Capitolo 1

Analisi dei suoni

4 ottobre 2005

1.1 Fondamenti matematici per l'elaborazione numerica dei segnali

TO DO [toglire ipotesi Ts=1]

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.

1.1.1 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):

$$\{x(n)\} \qquad -\infty < n < +\infty \tag{1.1}$$

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):

$$x(nT) \rightarrow x(n) \tag{1.2}$$

1.1.2 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:

$$x(n) \longrightarrow x(n-N) \tag{1.3}$$

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

$$x(n) = \sum_{-\infty}^{+\infty} x(n)\delta(n-k) \tag{1.4}$$

1.1.3 Sistemi

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

$$y(n) = \mathcal{T}[x(n)] \tag{1.5}$$

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 $\mathcal{T}$ ad esso associata verifica la seguente relazione:

$$\mathcal{T}[a_1x_1(n) + a_2x_2(n)] = a_1\mathcal{T}[x_1(n)] + a_2\mathcal{T}[x_2(n))] = a_1y_1(n) + a_2y_2(n) \tag{1.6}$$

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

$$y(n-k) = \mathcal{T}[x(n-k)] \quad \forall k \in \mathcal{N} \tag{1.7}$$

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

$$h_k(n) = \mathcal{T}[\delta(n-k)], \qquad k \in \mathcal{N} \tag{1.8}$$

La relazione ingresso/uscita del sistema $y(n) = \mathcal{T}[x(n)]$ può essere riscritta, applicando l'identità (8.4), come $y(n) = \mathcal{T}\left[\sum_{-\infty}^{+\infty} x(n)\delta(n-k)\right]$che, per la linearità del sistema si può anche esprimere come

$$y(n) = \sum_{-\infty}^{+\infty} x(n)\mathcal{T}[\delta(n-k)] = \sum_{-\infty}^{+\infty} x(n)h_k(n) \tag{1.9}$$

Se il sistema è anche tempo invariante, definita $h(n) = \mathcal{T}[\delta(n)]$, si ha: $h_k(n) = h(n-k)$, e quindi

$$y(n) = \sum_{-\infty}^{+\infty} x(n)\mathcal{T}[\delta(n-k)] = \sum_{-\infty}^{+\infty} x(n)h(n-k) \tag{1.10}$$

Convoluzione. La scrittura

$$y * x(n) = \sum_{-\infty}^{+\infty} x(n)y(n-k) \tag{1.11}$$

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:

$$\text{Stabilità BIBO} \Leftrightarrow \sum_{k=-\infty}^{+\infty} |h(k)| < \infty \tag{1.12}$$

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:

$$h(n) = 0, \qquad \forall n < 0 \tag{1.13}$$

1.1.4 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:

$$\sum_{k=0}^{N} a_k y(n-k) = \sum_{r=0}^{M} b_r x(n-r) \tag{1.14}$$

che può anche essere riscritta come:

$$y(n) = -\sum_{k=1}^{N} \frac{a_k}{a_0} y(n-k) + \sum_{r=0}^{M} \frac{b_r}{a_0} x(n-r) \tag{1.15}$$

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:

$$\textbf{FIR: } y(n) = \sum_{r=0}^{M} \frac{b_r}{a_0} x(n-r) \tag{1.16}$$

$$\textbf{IIR: } y(n) = -\sum_{k=1}^{N} \frac{a_k}{a_0} y(n-k) + \sum_{r=0}^{M} \frac{b_r}{a_0} x(n-r) \tag{1.17}$$

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

$$H(\omega) \stackrel{\Delta}{=} \mathcal{F}\left[h(n)\right](\omega) = \sum_k h(k) e^{-j\omega k} \tag{1.18}$$

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

$$w(n) = x * y(n) \xrightarrow{\mathcal{F}} W(\omega) = X(\omega)Y(\omega) \tag{1.19}$$

1.1.5 La trasformata $\mathcal{Z}$

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

$$X(z) \stackrel{\Delta}{=} \sum_k x(k) z^{-k} \qquad z \in \mathcal{C} \tag{1.20}$$

