

# Analisi Audio in Frequenza

---

Elaborazione dell'audio digitale

*Ingegneria del Cinema, Informatica e Telecomunicazioni*

**Antonio Servetti**

Internet Media Group

Dip. di Automatica ed Informatica

Politecnico di Torino

[servetti@polito.it](mailto:servetti@polito.it)

<http://media.polito.it>

# Sommario

---

- Sez.1 – ANALISI SPETTRALE
  - ✓ Suoni (toni) puri e complessi
  - ✓ Caratteristiche e analisi dello spettro di un segnale
  - ✓ Descrittori (features) spettrali
  
- Sez.2 – SEGMENTAZIONE
  - ✓ Importanza dell'intervallo temporale
  - ✓ Tempo varianza e stazionarietà locale
  - ✓ Attività – Analisi del segnale vocale: fonemi

# Riferimenti bibliografici

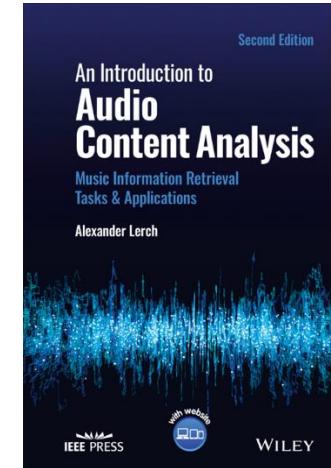
---

- [LERCH] Lerch, "An Introduction to Audio Content Analysis"
  - ✓ Ch.3.5 "Instantaneous Features"
    - Spectral Centroid, Spectral Spread, Spectral Rolloff, Zero Crossing Rate, ...
  - ✓ Code: <https://github.com/alexanderlerch/pyACA>
- [MULLER] Muller, "Fundamentals of Music Processing"
  - ✓ Ch.2 Fourier Analysis of Signals
  - ✓ Code: <https://audiolabs-erlangen.de/resources/MIR/FMP/C0/C0.html>
- [PRAAT] Praat Software – [praat.org](http://praat.org)  
Weenink, "Speech Signal Processing with Praat"  
<https://www.fon.hum.uva.nl/david/sspbook/sspbook.pdf>

# Lerch, "An Introduction to Audio Content Analysis"

---

- Part 1 – "Fundamentals of Audio Content Analysis"
  - ✓ Analysis of audio signals
  - ✓ Inference (classification, clustering, distance and similarity)
- Part 2 – "Music Transcription"
  - ✓ Tonal analysis (pitch detection, chord recognition)
  - ✓ Intensity (RMS, peak envelope, loudness perception)
  - ✓ Temporal analysis (onsets detection, tempo/beat detection)
  - ✓ Alignment (dynamic time warping, etc.)
- Part 3 – "Music Identification, Classification, ..."
  - ✓ Audio Fingerprinting, Music Similarity Detection, Music Genre Classification, Mood Recognition, Musical Instrument Recognition



2ed, 2022

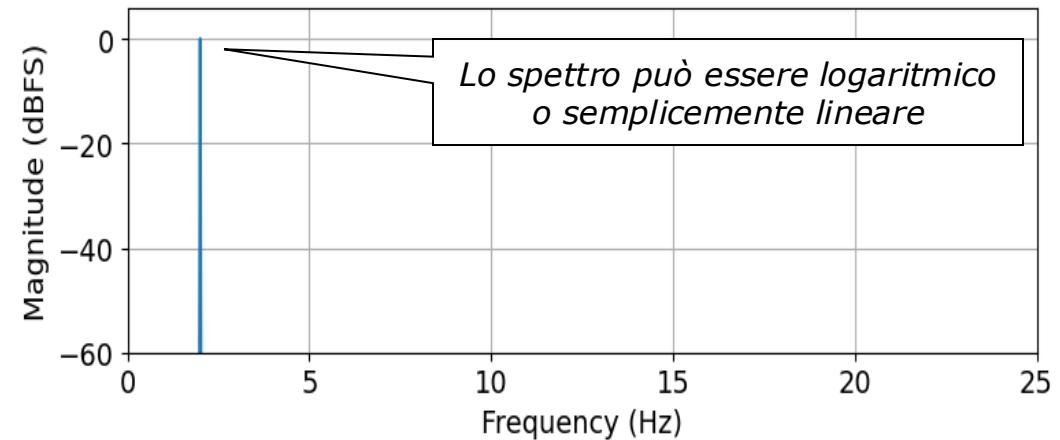
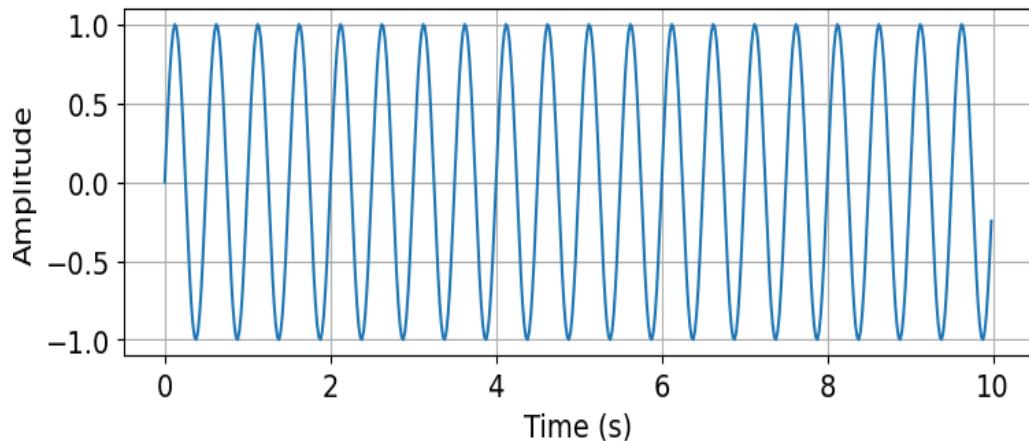
# Forma d'onda e spettro

## ■ Forma d'onda

- ✓ Ogni punto è un campione, rappresentato come l'ampiezza del segnale misurato in un istante temporale

## ■ Spettro (d'ampiezza)

- ✓ Ogni punto è una componente in frequenza del segnale, rappresentata come l'ampiezza della sinusoide a quella frequenza



---

## ANALISI SPETTRALE (con rudimenti di sintesi audio)

# ANALISI SPETTRALE

---

## ■ **Suoni (toni) puri**

- ✓ Il problema dell'aliasing
- ✓ Ritardo di fase

## ■ Caratteristiche e analisi dello spettro di un segnale

- ✓ Suoni periodici: altezza/pitch e armoniche
- ✓ Inviluppo spettrale

## ■ Descrittori (features) spettrali

# Sintesi (co-)sinusoide – tono puro

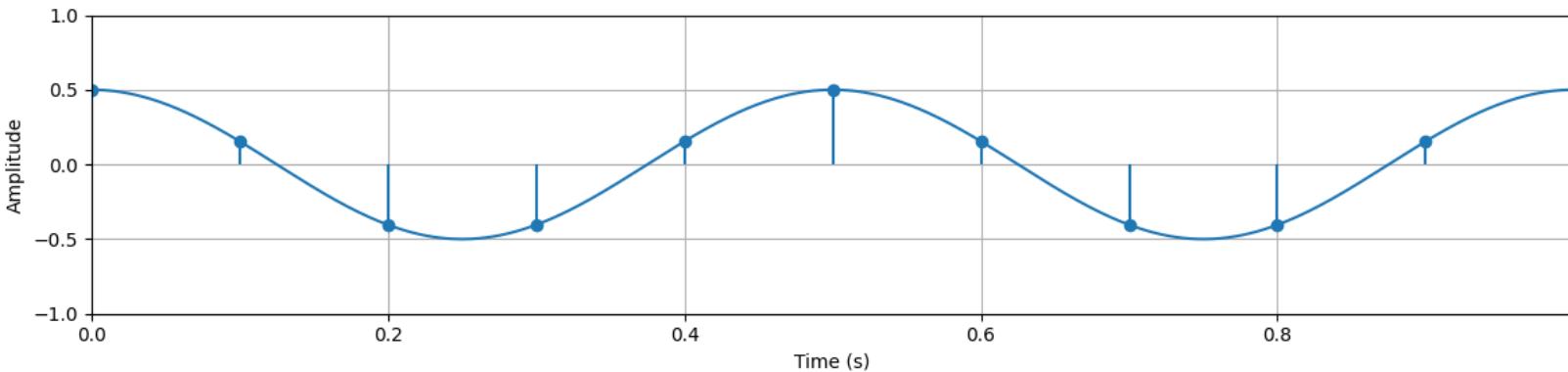
- Una sinusoide a frequenza  $f$  genera un tono puro

✓  $w = A \cdot \cos(2\pi \cdot f \cdot t)$

- Come generare una sinusoide digitale?

✓  $A = 0.5, f = 2$

Segnale audio con una frequenza sola  
(praticamente non esiste in natura)



Esempio didattico: usiamo frequenze molto basse, non udibili, ma più facili da visualizzare e con cui fare i calcoli.

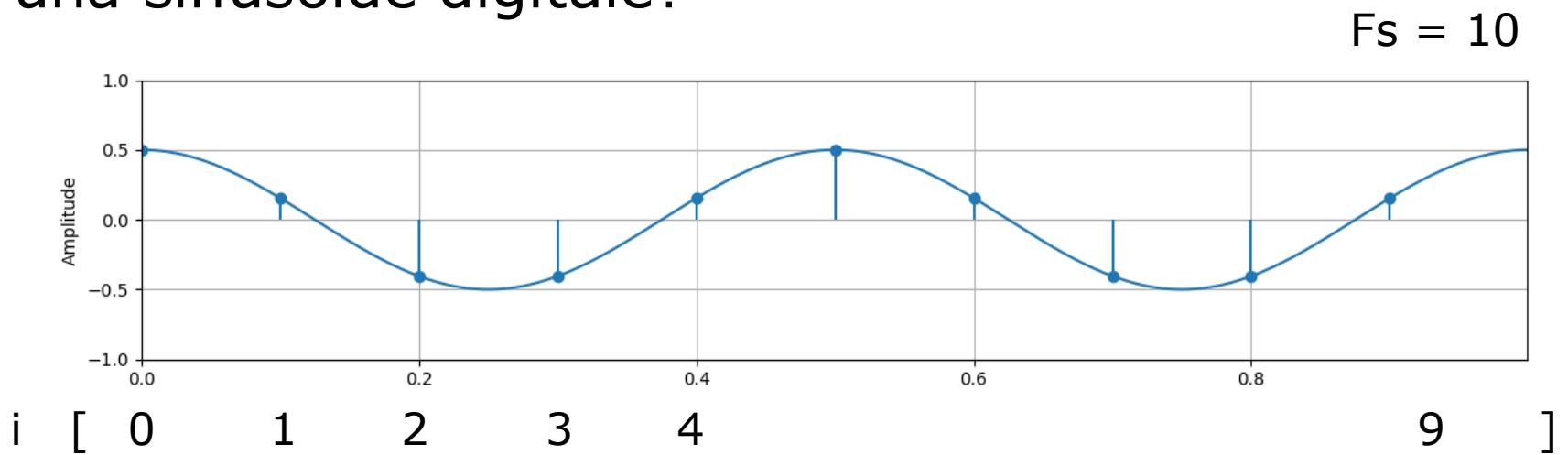
# Sintesi (co-)sinusoide – tono puro

- Una sinusoide a frequenza  $f$  genera un tono puro

✓  $w = A \cdot \cos(2\pi \cdot f \cdot t + \varphi_0)$

- Come generare una sinusoide digitale?

✓  $A = 0.5, f = 2$



- Occorre valutare il coseno negli istanti di campionamento

$$\hat{t}_i = i \cdot 1/F_s$$

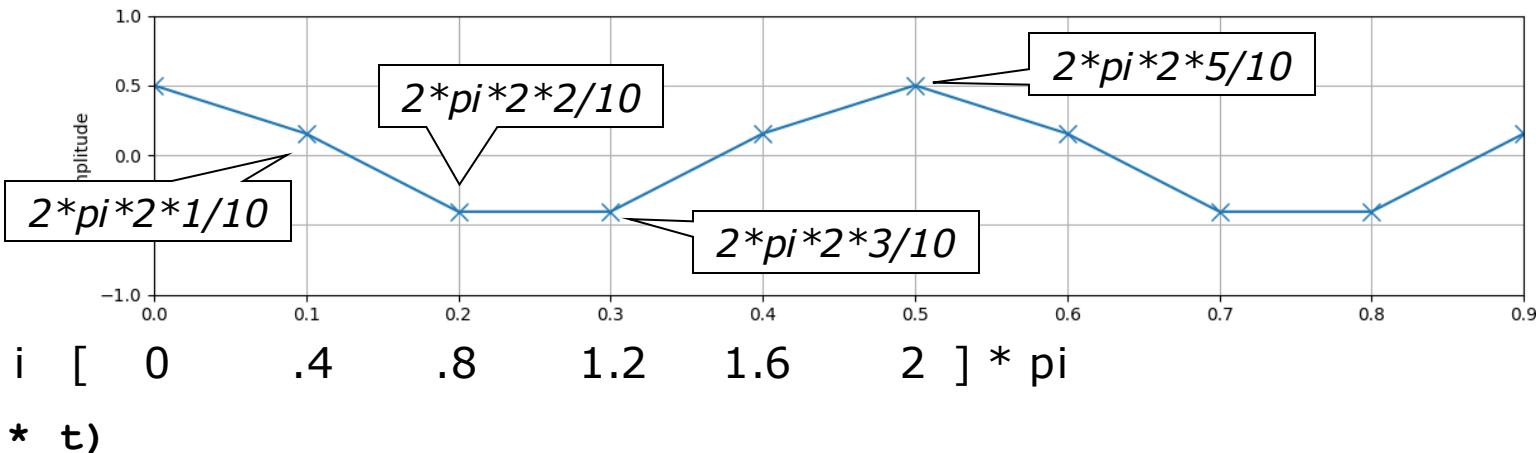
$F_s = 10 \Rightarrow 0, 0.1, 0.2, \dots$

Esempio didattico: usiamo frequenze molto basse, non udibili, ma più facili da visualizzare e con cui fare i calcoli.

# Sintesi tono puro

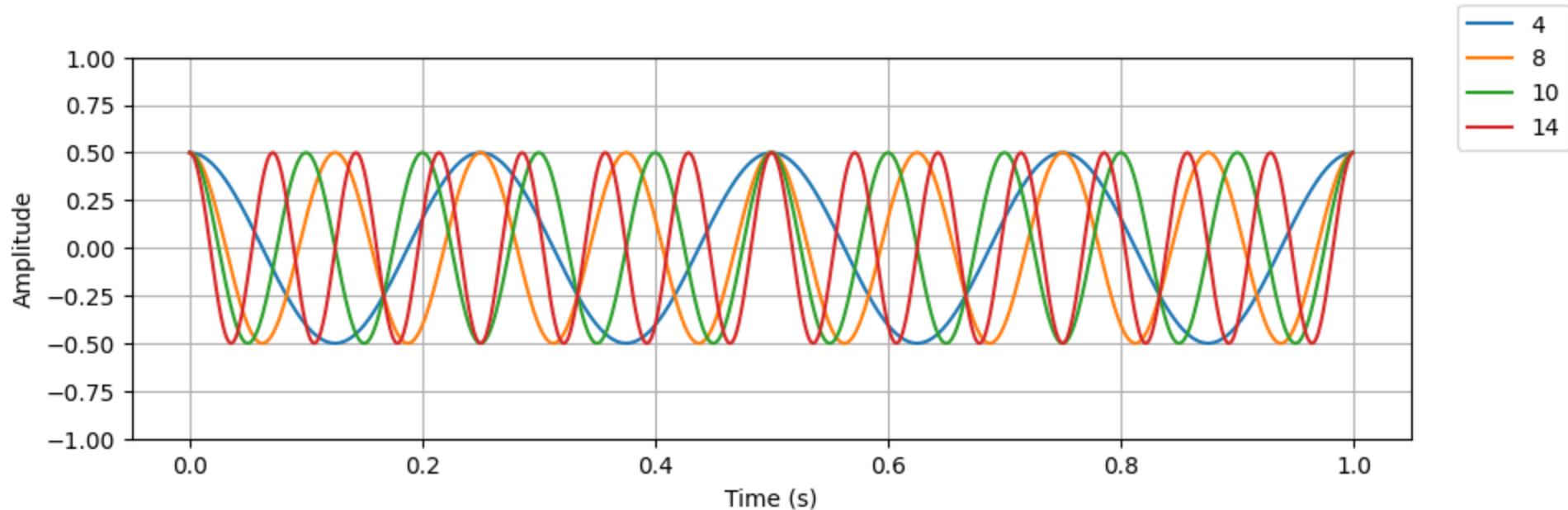
- Generazione degli istanti di campionamento
  - ✓ Genero N punti a distanza  $\Delta t = 1/F_s$ :  $t = np.arange(N) * 1/F_s$
  - ✓ La durata è data dal numero di campioni N per la distanza tra gli stessi  $\Delta t$  quindi  $N * 1/F_s$  (secondi)
- L'argomento del coseno è il valore della fase ISTANTANEA in rad.
  - ✓ Indica a che punto dell'oscillazione si trova l'onda all'istante  $t_i$

```
import numpy as np  
N = 10  
F_s = 10  
A = 0.5  
f = 2  
t = np.arange(N) / F_s  
s = A * np.cos(2 * np.pi * f * t)
```



# Esercizio

- Domanda – l'importanza della frequenza di campionamento
  - ✓ Continuiamo a supporre una frequenza di campionamento di 10 Hz
  - ✓ Quale forma hanno le sinusoidi corrispondenti alle frequenze di
    - 4, 8, 10, 14 Hz ?

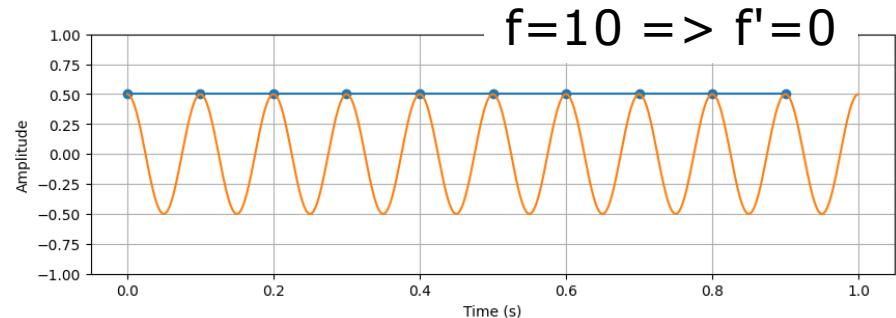


# Esercizio - aliasing

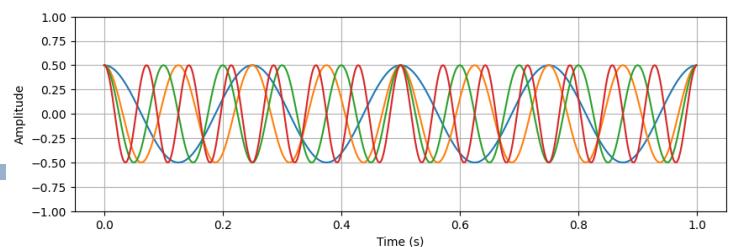
## ■ Domanda

- ✓ Continuiamo a supporre una frequenza di campionamento di 10 Hz
- ✓ Quale forma hanno le sinusoidi corrispondenti alle frequenze di
  - 4, 8, 10, 14 Hz ?

