Capitolo 4

Analisi ed Elaborazione del Suono

Riccardo Di Federico

4.1 Analisi di Fourier

L'analisi spettrale costituisce uno dei più potenti strumenti di indagine in molti campi dell'ingegneria. Il fatto di poter rappresentare segnali complessi come somma di funzioni semplici, tipicamente sinusoidi o esponenziali complessi, permette di evidenziare caratteristiche del segnale altrimenti difficili, se non impossibili, da rilevare. Ad esempio, parametri acustici quali pitch (altezza) e timbro sono generalmente ottenuti mediante algoritmi operanti nel dominio della frequenza. La decomposizione in funzioni semplici è di grande aiuto anche quando si deve modificare il segnale. Poter agire selettivamente su ogni singola componente permette di effettuare manipolazioni di caratteristiche del suono, quali il timbro, impraticabili con semplici interventi sulla forma d'onda. Una trattazione teorica rigorosa dell'analisi spettrale è al di là degli scopi di questi appunti. In questa sede ci concentreremo più sull'interpretazione e l'uso del più comune strumento di indagine spettrale: la Short Time Fourier Transform (STFT), definita come trasformata di Fourier dipendente dal tempo. La STFT è spesso sinonimo di *analisi tempo-frequenza*, locuzione con cui intendiamo uno studio *congiunto* delle caratteristiche temporali e spettrali del suono, cioè dell'evoluzione temporale dei parametri spettrali del segnale. Per arrivare a comprendere il significato della STFT verranno richiamati alcuni elementi teorici della trasformata di Fourier. A partire dalla serie di Fourier, definita per segnali analogici, periodici e di lunghezza infinita, verranno introdotte come estensioni l'integrale e la trasformata di Fourier. Rimuovendo quindi l'ipotesi di segnale analogico saranno definite la trasformata per segnali a tempo discreto e la trasformata discreta (cioè a tempi e frequenze discrete). Sulla base delle osservazioni fatte sulla trasformata di Fourier verranno infine discusse le principali problematiche relative all'impiego della Short Time Fourier Transform.

4.1.1 Segnale periodico, a tempo continuo, di estensione infinita - serie di Fourier

Sia $x(t)$ un segnale a tempo continuo, periodico di periodo T e di estensione infinita:

$$x(t) = x(t+mT) \qquad \forall m \in \mathcal{Z}, t \in \mathcal{R} \tag{4.1}$$

si dimostra allora che $x(t)$ può essere rappresentato da una somma pesata (e in generale infinita) di cosinusoidi le cui frequenze sono multiple intere di 1/T. Si ha cioè':

$$x(t) = \sum_{k=0}^{+\infty} C_k cos(\omega_0 kt + \phi_k) \qquad \omega_0 = \frac{2\pi}{T} \tag{4.2}$$

in cui il termine ϕ_k tiene conto della "posizione" della $k-esima$ cosinusoide. Una forma alternativa, più comoda per introdurre la trasformata di Fourier è la forma *complessa* della serie di Fourier, che si ottiene dalla (4.2) riscrivendo il coseno come somma di esponenziali complessi e riorganizzando i limiti della sommatoria:

$$x(t) = \sum_{k=-\infty}^{+\infty} F_k e^{(j\omega_0 kt)} \qquad \omega_0 = \frac{2\pi}{T} \tag{4.3}$$

in cui gli F_k sono legati ai C_k attraverso la relazione:

$$F_k = \frac{C_{|k|}}{2} e^{j\mathrm{sgn}(k)\phi_{|k|}} \qquad n \in \mathcal{Z} \tag{4.4}$$

quindi i coefficienti F_k contengono sia l'informazione di fase che quella di modulo relative alla k-esima parziale:

$$|F_k| = C_k/2, \angle F_k = \phi_k. \tag{4.5}$$

la determinazione di F_k si ottiene osservando che dato che tutti gli esponenziali presenti in (4.3) sono combinazioni di seni e coseni di periodo T, il loro integrale sul periodo è nullo; si ha in particolare:

$$\int_0^T e^{j\omega_0 nt} e^{-j\omega_0 mt} dt = \int_0^T e^{j\omega_0 (n-m)t} dt = \begin{cases} 0 & n \neq m \\ T & n = m \end{cases} \tag{4.6}$$

Se a $e^{j\omega_0 nt}$ sostituiamo $x(t)$, il valore dell'integrale diventa $F_m T$, pari cioè al coefficiente dell'unico esponenziale di x a pulsazione $\omega_0 m$. È quindi diretta la derivazione della seguente formula per il calcolo dei coefficienti della serie:

$$F_k = \frac{1}{T} \int_0^T x(t) e^{-j\omega_0 kt} dt \tag{4.7}$$

sostituendo infine la (4.7) nella (4.3) si ottiene l'identità:

$$x(t) = \sum_{k=-\infty}^{+\infty} \frac{1}{T} [\int_0^T x(t) e^{-j\omega_0 kt} dt] e^{(j\omega_0 kt)} \tag{4.8}$$

4.1.2 Segnale aperiodico, a tempo continuo, di estensione infinita. Integrale e trasformata di Fourier

Nel caso il segnale non sia perfettamente periodico non è possibile darne una rappresentazione periodica, ottenuta come somma di coseni a frequenze multiple della fondamentale. Tuttavia si può pensare di estendere l'intervallo considerato come periodo fino a comprendere l'intero asse reale. In questa operazione di limite la frequenza fondamentale tende a zero e così anche la distanza fra le armoniche. In pratica la somma nella (4.8) diventa un integrale con la sostituzione formale $1/T \to df, \omega = 2\pi f, d\omega = 2\pi f$ e i limiti di integrazione della (4.7) diventano $(-\infty, +\infty)$:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [\int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt] e^{(j\omega t)} d\omega \tag{4.9}$$

La (4.9) prende il nome di **integrale di Fourier**. Il termine entro parentesi quadre svolge il ruolo dei coefficienti complessi della serie di Fourier, e può quindi essere interpretato come l'ampiezza complessa, contenente cioè l'informazione di modulo e fase, alla frequenza ω. La funzione di ω che ne risulta è la **trasformata di Fourier**:

$$F(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt \tag{4.10}$$

La (4.9) ci fornisce anche la formula di inversione (trasformata inversa di Fourier):

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{j\omega t} d\omega \tag{4.11}$$

4.1.3 Segnale aperiodico, a tempo discreto, di estensione infinita. Trasformata di Fourier a tempo discreto (DTFT)

Il calcolo numerico della trasformata di Fourier richiede il campionamento del segnale da analizzare. Questa discretizzazione dei tempi implica la sostituzione del simbolo di integrale nella (4.10) con un simbolo di sommatoria:

$$F(\omega) = T_C \sum_{n=-\infty}^{+\infty} x(nT_C) e^{-j\omega n T_C} \tag{4.12}$$

oppure, ridefinendo $\omega_r = \omega T_C$ (e quindi passando da rad/s a rad) e $x(n) = x(nT_C)$:

$$F(\omega_r) = T_C \sum_{n=-\infty}^{+\infty} x(n) e^{-j\omega_r n} \tag{4.13}$$

è facile osservare che $F(\omega_r) = F(\omega_r + 2\pi)$ come era naturale attendersi dato che il campionamento in un dominio corrisponde alla periodicizzazione nel dominio duale. La trasformata a tempi discreti può quindi essere definita per valori di $\omega_r \in [0, 2\pi)$. Ammesso che siano rispettate le condizioni del teorema del campionamento (se non lo sono si può pre-processare il segnale con un filtro antialiasing), il segnale può essere recuperato dalla sua trasformata applicando la (4.11) con la sostituzione della variabile di integrazione $\omega \to \omega_r \quad (d\omega = d\omega_r / T_c)$

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega_r) e^{j\omega_r n} d\omega \tag{4.14}$$