Questa serie non è in generale convergente per ogni sequenza $x(n)$, nè 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:

$$z = re^{j\omega} \tag{1.21}$$

la (8.20) diventa

$$X(z) = \sum_k x(k) r^{-k} e^{-j\omega k} \tag{1.22}$$

che è la trasformata di Fourier del segnale $x(k)r^{-k}$. La convergenza assoluta della serie si ha quindi se $\sum_k |x(k)r^{-k}| < \infty$. 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 $x_a(n)$:

$$x(n) = x_c(n) + x_a(n) \tag{1.23}$$

$$x_c(n) = \begin{cases} x(n) & n >= 0 \\ 0 & n < 0 \end{cases} \tag{1.24}$$

$$x_a(n) = \begin{cases} 0 & n >= 0 \\ x(n) & n < 0 \end{cases} \tag{1.25}$$

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

$$r_c < |z| < r_a \tag{1.26}$$

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 trasformata $\mathcal{Z}$ è lineare (la verifica è immediata)
2. **teorema dello shift**: $x(n+N) \rightarrow z^N X(z)$
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 (8.14), trasformando ambo i membri e riordonando si ha:

$$H(z) = \frac{Y(z)}{X(z)} \quad = \quad \frac{\sum_{r=0}^{M} b_r z^{-r}}{\sum_{k=0}^{N} a_k z^{-k}} \tag{1.27}$$

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:

$$x(n) = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz \tag{1.28}$$

dove C è interna alla regione di convergenza.

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

$$H(z) = A \frac{\prod_{r=1}^{M} (1 - c_r z^{-1})}{\prod_{k=1}^{N} (1 - d_k z^{-1})} \tag{1.29}$$

dove c_r sono gli **zeri** e d_r sono i **poli** di $H(z)$, e $A = b_0/a_0$.

Si dimostra che un sistema caratterizzato da una funzione di trasferimento razionale del tipo (8.29) è stabile se e solo se tutti i poli sono interni alla circonferenza di raggio unitario:

$$|d_k| < 1, \qquad k = 1, 2, ...N \tag{1.30}$$

Per la dimostrazione basta applicare la definizione di stabilità BIBO alla scomposizione in frazioni parziali della (8.27). Nel caso di sistema FIR a risposta all'impulso finita (eq. 8.16), non essendoci poli, il sistema è sempre stabile e la funzione di trasferimento diventa (posto $a_0 = 1$)

$$H(z) = \sum_{r=0}^{M} b_r z^{-r}$$

1.2 Spectral Analysis of Discrete-Time signals

The spectral analysis is one of the powerful analysis tool in several fields of engineering. The fact that we can decompose complex signals with the superposition of other simplex signals, commonly sinusoid or complex exponentials, highlight some signal features that sometimes are very hard to discover with some other kind of analysis. For example, acoustical features such as *pitch* and *timbre*, are commonly obtained with algorithm that works in the frequency domain. Furthermore, the decomposition with simplex function is very useful when we want to modify a signal. In the frequency domain, the possibility to manipulate single spectral component give us the possibility to modify some fundamental feature of the sound, such as the timbre, that are hard, and sometimes impossible, to manipulate operating on the sound waveform.

A rigorous mathematical approach of the huge field of spectral analysis is out the scope of this book. In the next chapters we focus our attention on the most common and used spectral analysis tool: the **S**hort **T**ime **F**ourier **T**ransform (STFT).Sounds are time-varying signals in the real world and, indeed, all of their meaning is related to such time variability. Therefore, it is interesting to develop sound analysis techniques that allow to grasp at least some of the distinguished features of time-varying sounds, in order to ease the tasks of understanding, comparison, modification, and resynthesis. With STFT, often defined as the time-dependent Fourier Transform, we intend the joint analysis of the temporal and frequency features of the sound. In other word with this tool is possible to follow the temporal evolution of the spectral parameters of a sound.
The concept of STFT is based on the concept of the **D**iscrete-**T**ime **F**ourier **T**ransform, DTFT, that is the fundamental tool used to analyze a signal in the frequency domain. This is the discrete time version of the classical Fourier Transform commonly used for continuous-time signals. After a brief introduction on the DTFT, we will se how the DTFT on a periodic discrete-time signal specializes in the so called DFT that is at the bases of the STFT.

