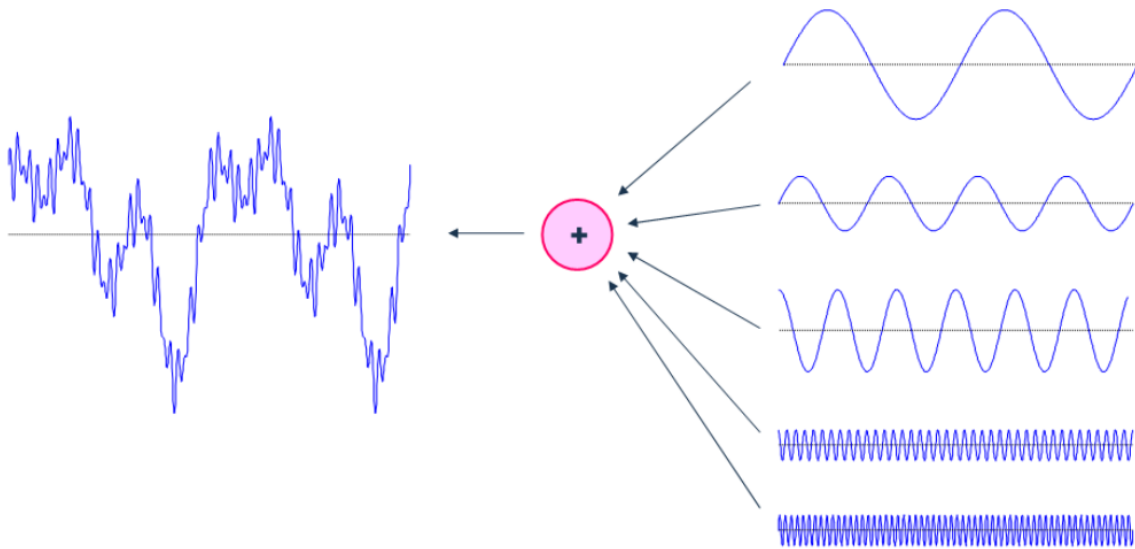
Si ha che $\varpi = 2\pi/T$ dove $1/T$ è la frequenza, e π è 180° .

Sviluppando i seni e i coseni si ha, con $a = A \sin(\varphi)$ e $b = A \cos(\varphi)$:

$$y = A \sin(\varpi x + \varphi) = a \cos(\varpi x) + b \sin(\varpi x)$$

$$y = A \cos(\varpi x + \varphi) = b \cos(\varpi x) + a \sin(\varpi x)$$

Le armoniche vengono combinate una per volta, avvicinandosi man mano alla funzione originaria.



5.1 Serie di Fourier

La serie di Fourier è una formula utile per approssimare la scomposizione. Rappresenta una funzione periodica mediante *combinazione lineare di funzioni sinusoidali*.

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{2\pi}{N} kx\right) + b_k \sin\left(\frac{2\pi}{N} kx\right)$$

$$N \rightarrow \text{periodo} \quad (1/N) \rightarrow \text{frequenza fondamentale } f_0 \quad (1/N)k \rightarrow \text{frequenze multiple } kf_0$$

a_k e b_k sono numeri reali, k è un numero intero che funge da fattore moltiplicativo e N è l'ampiezza della parte di funzione che si ripete periodicamente (l'inverso è la frequenza fondamentale). x è la variabile indipendente.

Le frequenze (infinite), quindi, sono una parte fondamentale della formula. Con tempo $T = 1/f$ il dominio del tempo è t , e il dominio della frequenza è Hz . In altre parole, il segnale passa dal dominio spazio-tempo alle frequenze, con coppie frequenza-peso associate alla sinusoidale (ampia A).

Estendendo la formula con il concetto di pulsazione, si ha che una pulsazione è $\varpi_k = 2\pi f_k$, e di conseguenza:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(2\pi k f_0 x) + b_k \sin(2\pi k f_0 x)$$

$$a_k = \frac{2}{N} \int_{-\frac{N}{2}}^{\frac{N}{2}} f(x) \cos(2\pi k f_0 x) dx \quad b_k = \frac{2}{N} \int_{-\frac{N}{2}}^{\frac{N}{2}} f(x) \sin(2\pi k f_0 x) dx$$

La variabile x e la funzione associata sono continue (periodiche), ma la variabile di frequenza è discreta (k intera), quindi dall'integrale si passa alla sommatoria.

Si ricorda che N è il periodo. Data la funzione $f(x)$ periodica, i coefficienti della serie sono **univocamente** determinati. I coefficienti sono i fattori moltiplicativi di seno e coseno, in relazione al tempo t .

Questo significa che esiste biunivocità, la trasformazione può essere effettuata da entrambi i versi senza perdere informazione.

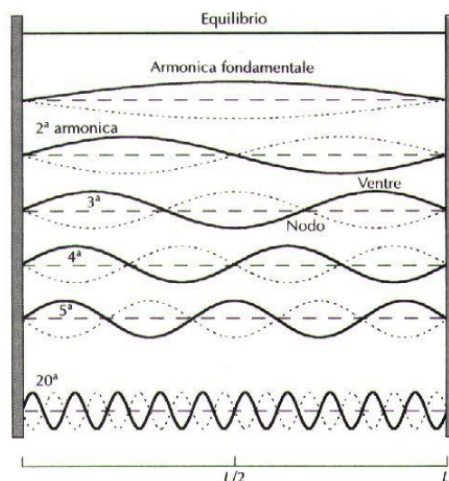
5.2 Fourier e i suoni

I suoni elementari hanno andamento sinusoidale, periodico e con estensione indefinita; la maggior parte dei suoni in natura sono però caratterizzati da forme d'onda diverse.

Un segnale si definisce **complesso** quando è formato da più funzioni sinusoidali combinate, invece che una sola.

Si può dimostrare che, fatte alcune ipotesi di regolarità sull'andamento della forma d'onda, un generico suono complesso può essere descritto come una *combinazione di suoni elementari* (armoniche).

Le armoniche sono funzioni che, combinate fra loro, permettono di determinare il timbro degli strumenti musicali. Insieme costituiscono forme d'onda.



5.3 Onda quadra

L'onda quadra è un caso particolare in cui tutte le armoniche pari sono nulle, e l'onda quadra è data dalla forma delle componenti $F_0, 3F_0, 5F_0, \dots$ cioè 200 Hz, 600 Hz, 1 Kz eccetera. Si ricorda che l'ampiezza di conseguenza è $1/3, 1/5, \dots$ cioè la parte **dispari** della sommatoria.

Ognuna di queste componenti ha un'ampiezza diversa rispetto a quella dell'armonica fondamentale. La funzione è asimmetrica, e i termini pari sono appunto i coseni: ci sono solo seni non nulli.



Esempio: un'onda quadra con periodo 5 ms può essere ottenuta sommando onde sinusoidali di opportuna frequenza, ampiezza e fase.

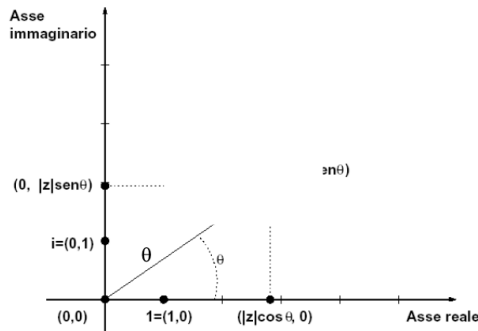
Il contributo più rilevante è dato dalla prima sinusoide, si frequenzia pari a quella dell'onda quadra (200 Hz, frequenza fondamentale F_0).

Per costruire l'onda quadra sono necessarie anche altre componenti elementari di frequenza maggiore, cioè tutte le armoniche multiple della frequenza F_0 .

5.4 Numeri complessi e serie di Fourier

I numeri complessi sono rappresentabili su un piano cartesiano come l'intersezione tra numeri reali sull'asse x e numeri complessi sull'asse y . Ogni numero complesso ha quindi una parte **reale** a e una **immaginaria** b .

$$z = ai + b \quad |z| = \sqrt{a^2 + b^2} = \pi$$



Forma polare:

$$a = \pi \cos(\theta)$$

$$b = \pi \sin(\theta)$$

$$\theta = \arctan\left(\frac{b}{a}\right)$$

Le coordinate polari corrispondono a quelle cartesiane (proiezione).

Applicando i numeri complessi alla serie di Fourier e quindi cambiando dominio, si possono usare gli esponenziali complessi e la formula di Eulero.

$$e^{j\theta} = \cos \theta + j \sin \theta \quad e^{-j\theta} = \cos \theta - j \sin \theta$$

Sia seno che coseno sono somme di due esponenziali complessi, con segno + e - rispettivamente. Il picco della frequenza è intero per il coseno, immaginario per il seno.

$$\cos \theta = \frac{1}{2}(e^{j\theta} + e^{-j\theta}) \quad \sin \theta = -\frac{j}{2}(e^{j\theta} - e^{-j\theta})$$

La parte del seno è immaginaria, il coseno è reale. Le formule sono le stesse, ma applicate al sovrainsieme dei complessi \mathbb{C} (j e i sono indicatori convenzionali della parte immaginaria).

Mettendo a sistema seno e coseno ottenuti in questo modo, si può estendere la serie di Fourier rappresentando le somme di contributi nello spazio degli esponenziali complessi.

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(2\pi f_k x) + b_k \sin(2\pi f_k x)$$

$$f(x) = \sum_{-\infty}^{\infty} R_k e^{j(2\pi f_k x)}$$

Si ricorda che $\theta = 2\pi f_k$. Quando θ cambia segno, l'angolo si sposta. Essendo il coseno pari, questo non cambia; il seno è dispari quindi è l'opposto.

Per passare dalle variabili discrete alle continue, si usa il coefficiente R_k che raccoglie le due sommatorie (fino a infinito da entrambi i versi) e si applica l'integrale.

$$R_k = \int_{-\frac{N}{2}}^{\frac{N}{2}} f(x) e^{-j(2\pi f_k x)} dx$$

Il coefficiente $\frac{A_n}{2}$ (altezza del segnale sull'asse y) è dato appunto dalla divisione del coefficiente R in 2 parti, seno e coseno: l'ampiezza sarà la metà, e ogni ampiezza viene sommata insieme alle funzioni.

C'è corrispondenza biunivoca fra dominio temporale e dominio delle frequenze. I seni hanno solo la parte complessa.

5.5 Trasformata di Fourier di funzioni continue

Per passare dalle funzioni periodiche a quelle reali, la sommatoria diventa un integrale. La trasformata di Fourier serve per rappresentare le funzioni continue come sinusoidi.

Ogni funzione continua $f(x)$ anche se non periodica (purché abbia area finita) può essere espressa come *integrale di sinusoidi complesse opportunamente pesate*. L'antitrasformata ha dominio spaziale (temporale, diretto) mentre la trasformata ha dominio delle frequenze.

$$\text{Antitrasformata di Fourier: } f(x) = \int_{-\infty}^{\infty} F(u) e^{j2\pi ux} dx$$

$$\text{Trasformata di Fourier: } F(u) = \int_{-\infty}^{\infty} f(x) e^{-j2\pi ux} dx$$

$F(u)$ sostituisce il coefficiente discreto R_k , u è la frequenza. La trasformata di Fourier non si può applicare a tutte le funzioni, ma solo a quelle con quadrato sommabile e **continue**.

Dalla trasformata di Fourier, la funzione originaria può essere ricostruita **senza perdita di informazione** attraverso il processo inverso: è possibile spostarsi dal dominio spaziale a quello delle frequenze e poi tornare indietro senza aver perso informazione.

Esiste un valore della trasformata (non più δ) per ogni funzione continua. Invertendo la funzione, quindi, se è a supporto finito non ci sarà 2δ .

Piuttosto che parte reale e immaginaria, si considerano modulo e fase, ricavabili direttamente. Il modulo indica la **potenza** del segnale, e la fase assicura la **biunivocità** della corrispondenza.

$$F(u) = F[f(x)] = \Re(u) + j\Im(u) = |F(u)| e^{j\phi(u)}$$

$$\text{Spettro: } |F(u)| = [\Re(u)^2 + \Im(u)^2]^{1/2}$$

$$\text{Fase: } \phi(u) = \tan^{-1} \left[\frac{\Im(u)}{\Re(u)} \right]$$

$$\text{Potenza (densità) spettrale: } |F(u)|^2 = \Re(u)^2 + \Im(u)^2$$

La trasformata di Fourier è un operatore **scomponibile**, quindi per passare da una dimensione a n semplicemente si applica n volte la formula. In altre parole, si scompone rispetto a una direzione la funzione.

5.6 Trasformata di Fourier di funzioni discrete

Il primo step è il campionamento: in questo caso si utilizza la sommatoria, ma la funzione continua è rappresentata da un numero discreto di campioni. La funzione è definita per campioni i , e la variabile indipendente x assume valore complesso $i\Delta x$.

$$F(u) = \int_{-\infty}^{+\infty} f(x)e^{-j2\pi ux} dx \implies F(u) = \sum_{i=-\infty}^{\infty} f(i)e^{-j2\pi ui\Delta x}$$

Limitando la funzione spazialmente, il passo Δx per convenzione diventa $1/N$, e la sommatoria arriva fino a $N-1$ invece che infinito. $f(i)$ è la funzione campionata (a partire da $f(x)$ continua): la trasformata è **periodica**. Si ha $f(i) = f(x_0 + i\Delta x)$.

$$F(u) = \frac{1}{N} \sum_{i=0}^{N-1} f(i)e^{-j2\pi u \frac{1}{N} i}$$

Per campionare le funzioni $f(x)$ e $F(u)$ si applicano le trasformate di Fourier discrete (DFT). Il segnale campionato è periodico, quindi la funzione inversa è comunque periodica con distanza $1/\Delta x$. Se il passo tende a 0, la funzione tende a infinito (senza repliche) e viceversa.

In altre parole, da una funzione campionata $\Delta x = 1/N$ si passa a una funzione continua e periodica di periodo N .

Il fattore $1/N$ davanti alla trasformata o all'antitrasformata è un fattore di normalizzazione per il dualismo fra spazio diretto e spazio trasformato. Di conseguenza, $\Delta u = \frac{1}{N\Delta x}$.

Un ragionamento analogo si applica per passare da una dimensione a due: la sommatoria è doppia e si divide per entrambi i fattori di normalizzazione. Entrambe le variabili rispettano la periodicità, cioè $F(u, v) = F(u + M, v + N)$ con $\Delta u = 1/M\Delta x$, $\Delta v = 1/N\Delta y$.

5.7 Elaborazione di immagini

Il segnale di un'immagine viene scomposto, come visto precedentemente, in modulo e fase. Il modulo raccoglie le *frequenza* di un'immagine, mentre la fase contiene l'*informazione posizionale* di ogni in un'immagine (le frequenze sono le stesse a prescindere dall'organizzazione spaziale).

L'ampiezza \angle contiene informazioni relative alle posizione o meno delle strutture nell'immagine.

$$F(u, v) = |F(u, v)| e^{j\phi(u, v)}$$

I pixel possono assumere valore da 0 a 255, mappati in una *trasformata di Fourier discreta*. Data la periodicità e la simmetria dello spettro rispetto all'origine, è possibile scegliere il periodo dello spettro centrato sull'origine $(0, 0)$.

In questo modo, il picco rappresenta la **componente continua** (valore medio della funzione), e le frequenze crescono allontanandosi dal picco. Per convenzione lo spettro (in alto a sinistra) si trasla in modo da essere al centro, e le alte frequenze sono verso gli angoli esterni.

Il processo è reversibile: l'immagine, di solito sotto forma di segnale, se viene ricostruita dalla trasformata è identica all'originale.

Un operatore puntuale ha una corrispondenza biunivoca non indipendente dall'ambiente, cioè ogni pixel dell'immagine originale viene mappato nel modulo nello stesso modo a prescindere dai pixel intorno.

Al contrario, la trasformata di Fourier funziona applicando **sommatorie** (per il segnale), quindi ogni valore dipende da tutti i precedenti. Ciò si definisce **operatore locale**: ogni frequenza ha una sua influenza, e sebbene ci possano essere corrispondenza tra linee e forme, ogni pixel nel dominio è ricavato in base a tutti i pixel nel codominio (e viceversa).

L'*istogramma di un'immagine a livelli di grigio* è una funzione discreta $h(r_k) = n_k$ dove r_k è il livello di grigio k -simo, e n_k è il numero di pixel nell'immagine con intensità r_k (per ogni livello).

Il numero di campioni stabilisce l'informazione che si può rappresentare: se è sufficientemente alto, la distribuzione dei toni di grigio sarà regolare. Questo non permette di trarre conclusioni sullo spazio, ma semplicemente su quanti campioni hanno un certo valore (e di conseguenza la bontà del quantizzatore).

Se ci sono picchi nella distribuzione, è dovuto alla ripetizioni di pattern nell'immagine (come lo sfondo). Il numero di campioni nella trasformata è uguale a quello nell'immagine, e con frequenza di campionamento bassa la Δf sarà piccolo e si verificherà l'aliasing.

Per rimuovere i rumori si acquisiscono porzioni di segnali più lunghe, eventualmente analizzando le loro medie in modo da preservare le componenti più importanti.

Lo **spettro** è il risultato della trasformata, di cui una parte è reale e una immaginaria. Lavorando sul modulo, è possibile individuare circonferenze che indicano dove si concentra la potenza del segnale: manipolando lo spettro il risultato dell'inversione sarà differente.

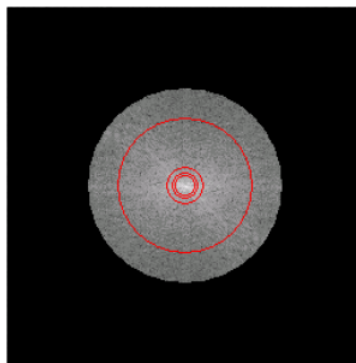
Alcune metodologie di filtraggio comportano l'applicazione di maschere, finestre (monodimensionali, rappresentate come un rettangolo nel piano) a sezione circolare. Modificare lo spettro, quindi, implica ricostruire una diversa immagine che sarà più o meno accurata rispetto all'originale.

Spesso si possono eliminare le ridondanze senza perdere informazioni. Su questi principi si basano gli algoritmi di compressione (es. JPEG), riducendo la complessità a discapito della qualità.

Si ha che il 99.5% dello spettro è contenuto in una circonferenza di dimensione limitata.



Originale



Spettro (99.5%)



Ricostruita

Eliminando parte dello spettro e delle frequenze, il filtro diventa parte integrante del segnale che viene antitrasformato: ciò introduce fenomeni di *ringing*, quindi è necessario osservare anche il comportamento dell'antitrasformata del filtro.

Per ottenere diversi effetti si selezionano diverse parti dello spettro da tenere, per esempio solo l'interno o l'esterno, o una banda di un certo intervallo (mosaico, *smoothing*).

5.8 Proprietà della trasformata

Un'importante proprietà è la **linearità**, che vale per ogni dominio: applicando la trasformata di Fourier a una funzione che è combinazione lineare (somma pesata) di funzioni, il risultato è uguale alla trasformata di Fourier di ogni funzione presa singolarmente.

$$F(f(x) + g(x)) = F(f(x)) + F(g(x)) = F(u) + G(u)$$

La linearità permette di ridurre (scomporre) la complessità della trasformata come *sovrapposizione di sistemi più semplici e monodimensionali*, senza introdurre dispersione del segnale.

Vale anche la **traslazione** nello spazio (tempo) e nelle frequenze, ma solo per la trasformata discreta: una traslazione equivale a una modulazione nelle frequenze con un esponente complesso. In altre parole, la trasformata del modulo è identica, ma la fase cambia di x_0 (sfasamento).

$$F(f(x + x_0)) = e^{j2\pi u x_0} \quad F^{-1}(F(u \pm \varpi)) = f(x) e^{\pm j2\pi \varpi x}$$

A oggetti più grandi nel dominio diretto, corrisponderanno oggetti simili ma compressi nel dominio trasformato e viceversa, a causa dell'aumento delle frequenze.

$$\text{Scala: } F(f(ax)) = \frac{1}{|a|} F\left(\frac{u}{a}\right) \implies f(\alpha x, \beta y) = \frac{1}{|\alpha\beta|} F\left(\frac{u}{\alpha}, \frac{v}{\beta}\right)$$

$$\text{Inversione: } F(f(-x)) = F(-u)$$

La trasformata di una funzione reale gode di *simmetria Hermitiana*: la parte reale e il modulo sono simmetrici rispetto all'origine (lo 0), mentre la parte immaginaria e la fase sono antisimmetriche rispetto all'origine.

Campionando il segnale di partenza, la trasformata diventa periodica e il periodo permette di ricostruire il segnale, a sua volta periodica.

$$\text{Periodicità (discreta): } F(u) = F(u + M) \implies F(u) = \frac{1}{M} \sum_{x=0}^{M-1} f(x) e^{-j \frac{2\pi}{M} u x} \quad u = 0, \dots, M-1$$

Altre proprietà sono già state enunciate, come l'**esistenza dell'inversa** della trasformata discreta e la **reversibilità**.

$$\text{Separabilità: } F(u, v) = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) e^{-j2\pi \left(\frac{ux}{N} + \frac{vy}{N}\right)}$$

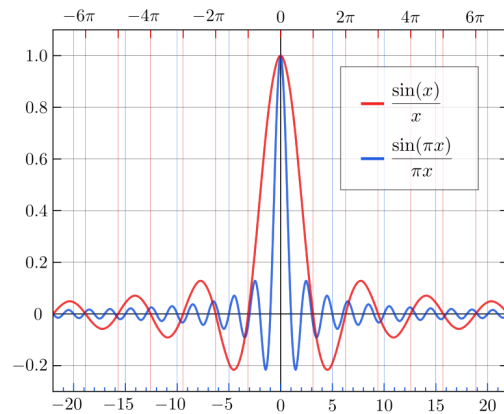
$$\text{Valore medio di una trasformata discreta: } F(0) = \frac{1}{M} \sum_{x=0}^{M-1} f(x) = \bar{f}(x)$$

$F(u_0)$ è il peso con cui l'onda complessa $e^{j2\pi u_0 x}$ di frequenza u_0 concorre a formare il segnale $f(x)$. Se $f(x)$ è un'onda complessa di frequenza u_k , $F(u)$ sarà 0 per ogni frequenza diversa da u_k .

5.9 sinc

sinc è un filtro passa-basso, utile per preservare la parte centrale di segnali con basse frequenze.

Nel dominio del coseno trasformato non ci saranno solo le δ se il segnale non è periodico e infinito, ma ci sarà anche l'effetto del *sinc* (finestra), dove $\text{sinc}(x) = \frac{\sin(x)}{x}$, con un fattore di normalizzazione π [immagine: Georg-Johann].



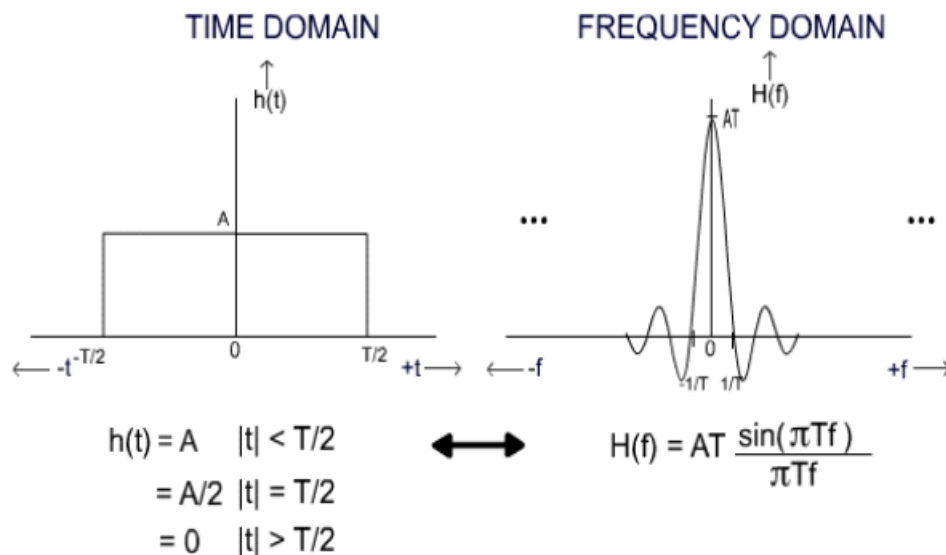
Questa funzione rappresenta il **limite** con x che tende a 0 della funzione seno, cioè 1. Nel processamento dei segnali è la trasformata di Fourier della funzione rettangolare (**finestra**) non scalata, ricostruendo il segnale continuo a partire dai suoi campioni.

Per il seno, c'è una δ positiva e una negativa, e se trasformando entrambe diventano positive non c'è più la possibilità di distinguerlo dal coseno.

In realtà le due funzioni sono la stessa traslata (proprietà di traslazione). La δ positiva della parte immaginaria in realtà corrisponde alla parte negativa.

L'altezza di δ nel punto 0 è uguale all'altezza del segnale, che varia sempre meno fino a diventare costante. La **trasformata di una costante**, quindi, è rappresentabile come un **coseno** le cui δ si sovrappongono nel centro (limite della funzione, con frequenza che tende a 0).