4.1.4 Segnale aperiodico, a tempo e frequenze discreti. Trasformata discreta di Fourier (DFT)

Come è facile rendersi conto, il calcolo numerico della trasformata inversa (4.14) richiede che, oltre ai tempi, anche le frequenze siano discretizzate. Questo passaggio si può ottenere campionando uniformemente l'asse frequenziale, e sostituendo quindi $\omega_r \to 2\pi/N$ (con $k \in [-N/2+1, N/2]$ per N pari e $k \in [-(N-1)/2, (N-1)/2]$ per N dispari). Il simbolo di integrale nella (4.14) viene sostituito da una sommatoria e $d\omega_r \to \Delta\omega_r = 2\pi/N$. Si ha quindi:

$$F(k) = T \sum_{n=-\infty}^{+\infty} x(n) e^{-j2\pi \frac{kn}{N}} \tag{4.15}$$

$$x(n) = \frac{1}{N} \sum_{k=-N/2-1}^{N/2-1} F(k) e^{-j2\pi\frac{kn}{N}} \qquad \text{N pari} \tag{4.16}$$

$$x(n) = \frac{1}{N} \sum_{k=-(N-1)/2}^{(N-1)/2} F(k) e^{-j2\pi\frac{kn}{N}} \qquad \text{N dispari} \tag{4.17}$$

Il campionamento nel dominio della frequenza induce sul segnale $x(n)$ una periodicizzazione di periodo N. Per evitare aliasing *nel tempo*, cioè per ricostruire in modo esatto il segnale, $x(n)$ **dovrà quindi avere un'estensione inferiore o uguale a** N, nel qual caso la somma nella (4.16) sarà limitata a $n \in [-N/2+1, N/2]$ e quella in (4.17) a $n \in [-(N-1)/2, (N-1)/2]$ per N dispari. I noti algoritmi di Fast Fourier Transform (FFT) non sono altro che implementazioni veloci della DFT, con la limitazione che la lunghezza del segnale deve essere limitata e pari a una potenza di due.

4.2 Short Time Fourier Transform (STFT)

Nel caso si voglia usare la DTFT per analizzare le proprietà tempo varianti di un segnale è necessario selezionare tratti di segnale sufficientemente corti da poter essere assunti stazionari. Una sequenza di questi spettri a breve termine costituisce la **Short Time Fourier Transform**.

4.2.1 Definizioni

Sia x un segnale a tempo discreto; definiamo come Short Time Fourier Transform di x:

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j\omega m} \tag{4.18}$$

in cui $w(n-m)$ è una sequenza reale di estensione finita, detta **finestra di analisi**, che ha la funzione di limitare, troncandola in modo più o meno brusco, la porzione di segnale sotto analisi. È evidente che la STFT di un segnale è una funzione di due variabili: la pulsazione ω (normalizzata a $[0, 2\pi)$), e il campione n a cui essa è valutata. La (4.18) può essere interpretata come una trasformata che "scorre" sul segnale (in effetti è la finestra che scorre sul segnale). Una forma alternativa della (4.18) si può ottenere con il cambio di indice nella somma $n-m \to m$:

$$X_n(e^{j\omega}) = e^{-j\omega n} \sum_{m=-\infty}^{\infty} w(n)x(n-m)e^{j\omega m} \tag{4.19}$$

In questo caso è il segnale che scorre sotto la finestra centrata intorno all'origine.

4.2.2 Interpretazione della STFT come Trasformata di Fourier e come banco di filtri

Considerando n fissato, la (4.18) si può vedere come la trasformata a tempo discreto di x intorno all'istante n, su un'estensione limitata dalla lunghezza della finestra di analisi. Applicando la formula di trasformata inversa alla (4.18) si può ricostruire $x(n)$ a partire dalla sua STFT:

$$w(n-m)x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X_n(e^{j\omega})e^{j\omega m} d\omega \quad \text{da cui, se } w(0) \neq 0 \tag{4.20}$$

$$x(n) = \frac{1}{2\pi w(0)} \int_{-\pi}^{\pi} X_n(e^{j\omega})e^{j\omega m} d\omega \tag{4.21}$$

Dal punto di vista della pulsazione, fissato il valore di ω, $X_n(e^{j\omega})$ si può interpretare come l'uscita di un filtro con risposta $w(\cdot)$ al cui ingresso viene immesso $x(n)$ demodulato dall'esponenziale $e^{-j\omega}$. In altre parole la porzione di spettro intorno alla pulsazione w viene riportata intorno all'origine e quindi 'vista' attraverso il filtro $w(n)$,che ha in genere una risposta di tipo passabasso (figura 4.1).

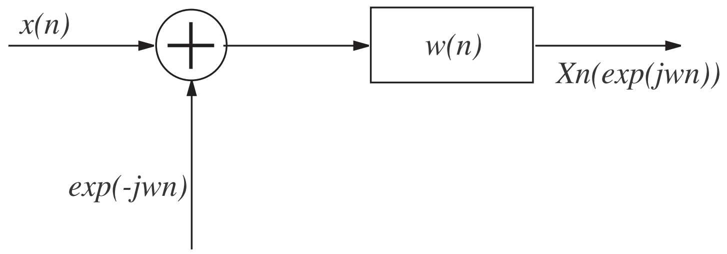

Figura 4.1: Interpretazione della STFT come banco di filtri

4.2.3 Influenza della finestra di analisi - principio di indeterminazione

Gli effetti del troncamento indotto dalla finestra di analisi possono essere evidenziati notando che, per il teorema della convoluzione, la trasformata di Fourier di un prodotto è la convoluzione delle trasformate. La STFT calcolata all'istante n è la convoluzione della trasformata $W_n(\omega)$ della finestra $w(n-m)$ e della trasformata del segnale $X(\omega)$:

$$X_n(\omega) = W_n * X(\omega) \tag{4.22}$$

in cui, detta $W(\omega)$ la trasformata di $w(n)$ e applicando le proprietà sulla traslazione e sull'inversione dell'asse dei tempi, si ha $W_n(\omega) = W(-\omega)e^{-j\omega n}$. Dato che $w(\cdot)$ è reale $|W(\omega)|$ è pari e quindi anche $|W_n(\omega)| = |W(\omega)|$. Ogni sinusoide (o esponenziale complesso) componente x dovrebbe essere rappresentato, in assenza della finestra, da un impulso ideale; l'effetto della finestra è di sostituire ad ognuno di questi impulsi la sua trasformata centrata alla frequenza dell'impulso stesso (figura 4.2). La scelta della lunghezza della finestra w va effettuata in base alle esigenze di risoluzione tempo-frequenziale. Prendiamo come esempio la finestra rettangolare (i ragionamenti che seguono si applicano ugualmente a tutte le finestre reali e pari che si usano normalmente nell'analisi spettrale). La trasformata di Fourier della finestra rettangolare è la funzione $sinc(\cdot)$, la cui estensione in frequenza cresce al diminuire della estensione temporale della finestra. Supponiamo di analizzare un segnale formato da due sinusoidi a frequenza diversa; se vogliamo una buona risoluzione temporale, l'intervallo di analisi deve essere il più corto possibile, in modo che i parametri del segnale si possano ritenere approssimativamente stazionari. La DTFT del segnale è data dalla convoluzione delle trasformate di x e della finestra. Dato che il segnale ha come trasformata una coppia di impulsi ideali e la finestra corrisponde a una fdt approssimativamente passabasso in frequenza, la trasformata globale consiste essenzialmente di due lobi centrati sulle frequenze dei seni. La larghezza di banda di questi lobi aumenta al diminuire della estensione temporale della finestra; se quest'ultima è troppo piccola i lobi sono così sovrapposti che non è più possibile distinguere le due componenti. In altre parole, una maggiore risoluzione temporale (piccola estensione della finestra) si paga con una peggiore risoluzione frequenziale (lobi larghi che si sovrappongono). Questo esempio può essere generalizzato nella definizione di un principio di indeterminazione secondo cui *non è possibile stimare con precisione arbitraria e simultaneamente i parametri temporali e frequenziali di un segnale*.