1.2.1 The Discrete-Time Fourier Transform: DTFT

First of all we will clarify the meaning of the variables which are commonly associated to the word "frequency" for signals defined in both the continuous and the discrete-time domain. The various symbols are collected in table 8.1, where the limits imposed by the Nyquist frequency are also indicated. With the term "digital frequencies" we indicate the frequencies of discrete-time signals.

Nyquist Domain	Symbol	Unit	
$[-F_s/2 \quad \ldots \quad 0 \quad \ldots \quad F_s/2]$	f	[Hz] = [cycles/s]	
$[-1/2 \quad \ldots \quad 0 \quad \ldots \quad 1/2]$	f/F_s	[cycles/sample]	digital freqs.
$[-\pi \quad \ldots \quad 0 \quad \ldots \quad \pi]$	$\omega = 2\pi f/F_s$	[radians/sample]	
$[-\pi F_s \quad \ldots \quad 0 \quad \ldots \quad \pi F_s]$	$\Omega = 2\pi f$	[radians/s]	

Tabella 1.1: Frequency variables

Recalling that for a continuous-time signal $x(t)$ the Fourier Transform is defined as:

$$F(\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t}dt \tag{1.31}$$

where $\omega = 2\pi f$ is the continuous-frequency expressed in radians, we can try to re-express this expression in the case of a discrete-time signal $x[n]$.
If we think about a discrete-time signal as the sampled version of a continuous-time signal, $x(t)$, with a sampling interval $T = \frac{1}{F_s}$, $x[n] = x(nT)$, we can define the DTFT starting from the Eq.8.31 where the integral is substituted by a summation:

$$X(f) = \sum_{n=-\infty}^{+\infty} x(nT)e^{-j2\pi \frac{f}{F_s} n} \ . \tag{1.32}$$

As we will see later, $X(f)$ is a periodic function of the continuous-frequency variable f, with period F_s. Now if we use the variable $\omega = 2\pi\frac{f}{F_s}$, a more compact expression arise from the Eq:8.32 :

$$X(\omega) = \sum_{n=-\infty}^{+\infty} x(nT)e^{-j\omega n} \ . \tag{1.33}$$

where the variable ω is called normalized frequency,and is expressed in radians.Eq.8.33 is the general expression used to compute the DTFT.
In general $X(\omega)$ is a complex function of the real variable ω and can be written in rectangular form as

$$X(\omega) = X_{re}(\omega) + jX_{im}(\omega) \quad , \tag{1.34}$$

where $X_{re}(\omega)$ and $X_{im}(\omega)$ are, respectively, the real and imaginary parts of $X(\omega)$, and are real function of ω. $X(\omega)$ can alternately be expressed in the polar form as

$$X(\omega) = |X(\omega)|e^{\theta(\omega)} \tag{1.35}$$

and

$$\theta(\omega) = arg[X(\omega)] \tag{1.36}$$

The quantity $|X(\omega)|$ is called the *magnitude function* and the quantity $\theta(\omega)$ is called the *phase function* with both function again being real function of ω.
As can be seen from the definition Eq.8.33, the Fourier Transform $X(\omega)$ of a discrete-time sequence is a periodic function in ω with a period 2π. Note that the periodicity, with period 2π in the domain of the normalized-frequency $\omega = 2\pi\frac{f}{F_s}$, is equivalent to a periodicity of F_s in the domain of absolute-frequency f.
It therefore follows that Eq.8.33 represent the Fourier series representation of the periodic function $X(\omega)$. As a result, the Fourier coefficients $x[n]$ can be computed from $x(\omega)$ using the Fourier Integral given by

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)\mathrm{e}^{j\omega n}d\omega \quad , \tag{1.37}$$

called the *inverse discrete-time Fourier transform*. Equations 8.33 and 8.37 together form a Fourier representation for the sequence $x[n]$. Equation 8.37, the Inverse Fourier Transform, is a *synthesis* formula. That is, it represent $x[n]$ as a superposition of infinitesimally small complex sinusoid of the form

$$\frac{1}{2\pi}X(\omega)\mathrm{e}^{j\omega n}d\omega \tag{1.38}$$

with ω ranging in the interval of length 2π and with $X(\omega)$ determining the relative amount of each complex sinusoidal component. Although in writing 8.37 we have chosen the range of values for ω between $-\pi$ and π, any interval of length 2π can be used. Equation 8.33, the Fourier Transform, is an expression for computing $X(\omega)$ from the sequence $x[n]$, i.e., for *analyzing* the sequence $x[n]$ to determine how much of each component is required to synthesize $x[n]$ using Eq.8.37.

