Scuola universitaria professionale della Svizzera italiana
Dipartimento tecnologie innovative
Istituto sistemi informativi e networking

SUPSI

Tecniche base di audio processing

Fondamenti di Multimedia Processing

Tiziano Leidi

05.10.2018

Processing di segnali audio

L'audio signal processing comprende lo studio, lo sviluppo e l'utilizzo di tecniche (di signal processing) per modificare in maniera intenzionale segnali audio.

In passato, si utilizzavano processori audio analogici di tipo elettronico (rari casi, come il phaser, utilizzavano tecniche meccaniche). Oggi, vengono impiegati soprattutto processori audio digitali.

Dal punto di vista del trattamento dei segnali, l'avvento del digitale ha introdotto molte nuove possibilità, ampliando le possibili forme di processing realizzabili.

Processing di segnali audio

Le tecniche di audio processing più comuni comprendono:

- equalizzazione e impiego di altri tipi di filtri
- modulazioni di ampiezza o di frequenza
- compressione, espansione e limiting
- aggiunta di delay e riverberi
- chorus, flanger, phaser
- pitch-shift e time-stretching
- effetti audio 3D

Alcuni dei possibili effetti vengono ottenuti con dei semplici trucchi. Altri richiedono l'impiego di tecniche avanzate di signal processing. Discuteremo l'impiego dell'FFT e dei filtri FIR/IIR.

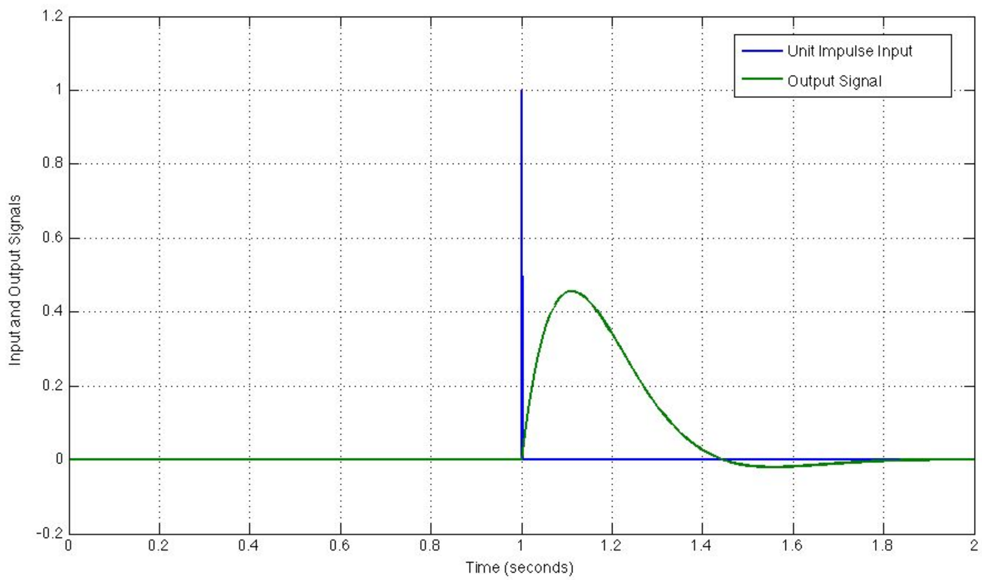
Processing di segnali audio

Per poter operare nell'ambito del processing di segnali audio sono necessari alcuni strumenti di analisi ...

Risposta impulsiva (impulse response)

La risposta impulsiva di un sistema è una funzione (del tempo) che coincide con l'output del sistema se sollecitato con un impulso.

Un impulso è un segnale idealmente istantaneo e di massima ampiezza (teoricamente infinita). In ambito analogico viene detto Dirac delta function, mentre in ambito digitale Kronecker delta function.



Risposta in frequenza

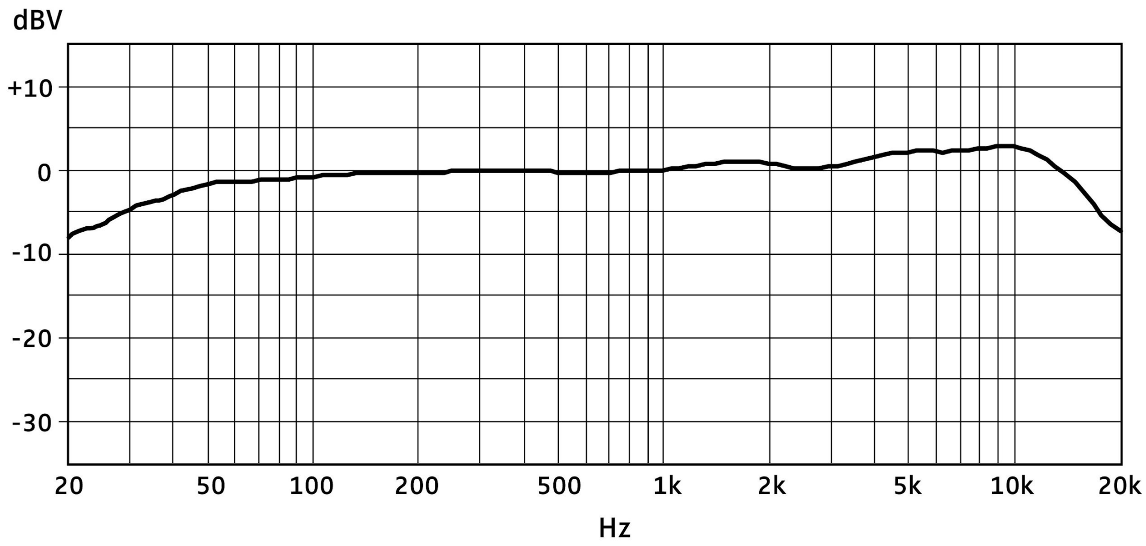
La risposta in frequenza è una misura quantitativa dello spettro di output di un sistema, utilizzata per caratterizzare la dinamica del sistema.

Solitamente la si misura sollecitando il sistema con uno sweep di ampiezza costante o con un segnale con ampio spettro di frequenze (come il rumore bianco).

Risposta in frequenza

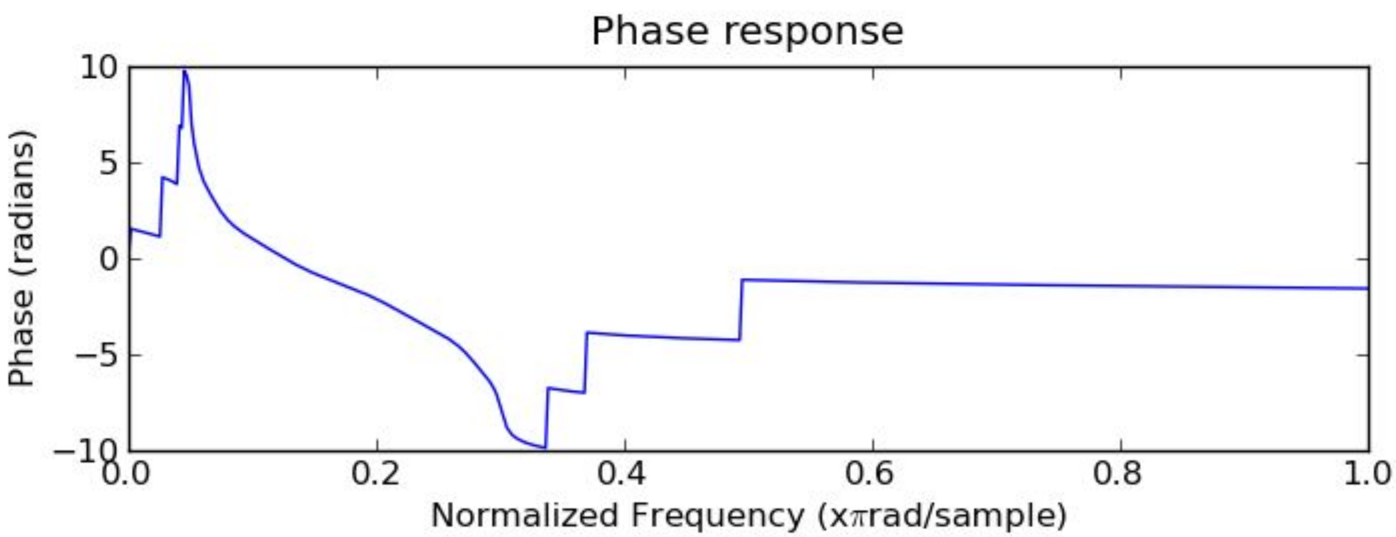
Per un sistema audio, una risposta in frequenza uniforme (piatta), significa assenza di alterazioni sonore (colorazione o distorsione).

Ad esempio, un amplificatore audio ad alta fedeltà ha di norma una risposta in frequenza piatta fra i 20 Hz e i 20,000 Hz, con tolleranze di soli ±0.1 dB nel range delle frequenze medie.



Risposta in fase

In maniera simile alla risposta in frequenza è possibile rappresentare la risposta in fase: mette in relazione le differenze di fase fra il segnale di input e quello di output di un sistema.



Trasformata di Fourier

I segnali possono essere continui o campionati, ma sono in ogni caso composti da numeri reali. Però nel signal processing si lavora spesso con numeri complessi.

Alcune caratteristiche dei segnali vengono meglio comprese se si opera nel dominio delle frequenze, dove ogni singola frequenza (sinusoide) è definita dalla sua ampiezza e dalla sua fase.

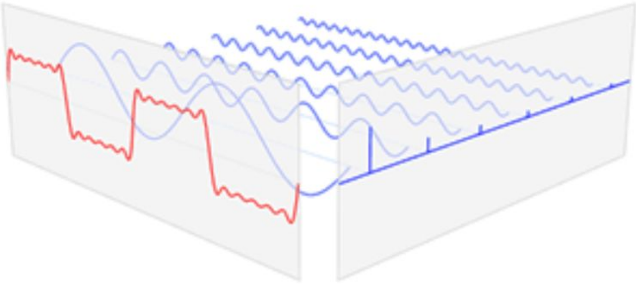
Ampiezza e fase vengono combinati dagli ingegneri in un unico numero complesso: *a cos(phi) + i a sin(phi).*

Quest'approccio introduce la trasformata di Fourier ...

Trasformata di Fourier

La trasformata di fourier scompone una funzione del tempo (quindi un segnale) nelle frequenze che la compongono. Il risultato è una funzione di numeri complessi, il cui valore assoluto rappresenta l'ampiezza di ogni frequenza del segnale, mentre l'argomento complesso è la differenza di fase della sinusoide di quella determinata frequenza.

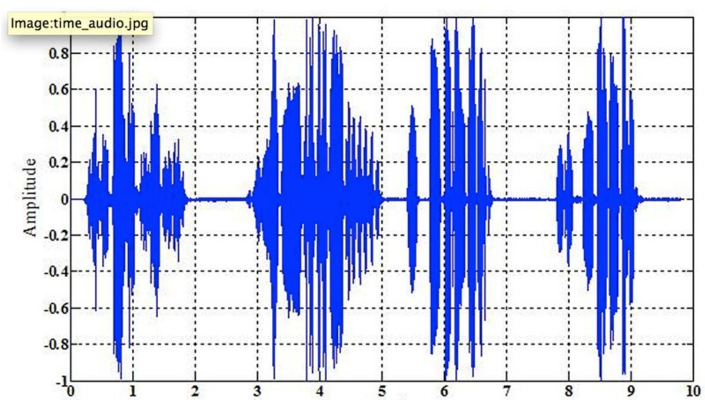
$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i x \xi}\, dx$$



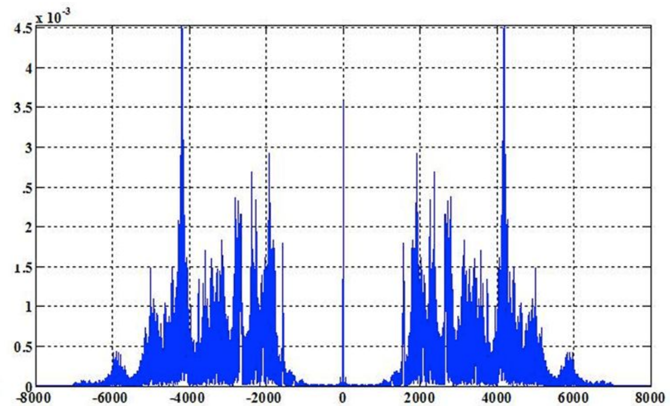
x rappresenta il tempo (in secondi), mentre la variabile ξ rappresenta la frequenza (in hertz)

Trasformata di Fourier

La trasformata di fourier permette quindi di passare dal dominio del tempo al dominio delle frequenze.



Time domain



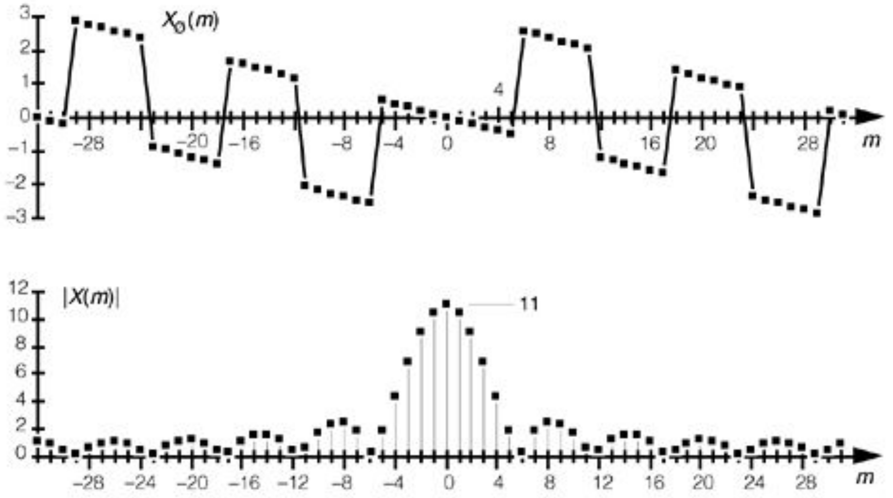
Frequency domain

La sua inversa viene chiamata sintesi di Fourier:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}\,(\xi)\; e^{2\pi i \xi x}\, d\xi$$

Trasformata discreta di Fourier

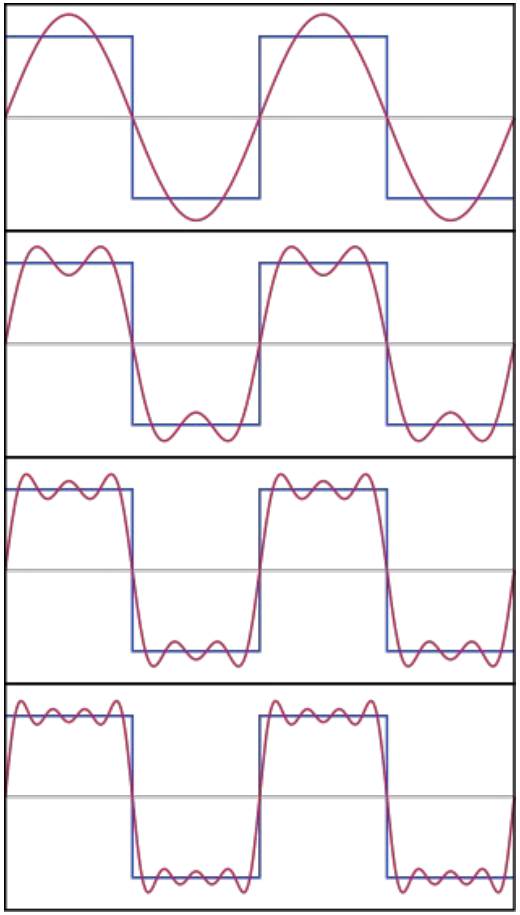
In digitale, per il calcolo della trasformata di Fourier viene utilizzata la trasformata discreta di Fourier - discrete Fourier transform (DFT) - che converte: una sequenza finita di campioni di una funzione, in una sequenza (della stessa lunghezza) di campioni della trasformata di Fourier. L’intervallo tra ogni frequenza è il reciproco della durata della sequenza in input.



Trasformata discreta di Fourier

L'inversa della DFT è una serie di fourier, che usa i campioni ottenuti tramite la DFT come coefficienti delle sinusoidi delle frequenze:

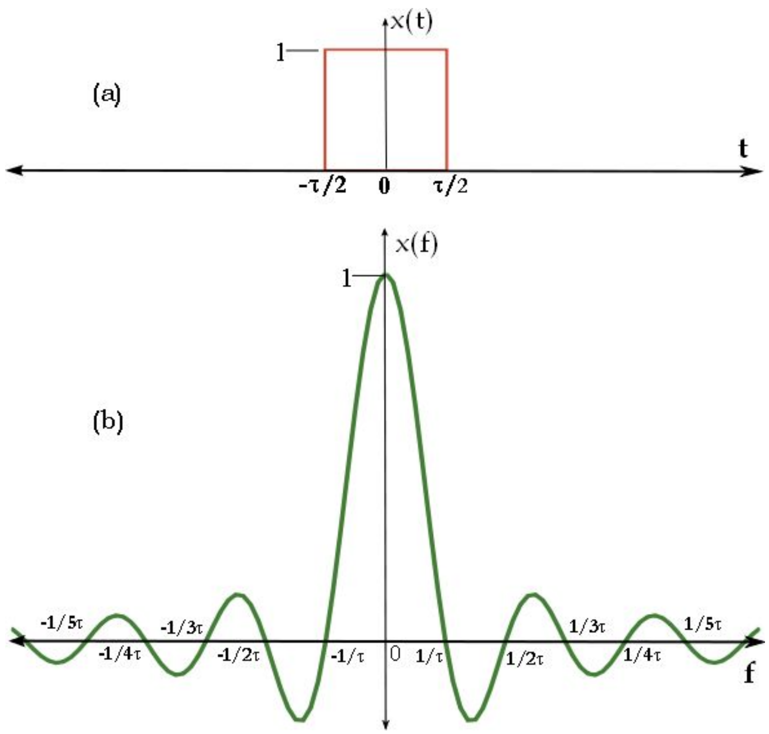
$$s_N(x) = \frac{A_0}{2} + \sum_{n=1}^{N} A_n \cdot \sin\left(\frac{2\pi n x}{P} + \phi_n\right)$$



Per il calcolo della DFT viene normalmente impiegato l'algoritmo di Fast Fourier Transform (FFT), che è particolarmente efficiente in termini di tempi di calcolo.

Trasformata di Fourier

Segnali che sono localizzati nel tempo hanno trasformate di Fourier che sono distribuite nel dominio delle frequenze e viceversa.



Trasformata di Fourier

Le operazioni lineari in uno dei due domini (tempo o frequenza) hanno corrispondenti operazioni nell'altro dominio, che sono spesso più facili da eseguire.

$$\begin{tikzcd}
\text{(problem)} \arrow[r, "\text{Fourier transform}"] \arrow[d, dotted, "\text{solve}"', "\text{(hard)}"] & \text{(transformed problem)} \arrow[d, "\text{solve}"', "\text{(easy)}"] \\
\text{(solution)} & \text{(transformed solution)} \arrow[l, "\text{inverse Fourier transform}"']
\end{tikzcd}$$

Ad esempio la convoluzione (un'operazione spesso utilizzata in audio processing) nel dominio del tempo corrisponde alla moltiplicazione nel dominio delle frequenze.

Trasformata di Fourier

Concretamente, ogni sistema dinamico lineare stazionario - linear time-invariant system (LTI) - come ad esempio un filtro, può essere espresso in maniera relativamente semplice come operazione sulle frequenze.

Dopo l'esecuzione delle operazioni si torna nel dominio del tempo per ottenere il risultato finale.

Una delle possibili implementazioni di filtri audio è quindi quella di utilizzare l'algoritmo di FFT ed eseguire il filtraggio direttamente nel dominio delle frequenze.

Filtri: sistemi LTI

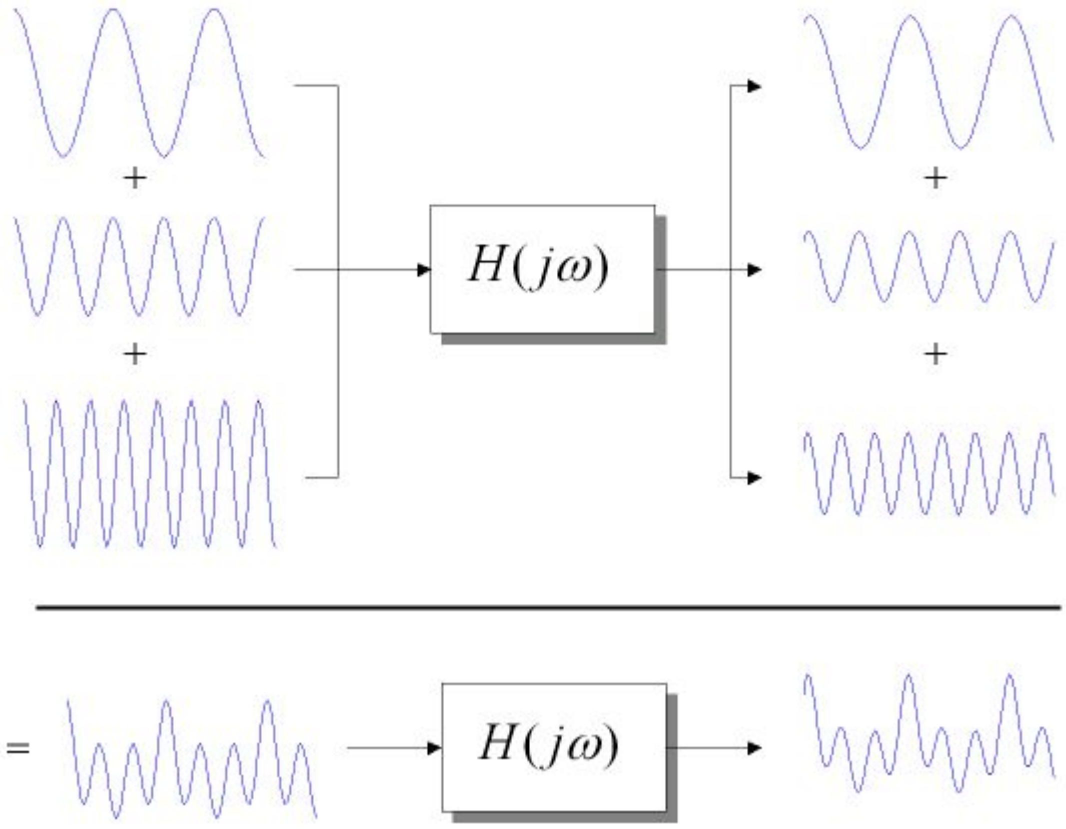
Linearità significa che, se il sistema ha due o più inputs e due o più corrispondenti outputs ad ogni combinazione lineare degli inputs