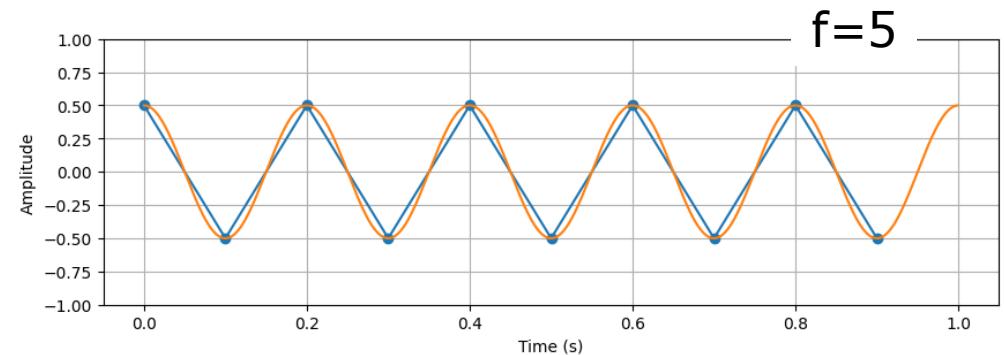
■ Con  $F_s = 10$  Hz non è possibile rappresentare frequenze superiori alla frequenza di Nyquist  $F_s/2 = 5$  che ha 2 campioni per periodo



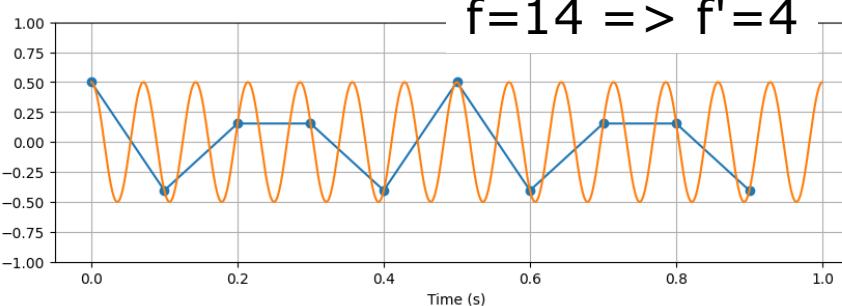
$$f=10 \Rightarrow f'=0$$



4  
8  
10  
14



$$f=5$$

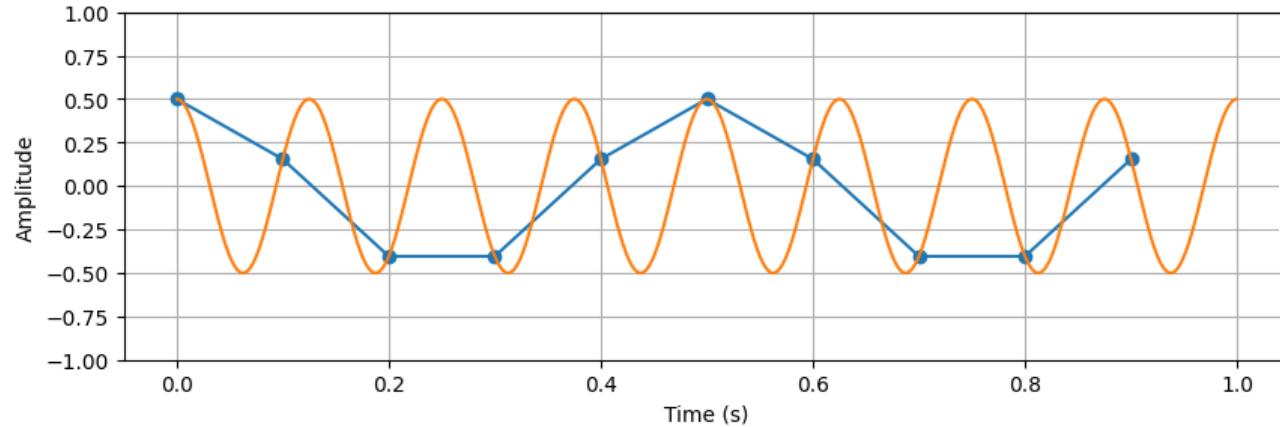


$$f=14 \Rightarrow f'=7$$

$$f=8 ?$$

# Esercizio - foldover

- Le frequenze tra  $F_s$  e  $F_s + F_s/2$
- Le frequenze tra  $F_s/2$  ed  $F_s$  subiscono foldover
  - ✓ Indicando con  $f$  la frequenza di sintesi e  $f'$  la frequenza risultante
  - ✓  $f \Rightarrow f' = [ \begin{array}{ccccccc} 0 \Rightarrow 0 & 1 \Rightarrow 1 & 2 \Rightarrow 2 & 3 \Rightarrow 3 & 4 \Rightarrow 4 & 5 \Rightarrow 5 \\ 6 \Rightarrow -4 & 7 \Rightarrow -3 & 8 \Rightarrow -2 & 9 \Rightarrow -1 & 10 \Rightarrow 0 & 11 \Rightarrow 1 \dots \end{array} ]$



$f=8 <> f=-2$

A "specchio" (negative)  
Coseno simmetrico  
Seno antisimmetrico

```
import numpy as np
N, Fs, A, f = 10, 10, 0.5, 5
t = np.arange(N) / Fs
s = A * np.cos(2 * np.pi * f * t)
```

# Sintesi tono puro

---

```
def gen_cos(dur=1, Fs=100, amp=1, freq=2, phase=0):
    num_samples = int(Fs * dur)
    t = np.arange(num_samples) / Fs
    x = amp * np.cos( 2*np.pi*freq*t) + phase)
    return x, t

s, t = gen_cos(dur=1, Fs=Fs, amp=A, freq=f, phase=phase)
```

# Effetto della fase

---

- Seno e coseno differiscono per la fase iniziale
  - ✓  $\cos(x) = \sin(x+\pi/2)$  e  $\sin(x) = \cos(x-\pi/2)$
- L'orecchio non è sensibile alla fase assoluta
  - ✓ Il tono (puro) di un seno o di un coseno è indifferente
  - ✓ Diventa sensibile nei segnali complessi se la fase ha un effetto significativo sull'inviluppo temporale del segnale
    - Caso classico somma di microfoni in controfase

Approfondimento => "Ritardo di fase / temporale"

---

# Ritardo di fase vs ritardo temporale

- L'argomento della funzione coseno può essere visto sia come *fase istantanea* sia come *istante temporale*

$$A \cos(\varphi) = A \cos(2\pi f t)$$

- Che relazione esiste?

✓ Uguagliando i punti a  $\varphi = \pi$  e a  $\varphi = 2\pi$  in figura otteniamo

$$A \cos(2\pi) = A \cos(2\pi f \cdot 0.5)$$

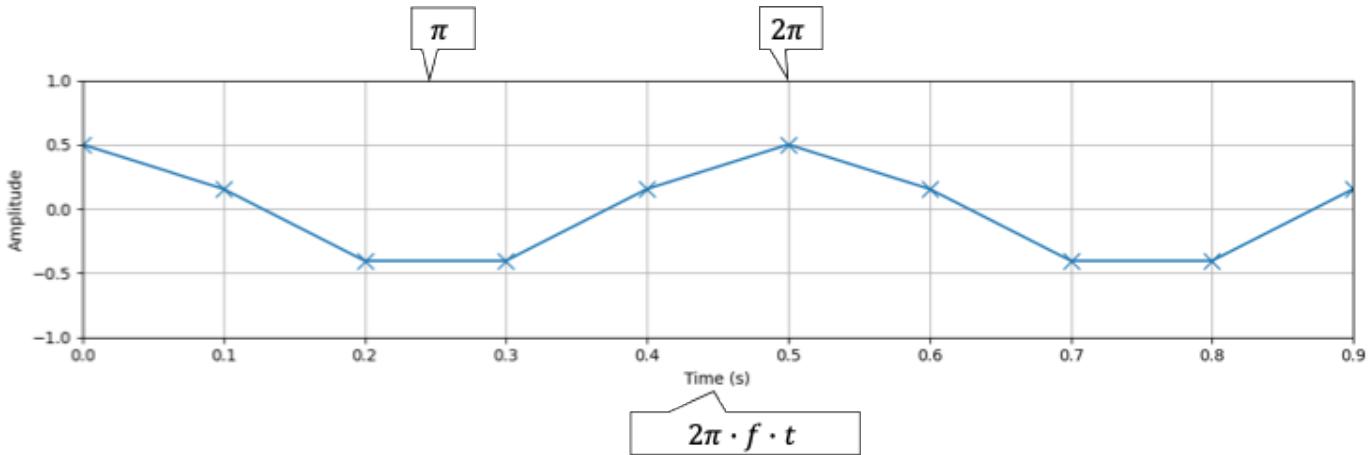
$$2\pi = 2\pi \cdot f \cdot 1/2$$

$$f = 2$$

✓ Cosa significa?

Perché 2?

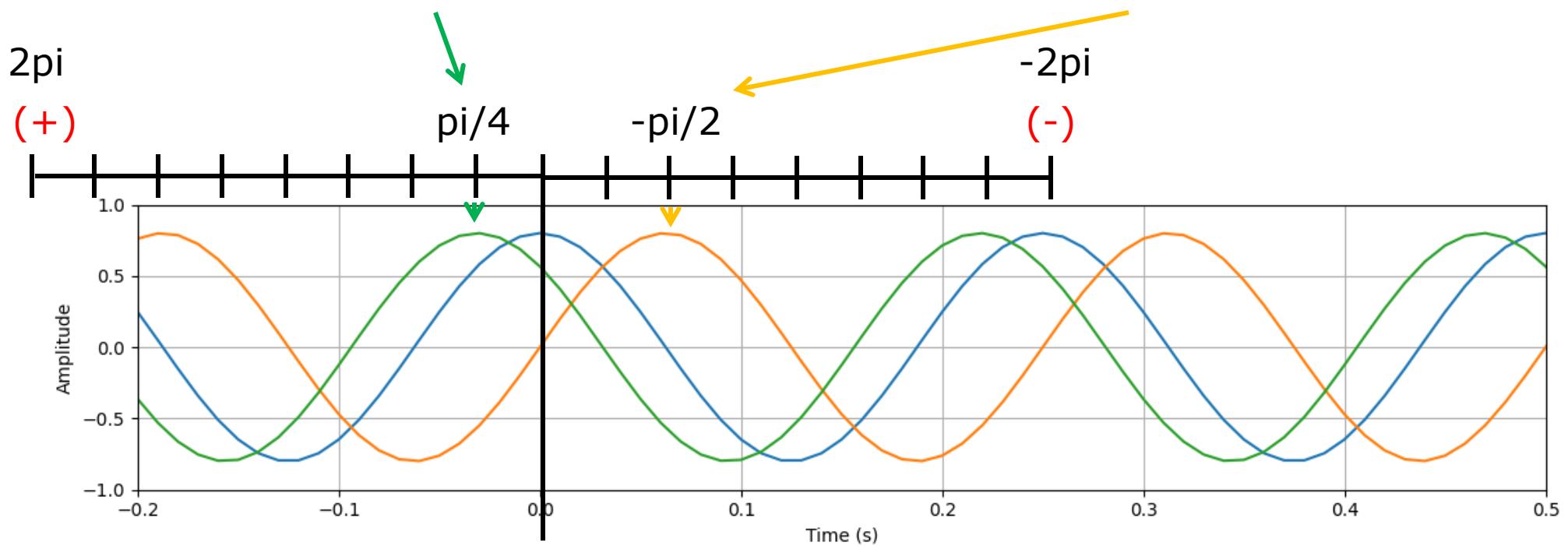
E se si rappresentasse un coseno a  $f = 4$ , cosa cambierebbe?



# Ritardo di fase

- Si può identificare la fase di un coseno localizzando il picco più vicino allo 0 e invertendone il segno

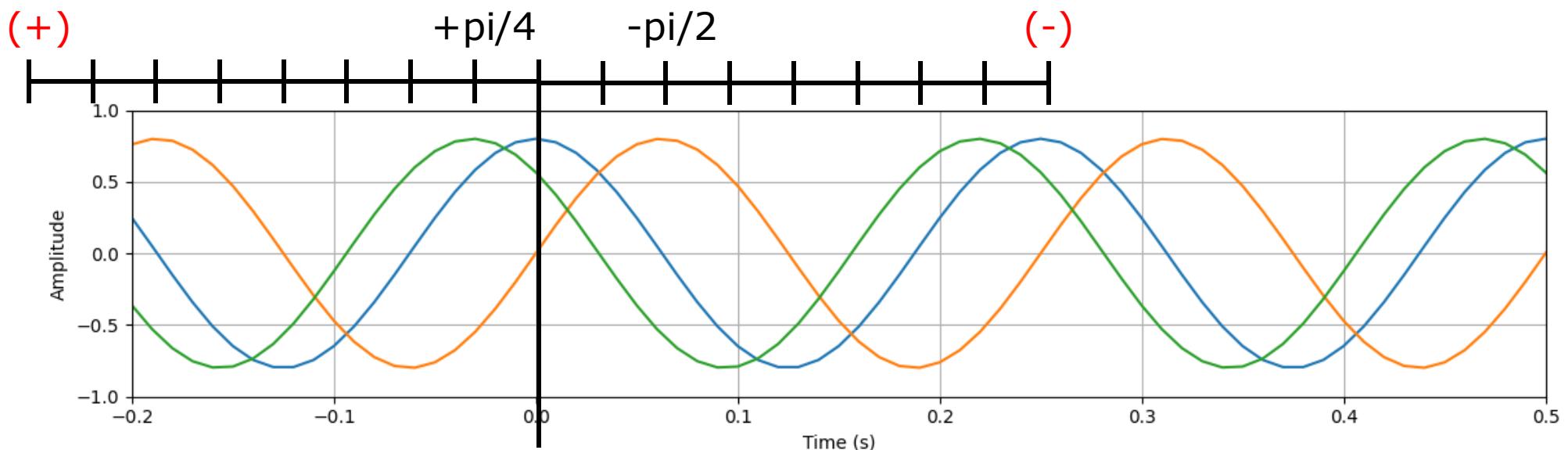
✓ verde:  $\cos(2\pi f t + \pi/4)$  arancio:  $\cos(2\pi f t - \pi/2)$



# Ritardo temporale

- Il ritardo di fase può anche essere visto come un ritardo temporale  
 $A \cos(2\pi f(t - t_0))$

- ✓ arancio: picco più vicino dopo lo zero "in ritardo",  $t_0$  positivo
  - ✓ verde: picco più vicino prima dello zero "in anticipo",  $t_0$  negativo
- la relazione tra ritardo di fase e ritardo temporale ha il segno opposto*



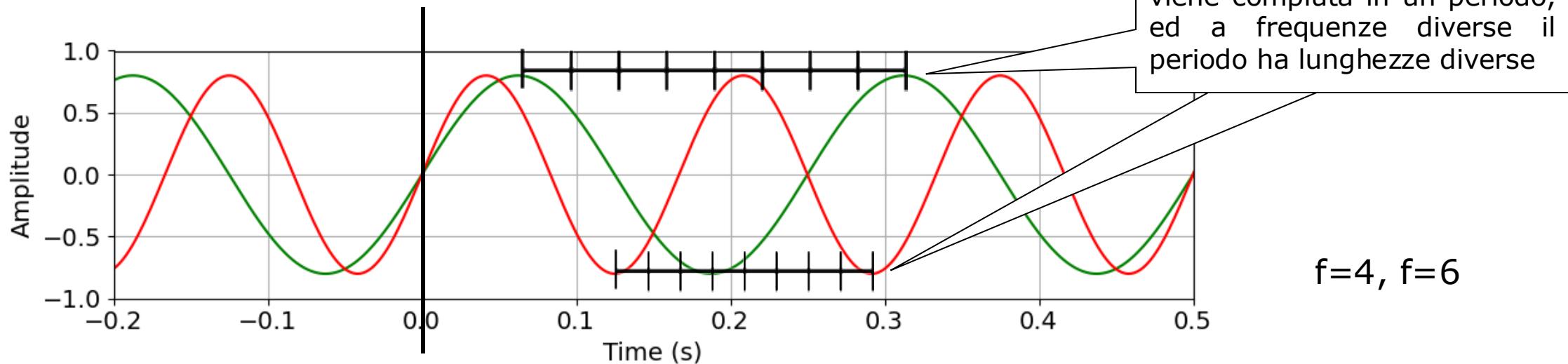
# Ritardo temporale vs ritardo di fase

- Uguagliando le espressioni si deriva la relazione

$$A \cos(2\pi f t + \varphi_0) = A \cos(2\pi f(t - t_0)) \Rightarrow t_0 = -\frac{\varphi}{2\pi f}$$

A parità di ritardo di fase il ritardo temporale è funzione della frequenza / periodo.

- In funzione del periodo  $T = 1/f \Rightarrow t_0 = -\frac{\varphi}{2\pi} T$



# Ritardo temporale vs ritardo di fase

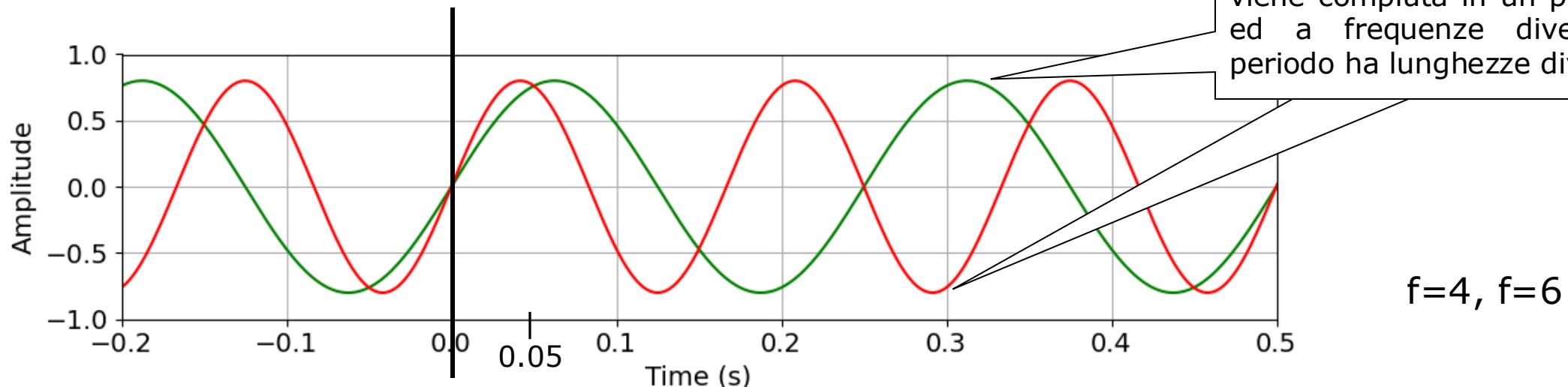
- In funzione del periodo  $T = 1/f \Rightarrow t_0 = -\frac{\varphi}{2\pi} T$

- Esempio

- ✓ Dato un ritardo di fase di  $-\pi/2$  il ritardo temporale
- ✓ In generale è positivo ed è  $(\pi/2 / 2*\pi) = 1/4$  del periodo
  - $f = 4, T = 0.25 \Rightarrow T/4 = 0.0625$
  - $f = 6, T = 0.16 \Rightarrow T/4 = 0.04$

A parità di ritardo di fase  
il ritardo temporale è  
funzione della frequenza  
/ periodo.

Infatti, una rotazione di  $2\pi$   
viene compiuta in un periodo,  
ed a frequenze diverse il  
periodo ha lunghezze diverse



# ANALISI SPETTRALE

---

- Suoni (toni) puri
  - ✓ Il problema dell'aliasing
  - ✓ Ritardo di fase
- **Caratteristiche e analisi dello spettro di un segnale**
  - ✓ Suoni periodici: altezza/pitch e armoniche
  - ✓ Involucro spettrale
- Descrittori (features) spettrali

# Suoni complessi periodici

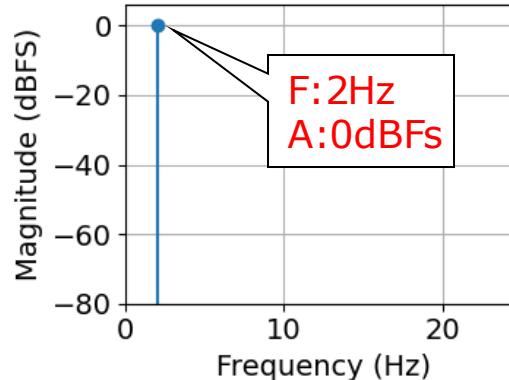
---

- Dei suoni complessi periodici è possibile determinarne il periodo  $T$  e di conseguenza l' **altezza** del suono  $F_0 = 1/T$
- **$F_0$**  NON è detta frequenza del suono (perché ce ne sono tante) ma **FREQUENZA FONDAMENTALE** o PITCH
- Per creare suoni complessi periodici occorre sommare (solo e soltanto) componenti (toni) con  $f_k = k * F_0$  dove con  $k$  INTERO
  - ✓  $F_0$  è quindi il M.C.D. delle frequenze dei toni presenti nel segnale
- Le componenti in frequenza (toni) dei suoni complessi periodici sono dette **armoniche**
  - 1° armonica:  $F_0$ , (in generale)  $k$ -esima armonica:  $f_k = k * F_0$

# Rappr. nel dominio della frequenza

## ■ Spettro (di Ampiezza o di Potenza)

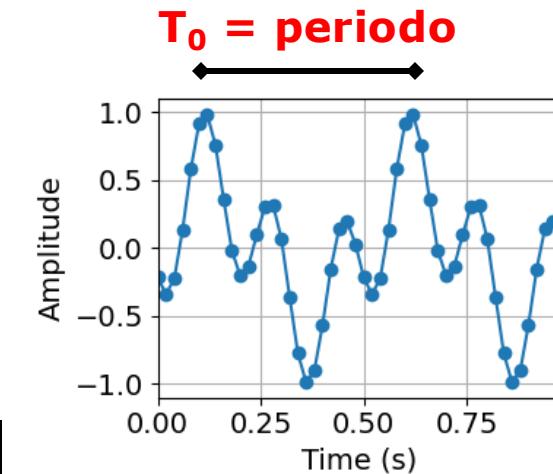
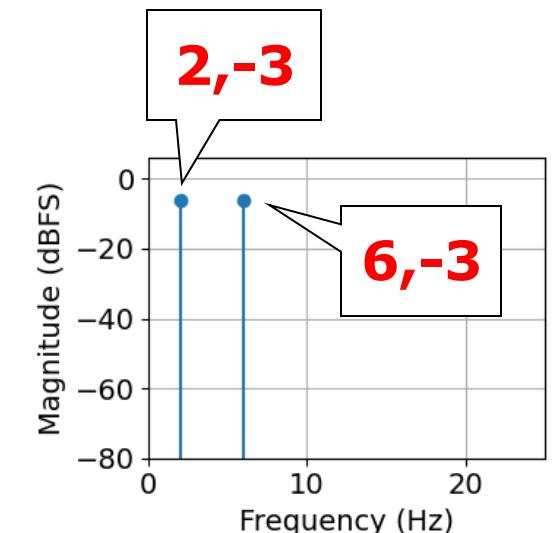
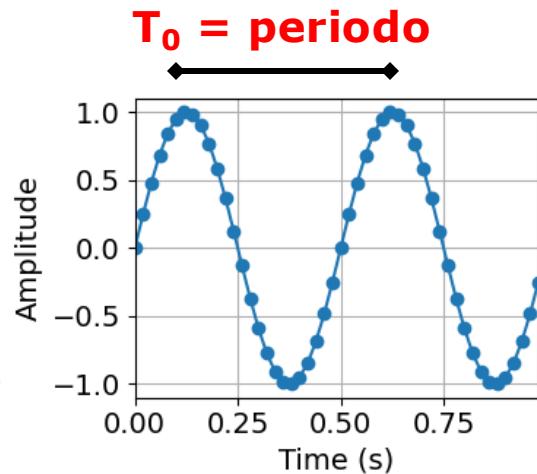
- ✓ Ogni punto è l'intensità (di una componente) del segnale (tono) ad una certa frequenza