1.2.2 The Discrete Fourier Transform: DFT

In the case of a finite-length sequence $x[n], \quad 0 \leq n \leq N-1$, there is a simpler relation between the sequence and its discrete-time Fourier tranform $X(\omega)$. In fact, for a lenght-N sequence, only N values of $X(\omega)$, called *frequency samples*, at N distinct frequency points, $\omega =$
$omega_k, \quad 0 \leq k \leq N-1$, are sufficient to determine $x[n]$, and hence $X(\omega)$, uniquely. This lead to the concept of the discrete Fourier Transform, a second transform-domain representation that is applicable only to a finite-length sequence.
The DFT is at the heart of digital signal processing, because it is a **computable** transformation. Although the Fourier, Laplace and z-transform are the analytical tools of signal processing as well as many other disciplines, it is the DFT that we must use in a computer program such as Matlab.

1.2.2.1 Definition

The simplest relation between a finite-length sequence $x[n]$, defined for $0 \leq n \leq N-1$, and its DTFT $X(\omega)$ is obtained by uniformly sampling $X(\omega)$ on the ω-axis between $0 \leq \omega \leq 2\pi$ at $\omega_k = 2nk/N$, $0 \leq k \leq N-1$. From Eq.8.33,

$$X[k] = X(\omega)|_{\omega=2\pi k/N} = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi}{N}kn}, \quad 0 \leq k \leq N-1 \tag{1.39}$$

Note that $X[k]$ is also a finite-length sequence in the frequency domain and is of length N. The sequence $X[k]$ is called the *Discrete Fourier transform (DFT)* of the sequence $x[n]$. Using the commonly used notation

$$W_N = e^{-j\frac{2\pi}{N}} \tag{1.40}$$

we can rewrite Eq.8.39 as

$$X[k] = \sum_{k=0}^{N-1} x[n]W_N^{kn}, \quad 0 \leq k \leq \tag{1.41}$$

The *inverse discrete Fourier Transform*(IDFT) is given by

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]W_N^{-kn}, \quad 0 \leq n \leq . \tag{1.42}$$

1.3 Short Time Fourier Transform (STFT)

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.

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**.

1.3.1 Definizioni

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

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j\omega m} \tag{1.43}$$

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. E' 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 (8.43) può essere interpretata come una trasformata che scorre sul segnale (in effetti è la finestra che scorre sul segnale). Una forma alternativa della (8.43) si può ottenere con il cambio di indice nella somma $n - m \to m$:

$$X_n(e^{j\omega}) = e^{-j\omega n} \sum_{m=-\infty}^{\infty} w(n)x(n-m)e^{j\omega m} \tag{1.44}$$

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

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

Considerando n fissato, la (8.43) 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 (8.43) si può ricostruire $x(n)$ a partire dalla sua STFT:

$$w(n-m)x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_n(e^{j\omega})e^{j\omega m}d\omega \quad \text{da cui, se } w(0) \neq 0 \tag{1.45}$$

$$x(n) = \frac{1}{2\pi w(0)}\int_{-\pi}^{\pi} X_n(e^{j\omega})e^{j\omega m}d\omega \tag{1.46}$$

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 8.1).

1.3.3 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

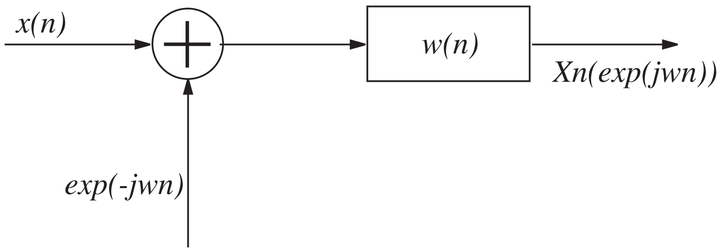

Figura 1.1: Interpretazione della STFT come banco di filtri

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 $X(w)$:

$$X_n(w) = W_n * X(\omega) \tag{1.47}$$

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 $W_n(\omega) = W(-\omega)e^{-jwn}$. 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 8.2). 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 cos 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.*