$$a_1 v_1 + \ldots + a_n v_n$$

corrisponde una combinazione lineare degli outputs.

Tempo-invariante (time invariant), significa che non è importante quando viene fatto partire il sistema. Per un determinato input, risulterà comunque sempre il medesimo output.

Filtri: sistemi LTI



Filtri: sistemi LTI

Nella realtà, non esiste nessun sistema che è veramente LTI. Alcune non-linearità possono essere introdotte ad esempio a causa di variazioni di temperatura o altri fattori esterni.

Molti sistemi possono però essere modellati per semplicità come sistemi LTI, trascurando le eventuali non-linearità.

I sistemi LTI sono importanti perché possono essere risolti con metodologie di signal processing.

Invece, i sistemi che non sono LTI sono rappresentati da equazioni differenziali non lineari, molte delle quali non sono risolvibili.

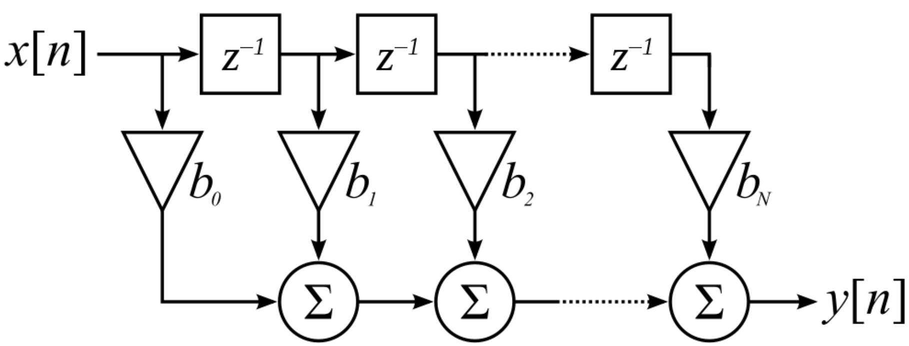
è possibile realizzare filtri con l'FFT (come descritto precedentemente. Alternative più efficaci sono i filtri FIR e IIR.

Filtri FIR

Un filtro con finite impulse response (FIR) è un filtro la cui risposta impulsiva (o risposta a qualsiasi input di lunghezza finita) è di durata finita, perché l'output va automaticamente a 0 in un tempo finito.

Ogni valore della sequenza di output è una somma pesata dei valori di input più recenti.