La finestra di osservazione nel tempo ha frequenza δ , con due picchi di seno e coseno, e il contributo di *sinc* cresce con l'aumento dell'ampiezza della finestra (ringing). Al contrario, *sinc* si avvicina sempre di più a δ con il restringimento.



6 Campionamento

Il **campionamento** è un processo che legge valori distanziati l'uno dall'altro con un **passo** Δt approssimato al tempo infinitesimo, ma in realtà discreto. La scelta dell'ampiezza del passo deve evitare sia lo spreco di risorse che la perdita di informazioni.

Il segnale, in pratica, passa dall'analogico al digitale e dal continuo al discreto, e viene rappresentato come una serie di impulsi.

A una funzione campionata (con passo Δx) corrisponde nel dominio trasformato uno **spettro periodico**. Il periodo è inversamente proporzionale al passo di campionamento (ricostruzione della funzione continua dalla trasformata).

Così come a una funzione continua nel dominio dello spazio ne corrisponde una nel dominio delle frequenze espressa come somma pesata di armoniche (analisi e sintesi), a una funzione campionata con passo Δx corrisponde nel dominio trasformato uno spettro periodico, con periodo in funzione del passo di campionamento.

$$F(u) = \sum_{k=-\infty}^{\infty} f(k\Delta x) e^{-j2\pi u k \Delta x}$$

La *frequenza di campionamento minima*, quindi, garantisce un periodo sufficientemente grande per contenere l'intero spettro. Esso è simmetrico, quindi $-f_{max} = f_{max}$ come visto nel dominio della trasformata di Fourier.

Se il passo di campionamento è troppo alto ($\Delta x' = 1/M > \Delta x = 1/N$), le repliche dello spettro si sovrappongono (periodo M troppo breve) e non è possibile risalire alla funzione originale.

Questo succede perché le alte frequenze che non vengono considerate (a causa dell'alta variabilità del segnale, che non può essere interpretata in un periodo) ma si sovrappongono alle più basse, causando una errata rappresentazione.

Al contrario, una bassa frequenza di campionamento produce fenomeni di *aliasing*: la funzione viene approssimata male, come costante o una differente sinusoide.

Il **teorema di Shannon** afferma che, data la frequenza massima del segnale f_{MAX} , la frequenza massima di campionamento deve essere:

$$F_s = 1/\Delta > 2f_{MAX}$$

Il periodo minimo di ripetizione dello spettro perché non vi sia sovrapposizione $N = 2f_{MAX}$. La f_{MAX} viene definita **frequenza di Nyquist**, e rappresenta la frequenza massima del segnale.

Lo spazio pertanto sarà $2f_{MAX}$, il che è il minimo valore per evitare le sovrapposizioni delle repliche. Nella formula viene utilizzata la disuguaglianza stretta per assicurare che non ci siano due frequenze nello stesso punto.

Per eliminare le frequenze alte e di conseguenza ridurre il passo di campionamento si ricorre al filtraggio (filtro anti-aliasing), che comunque non assicura il recupero della funzione originale ma evita le sovrapposizioni.

L'aliasing è appunto il fenomeno per cui il segnale originale non è ricostruibile dato che il campionamento è avvenuto con frequenza inferiore a quella di Nyquist, e il segnale risultante ha frequenza inferiore all'originale. Questo può gravemente compromettere la qualità di immagini e video: nella

pratica i segnali non sono limitati in frequenze, ma nel tempo, quindi le repliche di uno spettro vanno separate attentamente tramite filtri passa-basso.

Quando il segnale è campionato, la variabile tempo/spazio non è più esplicita e la frequenza f_N rappresenta il numero di cicli per campione.

6.1 Frequenze

Il range tra minimo e massimo di una frequenza di oscillazione è 1, tutto ciò che è al di fuori dell'intervallo in realtà si trova comunque normalizzato all'interno. Aumentando ω , infatti, dopo un giro (periodo 2π) si torna ad avere la stessa funzione.

Frequenza minima di oscillazione di una sinusoide a tempo discreto: $f_N = 0 \implies \omega = 0$

Frequenza massima di oscillazione di una sinusoide a tempo discreto: $f_N = \pm 1/2 \implies \omega = \pm \pi$

Frequenza di oscillazione di una sinusoide generica: $f = \frac{1}{n. \text{ campioni}} \implies \omega = 2\pi \cdot f_N$

I segnali sinusoidali a tempo discreto con pulsazioni separate da multipli di 2π sono identici:

$$\omega + k2\pi = \omega \text{ con } k \in \mathbb{Z}$$

$$x(n) = \cos(\omega n + \theta)$$

$$\cos((\omega + 2k\pi)n + \theta) = \cos(\omega n + \theta)$$

$$\cos(\omega n + kn2\pi + \theta) = \cos(\omega n + \theta)$$

Per convenzione si utilizza la frequenza di Nyquist, $f_n = [-1/2, 1/2]$, oppure intervalli di pulsazione $[-\pi, \pi]$ o $[0, 1] \rightarrow [0, 2\pi]$.

Il segnale sinusoidale è periodico solo se la sua frequenza normalizzata f è un numero razionale (rapporto tra due interi): $f = k/N$. I numeri razionali quindi impongono la periodicità.

7 Convoluzione

La **convoluzione** è un operatore matematico che descrive l'interazione tra il segnale e il filtro, e ne osserva il comportamento durante il movimento nello spazio. Si definisce nel dominio continuo fra due funzioni $g(x)$ e $f(x)$ tale che:

1. L'asse di rappresentazione di uno dei due assi è invertita, $g(t) \rightarrow g(-t)$;
2. Il segnale invertito viene fatto traslare tra ∞ e $-\infty$;
3. Per ogni traslazione si calcola il prodotto tra il segnale traslato e l'altro non traslato;
4. Si calcola l'**area** del prodotto, cioè *la somma degli infiniti prodotti* composti da una funzione che scorre e una statica.

La convoluzione è l'operatore $*$ con cui sono descritti i filtri lineari nel dominio spaziale, cioè l'applicazione di una funzione f a una funzione h chiamata **filtro** (*filter kernel*, descrive il sistema) per ogni valore di x . Gode della proprietà commutativa.

$$g * f = \int_{x=-\infty}^{\infty} g(x-s)f(s)ds$$

Nel caso di funzioni discrete, essa è definita come una **somma di prodotti** tra gli elementi di f e i coefficienti di h , con funzioni che sono in realtà sequenze:

$$g(x) = \sum_m f(m)h(x-m)$$

Se i segnali non si sovrappongono, la somma dei loro prodotti è 0. Quando iniziano a sovrapporsi progressivamente, se sono positivi, i valori crescono fino a raggiungere un massimo, per poi diminuire.

La lunghezza finale della funzione ottenuta dipende dai due segnali che la compongono: assumendo che non ci siano valori nulli, con A punti di $f(m)$ e B di $h(-m)$ si ha $g(x)$ con $A + B - 1$ punti.

Applicando la definizione di trasformata di Fourier, si può dimostrare il **teorema fondamentale della convoluzione**: *la trasformata della convoluzione di due funzioni è il prodotto delle trasformate delle due funzioni*.

$$G(u) = F[g(x)] = F[f(x) * h(x)] = F(u)H(u)$$

La somma di prodotti è più complessa computazionalmente rispetto alla trasformata, e il teorema della convoluzione permette di utilizzare quest'ultima. Ciò funziona grazie alla proprietà di traslazione.

Per la corrispondenza tra dominio spaziale e dominio delle frequenze, si hanno le seguenti relazioni:

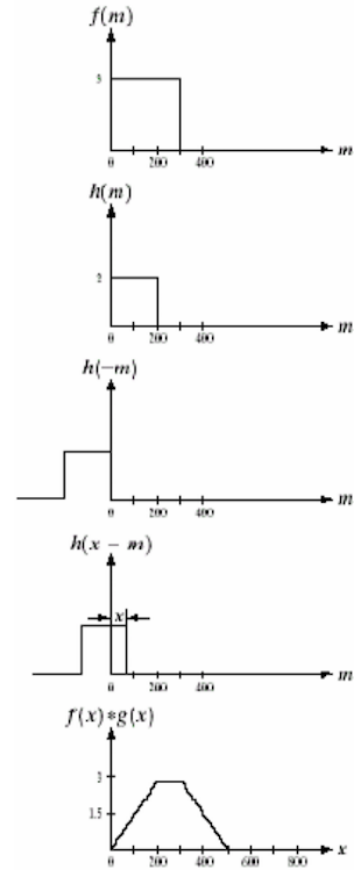
$$g(x) = f(x) * h(x) \iff G(u) = F(u)H(u)$$

$$g(x) = f(x)h(x) \iff G(u) = F(u) * H(u)$$

Un prodotto nello spazio-tempo, quindi, corrisponde a una convoluzione di trasformate. Per capire il comportamento del segnale è sufficiente osservarlo nel dominio trasformato.

La convoluzione è un operatore **lineare**, verificabile attraverso la relativa proprietà:

$$f(x) * [\alpha g_1(x) + \beta g_2(x)] = \alpha[f(x) * g_1(x)] + \beta[f(x) * g_2(x)]$$



8 Quantizzazione

La **quantizzazione** è un processo di discretizzazione dell'ampiezza: i segnali a tempo discreto sono convertiti a valori discreti (digitali), cioè appartenenti a un insieme limitato di possibili scelte.

Risoluzione: un campione reale che necessita ipoteticamente di un numero infinito di bit per essere rappresentato, è espresso su un numero finito. Il processo è **irreversibile**, a causa della perdita di informazione.

La caratteristica di un quantizzatore è pertanto la curva non lineare, ma a gradini. Per tutti i valori di input che appartengono a uno degli intervalli su cui sono definiti i gradini, l'output assume il valore del gradino corrispondente.

La funzione più semplice in cui l'input è uguale all'output è la *diagonale*: la funzione ottenuta sarà quindi simmetrica, con una curva caratteristica non uniforme (larghezza dei gradini). Il quantizzatore è uniforme se tutti i livelli sono distribuiti ugualmente rispetto all'asse delle ascisse. La dinamica $[-V, V]$ viene quindi divisa in sottointervalli della stessa ampiezza $\Delta = 2V/L$ dove L è il passo.

Così come il passo di campionamento, esiste anche il passo di quantizzazione, cioè la distanza tra il valore minimo e massimo. L'altro parametro è la risoluzione, cioè il **numero di livelli** proporzionale al numero di bit (con n bit, 2^n livelli).

Il processo consiste nell'associare a ciascun campione $x(m)$ il numero binario $x_q(m)$ corrispondente al livello quantizzato dell'intervallo in cui cade $x(m)$.

Tra tutti i livelli è necessario definire un criterio di scelta, essendo l'operazione irreversibile. Il quantizzatore è in grado di coprire solo un range L di valori.

Se L è maggiore del passo di quantizzazione Δ , il valore massimo $L\Delta$ sarà un intervallo troppo grande rispetto alla variazione del segnale e non verrà usato completamente (spreco).

Al contrario, se le variazioni del segnale sono molto ampie, il range sarà grande e andrà adattato a un numero ristretto di livelli.

Un quantizzatore è caratterizzato da una dinamica di ingresso, cioè un massimo range di valori ammissibili. Se il segnale supera gli estremi, esso viene modificato attraverso la **saturation** o la saturazione con azzeramento.

L'**intervallo di copertura**, definito dalla risoluzione e dal passo, è quindi $D_q = \Delta L = \Delta 2^n$. Si utilizza un range dinamico per avere un numero flessibile di livelli rappresentabili, convertibile in decibel: l'obiettivo è il riempimento di tutti i livelli.

Il **range dinamico** è il rapporto tra il numero minimo e massimo di valori rappresentabili (espressi dal numero di bit):

$$20 \log_{10} L$$

Se la risoluzione è 16 bit, il range dinamico è $20 \log_{10} 2^{16} \approx 96dB$.

Le strategie di quantizzazione sono il troncamento (Δ) e l'approssimazione ($\frac{\Delta}{2}$): entrambe causano un errore, e di conseguenza lo sfasamento dei gradini rispetto alla diagonale.

Si definisce **errore (rumore) di quantizzazione** la differenza tra il valore quantizzato e il valore reale del campione:

$$\varepsilon_q(n) = x_q(n) - x(n)$$

L'incidenza dell'errore nel segnale è misurata come **SRN** (*Signal-Noise Ratio*, potenza), e tipicamente si misura in dB. Il valore finale, naturalmente, non è influenzato solamente dalla quantizzazione ma anche da fattori esterni, quindi è difficile avere stime oggettive della dinamica del quantizzatore.

La qualità del segnale quantizzato si esprime come rapporto della potenza P_S media del segnale a tempo discreto $x(n)$ e la potenza media dell'errore di quantizzazione P_N .

$$SNR_Q = 10 \log_{10} \frac{P_S}{P_N}$$

Le valutazioni sono eseguite in valore assoluto, per eliminare la possibilità di annullamento dei termini. La potenza del rumore è legata al modulo della sequenza.

$$P_N = \frac{1}{N} \sum_{n=0}^{N-1} |\varepsilon_q(n)|^2 \quad P_S = \frac{1}{N} |x(n)|^2$$

Se il rumore è semplice e il segnale ha ampiezza nella dinamica del quantizzatore, varia nell'intervallo $\pm\Delta/2$, ed è equiprobabile e casuale con valor medio nulla (distribuzione uniforme). La potenza del rumore equivale alla varianza di ε , variabile casuale.

La situazione ideale ha un rapporto segnale-rumore **ottimale** (uguaglianza), cioè esso si adatta perfettamente ai livelli e la dinamica del quantizzatore è uguale a quella del segnale. Il range di valori è pari a $2A$, per un segnale sinusoidale.

$$P_N = \sigma_\varepsilon^2 = \frac{\Delta^2}{12} = \frac{A^2/3}{2^{2b}} \quad P_S = \frac{A^2}{2}$$

$$x_a(t) = A \cos(\omega_0 t) \text{ segnale sinusoidale analogico}$$

$$P_S = \frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2 \text{ segnale sinusoidale campionato}$$

La distanza dei livelli, di conseguenza, sarà $2A/2^b$, essendo la potenza del rumore strettamente legata all'ampiezza del segnale (aumenta in proporzione al numero di bit). Conoscendo il modo di calcolare P_S e P_N , insieme alla dinamica del quantizzatore $D_q = 2^b \Delta$, si ottiene il quantizzatore ottimale:

$$SNR_q = 20 \log \frac{\sqrt{P_S}}{D_q} + 10,8 + 6,02b$$

Il range del segnale sinusoidale è $D = 2A$, e con b bit di quantizzatore in 2^b livelli è possibile ricavare una nuova rappresentazione sostituendo nella formula generale:

$$SNR_Q = 1,76 + 6,02b$$

Per un insieme di segnali più ampio che si distribuisce sull'intero range dinamico del quantizzatore, si somma 1,25 invece che 1.76. Un incremento di un bit porta a un incremento di circa 6 dB, cioè l'intensità del segnale raddoppia rispetto all'intensità del rumore.

Dato che la quantizzazione provoca una perdita irreversibile di informazione, tenendo conto dell'errore, è importante stabilire in quali condizioni la distorsione introdotta è minima. Anche se il campionamento è ideale, la conversione D/A può invertire solo quest'ultimo senza considerare il rumore, quindi $x_r(t) \neq x(t)$.

Maggiore è la risoluzione (bit), più il segnale si avvicina all'originale, e l'ottimo del quantizzatore è il numero minimo dei livelli per avere una distorsione accettabile.

L'operazione che permette di ricostruire il segnale analogico $x(t)$ a partire dalla sequenza $x(NT_c)$ è detta **ricostruzione**. Essa è fedele se:

- Non c'è quantizzazione;
- Il campionamento è avvenuto nel rispetto del teorema di Shannon.

Nei convertitori digitale/analogico i punti del segnale vengono interpolati per generare il segnale analogico ricostruito. L'accuratezza della ricostruzione dipende dalla qualità del processo di conversione.

9 Sistemi lineari

Un **sistema** fisico è un apparato che riceve in input un segnale e ne produce un altro in output. Questa interazione è descritta da una funzione (operatore), le cui caratteristiche determinano il segnale in uscita. Il processo contiene una relazione di ingresso-uscita, o causa-effetto.

$$y(t) = S[x(t)]$$

Il campionamento con periodo o frequenza è un esempio di sistema, applicando un campionatore che legge istanti continui e li trasforma in discreti in base al passo Δt .

Un altro esempio è il quantizzatore, che associa al segnale $f(t)$ il segnale $Q(t)$ dove Q indica l'elemento in $V = \{x_1, \dots, x_n\}$ più vicino al numero reale x (valori finiti).

I sistemi possono avere più ingressi e uscite, come la somma $g(t) = f(t) + h(t)$. Mettendo insieme più sistemi semplici, ne vengono generati di più complessi in diverse combinazioni di componenti elementari:

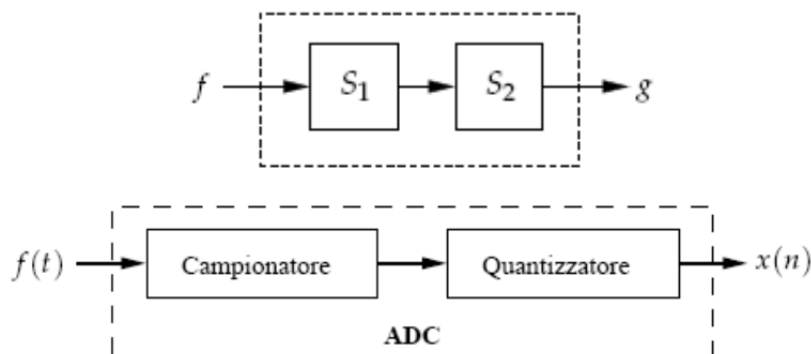
- **Composizione sequenziale** (a cascata), che combina più sistemi di cui l'output di uno è l'input di un altro (come processo di digitalizzazione);
- **Composizione parallela**, in cui lo stesso input viene immesso in più sistemi diversi;
- **Retroazione**, in cui l'output di un sistema viene modificato da un'altra funzione e torna come input della precedente (il risultato dipende dai valori precedenti).

Un sistema a tempo discreto è un dispositivo che trasforma una sequenza $x(n)$ in ingresso in una sequenza di uscita $y(n)$, attraverso un operatore $L[\cdot]$ (relazione di I/O). L'operatore è descrivibile tramite una sequenza, che definisce le proprietà del sistema.

9.1 Composizioni

9.1.1 Cascata

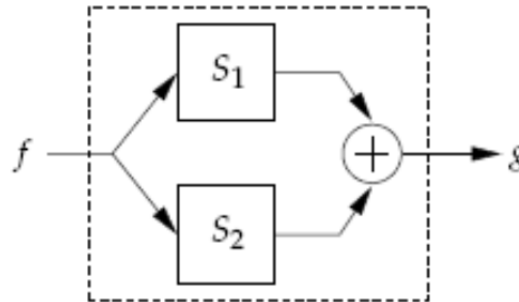
Dati due sistemi $S1 : F1 \rightarrow F2$ e $S2 : F2 \rightarrow F3$, la loro composizione sequenziale è il sistema $S3 : F1 \rightarrow F3$ ottenuto ponendo in ingresso a $S2$ la risposta di $S1$.



9.1.2 Parallela

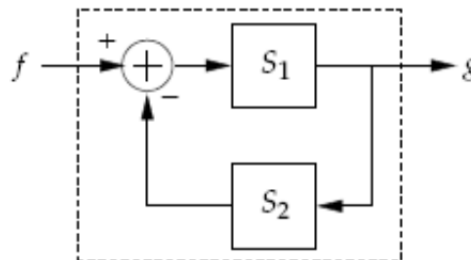
Dati due sistemi $S_1 : F_1 \rightarrow F_2$ e $S_2 : F_2 \rightarrow F_3$, la loro composizione parallela è il sistema che ha come risposta la somma delle risposte di S_1 e S_2 .

Perché si possa definire la composizione parallela, deve succedere che la somma di due segnali in F_2 sia ancora un segnale in F_2 .



9.1.3 Retroazione

Dati due sistemi $S_1 : F_1 \rightarrow F_2$ e $S_2 : F_2 \rightarrow F_3$, il sistema ottenuto per retroazione è il sistema S_3 che ha ingresso f e uscita g ottenuta ponendo in ingresso a S_1 la differenza tra f e la risposta di S_2 a g .



9.2 Tempo-invarianza

Un sistema complesso lineare può sempre essere scomposto in *componenti semplici*, e questo principio può essere applicato anche sugli input (più sistemi) con lo stesso operatore sui segnali. La relazione I/O soddisfa il principio di **sovrapposizione degli effetti**.

Dato il segnale in ingresso $x(n) = \alpha_1 x_1(n) + \alpha_2 x_2(n)$, la risposta a una combinazione lineare di ingressi è la combinazione lineare delle risposte del sistema ai singoli ingressi.

$$L[x(n)] = L[\alpha_1 x_1(n) + \alpha_2 x_2(n)] = \alpha_1 L[x_1(n)] + \alpha_2 L[x_2(n)]$$

Un'altra proprietà dei sistemi è la **tempo-invarianza**: sono stazionari e rispondono nello stesso modo in istanti differenti, quindi l'output è lo stesso anche con ritardo.

La relazione I/O produce un segnale di uscita $y(n)$ che dipende solo dalla forma del segnale di ingresso $x(n)$:

$$L[x(n - n_0)] = y(n - n_0)$$

Se l'ingresso al sistema è ritardato o anticipato di una quantità n_0 , allora anche la risposta viene ritardata o anticipata di n_0 .

Sistemi **causali**: presupponendo il campione attuale e i successivi, la risposta corrente $x(n)$ non dipende dai valori futuri $x(n + n_0)$ con costante intera $n_0 > 0$ (solo parte positiva).

Sistemi **anticausali**: considerando il campione corrente e i precedenti, la risposta corrente $y(n)$ non dipende dai valori passati $x(n - n_0)$ con costante intera $n_0 > 0$ (solo parte positiva). Non sono fisicamente realizzabili.

I sistemi **bilateri** hanno sia parte causale che anticausale.

I sistemi possono avere **memoria**, la cui grandezza è il numero di campioni che vengono immagazzinati per calcolare i risultati successivi. La risposta corrente in questo caso dipende dai valori dell'ingresso negli istanti di tempo precedenti a quello corrente n (altrimenti sono statici).

Un sistema a tempo discreto si dice *passivo* quando l'energia E_y dell'output $y(n)$ è minore o uguale all'energia E_x dell'input $x(n)$. Questo si rappresenta con il modulo quadro, osservando la conservazione.

$$\sum_{n=-\infty}^{+\infty} |y(n)|^2 \leq \sum_{n=-\infty}^{+\infty} |x(n)|^2 < \infty$$

Se la relazione precedente vale con il segno di uguaglianza, allora il sistema è detto *senza perdite*, in quanto conserva l'energia del segnale di ingresso.

9.3 Sistemi LTI

I **sistemi lineari tempo invariante** (LTI) sono definiti da operazioni I/O che soddisfano contemporaneamente la **linearità** e la **stazionarietà**. Se i coefficienti sono costanti, il segnale è stazionario perché non cambia in funzione al peso.

L'analisi dei LTI può essere condotta in tre diversi modi:

1. *Relazione I/O* (qualsiasi sistema);
2. *Equazione lineare alle differenze a coefficienti costanti* (LTI causale);
3. *Risposta all'impulso* nel dominio del tempo e delle frequenze (sistema LP).

In particolare per il caso di equazione alle differenze, l'output è espresso come combinazione lineare di tutti i valori (coefficienti reali, costanti positive o negative) della sequenza di input $x(n)$, che è combinazione lineare della stessa sequenza di output, a partire dal primo campione ritardato.

$$y(n) = -\alpha_1 y(n-1) - \alpha_2 y(n-2) - \dots - \alpha_M y(n-M) + \beta_0 x(n) + \beta_1 x(n-1) + \dots + \beta_N x(n-N)$$

L'uscita $y(n)$ dipende sia dai valori che l'ingresso $x(n)$ assume in un arco temporale fino a N istanti precedenti all'istante di tempo discreto n , sia dai valori assunti dal segnale di uscita $y(n)$ fino a M istanti di tempo precedenti a quello corrente n .

Queste operazioni possono essere raccolte come sommatoria con coefficienti costanti indipendenti dal tempo:

$$y(n) = \sum_{k=0}^N \beta_k x(n-k) - \sum_{j=1}^M \alpha_j y(n-j)$$

La y non parte da zero perchè altrimenti $y(n)$ dovrebbe dipendere da se stessa. Se l'equazione è ricorsiva (almeno un coefficiente α_j diverso da zero), la memoria è infinita.

Un'altra modalità per descrivere un sistema lineare a tempo invariante è la risposta all'impulso: la relazione I/O è semplice perchè si indica la reazione alla δ , dove il peso è il valore della sequenza in quel campione. Essendo lineare a tempo invariante, i segnali sono indipendenti dal tempo e esprimibili come combinazioni lineari.