1.3.4 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

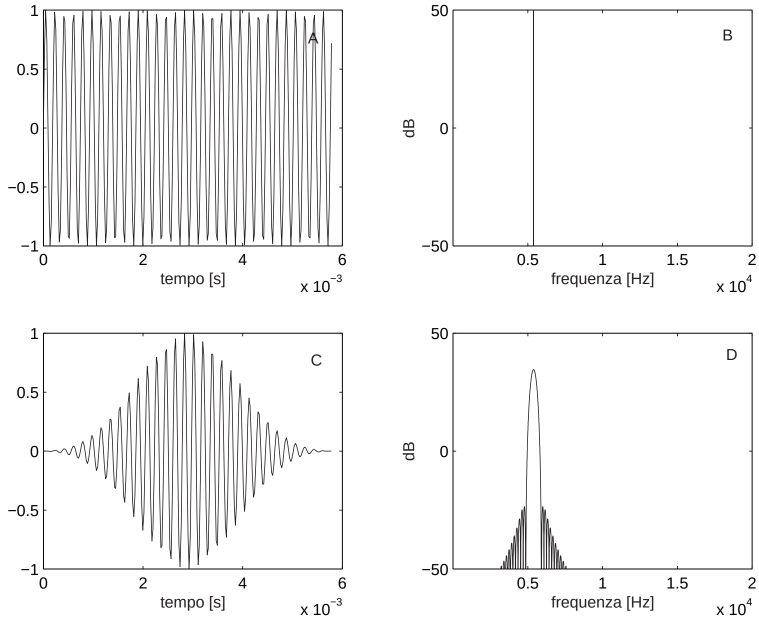

Figura 1.2: Effetto della finestra sulla trasformata di Fourier: a) segnale sinusoidale non troncato e b) modulo della sua trasformata; c) segnale dopo l'applicazione di una finestra e d) sua trasformata

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 8.3.

1.3.5 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*) affinchè 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 8.1), 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: $F_w \geq 2B$ [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

[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.

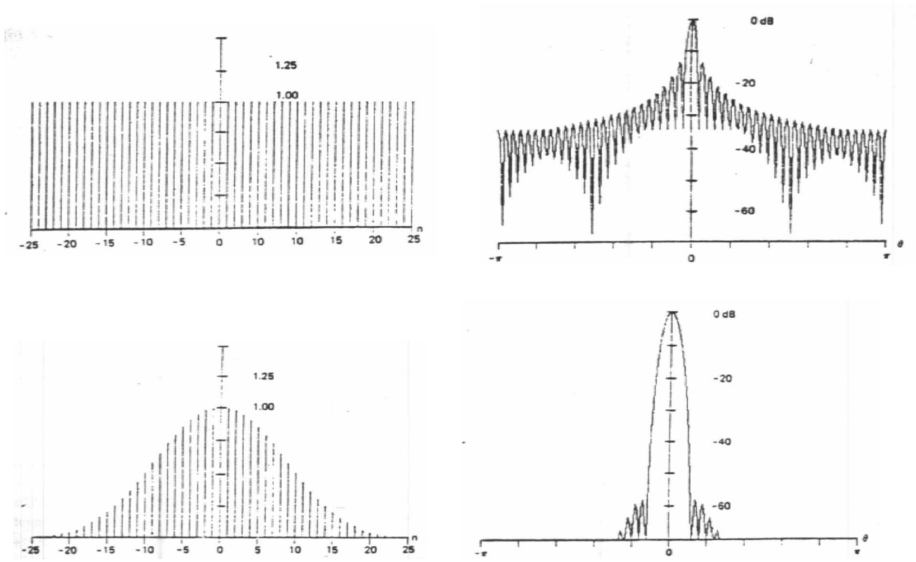

Figura 1.3: Confronto fra finestre di analisi con diverso compromesso larghezza del lobo principale/altezza delle code laterali. A)finestra rettangolare e B) modulo della trasformata in dB. C) finestra di Blackman e D) modulo della trasformata in dB.