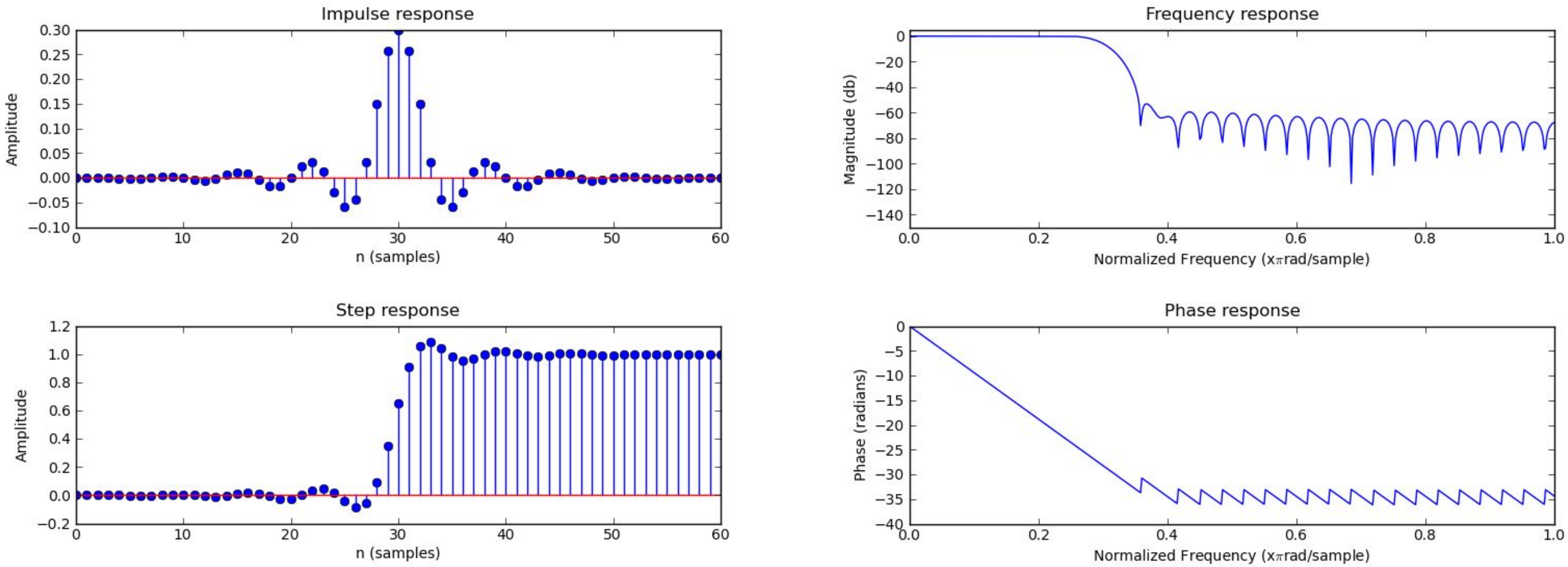
$$y[n] = b_0 x[n] + b_1 x[n-1] + \cdots + b_N x[n-N] = \sum_{i=0}^{N} b_i \cdot x[n-i]$$



x[n]: è il segnale di input
y[n]: è il segnale di output
N: è l'ordine del filtro
bi: sono i coefficienti del filtro

Filtri FIR

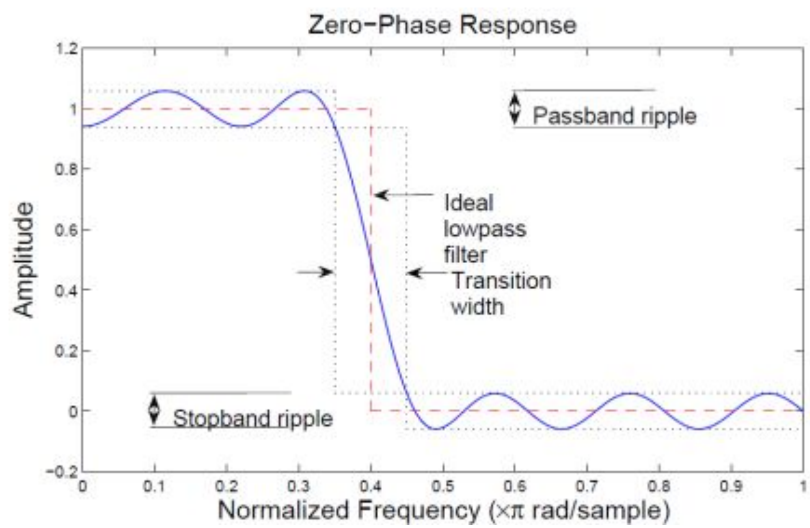
La risposta impulsiva di un filtro FIR di ordine N dura esattamente N + 1 campioni (dal primo elemento diverso da 0 all'ultimo elemento diverso da 0) prima che l'output vada a 0.



Filtri FIR

I filtri FIR possono garantire una risposta di fase lineare, caratteristica interessante ad esempio nei filtri di crossover degli altoparlanti.

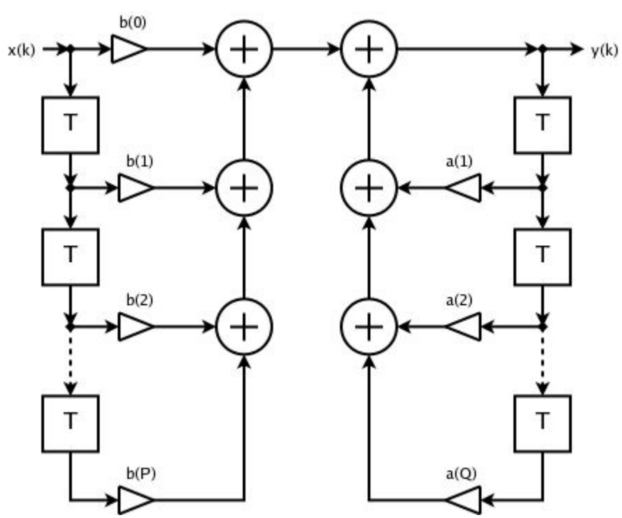
come controparte richiedono una quantità di calcoli maggiore rispetto ai filtri IIR, soprattutto per i tagli nelle basse frequenze. Inoltre è difficile rendere i filtri FIR parametrizzabili in tempo reale. Il calcolo dei coefficienti è un'operazione tipicamente complessa.