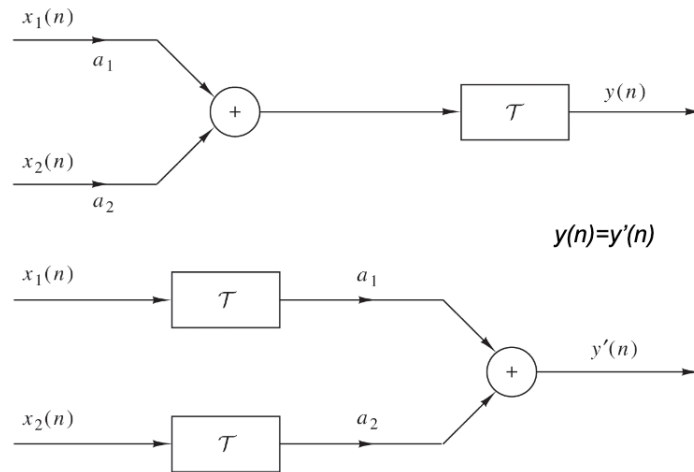
L'analisi della risposta del sistema all'impulso permette di valutare la risposta a regime, al fine di semplificarne l'analisi tempo-frequenza rappresentando la sequenza come somma di δ traslate e pesate.

$$x(n) = \sum_{i=-\infty}^{+\infty} x(i)\delta(n-i)$$

Per la proprietà di sovrapposizione degli effetti (linearità) e di stazionarietà, si ha:

$$y(n) = L\left[\sum_{i=-\infty}^{\infty} x(i)\delta(n-i)\right] = \sum_{i=-\infty}^{\infty} x(i)L[\delta(n-i)] = \sum_{i=-\infty}^{\infty} x(i)h(n-i) = x * h$$

h è il kernel, la sequenza che descrive il comportamento del filtro (risposta del sistema all'impulso). La formula si può ritrovare come **convoluzione** lineare discreta della sequenza con il filtro.

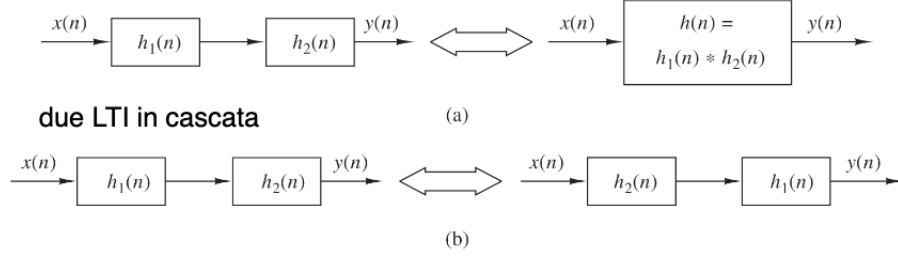


Se i segnali sono causali e limitati, il numero di valori sarà finito. La risposta all'impulso $h(n)$ è nulla per istanti di tempi $n < 0$, e l'uscita dipende dal valore corrente del segnale d'ingresso $x(n)$ e dei campioni già presenti nel sistema.

Se la risposta all'impulso $h(n)$ ha supporto temporale finito N_h , l'uscita del sistema LTI è:

$$y(n) = x(n) * h(n) = \sum_{i=0}^{N_h-1} h(i)x(n-i)$$

L'operatore di convoluzione è commutativo e associativo, quindi applicabile ai sistemi lineari a tempo invariante a cascata ottenendo lo stesso risultato.



Per la proprietà distributiva, due LTI $h_1(n)$ e $h_2(n)$ connessi in parallelo possono essere sostituiti da un singolo sistema $h(n) = h_1(n) + h_2(n)$.

L'interazione del segnale nel sistema è il prodotto delle trasformate di Fourier, semplificando la convoluzione. H è la risposta dell'impulso, che interagisce con X per generare Y .

Per il teorema della convoluzione,

$$y(n) = x(n) * h(n) \leftrightarrow Y(f) = X(f)H(f) \quad \text{dove} \quad H(f) = \frac{1}{M} \sum_{n=0}^{M-1} h(n)e^{-j\frac{2\pi}{M}fn}$$

$$y(n) = x(n) * h(n) \leftrightarrow Y(e^{j\omega}) = X(e^{j\omega})H(e^{j\omega})$$

La funzione continua $H(e^{j\omega})$ (trasformata discreta di $h(n)$) viene detta **risposta in frequenza** del sistema LTI, e può essere definita come:

$$H(e^{j\omega}) = \frac{Y(e^{j\omega})}{X(e^{j\omega})} = \Re(H(e^{j\omega})) + j\Im(H(e^{j\omega})) = |H(e^{j\omega})| e^{j\phi(H(e^{j\omega}))}$$

9.4 Filtri

Dato un generico sistema con una funzione $H(e^{j\omega})$, un **filtro** taglia alcune componenti in frequenza del segnale di ingresso lasciandone passare altre, secondo quanto specificato attraverso la risposta.

Le frequenze nell'intervallo vengono moltiplicate per 1, le altre per 0. Ci sono diverse tipologie di filtri ideali: passa-basso, passa-alto, passa-banda o attenua-banda, che corrispondono alla funzione *sinc* che è anticausale, pertanto non rappresentabile.

$$\text{Filtro ideale passa-basso } |H(e^{j\omega})| = \begin{cases} 1, & |\omega| \leq \omega_t \\ 0, & \omega_t < |\omega| \leq \pi \end{cases}$$

$$\text{Filtro ideale passa-alto } |H(e^{j\omega})| = \begin{cases} 0, & |\omega| \leq \omega_t \\ 1, & \omega_t < |\omega| \leq \pi \end{cases}$$

$$\text{Filtro ideale passa-banda } |H(e^{j\omega})| = \begin{cases} 0, & |\omega| \leq \omega_1 \\ 1, & \omega_1 < |\omega| \leq \omega_2 \\ 0, & \omega_2 < |\omega| \leq \pi \end{cases}$$

$$\text{Filtro ideale attenua-banda } |H(e^{j\omega})| = \begin{cases} 0, & |\omega| \leq \omega_1 \\ 1, & \omega_1 < |\omega| \leq \omega_2 \\ 0, & \omega_2 < |\omega| \leq \pi \end{cases}$$

I sistemi LTI a tempo discreto devono soddisfare alcune condizioni specifiche, al fine di poter essere impiegati in alcune applicazioni. Una di queste condizioni è la stabilità, di cui la più usata nella pratica è detta **BIBO** (Bounded-Input Bounded-Output).

Un sistema rispetta la stabilità BIBO se, per ogni ingresso $h(n)$ con energia limitata, anche l'output ha ampiezza limitata (non va a infinito, è sommabile in modulo). *sinc* tende a infinito, quindi non è stabile.

$$h_s = \sum_{-\infty}^{+\infty} |h(n)| < \infty$$

Un sistema LTI è fisicamente realizzabile se possiede una risposta all'impulso $h(n)$ reale e causale, dove la causalità implica $h(n) = 0 \ \forall \ n < 0$.

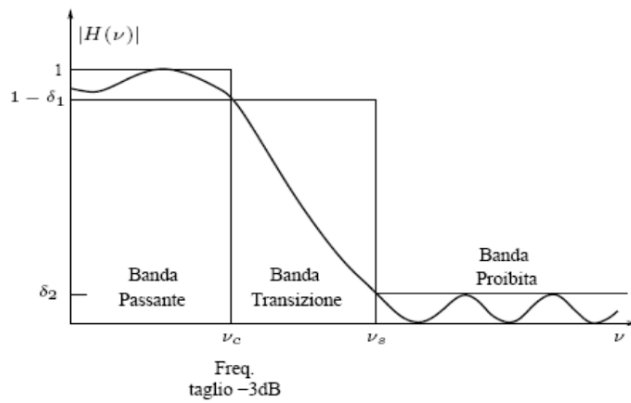
La funzione di trasferimento H nel dominio delle frequenze di un filtro ideale passa-basso possiede le seguenti caratteristiche:

- $|H|$ costante nella banda passante e nullo nella banda proibita;
- La banda passante e la banda proibita sono confinanti, separate dalla *frequenza di taglio*.

Un filtro reale, in confronto a uno ideale, non ha zone ben precise di banda passante e proibita: la transizione non è immediata (esiste una zona di transizione) e le frequenze non hanno peso uguale bensì fluttuazioni con ampiezza variabile.

Si rilevano oscillazioni non trascurabili ampie δ_1 . L'attenuazione si calcola come $20 \log_{10} \delta_2 \text{ dB}$, con δ_2 ampiezza della massima oscillazione in banda proibita.

La banda passante e la banda proibita sono separate appunto dalla **banda di transizione**, che inizia dalla frequenza di taglio v_c e termina alla frequenza di stop v_s (dimensione $v_s - v_c$). Questi parametri vengono dati come specifiche dei filtri.



La frequenza di taglio è la frequenza per la quale la potenza è il 50% rispetto al massimo, quindi c'è un'attenuazione della metà. Serve per identificare la fine della banda passante. Se $|H(v)|^2 \approx 1$, v_c è la frequenza per cui:

$$|H(v_c)|^2 = \frac{1}{2}$$

v_c è anche detta frequenza di taglio a 3 dB poichè $10 \log_{10}(1/2) \approx -3 \text{ dB}$.

10 Equazioni alle differenze

I LTI causali a tempo discreto possono essere rappresentati in modi differenti, tra cui la combinazione lineare con coefficienti costanti nel tempo (invariante).

Il comportamento di un sistema LTI a tempo discreto e causale può essere descritto anche da **equazioni alle differenze**, come visto precedentemente:

$$y(n) = -\alpha_1 y(n-1) - \alpha_2 y(n-2) - \dots - \alpha_M y(n-M) + \beta_0 x(n) + \beta_1 x(n-1) + \dots + \beta_N x(n-N)$$

$$y(n) = \sum_{k=0}^N b_k x(n-k) - \sum_{j=1}^M a_j y(n-j)$$

Si ricorda che i coefficienti sono costanti e indipendenti dal tempo, per assicurare la stazionarietà. Un'equazione è detta **ricorsiva** se almeno un coefficiente a_i è diverso da zero.

Quando il sistema ha una parte ricorsiva, dipende di fatto da più istanti precedenti dell'input, e non solo dai N istanti evidenti dalla sommatoria. M è l'ordine dell'equazione alle differenze (ordine del sistema).

Il ritardo su y indica di che **ordine** è il sistema: il primo ordine, per esempio, ha ricorsione con un campione in ritardo.

L'output di un sistema ricorsivo dev'essere calcolato in sequenza, al contrario di un sistema non ricorsivo in cui i campioni possono essere valutati con qualsiasi disposizione.

Esempi di sistemi:

- Identità $y(n) = x(n)$, non ricorsivo e causale;
- Ritardo (anticipo) unitario $y(n) = x(n \pm 1)$, non ricorsivo e (non) causale;
- Media mobile $y(n) = \frac{1}{3}[x(n-1) + x(n) + x(n+1)]$, non ricorsiva e non causale;
- Accumulatore $y(n) = \sum_{-\infty}^n x(k)$, causale ricorsivo con memoria infinita.

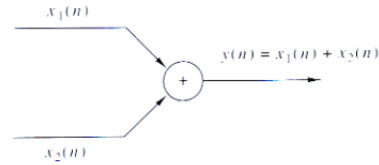
In quest'ultimo caso, l'output a un istante n dipende dal valore dell'input per tutti gli istanti precedenti, cioè dall'output ricavato con essi: bisogna conoscere anche le condizioni iniziali $y(n-1)$.

Se il sistema accumulatore $y(n) = y(n-1) + x(n)$ non ha subito alcuna eccitazione prima dell'istante n , $y(n-1) = 0$ e il sistema è scarico (per $n = -\infty$).

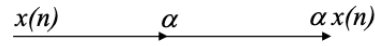
10.1 Diagrammi a blocchi

Ogni operazione di cui è composto un sistema viene rappresentata in un diagramma a blocchi, che contiene eventualmente anche una parte ricorsiva (freccia all'indietro). Si ricordano la linearità e la commutatività.

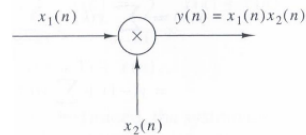
Sommatore: $y(n) = x_1(n) + x_2(n)$



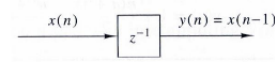
Moltiplicatore scalare: $y(n) = \alpha x(n)$



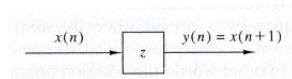
Moltiplicatore sequenze: $y(n) = x_1(n) \times x_2(n)$



Ritardo di un elemento: $y(n) = x(n-1)$



Anticipo di un elemento: $y(n) = x(n+1)$



10.1.1 Media cumulativa

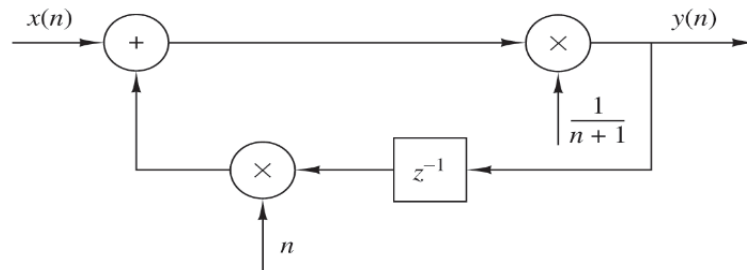
La media cumulativa è un sistema ricorsivo non stazionario, con normalizzazione effettuata in base a un parametro non costante.

$$y(n) = \frac{1}{n+1} \sum_{k=0}^n x(k)$$

Il calcolo di $y(n)$ richiede la conoscenza di tutti i valori di $x(n)$ precedenti, ma può essere semplicemente calcolato da $y(n-1)$, cioè il precedente output e l'input corrente: il valore di $y(n_0)$ si calcola a partire da una sequenza di input applicata per $n \geq n_0$ e dalla condizione iniziale $y(n_0 - 1)$.

$$(n+1)y(n) = \sum_{k=0}^n x(k) + x(n) = ny(n-1) + x(n)$$

$$y(n) = \frac{n}{n+1} y(n-1) + \frac{1}{n+1} x(n)$$

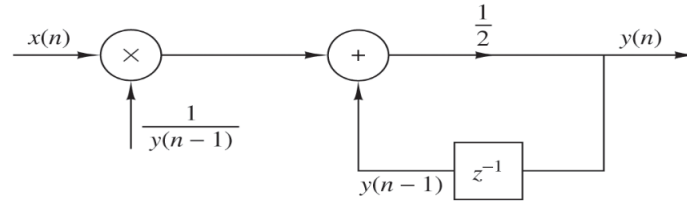


10.1.2 Radice quadrata

Questo sistema calcola la radice quadrata in modo ricorsivo, a partire da una condizione iniziale $y(-1)$ che approssima il valore cercato. Al crescere di n , la stima del valore migliora.

$$y(n) = 0.5 \frac{x(n)}{y(n-1)} + 0.5 y(n-1)$$

La radice quadrata non è un sistema lineare: non esistono sommatorie, ma il termine y compare al denominatore quindi non è rappresentabile come combinazione lineare.



11 Risposta all'impulso

La **risposta all'impulso** viene descritta attraverso un prodotto di convoluzione, come sequenza finita o infinita. I primi casi non hanno la ricorsione, sono campioni combinati con pesi: i valori di h sono i coefficienti di x .

In base alla risposta all'impulso, i sistemi possono essere divisi in due gruppi: **FIR** (*Finite Impulse Response*) e **IIR** (*Infinite Impulse Response*).

11.1 FIR

$$h(n) = \begin{cases} \neq 0 & 0 \leq n \leq M-1 \\ = 0 & n < 0 \wedge n \geq M \end{cases}$$

La risposta di un sistema FIR a un generico segnale è allora:

$$y(n) = x(n) * h(n) = \sum_{i=0}^{\infty} h(i)x(n-i) = \sum_{i=0}^{M-1} h(i)x(n-i)$$

Per un sistema FIR, l'output a qualsiasi istante n è semplicemente la somma di una combinazione lineare pesata degli M valori più recenti della sequenza di input.

Il sistema agisce come una finestra che vede solo M valori recenti, quindi ha memoria M .

I sistemi FIR possono essere sempre realizzati in modo **non ricorsivo**, essendo in funzione di M campioni pesati per la risposta all'impulso $h(n)$. Questo deriva dal fatto che una LTI non ricorsiva dipende solo da un numero finito di istanti nel passato.

11.2 IIR

$$h(n) = \begin{cases} \neq 0 & n \geq n_0 \\ = 0 & n < n_0 \end{cases}$$

La risposta di un sistema IIR causale a un generico segnale è allora:

$$y(n) = x(n) * h(n) = \sum_{i=0}^{\infty} h(i)x(n-i)$$

L'output a qualsiasi istante n dipende da tutti i valori della sequenza di input: il sistema ha memoria infinita e **non è realizzabile**, perché richiederebbe infinite locazioni di memoria.

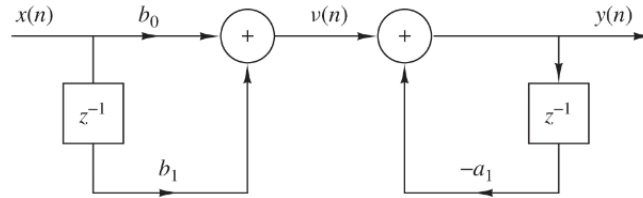
Se la sommatoria va a infinito, non c'è una diretta corrispondenza con il dominio delle frequenze, quindi non si individua $y(n)$ guardando la risposta ma si utilizzano le equazioni alle differenze e il dominio trasformato.

11.2.1 Prima forma diretta

Realizza un'equazione lineare alle differenze del primo ordine a coefficienti costanti, utilizzando due unità di ritardo (memoria) distinte come se fossero due sistemi a cascata. Il primo sistema non è ricorsivo, mentre il secondo lo è.

$$y(n) = -\alpha_1 y(n-1) + \beta_0 x(n) + \beta_1 x(n-1)$$

Per la proprietà commutativa, scambiando l'ordine dei sistemi il risultato non cambia, ma è necessario introdurre una variabile ausiliaria.



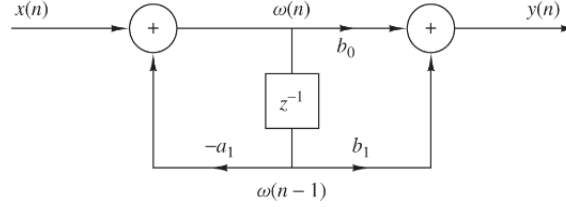
11.2.2 Seconda forma diretta

$$\omega(n) = -\alpha_1 \omega(n-1) + x(n) \quad y(n) = \beta_0 \omega(n) + \beta_1 \omega(n-1)$$

Le due unità di ritardo hanno entrambe in input $\omega(n)$ e in output $\omega(n-1)$, e possono quindi essere rappresentati con un'unica unità di ritardo. Ciò è efficiente nelle applicazioni pratiche in termini di memoria. Il passaggio dalla prima forma diretta alla seconda si può estendere al caso di una generica equazione alle differenze di ordine N :

$$y(n) = \sum_{k=0}^M \beta_k x(n-k) - \sum_{j=1}^N \alpha_j y(n-j)$$

Ciò richiede $N + M$ unità di ritardo e $M + N + 1$ moltiplicazioni. La seconda forma diretta è data dalla cascata di un sistema ricorsivo seguita da uno non ricorsivo, con $\max(N, M)$ ritardi.



12 Trasformata zeta

La **trasformata zeta** aiuta l'analisi di stabilità e causalità delle frequenze dei sistemi lineari a tempo invariante con i relativi filtri, essendo più ampia rispetto alla trasformata di Fourier.

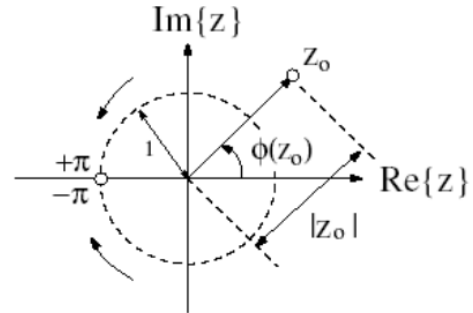
La trasformata discreta non è sufficiente in alcune situazioni perché non sempre esiste, descrive solo comportamenti di sistemi LTI scarichi e presume condizioni iniziali nulle.

La trasformata zeta è la **parte discreta** (digitale) della *trasformata di Laplace*, ed è rappresentata da una sommatoria con l'estensione delle frequenze nel mondo complesso, esprimibile in coordinate polari con un termine ρ qualsiasi che determina la distanza dall'origine. L'operatore è lineare.

Data una frequenza bilatera $x(n)$ con $-\infty < n < \infty$ si definisce trasformata zeta:

$$X(z) = Z[x(n)] = \sum_{-\infty}^{\infty} x(n)z^{-n}$$

Esiste una corrispondenza biunivoca tra $x(z)$ e $x(n)$ e i loro domini solo se viene definita la **regione di convergenza uniforme** ROC, cioè i valori della variabile complessa z tale che $x(z)$ è finita.



Nella regione di convergenza, $X(z)$ è una funzione analitica, cioè continua e indefinitamente derivabile con derivate continue in z .

La z è una pulsazione complessa con dominio \mathcal{C} , rappresentabile in termini di modulo e fase come $z = \rho e^{j2\pi f} = \rho e^{j\omega}$. Si può ridurre a trasformata di Fourier discreta togliendo il coefficiente complesso e applicando la formula a una nuova sequenza.

Deve valere la condizione di sommabilità in modulo, quindi la convergenza varia con la distanza e il piano si può dividere in circonferenze concentriche con comportamento diverso.

La regione ROC dipende solo dal modulo ρ delle pulsazioni complesse z , e non dalla loro fase: questo comporta le circonferenze come luoghi dei punti z a modulo costante.

$$\sum_{-\infty}^{\infty} |x(n)\rho^{-n}| < \infty \implies \sum_{-\infty}^{\infty} |x(n)| \rho^{-n} < \infty$$

In particolare, è importante trovare i valori di ρ che indicano la convergenza, e si ha che i valori sulla **circonferenza di raggio unitario** $\rho = 1$ sono i valori della trasformata di Fourier di frequenza 0. Quando c'è un ritardo, la regione sarà individuata con $[z \neq 0]$.

12.1 Trasformate e ROC

La trasformata zeta definisce una relazione biunivoca tra la sequenza $x(n)$ e una funzione della variabile complessa z . La biunivocità è garantita solo se si specifica la ROC di $X(z)$ e l'espressione analitica della trasformata.

Esistono relazioni specifiche tra la tipologia di una generica sequenza $x(n)$ e la ROC della relativa trasformata z :

- Le sequenze $x(n)$ *bilatere* hanno supporto su istanti di tempo discreto **negativi e positivi**;
- Le sequenze *causali* possiedono coefficienti identicamente nulli per istanti di tempo discreto **negativi**;
- Le sequenze *anticausali* possiedono coefficienti identicamente nulli per istanti di tempo discreto **positivi**.

12.1.1 Impulso δ

$$X(z) = \sum_{-\infty}^{\infty} \delta(n) z^{-n} = 1$$

Se l'impulso $x(n) = \delta(n)$ è una costante nel dominio trasformato, applicando la funzione zeta si ha un risultato simile: tutti i valori assumono 0 tranne quello per cui $n = 0$, quindi l'esponentiale diventa 1 così come tutta la trasformata.

La regione di convergenza è costante e limitata a 1, e rappresenta tutto il piano complesso per cui la serie geometrica converge.

$$\begin{aligned} x(n) = 2\delta(n+1) + \delta(n) + 4\delta(n-2) &= 2 \sum_{-\infty}^{\infty} 2\delta(n+1) z^{-n} + \sum_{-\infty}^{\infty} 2\delta(n) z^{-n} + 4 \sum_{-\infty}^{\infty} 2\delta(n-2) z^{-n} \\ &\implies 2z + 1 + 4z^{-2} \end{aligned}$$

Il risultato è ottenuto scomponendo le sommatorie e trovando i valori per cui il numero all'interno delle parentesi si annulla. Bisogna sempre tenere conto del segno negativo dell'esponente.

12.1.2 Gradino

$$X(z) = \sum_{-\infty}^{\infty} u(n) z^{-n} = \sum_0^{\infty} z^{-n} = \sum_0^{\infty} (z^{-1})^n = \frac{1}{1 - z^{-1}}$$

In una sequenza a gradino, la ROC corrisponde all'esterno della circonferenza di raggio unitario, cioè $[|z^{-1}| < 1 \geq |z| > 1]$. La stessa formula vale per il gradino anticausale, e la corrispondenza biunivoca si ritrova perché la ROC cambia in $|z| < 1$.