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:

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j\omega_k m} \tag{1.48}$$

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

$$X_n(k) = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j2\pi\frac{km}{N}} \tag{1.49}$$

1.3.5.1 Esempio: finestra di Hamming

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

$$w(n) = \begin{cases} 0.54 - 0.46cos(\frac{2\pi n}{L}) & 0 \leq n \leq L-1 \\ 0 & \text{altrove} \end{cases} \tag{1.50}$$

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

$$B \cong 2\frac{F_c}{L} \tag{1.51}$$

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

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 ($Fw << Fc$). 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)

$$SR = F_w \cdot L \cong 2BL \tag{1.52}$$

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

1.3.6 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.

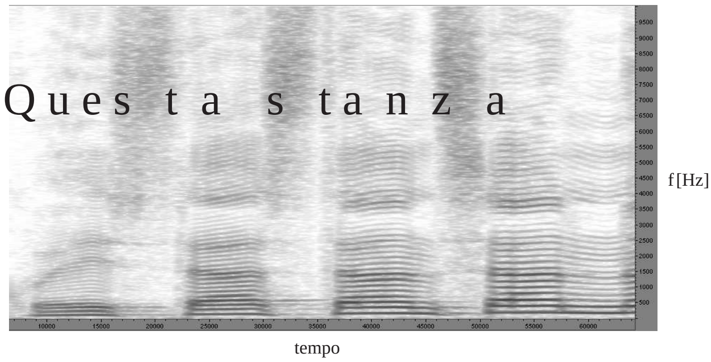

Figura 1.4: Sonogramma delle parole ...questa stanza... (cantate).

1.3.7 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)**.

1.3.7.1 Filter bank summation (FBS)

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

$$X_n(e^{j\omega_k}) = e^{-j\omega_k n} \sum_{m=-\infty}^{\infty} w_k(m)x(n-m)e^{-j\omega_k m} \tag{1.53}$$

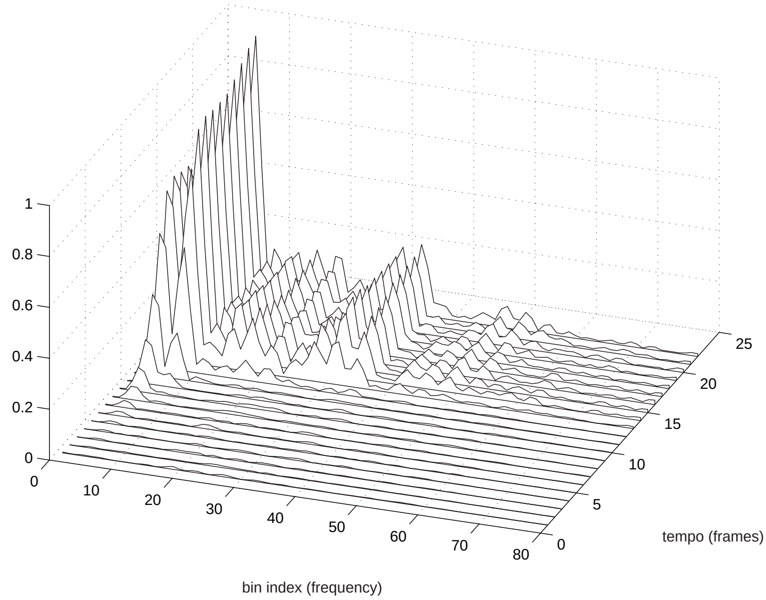

Figura 1.5: Visualizzazione spettrografica di tipo *waterfall*

definendo $h_k(n) = w_k(n)e^{-j\omega_k m}$ la (8.53) può essere espressa come:

$$X_n(e^{j\omega_k}) = e^{-j\omega_k n} \sum_{m=-\infty}^{\infty} h_k(m)x(n-m) \tag{1.54}$$

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

$$H_k(e^{j\omega}) = W_k(e^{j(\omega-\omega_k)}) \tag{1.55}$$

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

Definiamo adesso

$$y_k(n) = \sum_{m=-\infty}^{\infty} x(n-m)h_k(m) \tag{1.56}$$

l'uscita del filtro passabanda k-esimo; $y_k(n)$ può essere ricavata dalla STFT tramite la (8.54), 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

$$y_k(n) = \sum_{m=-\infty}^{\infty} y_k(n) \tag{1.57}$$

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

$$\tilde{H}(e^{j\omega}) = \sum_{k=0}^{N-1} H_k(e^{j\omega}) = \sum_{k=0}^{N-1} W(e^{j(\omega-\omega_k)}) \tag{1.58}$$