## ■ Esempio

- ✓ Punto a 2 Hz, tono puro 2Hz
- ✓ Punto a 2 Hz + punto a 6 Hz, tono complesso

```
w1 = np.cos(2 * np.pi * 2 * t)
w2 = np.cos(2 * np.pi * 6 * t + np.pi/2)
w = (w1 + w2)/2
```

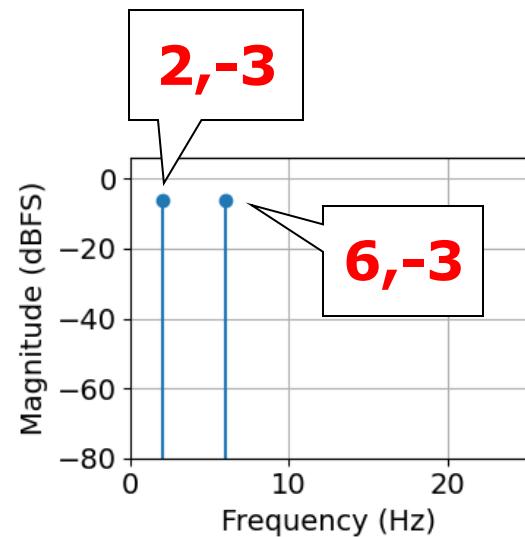


# Rappr. nel dominio della frequenza

## ■ Spettro (di Ampiezza o di Potenza)

- ✓ Ogni punto è l'intensità (di una componente) del segnale (tono) ad una certa frequenza

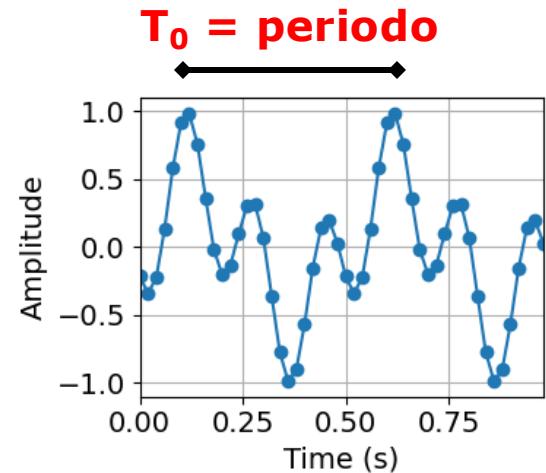
2 Hz prima armonica  
4 Hz seconda armonica  
— (assente)  
6 Hz terza armonica



## ■ Esempio

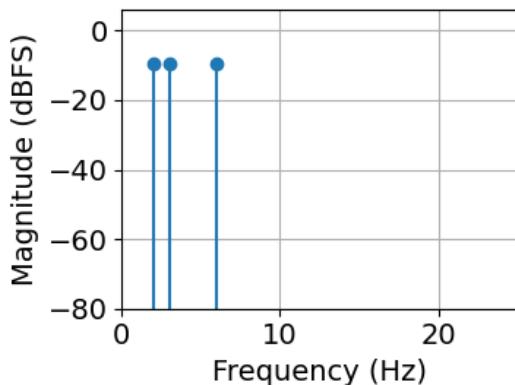
- ✓ Punto a 2 Hz, tono puro 2Hz
- ✓ Punto a 2 Hz + punto a 6 Hz, tono complesso

```
w1 = np.cos(2 * np.pi * 2 * t)
w2 = np.cos(2 * np.pi * 6 * t + np.pi/2)
w = (w1 + w2)/2
```



# Rappr. nel dominio della frequenza

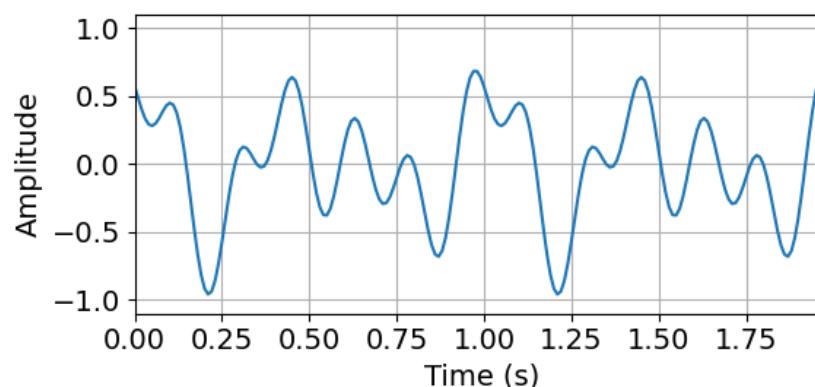
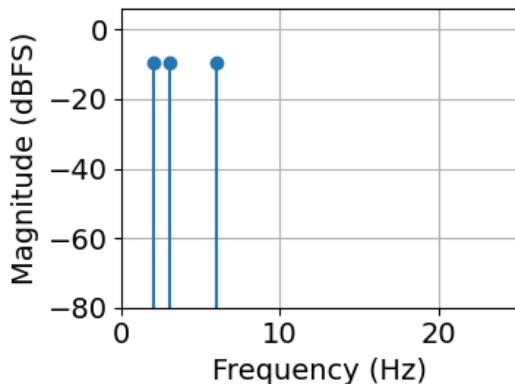
- Spettro (di Ampiezza o di Potenza)
  - ✓ Ogni punto è l'intensità (di una componente) del segnale (tono) ad una certa frequenza
- Domanda: cosa accade se aggiungo un "punto" a 3 Hz?



```
# Calcolate voi e plottate con plt_vs_time()  
  
w1 = np.cos(2 * np.pi * 2 * t)  
w2 = np.cos(2 * np.pi * 6 * t + np.pi/2)  
w3 = np.cos(2 * np.pi * 3 * t - np.pi/4)  
w = (w1 + w2 + w3)/3
```

# Rappr. nel dominio della frequenza

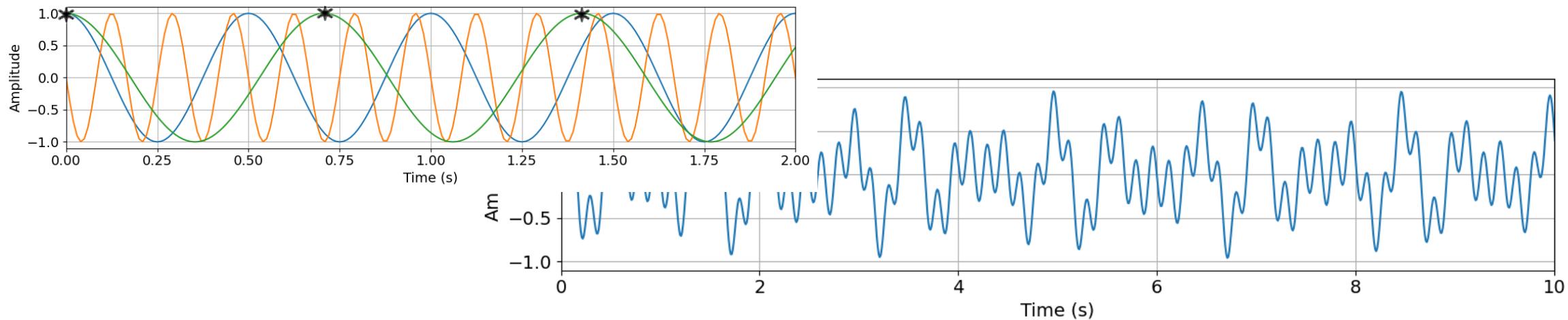
- Spettro (di Ampiezza o di Potenza)
  - ✓ Ogni punto è l'intensità (di una componente) del segnale (tono) ad una certa frequenza
- Domanda: cosa accade se aggiungo un "punto" a 3 Hz?



Quale è il periodo?  
Perchè?

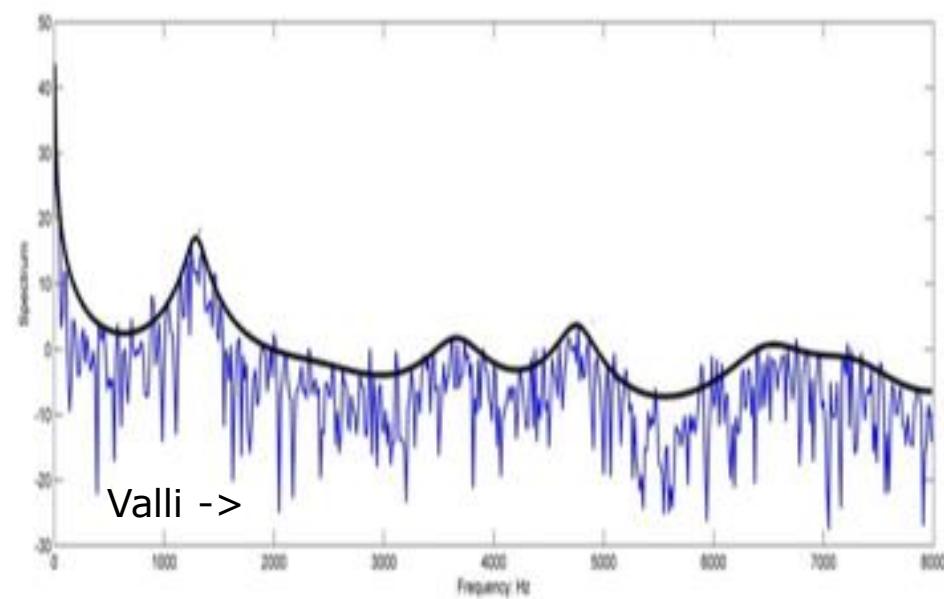
# Suono complesso a-periodico

- Componendo tre toni a  $2, 6, \sqrt{2}$  Hz non sono multipli interi di una frequenza fondamentale F0
- Non c'è quindi una chiara periodicità e il nostro orecchio non identifica una chiara altezza (pitch) del suono
  - ✓ N.B. se la componente a  $\sqrt{2}$  è molto bassa allora "non vale" perchè non è influente ☺



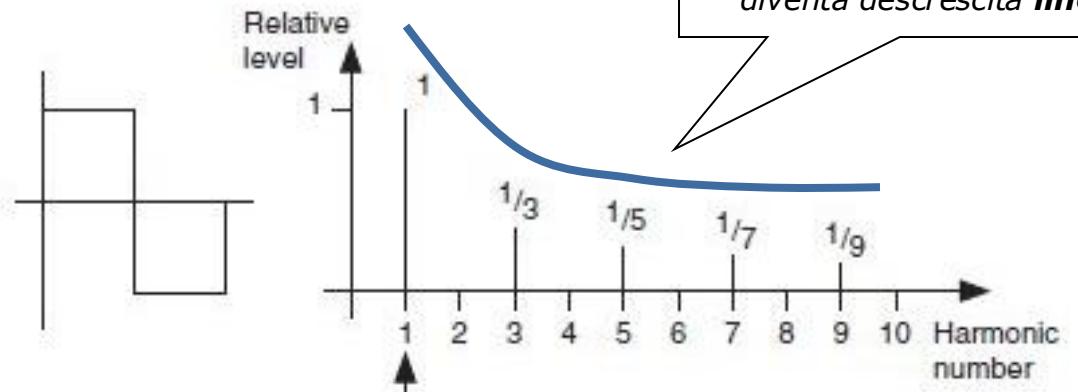
# Inviluppo spettrale

- L'inviluppo spettrale, praticamente il *profilo* dello spettro, è una funzione continua  $E(f)$  che descrive l'andamento "mediato" dell'ampiezza spettrale in modo da catturare la distribuzione dell'energia alle varie frequenze
- Si ignorano i dettagli "fini" delle singole componenti spettrali
- Si tende a modellare (seguire) i picchi e non le valli

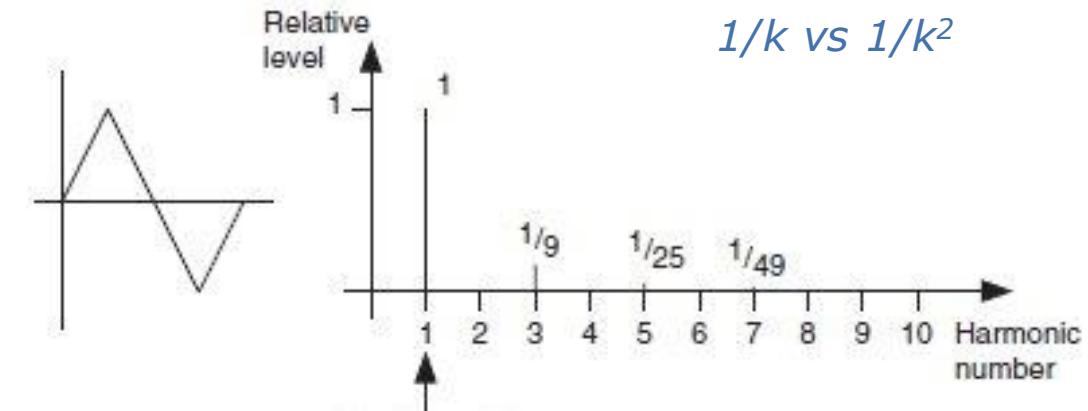


# Inviluppo spettrale - esempi

- Alcuni esempi dalla "teoria"
- Onda quadra
  - ✓ Contiene solo le armoniche dispari (1, 3, 5, ..)
  - ✓ L'ampiezza delle armoniche è definita come  $\sim 1/k$
- Onda triangolare
  - ✓ Contiene solo le armoniche dispari (1, 3, 5, ..)
  - ✓ L'ampiezza delle armoniche è definita come  $\sim 1/k^2$



In scala logaritmica l'ampiezza si dimezza ad ogni raddoppio della frequenza (-6 dB), diventa descrescita **lineare**



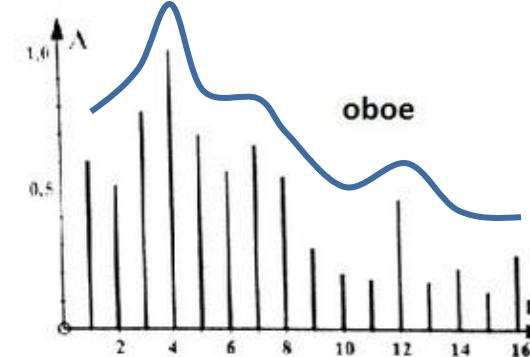
Il "profilo" è diverso  
 $1/k$  vs  $1/k^2$

# Inviluppo spettrale - esempi

- Alcuni esempi dagli strumenti musicali

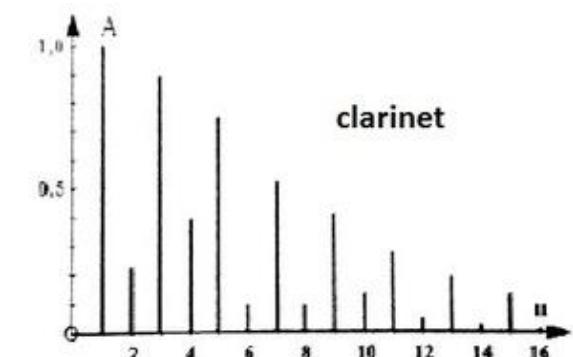
- Oboe

- ✓ Enfasi, maggiore ampiezza, delle armoniche "alte" (suono "nasale")



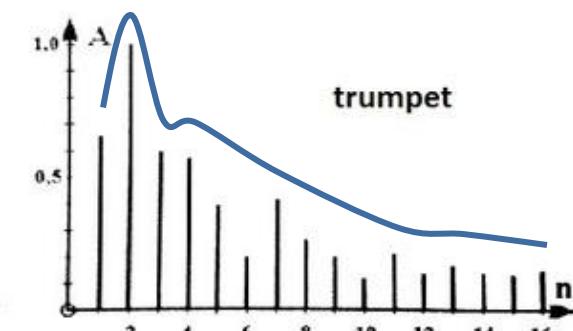
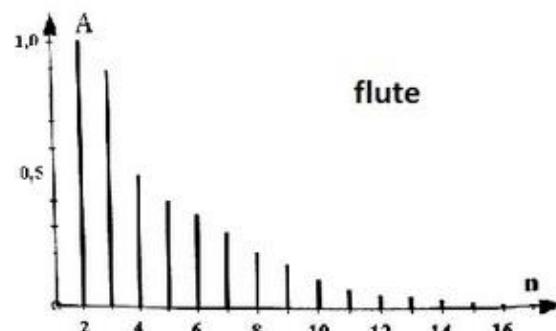
- Clarinetto

- ✓ Armoniche dispari molto più pronunciate (tono "sordo")



- Flauto

- ✓ Armoniche con decrescita costante



**N.B. Sia nell'oboe sia nella tromba  
L'ARMONICA PIU' FORTE NON E' LA PRIMA**

# Caratteristiche dello spettro

---

- Ricapitolando, lo spettro (il modulo) è caratterizzato da
  - ✓ Presenza (o mancanza) di un periodo  $T$  (e corrispondente  $F_0=1/T$ )
  - ✓ Se periodico può essere composto da una o più componenti in f.
    - Armoniche, se sono multiple intere di  $F_0$ 
      - Ed è quindi percepibile una altezza / pitch  $F_0$
    - Parziali, se non vale quanto sopra
  - ✓ Profilo o inviluppo dello spettro
    - Dato dalla relazione di ampiezza delle componenti in frequenza
- Aggiungere: evoluzione temporale per definire il TIMBRO

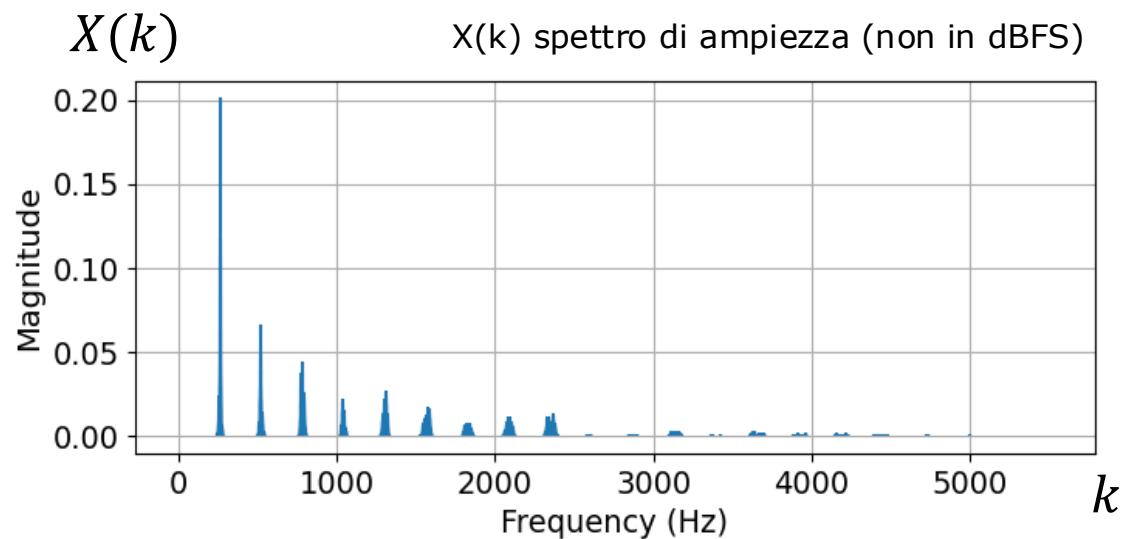
# ANALISI SPETTRALE

---

- Suoni (toni) puri
  - ✓ Il problema dell'aliasing
  - ✓ Ritardo di fase
- Caratteristiche e analisi dello spettro di un segnale
  - ✓ Suoni periodici: altezza/pitch e armoniche
  - ✓ Inviluppo spettrale
- **Descrittori (features) spettrali**

# Features spettrali

- A partire dal contenuto in frequenza è possibile estrarre dei descrittori / features di più "alto livello" che riassumono in un singolo valore (o pochi) alcuni caratteri dell'informazione spettrale
- Distribuzione dell'energia "lungo" lo spettro
  - ✓ Spectral centroid, variance
  - ✓ Spectral slope, roll-off,
  - ✓ High band energy ratio
- Armonicità
  - ✓ Harmonic-to-Noise Ratio
  - ✓ Spectral flatness



# Calcolo dello spettro (di ampiezza)

- Trasformata di Fourier (di un segnale reale):  $X(k)$

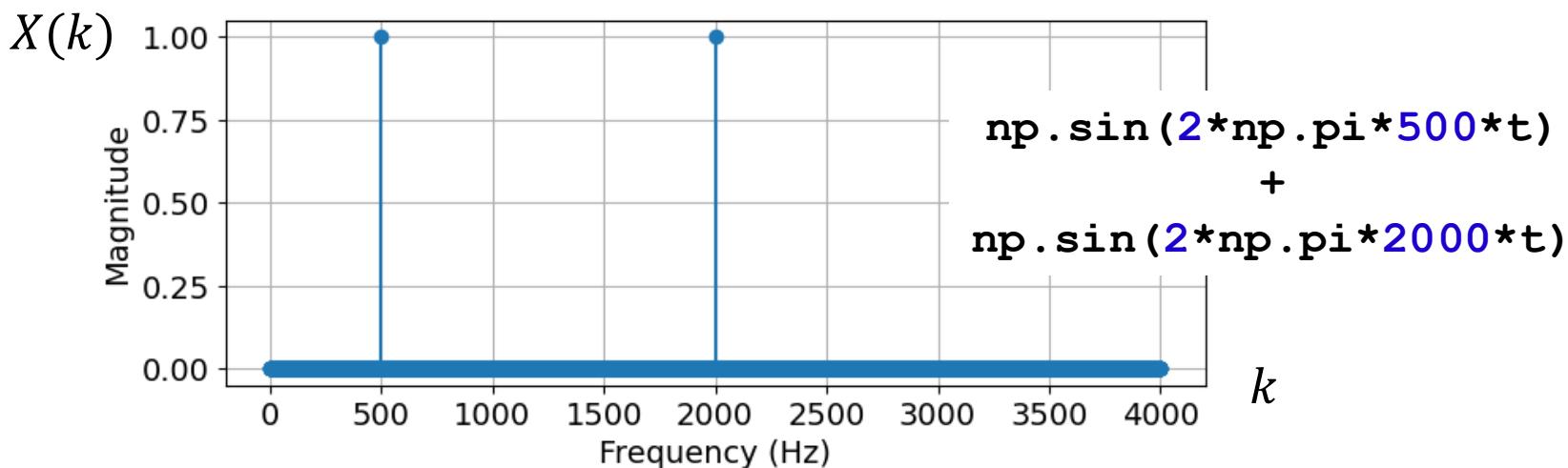
✓ `spectrum = np.abs(np.fft.rfft(y)) / (L/2)`

$X(k)$  lineare, non in dB  
Sinusoide con ampiezza A  
ha valore A.

