AISV Scuola Estiva 2008 - *Archivi di Corpora Vocali*

Analisi del segnale

Carlo Drioli
Dipartimento di Informatica dell'Università di Verona

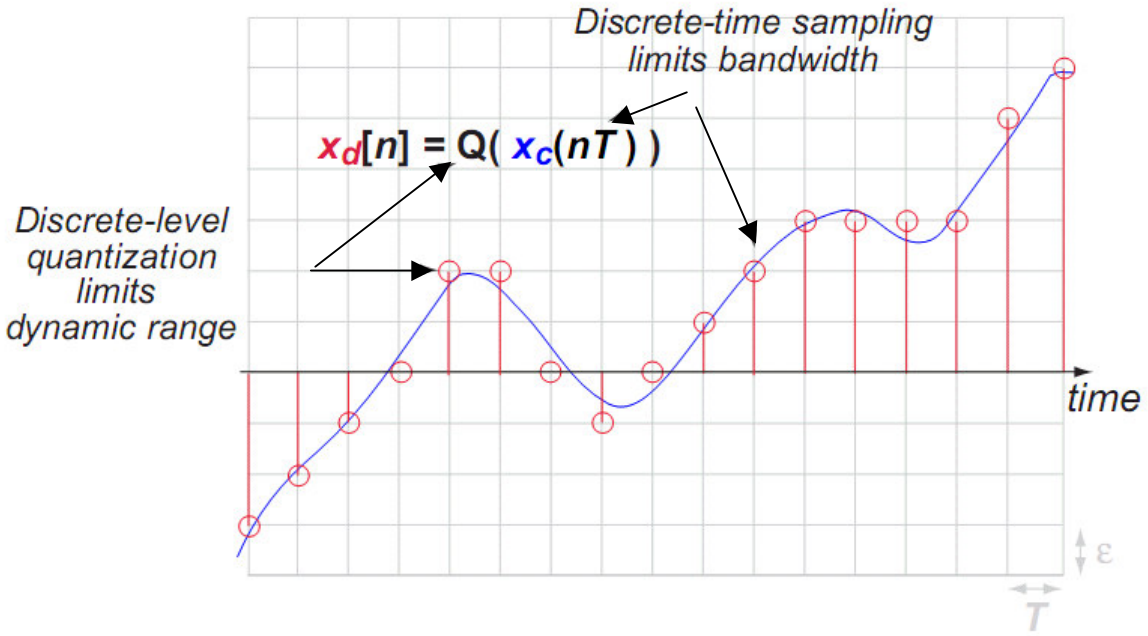
Outline

- Segnali digitali: campionamento e quantizzazione
- Analisi di Fourier e spettrogramma
- Analisi short-time e finestratura
- Filtraggio lineare
- Modelli sorgente-filtro della fonazione
- Analisi LPC
- Analisi cepstrale e mel-cepstrale

DSP: introduzione

Segnali digitali: dominio del tempo

- Discretizzazione dei tempi (**campionamento**) e delle ampiezze (**quantizzazione**): $x(t) \rightarrow x_c(nT) \rightarrow x_d[n]$



- T: intervallo di campionamento
- $F_s = \frac{1}{T}$ (Hz) , $\Omega_s = \frac{2\pi}{T}$ (rad/sec) : frequenza di campionamento
- Quantizzatore: $Q(x) = \epsilon \cdot \text{round}(\frac{x}{\epsilon})$

DSP: introduzione

Segnali digitali: quantizzazione

- Errore di quantizzazione: $\eta[n] = x_c[n] - x_d[n]$, con $-\frac{\epsilon}{2} \leq \eta \leq \frac{\epsilon}{2}$
- Nell'ipotesi che η sia rumore bianco unif. distribuito, si ha:

$$\bar{\eta} = 0, \qquad \bar{\eta^2} = \frac{2}{\epsilon}\int_0^{\epsilon/2} \eta^2 d\eta = \frac{\epsilon^2}{12}, \qquad \eta_{rms} = \sqrt{\frac{\epsilon^2}{12}} = \frac{\epsilon}{\sqrt{12}} \quad (1)$$

- Se il numero di bit usato per la quantizzazione è N_b, il valore max rappresentabile è $\epsilon \cdot 2^{N_b-1}$ e il **rapporto segnale rumore** è

$$\text{SNR} = 20\log_{10}\frac{\epsilon \cdot 2^{N_b-1}}{\epsilon/\sqrt{12}} = 20\log_{10} 2^{N_b}\sqrt{3} \approx 4.7 + 6N_b \text{ (dB)} \quad (2)$$

- Ogni bit in più incrementa di 6 dB il rapporto segnale/rumore

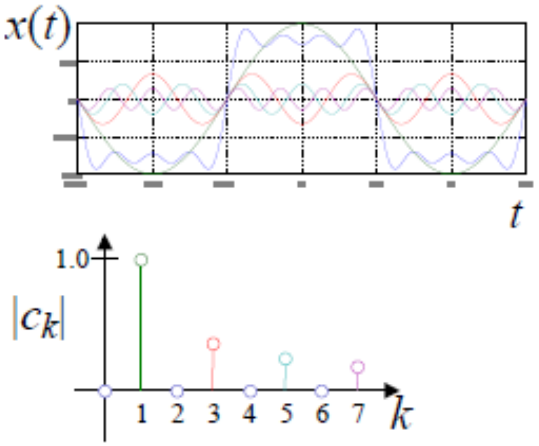
DSP: introduzione

Segnali continui: dominio della frequenza

- Serie di Fourier per segnali continui **periodici**

$$x(t) = \sum_k c_k \exp^{-jk\Omega_s t}$$

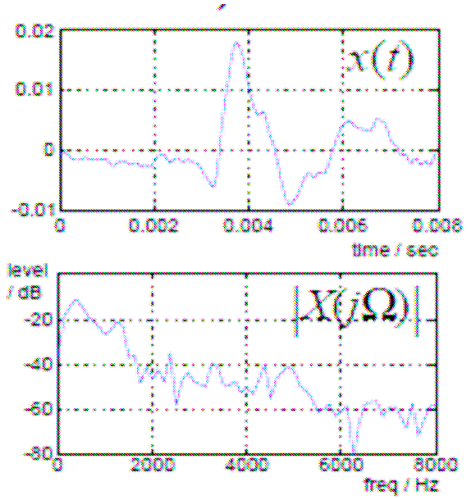
$$c_k = X[k] = \frac{1}{2\pi T} \int_{T/2}^{T/2} x(t) exp^{-jk\Omega_s t} dt$$



- Trasformata di Fourier per segnali continui **aperiodici**

$$x(t) = \frac{1}{2\pi} \int X(j\Omega) exp^{j\Omega t} d\Omega$$

$$X(j\Omega) = \int_{T/2}^{T/2} x(t) exp^{-j\Omega t} dt$$



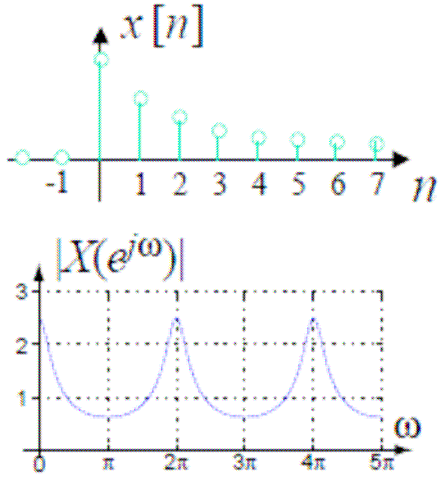
DSP: introduzione

Segnali digitali: dominio della frequenza

- Trasformata di Fourier per segnali discreti **aperiodici**

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(j\omega) exp^{j\omega n} d\omega$$

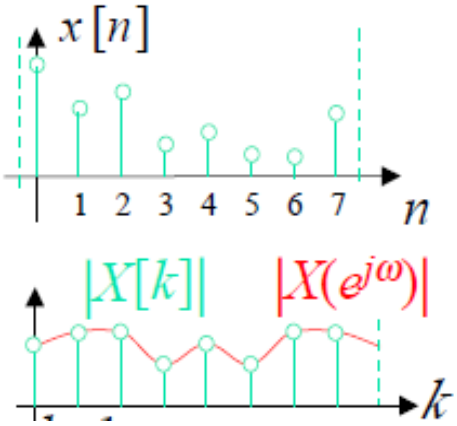
$$X[j\omega] = \sum_{n} x[n] exp^{-j\omega n}$$



- Serie di Fourier per segnali discreti **periodici** di lungh. N

$$x[n] = \sum_{k=0}^{N-1} X[k] exp^{j\frac{2\pi kn}{N}}, \quad n = [0, 1, ..., N-1]$$

$$X[k] = \sum_{n=0}^{N-1} x[n] exp^{-j\frac{2\pi kn}{N}}, \quad k = [0, 1, ..., N-1]$$



DSP: introduzione

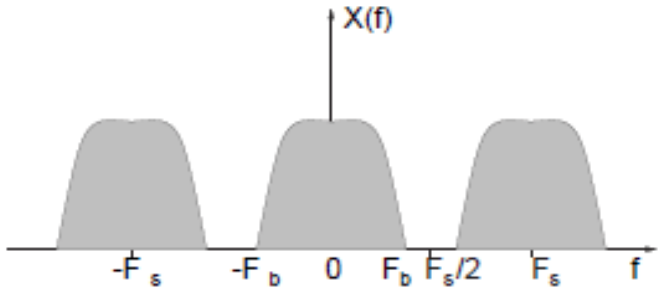
Campionamento e aliasing

- Il segnale discretizzato è uguale al segnale continuo **negli istanti di campionamento**: $x_d[n] = x_c(nT)$

- La trasformata di Fourier di una funzione di variabile discreta è una funzione della variabile continua ω, periodica di periodo 2π

- Il campionamento con frequenza F_s di un segnale continuo produce un segnale discreto il cui spettro di frequenze è una replica periodica dello spettro del segnale originale, con periodo F_s.

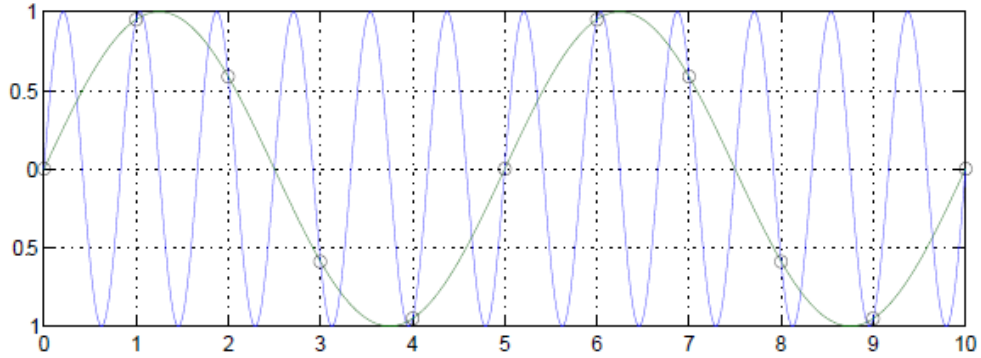
- Interpretazione in frequenza:
  ($\omega = 2\pi f / F_s$)



DSP: introduzione

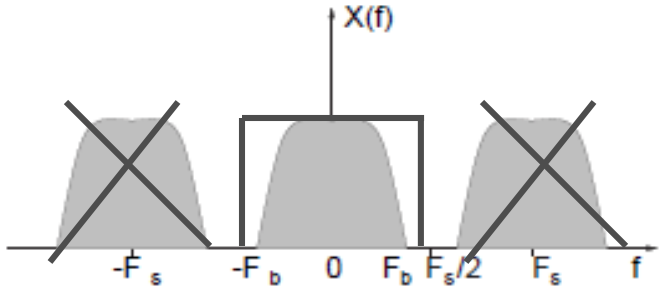
Campionamento e aliasing

- Il campionamento non può rappresentare variazioni troppo veloci:



- Teorema di **Nyquist**: Un segnale continuo con limite di banda F_b può essere ricostruito dal segnale campionato se la freq. di campionamento è $F_s > 2F_b$

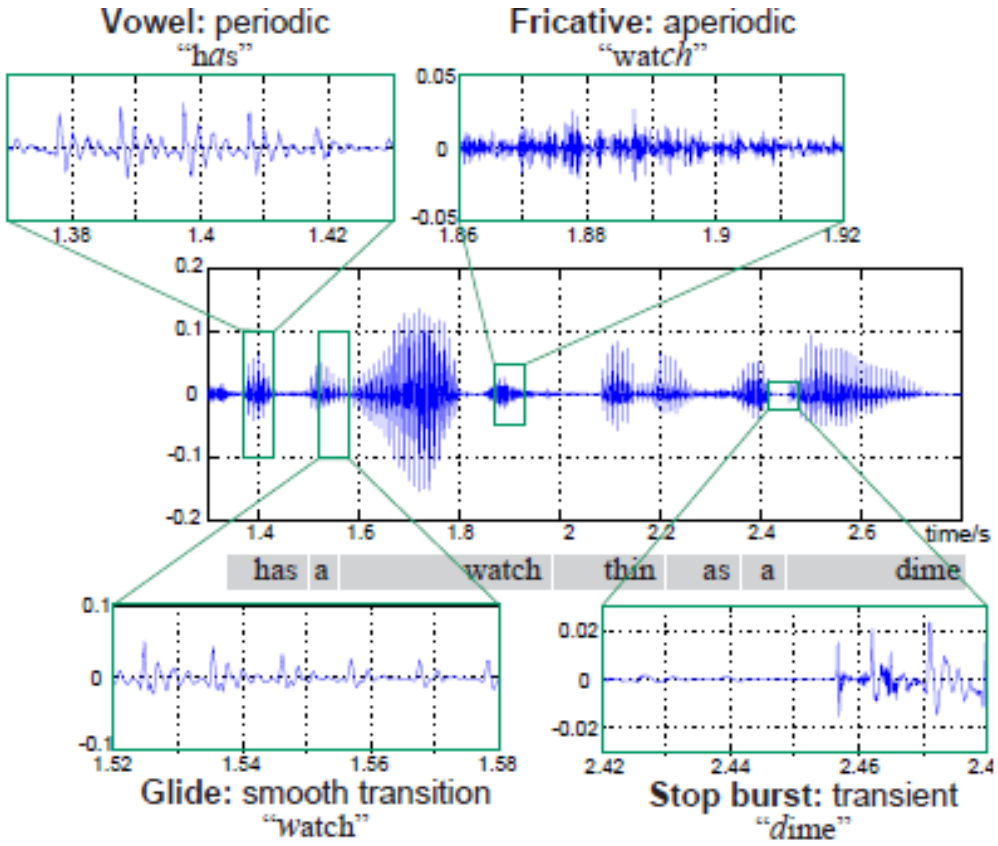
- La ricostruzione può avvenire mediante un filtro passa-basso ideale che elimina le repliche in frequenza (sinc)