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

$$Nw(0) = costante \tag{1.59}$$

e quindi

$$x(n) = y(n)/[Nw(0)] \tag{1.60}$$

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:

$$\begin{aligned} y(n) &= \sum_{k=0}^{N-1} X_n(e^{j\omega_k})e^{j\omega_k n} && (1.61) \\ x(n) &= y(n)/[Nw(0)] && (1.62) \end{aligned}$$

1.3.7.2 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

$$y_n(m) = w(n-m)x(m) \tag{1.63}$$

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:

$$y(n) = \sum_{r=-\infty}^{\infty}\left[\frac{1}{N}\sum_{k=0}^{N-1} Y_r(e^{j\omega_k})e^{j\omega_k n}\right] = \sum_{r=-\infty}^{\infty} x(n)w(rR-n) = x(n)\sum_{r=\infty}^{\infty} w(rR-n) \tag{1.64}$$

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 $W(e^{j0})/R$. Vale quindi la relazione:

$$x(n) = \frac{y(n)}{W(e^{j0})/R} \tag{1.65}$$

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.

1.3.8 Osservazioni sull'uso pratico della Short Time Fourier Transform

1.3.8.1 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:

$$\begin{aligned} x_M(n) &= w(n)x(n) \qquad n \in [-(M-1)/2, (M-1)/2] \\ x_N(n) &= \begin{cases} 0 & -(N-1)/2 \le n \le -(M+1)/2 \\ w(n)x(n) & -(M-1)/2 \le n \le -(M-1)/2 \\ 0 & (M+1)/2 \le n \le (N-1)/2 \end{cases} \end{aligned} \tag{1.66}$$

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

$$X_N(k) = \sum_{m=-(N-1)/2}^{(N-1)/2} x_N(m) e^{-j2\pi\frac{km}{N}} \tag{1.67}$$

$$= \sum_{m=-(M-1)/2}^{(M-1)/2} x_M(m) e^{-j2\pi\frac{km}{N}} \qquad k \in [\frac{-(N-1)}{2}, \frac{(N-1)}{2}] \tag{1.68}$$

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 8.6.

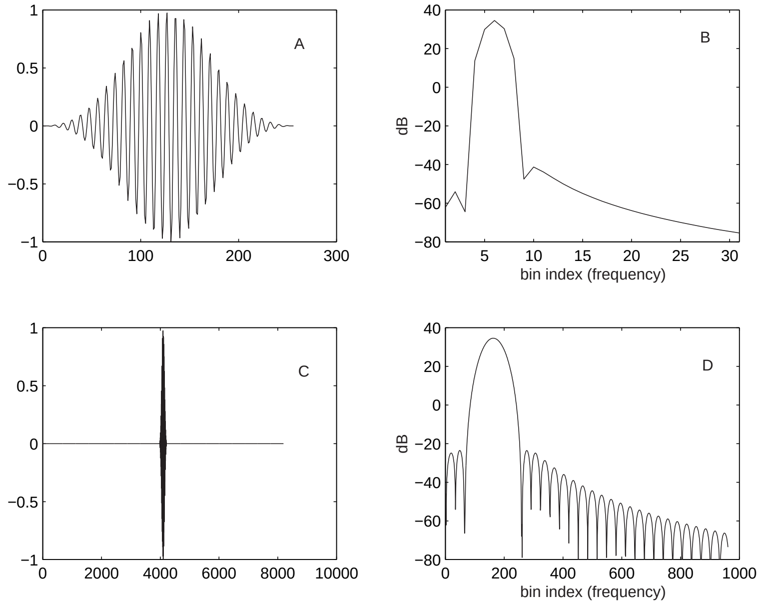

Figura 1.6: Illustrazione del procedimento di zero padding. A) sinusoide moltiplicata per la finestra di Blackman a 32 punti. B) FFT del segnale in A). C) frame di 512 punti ottenuto aggiungendo zeri a sinistra e a destra del segnale in A). D) trasformata del segnale in C)

Indice