- Valori delle frequenze su cui è calcolata:  $k$

✓ `freqs = np.fft.rfftfreq(L, 1/sr)`

( $L/2$ ) fattore  
di normalizzazione

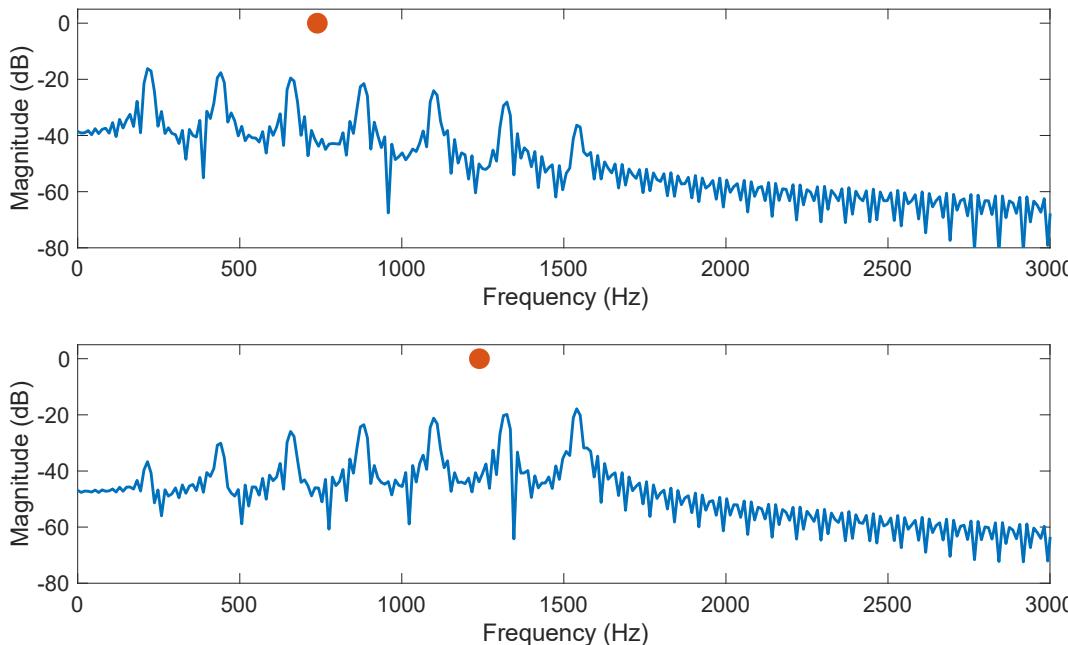


# Centroide spettrale

## ■ Centroide: centro di massa dello spettro

- ✓ Calcolato come la media pesata dell'energia delle frequenze presenti nel segnale

$$C = \frac{\sum_{k=1}^N (k \cdot X(k))}{\sum_{k=1}^N X(k)}$$



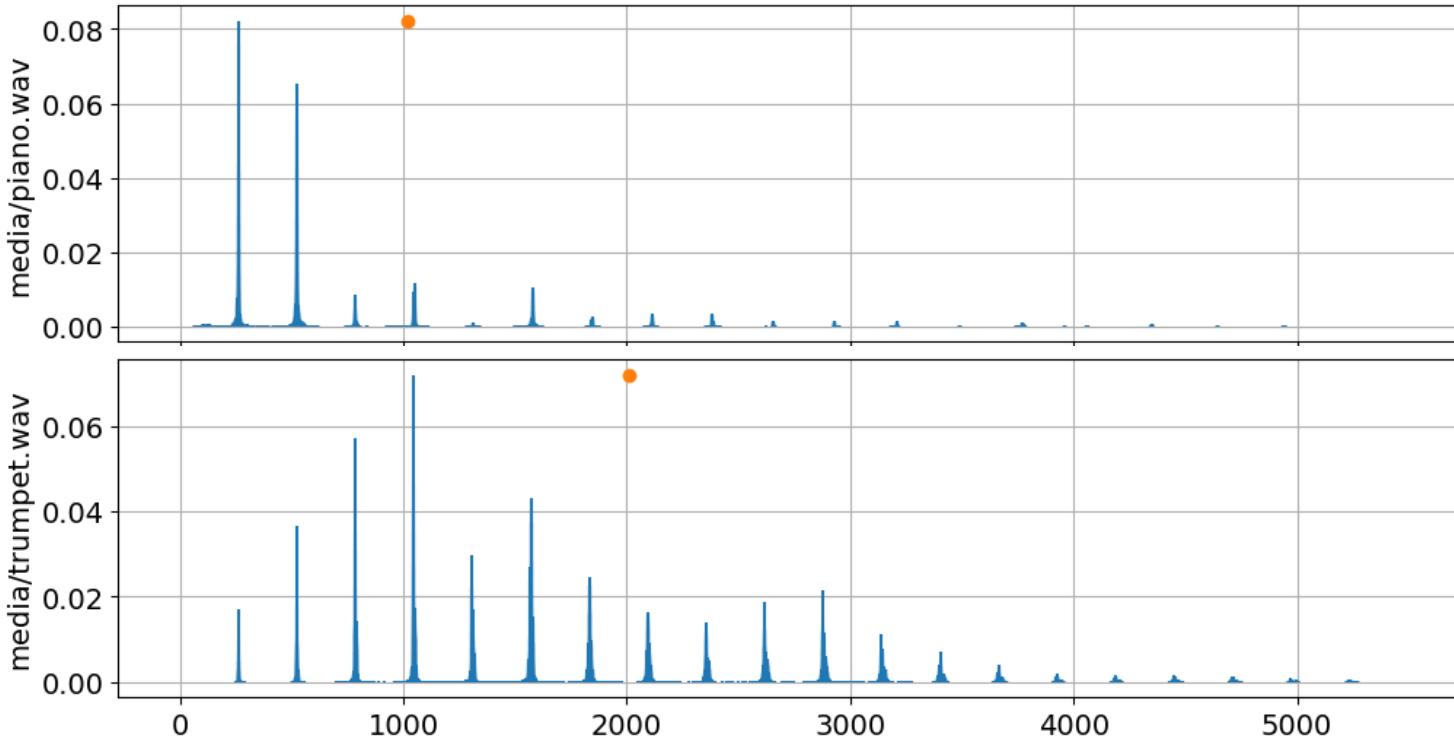
*Un buon preditore della "brillantezza" di un suono (armoniche superiori con elevata energia)*

# Centroide spettrale

$$C = \frac{\sum_{k=1}^N (k \cdot X(k))}{\sum_{k=1}^N X(k)}$$

- Centroide: centro di massa dello spettro

```
centroid = np.sum(freqs * spectrum) / np.sum(spectrum)
```



# Ascolto: centroide spettrale

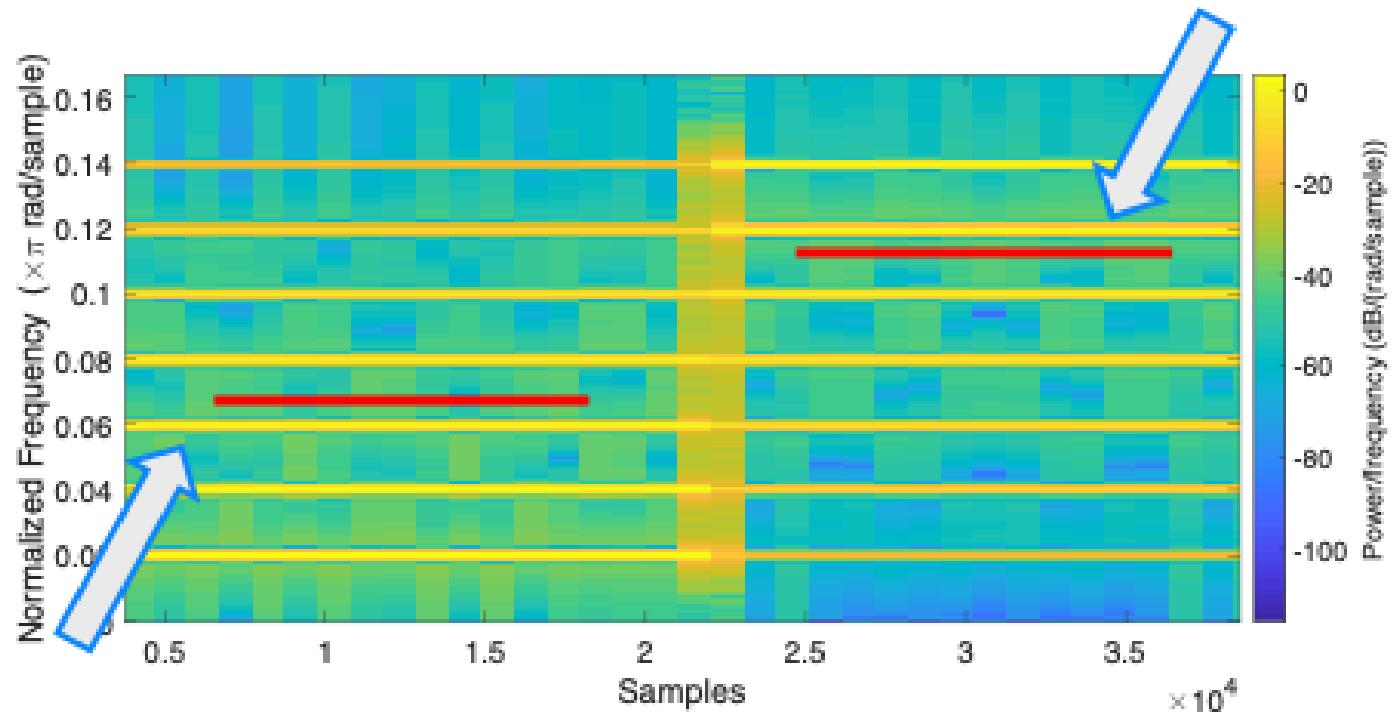


## ■ Esempio: 220LowHigh.wav

- ✓ Coppia di toni complessi con 7 componenti armoniche, stessa frequenza fondamentale 220 Hz

*Il secondo suono ha la maggior parte dell'**energia** nelle **armoniche alte** (**alto** centroide spettrale, suono **brillante**)*

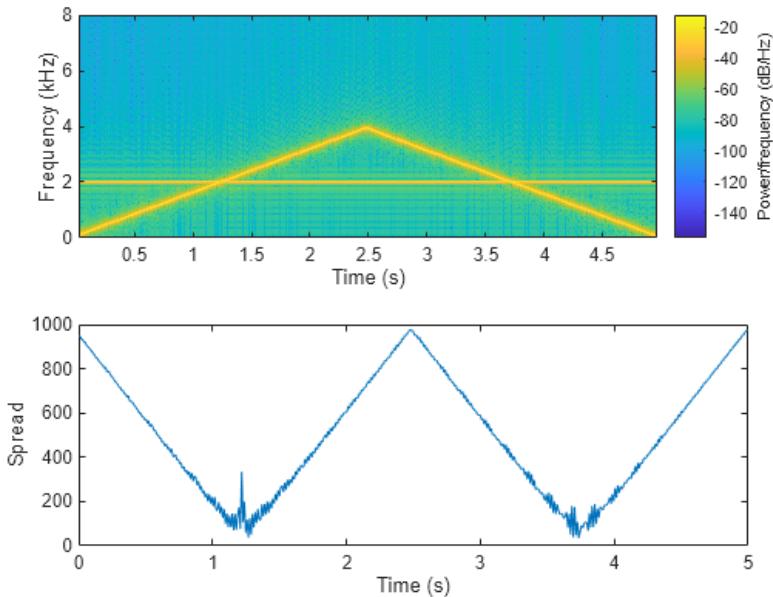
*Il primo suono ha la maggior parte dell'**energia** nelle **armoniche basse** (**basso** centroide spettrale, suono **cupo**)*



# Varianza spettrale (spectral spread)

- Varianza media dei coefficienti spettrali rispetto al centroide (secondo momento centrale)
- Comunemente associato al concetto di banda del segnale
- Esempi
  - ✓ Segnali noise-like hanno solitamente un a varianza spettrale elevata
  - ✓ Toni semplici o segnali tone-like con picchi concentrati hanno bassa varianza spettrale

$$S = \sqrt{\frac{\sum_{k=1}^N [(k - C)^2 \cdot X(k)]}{\sum_{k=1}^N [X(k)]}}$$

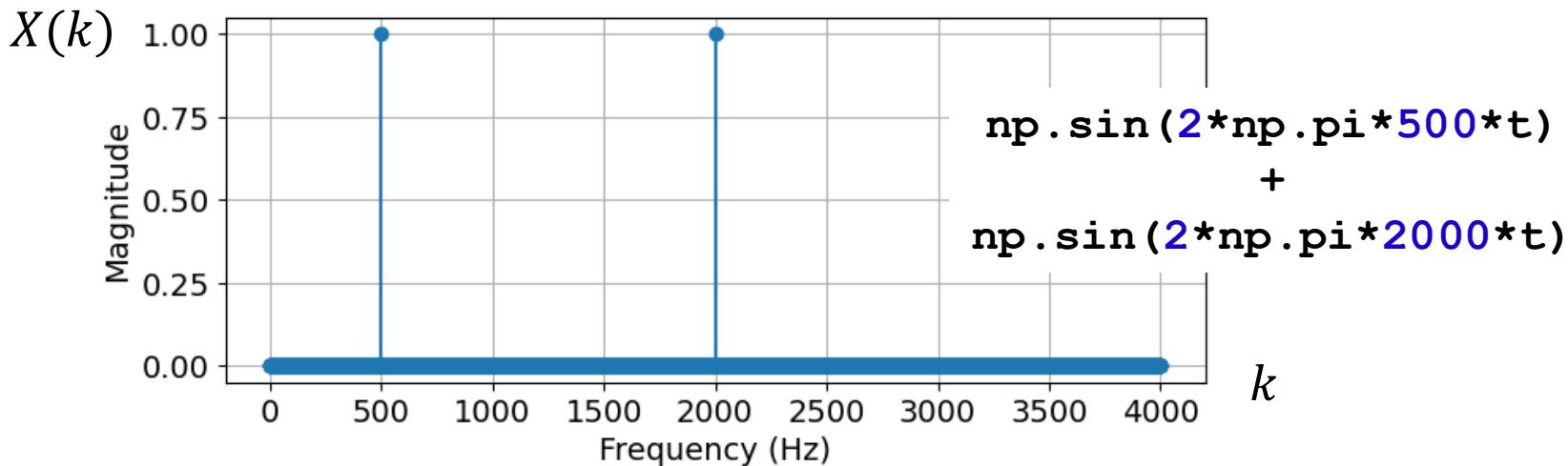


# Esercizio

## ■ Centroide e varianza spettrale

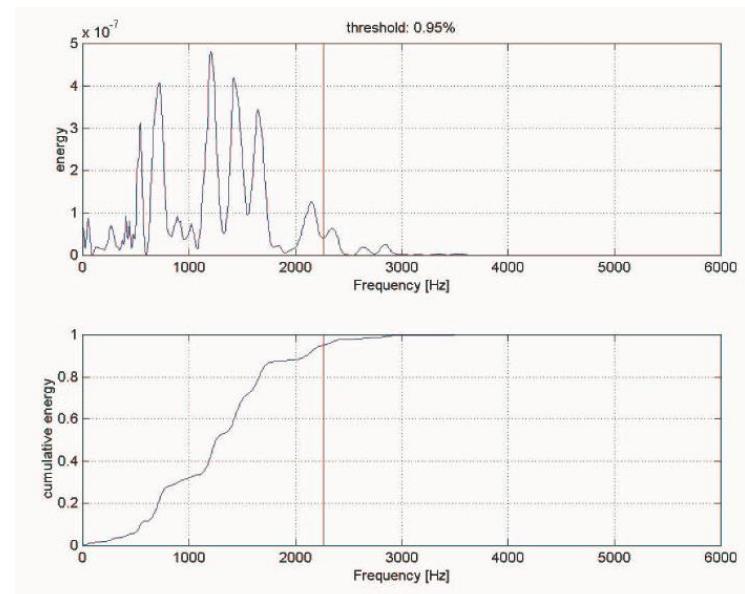
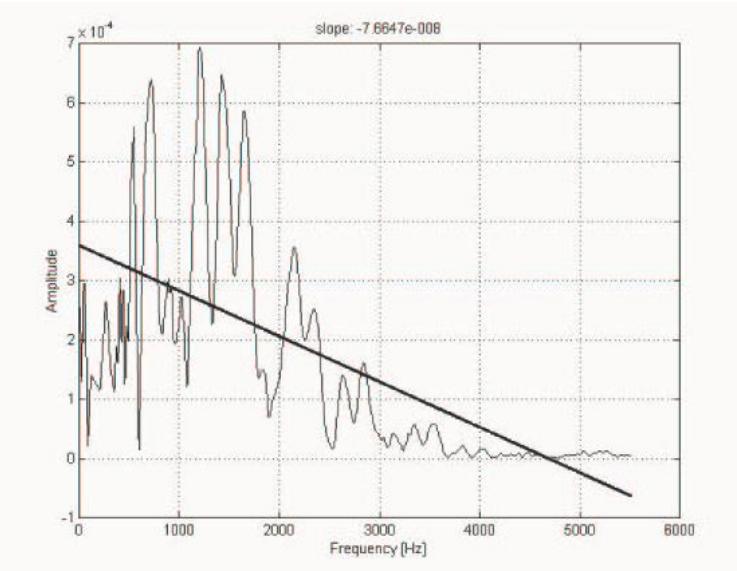
```
centroid = np.sum(freqs * spectrum) / np.sum(spectrum)
spread = np.sqrt(np.sum(((freqs - centroid) ** 2) * spectrum) / np.sum(spectrum))
print(f"Spectral centroid: {centroid:.1f} Hz")
print(f"Spectral spread: {spread:.1f} Hz")
```

Spectral centroid: 1250.0 Hz  
Spectral spread: 750.0 Hz



# Spectral slope, roll-off

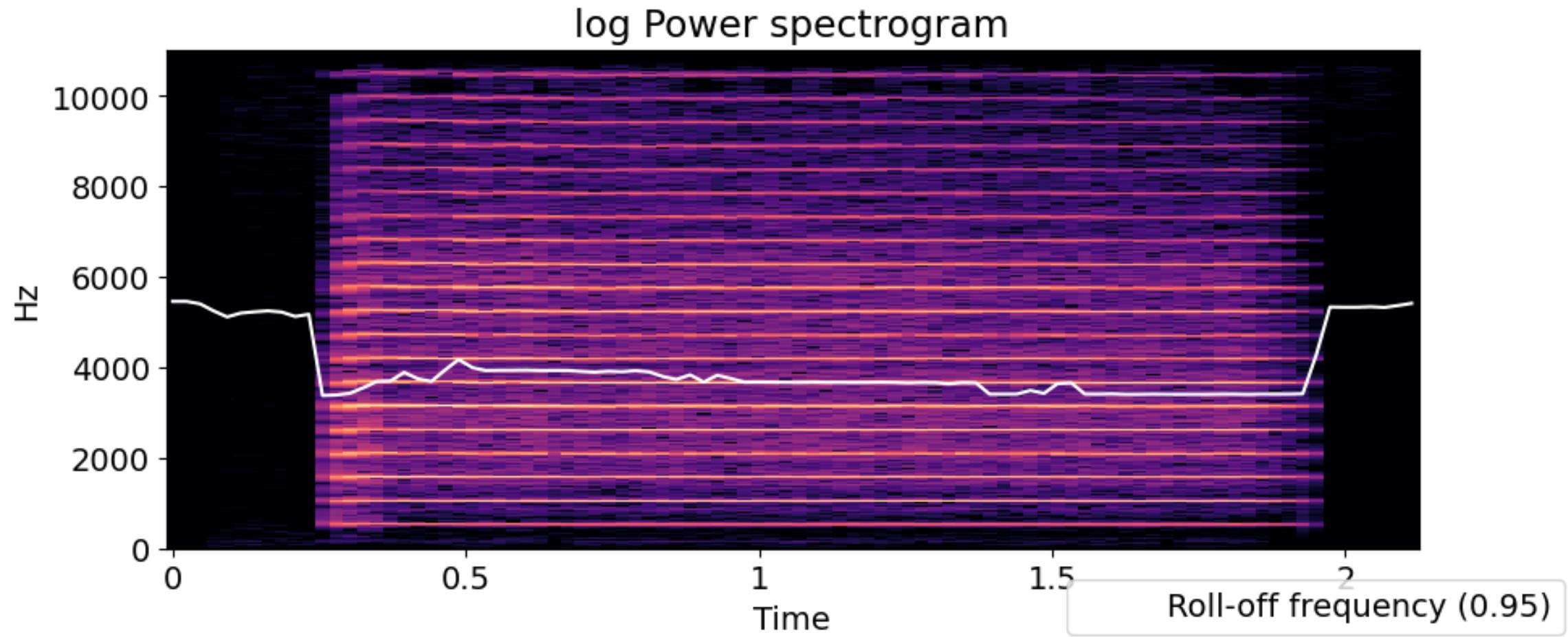
- Stima della riduzione dell'energia "muovendosi" verso le alte frequenze (cfr. [Lerch])
- Spectral slope / decrease
- Spectral roll-off
  - ✓ Frequenza al di sotto la quale è contenuto il 95% dell'energia