12.2 ROC per sequenze finite

Le sequenze $x(n)$ con supporto finito sono i segnali a tempo discreto che possiedono un numero finito di coefficienti non nulli. Queste ammettono sempre una trasformata $X(z)$ esprimibile in forma chiusa come un polinomio composto da un numero finito di variabili del tipo z^k , z^{-k} con k intero.

$$X(z) = \sum_{n=-M}^L x(n)z^{-n}$$

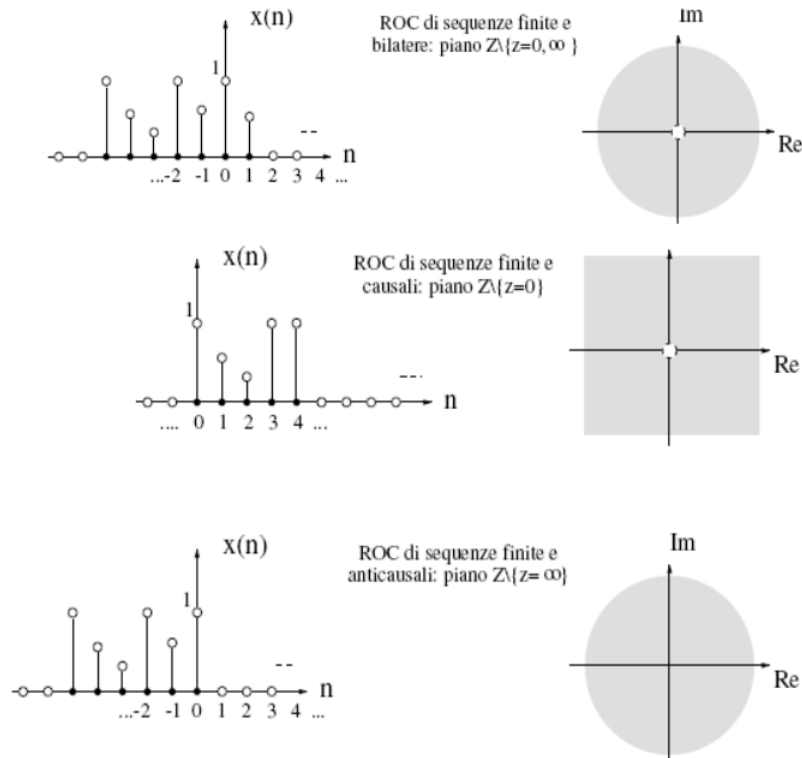
La trasformata converge per **qualsunque** z nel **piano complesso**, eccetto il punto $z = 0$ se esistono termini del tipo z^{-k} con $k > 0$, e il punto $z = \infty$ se esistono termini del tipo z^{+k} con $k > 0$.

Essendoci polinomi, sia il numeratore che il denominatore potrebbero annullarsi: nel primo caso si chiamano **zeri**, altrimenti **poli**. Per individuare i valori si può raccogliere i fattori arrivando all'ordine 1 nel denominatore.

ROC per segnali di durata finita:

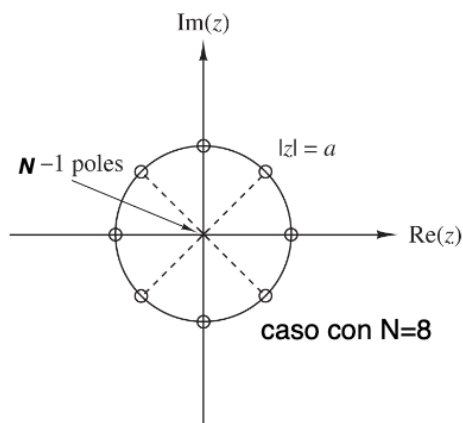
- ROC per segnali bilateri di durata finita: intero piano complesso \mathbb{C} tranne $[z = 0 \wedge z = \infty]$;
- ROC per sequenze finite e causali: intero piano complesso \mathbb{C} tranne $[z = 0]$;
- ROC per sequenze finite e anticausali: intero piano complesso \mathbb{C} tranne $[z = \infty]$.

In questo caso i campioni infiniti vengono rappresentati come rapporto finito di polinomi infiniti.



12.3 Poli e zeri

I **poli** sono le *radici del denominatore*, e gli **zeri** sono le *radici del numeratore* (possono trovarsi anche nelle pulsazioni complesse). Il loro numero è uguale.



Il numero di poli e quello degli zeri corrisponde all'ordine del polinomio che li origina (quindi inizialmente possono essere diversi, ma vengono poi bilanciati con poli/zeri aggiuntivi nell'origine).

$$X(z) = \frac{1 - \alpha^N z^{-N}}{1 - \alpha z^{-1}} = z^{-N+1} \frac{z^N - \alpha^N}{z - \alpha}$$

C'è una parte in comune $z = \alpha$ che annulla sia denominatore che numeratore, quindi gli zeri corrispondenti si semplificano e la sequenza resta finita. Ci sono $N - 1$ poli (e zeri) per $z = 0$.

Le radici si posizionano in modo uniforme a intervalli distanti uguali, nelle pulsazioni complesse, su una circonferenza di raggio α (polinomio di ordine N , $z^N - \alpha^N$).

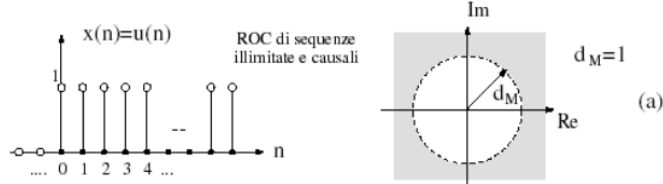
12.4 ROC per sequenze infinite

Le sequenze $x(n)$ illimitate sono i segnali a tempo discreto che possiedono un supporto temporale illimitato, ma convergono solo in una distinta parte del piano complesso.

Regione di convergenza per sequenze causali:

$|z| > d_M$, dove d_M è il modulo del polo più distante dall'origine.

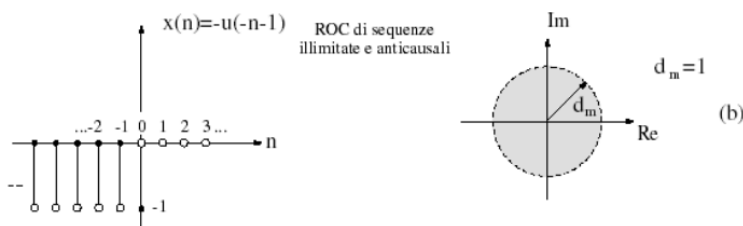
Esempio: sequenza gradino con ROC $|z| > 1$.



Regione di convergenza per sequenze anticausali:

$|z| < d_m$, dove d_m è il modulo del polo più vicino all'origine.

Esempio: sequenza gradino con ROC $|z| < 1$.



Per definizione, la **regione di convergenza non può contenere poli**.

Le sequenze causali, quindi, hanno poli all'esterno della circonferenza, mentre le sequenze anticausali all'interno.

12.5 Trasformata zeta razionale

Nei casi pratici di interesse, la funzione $X(z)$ è una funzione **razionale** di due polinomi: $X(z) = \frac{N(z)}{D(z)}$. $N(z)$ e $D(z)$ sono due polinomi nella variabile p_n e p_d . Scrivendo i due polinomi in forma estesa, si ottiene:

$$X(z) = \frac{\beta_0 + \beta_1 z^{-1} + \dots + \beta_{p_n}}{z^{-p_n}} \alpha_0 + \alpha_1 z^{-1} + \dots + \alpha_{p_d} z^{-p_d}$$

Esprimendo in termini di potenze positive di z e fattorizzando i due polinomi in base alle rispettive radici elementari, si ha:

$$X(z) = \frac{\beta_0}{\alpha_0} (z^{p_d - p_n}) \frac{\prod_{i=1}^{p_n} (z - c_i)}{\prod_{i=1}^{p_d} (z - d_i)}$$

I coefficienti $c_i \forall i = 1, \dots, p_n$ sono gli zeri non nulli di $X(z)$, mentre i coefficienti $d_i \forall i = 1, \dots, p_d$ sono i poli non nulli di $X(z)$.

Ci sono $p_n - p_d$ poli in $z = 0$ nel caso in cui $p_n > p_d$.

Ci sono $p_d - p_n$ poli in $z = 0$ nel caso in cui $p_d > p_n$.

13 Analisi dei sistemi

Come visto precedentemente, i sistemi LTI a tempo discreto possono essere descritti da equazioni lineari alle differenze a coefficienti costanti:

$$y(n) = -\alpha_1 y(n-1) - \alpha_2 y(n-2) - \dots - \alpha_M y(n-M) + \beta_0 x(n) + \beta_1 x(n-1) + \dots + \beta_N x(n-N)$$

Applicando la trasformata di Fourier discreta a ogni termine si ottiene:

$$Y(e^{j\omega}) = -\alpha_1 Y(e^{j\omega}) e^{-j\omega} - \dots - \alpha_M Y(e^{j\omega}) e^{-j\omega M} + \\ + \beta_0 X(e^{j\omega}) + \beta_1 X(e^{j\omega}) e^{-j\omega} + \dots + \beta_N X(e^{j\omega}) e^{-j\omega N}$$

Si ricorda che $F(f(x - x_0)) = e^{-j2\pi u x_0} F(u)$. Raccogliendo tutti i termini $Y(e^{j\omega})$ e $X(e^{j\omega})$, si ha:

$$Y(e^{j\omega}) (1 + \alpha_1 e^{-j\omega} + \dots + \alpha_M e^{-j\omega M}) = X(e^{j\omega}) (\beta_0 + \beta_1 e^{-j\omega} + \dots + \beta_N e^{-j\omega N})$$

Per il teorema della convoluzione, $y(n) = x(n) * h(n) \leftrightarrow Y(e^{j\omega}) = X(e^{j\omega}) H(e^{j\omega})$.

La risposta in frequenza $H(e^{j\omega})$ di un sistema LTI espresso tramite le equazioni alla differenze, può essere calcolata come segue:

$$H(e^{j\omega}) = \frac{Y(e^{j\omega})}{X(e^{j\omega})} = \frac{(\beta_0 + \beta_1 e^{-j\omega} + \dots + \beta_N e^{-j\omega N})}{(1 + \alpha_1 e^{-j\omega} + \dots + \alpha_M e^{-j\omega M})}$$

Questo per la relazione che lega trasformata di Fourier e zeta:

$$\sum_{n=-\infty}^{\infty} x(n) z^{-n} = \sum_{n=-\infty}^{\infty} (x(n) \rho^{-n}) e^{-j\omega n}.$$

La regione di convergenza della funzione $Y(z)$ coincide, nel caso generale, quando non avvengono cancellazioni tra poli e zeri in $X(z)$ e $H(z)$, con l'intersezione tra le regioni delle due funzioni.

Essendo la funzione di trasferimento $H(e^{j\omega})$ razionale, allora anche $H(z)$ lo è, quindi si può esprimere con una funzione razionale del tipo:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{(\beta_0 + \beta_1 z^{-1} + \dots + \beta_N z^{-N})}{(1 + \alpha_1 z^{-1} + \dots + \alpha_M z^{-M})}$$

Essa corrisponde all'equazione alle differenze, e al caso generale di un sistema LTI.

Un sistema LTI causale e non ricorsivo di tipo FIR ha una funzione di trasferimento del tipo:

$$H(z) = \beta_0 + \beta_1 z^{-1} + \dots + \beta_N z^{-N}.$$

Un sistema LTI causale e puramente ricorsivo di tipo IIR ha una funzione di trasferimento del tipo:

$$H(z) = \frac{1}{(1 + \alpha_1 z^{-1} + \dots + \alpha_M z^{-M})}.$$

14 Sistemi LTI e trasformata zeta

La funzione di trasferimento $H(z)$ può essere espressa in termini delle **radici** dei polinomi al numeratore e al denominatore fattorizzando i due polinomi $N(z)$ e $D(z)$:

$$H(z) = \frac{N(z)}{D(z)} = K z^{M-N} \frac{(z - c_1)(z - c_2) \dots (z - c_N)}{(z - d_1)(z - d_2) \dots (z - d_N)}$$

La risposta di un sistema LTI viene analizzata attraverso i poli e gli zeri della funzione di trasferimento $H(z)$.

I coefficienti nell'equazione alle differenze corrispondono ai coefficienti nella trasformata, considerando il cambiamento di segno. Al numeratore c'è la parte non ricorsiva, e al denominatore quella ricorsiva. Se non c'è il denominatore, la sequenza è FIR (risposta all'impulso finita).

I sistemi FIR hanno risposta d'impulso limitata, e si può immediatamente passare alla rappresentazione tramite equazione alle differenze dato che i coefficienti sono gli stessi. I sistemi non sono ricorsivi, dato che i FIR possono essere sempre realizzati con sistemi non ricorsivi.

Essi sono in funzione di soli M campioni della sequenza di ingresso, pesati per la risposta all'impulso $h(n)$: dipendono solo da un numero finito di elementi del passato. Dato $h(n)$, si ha subito la trasformata zeta.

$$\{1, 2, 0, 4\} \quad h(n) = \sum_{j=0}^{N-1} \beta_j \delta(n-j) = \delta(n) + 2\delta(n-1) + 4\delta(n-3)$$

$$y(n) = \sum_{k=-\infty}^{+\infty} h(k)x(n-k) = \sum_{k=0}^3 h(k)x(n-k)$$

$$y(n) = x(n) + 2x(n-1) + 4x(n-3) \quad H(z) = 1 + 2z^{-1} + 4z^{-3}$$

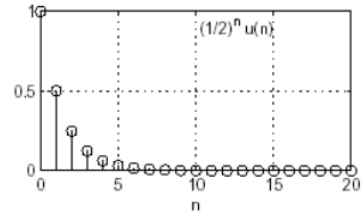
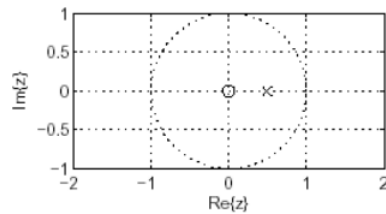
Studiando i poli si analizza la regione in cui la funzione viene amplificata e cresce in valore, gli zeri dove diminuisce. Insieme quindi danno l'andamento delle frequenze. Un paio di poli complessi e coniugati corrisponde a segnali causali con comportamento a oscillazione.

La frequenza è legata alla **posizione angolare**, quindi se i poli si trovano nello zero la loro influenza si riflette su tutte le frequenze o nessuna: il contributo è nullo.

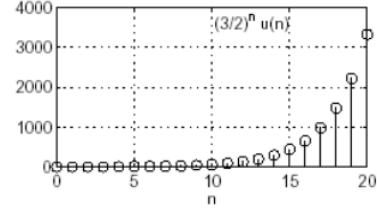
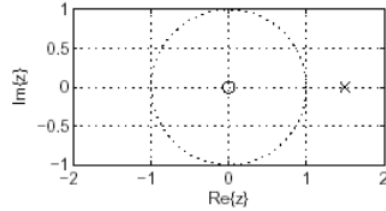
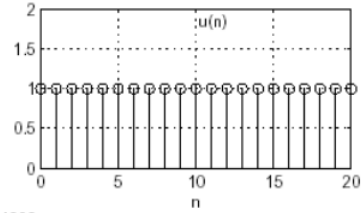
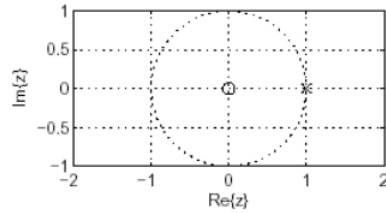
Il sistema è **stabile** se la trasformata di Fourier è interna alla regione di convergenza.

$$X(z) = \frac{1}{1 - \alpha z^{-1}} = z \frac{1}{z - \alpha} \quad \text{con ROC} \quad |\alpha z^{-1}| < 1 \rightarrow |z| > |\alpha|$$

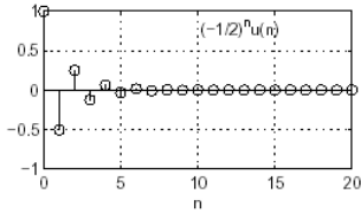
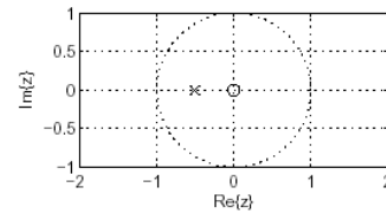
$\alpha > 0$



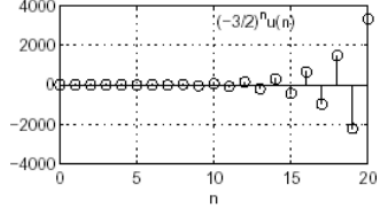
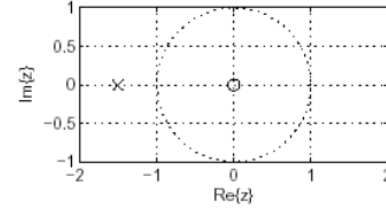
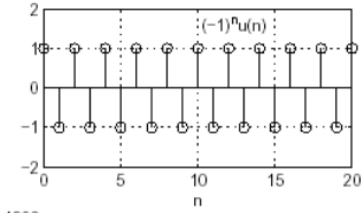
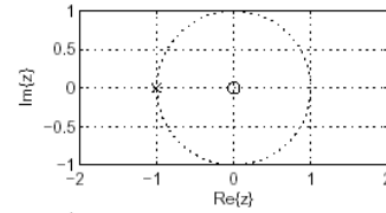
1 polo in α
1 zero in 0



$\alpha < 0$



1 polo in α
1 zero in 0



14.1 Stabilità BIBO

La **stabilità BIBO** di un sistema LTI impone che la risposta all'impulso $h(n)$ sia sommabile in modulo: $\sum_{n=-\infty}^{\infty} |h(n)| < \infty$. Se il sistema è causale, la condizione si traduce nel dominio z imponendo che la funzione di trasferimento $H(z)$ abbia *poli contenuti nella circonferenza di raggio unitario* del piano z .

Le caratteristiche della sequenza dipendono dalla posizione dei poli: la convergenza, e conseguentemente la stabilità, sono determinate dall'appartenenza al cerchio unitario.

I poli all'esterno del cerchio $|z| = 1$ conducono a una risposta all'impulso $h(n)$ con crescita esponenziale, e l'uscita del sistema LTI potrebbe divergere anche in presenza di segnali $x(n)$ d'ingresso costanti. Poli multipli sul cerchio unitario conducono a una crescita di tipo polinomiale.

Poli semplici su $|z| = 1$ possono condurre a una risposta divergente, e dunque a un sistema LTI **non stabile** secondo la definizione BIBO.

$H(z) = \frac{z}{z-3}$ ha un polo in $z = 3$, infatti ha $h(n) = 3^n u(n)$.

$H(z) = \frac{1}{z(z-1)^2}$ ha un polo di ordine 2 in $z = 1$, e risposta all'impulso $h(n) = u(n)$ di tipo rampa, quindi è instabile.

$H(z) = \frac{z}{(z-1)}$ ha un polo semplice in $z = 1$, e risposta all'impulso $h(n) = u(n)$ costante (con modulo che tende a infinito).

La circonferenza di raggio unitario dev'essere sempre contenuta nella regione di convergenza per garantire la stabilità del sistema:

- Se il sistema è causale, condizione necessaria e sufficiente per la stabilità BIBO è che la funzione di trasferimento $H(z)$ abbia tutti i poli contenuti nella circonferenza di raggio unitario;
- Se il sistema è anticausale, condizione necessaria e sufficiente per la stabilità BIBO è che la funzione di trasferimento $H(z)$ abbia tutti i poli all'esterno della circonferenza di raggio unitario, cioè con $|z| > 1$;
- Se il sistema è bilatero, condizione necessaria e sufficiente per la stabilità BIBO richiede che la circonferenza di raggio unitario $|z| = 1$ sia contenuta nell'anello circolare che identifica la regione di convergenza ROC nella funzione di trasferimento del sistema.

La condizione di stabilità implica che sia sempre possibile identificare la risposta in frequenza $H(e^{j\omega})$, essendo la circonferenza $|z| = 1$ contenuta nella regione di convergenza di $H(z)$.

Un sistema LTI fisicamente realizzabile deve possedere una risposta all'impulso $h(n)$ **causale** e con **coefficienti reali**. Ciò significa che:

1. La regione di convergenza $H(z)$ corrisponde all'esterno di un cerchio di raggio minore del polo di $H(z)$ di valore assoluto massimo (causalità);
2. Per ogni polo e zero complesso, deve essere presente anche il rispettivo complesso coniugato. Poli e zeri sull'asse reale possono essere singoli: hanno parte immaginaria nulla, quindi coincidono con i propri complessi coniugati (coefficienti reali).

Quando il polinomio è espresso tramite potenze negative, una costante corrisponde all'assenza di ritardo, quindi sia al numeratore che al denominatore si somma 1 per indicare l'istante corrente. Tutte le volte che la soluzione non è a coefficienti reali, è presente il complesso coniugato. I monomi vengono espressi in termini di potenze negative di z .

15 Filtri con poli e zeri

Avendo a disposizione la trasformata con i poli e gli zeri, conoscendo le proprietà di continuità della funzione analitica è possibile conoscere l'andamento delle frequenze.

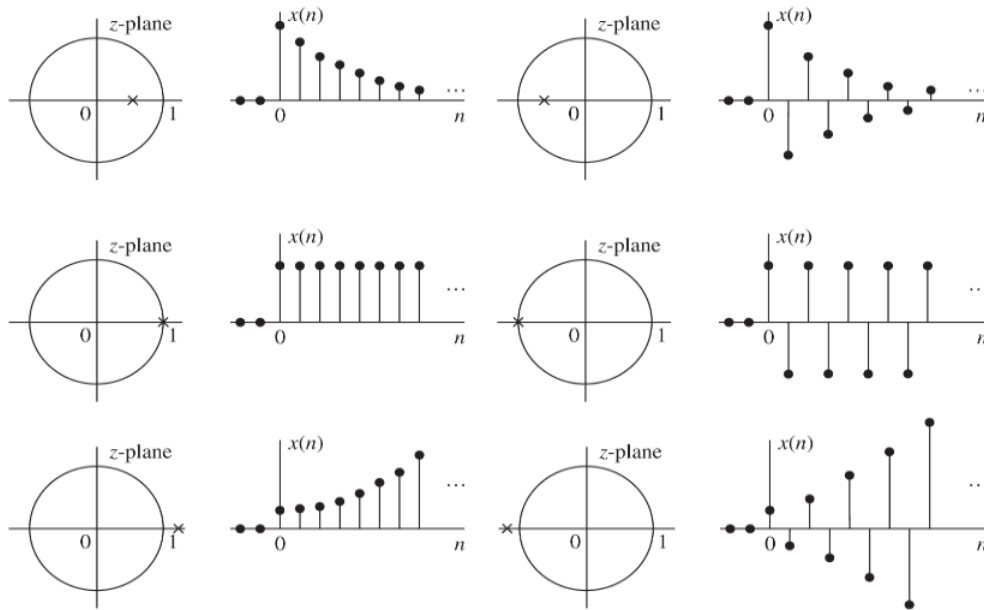
Se un polo reale è molto vicino al cerchio di raggio unitario, si ha che il denominatore si annulla con valori prossimi a $\rho = 1$, ma a $z = 1$ c'è un picco e la risposta cresce in funzione alla posizione del polo. In altre parole, ogni polo ha un'influenza sulle frequenze a seconda della loro vicinanza.

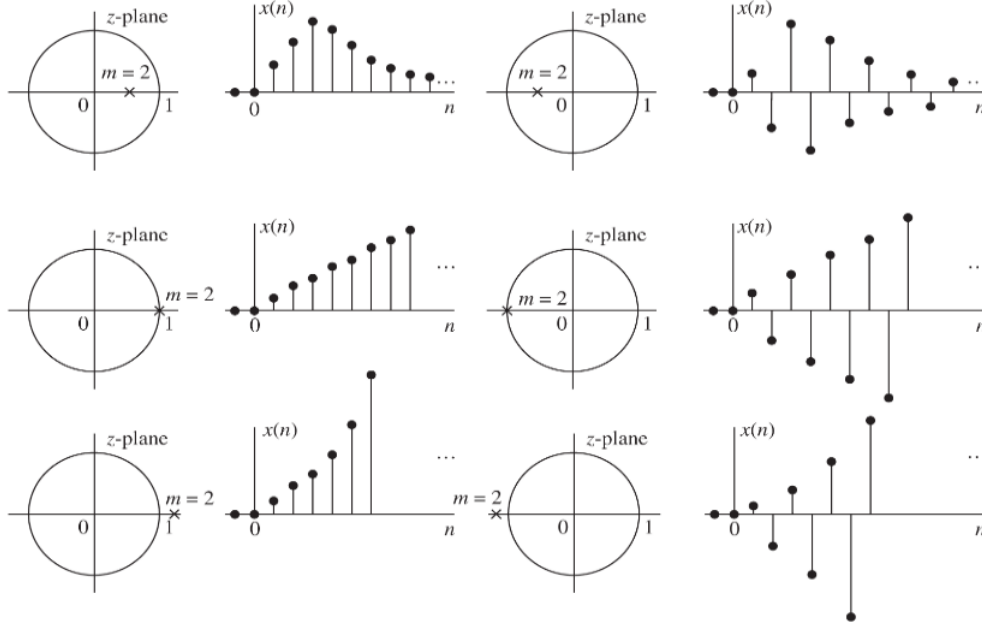
I **poli** devono essere posizionati in prossimità del cerchio di raggio unitario nelle pulsazioni complesse z corrispondenti alle componenti armoniche nel segnale d'ingresso $x(n)$ che devono essere **enfaticizzate**.