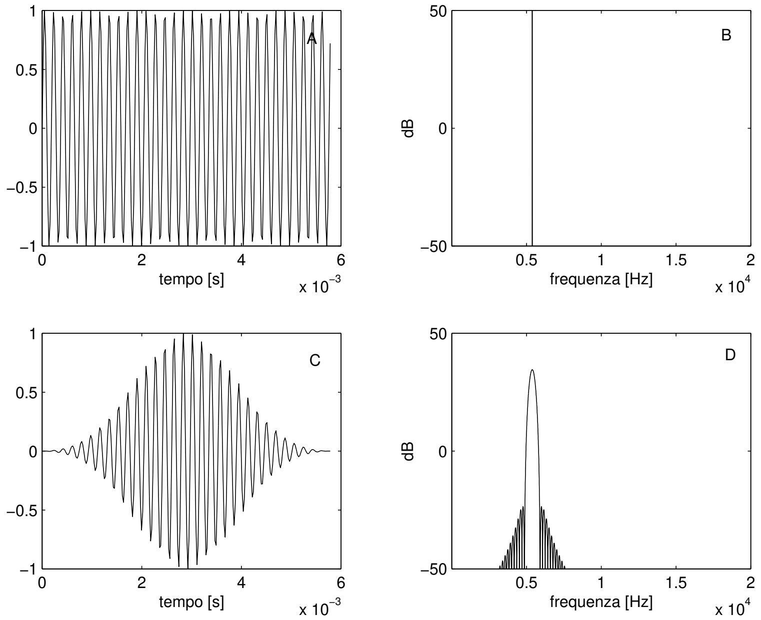

Figura 4.2: Effetto della finestra sulla trasformata di Fourier: a) segnale sinusoidale non troncato e b) modulo della sua trasformata; c) segnale dopo l'applicazione di una finestra e d) sua trasformata

4.2.4 Scelta del tipo di finestra da utilizzare

La finestra più semplice che si può pensare di utilizzare è quella rettangolare; in questo caso la porzione di segnale da analizzare viene semplicemente estratta mediante troncamento. Ci si può chiedere se non sia meglio in alcuni casi pesare in modo diverso l'inizio e la fine del frame di analisi. In effetti la finestra rettangolare è discontinua ai bordi e questo, come è noto, implica un decadimento delle code laterali della trasformata piuttosto lento. La conseguenza è che l'influenza della trasformata della finestra si sente anche a considerevole distanza sullo spettro. Se consideriamo invece una finestra che va a zero in modo 'dolce' agli estremi, le code laterali rimangono basse, producendo uno spettro più 'pulito'. Naturalmente questo miglioramento non è gratuito; il prezzo da pagare è in termini di larghezza del lobo principale. In generale una finestra che permette una buona risoluzione frequenziale (lobo principale stretto) ha le code laterali alte, e viceversa. Un esempio di finestre con diverso compromesso fra larghezza del lobo principale e altezza delle code laterali è mostrato in figura 4.3.

4.2.5 Frequenze di campionamento della STFT nel tempo e in frequenza

Se risulta ovvio che il segnale debba essere campionato con una frequenza che rispetti le condizioni del teorema del campionamento, meno banale è la definizione della frequenza di campionamento della SFTF, cioè dell'intervallo che deve intercorrere fra una DTFT e la successiva (*hop size*) affinché non ci sia perdita di informazione, in modo cioè che il segnale di ingresso possa essere ricostruito esattamente dalla sua STFT. Con riferimento all'interpretazione come banco di filtri (figura 4.1), possiamo osservare che la banda passante della STFT (per qualunque pulsazione ω considerata) è pari a quella della trasformata della finestra di analisi, che definiremo B; sarà quindi sufficiente porre la frequenza di campionamento F_w della STFT a un valore pari o maggiore a due volte la banda della finestra:

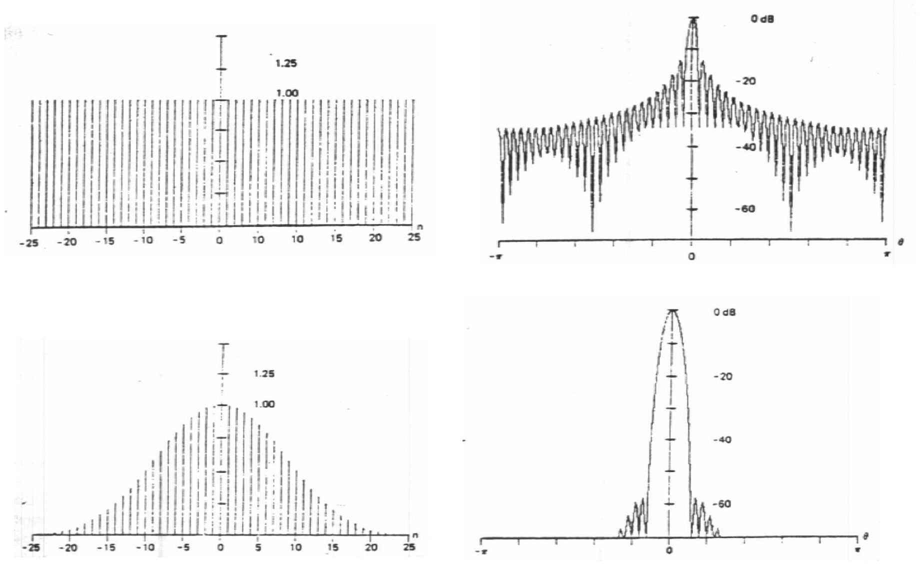

Figura 4.3: Confronto fra finestre di analisi con diverso compromesso larghezza del lobo principale/altezza delle code laterali. A)finestra rettangolare e B) modulo della trasformata in dB. C) finestra di Blackman e D) modulo della trasformata in dB.

$F_w \geq 2B$ [1]. Si dimostra che la finestra con banda minima, e quindi che richiede la minima frequenza di campionamento, è quella rettangolare. Finora abbiamo considerato la STFT come sequenza di spettri continui in frequenza; questa ipotesi non può essere rispettata nella realtà degli elaboratori in cui i calcoli devono essere effettuati su insiemi finiti (seppure molto vasti) di elementi. Dovendo quindi campionare l'asse frequenziale, sostituiamo alla DTFT la DFT. Il problema che rimane è capire quale frequenza di campionamento delle frequenze debba essere adottata. Ancora una volta, applicando il teorema del campionamento, ma questa volta scambiando i ruoli dei domini temporale e frequenziale si può affermare che è necessario adottare almeno $N \geq L$ campioni dell'asse frequenziale, se L è la lunghezza temporale della finestra di analisi. In definitiva, se tutte le condizioni sul campionamento sono soddisfatte, la STFT può essere espressa, in termini di DFT:

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j\omega_k m} \tag{4.23}$$

che con $\omega_k = 2\pi\frac{k}{N}$ diventa:

$$X_n(k) = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j2\pi\frac{km}{N}} \tag{4.24}$$

Esempio: finestra di Hamming

Supponiamo di adottare una frequenza di campionamento del segnale pari a $F_c = 44100$Hz, e usare come finestra di analisi la finestra di Hamming a 1024 punti (L=1024):

$$w(n) = \begin{cases} 0.54 - 0.46cos(\frac{2\pi n}{L}) & 0 \leq n \leq L-1 \\ 0 & \text{altrove} \end{cases} \tag{4.25}$$

[1]Le finestre che si usano normalmente per la STFT hanno estensione limitata nel tempo e quindi non limitata in frequenza. Ne consegue che qualunque sia la determinazione di B, l'ipotesi del teorema del campionamento è solo approssimata.

Si può vedere che la banda B della finestra di Hamming rispetta approssimativamente la seguente relazione:

$$B \cong 2\frac{F_c}{L} \tag{4.26}$$

E dovrà quindi essere $F_w >= 2B \cong 4F_c/L = 4 * 44100/1024 \cong 173Hz$

La STFT dovrà essere campionata nel tempo a circa 173 Hz cioè ogni $F_c/173 = 44100/173 = 254$ campioni del segnale. Parrebbe a questo punto corretto supporre un uso della STFT per comprimere il segnale ($Fw << F_c$). D'altra parte per le considerazioni sulla frequenza di campionamento dell'asse frequenziale si ha che ogni DFT deve essere rappresentata da almeno L campioni. Ne consegue una frequenza di campionamento totale (quantità di campioni al secondo)

$$SR = F_w \cdot L \cong 2BL \tag{4.27}$$

Nel caso della finestra di Hamming $SR = 2BL = 4F_c/L * L = 4F_c$! In generale $SR \geq F_c$, e il segno di uguaglianza vale solo nel caso di finestra rettangolare.