Reference: A large set of audio features for sound description (similarity and classification) in the CUIDADO project, 2004

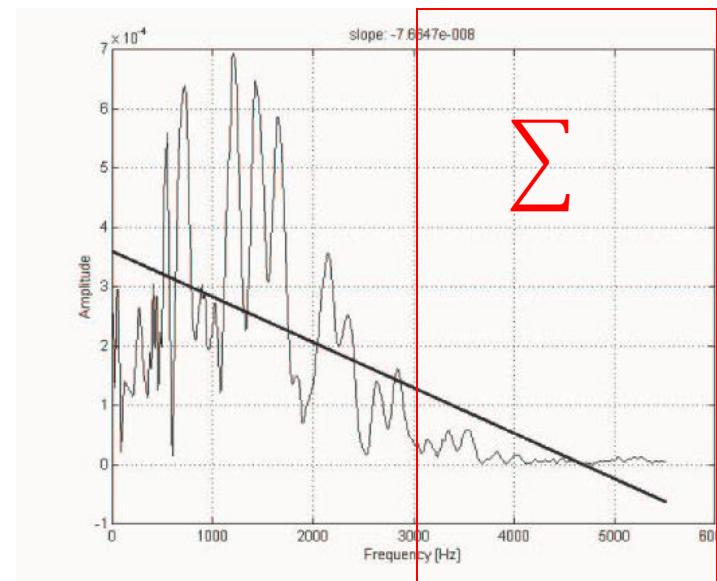
# Example: librosa spectral rolloff

- trumpet



# High Band Energy Ratio (HBER)

- Stima della riduzione dell'energia "muovendosi" verso le alte frequenze
- High Band Energy Ratio (HBER)
  - ✓ Una approssimazione della spectral slope/decrease
  - ✓ Rapporto tra l'energia del segnale sopra una soglia (e.g., 3 kHz) e l'energia totale del segnale



# Spectral flatness

---

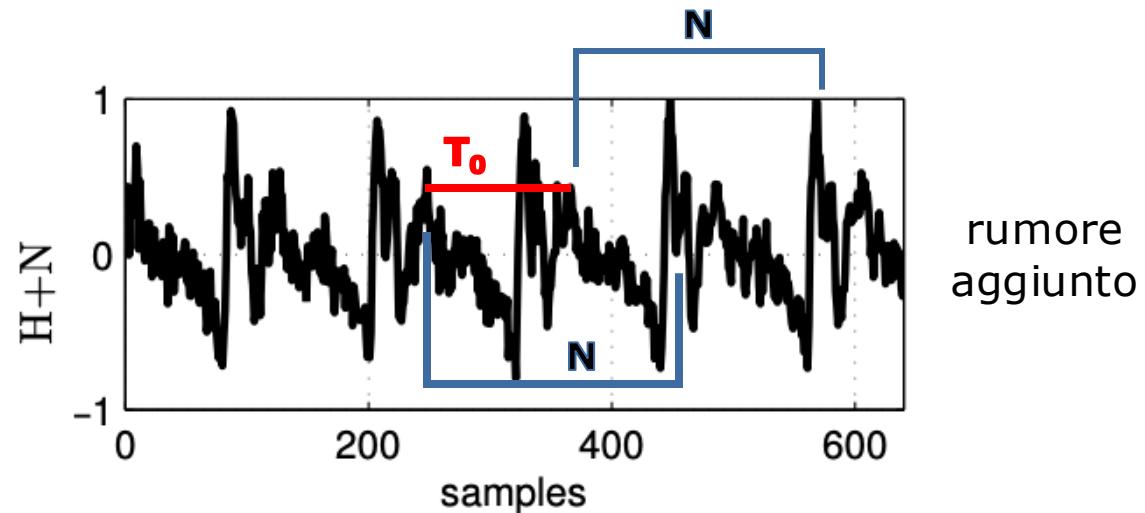
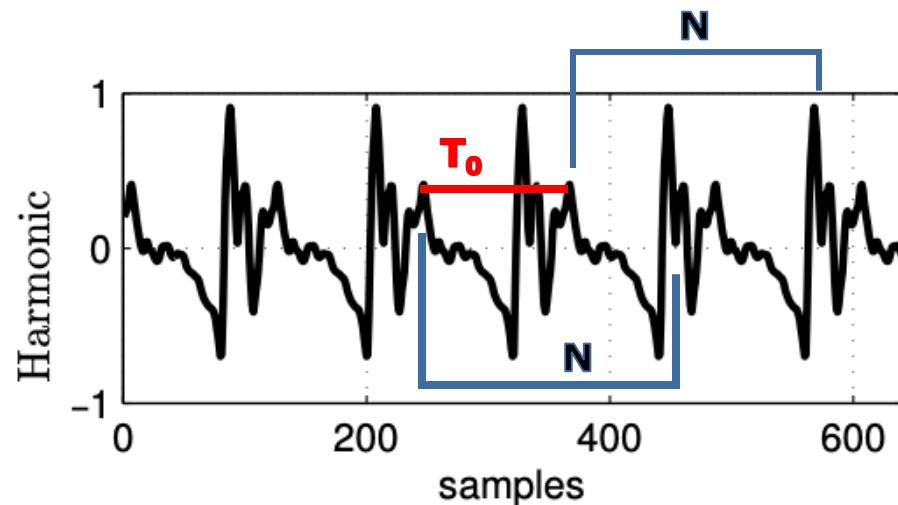
- Misura quanto un suono è di tipo rumore (1) rispetto a tono puro (0)
- Rapporto tra la media geometrica e la media aritmetica dei coefficienti spettrali
- La media geometrica è molto sensibile a valori piccoli
  - ✓ Se lo spettro è tendenzialmente non piatto e ci sono sia valori grandi sia valori piccoli la media geometrica tende ad assumere un valore basso
    - Quindi la spectral flatness si avvicina allo 0
  - ✓ E' invece vicino a 1 per lo spettro piatto, rumore bianco

$$SF = \frac{\sqrt[N]{\prod_{k=1}^N X(k)}}{1/N \sum_{k=1}^N X(k)}$$

# Harmonic to noise ratio

- Vengono confrontati tra di loro, tramite moltiplicazione campione per campione, frame del segnale di  $N$  campioni a distanza  $T_0$ 
  - ✓  $T_0$  è stimato come periodo "presunto" del segnale
  - ✓ Più i frame sono simili più il valore  $R_{xx}[T_0]$  è elevato

$$R_{xx}[T_0] = \frac{1}{N} \sum_{k=0}^{N-1} x[k]x[k - T_0]$$

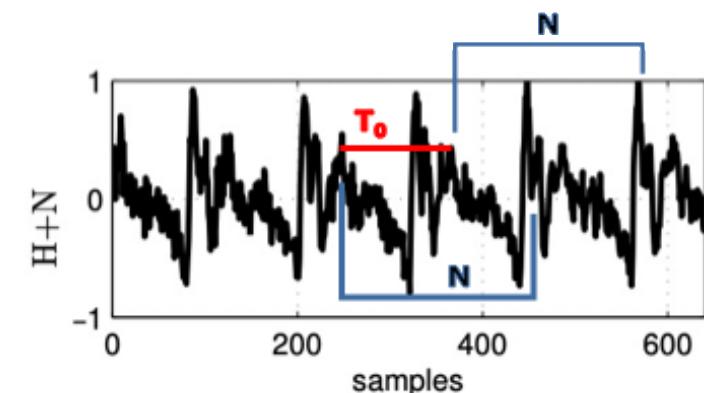
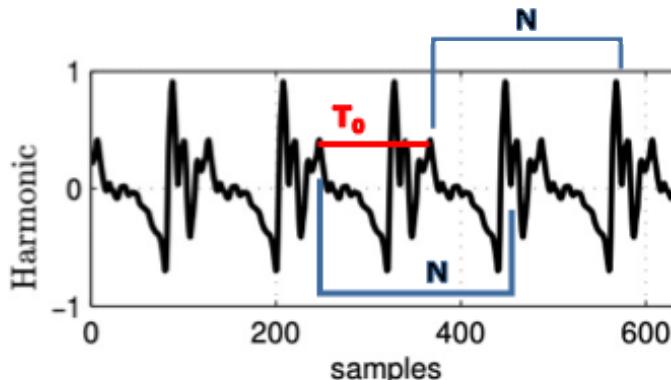


# Harmonic to noise ratio

$$R_{xx}[T_0] = \frac{1}{N} \sum_{k=0}^{N-1} x[k]x[k - T_0]$$

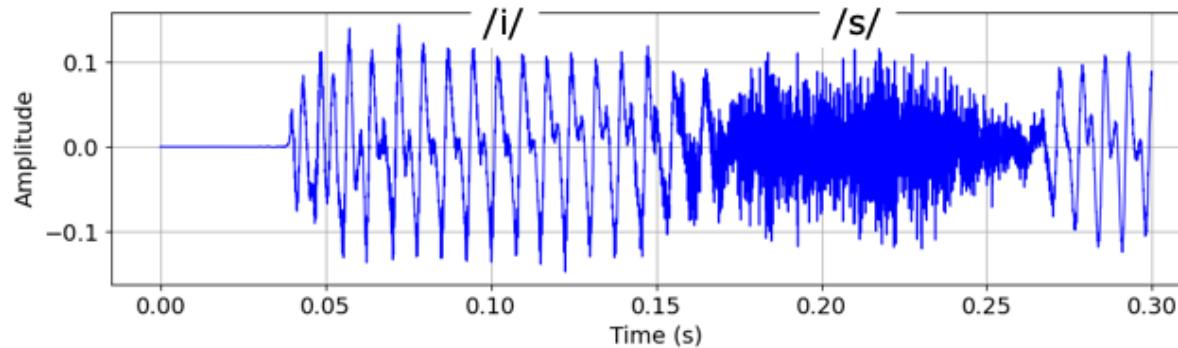
- Per calcolare l'HNR viene calcolato anche l'R<sub>xx</sub> "ideale" cioè quello del segnale con se stesso R<sub>xx</sub>[0] (che è la sua energia)
- L'HNR è dato quindi come il seguente rapporto calcolato in dB
  - ✓ Più il valore è elevato più il segnale è armonico, e viceversa

$$HNR = \frac{R_{xx}[T_0]}{R_{xx}[0] - R_{xx}[T_0]}$$



# Zero Crossing Rate (non spettrale)

- Lo zero crossing **rate** (ZCR) cioè il rapporto tra il numero di attraversamenti dello zero e il numero di campioni analizzati, è una misura della *rumorosità* di un segnale
  - ✓ Un segnale rumoroso, con la presenza di molta energia alle alte frequenze, attraversa frequentemente lo zero
  - ✓ Un segnale a bassa frequenza, attraversa lo zero raramente
    - Esempio: più una sinusoide ha frequenza elevata più frequentemente attraversa lo zero (due volte per periodo)



File: potter.wav /ils/i/

# Zero Crossing Rate (non spettrale)

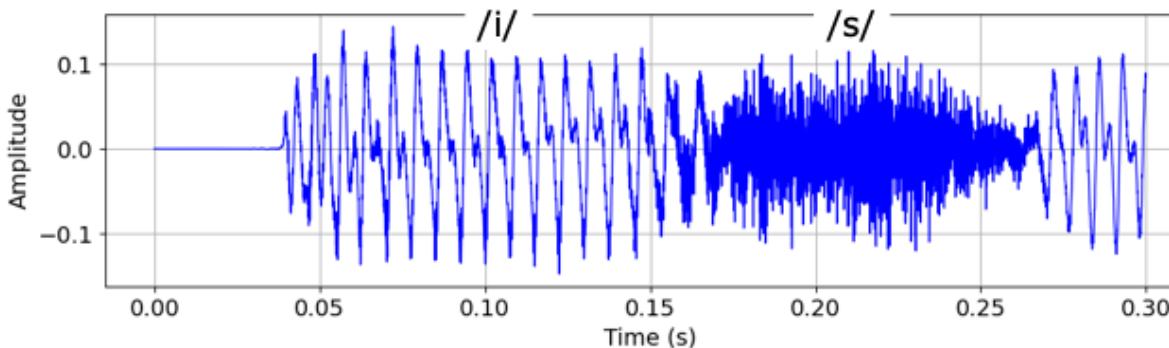
- Lo zero crossing **rate** (ZCR) cioè il rapporto tra il numero di attraversamenti dello zero e il numero di campioni analizzati, è una misura della *rumorosità* di un segnale

```
def zcr(x):  
    # x [1, -1, -0.5, 1, -1]  
    # sign [ 1. -1. -1. 1. -1.]  
    # diff [-2. 0. 2. -2.]  
    # count_nonzero 3  
    # zcr 0.6 # i.e. 3/5  
    return np.count_nonzero(np.diff(np.sign(x))) / len(x)
```

Implementazione basata sul conteggio della variazione del SEGNO.

La funzione sign vale (-1 negativo, +1 positivo).

Se tra una capione e il successivo il segno non cambia, la differenza è 0.

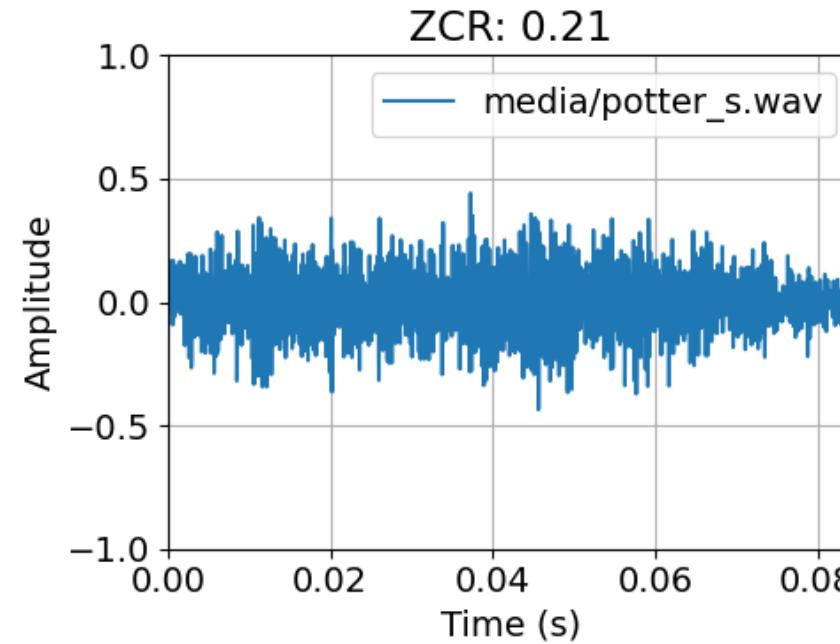
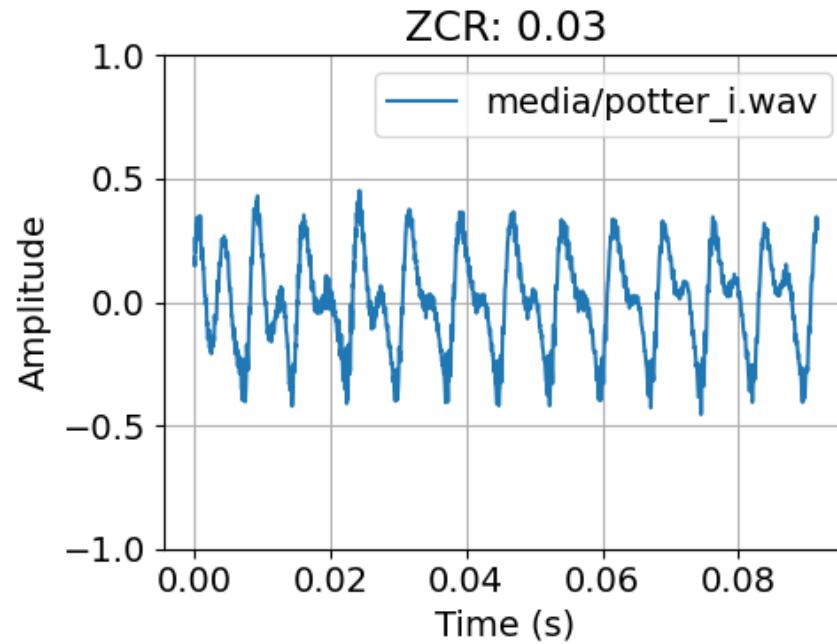


File: potter.wav /ils/i/

# Zero Crossing Rate

(non spettrale)

- Lo zero crossing **rate** (ZCR) cioè il rapporto tra il numero di attraversamenti dello zero e il numero di campioni analizzati, è una misura della *rumorosità* di un segnale



File: potter.wav /ilsi/

---

## SEGMENTAZIONE

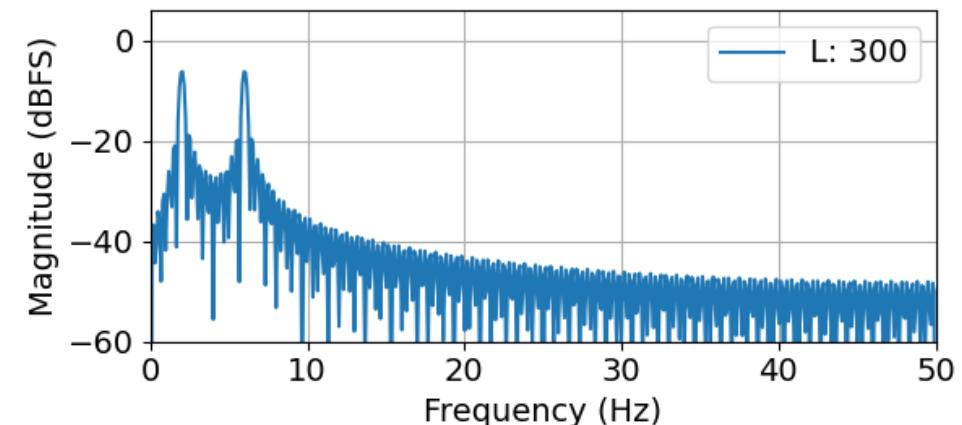
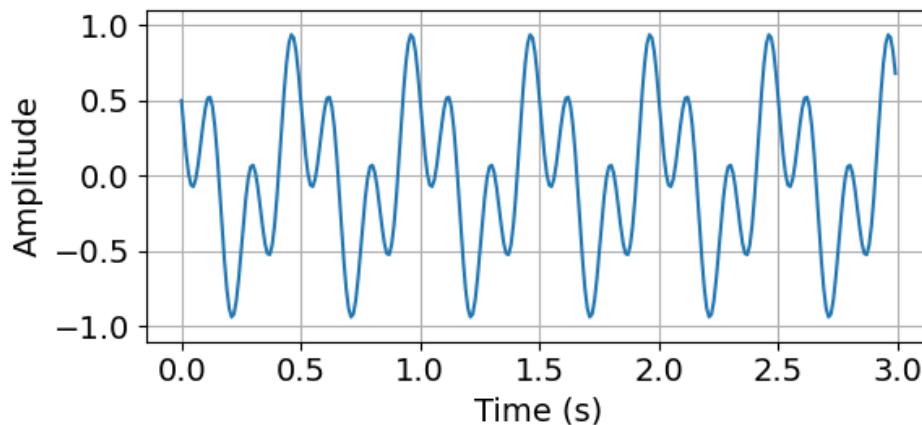
# SEGMENTAZIONE

---

- **Importanza dell'intervallo temporale**
  - ✓ La DFT è una MEDIA su un intervallo temporale
- Tempo varianza e stazionarietà locale
- ATTIVITA' – Analisi del segnale vocale: fonemi
  - ✓ Praat - <https://praat.org/>

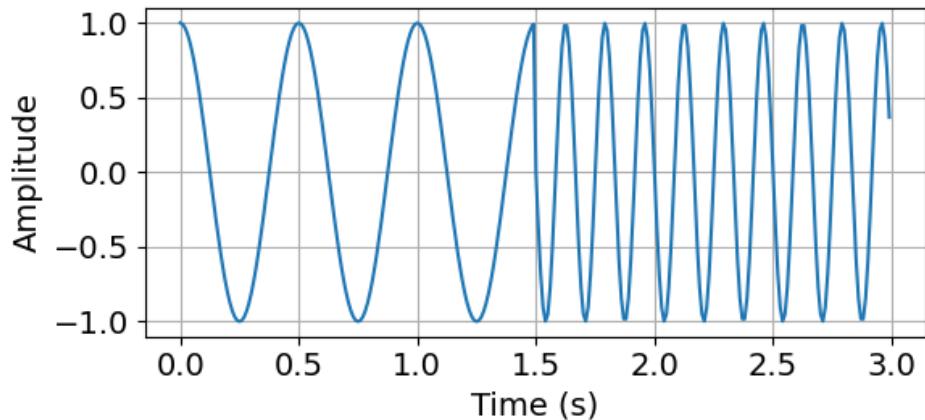
# Spettro di un segnale digitale

- Attenzione: Non si può calcolare lo spettro da un singolo campione! A rigore la T. di Fourier è definita su un segnale di lunghezza infinita
- Posso calcolare lo spettro di un segnale digitale applicando la Trasformata di Fourier (discreta) su L campioni, cioè su un intervallo temporale  $\Delta t$ 
  - ✓ Ottengo una versione "sporca" della trasformata teorica (cfr. approfondimento)



# Importanza dell'intervallo temporale

- La DFT è una "fotografia" del segnale audio su un intervallo temporale: riporta la media delle informazioni in frequenza, ma "nasconde" (nella fase) le informazioni temporali
- ESEMPIO
- ✓ Dato un segnale audio formato per i primi 2s da un tono a 2 Hz e per i successivi 2s da un tono a 6 Hz
  - ✓ Quale sarà la sua DFT?