Gli **zeri** devono essere posizionati vicino ai punti z del cerchio di raggio unitario corrispondenti alle componenti armoniche del segnale d'ingresso $x(n)$ che devono essere **attenuate**.

Per la stabilità e la realizzabilità fisica (causalità), tutti i poli del filtro devono cadere nella circonferenza di raggio unitario, mentre gli zeri possono essere posizionati in qualunque punto del piano. Entrambi devono essere presenti a coppie complesse e coniugate.

Introducendo un fattore di normalizzazione $1 - \alpha$, è possibile confrontare i filtri con valore di picco (guadagno) unitario. Se la frequenza è causale, il numero di zeri non deve superare il numero di poli.





Ricostruendo la frequenza a partire dal filtro, si ha $H(z) = k \frac{z}{z-\alpha} = \frac{k}{1-\alpha z^{-1}}$ con k costante. Se il guadagno è unitario, $|H(z)_{f=0 \wedge z=1}| = 1$ da cui $k = 1 - \alpha$ (filtro con sempre la stessa altezza).

Solitamente la massima oscillazione δ_2 permessa in banda proibita viene espressa in dB attraverso l'oscillazione $R - P$, dove $R_p = -10 \log(\delta_2^2)$.

La banda proibita è data dall'insieme delle frequenze per le quali il guadagno è inferiore a un'opportuna soglia di attenzione, che normalmente viene stabilita in funzione dell'applicazione. Indicando con v_s la frequenza di stop e con v_c la frequenza di taglio, le frequenze comprese tra esse costituiscono la banda di transizione.

15.1 Passa-basso e passa-alto

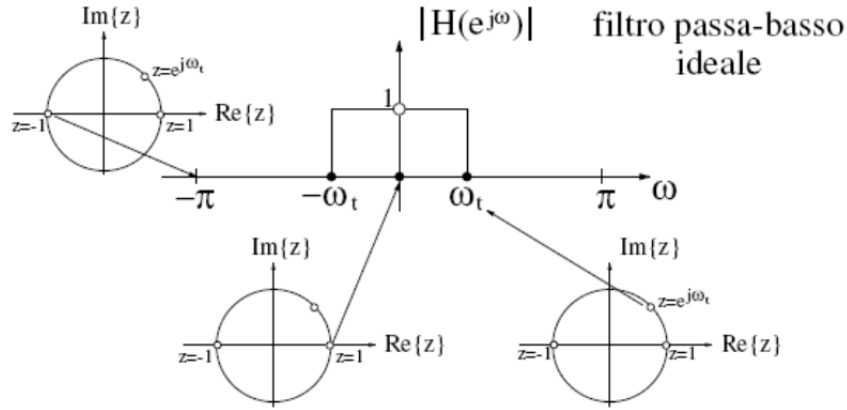
Tanto più i poli sono vicini all'asse reale, più influenzano le frequenze basse (filtro passa-basso).

I poli del filtro devono essere posizionati nelle pulsazioni complesse z corrispondenti alle frequenze della banda passante di $H(e^{j\omega})$: $|\omega| \in [\omega_t, \pi]$.

Gli zeri devono essere posizionati in prossimità, oppure sul cerchio $|z| = 1$ nelle pulsazioni complesse z corrispondenti alle frequenze $|\omega| \in [\omega_t, \pi]$.

Lo zero nell'origine $z = 0$ non ha alcun effetto sulla frequenza, essendo distante dalla circonferenza di raggio unitario: in questo caso la risposta in frequenza è determinata dalla posizione del polo.

Lo zero in $z = -1$, al contrario, enfatizza l'attenuazione alle alte frequenze. Essendo sulla circonferenza di raggio unitario, esso compare anche sulla risposta in frequenza del filtro in corrispondenza delle frequenze numeriche $\pm 1/2$.



Per enfatizzare le basse frequenze, è sufficiente inserire degli zeri per farle crescere di valore. Creando monomi con radici e aggiungendole al numeratore e al denominatore, si creano filtri in grado di alzare o abbassare le frequenze manipolando le posizioni di poli e zeri. Il numero di zeri non può essere superiore al numero di poli.

Nella realtà bisogna considerare che è impossibile azzerare completamente una parte di frequenza, ma ci saranno delle oscillazioni più o meno ampie.

Per realizzare un filtro passa-banda, è sufficiente unire un passa-basso e un passa-alto, per esempio posizionando due zeri in $z = 1$ e $z = -1$ con i rispettivi poli complessi e coniugati in $z = \rho e^{\pm j\pi/2}$ con $\rho < 1$.

15.2 Sistema inverso

Un **sistema inverso** ha lo scopo di compensare i poli e gli zeri, ed è un sistema che inverte il comportamento di un altro sistema LTI tale che la risposta rimanga inalterata.

Esempio: durante la trasmissione di dati attraverso il canale telefonico, il mezzo distorce il segnale trasmesso ottenendo in uscita un segnale $y(n) \neq x(n)$.

La compensazione della distorsione subita da $x(n)$ avviene in due fasi:

1. Stima della risposta in frequenza $H(e^{j\omega})$ del mezzo di trasmissione;
2. Filtraggio del segnale ricevuto $y(n)$ con un equalizzatore, sistema LTI che ha come risposta in frequenza la funzione $H^{-1}(e^{j\omega})$.

Se la stima del canale è condotta in modo ideale, si riesce a compensare perfettamente la distorsione introdotta dal mezzo di trasmissione.

La cascata di un sistema e del suo inverso ha una funzione di trasferimento pari a una costante unitaria nel piano z (il numeratore di uno dev'essere uguale al denominatore dell'altro e viceversa), definita sistema identità:

$$H_C(z) = H(z)H_I(z) = H(z)H^{-1}(z) = 1$$

Una classe di sistemi che ammettono l'inverso è quella dei sistemi caratterizzati da una trasformata z razionale.

La funzione di trasferimento del sistema inverso possiede zeri corrispondenti ai poli, e poli corrispondenti agli zeri del sistema diretto.

I sistemi LTI fisicamente realizzabili devono essere causali, con regione di convergenza in $|z| >$ polo di valore assoluto maggiore. Per un sistema stabile, il modulo dei poli dev'essere $|z_p| < 1$.

Se un sistema LTI con poli e zeri nel cerchio di raggio unitario rispetta queste condizioni, esso ammette un sistema inverso causale e stabile.

15.3 Sistemi passa-tutto

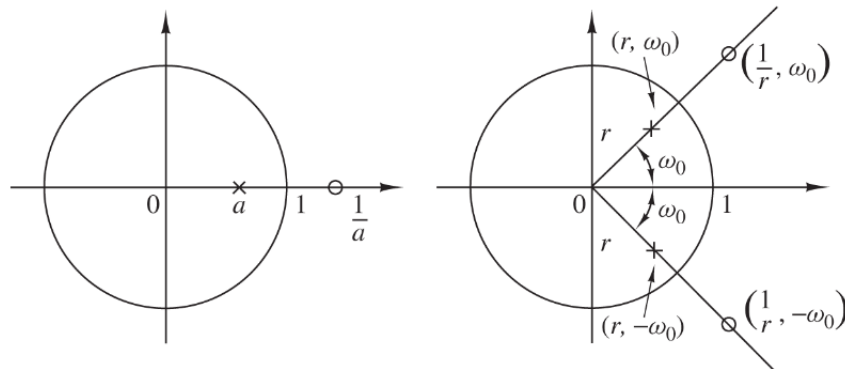
Questi filtri non modificano le frequenze del segnale, ma effettuano altre operazioni (ritardo, anticipo), e dev'esserci una precisa relazione tra i poli e gli zeri. La risposta all'impulso è sempre **costante**, per ogni frequenza: $H(e^{j\omega}) = 1$.

I coefficienti del polinomio al numeratore sono in ordine inverso rispetto a quelli del denominatore, e ciò viene garantito imponendo il modulo quadro unitario. Raccogliendo z , il numeratore resta invariato e il denominatore viene invertito.

Un polo, quindi, è il reciproco di uno zero (e viceversa):

- Se p_k è un polo di $H(z)$, allora $\frac{1}{p_k}$ è uno zero;
- Per la realizzabilità fisica, se p_k è un polo complesso dovrà essere polo anche p_k^* ;
- Di conseguenza, $\frac{1}{p_k}$ è uno zero.

Se uno è all'interno della circonferenza, l'altro è fuori: un sistema LTI passa-tutto causale e stabile presenta poli nel cerchio di raggio unitario nel piano z , e zeri all'esterno del cerchio $|z| = 1$. Di conseguenza, un sistema LTI passa-tutto stabile non ammette sistema inverso stabile.



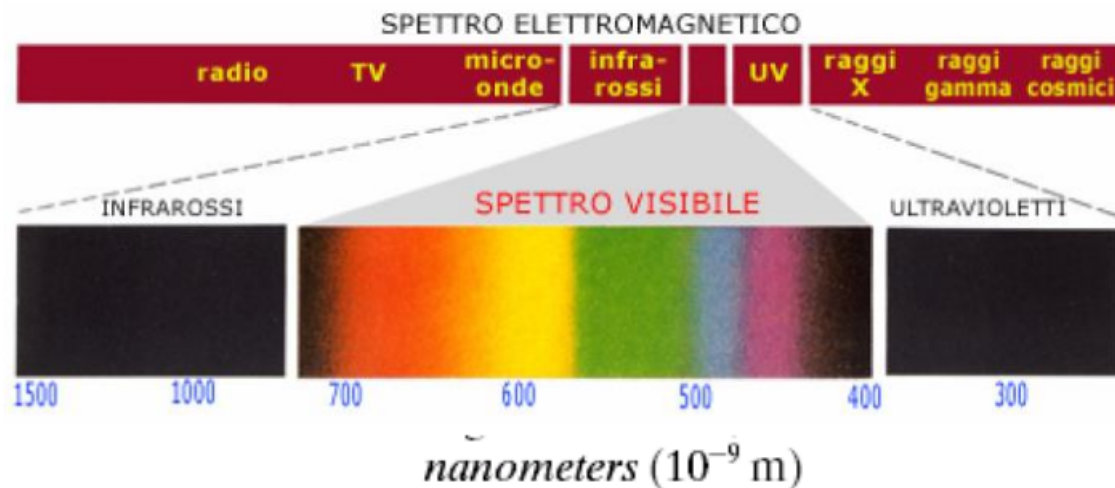
filtro passa tutto del
primo ordine
(1 polo 1 zero)

filtro passa tutto
del secondo ordine
2 poli comp. coniug.
2 zeri comp. coniug

16 Immagini

Un'immagine a colori viene percepita dal sistema visivo grazie all'impatto con le onde elettromagnetiche, caratterizzate dalla lunghezza d'onda λ (proporzionale all'inverso della frequenza, $\lambda \propto \frac{1}{\nu}$) e dall'ampiezza (intensità). La lunghezza d'onda è rapportata al periodo: indica quanto ci mette un ciclo a compiersi.

Lo spettro è in buona parte invisibile, con estremi alti che corrispondono alle radiazioni ultraviolette e bassi a radiazioni infrarosse. La porzione visibile è misurata in nanometri, con ordini di grandezza molto alti.



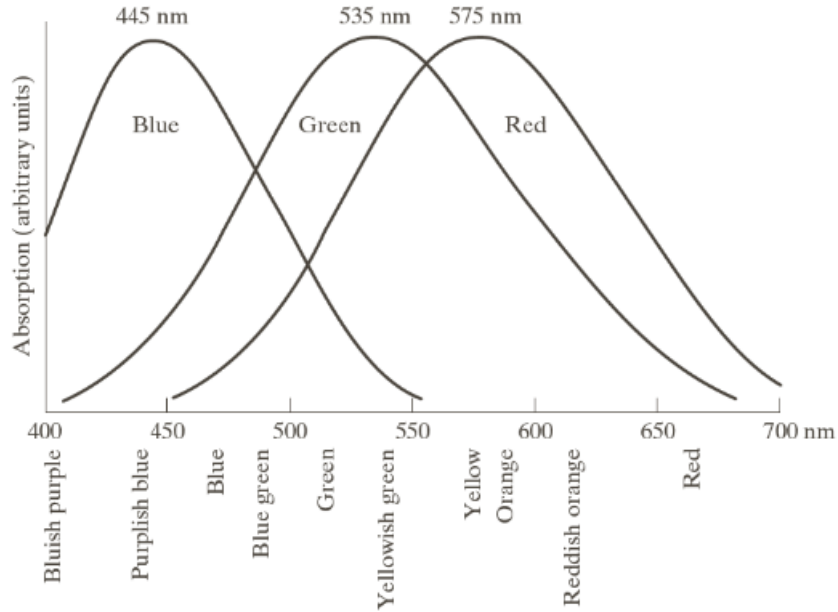
Il cristallino focalizza l'immagine sulla retina come una lente, e il colore è processato nella retina da coni e bastoncelli, in numerose tipologie con risposte differenti a seconda della lunghezza d'onda. Essi convertono le onde elettromagnetiche in uno stimolo elettrico.

I coni producono segnale per livelli di luminosità più alti, e possono essere:

- L (long waves), sensibili al rosso;
- M (middle waves), sensibili al verde;
- S (short waves), sensibili al blu.

I bastoncelli funzionano con le rappresentazioni in toni di grigio, esono in percentuale estremamente maggiore; essendo essi atti a misurare la variazione d'intensità, gli esseri umani hanno più sensibilità in condizioni di scarsa luce. La percezione ultima del colore è legata a intensità e cromaticità (tinta, saturazione), e le due dimensioni sono separabili.

L'occhio è più sensibile alle variazioni di luce nel centro dello spettro visibile, quindi lo spettro avrà la forma di una curva con funzione che varia al variare delle lunghezze d'onda. A seconda dei nanometri della radiazione, alcuni coni saranno attivati, eventualmente fondendo le risposte: il colore è combinazione di contributi a diverse lunghezze (spettro).



16.1 Segnale reale

Per descrivere esattamente il segnale che forma l'immagine a partire da una scena occorre:

- $E(\lambda)$, distribuzione spettrale dell'illuminante (SPD);
- $R(\lambda)$, riflettanza spettrale di ogni elemento della scena;
- $S(\lambda)$, sensitività dei recettori (coni e bastoncelli).

L'oggetto assorbe una parte della radiazione e ne ritrasmette un'altra parte (parametri legati alla capacità di riflettere), i cui segnali interferiscono con i recettori umani. La luce visibile è quindi $E(\lambda) \cdot R(\lambda)$, e ogni colore può essere descritto da tre parametri (RGB).

La maggior parte delle sorgenti luminose produce contributi di luce su più lunghezze d'onda:

$$L = \int E(\lambda)R(\lambda)l(\lambda)d\lambda$$

$$M = \int E(\lambda)R(\lambda)m(\lambda)d\lambda$$

$$S = \int E(\lambda)R(\lambda)s(\lambda)d\lambda$$

I numeri in output si definiscono valori di stimolo, che poi si traducono nei canali RGB. Le distribuzioni continue in funzione di λ si trasformano in valori discreti in spazio tridimensionale grazie all'integrale.

Le immagini acquisite da una fotocamera digitale sono differenti da quelle elaborate dall'occhio umano: il sistema visivo viene simulato tramite sensori e processamento. Ogni camera ha una

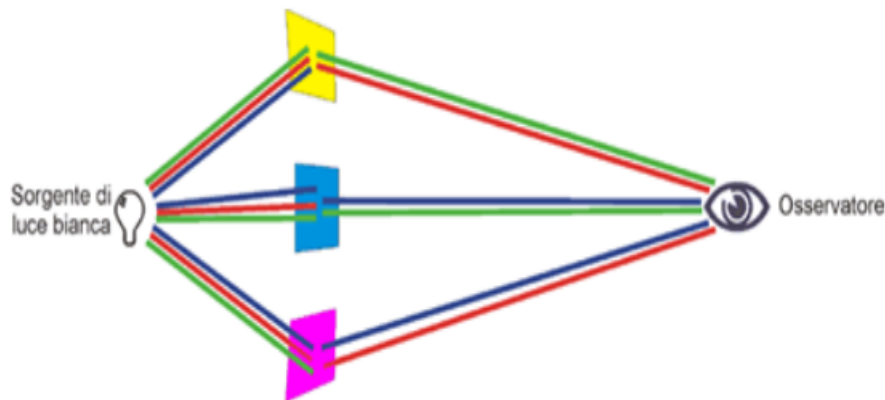
matrice che trasforma il segnale in modo da renderlo simile all'output del cervello, e i valori ottenuti dai sensori sono interpolati.

I dati sono fatti passare in filtri colore, disposti secondo Bayer pattern. Il verde è il colore recepito meglio, e permette di misurare maggiormente le variazioni di intensità (e conseguentemente i dettagli dell'immagine).

16.2 Sintesi

La somma dei contributi delle onde è chiamata sintesi additiva, e viene usata dai dispositivi che imitano l'occhio umano; esiste anche la sintesi sottrattiva, che fa differenze in base allo spettro e ai colori complementari di RGB (ciano, magenta, giallo).

Quest'ultima consiste nella sovrapposizione di più filtri: il colore che giunge alla vista è quello che riesce a passare per tutti i filtri, dato che ognuno di essi sottrae una parte della luce che lo attraversa.



Esistono rappresentazioni che separano i canali in base alla loro intensità, mettendola in evidenza in base alla frequenza delle onde. Alcuni esempi di modelli sono:

- RGB (red, green, blue), usato dalle camere e videocamere digitali;
- HSB (hue, saturation, brightness), corrisponde alla percezione umana del colore;
- HSV (hue, saturation, value), HSB con tinta definita su un angolo;
- CMYK (cyan, magenta, yellow, black), utilizzato per la sintesi sottrattiva nei sistemi di riproduzione;
- YIQ, per i segnali TV e VHS in America;
- YUV, per i segnali TV e VHS in Europa;
- YcbCr, per i video digitali.

Il formato grafico è la tecnologia utilizzata per memorizzare il file. Le immagini hanno due tipologie di formato:

- Raster, scalate in base a un numero prefissato di pixel e bit per pixel (bitmap) in una griglia di elementi, usate per digitalizzazione ed elaborazione di scene reali. L'accuratezza della rappresentazione dipende dal numero di pixel e dalla loro condifica (quantizzazione);
- Vettoriale, rappresentate con formule matematiche e oggetti, applicabili solo nel mondo geometrico ma con qualità maggiore. Le immagini sono più complesse e non viene persa definizione, ma sono più difficilmente modificabili.

Un'immagine in meta formato è una combinazione di questi formati, cioè contiene informazioni di ogni tipo: solitamente viene usato raster con alcuni elementi vettoriali.

16.3 Risoluzione

Le immagini hanno dimensioni espresse in pixel, mentre le dimensioni fisiche dipendono dal dispositivo di acquisizione.

La risoluzione dà informazioni sulla quantità o la densità di pixel contenuti in un'immagine (ppi, pixel per inches). In caso di stampa o acquisizione con scanner si misura in dpi, dot per inches (gocce di inchiostro). Più alta è la risoluzione, più piccola è la dimensione fisica dei pixel: tipicamente una qualità alta di stampa ha una risoluzione intorno ai 300 ppi.

La risoluzione ottima dipende dallo scopo dell'immagine, tenendo conto di un fattore moltiplicativo di ingrandimento: esso rappresenta il rapporto tra il lato dell'immagine di destinazione e la sua dimensione fisica.

Esempio: diapositiva di 24×36 mm, da stampare in 10×15 in 300 ppi:

$$\frac{100\text{mm}}{24\text{mm}} \cdot 300\text{ppi} = 1248\text{ppi}$$

Le immagini sugli schermi video solitamente sono acquisite in 72 dpi.

$$\text{File size} = \frac{\text{height} \cdot \text{width} \cdot \text{bit depth} \cdot \text{dpi}}{8} = \frac{\text{pixel dimensions} \cdot \text{bit depth}}{8}$$

16.4 Quantizzazione cromatica

La profondità del colore è il numero di bit per pixel:

- 1 bit rappresenta 2 colori (bianco e nero);
- 8 bit rappresentano 256 colori (colormap, livelli di grigio);
- 24 bit rappresentano quasi 17 milioni di colori (true color).

Per risparmiare sulla profondità si usano palette contenenti il sottoinsieme dei colori rappresentabili che compare in una foto. Una lookup table individua i corrispondenti valori di rosso, verde e blu.

Le dimensioni, la risoluzione e il numero di bit determinano l'occupazione in memoria, proporzionale a essi. Per una visualizzazione in scala 1 : 1 l'immagine deve avere la stessa risoluzione del display.

Un istogramma rappresenta l'insieme dei valori che un segnale può assumere, espresso in livelli di quantizzazione. Non dà informazioni spaziali, ma solo sul contrasto e sul numero di bit: indica quanti

campioni assumono un certo valore (quantizzatore ottimo), e in base a ciò è possibile decidere come applicare la compressione (probabilità normalizzata).

Il dynamic range è l'intervallo tra ampiezza massima e minima di un segnale, legato al numero di bit e al contrasto: se esso viene sfruttato appieno, l'immagine è ben contrastata. L'occhio umano può distinguere valori di luminosità massima/minima con un rapporto di circa 10.000 : 1, mentre i monitor riproducono a 256 : 1.

Per rappresentare meglio immagini con bianco o nero predominante, si utilizzano quantizzatori con pochi bit tagliando tutti i valori che non vengono utilizzati, in modo da avere una distribuzione più uniforme.

L'occupazione del range dinamico dipende dal tempo di esposizione: più è breve, più la fotografia è luminosa. Il sistema HDR usa n bit per catturare una porzione del range dinamico con differente esposizione, per poi sovrapporre tutte le immagini ottenute in modo da avere una visione più chiara.

Per visualizzare scene con alto range dinamico (HDR) sui monitor, è necessario effettuare il tone mapping, cioè la riduzione in scala inferiore (8 bit). Il dithering è la conversione a una profondità colore inferiore: i pixel sono divisi in blocchi, ognuno dei quali viene riempito con diversi pattern di nero allo scopo di simulare i livelli di grigio. L'effetto è la perdita di risoluzione.

Per evitare i "salti", oggi pixel è sostituito con un pattern più grande, in modo che il numero di punti stampati possa approssimare il livello di intensità (half tone printing). Le dimensioni, pertanto, sono più grandi. Per le immagini a colori, si usano vicinanze di colori diversi per creare un altro osservando a distanza.

Effettuando il dithering bisogna anche aggiungere rumore casuale: in questo modo le basse frequenze vengono nascoste dalle alte, e interferiscono con il rumore di quantizzazione ottenendo un disturbo più regolare e quindi percepito in misura minore.

17 Audio

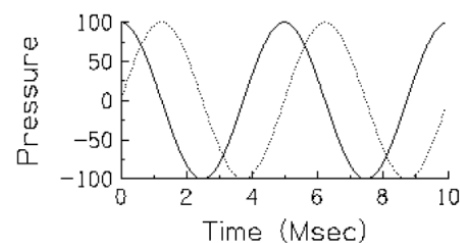
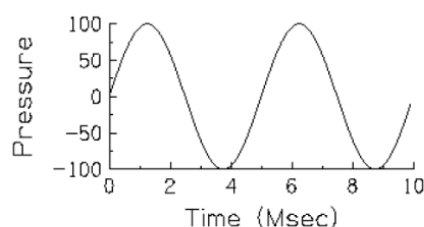
Un segnale sonoro è la variazione di pressione sul timpano, percepito nel tempo attraverso l'aria (mezzo di propagazione).

Il suono è un'onda di pressione come la luce (fenomeno ondulatorio), ma macroscopico: le molecole dell'aria vengono compresse ed espanse, e la sorgente sonora vibra in modo longitudinale nella stessa direzione di propagazione del suono.

$$A(t) = A_{max} \cdot \sin(2\pi ft + \varphi_0)$$

I suoni elementari hanno andamento sinusoidale, periodico e con estensione indefinita.

La maggior parte dei suoni natura sono caratterizzati da forme d'onda diverse, ma possono essere scomposte come una combinazione di suoni elementari.



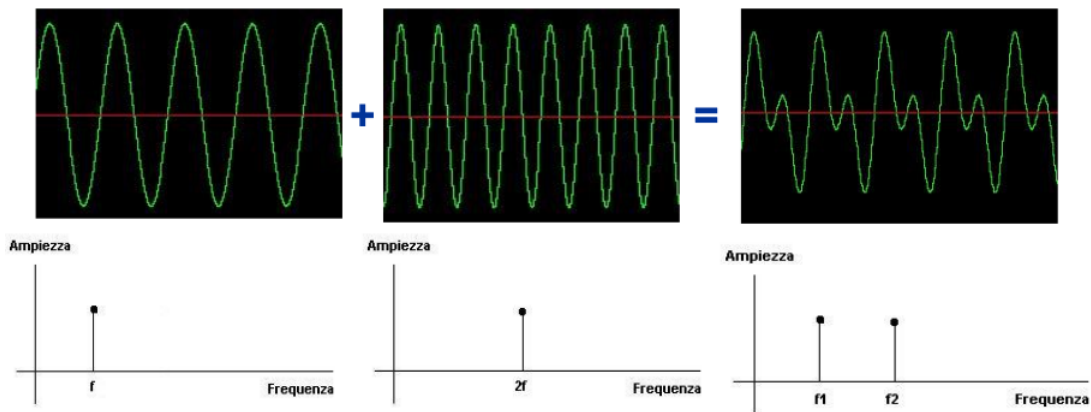
Le componenti di un segnale sonoro sono individuate da:

- Ampiezza A , che si misura rispetto al valore medio della pressione dell'aria ed è espressa in dB;
- Periodo T , la durata nel tempo di ogni ciclo dell'oscillazione, espresso in secondi;
- Frequenza F , velocità con cui i valori di pressione fluttuano ciclicamente, espressa in numero di cicli al secondo (onde e Hz).

Il valore 0 indicato sull'asse della pressione corrisponde al valore medio della pressione nell'aria. La differenza di fase ha unicamente a che vedere con il fatto che due funzioni siano diversamente allineate rispetto al tempo.

17.1 Analisi di Fourier

L'analisi di Fourier permette la rappresentazione del segnale sonoro nel dominio delle frequenze a partire dal tempo, esplicitando f .



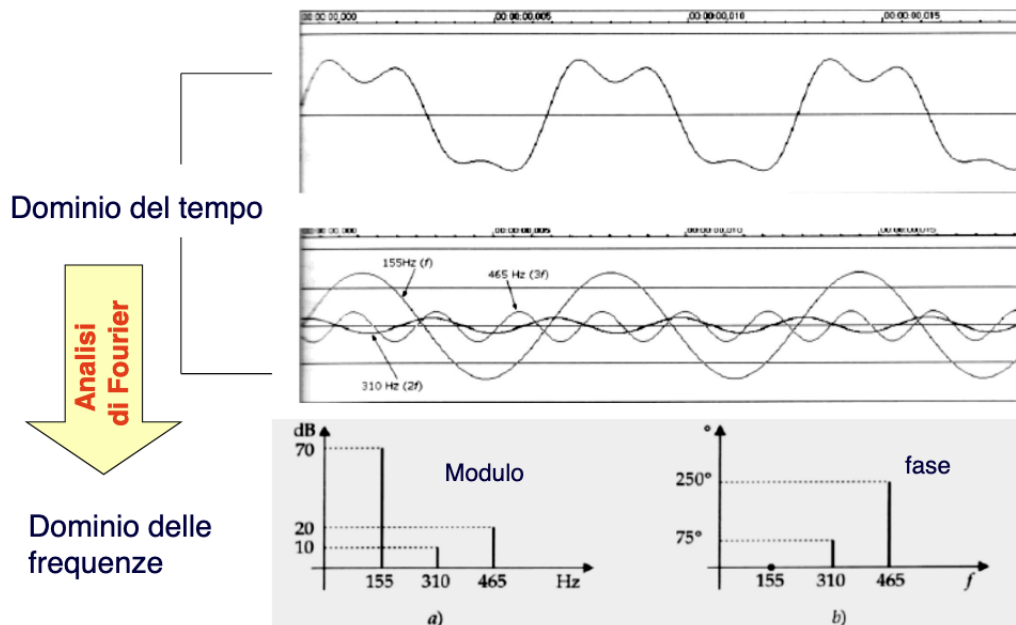
La presenza di una linea nello spettro in frequenza indica la presenza di un segnale esattamente sinusoidale periodo, tuttavia i suoni caratterizzati da uno spettro discreto sono pochi. I suoni normalmente uditi hanno un inizio e una fine precisi, cioè sono contenuti in un intervallo temporale finito.

Ai parametri che descrivono un segnale ondulatorio possono essere associate le tre grandezze percettive che descrivono ogni suono:

- Altezza, che rappresenta la tonalità dell'audio e ha come parametro la sequenza;
- Intensità, il volume, con parametro fisico l'ampiezza;
- Timbro, cioè la tipologia di strumento, con parametro fisico lo spettro.

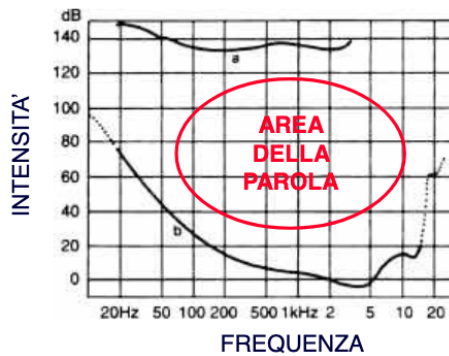
A parità di frequenza fondamentale e intensità, due suoni possono differire per timbro (la sovrapposizione delle onde sinusoidali può essere diversa).

La frequenza fondamentale è proporzionale all'altezza del suono, cioè della sensazione di acutezza o gravità. L'aumento della frequenza non è linearmente in rapporto con l'ampiezza, ma si comporta



seguendo una scala di tipo logaritmico. Affinché in un suono sia possibile individuare un'altezza, esso deve essere periodico.

17.2 Grandezze fisiche e grandezze percettive



Ampiezza e frequenza della forma d'onda hanno effetto sul suono percepito: variazioni di piccola ampiezza producono suoni di bassa intensità e viceversa, e al crescere della frequenza aumenta il tono (non linearmente).

Le energie in gioco nei fenomeni acustici sono irrilevanti rispetto a quelle nel fenomeno luminoso.

Il range di suoni in grado di essere percepite dagli umani (tra i 20 e i 20.000 Hz) è minore dell'insieme delle frequenze possibili, oltre esso ci sono ultrasuoni e infrasuoni.

Il volume aumenta man mano che l'ampiezza cresce, e l'incremento di tono è sempre più piccolo al crescere della frequenza.

Il campo di udibilità è determinato dai valori limite di intensità e frequenza. Il limite inferiore per l'intensità è costituito dalla curva di soglia di udibilità, mentre quello superiore dalla soglia del dolore.

Le curve isofone rappresentano i suoni all'interno di una soglia di udibilità, che può essere superata fino al fastidio. Esse dipendono dall'intensità e dalla frequenza.

Vengono usate scale di rappresentazione che riflettono uguali differenze percettive. La scala di Mei indica la relazione tra sensazione di altezza e frequenza, uguali differenze sulla scala corrispondono

a uguali differenze percepite di tono ma non di frequenza.

Le ampiezze sono rappresentate in scala logaritmica e hanno come unità di misura il dB. Incrementi di 1 dB corrispondono a JND (Just Noticeable Differences).

17.3 Campionamento

Le variazioni di pressione vengono tradotte in variazioni di tensione/corrente elettrica dal microfono, che effettua una trasduzione elettroacustica. Il filtro antialiasing è applicato tramite amplificazione e filtraggio, per poi effettuare conversione A/D e registrazione su supporto.

La funzione a gradini generata dal convertitore viene filtrata con un filtro passa-basso prima del campionamento, per limitare le frequenze e l'aliasing. L'intervallo di frequenze che viene mantenuto dipende dall'applicazione.

Il segnale vocale può avere componenti fino a 10 KhZ, ma in genere se ne utilizzano solo 8, quindi questa è la frequenza di campionamento del segnale telefonico. Il segnale musicale ha range tra 20 Hz e 20 kHz.

Dopo la conversione DA, nell'output possono essere di nuovo presenti alte frequenze a causa del campionamento e della quantizzazione, quindi viene applicato un ulteriore filtro. La frequenza di campionamento standard è di 44,1 kHz, per registrare audio in formati differenti. Al di fuori di tale soglia si incorre in sovracampionamento o sottocampionamento (aliasing).

La maggior parte dell'informazione del segnale vocale è contenuta in 4 kHz, ma esso andrebbe campionato a 40 kHz. Esiste anche per il campionamento il SNR (dB), con potenza del segnale proporzionale al quadrato della tensione.

$$SNR = 10 \log_{10} \frac{V^2_{\text{segnale}}}{V^2_{\text{rumore}}} = 20 \log_{10} \frac{V_{\text{segnale}}}{V_{\text{rumore}}}$$

17.4 Dimensione e codifica

Lo spazio in memoria (kB) occupato da un file audio si calcola con la seguente formula:

$$\frac{f_c \cdot D \cdot b \cdot N_c}{8 \cdot 1024}$$

f_c è la frequenza di campionamento, D è la durata in secondi, b è il numero di bit del quantizzatore e N_c è il numero di canali (mono o stereo).

I dati non compressi crescono al crescere della risoluzione del quantizzatore, dato che aumenta il numero di bit. Un segnale stereo raddoppia la banda per la trasmissione.

La codifica può avvenire in due modi:

- PCM (Pulse Code Modulation), modulazione del codice dell'impulso;
- PAM (Pulse Amplitude Modulation), modulazione dell'ampiezza dell'impulso.

La PCM converte in forma digitale i segnali analogici, trasformando le forma d'onda attraverso campionamento, quantizzazione e codifica. I valori numerici diventano e di 4 bit, e la rappresentazione può essere lineare o non.

Il trasmettitore digitale associa una forma d'onda a ogni cifra in uscita dal PCM, e all'uscita del canale le onde sono riconosciute dal ricevitore digitale e riconvertite nelle cifre originarie.

PAM è un modo per associare forme d'onda a simboli in uscita da una qualsiasi sorgente discreta. Essa prevede che si associ a ogni numero nella sequenza di bit la stessa pulsazione base, con un'ampiezza che dipende dal valore trasmesso.

I sistemi PAM inviano sul canale una sequenza di forme d'onda diverse tra loro, ma di diversa ampiezza. In ricezione, osservando l'ampiezza si può risalire al messaggio originale.

Nel processo di digitalizzazione, la qualità del suono è legata al tentativo di soddisfare due esigenze opposte:

1. Fedeltà nella riproduzione;
2. Dimensione del file.

Le onde creano interferenze nello spazio, quindi è necessario creare l'effetto di più suoni generati in punti distinti e il numero di canali è un fattore importante. Le configurazioni più usate sono stereo e surround 5 + 1, con 5 canali contenenti frequenze medie e alte, e un subwoofer per quelle basse.

In fase di scrittura su disco, i campioni provenienti da diversi canali vengono trasmessi in successione. Questa tecnica si chiama interleaving, o multiplexing in caso di trasmissione. Al momento della riproduzione, occorre utilizzare buffer per la sincronizzazione.

18 Video

Un video viene percepito dal nostro sistema visivo con la persistenza della visione, e per ottenere il movimento apparente (serie di immagini statiche) è necessario individuare la frequenza minima per avere una visione continua.

La persistenza della visione è un fenomeno per il quale un'immagine rimane per qualche frazione di secondo in memoria, attribuito a livello cerebrale. Ciò permette di ottenere un movimento continuo.

Esistono frequenze di campionamento prestabilite per percepire una serie di immagini come un video continuo:

- Una teleconferenza è a 10 fps;
- Un film muto è a 16 fps (limite del movimento a scatti);
- Un film sonoro è a 24 fps;
- Le televisioni sono a 25-60 fps, a seconda della tipologia.

Dalla conoscenza della risposta si determina la progettazione di un dispositivo di riproduzione video. La risposta dell'HVS dipende dal contenuto frequenziale, rispetto al tempo e allo spazio.

Per definire il numero di immagini al secondo (frame rate e refresh rate) bisogna definire una soglia critica legata alla frequenza di Flicker, ottenuta in base a una serie di parametri, quindi variabile nel tempo e nello spazio.

Se la frequenza è troppo alta, l'immagine non è più nitida: il range ha anche un limite superiore, con estremi tra i 20 e gli 80 Hz, definendo una Critical Flicker Frequency alla quale lo stimolo passa da intermittente a continuo. Essa varia in funzione di:

- Luminosità media del display;
- Luminosità dell'ambiente;
- Distanza dallo schermo.

Un video field è un insieme di campioni in un'immagine, composto da linee alternate. I vari field sono campionati a istanti diversi. L'aumento delle frequenza di refresh per la televisione è ottenuto dividendo l'immagine in due campi, le cui righe non vengono scansionate sequenzialmente ma sono alternate tra pari e dispari. Il refresh include anche la ripetizione della stessa immagine, riducendo il flicker.

La scansione progressiva consiste nella trasmissione in sequenza e in modo continuo di tutte le linee. Al contrario, la scansione interlacciata divide il frame in field trasmessi in metà del tempo totale.

I 24 fps del cinema non sono sempre sufficienti, quindi i proiettori possono arrivare a 48 o addirittura 72. Nella trasmissione analogica, invece, la frequenza viene virtualmente raddoppiata con l'interfacciamento: le risposte in frequenza vengono combinate. Nei computer si utilizza un rate di 72 fps.

Il sottocampionamento introduce aliasing quando il movimento è molto veloce, causando differenze di percezione ovviate da opportuni filtri passa-basso (smoothing).

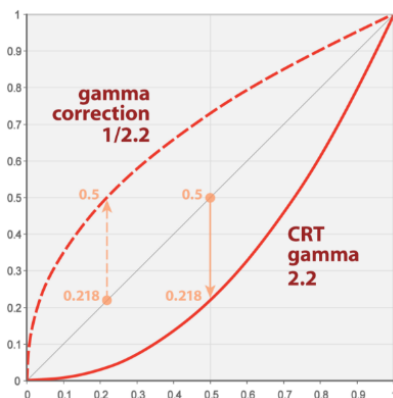
18.1 Riproduzione

Un segnale analogico ha risoluzione definita dal numero di linee di scansione.

Un segnale digitale ha risoluzione definita dal numero di pixel.

La riproduzione del colore cambia a seconda dello spazio, delle modalità di trasmissione e della quantizzazione per il segnale digitale. Bisogna rappresentare il colore mantenendo un segnale di luminanza compatibile, le stesse temporizzazioni e la stessa banda.

Per codificare le immagini, si cambia spazio colore in modo da separare luminanza da cromaticità. Non ci sono limitazioni relative ai canali e ai componenti sulla stessa portante.



Il video fornisce informazioni sulla luminosità dell'immagine usando un segnale di luminanza Y ottenuto sommando in modo pesato le componenti RGB, a cui viene in seguito applicata una correzione gamma.

$$Y = 0.299R + 0.587G + 0.114B$$

La maggioranza del segnale è composta dal verde; l'informazione colore è codificata in due canali aggiuntivi, ottenuti sottraendo la luminanza dalle componenti rosso e blu.

$$U = 0.492(B - Y) \quad V = 0.877(R - Y)$$

Queste due componenti sono chiamate *chrome*, e la rappresentazione di un valore indipendente dalla luminanza è definita *crominanza*.

I coefficienti associati a ogni colore cambiano in base alla posizione geografica per lo standard analogico, mentre per il digitale è globalmente YCrCb. Nella combinazione lineare, il canale più importante è quello del verde.

I più importanti standard di video analogico sono NTSC in Nord America e Giappone, PAL e SECAM in Europa. I segnali analogici usati sono YUV e YIQ. Le informazioni su colore UV e IQ sono combinate insieme in un segnale di chroma, che a sua volta è combinato con la luminanza Y.

Un altro aspetto da considerare è la dipendenza tra l'intensità del monitor e la tensione, la quale ha un fattore esponenziale: il monitor tende a comprimere gli scuri, mostrandoli in livello maggiore.

$$B_d = v_d^{\gamma_d} \quad \begin{cases} v_d = \text{tensione} \\ B_d = \text{brightness} \\ \gamma_d = \text{gamma} \end{cases}$$

Per questo motivo si utilizza la gamma correction, che aumenta l'intensità dei grigi. Essa compensa le caratteristiche di non linearità dei display, per mantenere la compatibilità con i televisori in bianco e nero.

Se il valore del canale rosso è R , lo schermo emette una luce proporzionale a R^γ , con $\gamma \approx 2.2$. In genere il segnale normalizzato viene corretto prima della trasmissione, elevando a $1/\gamma$.

L'HDTV (alta definizione) è concepita per avere una risoluzione doppia rispetto alla TV analogica, con dimensioni in pixel molto più grandi che però permettono l'adattamento agli standard precedenti, aumentando il refresh rate.

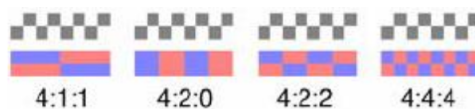
Poiché il sistema visivo è più sensibile all'intensità luminosa che alla croma, si riduce la risoluzione spaziale delle componenti cromatiche. Il campionamento del colore è espresso come $x : y : z$, dove:

- x è il numero di campioni di luminanza;
- y è il numero dei campioni di chroma per ogni linea dispari;
- z è il numero di campioni di chroma per le altre linee.

Nella digitalizzazione, i canali Cr e Cb sono di solito campionati orizzontalmente. Lo schema di sottocampionamento 4 : 4 indica che non c'è sottocampionamento sulle componenti cromatiche, quindi i valori YCbCr sono tutti trasmessi.

Per eliminare le coppie, si utilizzano schemi del tipo 4 : 2 : 2 o 4 : 1 : 1, che trasmettono solamente alcuni valori campionando orizzontalmente.

La compressione JPEG usa 4 : 2 : 0, il quale elimina un fattore 2 in entrambe le dimensioni posizionando fra righe e colonne un pixel con croma media. Per ogni 4 campioni di luminanza, ci sono 2 chroma sulle linee dispari.



Chroma Sampled 8x: il sottocampionamento non è percepito, utilizzando solo il canale di intensità.

19 DCT wavelets

L'obiettivo dell'analisi è la riduzione delle ridondanze, spostandosi in uno spazio dove le informazioni e i canali sono separati.

Nel dominio diretto le componenti di un segnale X sono tra loro significativamente correlate, quindi la stessa informazione è ridondante. Spostandosi nel dominio trasformato $Y = T[X]$, si cerca una rappresentazione dove le componenti siano molto meno correlate.

Y può essere codificato in modo più efficiente di X , utilizzando solo le componenti del segnale che descrivono l'informazione. Lo spettro, per esempio, ha un contributo maggiore espresso con le basse frequenze al centro, quindi è possibile codificare solo una porzione determinata di energia.

In termini di compressione, sono necessari solo i bit delle componenti da tenere. Il modello di compressione permette di quantizzare con perdite trascurabili, in un cambio di spazio secondo due strategie:

- Codificatore di sorgente, che riduce le ridondanze;
- Codificatore di canale, che incrementa l'immunità al rumore.

19.1 Codifica con trasformate

Questo algoritmo introduce perdita ed è computazionalmente costoso, a causa del calcolo delle trasformate. Una trasformata lineare e reversibile è usata per il mapping dell'immagine in un set di coefficienti che vengono poi quantificati e codificati.

Il mapping può essere effettuato secondo diverse metodologie, a seconda della capacità di decorrelazione dei dati, semplicità di realizzazione e altri fattori.

La KLT (analisi alle componenti principali) è la trasformata ottima, ma è computazionalmente inefficiente. DCT, invece, approssima il comportamento ottimo, ed è usata negli algoritmi di codifica più utilizzati (jpeg, mpeg).

La trasformata wavelet permette la multirisoluzione (jpeg2000), e funziona secondo il principio di determinazione di Heisenberg, trovando un compromesso tra le frequenze e il dominio temporale lavorando con tradeoff.

Data una funzione $f(x)$, la sua DFT $F(u)$ è:

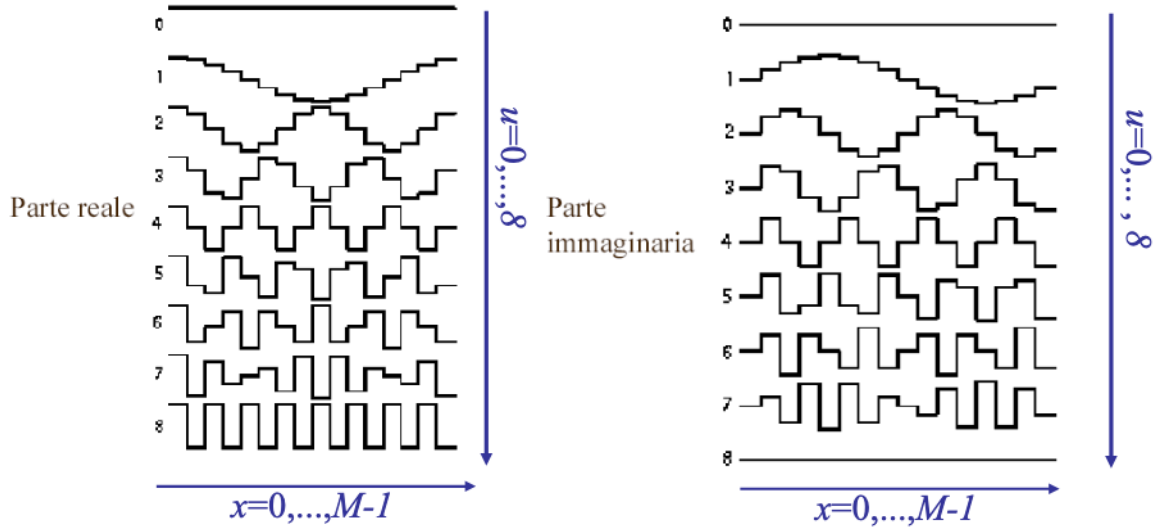
$$F(u) = \frac{1}{M} \sum_{x=0}^{M-1} f(x) e^{-j \frac{2\pi}{M} ux} = \frac{1}{M} \sum_{x=0}^{M-1} f(x) g(u, x)$$

$$g(u, x) = e^{-j \frac{2\pi}{M} ux}$$

La trasformata è lineare e invertibile, e $g(u, x)$ è detto kernel della trasformazione diretta, da cui dipendono le proprietà. I kernel devono essere invertibili, lineari e separabili.

Trasformata di Fourier discreta 1D:

$$e^{-j \frac{2\pi}{M} ux} = \cos\left(\frac{2\pi}{M} ux\right) + j \sin\left(\frac{2\pi}{M} ux\right)$$



Il segnale di partenza è proiettato in base al suo tempo e spazio con le sue parti reale e immaginaria. Seno e coseno hanno simmetria rispettivamente dispari e pari, e trasformando si osserva l'influenza di ciascuna componente.

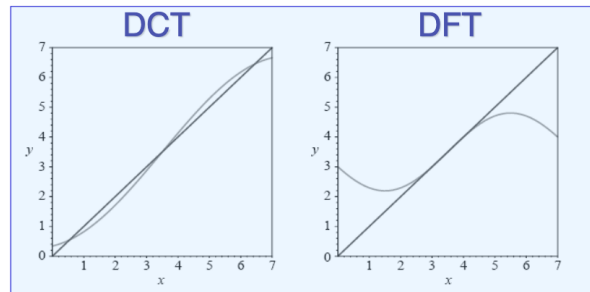
La trasformata coseno discreta (DCT) è lineare, con kernel diretto uguale a quello inverso (simmetria speculare), separabile e simmetrico. La componente continua è legata al valore medio dell'immagine (0, 0).

$$T(u) = \sum_{x=0}^{M-1} f(x)g(u, x) \quad g(x, u) = \alpha(u) \cos \left[\frac{(2x+1)\pi u}{2N} \right]$$

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & u = 0 \\ \sqrt{\frac{2}{N}} & u = 1, 2, \dots, N-1 \end{cases}$$

La separabilità permette di calcolare la trasformata 2D tramite applicazioni successive della trasformata 1D alle righe e alle colonne, senza perdita di informazione. Ciascun blocco è costituito da $N \times N$ sottoblocchi.

Una maggiore quantità di informazione è presente nei primi coefficienti della DCT, rispetto allo stesso numero di coefficienti della DFT.



19.2 Analisi multirisoluzione

Le immagini sono generalmente costituite da regioni connesse che formano gli oggetti, omogenee rispetto a una qualche proprietà.

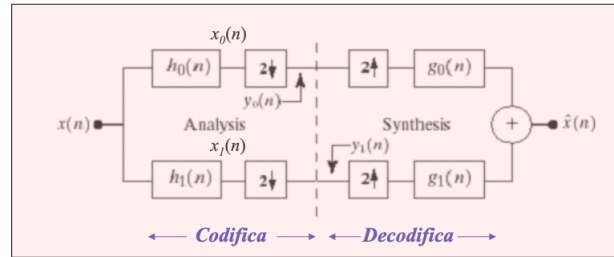
L'analisi multirisoluzione permette di mettere in evidenza sia le immagini a bassa frequenza che quelle ad alta, con segnali e campioni di dimensioni diverse, concentrandosi su differenti posizioni nello spettro.

Le caratteristiche locali di un'immagine sono contraddistinte da variazioni statistiche locali, dovute discontinuità fra regioni omogenee. Caratteristiche nascoste a una data risoluzione possono essere individuabili a un'altra.

La tecnica più frequente è l'analisi piramidale, che parte dalla risoluzione massima e progressivamente sottocampiona, evidenziando i cambiamenti tra i dettagli e le perdite tra un livello e l'altro ricostruendo ogni volta l'immagine con meno dettagli. I pixel mancanti vengono approssimati tramite wavelet.

Il segnale è decomposto in un insieme di sottosegnali (sottobanda, analisi), passando attraverso filtri complementari, ciascuno dei quali agisce su una fascia di frequenze. Ricampionando, filtrando e ricombinando le sottobande (sintesi) si ottiene un'approssimazione $\hat{x}(n)$ dell'originale, non incorrendo in aliasing.

Ciascuna sottobanda ($y_0(n), y_1(n)$) è ottenuta filtrando con un passa-banda ($h_0(n), h_1(n)$) il segnale originale. Poiché la sottobanda ha spettro limitato e pari a metà dell'originale, è possibile sottocampionare senza perdita di informazione.

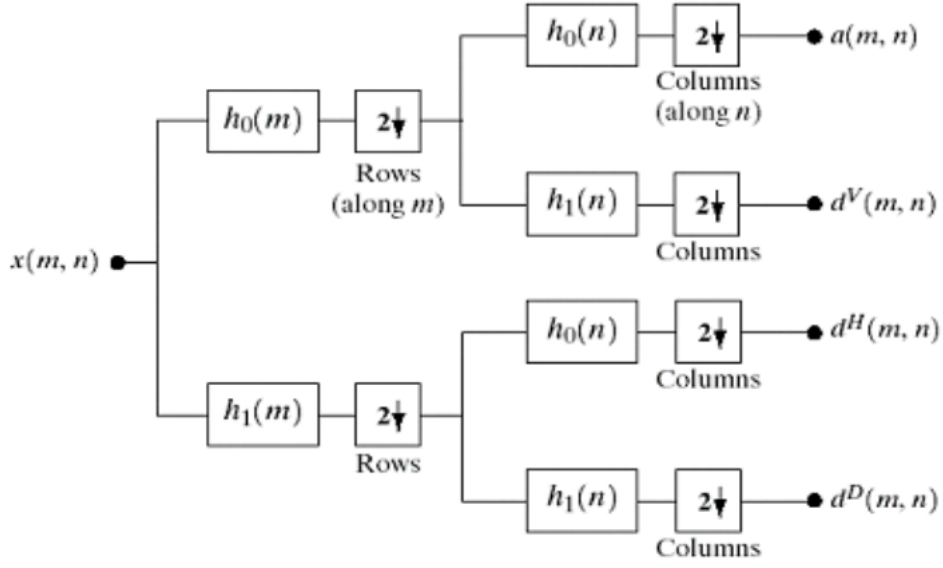


Imponendo $\hat{x}(n) = x(n)$, si trovano le opportune coppie di filtri di sintesi e corrispondenti filtri di analisi che garantiscono una perfetta ricostruzione. In caso di più dimensioni (separabili), possono esserci n filtri per righe e colonne che mettono in evidenza diverse porzioni di frequenze rispetto alla direzione. Flitraggio e sottocampionamento avvengono quindi in due fasi successive.

Si ottengono 4 immagini di output: $dV(m, n)$, $dH(m, n)$, $dD(m, n)$ le immagini nelle 3 dimensioni per le alte frequenze, e $a(m, n)$ l'immagine approssimata. Ciascuna sottobanda a sua volta può essere scomposta in 4 ulteriori sottobande.

Visivamente, è immediata la correlazione fra filtri di analisi e sintesi corrispondenti: sono ortonormali, cioè ortogonali a norma unitaria, e permettono una ricostruzione error-free in ciascun livello che a sua volta viene approssimato.

Dato che le statistiche locali sono facilmente modellabili e presentano molti valori nulli, questa codifica è particolarmente vantaggiosa per la compressione: non tutto l'istogramma viene occupato,



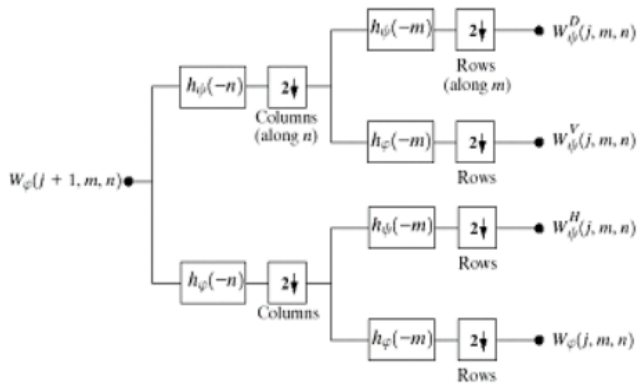
oppure ci sono poche alte frequenze che possono essere eliminate. L'approccio di denoising per ridurre il rumore è utilizzato nelle applicazioni mediche.

Nell'analisi multirisoluzione, il filtro passa-basso è una funzione di scala (MRA), mentre il passa-alto è una wavelet generata a partire da una wavelet madre. Ogni approssimazione differisce dalla più vicina di un fattore 2.

Le wavelet descrivono la differenza di informazione fra due approssimazioni adiacenti. La formula in notazione monodimensionale è:

$$f(x) = \frac{1}{\sqrt{M}} \sum_k W_\varphi(j_0, k) \varphi_{j_0, k}(x) + \frac{1}{\sqrt{M}} \sum_{j=j_0}^J \sum_k W_\psi(j, k) \psi_{j, k}(x)$$

$\varphi(x)$ è la funzione di scaling, con i relativi coefficienti di approssimazione. $\psi(x)$ è la wavelet madre, con i relativi coefficienti di dettaglio.



Tramite imposizione di equivalenze tra prodotti, è possibile misurare le variazioni lungo le righe, le colonne e le diagonali.

Trasformando con la wavelet, è possibile pesare ogni contributo per capire quali frequenze e bande hanno influenza maggiore, valutando le features che caratterizzano l'immagine.

Per calcolare gli edge, si ha il segnale di partenza in una somma di contributi: la parte interessata (per esempio il filtro passa alto ψ) viene tenuta applicando la

formula, mentre le altre componenti con frequenze basse vengono annullate.

Annullando anche i termini di edge orizzontali a tutte le scale e ricostruendo (sintesi) a partire da questi dati, si isolano i soli edge verticali.

Un altro approccio è il denoising, a partire dalle wavelet. La ricostruzione è effettuata dopo aver posto una soglia alta sui coefficienti a tutte le risoluzioni, accettando solo i valori al di sopra.

L'immagine originale è una risonanza magnetica con un rumore bianco additivo o moltiplicativo, a cui viene applicata la procedura di denoising:

1. Scelta della funzione wavelet;
2. Scelta del numero di livelli;
3. Sogliatura dei coefficienti di wavelet ed eliminazione di quelli inferiori alla soglia;
4. Ricostruzione a partire dall'immagine approssimata all'ultima scala.

Riducendo il rumore, si ha notevole perdita di dettaglio anche sugli edge, applicando soglie a tutte le risoluzioni. Agendo solo sulla risoluzione massima, si ha una perdita minore.

Per il principio di indeterminazione di Heisenberg, la distanza per ogni punto esatto di una funzione è infinitesima: pertanto, la risoluzione nelle frequenze è indefinita. Nel dominio trasformato, è nota l'informazione per ogni valore di f , ma δx è infinito.

In altre parole, in base al dominio, una dimensione è indeterminata. Il prodotto delle due, che corrisponde all'area, è sempre maggiore di un valore K :

$$\delta t \times \delta v \geq K$$

Nelle wavelet, è possibile determinare precedentemente una delle due dimensioni, avendo contemporaneamente informazione spaziale e frequenziale.

20 Tecniche di compressione

Il costo di un segnale dipende da campionamento e quantizzazione, fino ad arrivare a una quantità troppo elevata di dati per la memorizzazione ad alta qualità.

Per rappresentare l'informazione si ricorre alla compressione, cioè la trasformazione in un insieme di dati statisticamente incorrelati che garantiscano un grado di fedeltà (qualità) rispetto all'originale, e abbiano un accettabile peso computazionale per il tipo di applicazione.

Le tecniche di compressione si dividono in due grandi famiglie: lossy (con perdita accettabile a seconda dell'applicazione) e lossless (senza perdita). I dati ridondanti vengono rimossi, e il risparmio viene misurato tramite un rapporto di compressione in bit, dipendente appunto dalla ridondanza.

I dati sono gli strumenti tramite i quali è rappresentata l'informazione, e quest'ultima può essere associata a diversi quantitativi di dati. Il principio della compressione è appunto minimizzare il numero di bit utilizzati.

$$\text{rapporto di compressione } C = \frac{b_1 \text{ bit prima della compressione}}{b_2 \text{ bit dopo la compressione}}$$

$$\text{ridondanza relativa } R = 1 - \frac{1}{C}$$

Se $b_1 = b_2$, allora $R = 0$ e non ci sono dati ridondanti tra le due rappresentazioni. Un rapporto tipico di compressione è $10 : 1$, con corrispondente ridondanza 0.9 .

20.1 Ridondanza

Possono essere individuati diversi tipi di ridondanza, di cui almeno uno va ridotto:

1. Della codifica;
2. Spaziale o temporale (correlazione inter-campione);
3. Percettiva, imponendo quantizzazioni più o meno spinte.

I primi due consistono nella ridondanza statistica: i vicini possono essere correlati o dipendenti, quindi una parte dell'informazione è ripetuta.

20.1.1 Ridondanza della codifica

La ridondanza della codifica non introduce perdita, quindi è reversibile e ha una soglia massima di compressione che può essere raggiunta. Quest'ultima dipende dal tipo di segnale.

L'idea è costruire un quantizzatore che venga usato nel miglior modo possibile, distribuendo i valori in modo equiprobabile tra i livelli. Alcuni livelli, secondo l'istogramma a livelli di grigio, hanno una maggiore probabilità di essere occupati, quindi è possibile calcolare quanto comprimere in base al numero medio normalizzato di bit necessari per ogni frequenza.

Se i valori sono sbilanciati, si ricorre alla codifica a lunghezza variabile, con numero di bit medio necessario per descrivere l'immagine:

$$L_{avg} = \sum_{k=0}^{L-1} l(r_k) p_r(r_k)$$

$l(r_k)$ è il numero di bit necessario per descrivere il k -esimo livello r , con frequenza (probabilità) p_r . Essendo gli standard a 256 livelli di grigio, la frequenza sarà costante e la sommatoria sarà 1, pertanto le immagini saranno codificate a $M \times N \cdot 8$ bit.

VLC (Variable-Length Coding) è una strategia di riduzione della ridondanza che impiega un numero minor di bit per rappresentare i livelli più probabili, e viceversa. Il numero medio di bit è minore rispetto alla lunghezza fissa.

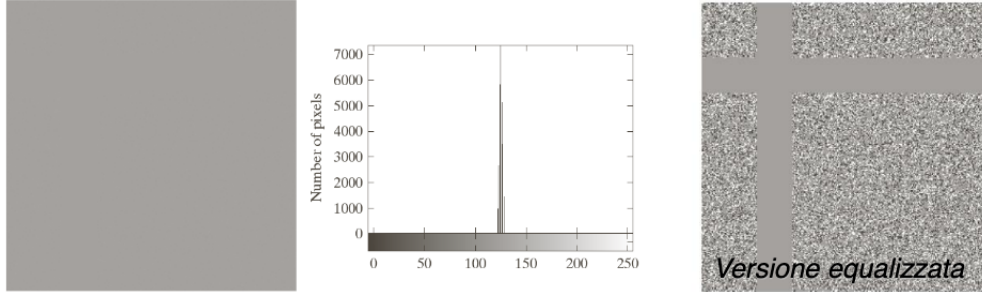
20.1.2 Ridondanza spaziale

Quando l'istogramma è uniforme, però, la VLC non ha effetto. Tutti i valori sono equiprobabili, ma ciò li rende fortemente correlati (e pertanto ridondanti) spazialmente. Questo succede per esempio in immagini a righe, mentre lungo le colonne i valori sono incorrelati.

Si può ridurre la ridondanza spaziale individuata introducendo coppie run-length: il primo componente della coppia individua l'intensità, mentre il secondo il numero di volte con cui si ripete. L'approccio è lossless.

20.1.3 Ridondanza percettiva

L'immagine è percepita come se avesse un valore di grigio uniforme, ma l'istogramma è discordante da questa rappresentazione.



Le informazioni ignorate dal sistema visivo umano possono essere eliminate, sostituite da un valore medio. Il processo è eseguito tramite quantizzazione, ma c'è perdita.

20.2 Entropia

L'entropia è la misura della quantità di dati minima necessaria per codificare senza perdita una sorgente di informazione, modellando la sorgente come processo probabilistico con eventi statisticamente indipendenti.

Informalmente, rappresenta il valore dell'informazione attraverso incertezza e probabilità di un certo simbolo. La sorgente deve emettere valori tra di loro incorrelati, quindi non deve avere memoria.

Nel caso delle immagini, c'è molto spesso dipendenza tra pixel contigui, ma l'entropia è la soluzione ottima a cui avvicinarsi il più possibile.

Un evento casuale E con probabilità $p(E)$ ha un grado di incertezza (e una quantità di informazione) pari a:

$$I(E) = \log \frac{1}{p(E)}$$

La base del logaritmo dipende dall'unità di misura dell'informazione, in questo caso 2 (bit). Se un evento è certo, l'incertezza è nulla.

Essendo l'entropia proporzionale all'incertezza, l'entropia di una sorgente con alfabeto $S = \{s_1, s_2, \dots, s_M\}$ è:

$$H(S) = \sum_{k=1}^M p_k \log_2 \frac{1}{p_k} = - \sum_{k=1}^M \log_2 p_k$$

p_k è la probabilità del k -esimo simbolo dell'alfabeto, e $\log_2 \frac{1}{p_k}$ è la quantità di informazione contenuta nel simbolo s_k , cioè il numero di bit necessario per la codifica. La situazione di equiprobabilità genera entropia massima.

L'entropia indica pertanto un limite inferiore (valore medio) per il numero di bit necessari per codificare un certo alfabeto, ipotizzando l'equiprobabilità di ogni livello e fornendo un punto di riferimento per i codici a lunghezza variabile.

L'obiettivo di VLC è trovare il codice di lunghezza minima per descrivere il segnale. Se un codice c_1, c_2, \dots, c_M ha parole con lunghezza b_1, b_2, \dots, b_M , il numero medio di bit richiesti è:

$$R = \sum_{k=1}^M b_k p_k$$

Il problema della progettazione di codice è la ricerca di parole con lunghezza media vicina all'entropia della sorgente del segnale, cioè r vicino a H , e la stima della precisione dei valori di probabilità.

La massima entropia implica un contenuto informativo visivo basso, essendo i valori equiprobabili. Immagini diverse possono avere entropia simile, ma essa serve come valore di riferimento solo con sorgenti senza memoria, quindi il rapporto di compressione può essere molto differente in caso di fotografie o immagini con pattern.

20.2.1 Codici di Huffman

La codifica di Huffman funziona come VLC: ai caratteri più frequenti vengono assegnate parole di pochi bit, quindi la lunghezza è inversamente proporzionale alla probabilità.

La compressione è senza perdita, e i codici possono essere costruiti con una struttura da albero binario, generando un codice ottimale minimizzando L_{avg} in modo che sia il più vicino possibile all'entropia della sorgente.

Se il segnale viene letto sequenzialmente, non si hanno informazioni precise riguardo alla distribuzione dei simboli: essa viene stimata tramite strumenti che aggiornano a ogni valore la possibile frequenza.

20.3 RLC

La correlazione tra bit può portare a un elevato risparmio, ma per ottenerla si deve ricorrere a sorgenti con memoria e algoritmi come RLC (Run-Length Coding).

Se ci sono raggruppamenti, è possibile sfruttare la memoria codificando il simbolo con la lunghezza del corrispondente gruppo. Il segnale si presenta con tanti 0 in sequenza intervallati da 1, e viene trasmessa la coppia di valore assunto e numero di valori.

Gran parte del contenuto informativo di un pixel è ridondante, e può essere ottenuto a partire dai pixel vicini passando a un mapping più efficiente, ma generalmente non interpretabile visivamente.

RLC considera singolarmente le linee dell'immagine, collezionando le coppie (g, w) che indicano rispettivamente il valore e la lunghezza.

Questa tecnica è impiegata per la codifica di immagini a colori, ma è poco efficiente per immagini reali a causa delle imprecisioni e del rumore.

Il bitmap prevede una modalità di compressione senza perdita utilizzando appunto RLC, codificando le informazioni sequenziali ripetute. Il rapporto di compressione può essere minore di 1 in caso che nuovi bit vengano sprecati per creare coppie di valori diversi da quelli non contigui.

20.4 Differential coding

Questa tecnica viene applicata quando RLC è inefficiente, e ha come scopo la riduzione della ridondanza in simboli consecutivi di un datastream. Nel caso di audio, si utilizza la variante DPCM (Differential Pulse Code Modulation).

PCM (Pulse Code Modulation) converte forme d'onda analogiche in segnali digitali attraverso:

- Campionamento;
- Quantizzazione;
- Codifica, in genere binaria.

Osservando il segnale in termine di differenza tra frequenze, una forte correlazione spaziale viene sfruttata tramite codifica a lunghezza variabile sui picchi dell'istogramma.

DCPM calcola la differenza tra frequenze adiacenti, ed essa viene codificata. Il segnale viene ricostruito sommando le differenze a partire dal valore iniziale (noto).

Nel caso delle immagini, si opera generando un'immagine differenza fra pixel contigui applicando un operatore che approssima la derivata prima (gradiente) o seconda (Laplaciana):

$$d(x, y) = I(x, y) - I(x - 1, y)$$

$$d(x, y) = 4I(x, y) - I(x, y - 1) - I(x, y + 1) - I(x + 1, y) - I(x - 1, y)$$

Per sfruttare la ridondanza tenendo conto delle differenze, è possibile anche sfruttare la differenza tra il valore attuale e quello predetto. Se la regione è uniforme in base a determinate regole, sarà immediato il calcolo del valore successivo, sfruttando l'interpolazione.

La differenza, con valori accurati, tenderà a 0 con un istogramma stretto, e di conseguenza un'entropia minore (MPEG) in presenza di ridondanza spaziale.

20.5 Algoritmi lossless

Lossless JPEG non usa la DCT e non introduce perdita utilizzando la codifica predittiva, sfruttando l'uniformità delle regioni.

Valuta le differenze tra il valore effettivo del pixel x e il suo valore predetto dai pixel adiacenti sopra e a sinistra. La sequenza di valori viene poi codificata con VLC, riportando il primo pixel identicamente.

Gli algoritmi lossless universali non richiedono la conoscenza a priori della distribuzione di probabilità dei simboli: in genere riescono a modellare dinamicamente le caratteristiche dei dati, adeguando la codifica.

Alcuni esempi di algoritmi sono Adaptive Huffman Coding, Lempel-Ziv e Arithmetic Coding. In funzione di com'è costruita l'immagine, ciascun algoritmo sarà più o meno efficiente.

20.5.1 LZW

LZW rimuove ridondanza di codifica e spaziale senza necessità di conoscenza a priori della probabilità di occorrenza dei simboli.

Viene inizialmente costruito un dizionario che contiene i simboli da codificare, con un ulteriore bit che permette di identificare l'occorrenza dei valori. Tutte le volte che un valore si ripete, la codifica può essere riutilizzata usando la coppia e associandola a un livello (analisi delle frequenze).

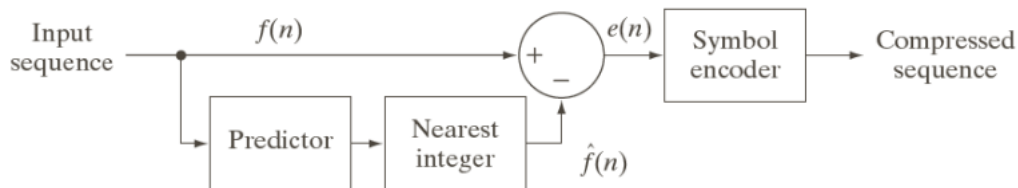
Dictionary Location	Entry
0	0
1	1
⋮	⋮
255	255
256	—
⋮	⋮
511	—

20.6 Codifica video

Nei video esiste una correlazione non solo tra i pixel dello stesso fotogramma, ma anche tra fotogrammi adiacenti. In base alla ridondanza, è possibile applicare:

- Compressione spaziale, intra-frame applicata a ciascuna immagine;
- Compressione temporale, inter-frame.

La codifica predittiva rimuove entrambe le tipologie, con un errore di predizione $e(n)$ ottenuto tramite VLC e diversi modelli di predizione.



20.7 Ridondanza percettiva

In presenza di un segnale sonoro a una generica frequenza, si ha un'alterazione della soglia di udibilità per le frequenze limitrofe al segnale, che non vengono percepite: non tutta l'informazione ha la stessa importanza.

La riduzione di ridondanza percettiva comporta quantizzazione, quindi è irreversibile e lossy dato che tiene conto solo delle componenti effettivamente udibili.

Gli schemi di compressione percettivi comprimono il segnale eliminando la parte non percepita dall'orecchio umano. Il masking può essere delle frequenze o temporale: quest'ultimo è effettuato saturando il sistema uditivo in modo da alterare le condizioni uditive (rumori forti alternati da deboli).

La ridondanza psicovisuale è un fenomeno legato alla legge di Weber: la sensibilità alle variazioni di luminosità diminuisce con la diminuzione della luminosità (immagine scura). Questo permette una maggiore quantizzazione nelle zone più scure.

L'occhio umano è più sensibile al rumore nelle regioni a bassa frequenza (uniformi) piuttosto che ad alta, e alle variazioni di luminanza piuttosto che cromatiche.

La quantizzazione IGS permette di convertire l'errore di quantizzazione alle basse frequenze in rumore ad alte frequenze, meno percepibile. L'effetto è l'eliminazione dei falsi contorni (dithering).

I canali cromatici possono essere compressi mantenendone l'intensità, per poi ricostruire in base a questa informazione.

21 Valutazione della qualità di compressione

Un sistema per la compressione di immagini è generalmente formato da unità strutturali distinte:

- Codificatore o compressore:
 - Di sorgente, che riduce la ridondanza del segnale operando sulle ridondanze spaziali e temporali, percettive e di codifica;
 - Di canale, che incrementa l'immunità al rumore;
- Decodificatore, che svolge le operazioni inverse:
 - Di sorgente;
 - Di canale.

Le tecniche lossless non consentono di raggiungere rapporti di compressione superiori a 10 : 1, effettuando solo mapping e codifica di simbolo senza quantizzazione.

La qualità della compressione si valuta una volta introdotta la quantizzazione nella fase di codifica, perché è l'unico caso in cui viene introdotta perdita di dati non ripetuti.

La perdita di informazione permette di raggiungere rapporti fino a 100 : 1, cercando un compromesso tra accuratezza e dimensioni. La qualità deve essere misurata quantitativamente secondo criteri oggettivi, considerando la presenza di ridondanze.

Alcuni criteri sono:

$$\text{Segnale errore: } e(xy) = \hat{f}(x, y) - f(x, y)$$

$$\text{Errore totale: } \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} e(x, y) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \hat{f}(x, y) - f(x, y)$$

$$\text{Root Mean Square error: } e_{rms} = \left[\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [\hat{f}(x, y) - f(x, y)]^2 \right]^{\frac{1}{2}}$$

$$\text{Signal to Noise Ratio: } SNR = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \hat{f}(x, y)^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [\hat{f}(x, y) - f(x, y)]^2}$$

Il segnale errore restituisce un'immagine. RMS, il più comune, serve a normalizzare l'errore totale, dato che sommare valori di segno opposto potrebbe risultare in un numero prossimo a 0. SNR è già adottato per la quantizzazione, ma non è modellabile: è definito come il rapporto tra il segnale originale e la differenza, da minimizzare.

Le metriche non tengono in considerazione la percezione soggettiva: un algoritmo può commettere errori specifici non individuabili dalle formule, ma soltanto dall'occhio umano: sono necessari test soggettivi con valutazioni della qualità su scala predefinita e correlata con valori oggettivi. A parità di errore, una diversa tecnica di compressione potrebbe essere migliore o peggiore.

21.1 Codifica con trasformate

La codifica con trasformate è un algoritmo lossy utilizzato per la compressione JPEG, che opera nel dominio trasformato.

Trasformata diretta e inversa:

$$T(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) g(x, y, u, v)$$

$$f(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} T(u, v) h(x, y, u, v)$$

La codifica attraverso i coefficienti della trasformata prepara alle successive fasi di compressione, in cui i coefficienti con ampiezza poco significativa sono quantizzati.

Il valore medio viene sostituito ai pixel che vengono rimossi: nel caso 2×2 ci sono 4 coefficienti che vengono sostituiti da un componente, ma all'aumento del numero la correlazione tra pixel adiacenti diminuisce fino a rendere impossibile la riduzione senza perdite significative.

Diverse tipologie di coppie di trasformata e antitrasformata impiegabili:

- KLT, analisi alle componenti principali, decorrela i dati ma computazionalmente inefficiente;
- DCT, approssima meglio il comportamento ottimo;
- DFT;
- ...

Viene utilizzata la trasformata lineare reversibile coseno 8×8 con eliminazione di una parte dei coefficienti in base all'ampiezza, perché è il metodo che assicura un errore minore approssimando meglio il segnale a parità di numero di coefficienti.

La dimensione n della sotto-immagine è una potenza di 2 per semplicità computazionale, e all'aumentare di n viene persa efficienza fino a un massimo valore di 20. Per questo motivo si preferisce 8 a 16.

64 bit permettono di rappresentare un numero di colori sufficiente, e al crescere di n diminuisce la correlazione tra i pixel.

21.2 Compressione wavelet

I coefficienti della trasformata decorrelano il contenuto informativo dei pixel di un'immagine, permettendo una codifica più efficiente anche se con un maggior numero di operazione.

La maggior parte dell'informazione è contenuta in un numero limitato di coefficienti, e i restanti possono essere opportunamente quantizzati con scarsa influenza sulla distorsione.

Scelto il livello J di analisi, la trasformata wavelet decompone l'immagine evidenziando i coefficienti legati alle caratteristiche di ogni dimensione, effettuando un'analisi statistica che evidenzia contenuto informativo legato alle basse frequenze.

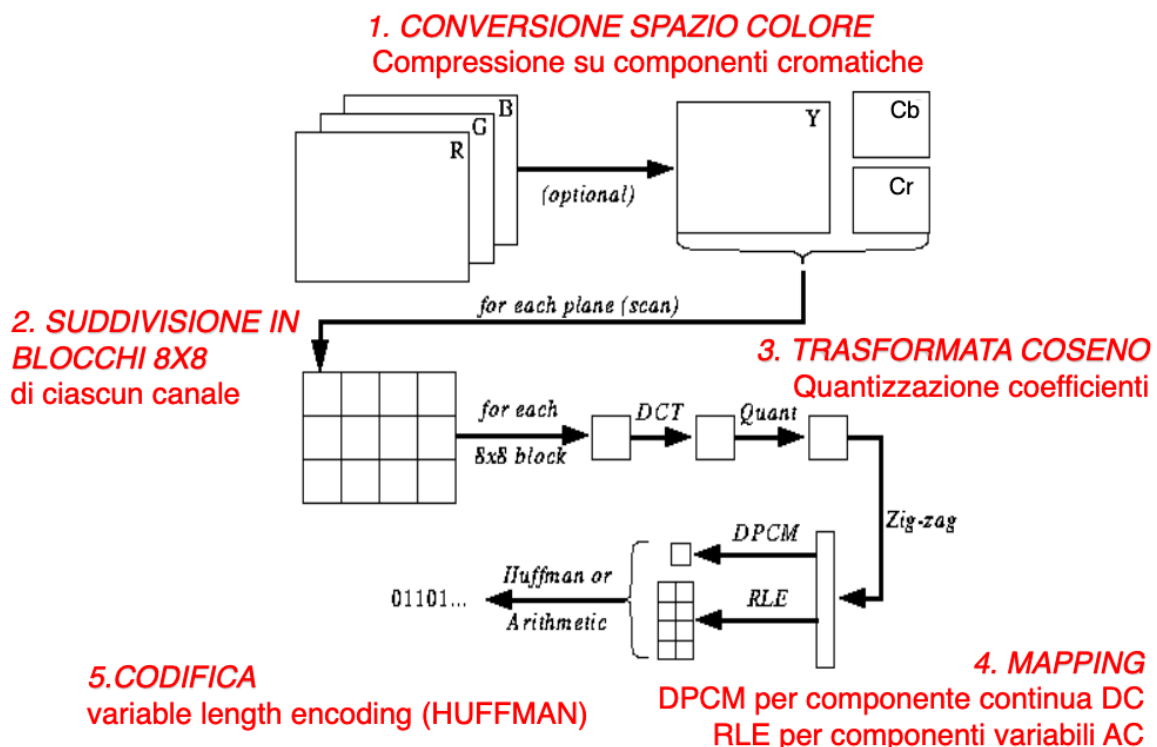
Al crescere del rapporto di compressione, si ha una crescente perdita dei dettagli, in particolare texture ed edge, con conseguente smoothing, ma la qualità rispetto a DCT rimane superiore grazie all'assenza di blocchettizzazione.

Il numero di operazioni dipende dalla metodologia utilizzata (Haar, Daubechies, symlet, C-D-F) ed è proporzionale all'efficienza. Un altro fattore importante è il numero di livelli della decomposizione wavelet, generalmente non superiore a 3.

22 Compressione JPEG

JPEG (Joining Photographic Experts Group) è lo standard di memorizzazione delle immagini fotografiche, che permette un'elevata compressione lossy con accettabile degradazione della qualità.

JPEG definisce una serie di elaborazioni flessibili da seguire sulle immagini, che possono eventualmente essere saltate. Le regole sono però definite in modo rigido, includendo la decompressione.



La compressione JPEG si articola in 5 fasi:

1. Conversione spazio colore, comprimendo le parti cromatiche;
2. Suddivisione in blocchi 8×8;
3. Trasformata coseno e quantizzazione coefficienti;
4. Mapping con DCPM e RLE;
5. Codifica con Huffman.

La perdita avviene nelle fasi di sottocampionamento del chroma e quantizzazione.

22.1 Conversione spazio cromatico

Si può sfruttare la caratteristica psicovisuale per cui il sistema visivo umano è più sensibile alle variazioni di luminanza che non a quelle di cromaticanza, quantizzando maggiormente queste ultime.

In questo modo, la ridondanza psicovisuale diminuisce, con poca perdita. L'immagine viene convertita da RGB a un altro spazio come YCbCr, esprimendo il colore in termini di $x : y : z$ e campionando orizzontalmente.

La compressione lossy sulle coordinate cromatiche è opzionalmente effettuata riducendo le dimensioni sottocampionando tramite media a due a due tra i pixel adiacenti.

22.2 Suddivisione in blocchi

Ciascun canale (Y, Cb, Cr) dell'immagine è diviso in blocchi quadrati, di cui ciascuno è costituito da 8×8 pixel.

22.3 Analisi in frequenza

L'idea per la quantizzazione si basa sull'utilizzo di una matrice dei coefficienti che genera coefficienti quantizzati. Le frequenze vengono pesate diversamente tagliando le alte frequenze, quindi dividendo con valori più alti, e salvando le basse applicando valori bassi.

In JPEG viene utilizzata la DCT, e i coefficienti rappresentano le ampiezze dei segnali armonici che sommati ricostruiscono il segnale.

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) g(u, v, x, y)$$

$$f(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) \cos \left[\frac{(2x+1)\pi u}{2N} \right] \cos \left[\frac{(2y+1)\pi v}{2N} \right]$$

$$g(x, y, u, v) = \alpha(u) \alpha(v) \cos \left[\frac{(2x+1)\pi u}{2N} \right] \cos \left[\frac{(2y+1)\pi v}{2N} \right]$$

Per ogni canale, a ogni blocco nello spazio corrisponde un blocco di coefficienti nel dominio della frequenza, in cui le basse frequenze sono in alto a sinistra e le alte sono in basso a destra. Il primo coefficiente del blocco trasformato è correlato al valore medio dei valori del blocco originario.

Viene poi definita una tabella di quantizzazione Q che divide la matrice dei coefficienti F e genera coefficienti quantizzati:

$$\hat{F}(u, v) = \text{round}\left(\frac{F(u, v)}{Q(u, v)}\right)$$

Le tabelle $Q(u, v)$ sono calcolate tramite esperimenti psicovisuali, scegliendo numeri che permettono una minore complessità computazionale.

Poiché i valori della tabella di quantizzazione sono abbastanza elevati, i coefficienti quantizzati sono più piccoli e con una varianza minore. Si può variare il rapporto di compressione riscalandolo la tabella di quantizzazione (quality factor), che vengono poi moltiplicati in fase di decodifica:

$$\tilde{F}(u, v) = \hat{F}(u, v)Q(u, v)$$

I coefficienti quantizzati sono ottenuti arrotondando all'intero più vicino, e quelli meno significativi tendono ad azzerarsi.

Nelle regioni disomogenee (texture), i valori adiacenti sono distinti fra loro, e i coefficienti della matrice saranno in quantità maggiore e molto distanti dallo zero. A parità di coefficienti, la perdita è più elevata. La qualità, in generale, è buona con 16 coefficienti diversi da zero.

22.4 Mapping dei coefficienti

Il tipo di mapping impiegato si differenzia per i coefficienti supersintiti, che possiedono una componente continua, indicata come DC, e alcune variabili, AC. Queste ultime vengono ordinate tramite zig-zag scan e codificati con RLE.

Sul valore DC invece viene applicata una tecnica DCPM (Differential Pulse Code Modulation), che codifica ogni blocco come differenza rispetto al valore del blocco precedente, sfruttando la correlazione statistica.

$$\delta_k = DC(0, 0)_k - DC(0, 0)_{k-1}$$

22.5 Codifica dei dati

L'ultima codifica applicata ai dati utilizza VLE (lunghezza variabile) in base all'entropia. Viene analizzata la frequenza statistica di ciascuna parola e a ognuna è assegnata una stringa del codice.

$$p_r(r_k) = \frac{n_k}{n} \quad L_{avg} = \sum_{k=0}^{L-1} l(r_k)p_r(r_k)$$

- Codifica sequenziale: i blocchi DCT sono codificati e trasmessi uno dopo l'altro, alternando le diverse componenti a colori (pagine web);
- Codifica progressiva: vengono trasmesse prima le porzioni più significative, visualizzando un'anteprima con qualità ridotta.

Per misurare la qualità della compressione, si usa il rapporto di compressione $C = \frac{bit_1}{bit_2}$ oppure il numero di bit per pixel. Questi valori sono dipendenti dal fattore di qualità.

22.6 JPEG-20

La compressione con bit-stream ha una bassa complessità ed è efficiente, ma non esistono regioni di interesse e non è possibile differenziare la qualità.

JPEG-20 permette diverse risoluzioni, evitando i blocking artifact sfruttando una decomposizione wavelet a 3 livelli. A parità di fattore di compressione, l'immagine appare con meno blocchetti, quindi migliore.

Le zone più importanti, o meno uniformi, possono essere priorizzate in termini di qualità, applicando una maggior compressione sullo sfondo.

23 Compressione video

La compressione video deve necessariamente essere lossy, dato che la quantità di ridondanze è generalmente elevata e le dimensioni dei video non compresse sono ingestibili dai supporti di memoria. Il tipo di processing adottato è intra-frame o inter-frame.

Ci sono soluzioni che considerano solo la ridondanza spaziale o temporale, e altre che sfruttano la ridondanza percettiva attraverso le limitazioni del sistema visivo/uditivo.

Il procedimento più semplice applica codifiche predittive oppure JPEG frame per frame, ma queste tecniche sono legate solo alle caratteristiche spaziali: esse devono essere estese valutando anche la ridondanza temporale, in termini di differenze tra i frame con predizioni (istogramma piccato).

Il primo step è naturalmente la differenza tra frame consecutivi, a cui è associata un'entropia: ciò funziona in modo ottimale quando le zone delle immagini cambiano poco, ma non è efficace quando la camera non è perfettamente ferma, o ci sono cambi di luce, o le scene sono in movimento.

Generalmente anche in presenza di queste condizioni c'è correlazione, ma se la differenza tra due fotogrammi è il rumore o è dovuta al cambio di scena il segnale è poco correlato. In alcuni casi è possibile predire anche il movimento.

23.1 Motion compensation

La motion compensation è una codifica predittiva che consiste nella stima dei vettori di moto con cui le porzioni di frame vengono spostati, e nella loro applicazione per le variazioni future.

Lo scopo è l'individuazione e la compensazione del moto, studiando il disallineamento dei pixel. La stima riguarda gli spostamenti dei macro blocchi 16×16 , ma questa tecnica non funziona in caso di differenze significative.

Tipologie di frame:

- Target frame, da codificare;
- Reference frame, utilizzato per il calcolo delle differenze;
- I-frame (intra/independent frame, compressi con JPEG), frame di partenza per la predizione;
- P-frame, risultati:

- Predetti solo in base al reference frame;
- Predetti in base al frame precedente e ai frame successivi.

Il disallineamento fra il reference frame e il target è chiamato MV (Motion Vector). Utilizzare le informazioni dei frame consecutivi aumenta la precisione.

Ogni gruppo non sovrapposto di pixel viene assegnato a un vettore di moto comune. L'intensità è 16, con 8 canali cromatici ($1/4$) per semplificare il campionamento $4 : 2 : 0$.

Dato un macroblocco del target frame, viene cercato il più simile macroblocco nel reference frame tra i precedenti e successivi. Le dimensioni della finestra di ricerca variano in base allo standard utilizzato. Il MV consiste nelle sottrazioni, espresse tramite coppie di valori sugli assi.

L'encoder calcola la somiglianza tra il frame predetto e quello ottenuto, detta motion compensated difference frame $MCD(n)$. MCD ha una correlazione fra pixel inferiore a quella risultante dalla differenza tra target e difference.

Il decoder prende in input MCD, il frame reference e i vettori di moto, per ricostruire il frame predetto e in seguito il frame target, con perdita. I blocchi predetti sono organizzati su una griglia, al contrario di quelli utilizzati per la predizione.

Le possibilità per la compressione sono la codifica della differenza oppure la compensazione della differenza attraverso il moto.

Riassumendo:

1. Individuazione del target frame;
2. Scorrimento dei macroblocchi e ricerca del più simile al blocco del reference frame, cercando il vettore di moto;
3. Sostituzione con il più simile traslato di un vettore di moto (compensazione);
4. Calcolo della differenza tra predizione e target (errore);
5. Codifica della differenza.

La compensazione del moto viene applicata in modo efficiente ai canali di luminanza, mentre i valori di cromaticanza sono meno correlati: i vettori di moto si ottengono sottocampionando i risultati sulla luminanza.

Se il moto è da sinistra verso destra, le zone di sinistra hanno l'errore di predizione più alto perché sono zone che non esistono nei frame precedenti.

Il vettore di moto di ciascun macroblocco è calcolato trovando nel reference frame la posizione più probabile di quel macroblocco del target frame: il reference frame può essere tra i precedenti o i successivi. Un frame codificato usando entrambi si indica come B-frame (bidirectional).

La posizione più probabile è quella che minimizza una misura di errore tra reference e target, all'intorno della regione definita.

Si definisce MAD la distorsione media assoluta tra pixel di due macroblocchi di dimensioni $N \times N$ disallineati, da minimizzare.

La stima del moto con frame successivi implica un ritardo nei calcoli finché le informazioni sono incomplete. Per semplificare le operazioni, la ricerca del macroblocco è effettuata in un intorno, e non ha standard definiti:

- Ricerca logaritmica, valutando la MAD su 9 punti (3×3), selezionando il più vicino e dividendolo a sua volta, fino ad arrivare a un errore tollerabile, nonostante il rischio di concentrazione sui punti locali tralasciando il possibile ottimo globale;
- Multirisoluzione, in tre passi;
 1. Stima iniziale sul frame a risoluzione minore;
 2. Raffinazione a risoluzione intermedia;
 3. Determinazione del vettore di moto finale sulla risoluzione massima.

23.2 Standard per la codifica

Gli standard di codifica video digitale sono regolati da:

- ITU (raccomandazioni), con un carattere generale e orientato alle applicazioni real-time (H.xx);
- ISO, MPEG e derivati, per necessità più ampie come storage e broadcast.

Gli IP frames sono particolari sequenze di frame, la cui combinazione è la base di H-261:

- I (intra) servono per predire i frame successivi, codificati con JPEG con riduzione della ridondanza e DCT;
- P (picture), codificati tramite MC e predizione, e in seguito DCT insieme al vettore di moto.

La sequenza è I-P-P con finestra temporale stabilita dall'encoder e disallineamento misurato in pixel.

Nella quantizzazione per coefficienti DCT, non vengono applicate matrici di quantizzazione ma si usa una costante per tutti i coefficienti, calcolata per approssimazione tra DCT e step-size.

Il rapporto di compressione è più alto perché questa metodologia non sfrutta solo la ridondanza spaziale, ma riduce anche quella temporale. I frame B sono quelli su cui la perdita è maggiore, perché vengono codificati agendo su entrambe le direzioni.

Non sempre i macroblocchi sono codificati nello stesso modo: a seconda del tipo di frame e dell'errore di predizione essi saranno determinati diversamente. Se nessuna strategia è la migliore, si utilizza l'interpolazione.

23.2.1 MPEG

MPEG (Motion Picture Experts Group) è uno standard creato da ISO per la codifica di informazioni audio e video in formato digitale compresso, con bitstream predefiniti.

MPEG-1 è il primo standard per i filmati digitali, non interlacciato. I blocchi hanno dimensioni 8×8 , e l'unità elementare di codifica è il macroblocco 16×16 .

I dati hanno una struttura gerarchica:

- Più macroblocchi formano una slice, codificata in modo indipendente con differenti fattori di scala, senza propagazione dell'errore;
- Un'immagine del frame codificato consiste in una picture;
- Più pictures sono definite GOP;
- Più GOP sono inclusi in una sequence.

La predizione è solamente di tipo forward, e ci sono casi in cui il frame precedente non è sufficiente per una buona stima del vettore di moto: si introduce backward prediction.

Analogamente a IP, lo standard MPEG definisce tre tipi di picture:

1. I, codificate sfruttando la ridondanza;
2. P, codificati con predizione compensata del moto rispetto al picture precedente;
3. B, codificati con predizione compensata del moto rispetto al picture precedente, successivo o entrambi (buffering).

Per ogni B-frame sono calcolati due vettori di moto per ogni macroblocco: se il matching è buono, viene fatta una media delle predizioni. La compressione di I è peggiore, mentre B ha priorità minore ma compressione maggiore.

Non sempre i macroblocchi sono codificati allo stesso modo: l'encoder sceglie l'approccio più adatto in base all'informazione contenuta.

- I sono intra-coded, con matrice di quantizzazione variabile;
- P sono codificati come intra se non è possibile avere una predizione sufficientemente accurata;
- B sono determinati dal precedente, dal successivo o dalla media, a seconda dell'errore di predizione.

MPEG-2 è un'evoluzione di MPEG-1, con la stessa struttura base ma dedicato a video di qualità superiore, con bitrate maggiori di 4 Mbps. Questo standard implementa diverse misure di adattabilità con livelli di decodifica in base ai frame rate, pertanto è più scalabile.

Le versioni sono a qualità variabile, trattando i coefficienti secondo la perdita indicata. I video sono interlacciati, gestendo i field (righe dispari e pari), ottenendo un errore più basso. L'approccio può essere frame prediction o field prediction, codificando top e bottom separatamente con due vettori di movimento.

L'utilizzo dei field ha un effetto diretto sulla quantizzazione: il campo è 16×8 , il pattern a zig-zag è opzionale e sono ammessi anche pattern alternativi. Il sottocampionamento è solo in una dimensione, essendo i sottoinsiemi incorrelati.

Non tutti i livelli ammettono le stesse variazioni: esse cambiano in base alla dimensione e i frame al secondo, campionando con diversi profili di chroma.

Il concetto di scalabilità prevede alcune caratteristiche:

- Livelli base a qualità inferiore;
- Livelli base indipendenti per fornire un video di qualità inferiore;
- Codifica e decodifica dei livelli superiori dipendenti dai precedenti.

I layers sono definiti in base ad alcune possibilità:

- SNR, con differente quantizzazione a seconda del livello;
- Scalabilità spaziale, dividendo i dati relativi alle immagini sotto-campionati da quelli che permettono di ricostruire l'immagine completa;
- Temporale;
- Partizionamento dei dati con le frequenze.

Questi modi possono eventualmente essere combinati in una scalabilità ibrida.

MPEG-4 pone l'attenzione sull'interattività con l'utente, introducendo ulteriori caratteristiche basate sugli oggetti e le loro relazioni (anche tridimensionali). La trasmissione è possibile anche su mezzi a banda limitata, garantendo efficienza e accesso universale.

Esistono anche possibili oggetti precodificati, come le animazioni che permettono di inserire effetti speciali nei video. Ogni oggetto può essere trasmesso in uno stream separato, ed essi sono ricomposti dopo la decodifica. Tutti i componenti di una presentazione sono accessibili tramite Object Descriptors, posti in mutua relazione tramite scene description e relative informazioni spatio-temporali. La ricostruzione è possibile grazie a BiFS, linguaggio di descrizione.

Alcuni standard di decompressione introducono un filtro di denoising, applicato dopo la decodifica per rimuovere gli artefatti di blocchettizzazione.

MPEG-7 definisce gli strumenti necessari per descrivere il contenuto audiovisuale dei dati multimediali e consentire l'interoperabilità tra diversi sistemi e applicazioni.

Permette la migliore gestione, distribuzione e fruizione delle descrizioni, aiutando utenti e applicazioni ad identificare e filtrare le informazioni rilevanti.

Ogni scena viene caratterizzata e classificata secondo un numero di features che impongono un grado di correlazione tra gli oggetti, tramite sintassi e semantica.

Ogni descrittore è una rappresentazione di una feature, per caratteristiche di basso livello. Può essere atomico o composito, e ogni feature può averne uno o più.

24 Formati immagine

Il formato è la regola con la quale una descrizione dell'immagine è memorizzata in un file elettronico, che comprende sia i dati che le modalità di lettura e interpretazione.

24.1 Bitmap

Bitmap è il formato più indicato per le immagini a 3 canali di 8 bit, ma supporta anche profondità differenti. Il numero di bit necessari dipende dal numero di pixel, con un fattore moltiplicativo di 24.

A livelli di grigio utilizza i tre colori uguali, memorizzando le immagini dal basso a sinistra riga per riga, con header e mappa di colori.

La compressione non introduce perdita, utilizzando RLE: questo metodo può eventualmente fallire in caso di pattern disomogenei.

24.2 GIF

GIF (Graphics Interchange Format) permette la sovrapposizione tra immagini, attivando la trasparenza uniforme 0-1. Può contenere più immagini, quindi può essere utilizzato per le animazioni.

Si usa quando l'immagine originale è a scala di colore, con un massimo di 256 colori. La compressione è senza perdita, ma la limitazione di 256 colori (8 bit) quantizza pesantemente il numero di combinazioni a disposizione.

24.3 TIFF

TIFF (Tagged Image File Format) gestisce immagini con contenuti ad altissime frequenze, sfruttando algoritmi che cercano ripetizioni di pattern (loghi) a discapito delle dimensioni.

Permette di salvare caratteristiche aggiuntive memorizzando le informazioni colore in spazi diversi, definendo profili standard per allineare risposte di devices differenti.

Può essere salvato con o senza compressione, in caso adottando conservazione come LZW.

TIFF contiene meta informazioni in locazioni di memoria chiamate tag, che includono per esempio risoluzione, compressione o modello di colore.

24.4 PNG

PNG (Portable Network Graphics) è una valida alternativa a GIF con diverse profondità colore, fino a 48 bit. Un ulteriore passaggio è l'alpha-channel per gestire la trasparenza, estendendo 0-1 a più livelli.

PNG è utile per lo scambio di informazioni su internet, senza la limitazione colore. La compressione è lossless con dimensioni inferiori rispetto a TIFF, utilizzando un alpha channel per la trasparenza.

L'alpha-channel assume valori tra 0 e 1, e indica le modalità di blending tra più immagini, secondo la formula:

$$\alpha I_A + (1 - \alpha) I_B$$

24.5 JPEG

JPEG (Joint Photographic Experts Group) memorizza immagini fotografiche con perdita di informazioni, tramite RGB, CMYK o scale di grigio.

Non supporta la trasparenza e tratta solo immagini statiche, mentre per i filmati viene usato MPEG. Come formato intermedio per una sequenza di passi di manipolazione non è indicato, mentre TIFF è ottimale.

24.5.1 JPEG-2000

JPEG-2000 è basato su tecnologie wavelet per una migliore compressione, senza artefatti a blocchi. Ha il supporto per multi-spectral imagery e la capacità di ricostruzione progressiva dell'immagine codificando regioni di interesse.

24.6 Postscript

Postscript è basato su una descrizione vettoriale, definendo gli elementi di pagina come vettori. Immagini bitmap possono essere inclusi nei file, insieme a informazioni aggiuntive memorizzate senza compressione (ASCII).

24.7 PDF

PDF (Portable Document Format) consiste in una o più pagine che conservano font, formattazione e colori del sorgente. Può contenere grafica vettoriale, immagini, video e link a ipertesti.

Supporta la compressione JPEG, RLE e LZW, con profili ICC.

24.8 SVG

SVG è un linguaggio per la grafica 2D che sostituisce Postscript, per visualizzare sul web immagini vettoriali, bitmap e testo.

24.9 EXIF

EXIF è un formato per immagini da camera digitale. Contiene numerosi tag, come TIFF, con informazioni sulla camera e sulle condizioni di acquisizione. Lo standard EXIF include anche specifiche per l'audio.

I tag si suddividono in:

- Tag che riguardano la fotocamera;
- Tag che riguardano l'immagine;
- Altri tag.

25 Compressione audio

La compressione audio è legata a strategie predittive tramite interpolazione dei valori o modelli basati su come alcuni suoni possono apparire (es. parlato), senza introdurre perdita e con rapporti di compressione dipendenti dal tipo di audio: il silenzio può essere eliminato, e i suoni artificiali vengono predetti più difficilmente.

Un modo più intelligente tiene conto del sistema percettivo, eliminando le frequenze fuori dal range di udibilità applicando un filtro passa-banda e rimuovendo gli alias.

Un altro aspetto da sfruttare riguarda gli altri meccanismi dei sistemi umani: il mascheramento e la non linearità degli effetti.

MPEG coding algorithm: il tono mascherato viene modificato una volta individuata la banda con il tono più alto, osservando la potenza nelle bande vicine. Considerato un frame, si trova il tono mascherante, e in base a esso si osserva come viene modificata la soglia di udibilità in modo da poter eliminare i suoni non percepiti.

La codifica di Huffman viene in seguito applicata