4.2.6 Esempi di rappresentazione della STFT

La serie delle DFT che costituiscono la STFT può essere visualizzata in modo da fornire un'immagine complessiva dell'evoluzione temporale dello spettro del segnale. Un importante esempio è costituito dal **sonogramma** (o **spettrogramma**), nel quale le DFT vengono accostate l'una all'altra in modo che l'asse orizzontale rappresenti il tempo e l'asse verticale le frequenze. Ad ogni punto del grafico viene assegnata una sfumatura di colore legata all'ampiezza dello spettro. In figura 4 è riportato un esempio di sonogramma di un brano cantato. Risultano evidenti le righe corrispondenti alle armoniche delle vocali e la localizzazione dei formanti. Si nota inoltre la distribuzione spettrale delle consonanti sorde 's' e 'z' (con andamento di tipo passa alto) che risultano prive di struttura armonica.

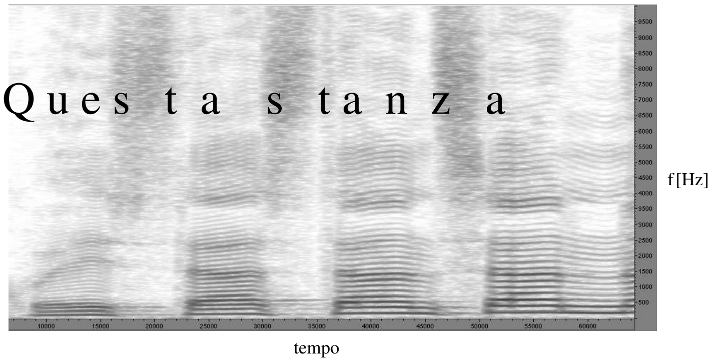

Figura 4.4: Sonogramma delle parole "...questa stanza..." (cantate).

4.2.7 Sintesi

Posto che la fase di analisi sia condotta con le condizioni di ricostruibilità del segnale, la sintesi del segnale a partire dalla sua STFT può avvenire in due modi: **filter bank summation (FBS)** e **overlap and add (OLA)**.

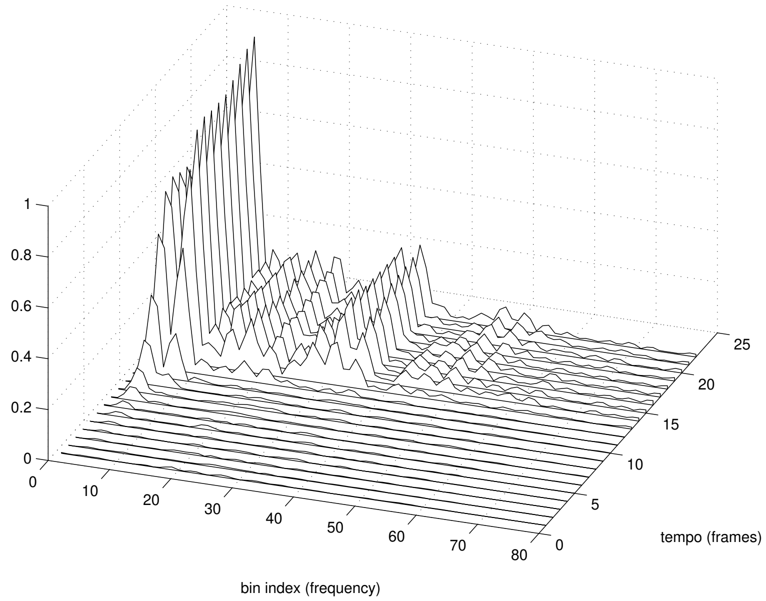

Figura 4.5: Visualizzazione spettrografica di tipo *waterfall*

Filter bank summation (FBS)

Nel primo caso si considera l'interpretazione della STFT a banco di filtri; ricordando che alla pulsazione ω_k

$$X_n(e^{j\omega_k}) = e^{-j\omega_k n} \sum_{m=-\infty}^{\infty} w_k(m)x(n-m)e^{-j\omega_k m} \tag{4.28}$$

definendo $h_k(n) = w_k(n)e^{-j\omega_k m}$ la (4.28) può essere espressa come:

$$X_n(e^{j\omega_k}) = e^{-j\omega_k n} \sum_{m=-\infty}^{\infty} h_k(m)x(n-m) \tag{4.29}$$

$h_k(n)$ rappresenta la risposta all'impulso di un filtro passabanda la cui fdt risulta quella della finestra centrata sulla pulsazione ω_k:

$$H_k(e^{j\omega}) = W_k(e^{j(\omega-\omega_k)}) \tag{4.30}$$

Essa è infatti la risposta della finestra traslata in frequenza (modulazione indotta dell'esponenziale).

Definiamo adesso

$$y_k(n) = \sum_{m=-\infty}^{\infty} x(n-m)h_k(m) \tag{4.31}$$

l'uscita del filtro passabanda k-esimo; $y_k(n)$ può essere ricavata dalla STFT tramite la (4.29), moltiplicando primo e secondo membro per $e^{j\omega_k n}$ (cioè modulando). L'idea è di ricavare x sommando tutti i contributi y_k. Posto uguale a $N(\geq L)$ il numero di filtri definiamo

$$y_k(n) = \sum_{m=-\infty}^{\infty} y_k(n) \tag{4.32}$$

la f.d.t. che lega $y(n)$ a $x(n)$ risulta essere la somma delle f.d.t. di tutti i filtri:

$$\tilde{H}(e^{j\omega}) = \sum_{k=0}^{N-1} H_k(e^{j\omega}) = \sum_{k=0}^{N-1} W(e^{j(\omega-\omega_k)})) \tag{4.33}$$

si può dimostrare che nell'ipotesi di corretto campionamento dell'asse frequenziale

$$Nw(0) = costante \tag{4.34}$$

e quindi

$$x(n) = y(n)/[Nw(0)] \tag{4.35}$$

È da notare che la formula di sintesi non dipende dalla forma della particolare finestra impiegata.

Riassumendo, il metodo di sintesi con banco di filtri si può esprimere tramite le seguenti relazioni:

$$y(n) = \sum_{k=0}^{N-1} X_n(e^{j\omega_k})e^{j\omega_k n} \tag{4.36}$$

$$x(n) = y(n)/[Nw(0)] \tag{4.37}$$

Overlap and add

Il punto di vista duale sulla sintesi si ha adottando l'interpretazione della STFT come successione di normali DFT. In questo caso la formula di inversione ci dice che i campioni del segnale all'interno della finestra di analisi possono essere recuperati tramite, appunto, una trasformata inversa, che produce

$$y_n(m) = w(n-m)x(m) \tag{4.38}$$

e quindi dividendo per la finestra $w(n-m)$. da ogni singola DFT è possibile estrarre L valori di x, esauriti i quali n può essere incrementato di L e il procedimento viene iterato. In questo modo si avrebbe un *hop size* pari a L; dalle considerazioni sulla giusta misura dell'hop size è chiaro che questo in modo la STFT è sottocampionata e quindi piuttosto sensibile a problemi di aliasing. Anche se in linea di principio è possibile estrarre i valori di x da una singola DFT, una piccola variazione dello spettro sarebbe in questo caso una potenziale fonte di distorsione della ricostruzione. Dato che in generale la *hop size*, che da ora in poi chiameremo R, è generalmente inferiore alla lunghezza della finestra, i segmenti analizzati saranno sovrapposti l'uno all'altro. Sia $Y_r(e^{j\omega_k})$ la STFT di x calcolata ogni R campioni: $Y_r(e^{j\omega_k}) = X_{rR}(e^{j\omega_k})$.

L'equazione di sintesi risulta essere:

$$y(n) = \sum_{r=-\infty}^{\infty}[\frac{1}{N}\sum_{k=0}^{N-1} Y_r(e^{j\omega_k})e^{j\omega_k n}] = \sum_{r=-\infty}^{\infty} x(n)w(rR-n) = x(n)\sum_{r=\infty}^{\infty} w(rR-n) \tag{4.39}$$

Se R è sufficientemente piccolo da evitare time aliasing, la sommatoria nell'equazione precedente è circa costante al variare di n, e in particolare circa uguale a $W(e^{j0})/R$ [2].

Vale quindi la relazione:

$$x(n) = \frac{y(n)}{W(e^{j0})/R} \tag{4.40}$$

In generale non è necessario sommare infiniti termini nella sommatoria dell'ultimo membro della (34). Infatti, dato che l'estensione della finestra è L, basterà sommare L/R campioni della finestra. Per la finestra di Hamming, ad esempio, servono 4 termini.

[2]Per dimostrare che $\sum_{r=\infty}^{\infty} w(rR-n) \cong W(0)/R$, basta osservare che:

- $w(rR-n)$ è una versione sottocampionata di $w(n)$ di un fattore R.
- Se R è abbastanza piccolo da rispettare la condizione di campionamento della STFT non c'è aliasing in frequenza.
- La somma dei campioni nel tempo non è altro che la componente continua $W(0)$ moltiplicata per la lunghezza della finestra R, cioè $R\sum_{r=\infty}^{\infty} w(rR-n) \cong W(0)$

4.2.8 Osservazioni sull'uso pratico della Short Time Fourier Transform

Adattamento della lunghezza della FFT: zero padding

Uno dei problemi che si incontra spesso usando la STFT è quello di svincolare la lunghezza della finestra di analisi dal numero di punti sul quale viene calcolata la FFT. L'algoritmo di FFT realizza infatti una mappa di N numeri in N numeri e questo, volendo mantenere costante la granularità in frequenza ($2\pi \cdot freq.di campionamento/N$), ci obbliga a usare sempre la stessa lunghezza (N) per la finestra temporale. In certi casi può tuttavia essere comodo poter regolare la quantità di segnale da trasformare in base ad altre considerazioni. Ad esempio, quando si ha a che fare con segnali (quasi) armonici, un buon compromesso fra l'ipotesi stazionarietà del segnale e la risoluzione in frequenza è quello di usare per l'analisi tre o quattro (pseudo) periodi; la lunghezza della finestra risulta quindi funzione di una proprietà tempovariante del segnale, il periodo. Per riuscire a mantenere questo compromesso e non essere costretti a cambiare il numero di punti della FFT si può applicare il procedimento di *zero padding*. Dapprima si moltiplica il segnale per la finestra prescelta di lunghezza (tempovariante) M, quindi si aggiungono un ugual numero di zeri a sinistra e a destra in modo da formare un frame di lunghezza N, pronto per essere trasformato mediante FFT. Non è difficile vedere che questo procedimento ha come unico effetto quello di interpolare da M a N punti lo spettro del segnale. Infatti, se indichiamo con $x_M(n)$ la porzione di segnale selezionata dalla finestra di lunghezza M e con $x_N(n)$ la sua versione estesa dallo zero padding:

$$
\begin{aligned}
x_M(n) &= w(n)x(n) \qquad n \in [-(M-1)/2, (M-1)/2] \\
x_N(n) &= \begin{cases} 0 & -(N-1)/2 \le n \le -(M+1)/2 \\ w(n)x(n) & -(M-1)/2 \le n \le -(M-1)/2 \\ 0 & (M+1)/2 \le n \le (N-1)/2 \end{cases}
\end{aligned} \tag{4.41}
$$

la trasformata di $x_M(n)$, $X_M(k)$ è uguale a quella di $x_N(n)$, $X_N(k)$:

$$X_N(k) = \sum_{m=-(N-1)/2}^{(N-1)/2} x_N(m) e^{-j2\pi\frac{km}{N}} \tag{4.42}$$

$$= \sum_{m=-(M-1)/2}^{(M-1)/2} x_M(m) e^{-j2\pi\frac{km}{N}} \quad k \in [\frac{-(N-1)}{2}, \frac{(N-1)}{2}] \tag{4.43}$$

L'asse delle frequenza rimane comunque campionato su N punti. È importante notare che il procedimento di zero padding produce un interpolazione dell'asse delle frequenze ma non migliora in alcun modo la capacità di discriminare sinusoidi con frequenze vicine, che dipende esclusivamente dalla larghezza del lobo principale e quindi dal tipo e dalla lunghezza M della finestra di analisi $w(n)$. Un esempio di trasformata senza e con zero padding è presentato in figura 4.6.

Corretta valutazione della fase: finestre a fase nulla

In molte applicazioni della STFT interessa conoscere solo la distribuzione spettrale dell'energia di un suono, e quindi principalmente lo spettro di ampiezza. Esistono però situazioni, come ad esempio nel caso del phase vocoder, in cui è necessario stimare con precisione anche la fase delle componenti spettrali. Vedremo in questo paragrafo come questo problema non sia banale quando la stima debba essere ottenuta tramite una STFT i cui frames siano ricavati mediante FFT su un numero pari di punti (che sfortunatamente è il caso più comune). Un generico frame di STFT è una DFT su N punti. La questione della stima della fase può quindi essere ricondotta, senza perdita di generalità,

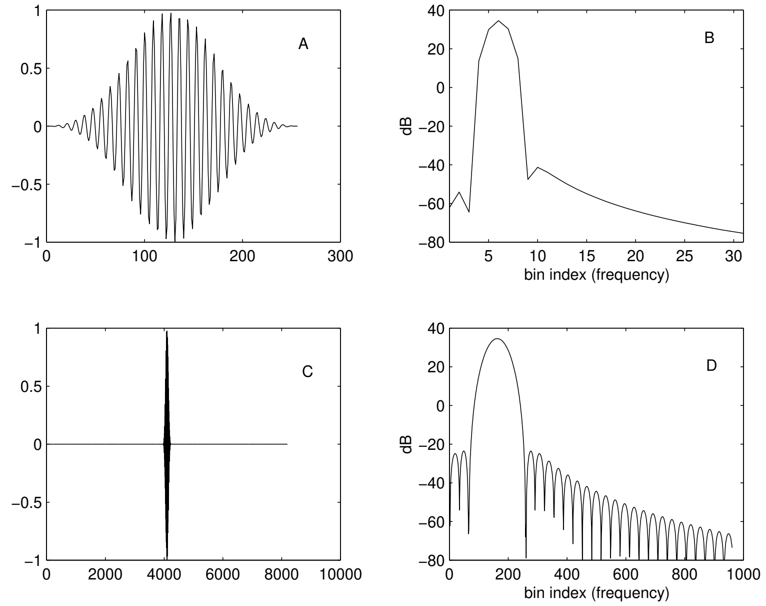

Figura 4.6: Illustrazione del procedimento di zero padding. A) sinusoide moltiplicata per la finestra di Blackman a 32 punti. B) FFT del segnale in A). C) frame di 512 punti ottenuto aggiungendo zeri a sinistra e a destra del segnale in A). D) trasformata del segnale in C)

alla valutazione della fase di una porzione del segnale $x(m)$ intorno all'origine, 'visto' attraverso la finestra di analisi $w(m)$. La DFT di x risulta essere

$$X(k) = \sum_{m=-N/2+1}^{N/2} w(m)x(m)e^{-j2\pi\frac{km}{N}} \qquad \text{N pari} \tag{4.44}$$

$$X(k) = \sum_{m=-(N-1)/2}^{(N-1)/2} w(m)x(m)e^{-j2\pi\frac{km}{N}} \qquad \text{N dispari} \tag{4.45}$$