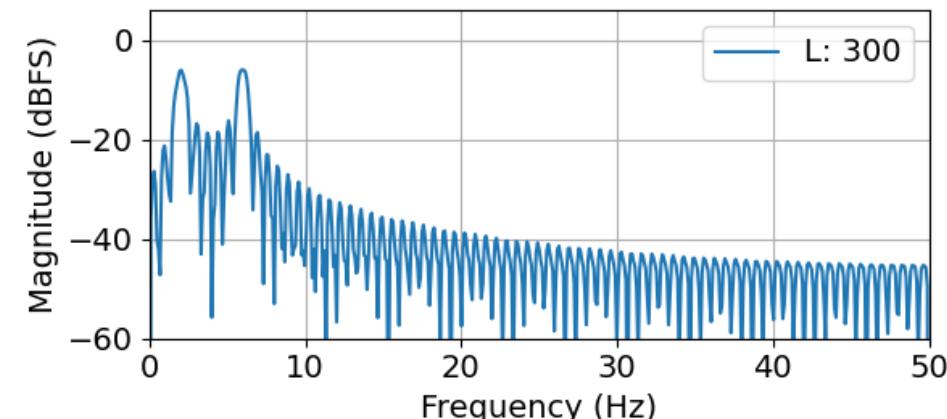
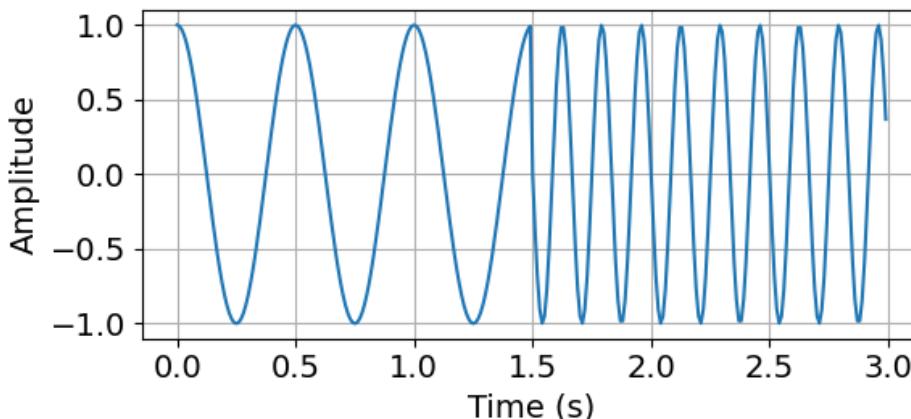


# Importanza dell'intervallo temporale

- La DFT è una "fotografia" del segnale audio su un intervallo temporale: riporta la media delle informazioni in frequenza, ma "nasconde" (nella fase) le informazioni temporali

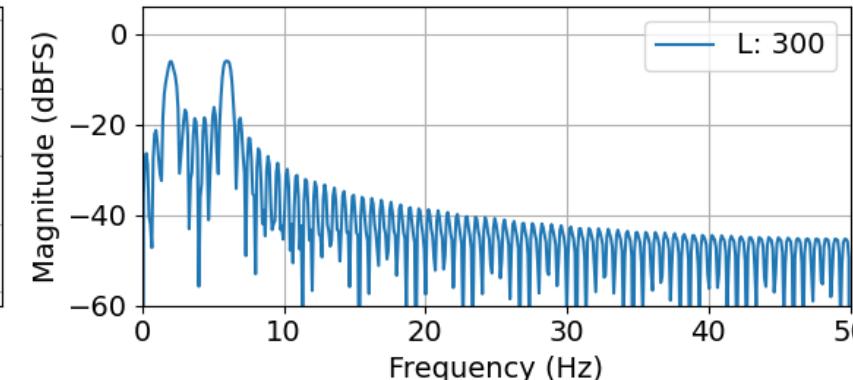
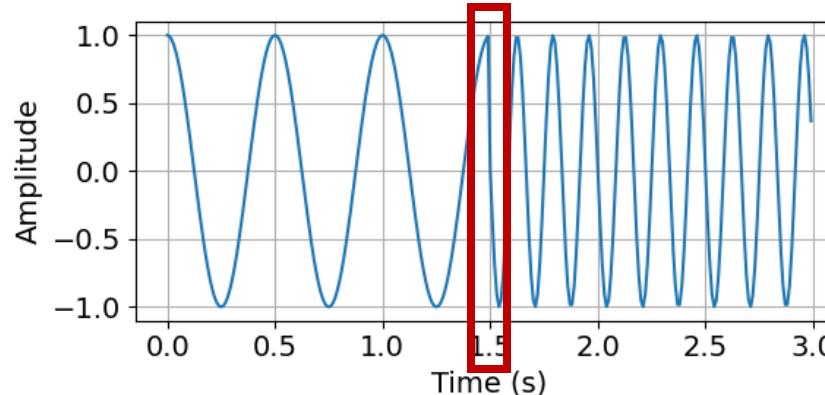
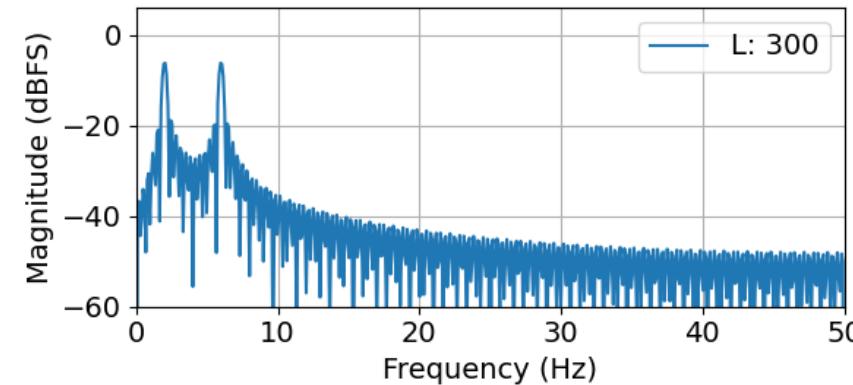
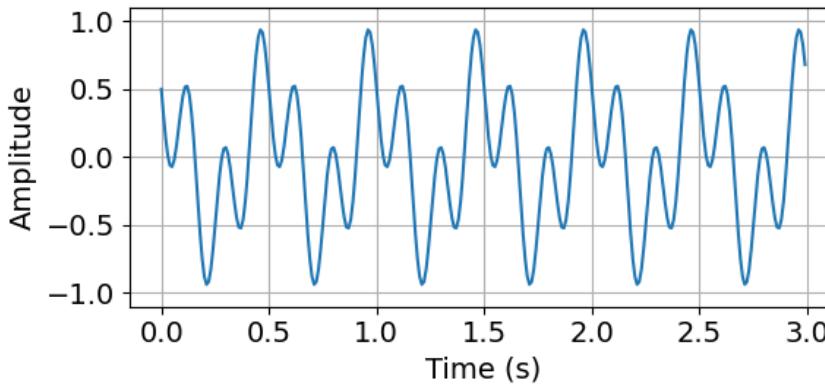
## ESEMPIO

- ✓ Dato un segnale audio formato per i primi 2s da un tono a 2 Hz e per i successivi 2s da un tono a 6 Hz
- ✓ **La DFT dell'intera sequenza riporta entrambe le frequenze perché entrambe sono contenute nella sequenza analizzata**



# Importanza dell'intervallo temporale

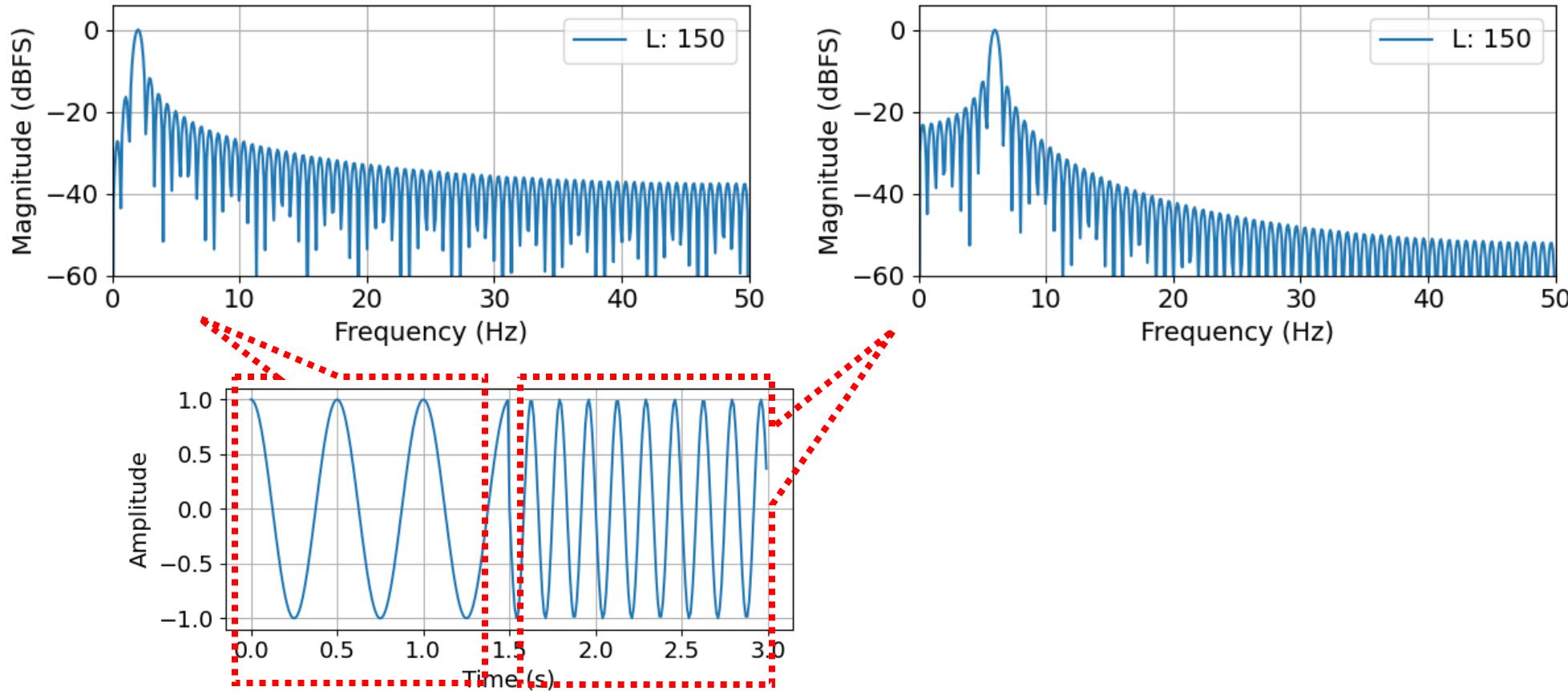
- Lo spettro della concatenazione di due toni a 2 e 6 Hz non è tanto diverso da quello della somma di due toni a 2 e 6 Hz



Più rumoroso  
a causa della  
discontinuità  
nel mezzo

# Importanza dell'intervallo temporale

- Occorre fare attenzione ai "cambiamenti" ed effettuare la trasformata sulle due parti separatamente

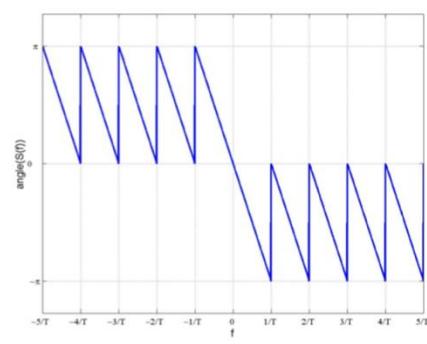
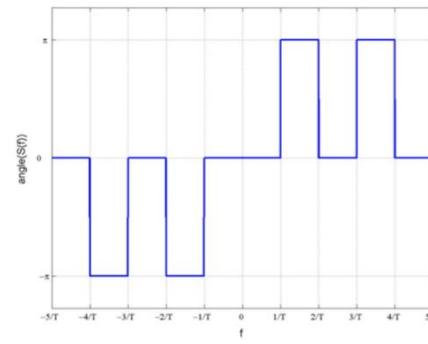
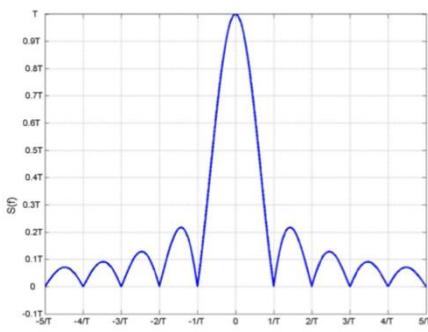
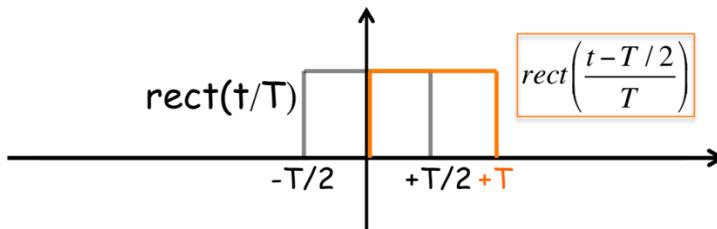


# Time shift property

- A *delay* in the time domain  $x(n - \Delta)$  corresponds to a *linear phase term* in the frequency domain  $e^{j\omega_k \Delta} X(j\omega_k)$

✓ The spectral magnitude is **unaffected** by a linear phase term

$$|e^{j\omega_k \Delta} \cdot X(j\omega_k)| = |X(j\omega_k)|$$



Reference: [https://www.dsprelated.com/freebooks/mdft/Shift\\_Theorem.html](https://www.dsprelated.com/freebooks/mdft/Shift_Theorem.html)

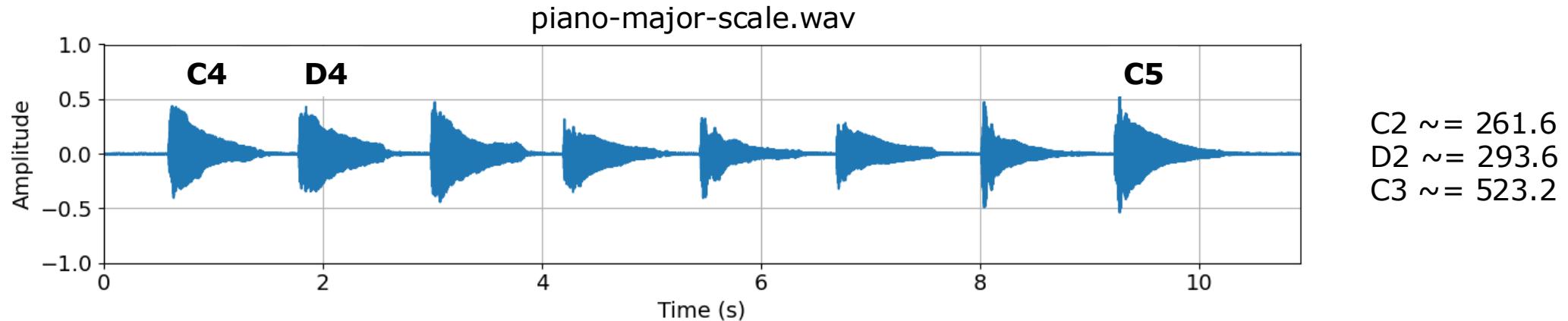
# SEGMENTAZIONE

---

- Importanza dell'intervallo temporale
  - ✓ La DFT è una MEDIA su un intervallo temporale
- **Tempo varianza e stazionarietà locale**
- ATTIVITA' – Analisi del segnale vocale: fonemi
  - ✓ Praat - <https://praat.org/>

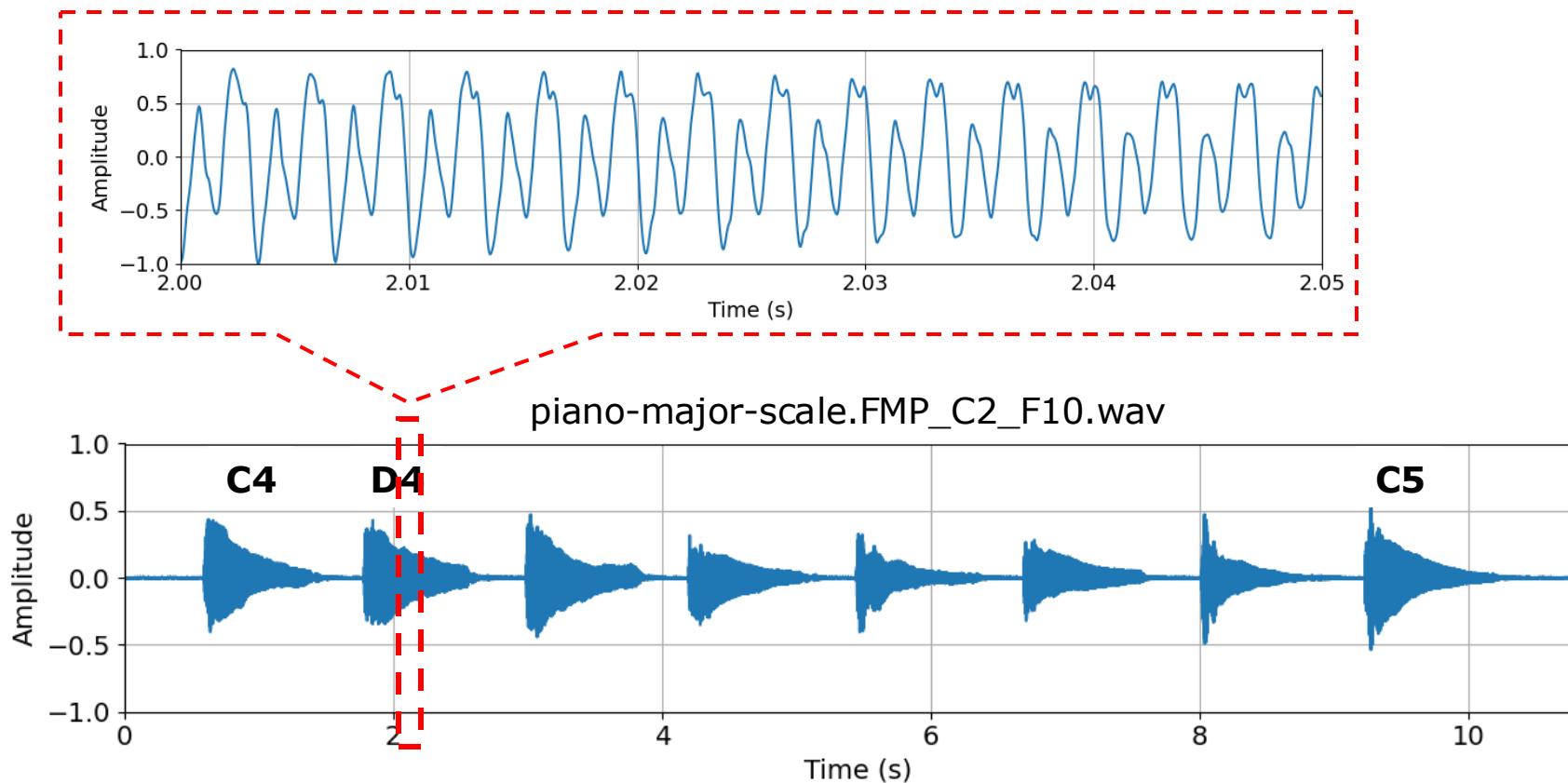
# Il segnale audio è tempo-variante

- Il segnale audio è un segnale "tempo variante" perché le sue proprietà statistiche (momenti di ordine n) cambiano nel tempo
  - ✓ La media è generalmente nulla
  - ✓ Ma ad esempio cambiano nel tempo
    - Varianza (cioè l'energia), il periodo, l'intensità delle componenti in frequenza, ecc.