Filtri IIR

I filtri IIR sono filtri con risposta impulsiva infinita. A differenza dei filtri FIR hanno un feedback (una parte ricorsiva del filtro). Per questo motivo la risposta in frequenza dei filtri IIR è migliore di quella dei filtri FIR dello stesso ordine.

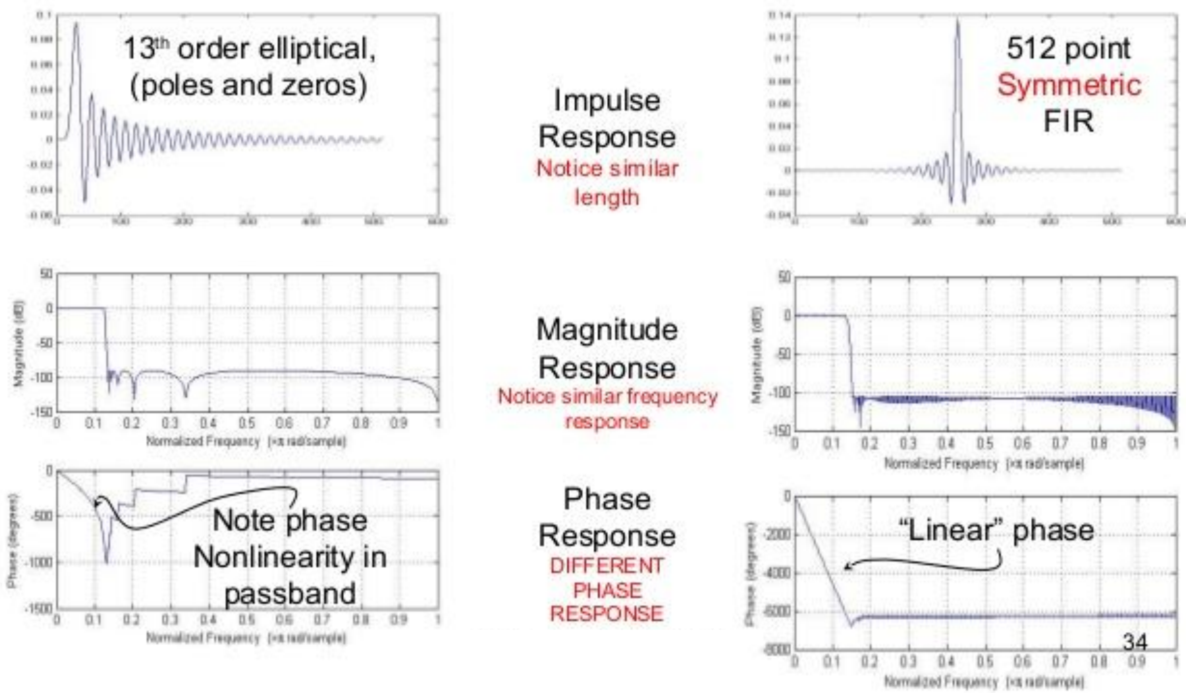
$$y[n] = \frac{1}{a_0}(b_0 x[n] + b_1 x[n-1] + \cdots + b_P x[n-P] - a_1 y[n-1] - a_2 y[n-2] - \cdots - a_Q y[n-Q])$$



x[n]: è il segnale di input
y[n]: è il segnale di output
P: è l'ordine della parte in feedforward
bi: sono i coefficienti di feedforward
Q: è l'ordine della parte in feedback
ai: sono i coefficienti di feedback

Filtri IIR

I filtri IIR sono più efficienti e possono essere facilmente parametrizzati in tempo reale. La risposta di fase però non è lineare. Inoltre, se non vengono disegnati correttamente, possono diventare instabili a causa dell'anello di feedback.



Processori di dinamica

Le trasformazione eseguite dai processori di dinamica come compressori e limiter sono di principio tutte non-lineari. Quindi, non è possibile utilizzare gli strumenti del signal processing. Le possibilità di trattamento sono quindi relativamente limitate.

Nella maggior parte dei casi questi tipi di processori mantengono in memoria un certo numero di dati recenti che usano per adattare dinamicamente la dinamica (quindi l'inviluppo) del segnale d'uscita.