Supponiamo per semplicità che $x(m)$ sia un esponenziale complesso a frequenza (normalizzata) k_0 e fase ϕ :

$$x(n) = e^{j2\pi\frac{k_0 n}{N}+\phi} \tag{4.46}$$

sostituendo la (4.46) nella (4.44) e nella (4.45) si ottiene:

$$X(k) = \sum_{m=-N/2+1}^{N/2} w(m)e^{j[2\pi\frac{k_0 m}{N}+\phi]}e^{-j2\pi\frac{km}{N}} \qquad \text{N pari} \tag{4.47}$$

$$X(k) = \sum_{m=-(N-1)/2}^{(N-1)/2} w(m)e^{j[2\pi\frac{k_0 m}{N}+\phi]}e^{-j2\pi\frac{km}{N}} \qquad \text{N dispari} \tag{4.48}$$

A questo punto è bene evidenziare che $w(m)$ può avere simmetria pari solo se la sua lunghezza è dispari. Infatti, la relazione $w(n) = w(-n)$ impone che esistano un ugual numero di campioni con indice positivo e negativo; aggiungendo quindi il campione nell'origine si ottiene una lunghezza della finestra dispari. Nel caso la lunghezza della finestra sia pari è comunque valida la relazione

$w(n) = w(-n+1)$. Sfruttando la simmetria di w le equazioni diventano

$$X(k) = e^{j\phi} \sum_{m=1}^{N/2} w(m)[e^{j2\pi\frac{(k_0-k)m}{N}} e^{-j2\pi\frac{(k_0-k)(m-1)}{N}}] \qquad \text{N pari} \tag{4.49}$$

$$X(k) = e^{j\phi} \left\{ w(0) + \sum_{m=1}^{(N-1)/2} w(m)[e^{j2\pi\frac{(k_0-k)m}{N}} e^{-j2\pi\frac{(k_0-k)m}{N}}] \right\} \qquad \text{N dispari} \tag{4.50}$$

e riorganizzando gli esponenziali

$$X(k) = e^{j\phi} e^{j2\pi\frac{(k_0-k)m}{N}} \sum_{m=1}^{N/2} w(m) cos[2\pi\frac{(k_0-k)}{N}(m-1/2)] \qquad \text{N pari} \tag{4.51}$$

$$X(k) = e^{j\phi} \sum_{m=0}^{(N-1)/2} w(m) cos[2\pi\frac{(k_0-k)}{N}(m)] \qquad \text{N dispari} \tag{4.52}$$

Le sommatorie contengono solo addendi reali e producono quindi numeri reali. La fase di $X(k)$ è quindi quella degli esponenziali complessi:

$$\angle X(k) = \phi + 2\pi\frac{(k_0-k)}{N} \qquad \text{N pari} \tag{4.53}$$

$$\angle X(k) = \phi \qquad \text{N dispari} \tag{4.54}$$

$$\tag{4.55}$$

Entrambe le stime producono il valore corretto di fase ϕ per $k_0 = k$. Bisogna però notare che nel caso la finestra abbia lunghezza pari è presente anche un termine di fase lineare. Questo termine introduce un errore nella stima della fase quando k_0 è non intero (cioè quasi sempre), nel qual caso la valutazione della DFT può essere fatta solo sull'indice appena inferiore o appena superiore a k_0. In figura 4.7 sono riportate le risposte di fase nei due casi di N pari ed N dispari. Il problema che si pone è quindi come usare gli algoritmi di Fast Fourier Transform, che funzionano normalmente con N pari, con finestre di lunghezza dispari. L'idea è di applicare una finestra di lunghezza dispari e quindi eliminare il campione meno significativo prima di effettuare il calcolo della FFT. Usando una delle finestre classiche, che vanno a zero (o almeno decrescono) verso i bordi, basta eliminare il primo o l'ultimo campione. Bisogna inoltre notare che la DFT è definita su intervalli simmetrici intorno all'origine ($[-(N-1)/2,(N-1)/2]$ per N dispari o $[-N/2+1,N/2]$ per N pari), mentre gli algoritmi di FFT operano sull'intervallo $[0,N-1]$. Occorre quindi applicare una rotazione di $N/2$ punti del frame di analisi (di fatto uno scambio delle due metà del frame) in modo da riportare il campione centrale della finestra $w(m)$ sull'origine (date le ipotesi di stazionarietà e continuità al di fuori del frame, questa operazione è equivalente a uno shift del segnale indietro di mezza finestra). In questo modo i valori di fase calcolati saranno riferiti al centro del frame. Un esempio di analisi con finestra a fase nulla è riportato nelle figure 4.8 e 4.8.

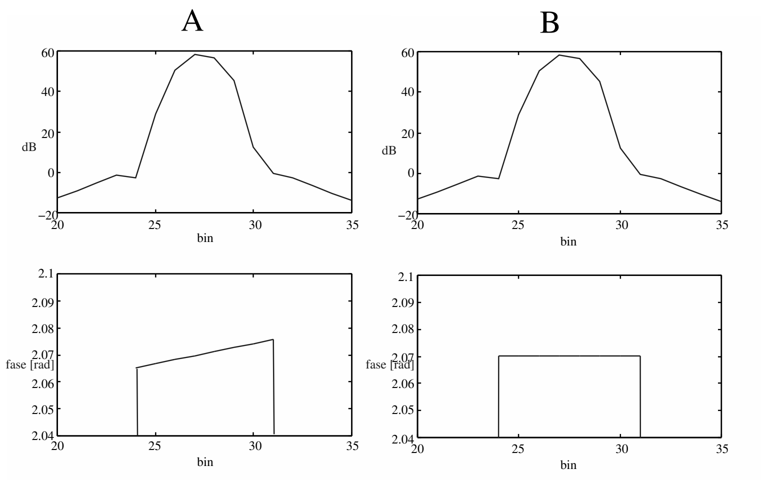

Figura 4.7: Risposte di fase calcolate sul segnale $x(m) = 1000e^{j[2\pi\frac{26.3256}{N}m+2.0701]}$nell'intorno del 26° bin. A) N=2048, B)N=2049. Si noti l'andamento lineare sovrapposto nel caso di N pari.

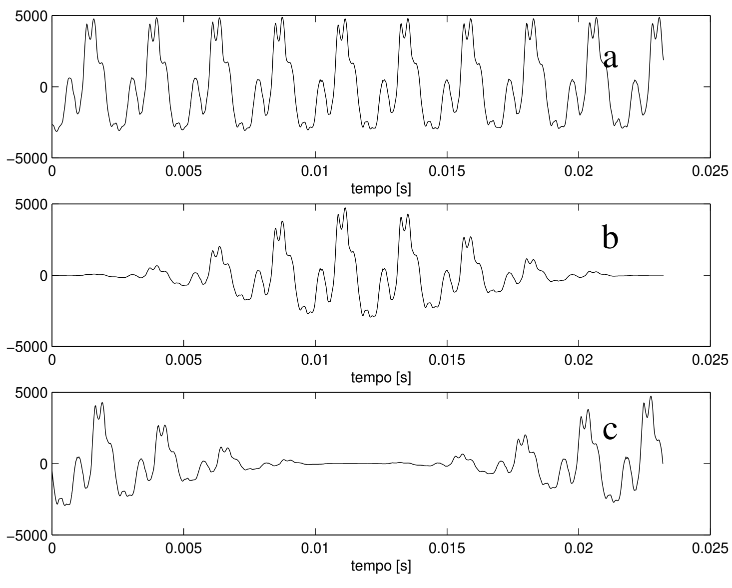

Figura 4.8: Applicazione di una finestra a fase nulla su un frame di segnale. A)Frame originale (1025 punti). B)frame dopo l'applicazione di una finestra di Blackman a 1025 punti. C) eliminazione del 1025° campione e rotazione di 512 campioni.

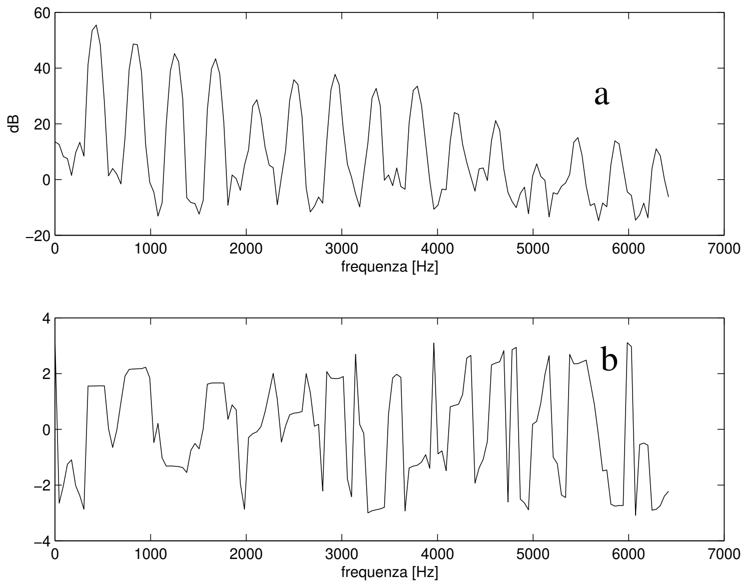

Figura 4.9: Diagrammi di modulo (a) e fase (b) della FFT calcolata sul segnale della figura precedente. Si noti che mano a mano che il modulo decresce la riposta di fase risulta sempre più disturbata dalle interferenze laterali.

4.3 Il Modello Sinusoidale

E' noto (vedi capitolo successivo) che la sintesi additiva è uno dei metodi più potenti per la generazione del suono. La possibilità di agire in modo indipendente sui parametri delle singole sinusoidi (parziali) componenti permette di controllare il risultato sonoro in modo semplice e accurato. Lo svantaggio storico di questo tipo di sintesi risiede nella sua complessità computazionale. La rapida evoluzione dell'hardware ha però portato ai personal computer odierni, che permettono di sintetizzare, via software, anche molte centinaia di sinusoidi in tempo reale.

Il rinnovato l'interesse per la sintesi additiva ha prodotto una serie di importanti risultati, fra cui l'estensione del metodo anche all'elaborazione del suono. L'idea di fondo è che se si possiede la descrizione del suono in termini di sinusoidi tempovarianti è possibile effettuare una grande quantità di trasformazioni, semplicemente agendo sui parametri di ampiezza frequenza e fase della rappresentazione sinusoidale.

Nell'articolo che segue viene presentato uno dei più recenti modelli per rappresentazione del suono mediante sinusoidi.

4.4 Fondamenti Matematici per l'Elaborazione del Suono

Questi appunti richiamano brevemente alcune nozioni sui segnali numerici, con lo scopo di introdurre gli elementi necessari alla presentazione dei fondamenti dell'elaborazione numerica dei segnali.

4.4.1 Definizioni

Ricordiamo che un segnale può essere definito come una funzione o grandezza, solitamente variabile nel tempo, che comunica informazione. Una classificazione dei segnali può essere la seguente:

1. segnali a tempo continuo: $x(t), \qquad t \in \mathcal{R} \qquad e \qquad x(t) \in \mathcal{R}$
2. segnali a tempo discreto: $x(n), \qquad n \in \mathcal{N} \qquad e \qquad x(n) \in \mathcal{R}$
3. segnali numerici: $x(n), \qquad n \in \mathcal{N} \qquad e \qquad x(n) \in \mathcal{N}$

I segnali a tempo discreto possono essere studiati come treni di impulsi ideali a tempo continuo; è tuttavia più pratico introdurre una rappresentazione *ad hoc*.

Definamo come **sequenza** un insieme di valori ordinato secondo un indice (che rappresenta l'asse temporale):

$$\{x(n)\} \qquad -\infty < n < +\infty \tag{4.56}$$

Esempi notevoli di sequenze:

- **Sequenza sinusoidale**: $x(n) = A\cos(\omega_0 n + \phi)$
- **Sequenza impulso unitario**: $\delta(n) = \begin{cases} 1/T & n = 0 \\ 0 & n \neq 0 \end{cases}$
- **Sequenza gradino**: $\delta_{-1}(n) = \begin{cases} 1/T & n \geq 0 \\ 0 & n < 0 \end{cases}$

Si dimostra che $\delta_{-1}(n) = \sum_{k=-\infty}^{n} T\delta(k)$ e viceversa $\delta(n) = \delta_{-1}(n) - \delta_{-1}(n-1)$.

Per la trattazione dei segnali numerici si può sottintendere la dipendenza dal quanto temporale di campionamento T, ed assumere una rappresentazione normalizzata (T=1):

$$x(nT) \to x(n) \tag{4.57}$$

4.4.2 Proprietà dei segnali numerici

Periodicità. Una sequenza $x(n)$ è detta periodica se $\exists N : x(n) = x(n+N) \forall n \in \mathcal{N}$.
Traslazione. La traslazione in avanti di N campioni di un segnale $x(n)$ si esprime mediante la seguente relazione:

$$x(n) \longrightarrow x(n-N) \tag{4.58}$$

Ogni sequenza può essere vista come somma di impulsi unitari scalati e traslati:

$$x(n) = \sum_{-\infty}^{+\infty} x(n)\delta(n-k) \tag{4.59}$$

4.4.3 Sistemi.

Definiamo come **sistema** una qualunque trasformazione univoca che mappa una sequenza x(n) in un'altra y(n):

$$y(n) = T[x(n)] \tag{4.60}$$

Linearità. Una sistema si dice lineare se, per ogni coppia di segnali $x1(\cdot), x2(\cdot)$ e $\forall a_1, a_2 \in \mathcal{R}$, la trasformazione T ad esso associata verifica la seguente relazione:

$$T[a_1x_1(n) + a_2x_2(n)] = a_1T[x_1(n)] + a_2T[x_2(n))] = a_1y_1(n) + a_2y_2(n) \tag{4.61}$$

Tempo invarianza. Un sistema si dice tempo invariante se la traslazione dell'ingresso induce la medesima traslazione sull'uscita:

$$y(n-k) = T[x(n-k)] \quad \forall k \in \mathcal{N} \tag{4.62}$$

Risposta all'impulso. Per i sistemi lineari è possibile definire la risposta all'impulso:

$$h_k(n) = T[\delta(n-k)], \qquad k \in \mathcal{N} \tag{4.63}$$

La relazione ingresso/uscita del sistema

$$y(n) = T[x(n)] \tag{4.64}$$

può essere riscritta, applicando l'identità (4.59), come

$$y(n) = T[\sum_{-\infty}^{+\infty} x(n)\delta(n-k)] \tag{4.65}$$

che, per la linearità del sistema si può anche esprimere come

$$y(n) = \sum_{-\infty}^{+\infty} x(n)T[\delta(n-k)] = \sum_{-\infty}^{+\infty} x(n)h_k(n) \tag{4.66}$$

Se il sistema è anche tempo invariante, definita $h(n) = T[\delta(n)]$, si ha:

$$h_k(n) = h(n-k), \qquad \text{e quindi} \tag{4.67}$$

$$y(n) = \sum_{-\infty}^{+\infty} x(n)T[\delta(n-k)] = \sum_{-\infty}^{+\infty} x(n)h(n-k) \tag{4.68}$$

Convoluzione. La scrittura

$$y * x(n) = \sum_{-\infty}^{+\infty} x(n)y(n-k) \tag{4.69}$$

è detta **convoluzione** dei segnali $x(n)$ e $y(n)$. Si può verificare che la convoluzione è un'operazione lineare e gode della proprietà commutativa.