# (quasi) Stazionarietà locale

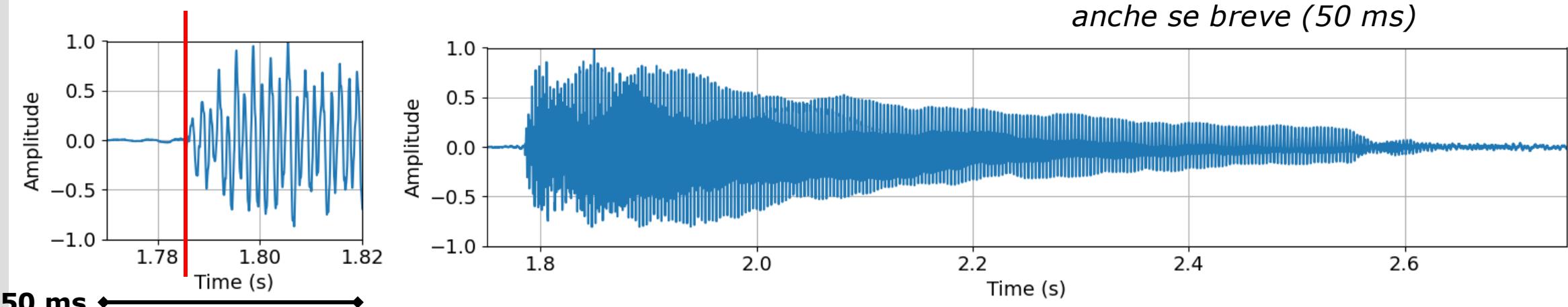
- Il segnale audio è quindi NON stazionario
- Ma può essere considerato localmente quasi-stazionario



$$\begin{aligned} C2 &\approx 261.6 \\ D2 &\approx 293.6 \\ C3 &\approx 523.2 \end{aligned}$$

# Segmentazione

- "Localmente" significa su un segmento di "breve" durata
- Ma occorre fare attenzione a
  - ✓ Cosa si intende per breve durata (cambia con il contesto)
    - Ad esempio qui una nota ha una durata di più o meno un secondo, ma la stazionarietà non vale per tutta la durata
  - ✓ La posizione in cui viene estratto (transienti)
    - Se il segmento è preso "a cavallo" del punto di attacco non è stazionario



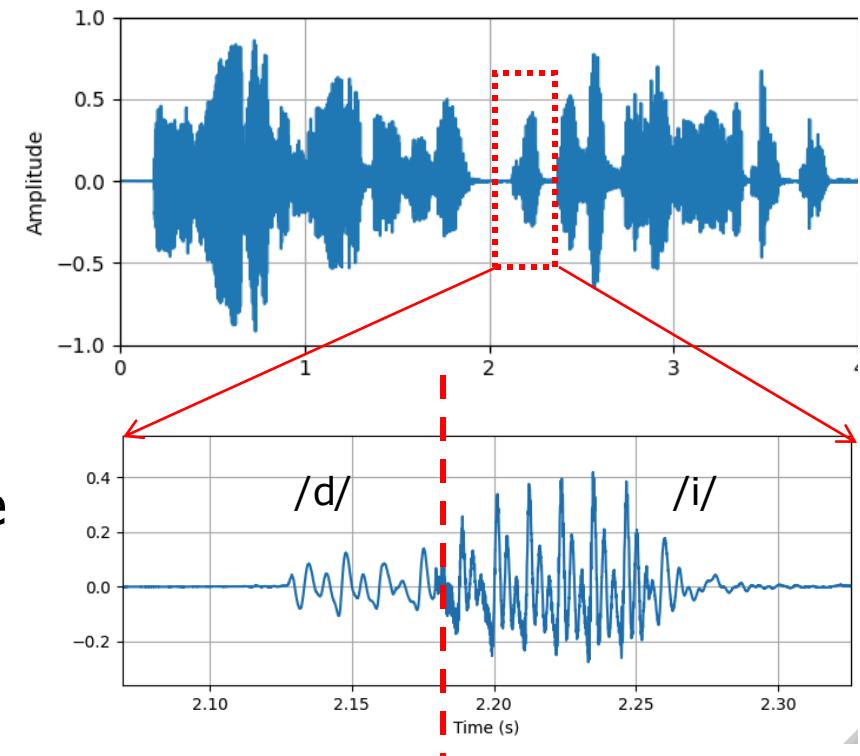
# Stazionarietà e spettro

---

- Nel caso dell'audio per "quasi-stazionarietà" intendiamo più formalmente "wide-sense stationarity (WSS)", che si limita all'invarianza dei primi due momenti
  - ✓ La media è costante
  - ✓ L'autocorrelazione dipende solo dalla differenza temporale e non dal tempo assoluto
- Per un processo WSS la PSD (densità di potenza spettrale) è la trasformata di Fourier dell'autocorrelazione
  - ✓ Non dipendendo l'autocorrelazione dal tempo assoluto (punto in cui viene calcolata) anche lo spettro è uguale in qualsiasi punto venga calcolato, i.e., costante

# Esempio: segnale vocale

- Il segnale vocale può essere decomposto in **SEGMENTI**
  - ✓ Nel caso del parlato sono i *fonemi* le "note" del nostro apparato vocale
- Ciascun SEGMENTO ha delle caratteristiche proprie in termini di
  - ✓ Statistica dei valori dei campioni
  - ✓ Contenuto in frequenza
- Esempio
  - ✓ Non è stazionario se considero /d/ e /i/ insieme
  - ✓ Ma è (quasi) valido se considero solo /d/ o solo /i/



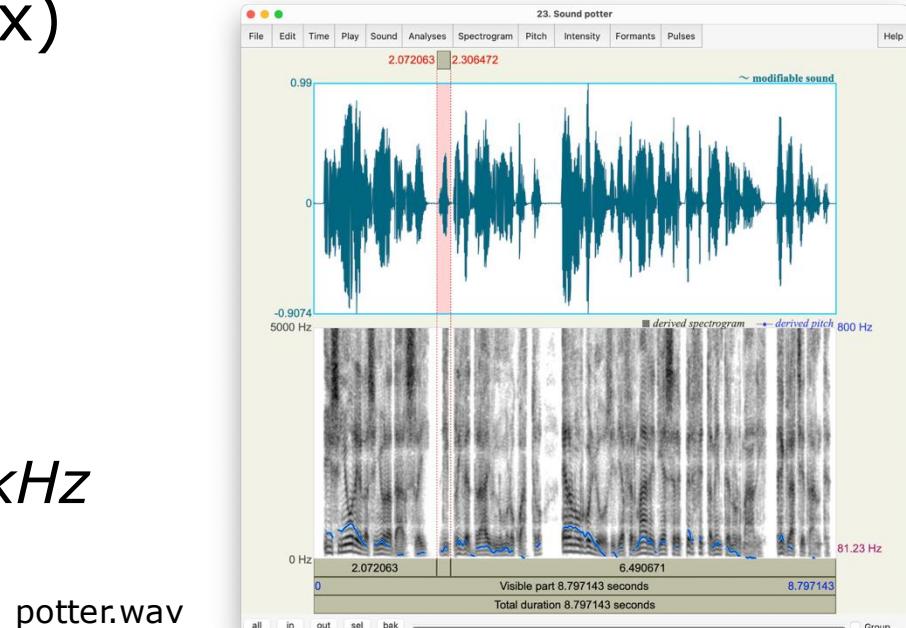
# SEGMENTAZIONE

---

- Importanza dell'intervallo temporale
    - ✓ La DFT è una MEDIA su un intervallo temporale
  - Tempo varianza e stazionarietà locale
- 
- 
- **ATTIVITA' – Analisi del segnale vocale: fonemi**
    - ✓ Praat - <https://praat.org/>

# Analisi segnale vocale (Praat)

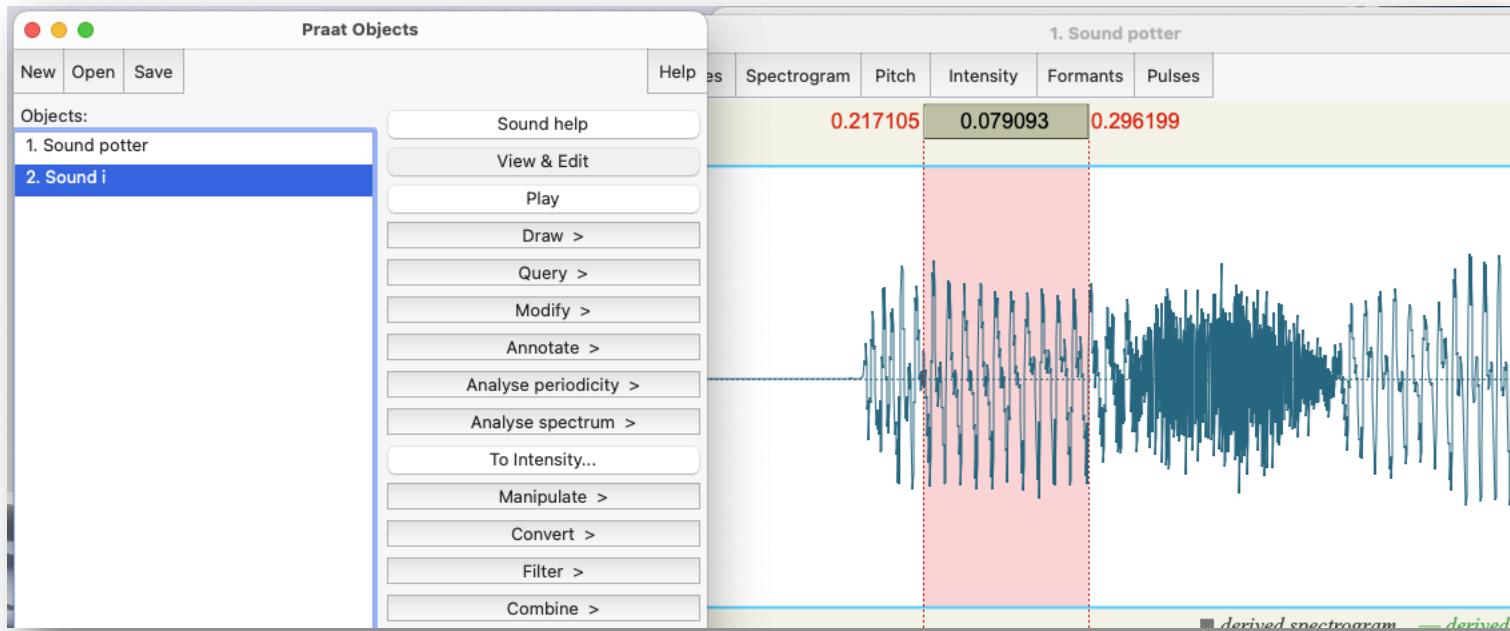
- Normalizzazione: *(sound) Modify > Scale peak: 0.99*
- *(sound) View & Edit*
- Riproduzione con tasto tab o click su barre selezione, visibile, tutto
- Selezione e zoom (bottoni in basso a sx)
- Esportazione (e.g. 100ms)  
*File > Selected sound as WAV file*
  
- Spectrogram  
*Spectrogram > Settings > View range: 8 kHz*



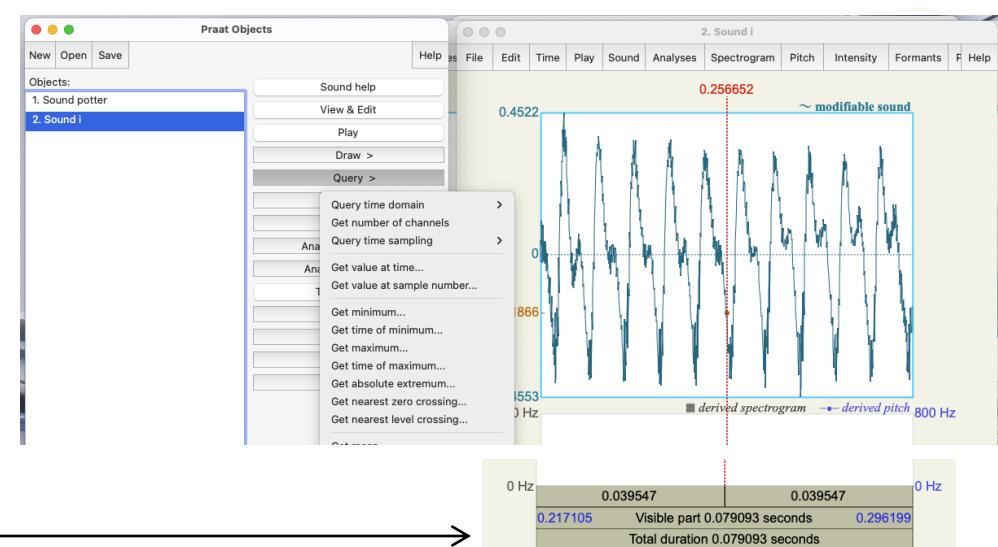
Reference: Praat Tutorial (videos) <https://www.youtube.com/@linguistiklaboralbert-ludw3514/videos>

# Selezione ed "estrazione"

- Selezionare l'area di interesse
  - ✓ *Sound > Extract Selected Sound (preserve times)*
- Rinominare il nuovo oggetto "sound"



# Analisi temporale (sound)

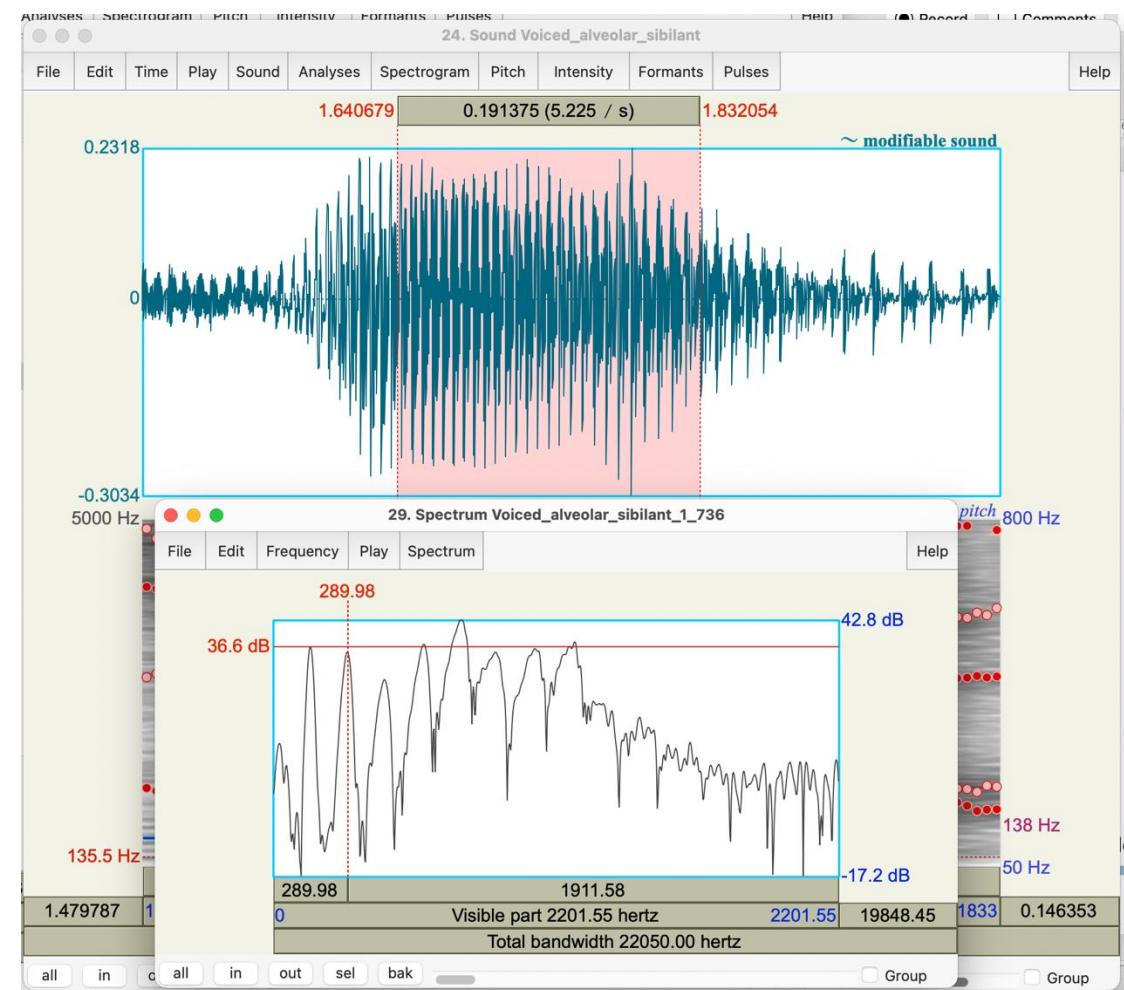
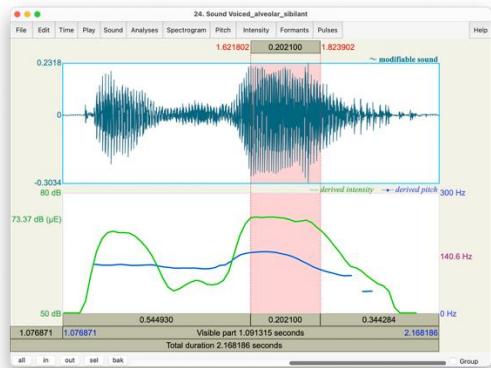
- Intensità: *(sound) Query > Get Intensity (dB)*
  - Intensity settings
    - ✓ Visualizza il valore in SPL non in dBFS (ma non è affatto tarato). Al denominatore invece di 1 ci sono  $(2 \cdot 10^{-5} \text{ Pa})^2$ , quindi si ottiene un valore positivo +94 dB rispetto ai dBFS
  - RMS:  
*(sound) Query*  
    *> Get root-mean-square (dB)*
  - Durata:  
*(sound) Query*  
    *> Query time domain*  
    *> Get total duration*
- 

# Spettro (Praat)

## ■ Su una selezione "piccola"

- ✓ *Spectrogram > View Spectral Slice*
- ✓ Automaticamente crea oggetto spettro

## ■ Pitch > Show pitch



# Fonema

- E' definito come un suono prodotto dall'apparato fonatorio umano che corrisponde a una unità linguistica dotata di valore distintivo
- E' un concetto differente dalla lettera o dal grafema
- E' rappresentato in forma scritta per mezzo delle trascrizioni fonetiche
  - ✓ Alfabeto fonetico internazionale

- Praat – simboli fonetici
  - ✓ [https://www.fon.hum.uva.nl/praat/manual/Phonetic\\_symbols.html](https://www.fon.hum.uva.nl/praat/manual/Phonetic_symbols.html)

- ✓ Simboli utilizzati per la lingua italiana:  
[https://it.wikipedia.org/wiki/Aiuto:IPA\\_per\\_l%27italiano](https://it.wikipedia.org/wiki/Aiuto:IPA_per_l%27italiano)

Consonanti <sup>[N 1]</sup>		Vocali <sup>[N 7]</sup>	
IPA	Esempi	IPA	Esempi
b	banca; cibo	a	alto; sarà
d <sup>[N 2]</sup>	dove; idra	e	vero; perché
dz <sup>[N 3]</sup>	zanzara; zaino; razzo <sup>[N 4]</sup>	ɛ	etto; cioè
dʒ <sup>[N 3]</sup>	gelato; giungla; magia; jeans	i	imposta; colibrì <sup>[N 5]</sup>
f	fatto; cofano	o	ombra; gogó
g	gatto; glifo; ghetto; lingua	ɔ	notte; sarò
j	ieri; scoiattolo; più; yacht	u	ultimo; putipù <sup>[N 8]</sup>
k	cane; scritto; anche; quei; kaiser		

Reference:

# Fonema

- E' definito come un suono prodotto dall'apparato fonatorio umano che corrisponde a una unità linguistica dotata di valore distintivo
- E' un concetto differente dalla lettera o dal grafema
- E' rappresentato in forma scritta per mezzo delle trascrizioni fonetiche
  - ✓ Alfabeto fonetico internazionale

- Praat – simboli fonetici
  - ✓ [https://www.fon.hum.uva.nl/praat/manual/Phonetic\\_symbols.html](https://www.fon.hum.uva.nl/praat/manual/Phonetic_symbols.html)

- ✓ Simboli utilizzati per la lingua italiana:  
[https://it.wikipedia.org/wiki/Aiuto:IPA\\_per\\_l%27italiano](https://it.wikipedia.org/wiki/Aiuto:IPA_per_l%27italiano)

Consonanti <sup>[N 1]</sup>		Vocali <sup>[N 7]</sup>	
IPA	Esempi	IPA	Esempi
b	banca; cibo	a	alto; sarà
d <sup>[N 2]</sup>	dove; idra	e	vero; perché
dz <sup>[N 3]</sup>	zanzara; zaino; razzo <sup>[N 4]</sup>	ɛ	etto; cioè
dʒ <sup>[N 3]</sup>	gelato; giungla; magia; jeans	i	imposta; colibrì <sup>[N 5]</sup>
f	fatto; cofano	o	ombra; gogó
g	gatto; glifo; ghetto; lingua	ɔ	notte; sarò
j	ieri; scoiattolo; più; yacht	u	ultimo; putipù <sup>[N 8]</sup>
k	cane; scritto; anche; quei; kaiser		

Reference:

# Vocali

---

- In italiano abbiamo 7 suoni vocalici: /a/, /e/, /ɛ/, /i/, /o/, /ɔ/, /u/
- A fronte di 5 grafemi: a, e, i, o, u
- Alcuni grafemi hanno una doppia pronuncia (chiusa, aperta)
- Tutte le vocali sono *voiced*, sonore, cioè con vibrazione delle corde vocali, e quindi con F0

*E il capitano disse: "certe volte le cose capitano". E poi disse al marinaio: "esca e vada a pesca di una pesca, e mi raccomando, per la pesca di una pesca ci vuole l'esca"*



Video (scuola doppiaggio Brescia): <https://www.youtube.com/watch?v=dzVbfZEiaiI>

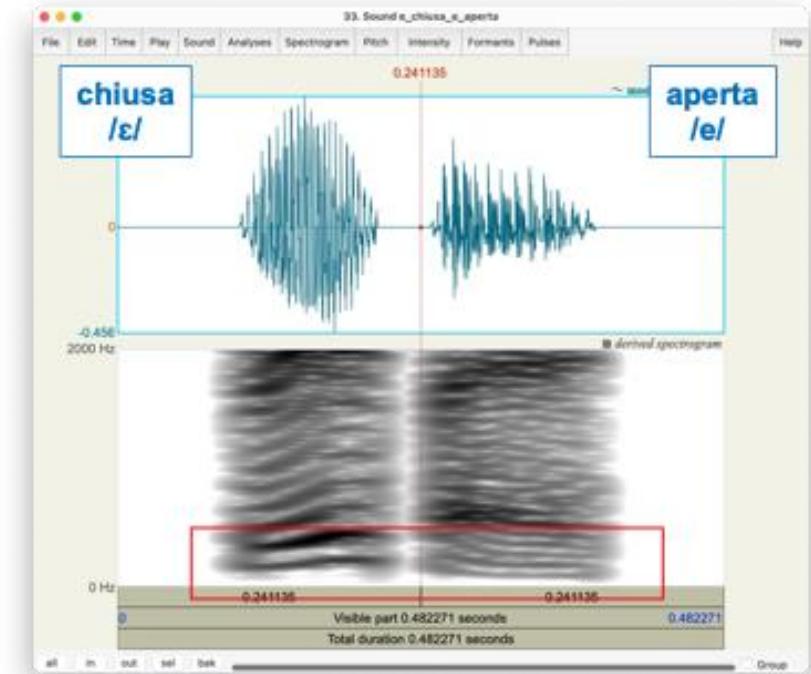
# Vocali aperte e chiuse

---

- Sono "doppie"
  - ✓ La 'e' chiusa /e/ ['peska] (péscia, fishing), perché (accento acuto), ésca (all'amo), e
  - ✓ La 'e' aperta /ɛ/ ['pɛska] (pèsca, peach), poichè (accento grave), è, èsca (uscire)
  - ✓ La 'o' chiusa /o/ voce, cicogna, rasoio
  - ✓ La 'o' aperta /ɔ/ carrozza, suono, però (parola termina ò accentata)
- Nota:
  - ✓ Dove non cade l'accento (atona) è sempre chiusa, con l'accento grave è aperta, con quello acuto è chiusa