DSP: segnali vocali

Il segnale vocale: dominio del tempo

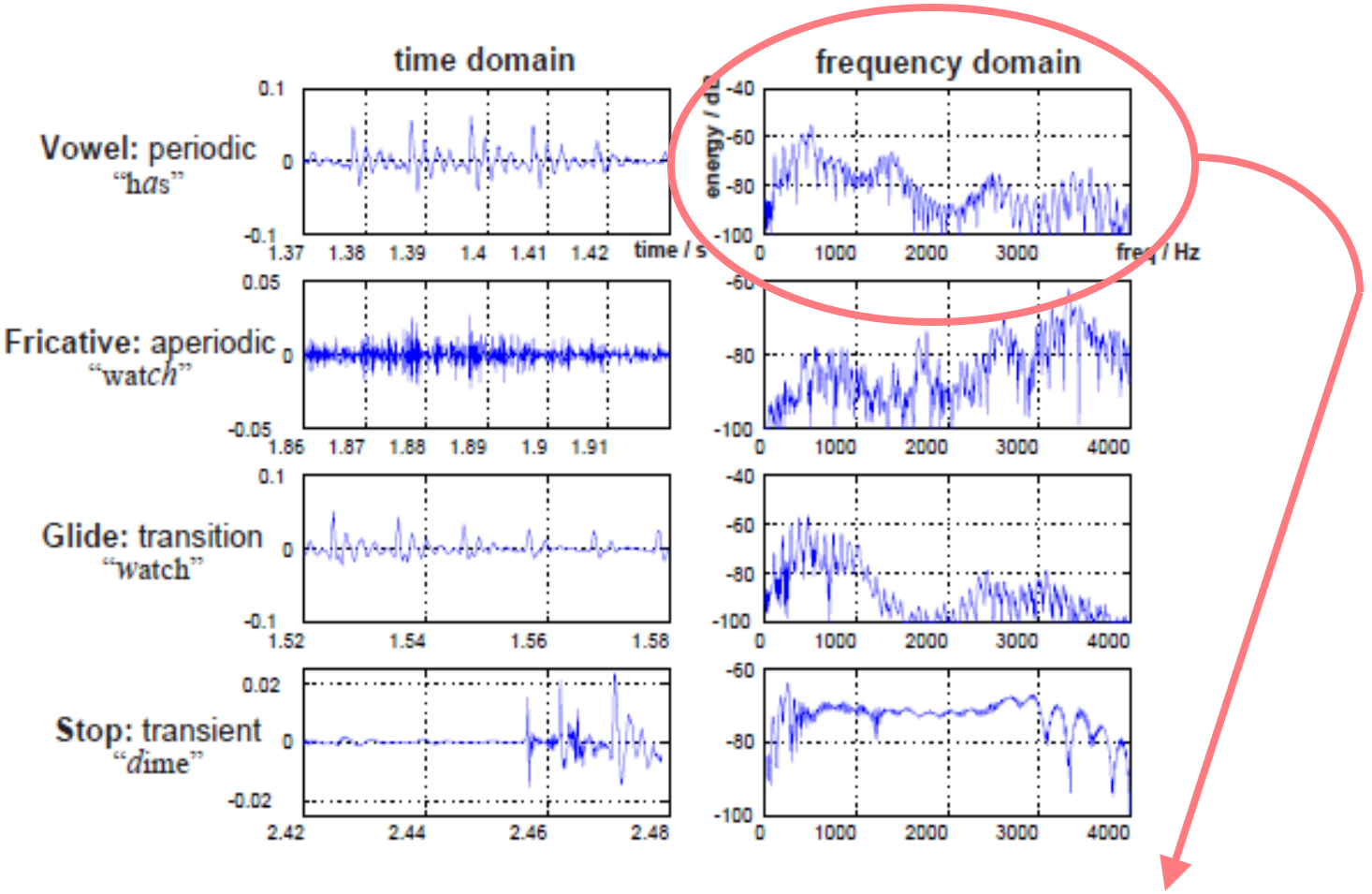
- Il segnale vocale è caratterizzato da notevole tempo-varianza



- E' necessario considerare brevi frammenti in cui può valere l'ipotesi di periodicità $\Rightarrow$ **Short Time Fourier Transform**

DSP: segnali vocali

Il segnale vocale: dominio della frequenza

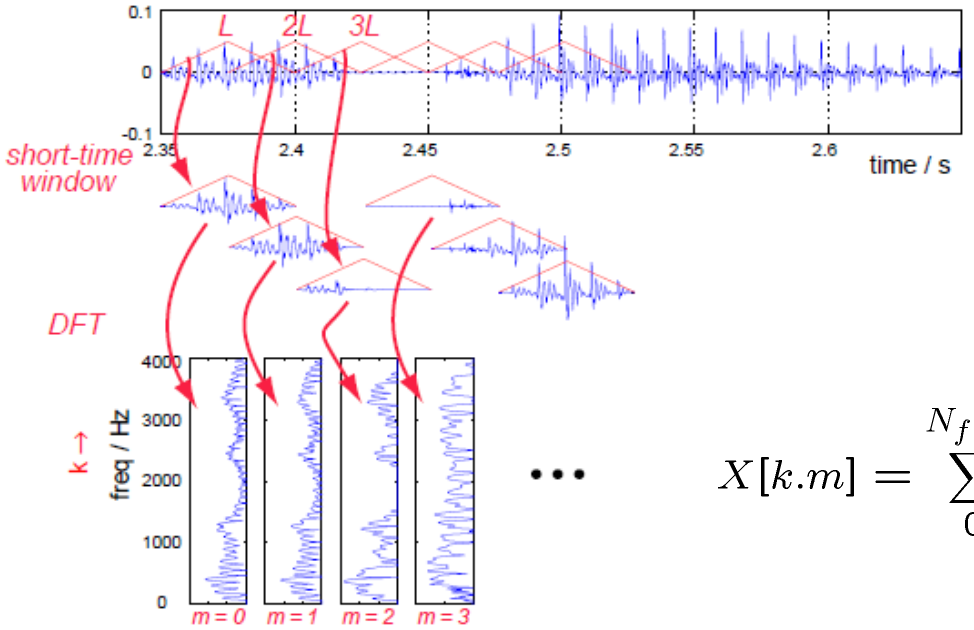


- Caratteristica notevole dei segmenti vocalici: presenza di **componenti periodiche** e di **formanti** nell'inviluppo spettrale

DSP: introduzione

Short Time Fourier Transform (STFT)

- Il segnale è segmentato in brevi segmenti (frame) di lunghezza N_f, con possibilità di sovrapposizione (L: hopsize)
- I frame sono moltiplicati con una funzione di finestratura $w[n]$ per attenuare gli effetti ai bordi
- Ogni frame è trasformato con una DFT



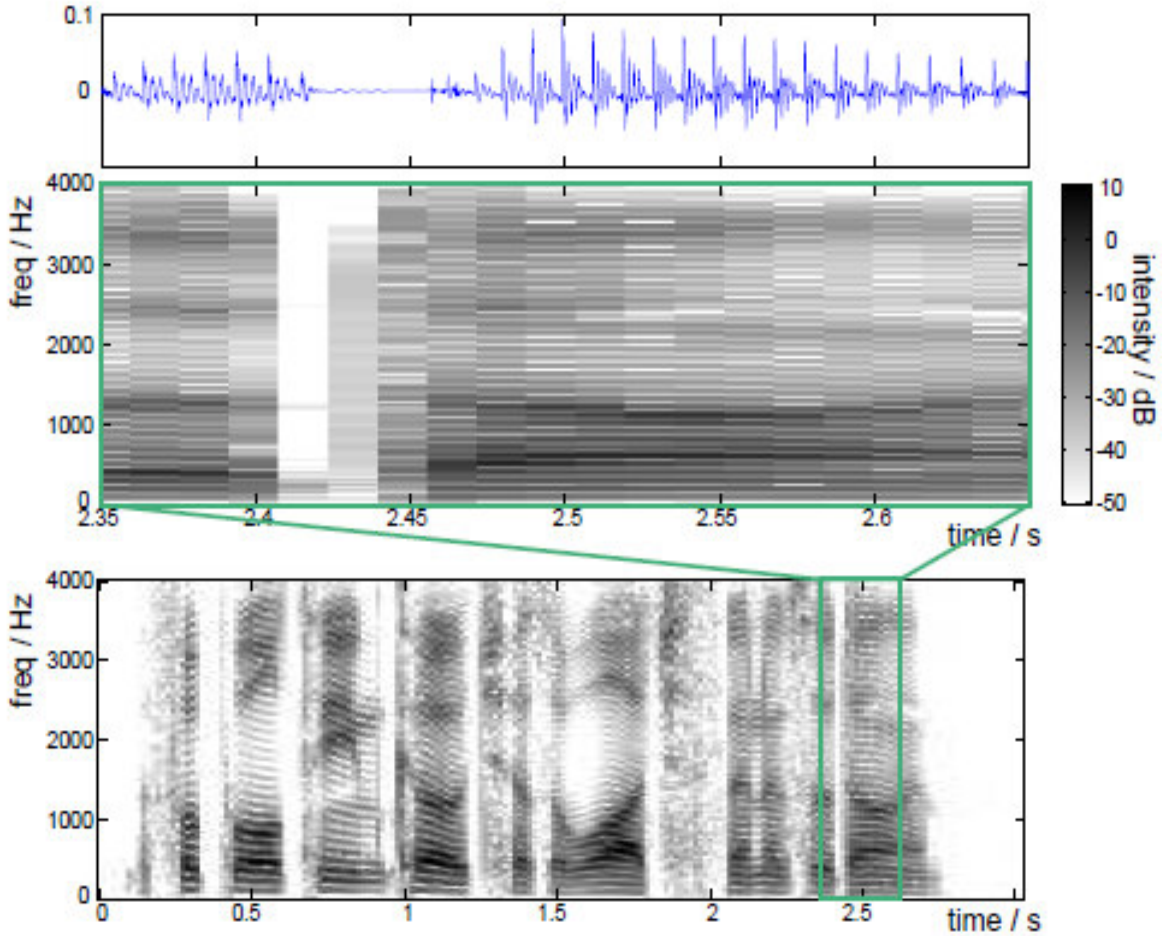
$$L = N_f/2$$

$$\cdots \qquad X[k.m] = \sum_{0}^{N_f-1} x[n]w[n-mL] \exp^{-j\frac{2\pi k(n-mL)}{N_f}}$$

DSP: introduzione

Spettrogramma

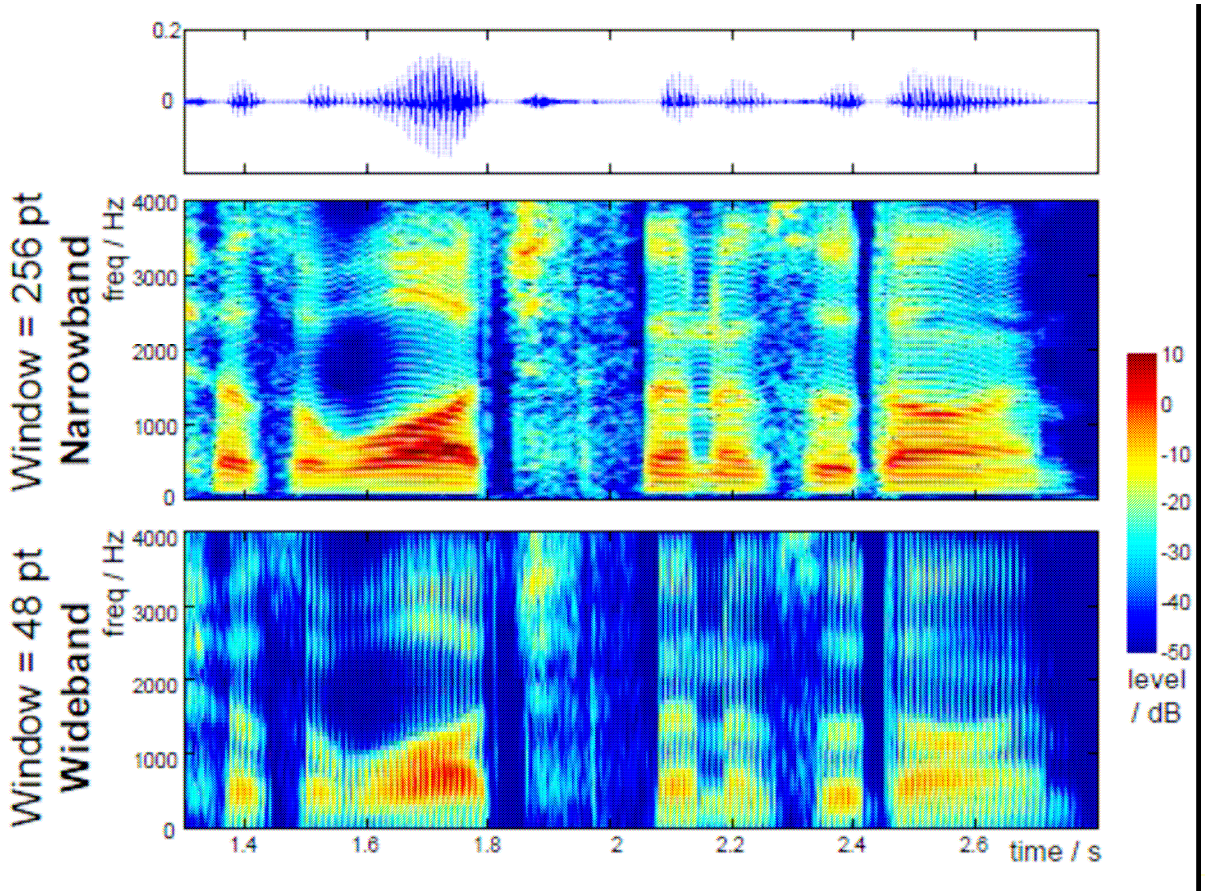
Rappresentazione grafica della sequenza dei **moduli** di $X[k, m]$
(intesità dei colori $\leftrightarrow$ valore del modulo per diversi istanti e frequenze)



DSP: introduzione

Rappresentazione tempo-frequenza: lunghezza di finestra

Ad una lunghezza maggiore della finestra di analisi corrisponde una migliore risoluzione in frequenza a scapito della risoluzione nel tempo.



DSP: introduzione

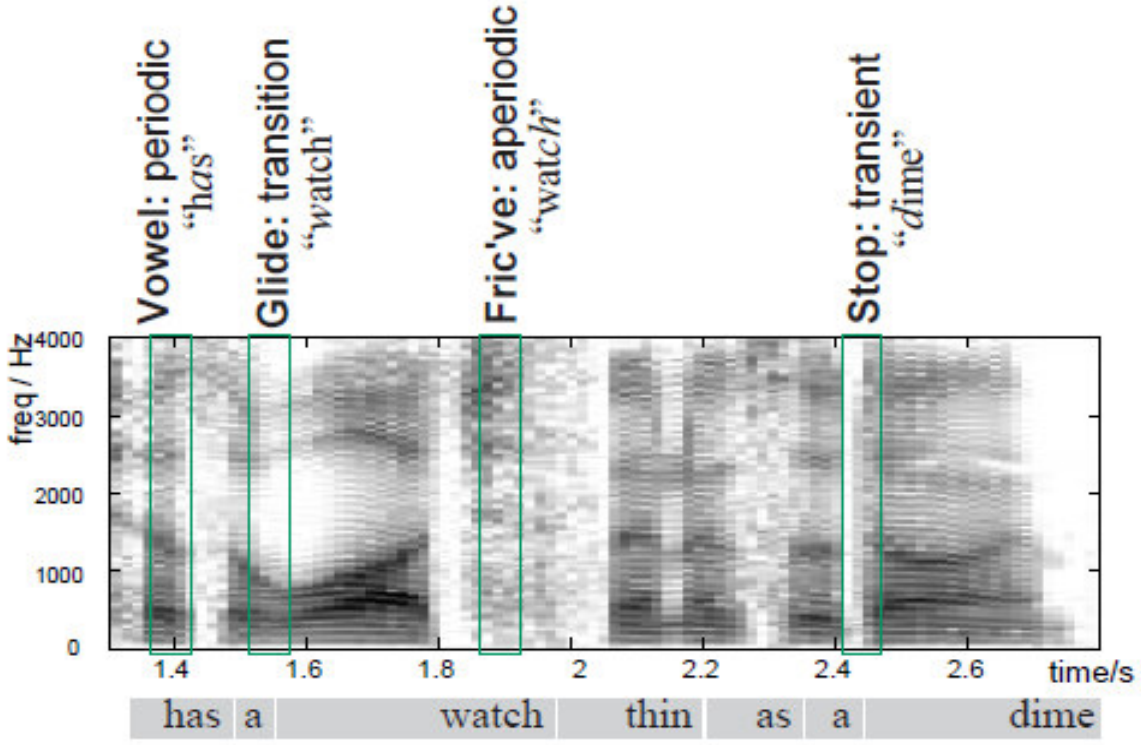
Spettrogramma: istruzioni Matlab

```
[s,Fs,nbit]=wavread('IlColombre_init.wav');

WindowSize=1024/2;
OverlapLen=1024/2;
Nfft=1024/2;

[y,f,t,p] = spectrogram(s,WindowSize,OverlapLen,Nfft,Fs,'yaxis');

surf(t,f,10*log10(abs(p)),'EdgeColor','none');
axis tight; colormap(jet); view(0,90);
xlabel('Time'); ylabel('Frequency (Hz)');
```

DSP: segnali vocali

Spettrogramma di suoni vocali

- Principale strumento di analisi in fonetica acustica
- Lo spettrogramma a banda stretta (alta risoluzione frequenziale) evidenzia le caratteristiche armoniche (es., sorgente periodica)
- Lo spettrogramma a banda larga (alta risoluzione temporale) mette in evidenza le formanti (tratto vocale)

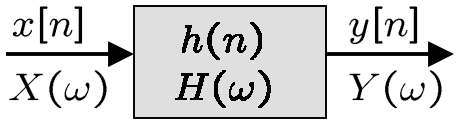


DSP: sistemi discreti e filtri digitali

Sistemi lineari tempo-invarianti (LTI) a tempo discreto

- Un blocco di elaborazione che trasforma una sequenza di campioni in ingresso $x[n]$ in una sequenza di uscita $y[n]$

- Un sistema LTI può essere descritto dalla sua risposta impulsiva $h(n)$ o dalla risposta in frequenza $H(\omega)$



- La relazione ingresso uscita è fornita dalla **convoluzione**:

$$y[n] = (h * x)[n] = \sum_{i=-\infty}^{+\infty} h[i]u[n-i]$$

- Teorema della **convoluzione in frequenza**:

$$y[n] = (h * x)[n] \Leftrightarrow Y(\omega) = H(\omega)X(\omega)$$

DSP: sistemi discreti e filtri digitali

Equazioni alle differenze e trasformata Z

- La relazione I/O di un sistema LTI si può descrivere con un'equazione alle differenze:

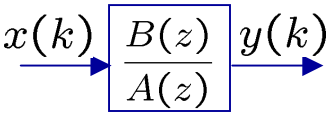
$$y[n]+a_1y[n-1]+a_2y[n-2]-\ldots = b_0x[n]+b_1x[n-1]+b_2x[n-2]+\ldots$$

$$\Downarrow Z-transform$$

$$Y(z)+a_1z^{-1}Y(z)+a_2z^{-2}Y(z)-\ldots = b_0X(z)+b_1z^{-1}X(z)+b_2z^{-2}X(z)+\ldots$$

$$\Downarrow$$

$$H(z)=\frac{Y(z)}{X(z)}=\frac{b_0+b_1z^{-1}+b_2z^{-2}+\ldots}{1+a_1z^{-1}+a_2z^{-2}-\ldots}$$



- La trasformata Zeta (corrispondente alla trasformata di Laplace nel continuo) permette di descrivere un sistema LTI con una f.d.t. polinomiale a coefficienti costanti
- Con questa notazione un sistema discreto LTI è caratterizzato da poli (radici del polinomio denominatore) e zeri (radici del numeratore)

DSP: sistemi discreti e filtri digitali

Filtri digitali

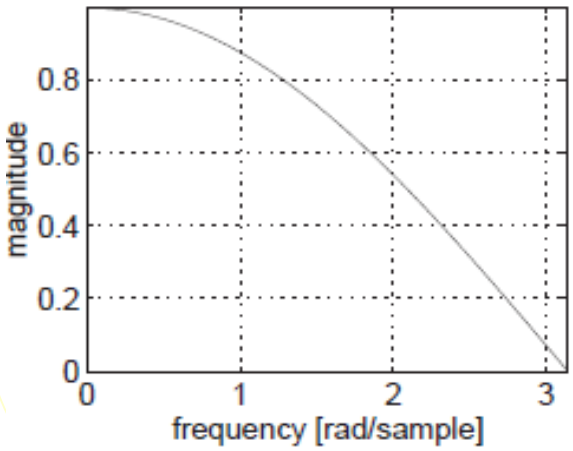
- Un filtro digitale è un sistema LTI operante su segnali discreti
- Comunemente si dividono i filtri digitali in **FIR** (finite impulse response) e **IIR** (infinite impulse response)

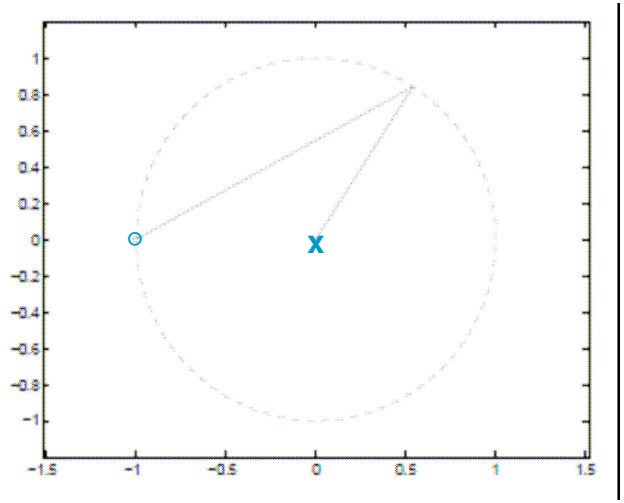
- Filtri **FIR**: $y[n] = \sum_{i=0}^{N-1} b_i x[n-i]$ $\qquad$ $H(z) = \sum_{i=0}^{N-1} b_i z^{-i}$
- **Esempio**: Filtro FIR passa basso del I ordine:

$$y[n] = 0.5x[n] + 0.5x[n-1]$$

$$H_{LP}(z) = 0.5 + 0.5z^{-1}$$

$$|H_{LP}(\omega)| = cos(\omega/2)$$





DSP: sistemi discreti e filtri digitali

Filtri digitali

- Filtri **IIR**:

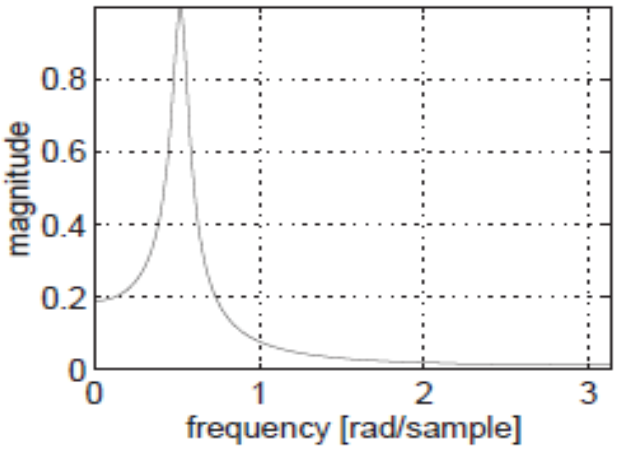
$$y[n] = \sum_{i=1}^{N_a} a_i y[n-i] + \sum_{i=0}^{N_b} b_i x[n-i]$$

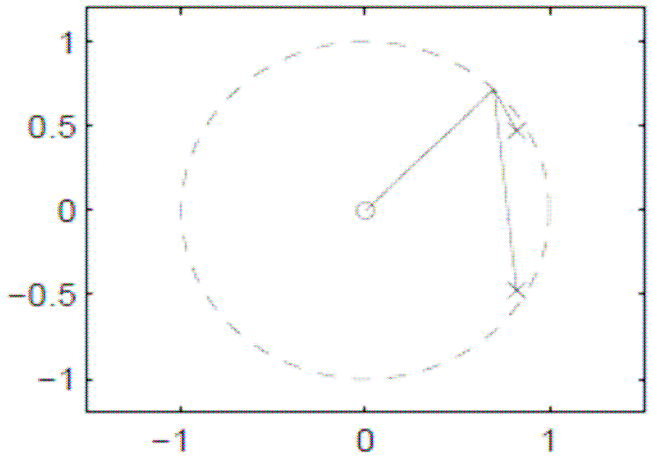
$$H(z) = \frac{\sum_{i=0}^{N_b-1} b_i z^{-i}}{1 + \sum_{i=1}^{N_a-1} a_i z^{-i}}$$

- **Esempio**: il filtro IIR del II ordine a soli poli (passa-banda)

$$y[n] = b_0 x[n] - a_1 y[n-1] - a_2 y[n-2]$$

$$H(z) = \frac{b_0}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

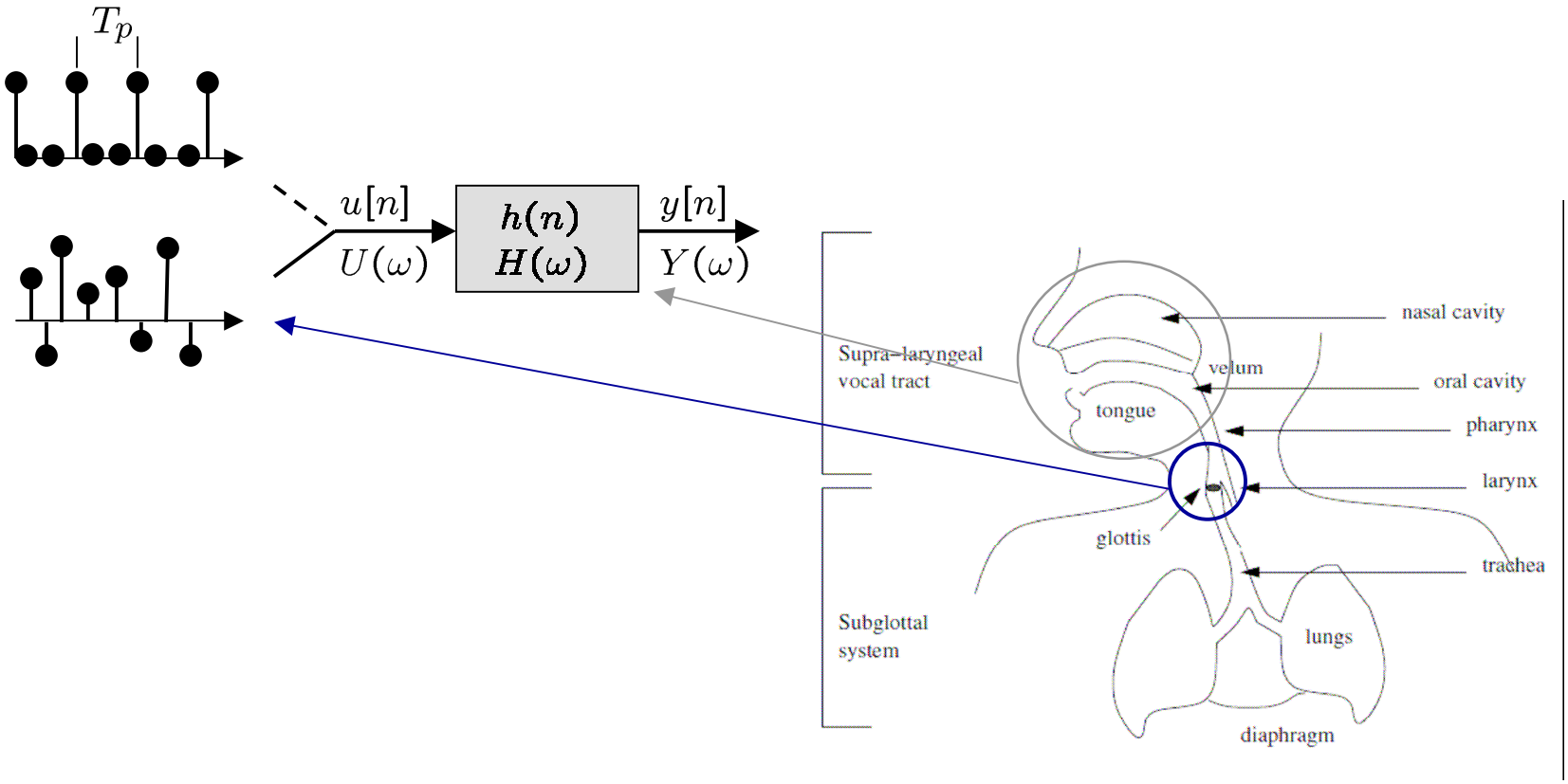




DSP: segnali vocali

Fonazione: modello sorgente-filtro

- La fonazione è rappresentata come il risultato del filtraggio da parte del tratto vocale di una sorgente impulsiva periodica (per suoni vocalici) o rumorosa (per le consonanti)

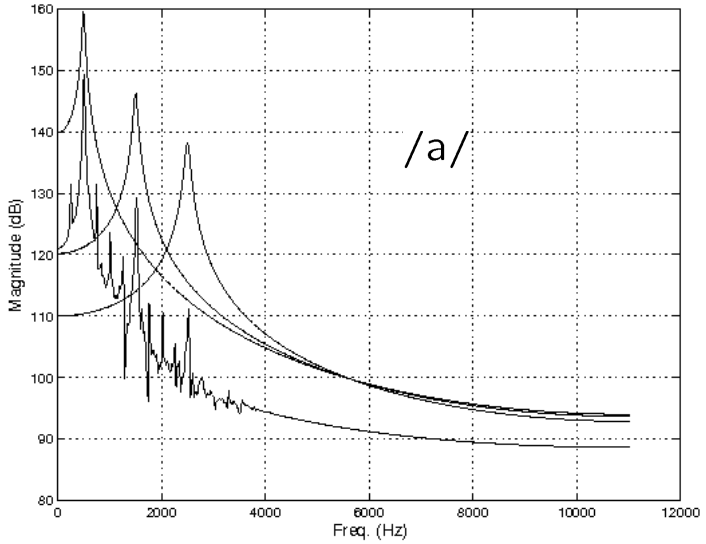


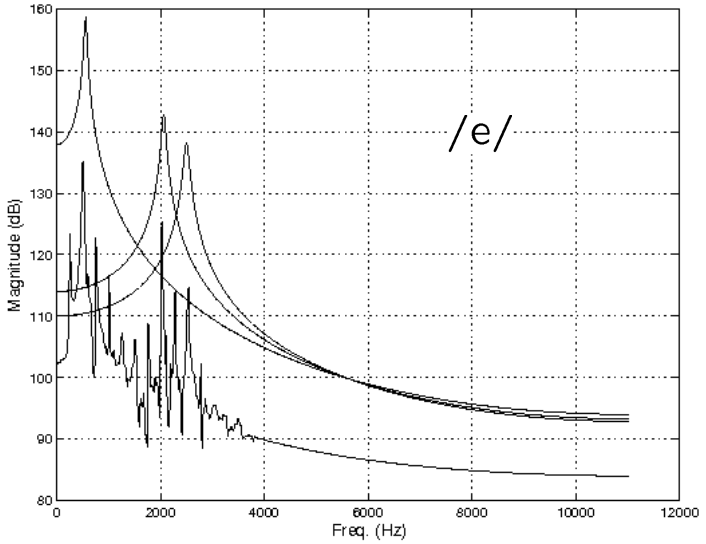
DSP: segnali vocali

Fonazione: modello sorgente-filtro

- Il tratto vocale presenta risonanze che si prestano ad essere rappresentate con filtri IIR del second'ordine

- **Esempio**: sintesi per formanti delle vocali /a/ ed /i/ usando una sorgente impulsiva e tre filtri IIR del II ordine:





Segnali vocali – analisi LPC

Analisi LPC: stima del filtro del tratto vocale dal segnale alle labbra

- Modello predittivo: stima del campione presente sulla base dei campioni passati.

$$y(k) = \sum_{i=1}^{n_a} a_i y(k-i) + e(k)$$

$e(k)$ → errore di predizione

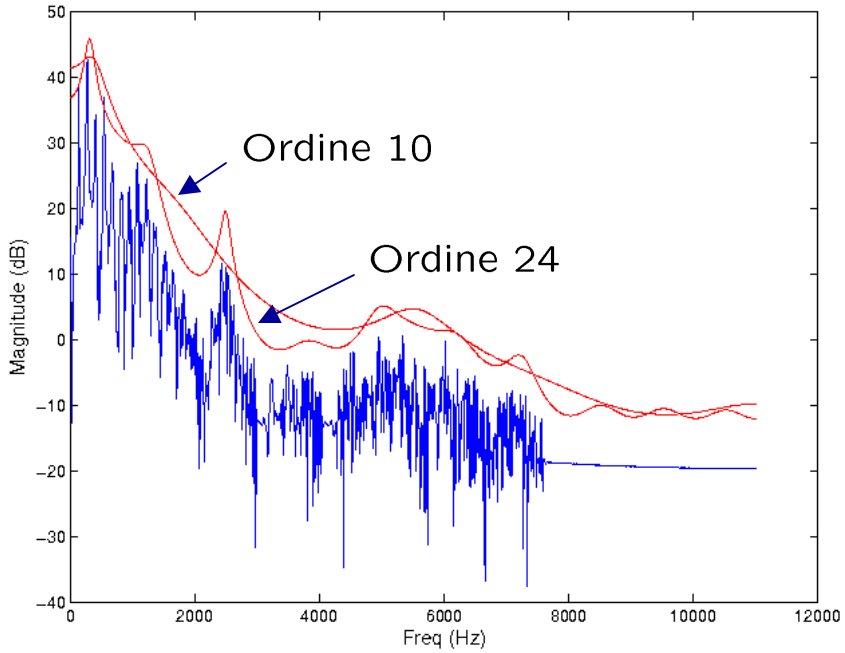
- La stima dei parametri $\{a_i\}$ avviene mediante calcolo della funzione di autocorrelazione del segnale e decorrelazione dell'errore di predizione (algoritmo di Levinson-Durbin)

- I coefficienti $\{a_i\}$ costituiscono una parametrizzazione compatta dell'inviluppo spettrale e contengono informazioni sulle formanti

Segnali vocali – analisi LPC

Analisi LPC

- L'ordine del filtro di analisi regola l'ordine di accuratezza con cui l'inviluppo spettrale viene rappresentato.



Segnali vocali – analisi LPC

LPC: istruzioni Matlab

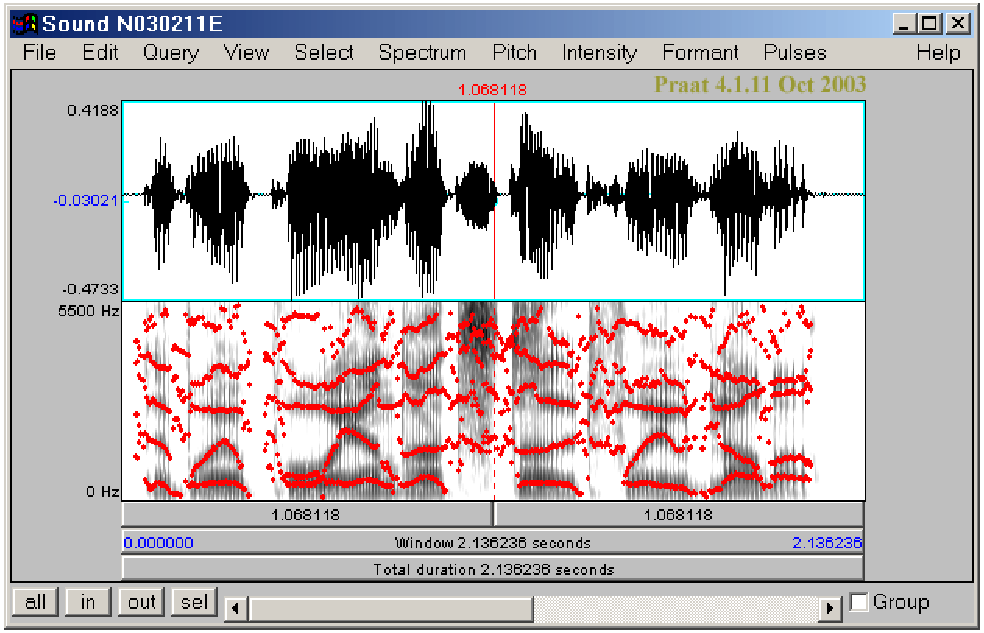
```
[s,Fs,nbit]=wavread('IlColombre_init.wav');

ti=3000;
Nwin=2048;
s_sel=s(ti:ti+Nwin-1);
S=rfft(s_sel,Nwin);

LpcOrd=20;
[A,g]=lpc(s_sel,LpcOrd);
[Hlpc,f]=freqz(g,A,Nwin/2+1,Fs);

figure
freqaxis=[0:Nwin/2]./Nwin*Fs;
plot(f,db(S))
hold on
plot(f,db(Hlpc)+100,'r')
```

Segnali vocali – analisi LPC

Analisi LPC e tracking delle formanti

- Gli algoritmi più noti di stima delle formanti si basano sulla analisi LPC del segnale secondo il seguente schema:
  - analisi LPC a intervalli regolari e stima delle formanti dai picchi di $1/A(z)$
  - applica criteri di inseguimento delle traiettorie e confronto con valori medi noti delle formanti

Esempio di inseguimento delle formanti ottenuto con il programma Praat



Segnali vocali – analisi LPC

LPC e tracking delle formanti: considerazioni

- La stima e l'inseguimento delle formanti risultano robusti prevalentemente per segmenti vocalici del segnale.

- Il genere del parlatore influisce sulla qualità della stima (la voce femminile presenta formanti meno definite di quelle della voce maschile)

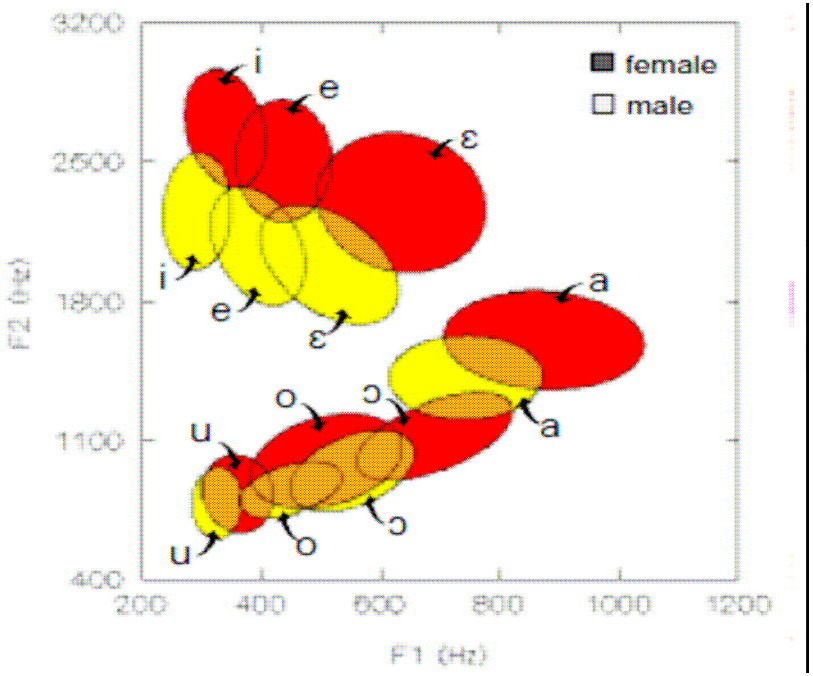
- Per vocali nasali si osserva la presenza di antirisonanze nello spettro. Queste in genere degradano la stima delle formanti dall'analisi LPC poiché possono provocare cancellazioni delle risonanze orali.

- Le vocali posteriori sono generalmente più difficili da analizzare poiché F_1 e F_2 si sovrappongono.

Segnali vocali – analisi LPC

Formanti e riconoscimento di vocali

- Dalla stima delle formanti è possibile stimare la vocale pronunciata dal parlatore, basandosi sui dati medi dei valori di F_1, F_2 ed F_3 noti in letteratura (traingolo vocalico)

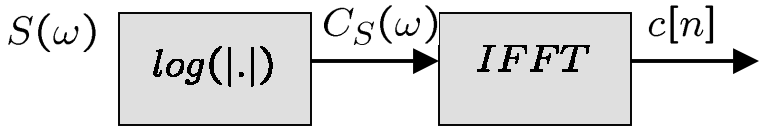


Esempio di triangolo vocalico
per l'italiano
(da Cosi, Ferrero e Vagges, 1995)

DSP – analisi mel-cepstrale

Mel-frequency cepstral coefficients analysis (MFCCs)

- Rappresentazione dell'inviluppo spettrale su base percettiva
- E' basata sui principi dell'analisi cepstrale. Il **cepstrum** $c(n)$ è calcolato secondo lo schema:



- Applicato alla voce ha la proprietà di deconvolvere naturalmente le componenti della sorgente e del tratto vocale:

$$s[n] = h[n] * e[n]$$
$$\Downarrow$$
$$|S(\omega)| = |H(\omega)||E(\omega)|$$
$$\Downarrow$$
$$log|S(\omega)| = log|H(\omega)| + log|E(\omega)|$$
$$\Downarrow$$
$$s_c[n] = h_c[n] + e_c[n]$$

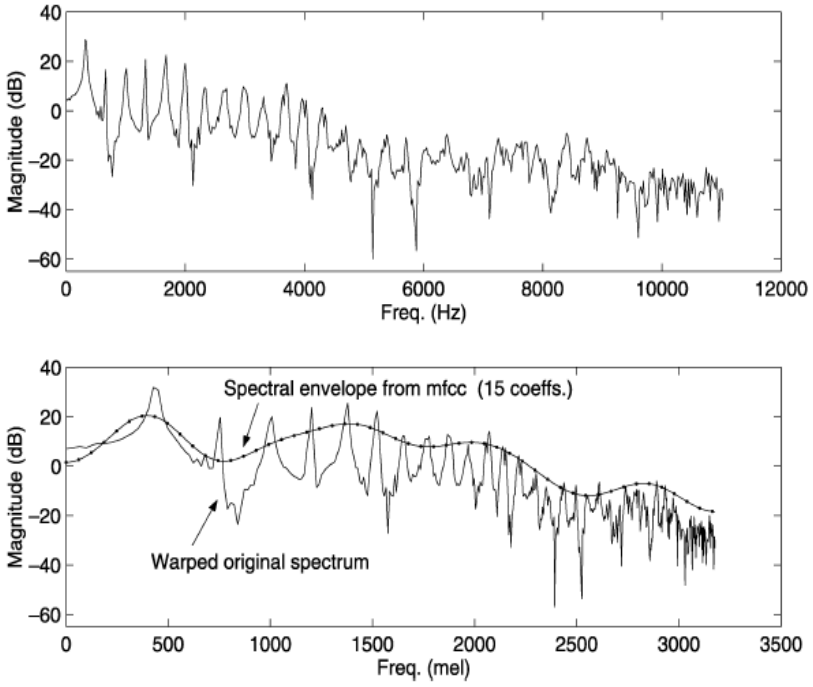
Segnali digitali – analisi melcepstrale

Mel-frequency cepstral coefficients analysis (MFCCs)

- Il **mel-cepstrum** è calcolato modificando lo schema di calcolo del cepstrum con l'introduzione di uno stadio di filtri percettivi su scala mel

- Permette di concentrare l'enfasi del modello sulle zone dello spettro percettivamente più interessanti

Rappresentazione dell'inviluppo spettrale mediante coefficienti mfcc



Segnali vocali – analisi LPC

Coefficienti mfcc: calcolo con Matlab

```
[s,Fs]=wavread('ah.wav');

p=70;
n=1024;
Nc=25; %n. di coefficienti
x=melbankm(p,n,Fs);
f=fft(s,n);

n2=1+floor(n/2);
frq = [0:n2-1]*Fs/n;
mel = frq2mel(frq);

z=log10(x*abs(f(1:n2)).^2);
c=dct(z); %mfcc's

melEnv=n2/Nc*idct(cz);
```

Analisi del segnale vocale – front-end acustici

Ulteriori parametri acustici per l'analisi e il riconoscimento vocale

- Parametri legati alla frequenza fondamentale (**pitch**) e altri parametri derivati (**shimmer**, **jitter**, **HNR**). Gli algoritmi di pitch detection si basano sull'individuazione di serie armoniche di righe nel dominio dell frequenza, sull'analisi di picchi nel residuo LPC, o sull'uso fi funzioni di autocorrelazione.

- Parametri legati all' intensità del segnale: **potenza**, **energia**, **RMS**.

- Varianti dell'analisi LPC: **LSP** (robusti rispetto a interpolazione), **PLP** (perceptual linear prediction), **Rasta** (per segnali rumorosi).

- Parametri **Delta** e **Delta-Delta**: derivate prime e seconde dei parametri acustici (scalari o vettoriali).

Segnali digitali – strumenti software

Alcuni strumenti software per l'analisi del segnale vocale

- **Matlab**. Sono disponibili Toolbox per le principali funzioni di analisi del segnale in generale (signal processing toolbox) e del segnale vocale (ad es., il toolbox voicebox)

- **Praat**. Fornisce routine robuste per il calcolo del pitch e delle formanti

- **Wavesurfer**. Strumento per la visualizzazione e manipolazione dei suoni vocali. Fornisce algoritmi per il calcolo del pitch e strumenti per la trascrizione.

Riferimenti bibliografici

J. R. Deller, J. G. Proakis, and J.H.L. Hansen, *Discrete-time processing of speech signals*, Prentice Hall, 1987.

J. W. Picone, "Signal modeling techniques in speech recognition," *Proceedings of the IEEE*, vol. 81, no. 9, pp. 1215–1247, September 1993.