Stabilità BIBO. Un sistema si dice stabile nel senso Bounded Input Bounded Output (BIBO) se per ogni sequenza di ingresso limitata l'uscita risulta limitata. Si dimostra che la stabilità BIBO equivale ad avere una risposta all'impulso assolutamente sommabile:

$$\text{Stabilità BIBO} \Leftrightarrow \sum_{k=-\infty}^{+\infty} |h(k)| < \infty \tag{4.70}$$

ad esempio il sistema caratterizzato dalla risposta a scalino $h(n) = \delta_{-1}(n)$ (integratore) non è stabile.
Causalità. Dicamo che un sistema è causale quando la sua uscita dipende solo da valori passati o presenti dell'ingresso. Si dice anche che il sistema è non anticipatorio. La definizione di causalità può essere espressa in termini di risposta impulsiva affermando che un sistema è causale quando la sua risposta impulsiva è nulla per tempi negativi:

$$h(n) = 0, \qquad \forall n < 0 \tag{4.71}$$

4.4.4 Sistemi lineari tempo invarianti (LIT).

Tra i sistemi lineari e tempo invarianti, la classe di maggior interesse per l'elaborazione numerica dei segnali è costituita dai sistemi razionali, caratterizzati dalla seguente equazione alle differenze a coefficienti costanti, che rappresenta la relazione ingresso uscita:

$$\sum_{k=0}^{N} a_k y(n-k) = \sum_{r=0}^{M} b_r x(n-r) \tag{4.72}$$

che può anche essere riscritta come:

$$y(n) = -\sum_{k=1}^{N} \frac{a_k}{a_0} y(n-k) + \sum_{r=0}^{M} \frac{b_r}{a_0} x(n-r) \tag{4.73}$$

L'uscita del sistema all'istante n dipende dagli N valori precedenti dell'uscita, da M valori precedenti dell'ingresso e dal valore attuale dell'ingresso.

I sistemi razionali possono essere di tipo Infinite Impulse Response (IIR) o di tipo Finite Impulse Response (FIR), a seconda che sia presente o meno la dipendenza da valori precedenti dell'uscita:

$$\textbf{FIR: } y(n) = \sum_{r=0}^{M} \frac{b_r}{a_0} x(n-r) \tag{4.74}$$

$$\textbf{IIR: } y(n) = -\sum_{k=1}^{N} \frac{a_k}{a_0} y(n-k) + \sum_{r=0}^{M} \frac{b_r}{a_0} x(n-r) \tag{4.75}$$

Risposta in frequenza. La risposta in frequenza di un sistema è definita come trasformata di Fourier della risposta all'impulso:

$$H(\omega) \triangleq \mathcal{F}\left[h(n)\right](\omega) = \sum_k h(k) e^{-j\omega k} \tag{4.76}$$

Proprietà della convoluzione. La trasformata di Fourier della convoluzione di due sequenze è il prodotto delle trasformate delle singole sequenze:

$$w(n) = x * y(n) \xrightarrow{\mathcal{F}} W(\omega) = X(\omega)Y(\omega) \tag{4.77}$$

Dimostrazione:

$$w(n) = \sum_i x(n-i) y(i),$$

$$\begin{aligned}
& W(\omega) = \textstyle\sum_n [\sum_i x(n-i) y(i)] e^{-j\omega n} \\
\text{per la linearità:} \quad & W(\omega) = \textstyle\sum_i y(i) \sum_n x(n-i) e^{-j\omega n} \\
\text{ponendo } n-i=k: \quad & W(\omega) = \textstyle\sum_i y(i) \sum_k x(k) e^{-j\omega(k+i)} \\
& W(\omega) = \textstyle\sum_i y(i) e^{-j\omega i} \sum_k x(k) e^{-j\omega k} \\
& W(\omega) = X(\omega)Y(\omega)
\end{aligned}$$

ne viene che:

$$y(n) = h * x(n) \xrightarrow{\mathcal{F}} Y(\omega) = H(\omega)Y(\omega) \tag{4.78}$$

4.4.5 La trasformata $\mathcal{Z}$

Definiamo trasformata $\mathcal{Z}$ di una sequenza x(n) la quantità

$$X(z) \triangleq \sum_k x(k)z^{-k} \qquad z \in \mathcal{C} \tag{4.79}$$

Questa serie non è in generale convergente per ogni sequenza x(n), né per ogni valore di z. Si può dimostrare che data una sequenza, la regione di convergenza è una corona circolare nel piano complesso. Infatti, se esprimiamo z in forma polare:

$$z = re^{j\omega} \tag{4.80}$$

la (4.79) diventa

$$X(z) = \sum_k x(k)r^{-k}e^{-j\omega k} \tag{4.81}$$

che è la trasformata di Fourier del segnale $x(k)r^{-k}$. La convergenza assoluta della serie si ha quindi se

$$\sum_k |x(k)r^{-k}| < \infty \tag{4.82}$$

Quindi, se la trasformata zeta converge in un punto z del piano complesso allora converge su tutta la circonferenza di raggio pari a $|z|$. Notiamo ora che una qualunque sequenza può essere scomposta in una parte causale $x_c(n)$ e una parte anticausale $x_a(n)$:

$$x(n) = x_c(n) + x_a(n) \tag{4.83}$$

$$x_c(n) = \begin{cases} x(n) & n >= 0 \\ 0 & n < 0 \end{cases} \tag{4.84}$$

$$x_a(n) = \begin{cases} 0 & n >= 0 \\ x(n) & n < 0 \end{cases} \tag{4.85}$$

Non è difficile a questo punto verificare che se la trasformata di una sequenza causale esiste per $z = z_1$ allora esiste anche per $|z| > |z_1|$. Analogamente, se la trasformata di una sequenza anticausale esiste per $z = z_2$ allora esiste anche per $|z| < |z_2|$. In generale quindi, la regione di convergenza della trasformata è una corona circolare del tipo

$$r_c < |z| < r_a \tag{4.86}$$

in cui r_c e r_a sono rispettivamente il raggio minimo della regione di esistenza della parte causale e il raggio massimo della regione di esistenza della parte anticausale.

Proprietà della trasformata $\mathcal{Z}$

1. **linearità**. La trasfomata $\mathcal{Z}$ è lineare (la verifica è immediata)

2. **teorema dello shift**: $x(n+N) \to z^N X(z)$

3. **trasformata della convoluzione**: $\mathcal{Z}[x * y(n)](z) = \mathcal{Z}[x] \cdot \mathcal{Z}[y](z)$

Funzione di trasferimento Sia data la relazione ingresso/uscita $Y(z) = H(z)X(z)$. $H(z) = \mathcal{Z}[h(n)](z)$ è detta **funzione di trasferimento** del sistema.

Nel caso di sistema razionale descritto dall'equazione (4.73), trasformando ambo i membri si ha:

$$\begin{aligned}
\mathcal{Z}\left[\sum_{k=0}^{N} a_k y(n-k)\right] &= \mathcal{Z}\left[\sum_{r=0}^{M} b_r x(n-r)\right] \\
\sum_{k=0}^{N} a_k \mathcal{Z}[y(n-k)] &= \sum_{r=0}^{M} b_r \mathcal{Z}[x(n-r)] \\
\sum_{k=0}^{N} a_k z^{-k} Y(z) &= \sum_{r=0}^{M} b_r z^{-r} X(z) \\
\frac{Y(z)}{X(z)} &= \frac{\sum_{r=0}^{M} b_r z^{-r}}{\sum_{k=0}^{N} a_k z^{-k}} = H(z)
\end{aligned}$$

Osservazione Si vede subito che $H(e^{j\omega}) = \mathcal{F}[h(n)](\omega)$. Questo è conseguenza del fatto che per $z = e^{j\omega}$ la trasformata $\mathcal{Z}$ coincide con la trasformata di Fourier. Dato un sistema descritto da un'equazione alle differenze è quindi immediato calcolare la trasformata zeta e quindi la risposta in frequenza. Si osserva che $\mathcal{Z}[x(n)]$ è una serie di Laurant. Pertanto, la formula di inversione è data da:

$$x(n) = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz \tag{4.87}$$

dove C è interna alla regione di convergenza.

Se il sistema è razionale, $H(z)$ può esprimersi come:

$$H(z) = \frac{A \prod_{r=1}^{M} (1 - c_r z^{-1})}{\prod_{k=1}^{N} (1 - d_k z^{-1})} \tag{4.88}$$

dove c_r sono gli **zeri** e d_r sono i **poli** di $H(z)$.

Si dimostra che un sistema caratterizzato da una funzione di trasferimento razionale del tipo (4.88) è stabile se e solo se tutti i poli sono interni alla circonferenza di raggio unitario:

$$|d_k| < 1, \qquad k = 1, 2, \ldots N \tag{4.89}$$

Per la dimostrazione basta applicare la definizione di stabilità BIBO alla scomposizione in frazioni parziali della (4.88).