# Vocali aperte e chiuse

- Dal punto di lista della fonazione si distinguono perché la versione aperta è pronunciata con la bocca più aperta della versione chiusa
- Dal punto di vista acustico nelle vocali chiuse abbiamo più energia alle basse frequenze che non alle alte
- Wikipedia
  - ✓ [https://it.wikipedia.org/wiki/Vocale\\_anteriore\\_semichiusa\\_non\\_arrotondata](https://it.wikipedia.org/wiki/Vocale_anteriore_semichiusa_non_arrotondata)
  - ✓ [https://it.wikipedia.org/wiki/Vocale\\_anteriore\\_semiaperta\\_non\\_arrotondata](https://it.wikipedia.org/wiki/Vocale_anteriore_semiaperta_non_arrotondata)



e\_chiusa\_e\_aperta.wav

# Fonemi sordi e sonori

---

- Non solo le vocali sono sonore, ma anche alcune consonanti sono sonore, cioè fanno vibrare le corde vocali
- Occlusive sonore: /b/: baco, /d/: dado, /g/: gatto
- Fricative sonore: /v/: vino, /z/: casa
- Nasali: /m/: mano, /n/: nave
- Altre: /l/: luna, /r/: riso

# 's' sonora o sorda

---

- Alcune consonanti hanno un suono "doppio"
- La s è sorda /s/
  - ✓ In inizio di parola: specie, sette
  - ✓ Quando viene raddoppiata: rosso, gesso
  - ✓ Aggettivi in -oso: affettuoso, vanitoso
  - ✓ Ma anche: trentasei, disegno, casa
- La 's' è sonora /z/
  - ✓ Davanti alle consonanti (b/d): sbadato, sdentato, sbaglio, sdegno
  - ✓ Nelle finali in -asi/e -esimo: rimasi, centesimo, umanesimo

Video (scuola doppiaggio Brescia): <https://www.youtube.com/watch?v=tvqle7Xo7Ds>

---

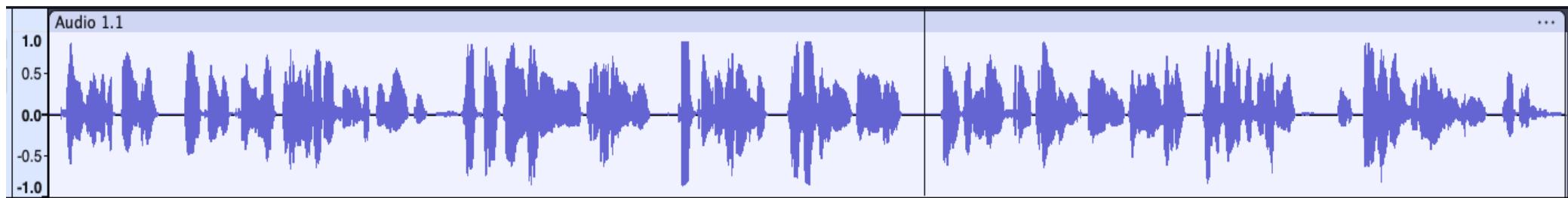
# 's' sonora o sorda

---

## ■ Esempio

- ✓ "dizione-italiana-la-s-sorda-e-sonora.wav "

*Chiese la sposa in quale chiesa si sarebbero sposati e pretese un impresario che con riserbo trovasse in paese trentasei francesi e un cinese della buona borghesia per presiedere l'ennesimo matrimonio.*

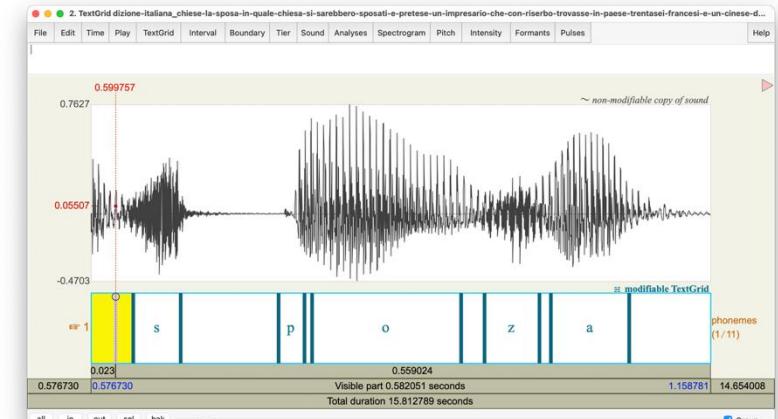


Video (scuola doppiaggio Brescia): <https://www.youtube.com/watch?v=tvqle7Xo7Ds>

---

# Esercizio: analisi fonemi

- Analizzare la parola "sposa" all'interno della frase "dizione-italiana-la-s-sorda-e-sonora.wav "
- /a/ ed /o/ - vocale (sonora):
  - ✓ F0, energia concentrata alle basse frequenze
- /s/ - consonante fricativa (sorda, non sonora)
  - ✓ No F0, energia alle alte frequenze
- /z/ - consonante fricativa (sonora)
  - ✓ F0, energia sia basse sia alte frequenze
- /p/ - occlusiva (sorda)
  - ✓ No F0, energia per una durata molto breve (burst)



"/spoza/"

Reference:

# Esercizio: analisi fonemi

## ■ Tramite Praat

- ✓ Annotiamo l'audio identificando le parti centrali dei fonemi
- ✓ Calcoliamo alcune caratteristiche audio dei fonemi

## ■ Tempi: inizio, fine e durata

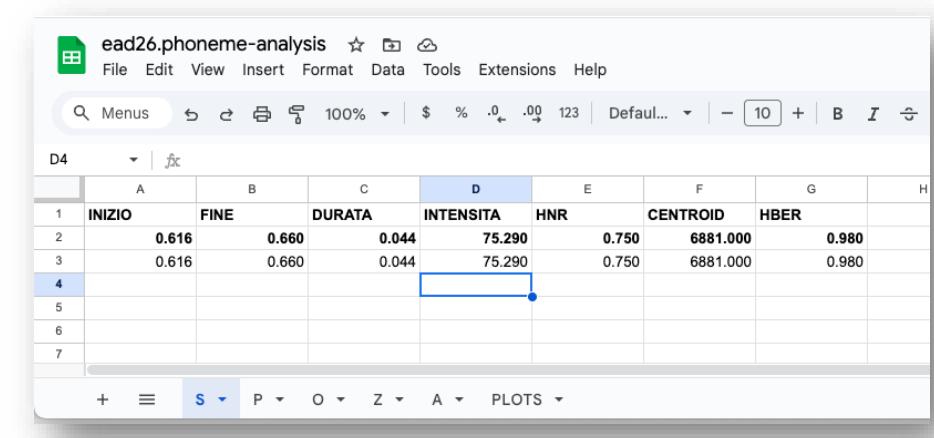
## ■ Energia: energia media (RMS)

## ■ Distribuzione delle frequenze

- ✓ Centroide Spettrale (Centre of Gravity)
- ✓ High Band Energy Ratio (HBER)

## ■ Armonicità

- ✓ Harmonic-to-Noise Ratio (HNR)



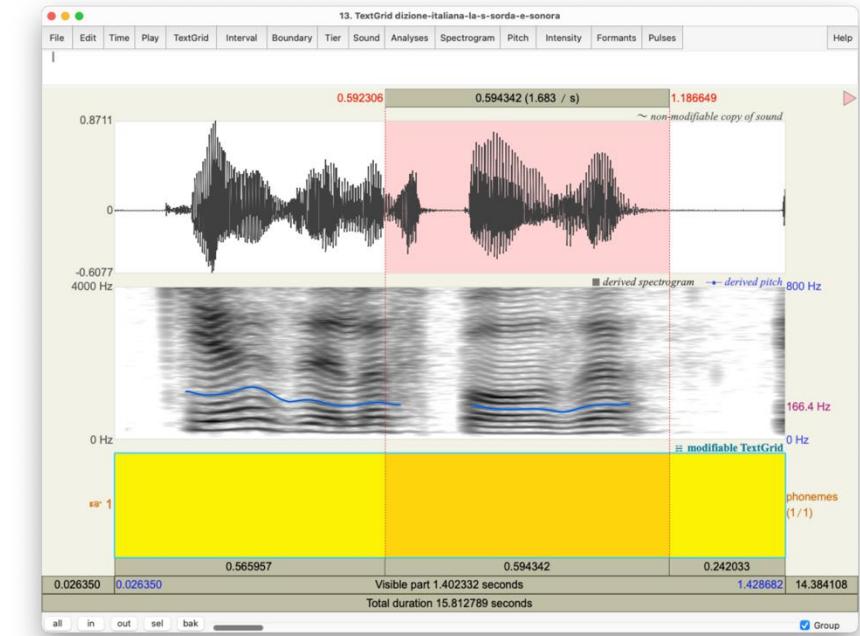
The screenshot shows a spreadsheet titled "ead26.phoneme-analysis" with data for four phonemes. The columns represent features: INIZIO, FINE, DURATA, INTENSITA, HNR, CENTROID, and HBER. The data is as follows:

	A	B	C	D	E	F	G	H
1	INIZIO	FINE	DURATA	INTENSITA	HNR	CENTROID	HBER	
2	0.616	0.660	0.044	75.290	0.750	6881.000	0.980	
3	0.616	0.660	0.044	75.290	0.750	6881.000	0.980	
4								
5								
6								
7								

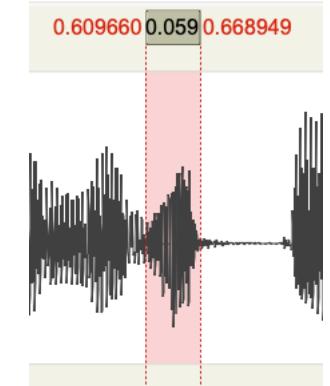
Reference: ead<YY>.phoneme-analysis

# Esercizio: analisi fonemi (Praat)

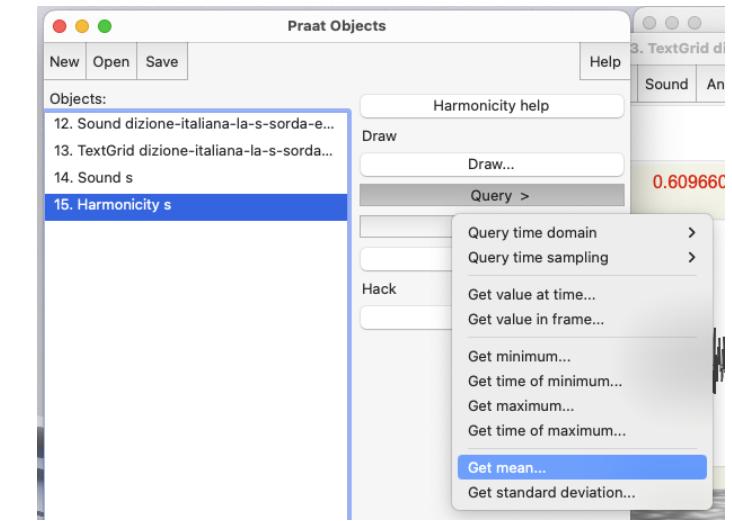
- Creazione traccia per annotazioni
  - ✓ Sound > Annotate > To Text Grid
  - ✓ All tier names: phonemes / Which .. point tiers?: <empty>
- Selezionare entrambi gli oggetti sound e textgrid
  - ✓ View & Edit
- Annotazione
  - ✓ Selezionare l'intervallo, ascoltarlo
  - ✓ Boundary
    - > "Add on selected tier":  
<simbolo fonema>



# Esercizio: analisi fonemi (Praat)



- Inizio, fine e durata si vedono dalla selezione
- Creare oggetti "sound" separati per ogni fonema
  - ✓ Selezionare il fonema (porzione audio)
  - ✓ Sound > Extract selected sound (preserve time)
  - ✓ Rinominare l'oggetto con il simbolo del fonema



- Harmonicity to Noise Ratio
  - ✓ (sound) Analyse periodicity  
    > To Harmonicity (cc) ...
  - ✓ (harmonicity) Query > Get mean

# Esercizio: analisi fonemi (Praat)

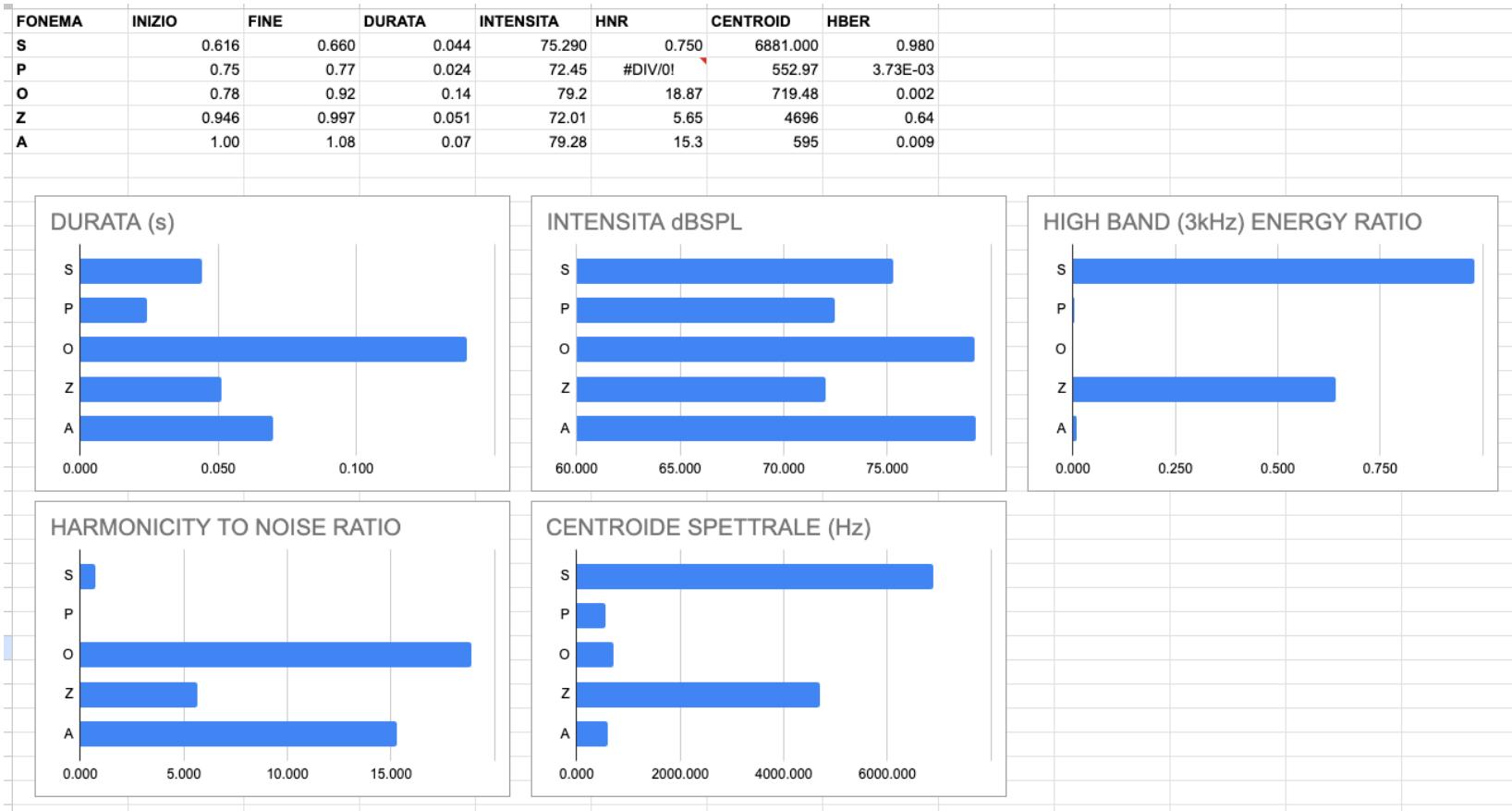
---

- Spectral Centroid
  - ✓ (sound) Analyse spectrum > To Spectrum ...
  - ✓ (spectrum) Query > Centre of Gravity
  
- High Band Energy Ratio
  - ✓ HB = (spectrum) Query > Get band energy: 3000 Hz – <sr> Hz
  - ✓ Total = (spectrum) Query > Get band energy: 0 Hz - <sr> Hz
  - ✓ HBER = HB/Total

Reference:

---

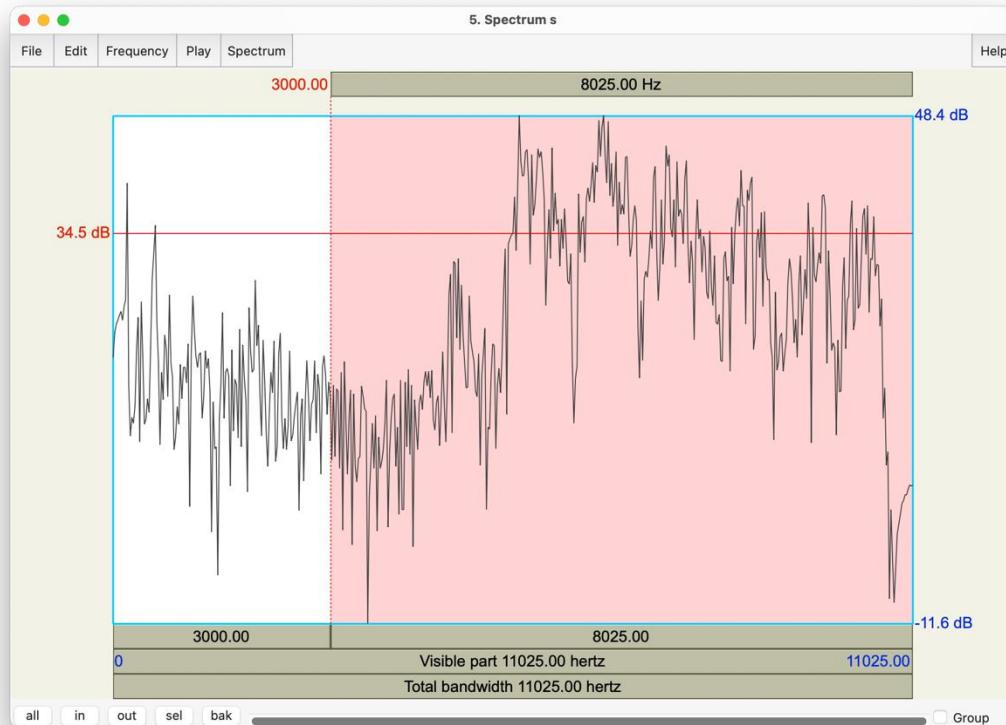
# Analisi dei risultati



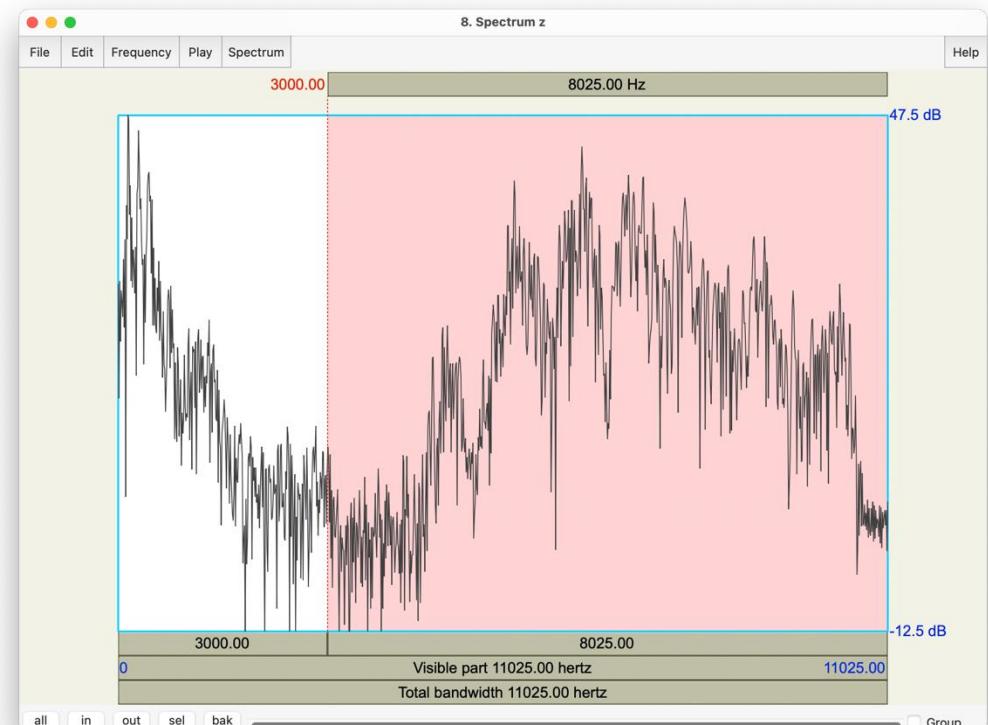
Reference: ead<YY>.phoneme-analysis

# 's' sorda /s/ e sonora /z/

/s/



/z/



Reference: