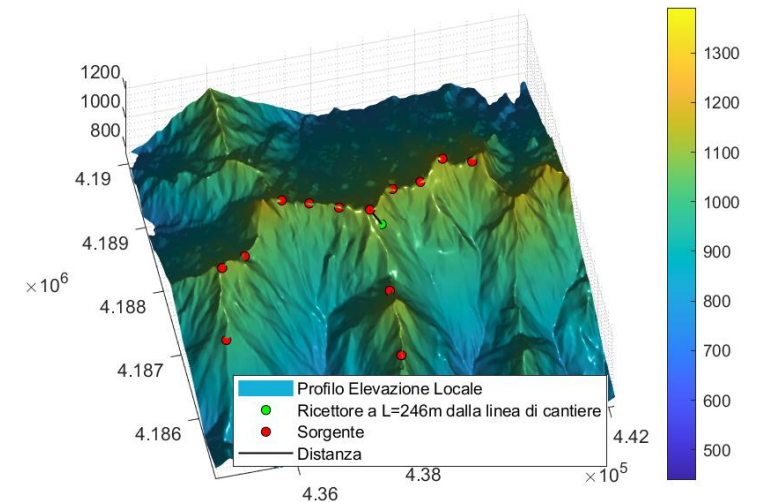
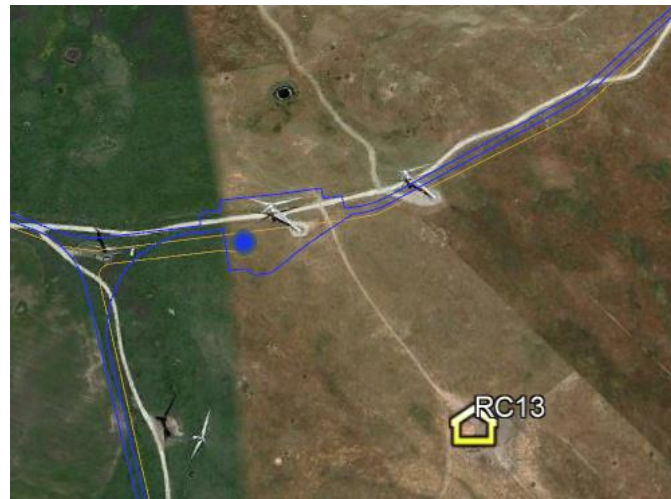
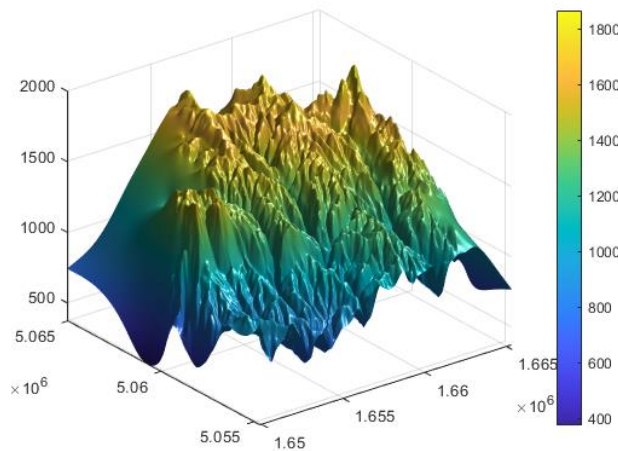


Impatto vibrazionale – una possibile procedura

Data una sorgente che induce vibrazioni, è possibile stimare un eventuale il disturbo presso un ricettore?

E' possibile stimare il livello vibrazionale presso un ricettore

SPETTRO DI ECCITAZIONE
DISTANZA SORGENTE-RICETTORE
CARATTERISTICHE DEL TERRENO
SCELTA DI UN MODELLO DI ATTENUAZIONE



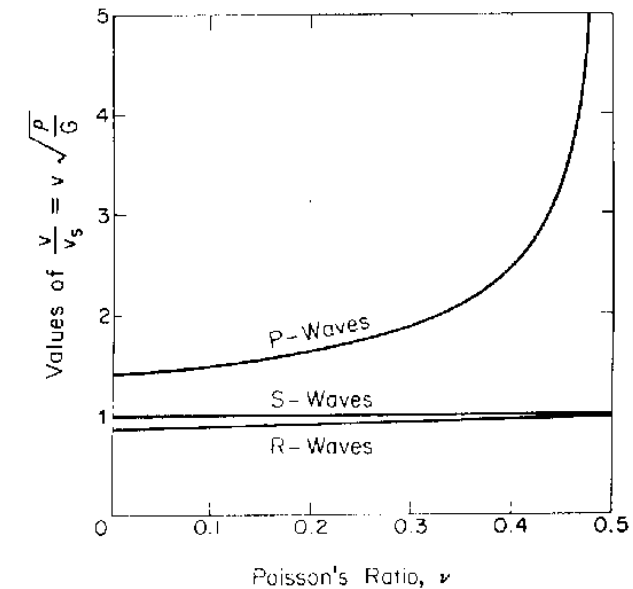
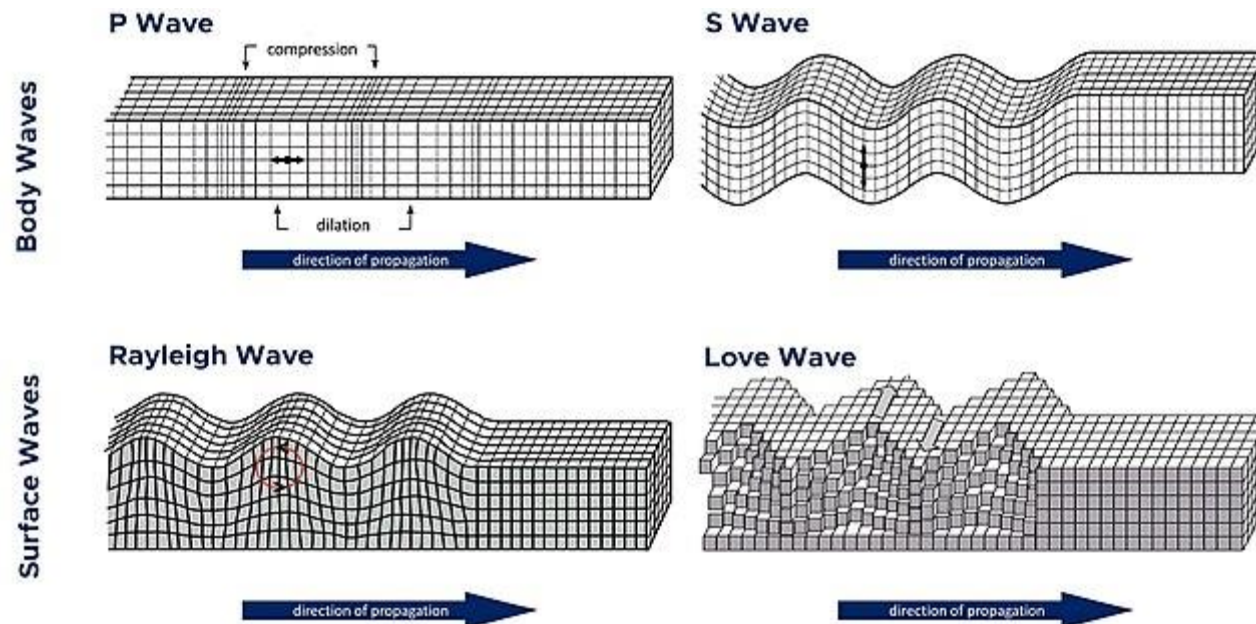
Impatto vibrazionale – una possibile procedura

MODELLO DI BASSA VELOCITA'

Il disturbo vibrazionale ai ricettori dipende da meccanismi di propagazione relativi a: spettro di emissione della sorgente, distanze e tipologia di suolo, risposta strutturale dell'edificio e sensibilità fisiologica

MODELLO CTS (COMPRESSIONE, TAGLIO E SUPERFICIE)

- Onde di corpo
- Onde di superficie



Impatto vibrazionale – una possibile procedura

MODELLO DI BASSA VELOCITA'

Il disturbo vibrazionale ai ricettori dipende da meccanismi di propagazione relativi a: spettro di emissione della sorgente, distanze e tipologia di suolo, risposta strutturale dell'edificio e sensibilità fisiologica

MODELLO CTS (COMPRESSIONE, TAGLIO E SUPERFICIE)

- Onde di corpo
- Onde di superficie

$$L = 20 \cdot \log \left[10^{\frac{L_c}{20}} + 10^{\frac{L_t}{20}} + 10^{\frac{L_s}{20}} \right]$$

$$\begin{aligned} L_c &= L_0 + 20 \cdot \log(\beta_c) - k_c \cdot \log\left(\frac{R}{R_0}\right) - \alpha_c (R - R_0) \frac{f}{V_c} \\ L_t &= L_0 + 20 \cdot \log(\beta_t) - k_t \cdot \log\left(\frac{R}{R_0}\right) - \alpha_t (R - R_0) \frac{f}{V_t} \\ L_s &= L_0 + 20 \cdot \log(\beta_s) - k_s \cdot \log\left(\frac{R}{R_0}\right) - \alpha_s (R - R_0) \frac{f}{V_s} \end{aligned}$$

- L ed L0 rappresentano i livelli di vibrazione in decibel alla velocità V e V0, ovvero alla velocità di transito e alla velocità di riferimento
- R distanza plano-altimetrica, riferita all'asse della linea, del ricettore
- R0 distanza plano-altimetrica, riferita all'asse della linea, della sorgente (pt. di riferimento)
- Vc Velocità per le onde di compressione
- Vt Velocità per le onde di taglio
- Vs Velocità per le onde di superficie

Impatto vibrazionale – una possibile procedura

MODELLO DI BASSA VELOCITA'

$$L = 20 \cdot \log \left[10^{\frac{L_c}{20}} + 10^{\frac{L_t}{20}} + 10^{\frac{L_s}{20}} \right]$$

$$L_c = L_0 + 20 \cdot \log(\beta_c) - k_c \cdot \log\left(\frac{R}{R_0}\right) - \alpha_c (R - R_0) \frac{f}{V_c}$$

$$L_t = L_0 + 20 \cdot \log(\beta_t) - k_t \cdot \log\left(\frac{R}{R_0}\right) - \alpha_t (R - R_0) \frac{f}{V_t}$$

$$L_s = L_0 + 20 \cdot \log(\beta_s) - k_s \cdot \log\left(\frac{R}{R_0}\right) - \alpha_s (R - R_0) \frac{f}{V_s}$$

- L ed L0 rappresentano i livelli di vibrazione in decibel alla velocità V e V0, ovvero alla velocità di transito e alla velocità di riferimento
- R distanza plano–altimetrica, riferita all'asse della linea, del ricettore
- R0 distanza plano–altimetrica, riferita all'asse della linea, della sorgente (pt. di riferimento)
- Vc Velocità per le onde di compressione
- Vt Velocità per le onde di taglio
- Vs Velocità per le onde di superficie

Tipo di sorgente	Tipo di onde di corpo	Tipo di onde di superficie
Puntuale	K = 20	K = 10
Lineare	K = 10	K = 0

Tipo di onda	Rilevato Raso	Viadotto	Galleria Profonda
Compressione	bc = 0.20	bc = 0.35	bc = 0.40
Taglio	bt = 0.20	bt = 0.35	bt = 0.40
Superficie	bs = 0.60	bs = 0.30	bs = 0.20

Impatto vibrazionale – una possibile procedura

MODELLO DI ATTENUAZIONE ESPONENZIALE

Nel caso del modello di attenuazione esponenziale, i parametri dimensionanti sono le distanze tra sorgente e ricettore, lo spettro di emissione, il terreno e la velocità di propagazione delle onde nel terreno

Il fattore $\rho = \frac{\eta}{c}$ può essere ricavato da input tabellari

$$a(d, f) = a(d_0, f) \cdot \left(\frac{d_0}{d} \right)^n \cdot e^{-2 \cdot \pi \cdot f \cdot \eta / c \cdot (d - d_0)}$$

Classe	Descrizione tipologia di terreno	Coefficiente di attenuazione	ρ
I	Cedevole o tenero (terreno che può essere scavato facilmente)	0.003-0.01	$2 \times 10^{-4} - 6 \times 10^{-4}$
II	Consolidato (terreno che può essere scavato utilizzando una pala)	0.001-0.003	$6 \times 10^{-5} - 2 \times 10^{-4}$
III	Duro (terreno che non può essere scavato con una pala ma necessità di un piccone)	0.0001-0.001	$6 \times 10^{-6} - 6 \times 10^{-5}$
IV	Duro consolidato (terreno che scavato difficilmente utilizzando un martello)	<0.0001	< 6×10^{-6}

Impatto vibrazionale – una possibile procedura

MODELLO DI ATTENUAZIONE ESPONENZIALE

Nel caso del modello di attenuazione esponenziale, i parametri dimensionanti sono le distanze tra sorgente e ricevitore, lo spettro di emissione, il terreno e la velocità di propagazione delle onde nel terreno

Il fattore $\rho = \frac{\eta}{c}$ può essere ricavato da input tabellari

$$a(d, f) = a(d_0, f) \cdot \left(\frac{d_0}{d} \right)^n \cdot e^{-2 \cdot \pi \cdot f \cdot \eta / c \cdot (d - d_0)}$$

Tipo di sorgente	Onda	Strato	n
Linea	Superficie	Superficie	0
	Volume	Superficie	1
Punto	Rayleigh	Superficie	0.5
	Volume	Superficie	2.0
Linea sotterranea	Volume	Profondo	0.5
Punto sotterraneo	Volume	Profondo	1.0

Impatto vibrazionale – una possibile procedura

Caratterizzazione della sorgente

Il profilo di emissione della sorgente può essere determinato per via sperimentale o numerica, a partire da dati di letteratura.

E' possibile caratterizzare il profilo di emissione e determinare la legge di attenuazione del terreno in esame

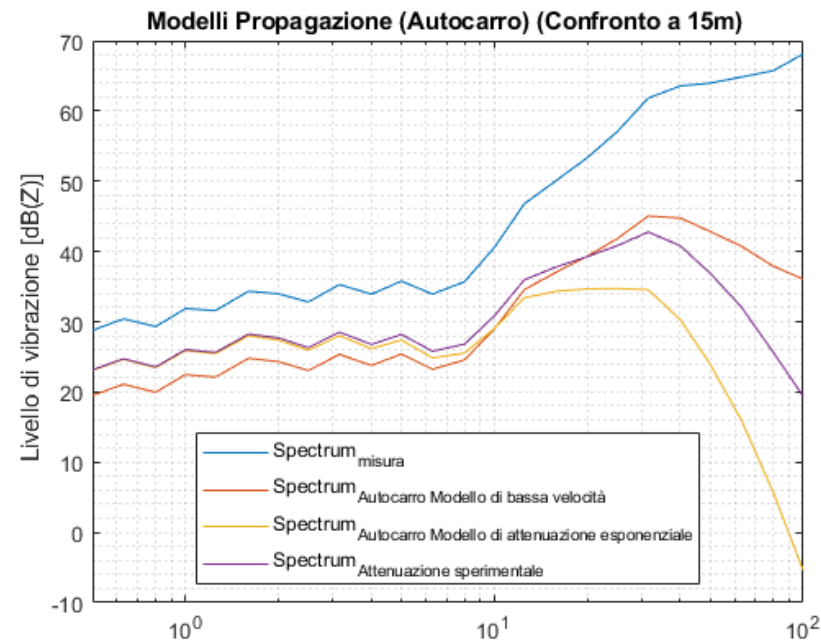
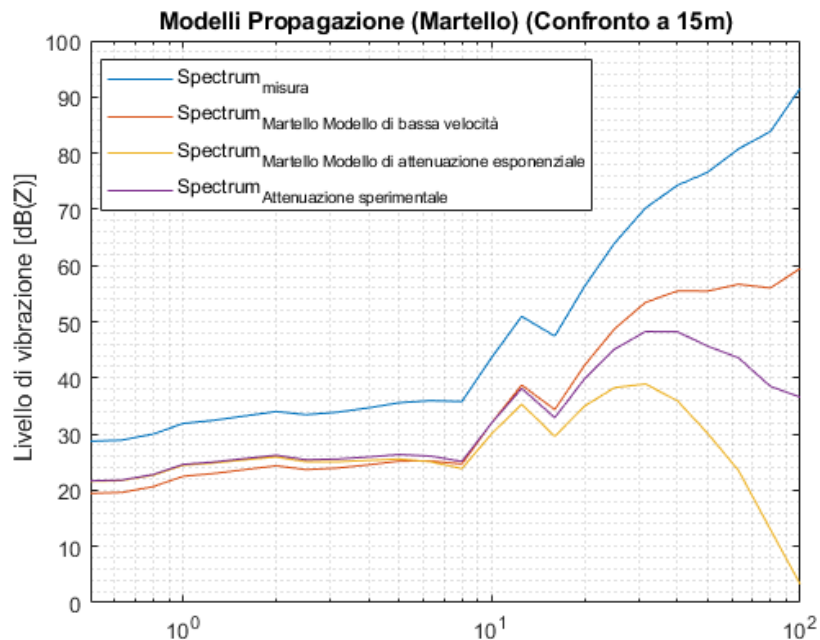


Impatto vibrazionale – una possibile procedura

Caratterizzazione della sorgente e terreno

Il profilo di emissione della sorgente può essere determinato per via sperimentale o numerica, a partire da dati di letteratura.

E' possibile caratterizzare il profilo di emissione e determinare la legge di attenuazione del terreno in esame

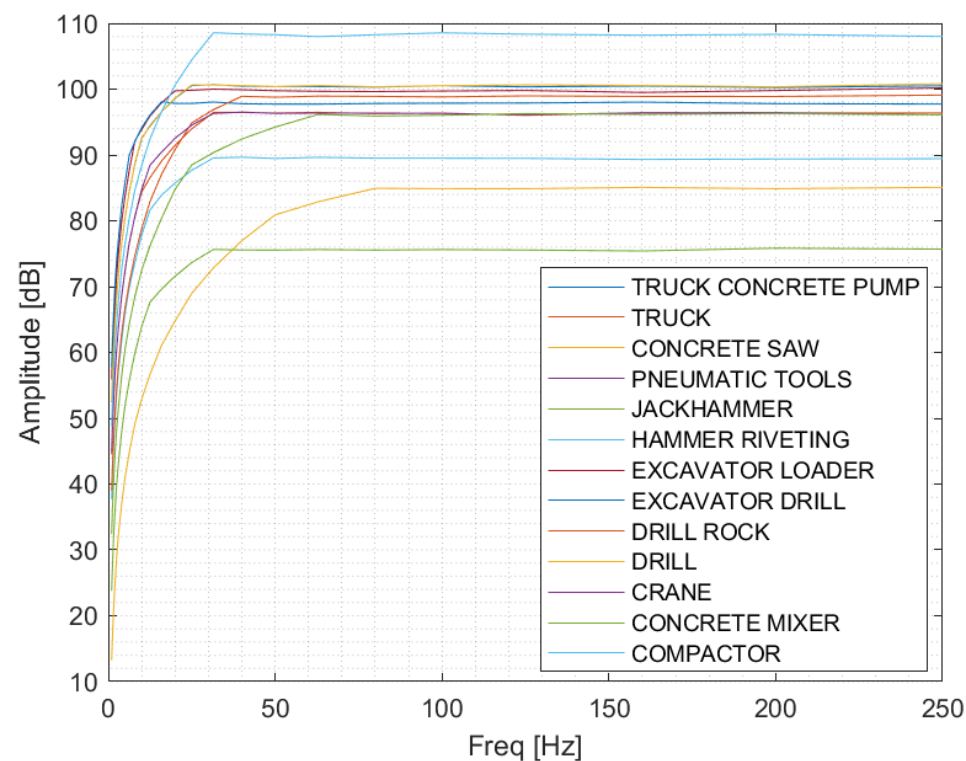


Impatto vibrazionale – una possibile procedura

Caratterizzazione della sorgente mediante profili di letteratura

Il profilo di emissione della sorgente può essere determinato per via sperimentale o numerica, a partire da dati di letteratura.

Sono presenti delle banche dati di profili vibrazionali acquisiti su suolo cementizio

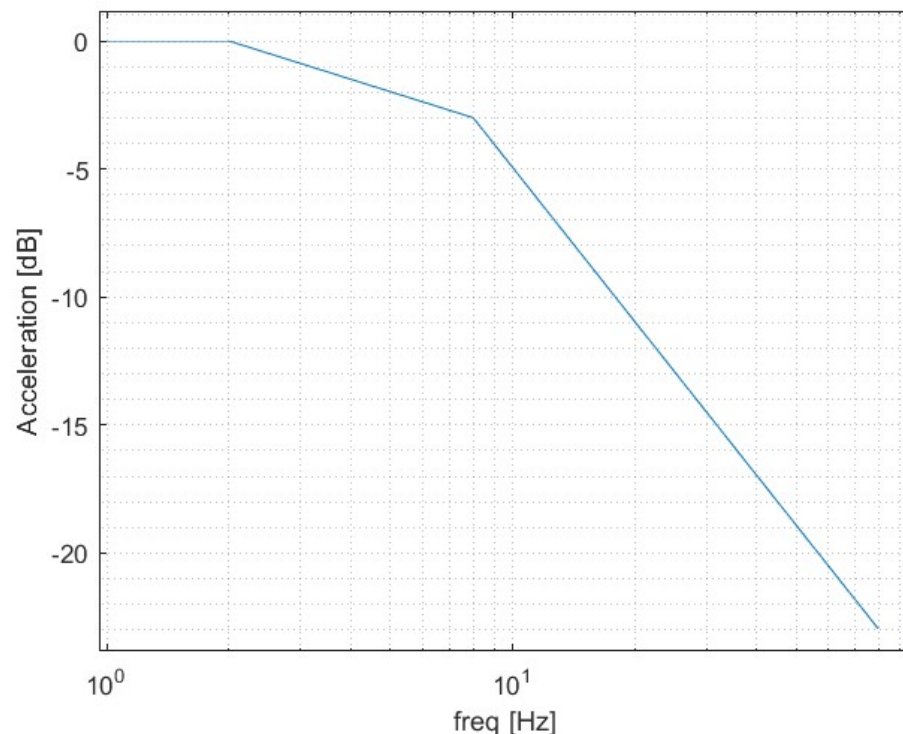


Impatto vibrazionale – una possibile procedura

Caratterizzazione della sorgente mediante profili di letteratura

Il profilo di emissione della sorgente può essere determinato per via sperimentale o numerica, a partire da dati di letteratura.

E' opportuno considerare il filtro di risposta del corpo umano al disturbo vibrazionale



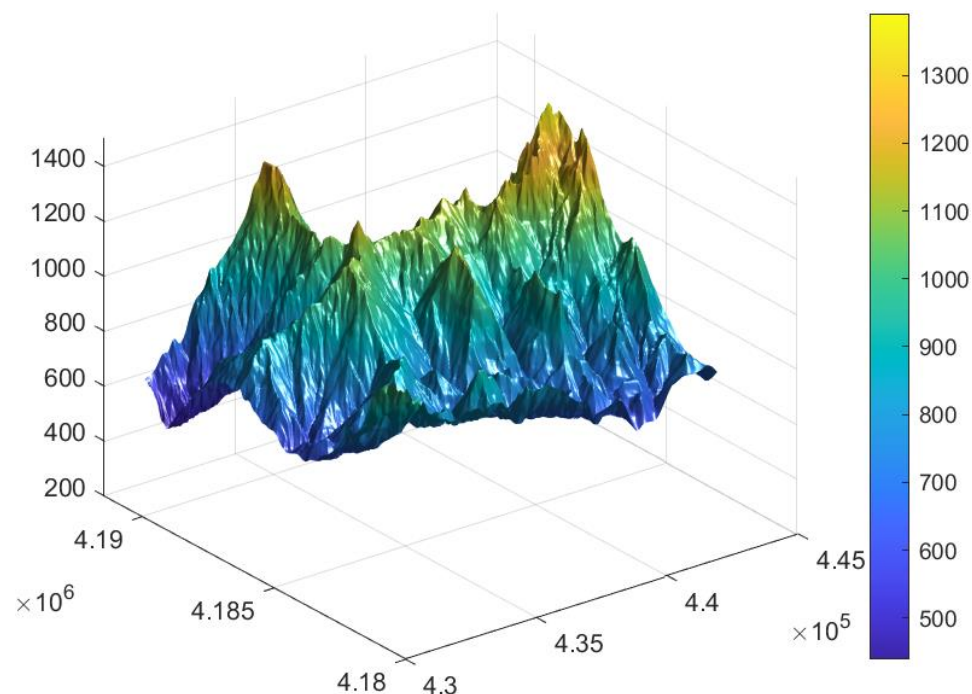
$$A_{TOT,f} = \sqrt{A_1(f,d)^2 + A_2(f,d)^2 + \dots + A_N(f,d)^2}$$

Impatto vibrazionale – una possibile procedura

Ricostruzione del profilo altimetrico locale per determinazione distanza piano - altimetrica

Entrambi i modelli di attenuazione prevedono
l'utilizzo della distanza piano-altimetrica

Funzione per la ricostruzione del profilo di elevazione
locale a partire da dati georeferenziati

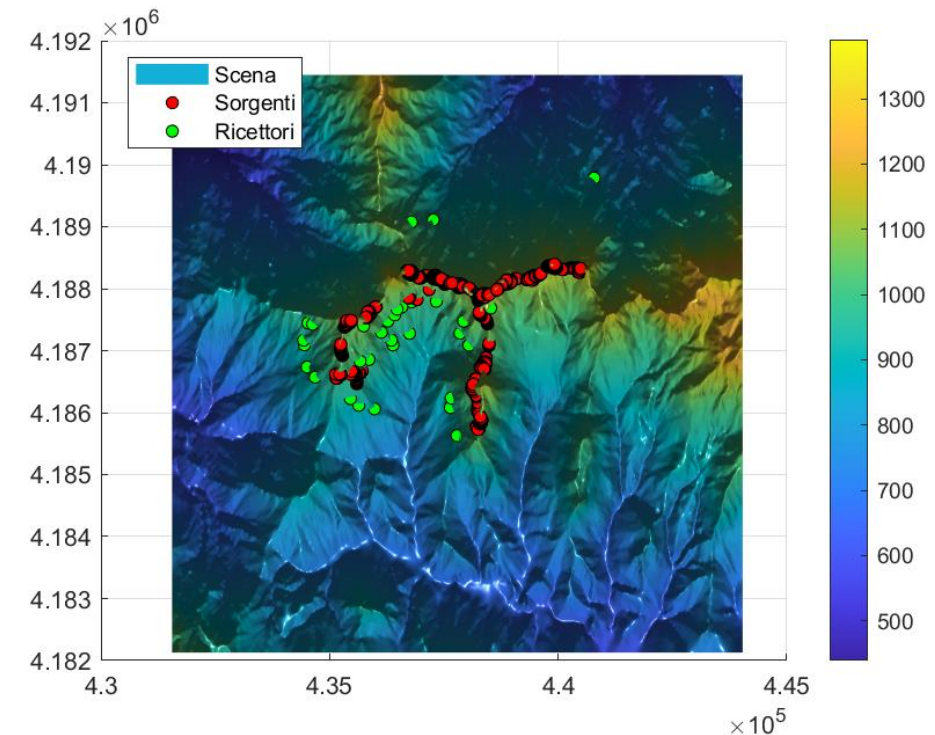
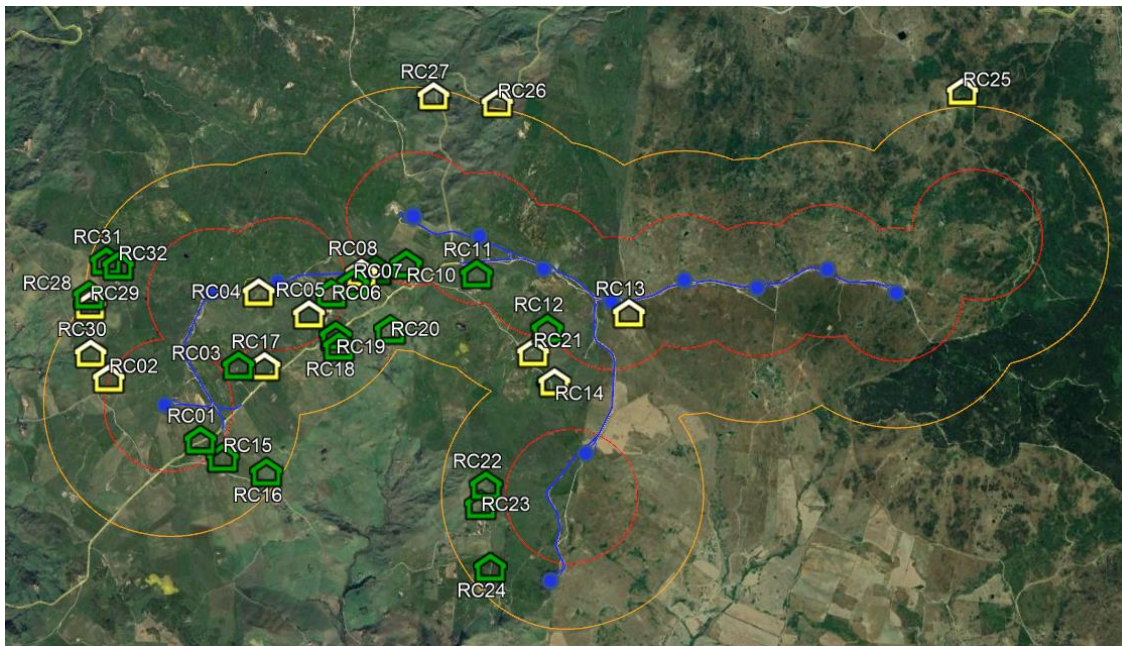


Impatto vibrazionale – una possibile procedura

Ricostruzione del profilo altimetrico locale per determinazione distanza plano - altimetrica

Entrambi i modelli di attenuazione prevedono
l'utilizzo della distanza plano-altimetrica

Importazione layout ricettori e linea di cantiere

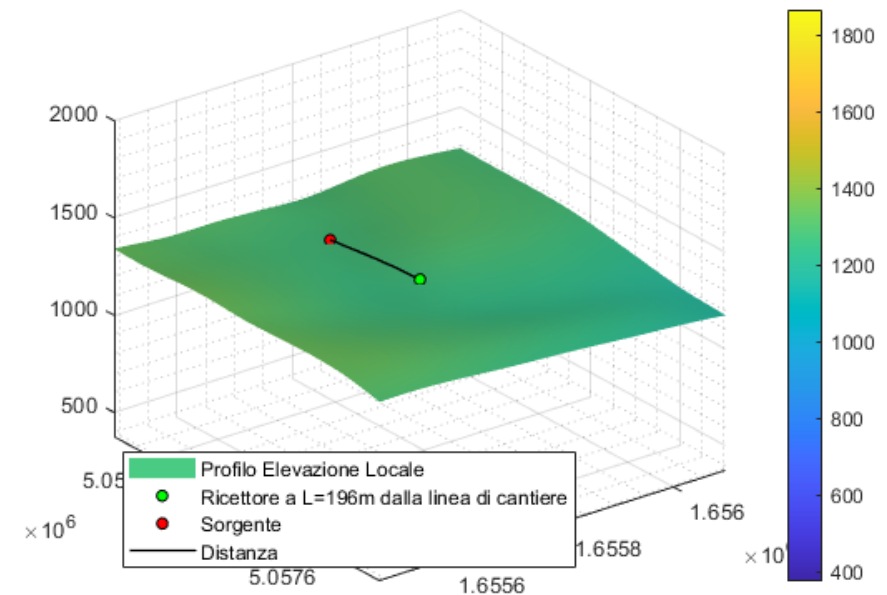
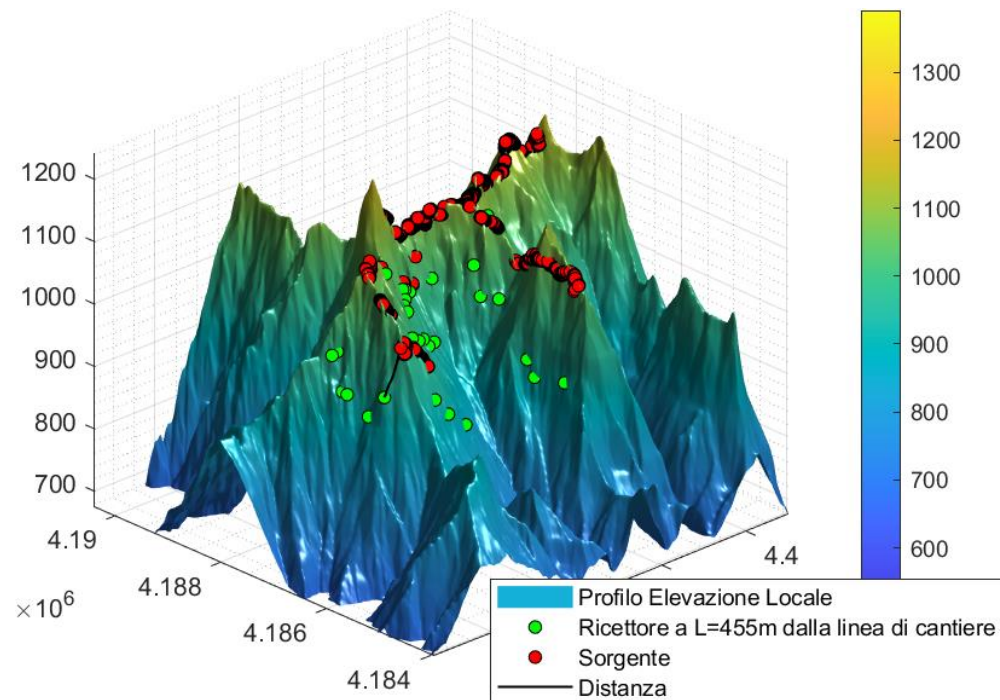


Impatto vibrazionale – una possibile procedura

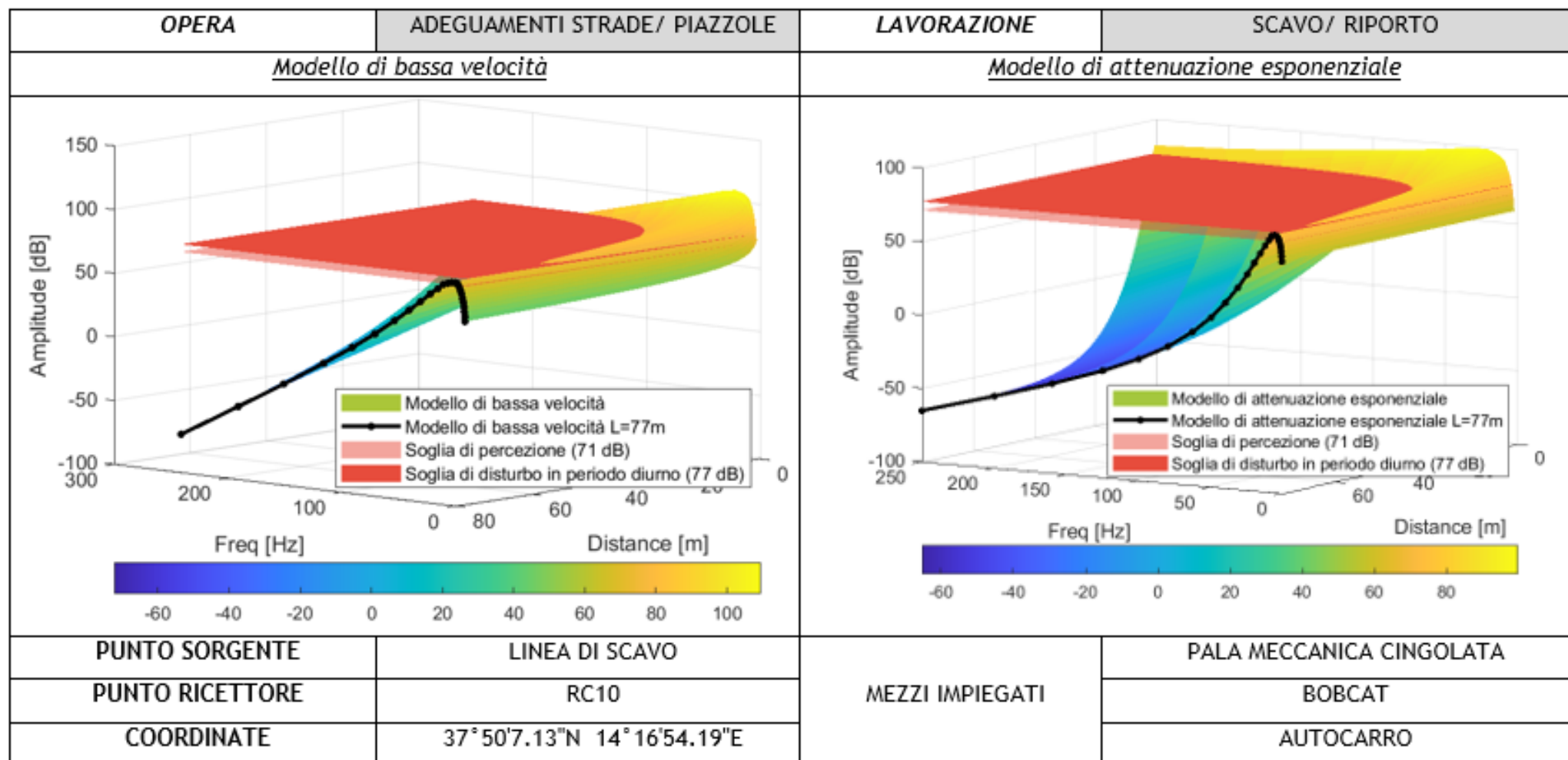
Ricostruzione del profilo altimetrico locale per determinazione distanza piano - altimetrica

Entrambi i modelli di attenuazione prevedono
l'utilizzo della distanza piano-altimetrica

Determinazione distanza tra linea di cantiere e
ricettore su profilo di elevazione locale



Impatto vibrazionale – una possibile procedura



Alcuni spunti di riflessione



**Siamo realmente sicuri di
ciò che misuriamo?**

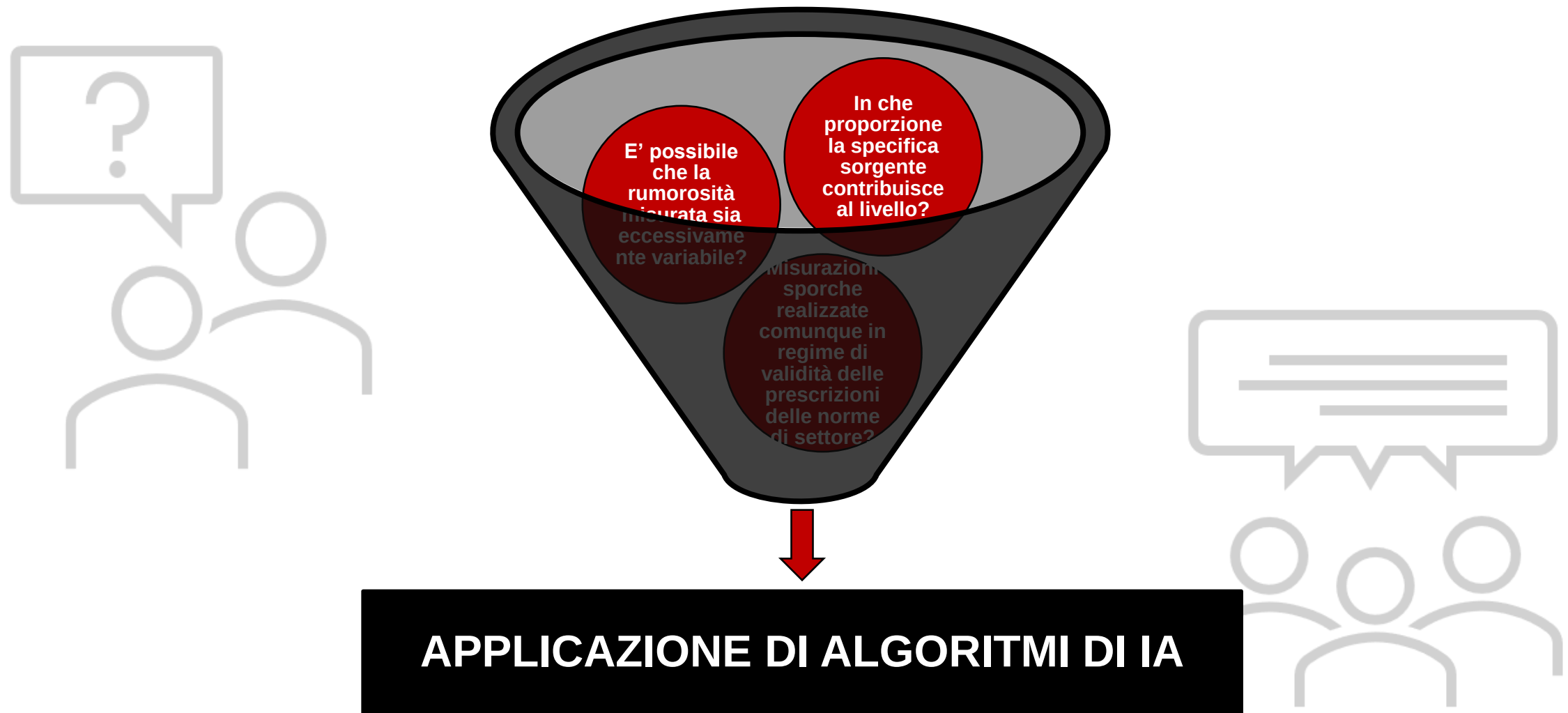
```
graph TD; A[Siamo realmente sicuri di ciò che misuriamo?] --- B[In che proporzione la specifica sorgente contribuisce al livello?]; A --- C[In caso di fenomeni a bassa rumorosità e lunga durata, è possibile che la rumorosità misurata sia eccessivamente variabile?]; A --- D[Cosa fare per misurazioni sporche (e.g. forte vento) realizzate comunque in regime di validità delle prescrizioni delle norme di settore?];
```

**In che proporzione la
specifica sorgente
contribuisce al livello?**

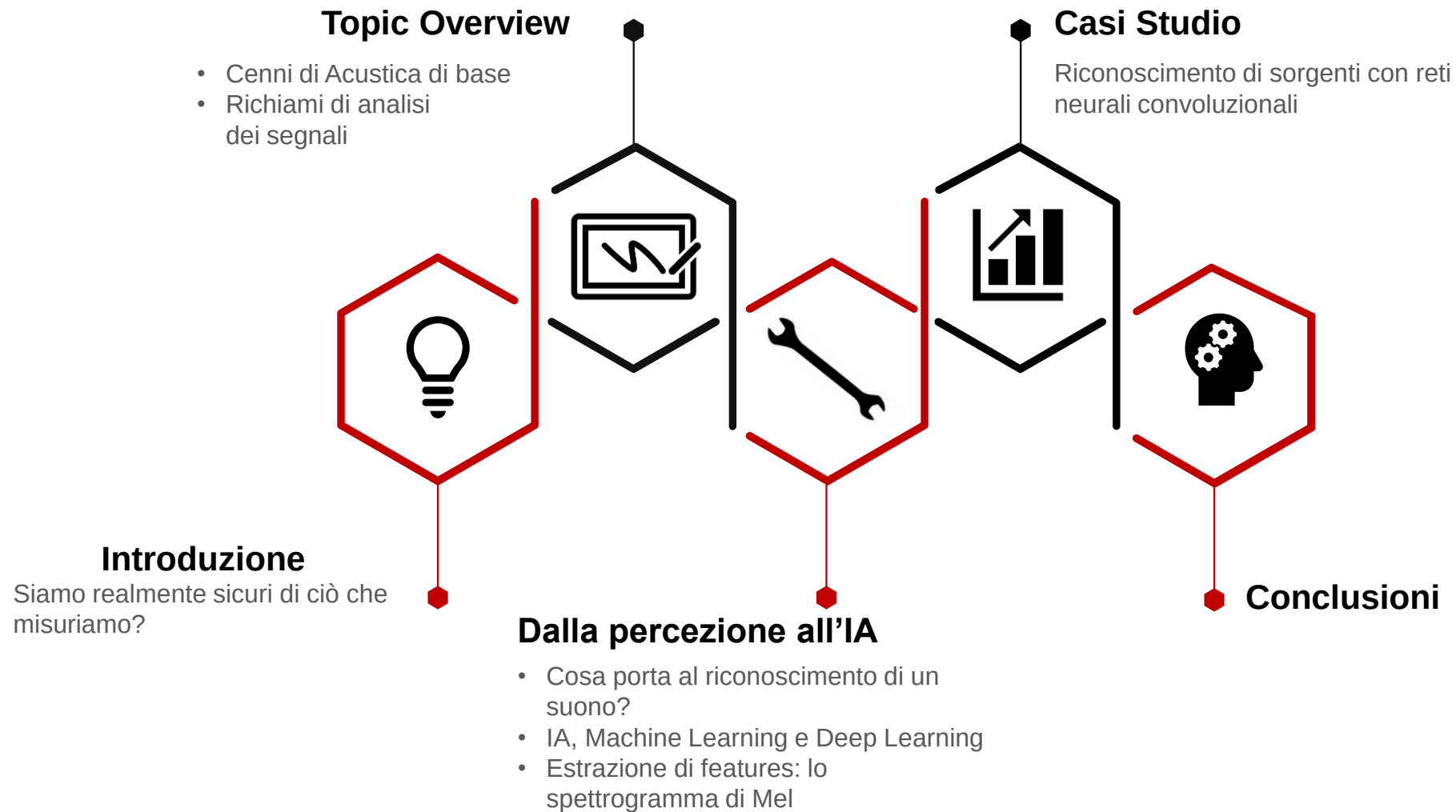
**In caso di fenomeni a
bassa rumorosità e lunga
durata, è possibile che la
rumorosità misurata sia
eccessivamente variabile?**

**Cosa fare per misurazioni
sporche (e.g. forte vento)
realizzate comunque in
regime di validità delle
prescrizioni delle norme di
settore?**

Alcuni spunti di riflessione



DI COSA PARLIAMO

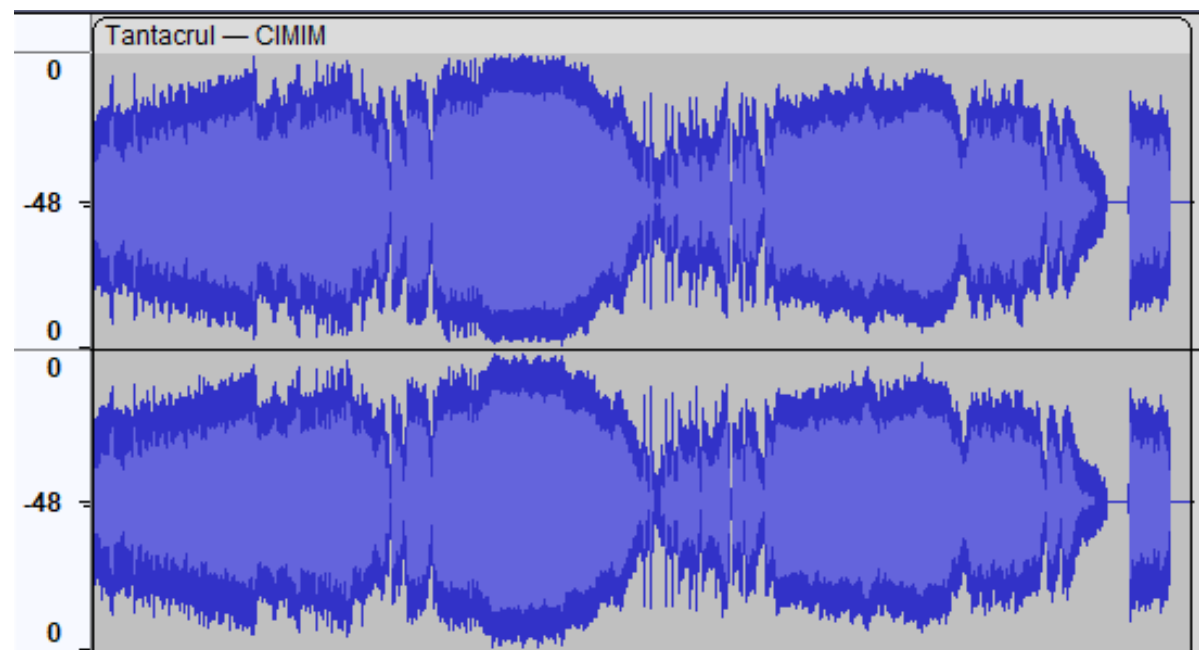
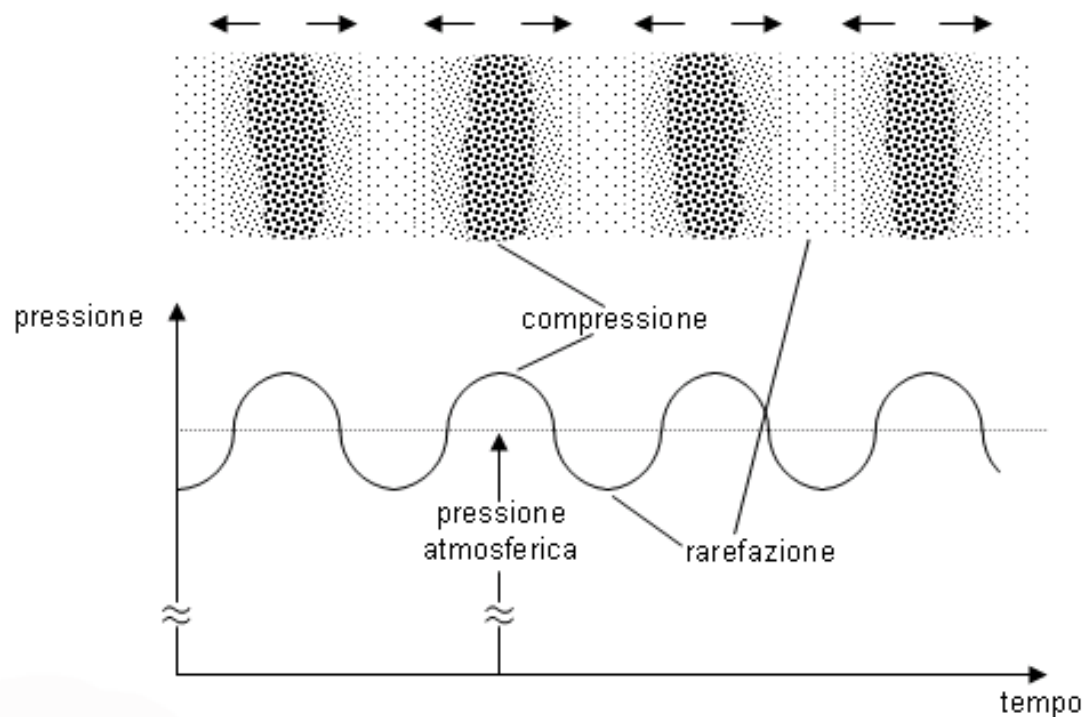


Cenni di Acustica di base

Suono: «Il prodotto di una oscillazione meccanica di un mezzo elastico».

Un disturbo di pressione che si propaga da una sorgente ad un ricevitore

Un'onda meccanica, ovvero un'oscillazione, che propaga energia in un mezzo attraverso una compressione/rarefazione dello stesso

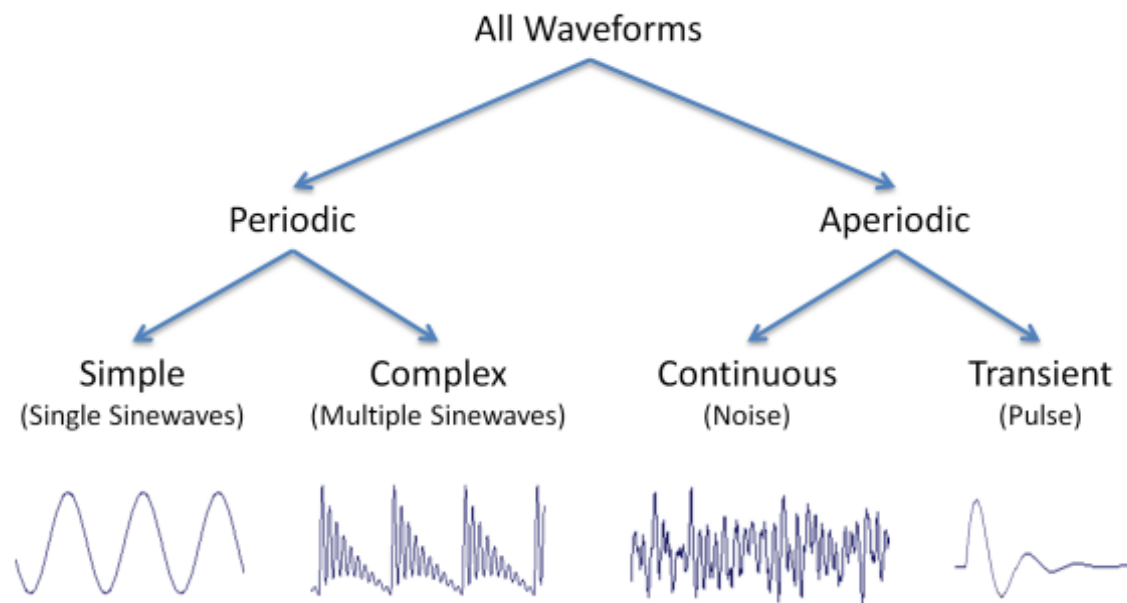
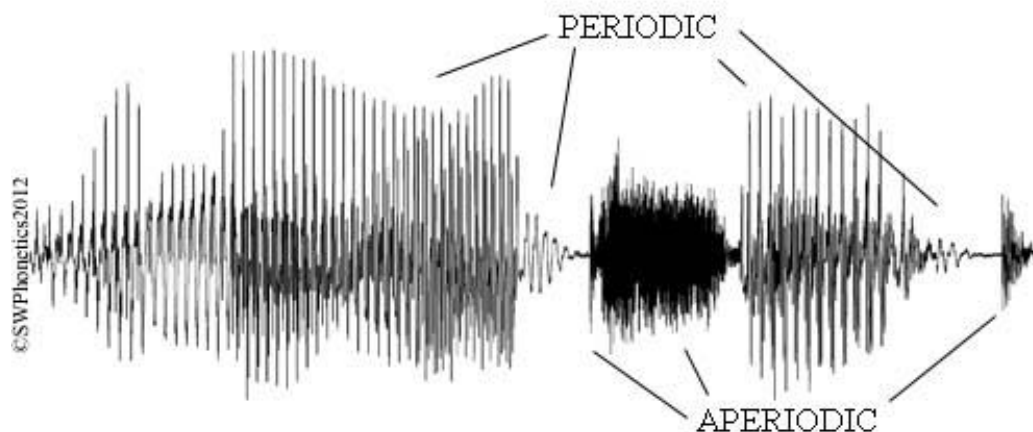


Cenni di Acustica di base

Suono: «Il prodotto di una oscillazione meccanica di un mezzo elastico».

Intensità, timbro, frequenza, inviluppo, durata, ...

FEATURES



Cenni di Acustica di base

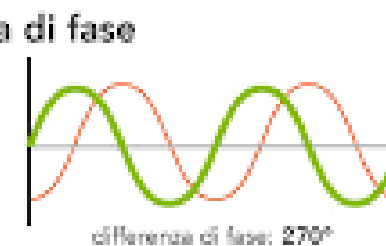
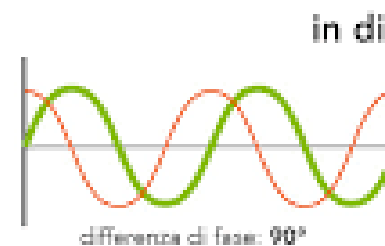
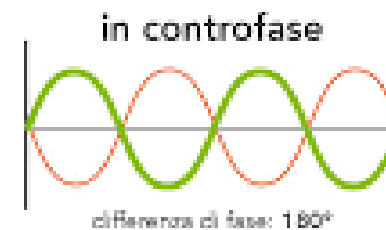
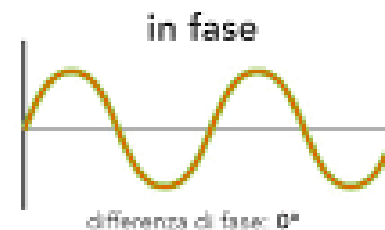
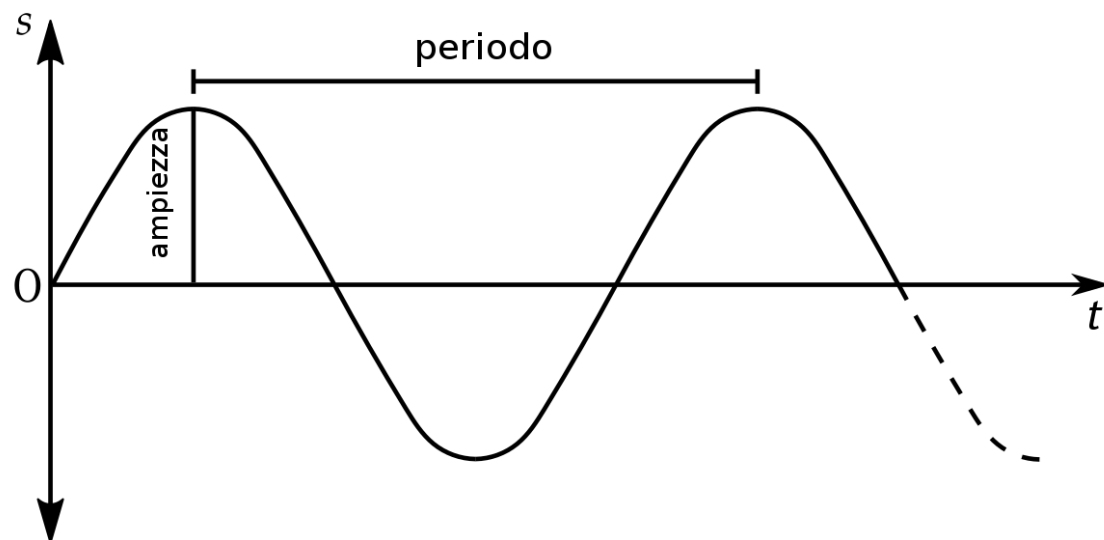
Suono: «Il prodotto di una oscillazione meccanica di un mezzo elastico».

Caratteristiche principali di un'onda

AMPIEZZA

FREQUENZA

FASE

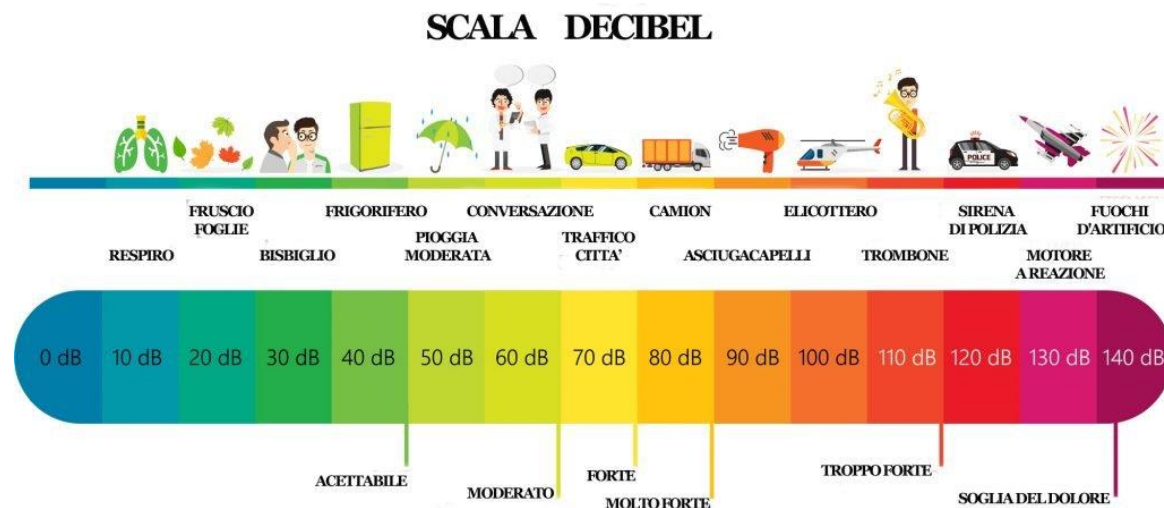
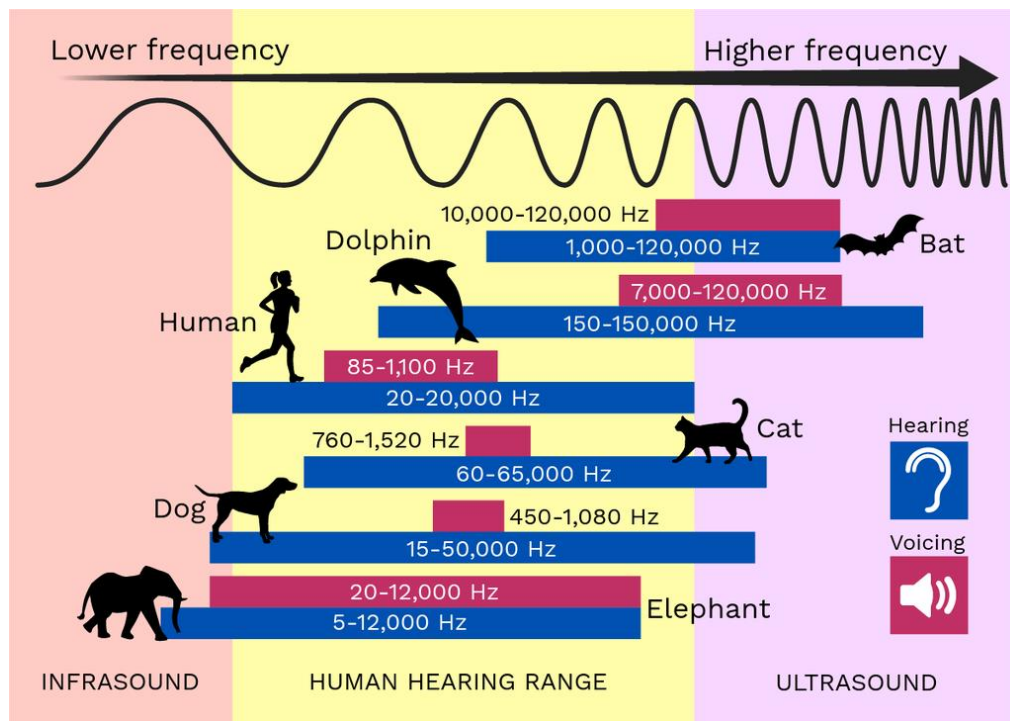


Cenni di Acustica di base

Suono: «Il prodotto di una oscillazione meccanica di un mezzo elastico».

L'orecchio umano è un sistema meccanico dinamico dotato di una sua risposta!

Diverso comportamento, non lineare, al variare dell'ampiezza e della frequenza.

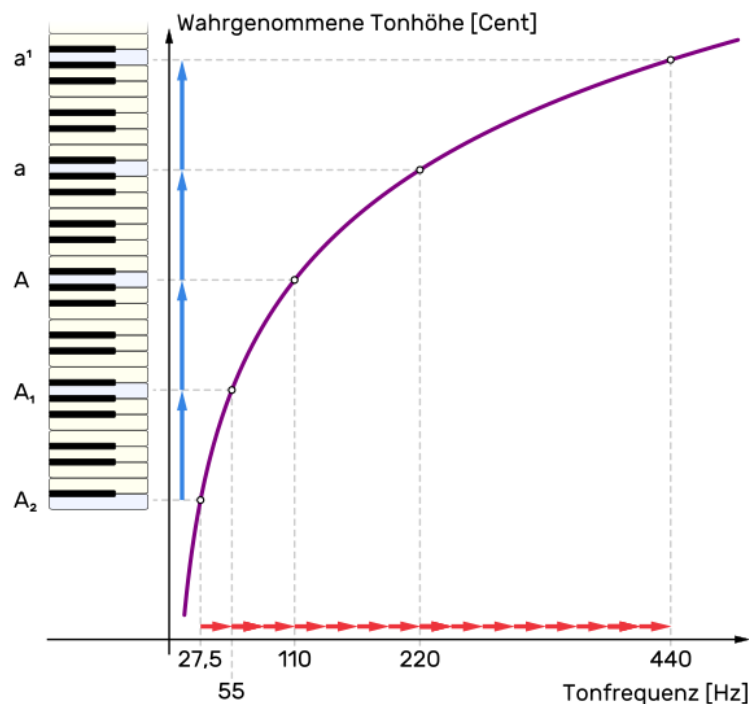


Cenni di Acustica di base

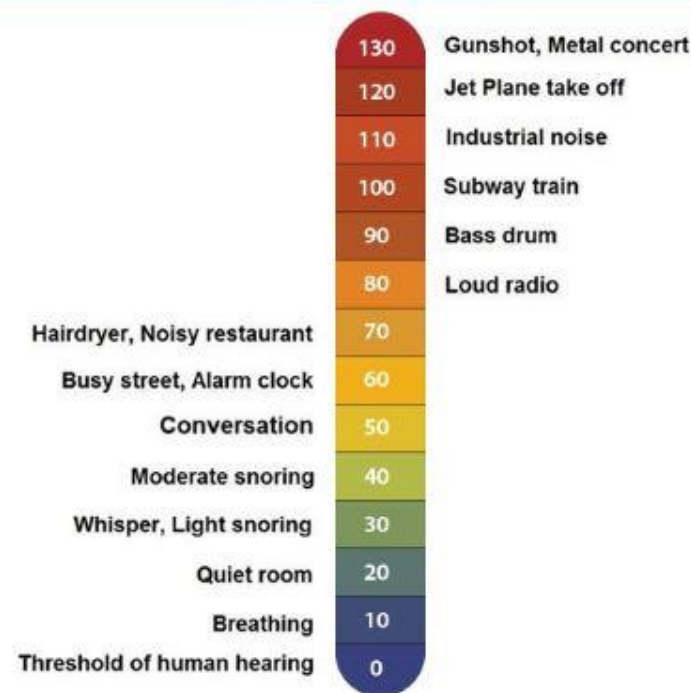
Suono: «Il prodotto di una oscillazione meccanica di un mezzo elastico».

Diverso comportamento, non lineare, al variare dell'ampiezza e della frequenza.

Percezione logaritmica della frequenza e dell'ampiezza



DECIBEL SCALE

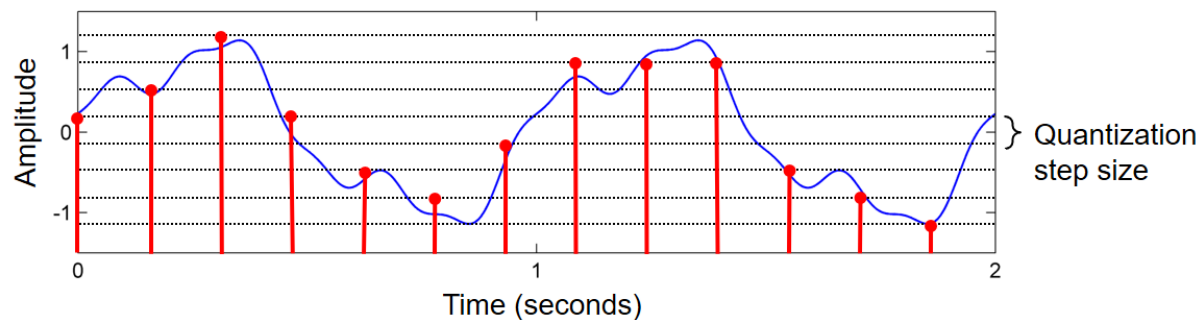
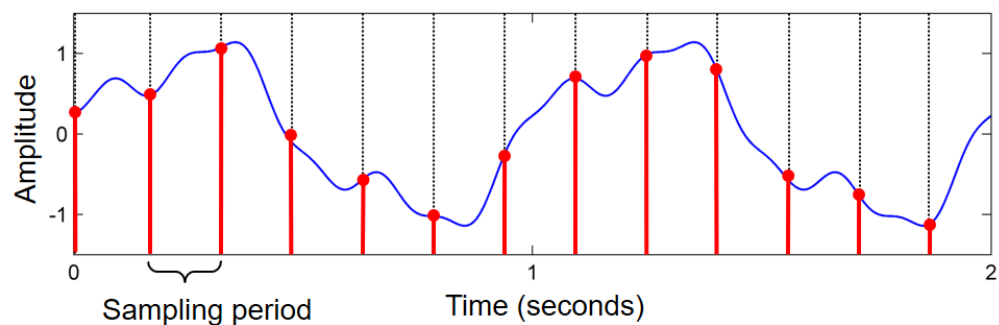


Richiami di analisi dei segnali

E' necessaria una rappresentazione del suono/
rumore che stiamo misurando

Conversione di un segnale analogico rappresentativo di
un'onda meccanica in un segnale digitale che conserva tutte
le caratteristiche «percettive» ad esso associate

Figure 2.13 from [Müller, FMP, Springer 2015]



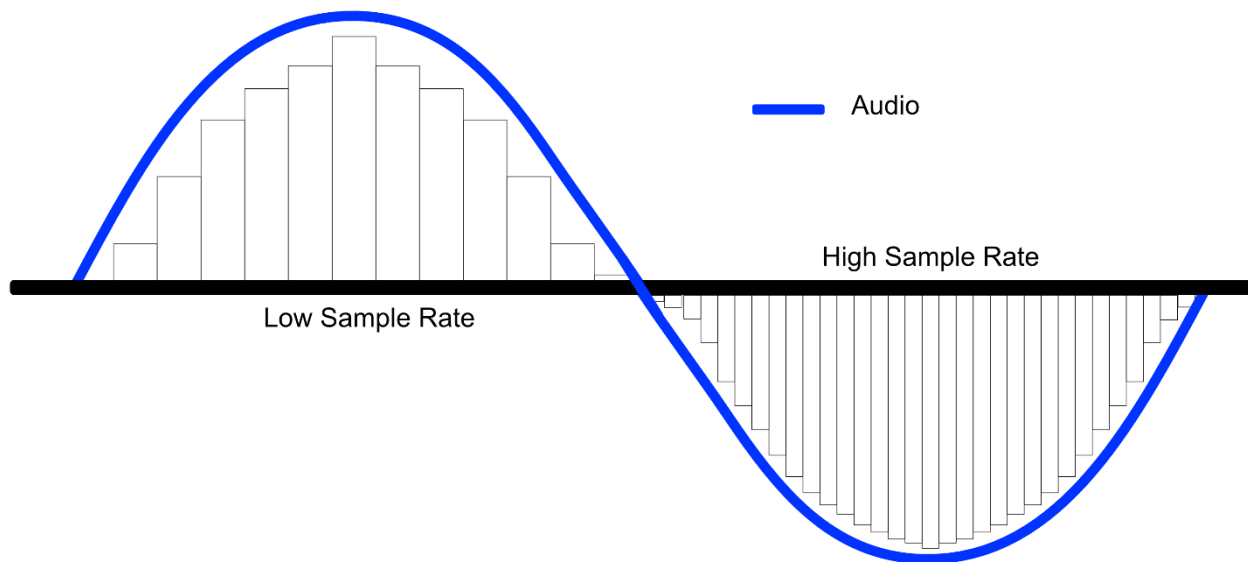
CAMPIONAMENTO O SAMPLING

QUANTIZZAZIONE

Richiami di analisi dei segnali

CAMPIONAMENTO O SAMPLING

Tecnica di discretizzazione di un segnale continuo che consiste nel valutare l'ampiezza dello stesso ad intervalli temporali o spaziali solitamente regolari



Frequenza di campionamento

$$f_s = \frac{1}{T}$$

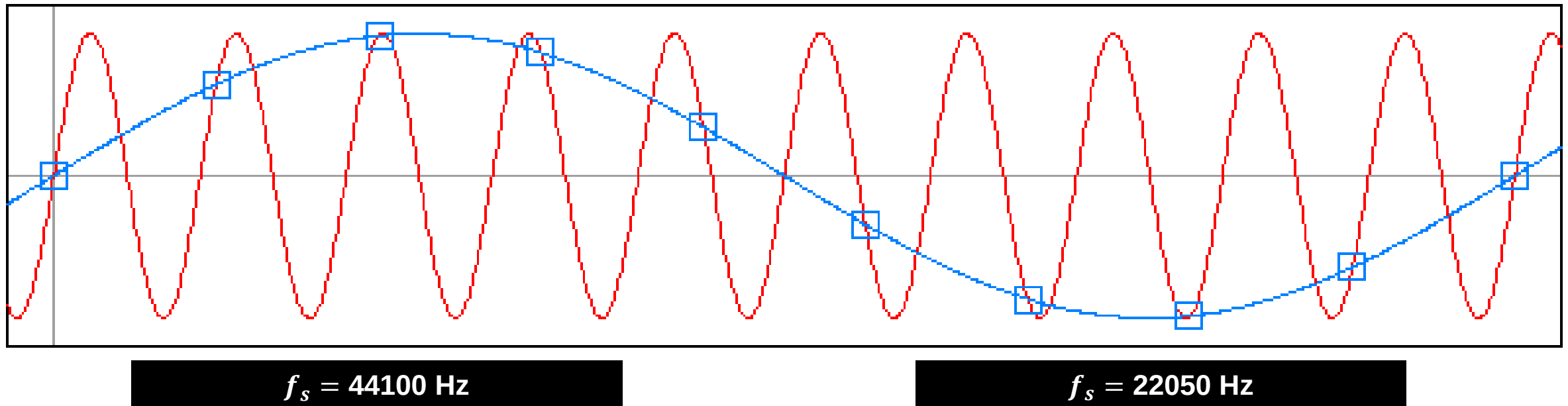
Teorema del campionamento

$$f_s > 2f_s$$

Richiami di analisi dei segnali

Ma quale è il rischio di sotto campionare un segnale?

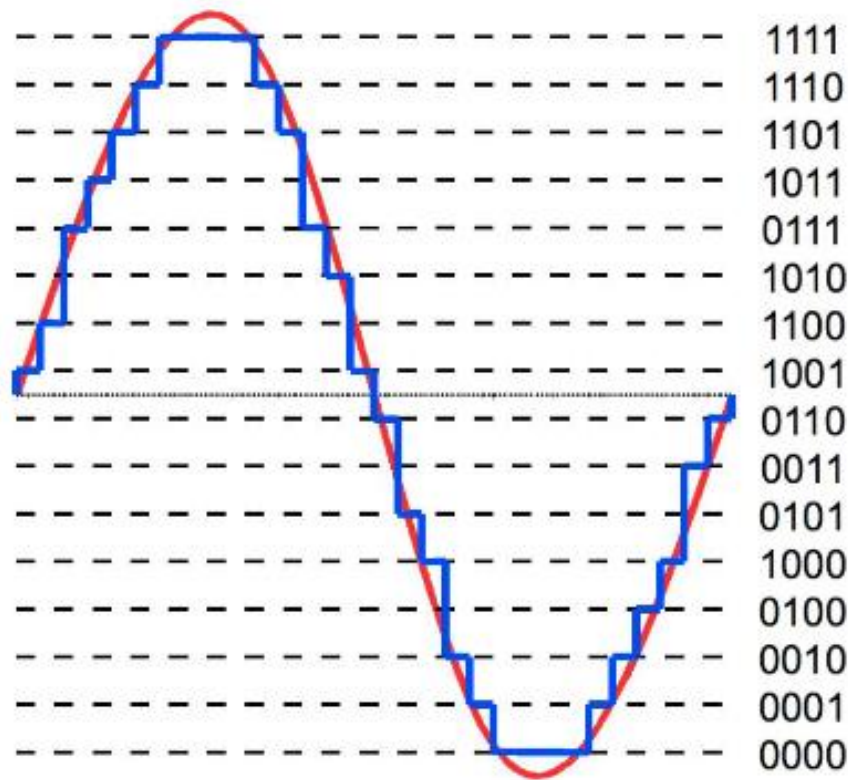
ALIASING ovvero produzione di frequenze non proprie del segnale originario oppure, nel dominio del tempo distorsione e perdita di fedeltà del segnale



Richiami di analisi dei segnali

Quantizzazione

La quantizzazione è il processo di prelievo di tali campioni e la loro conversione in un valore digitale di precisione limitata, con solo un numero fisso di valori possibili.



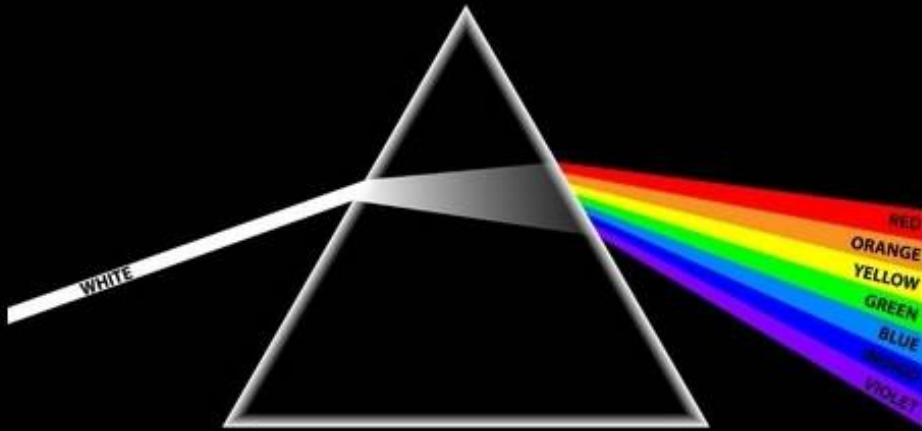
CD 16 bit

$2^{16} = 65536$ possibili valori
di ampiezza per
discretizzare il segnale

Richiami di analisi dei segnali

Necessità e possibilità di analizzare un segnale nel dominio delle frequenze piuttosto che nel solo dominio temporale

TRASFORMATTA DI FOURIER: ovvero a quell'operatore che permette di scrivere una funzione dipendente dal tempo come combinazione lineare (eventualmente continua) di funzioni di base esponenziali.



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

DFT

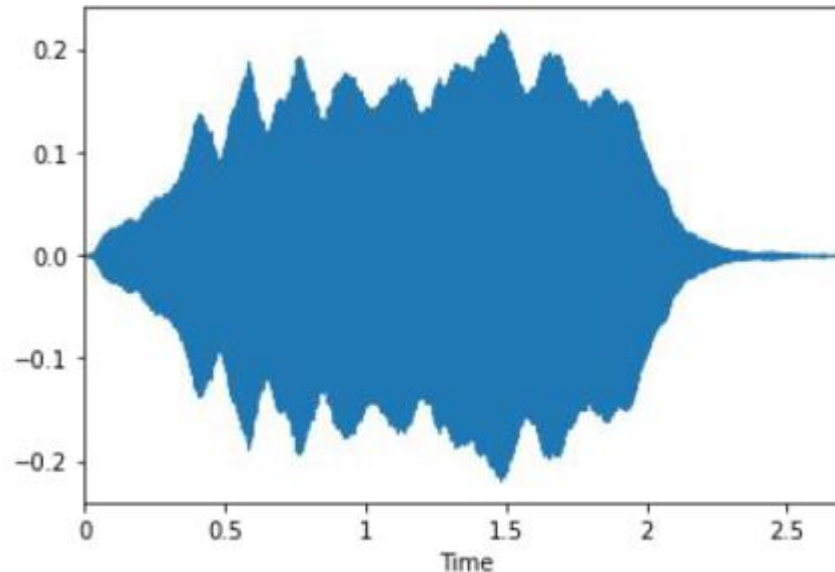
$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}$$

teropa.info/harmonics-explorer/

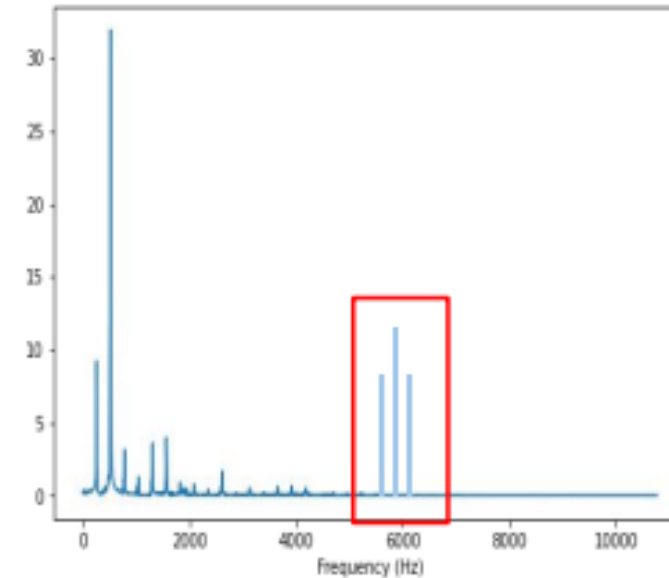
Richiami di analisi dei segnali

Stiamo considerando l'intero segnale, ma chi ci assicura che questo segnale sia formato da un numero finito di periodi?

PROBLEMA DELLO SPECTRAL LEAKAGE. Quando il segnale analizzato non è composto da un numero finito di periodi, sussistono delle discontinuità ai bordi. Una volta applicata la DFT al segnale, potrebbero risultare delle componenti di alta frequenza non fisiche



DFT



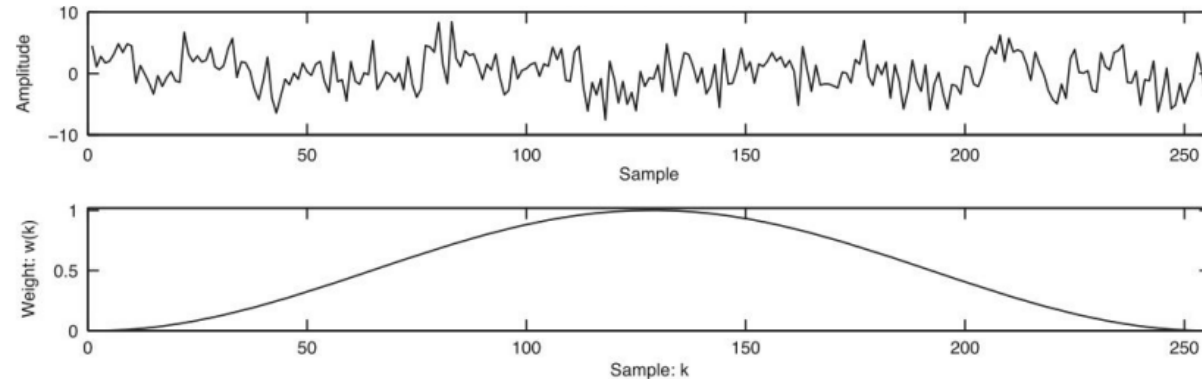
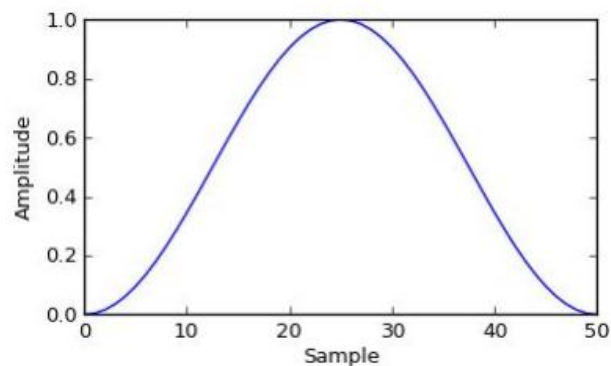
teropa.info/harmonics-explorer/

Richiami di analisi dei segnali

ELIMINARE LO SPECTRAL LEAKAGE

FINISTRATURE DEL SEGNALE: Per ovviare a questo problema, sono in uso delle tecniche di finestatura del segnale che permettono di eliminare le sezioni estremali dello stesso o di alcune sue parti per riprodurre la condizione di continuità ai bordi.

$$w(k) = 0.5 \cdot \left(1 - \cos\left(\frac{2\pi k}{K-1}\right)\right), k = 1 \dots K$$

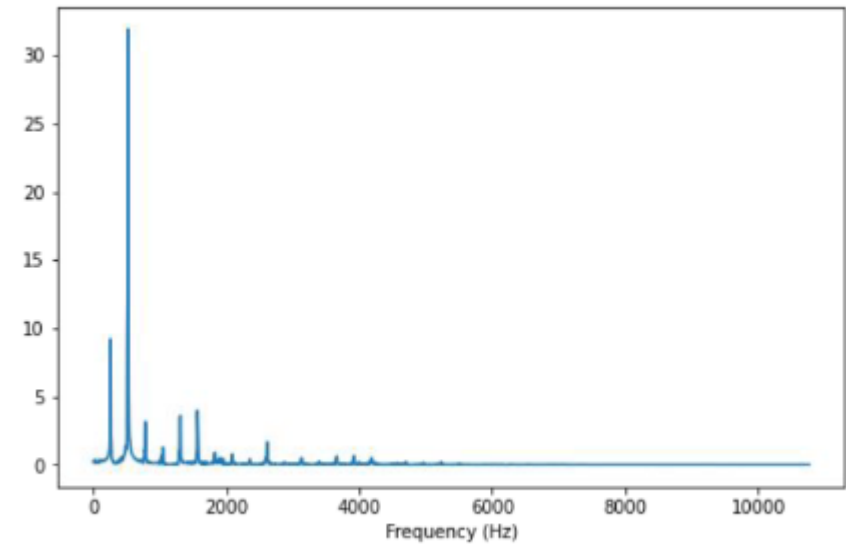
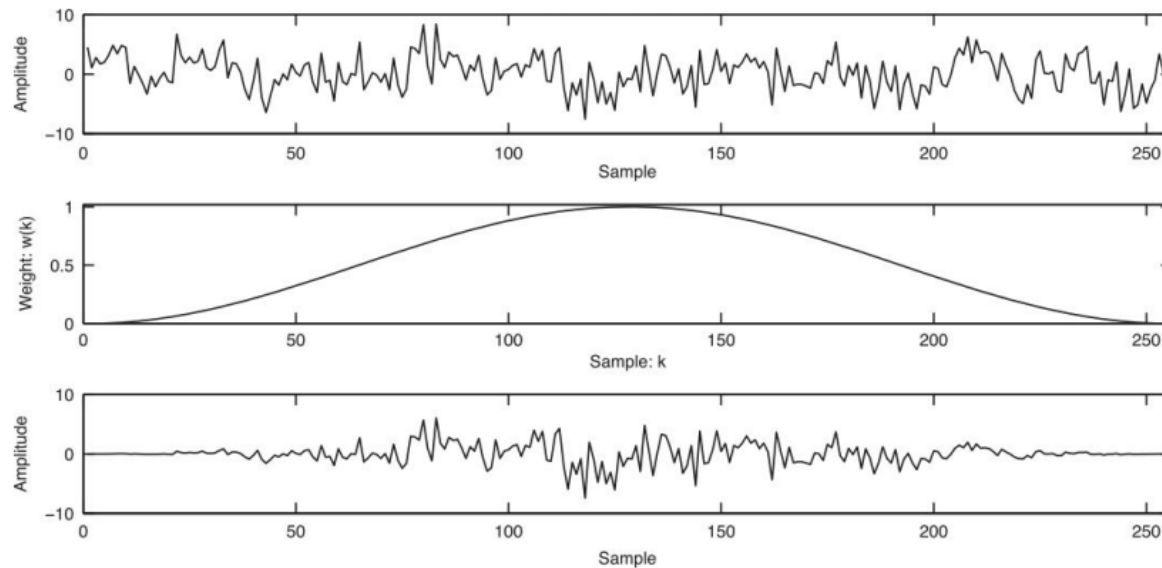


$$s_w(k) = s(k) \cdot w(k), k = 1 \dots K$$

Richiami di analisi dei segnali

Segnale continuo ai bordi

Eliminazione dello spectral leakage



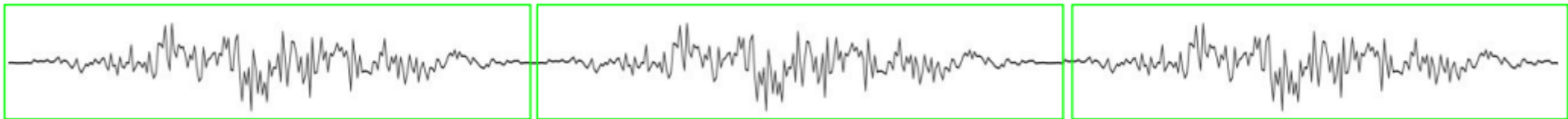
Richiami di analisi dei segnali

E' possibile applicare la DFT a porzioni definite di segnale?

SHORT TIME FOURIER TRANSFORMATION

$$S(m, k) = \sum_{n=0}^{N-1} x(n + mH) \cdot w(n) \cdot e^{-i2\pi n \frac{k}{N}}$$

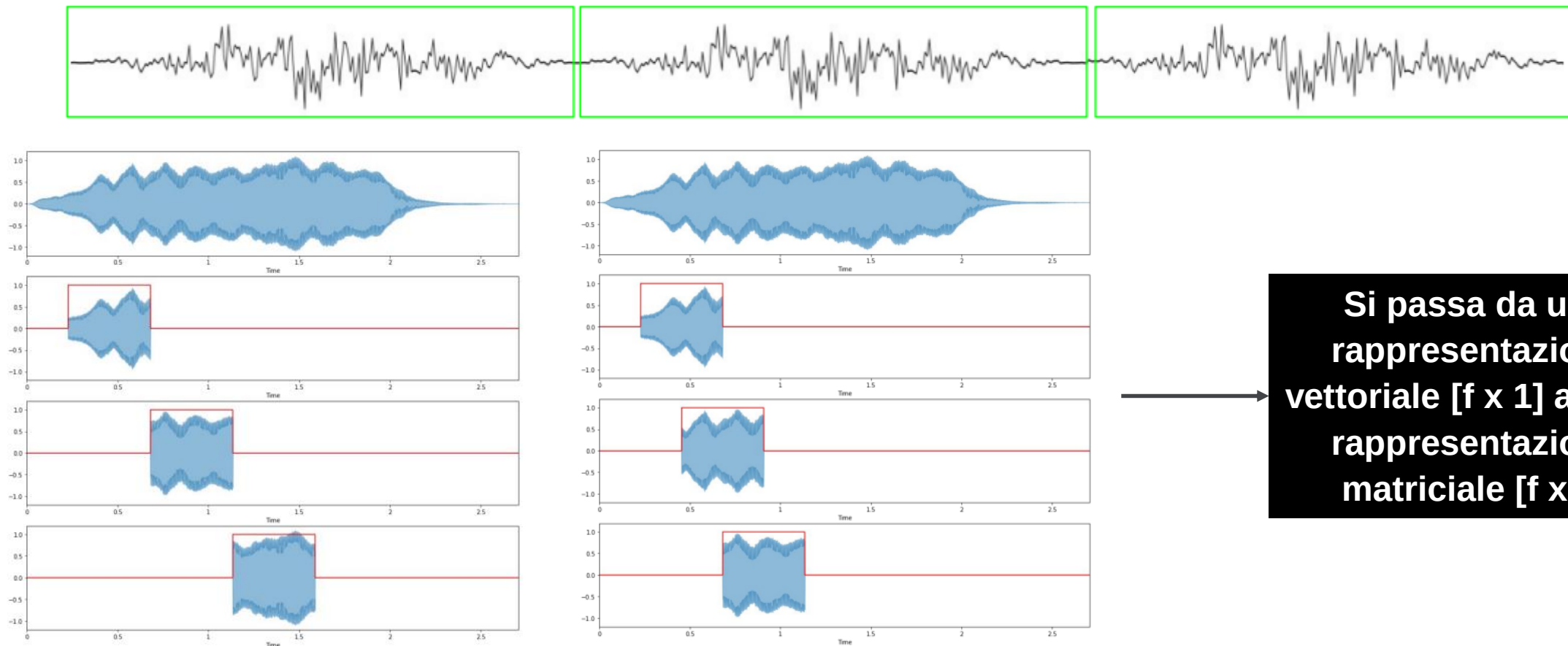
Fa riferimento al m-esimo frame di lunghezza $n+mH$ dove mH è il sample di inizio del frame ed n è ciascun sample presente all'interno del frame



Richiami di analisi dei segnali

E' possibile applicare la DFT a porzioni definite di segnale?

SHORT TIME FOURIER TRANSFORMATION

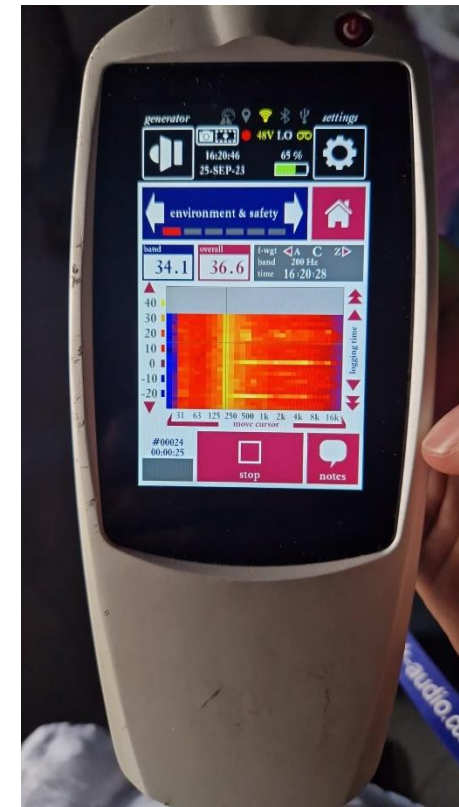
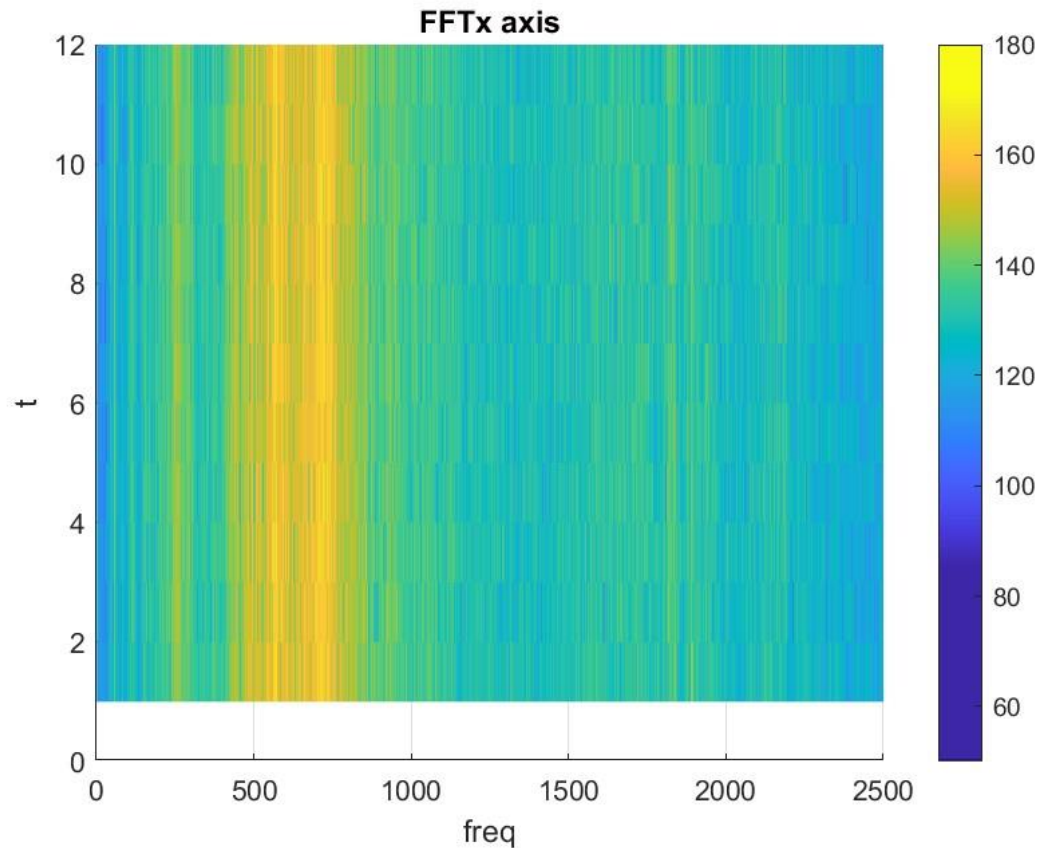


Si passa da una
rappresentazione
vettoriale $[f \times 1]$ ad una
rappresentazione
matriciale $[f \times n]$

Richiami di analisi dei segnali

SHORT TIME FOURIER TRANSFORMATION

HO UN'INDICAZIONE SIA NEL DOMINIO DEL TEMPO CHE DELLE FREQUENZE



Cosa porta al riconoscimento di un suono?

Come riconosciamo un suono?

Abbiamo bisogno di individuare alcune delle caratteristiche primarie che portano al riconoscimento di un suono

Intensità sonora

Loudness

Timbro



Cosa porta al riconoscimento di un suono?

Intensità sonora

L'intensità rappresenta la potenza trasferita attraverso un'unità d'area.

Potenza sonora [W]

Intensità sonora [W/m²]



Cosa porta al riconoscimento di un suono?

Intensità sonora

Abbiamo bisogno di individuare alcune delle caratteristiche primarie che portano al riconoscimento di un suono

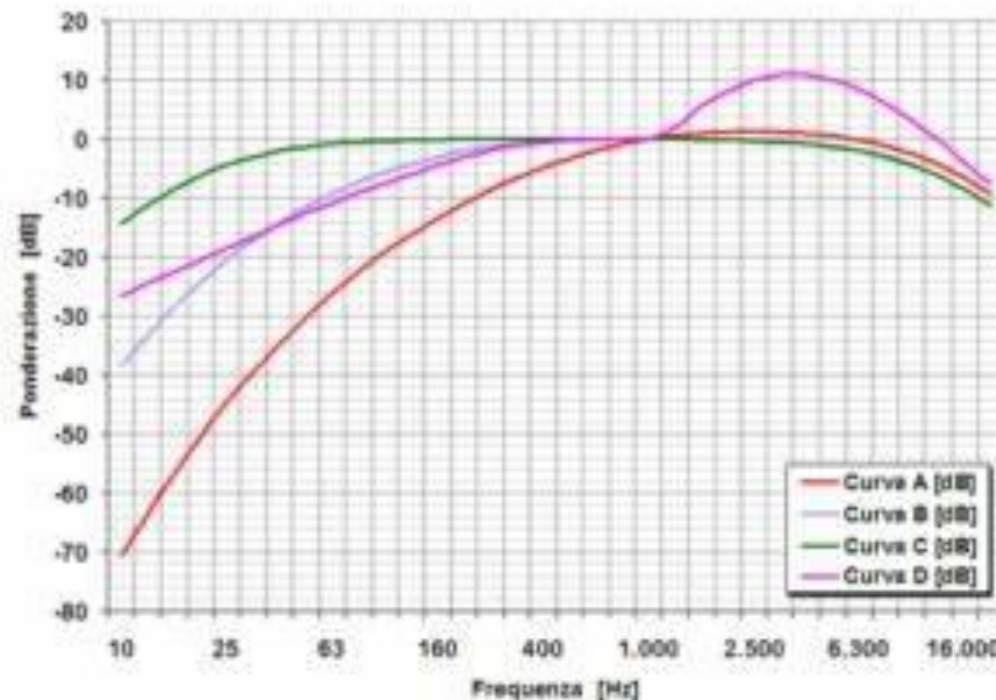
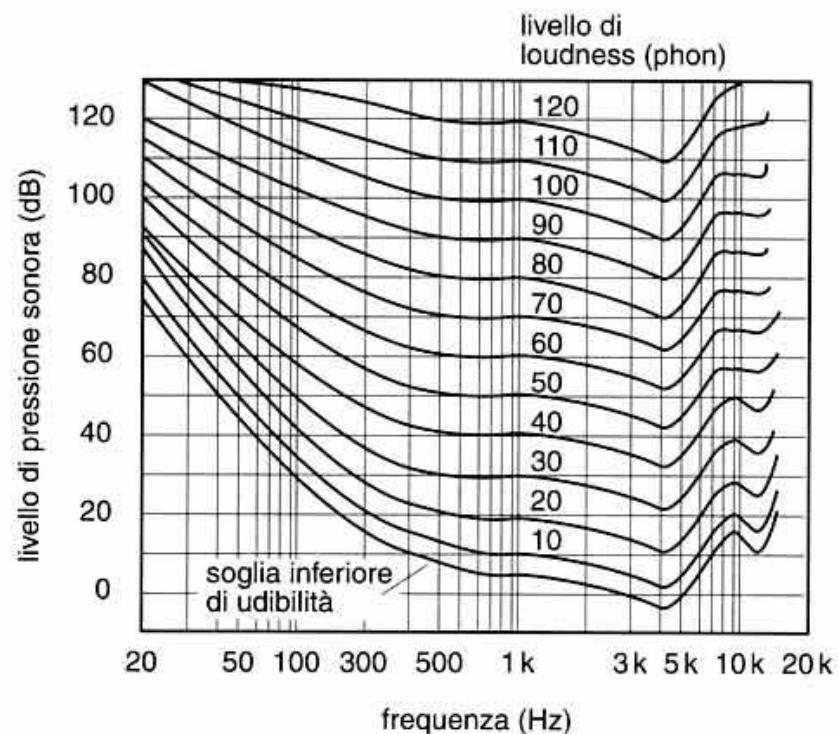
Source	Intensity	Intensity level	× TOH
Threshold of hearing (TOH)	10^{-12}	0 dB	1
Whisper	10^{-10}	20 dB	10^2
Pianissimo	10^{-8}	40 dB	10^4
Normal conversation	10^{-6}	60 dB	10^6
Fortissimo	10^{-2}	100 dB	10^{10}
Threshold of pain	10	130 dB	10^{13}
Jet take-off	10^2	140 dB	10^{14}
Instant perforation of eardrum	10^4	160 dB	10^{16}

Table 1.1 from [Müller, FMP, Springer 2015]

Cosa porta al riconoscimento di un suono?

LOUDNESS

La Loudness è un parametro percettivo! Si misura in phon il quale definisce il livello di pressione sonora che il suono di riferimento a 1kHz deve avere al fine di provare la medesima sensazione uditiva di un suono alla generica frequenza f

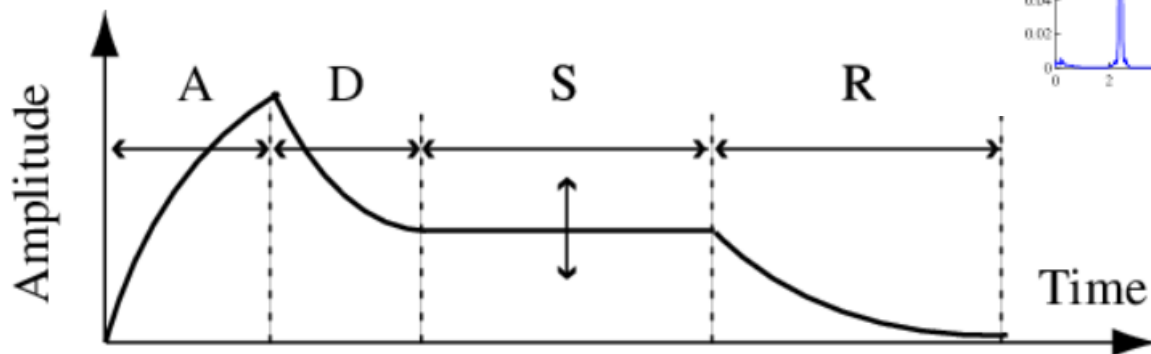


Cosa porta al riconoscimento di un suono?

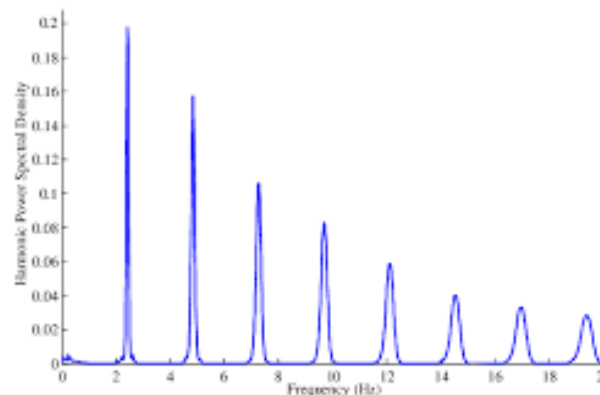
TIMBRO

Non esiste una definizione univoca del concetto di TIMBRO, potrebbe essere descritto come il colore del suono?

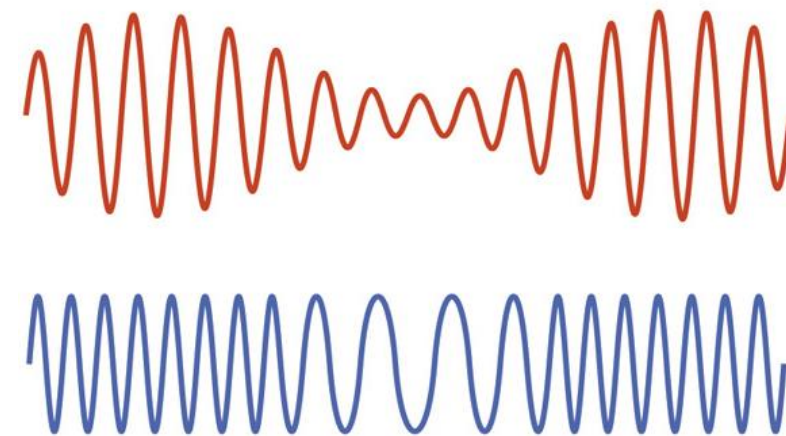
Inviluppo sonoro



Contenuto Armonico



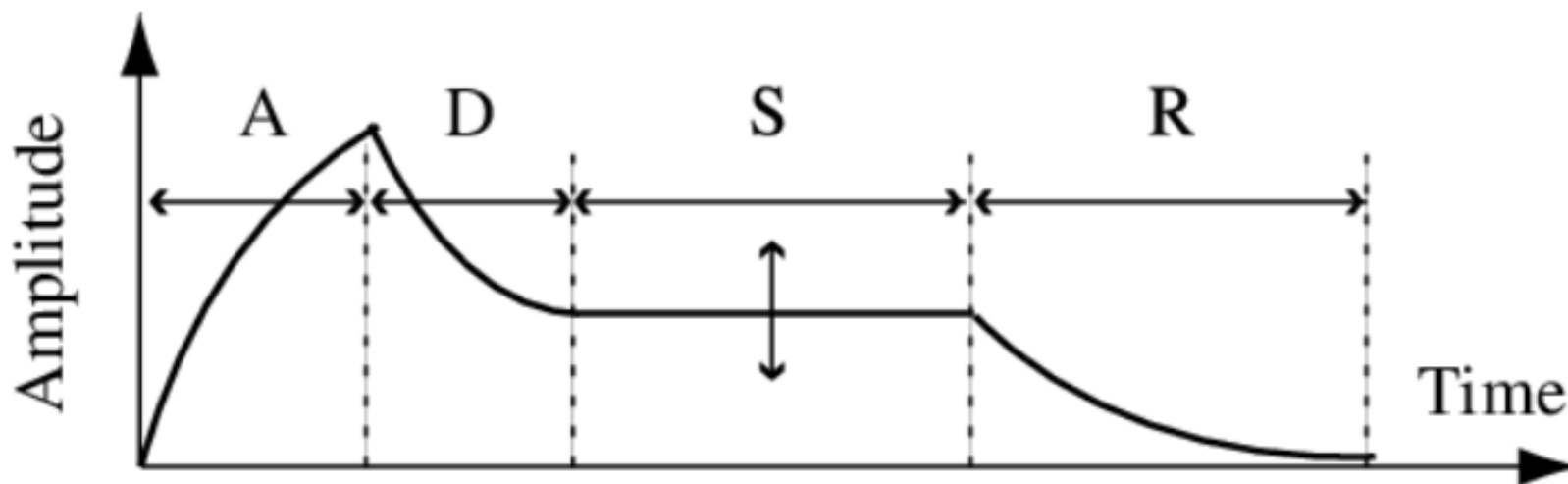
Modulazione in ampiezza e frequenza



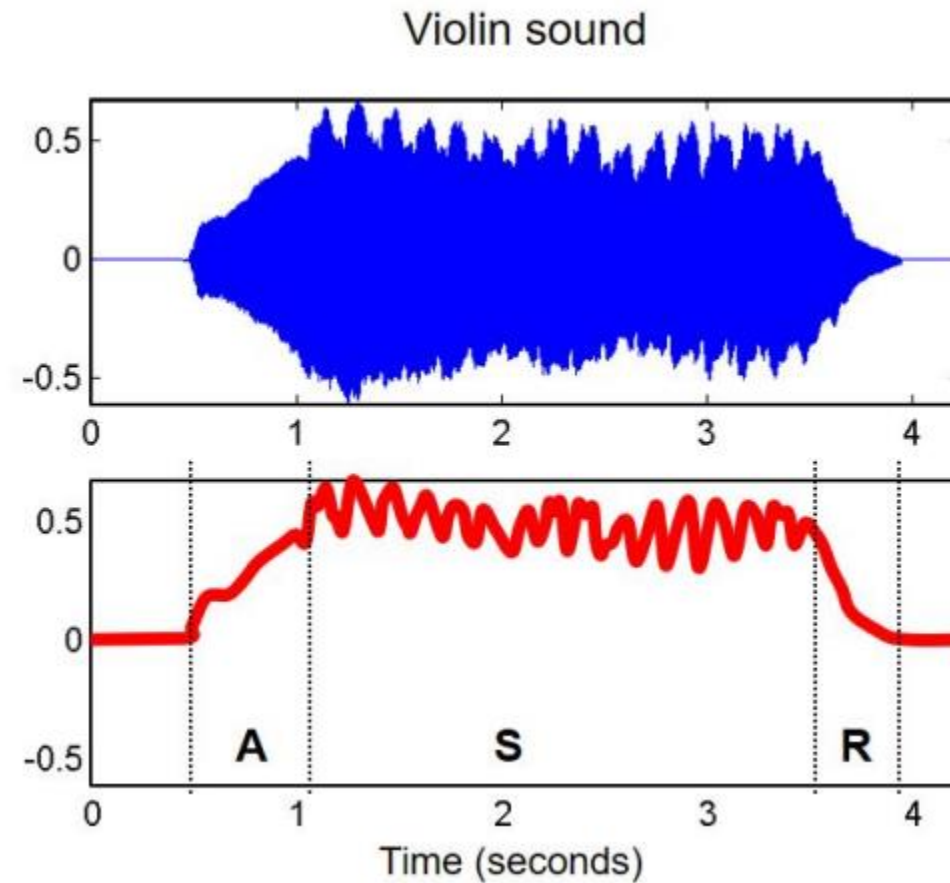
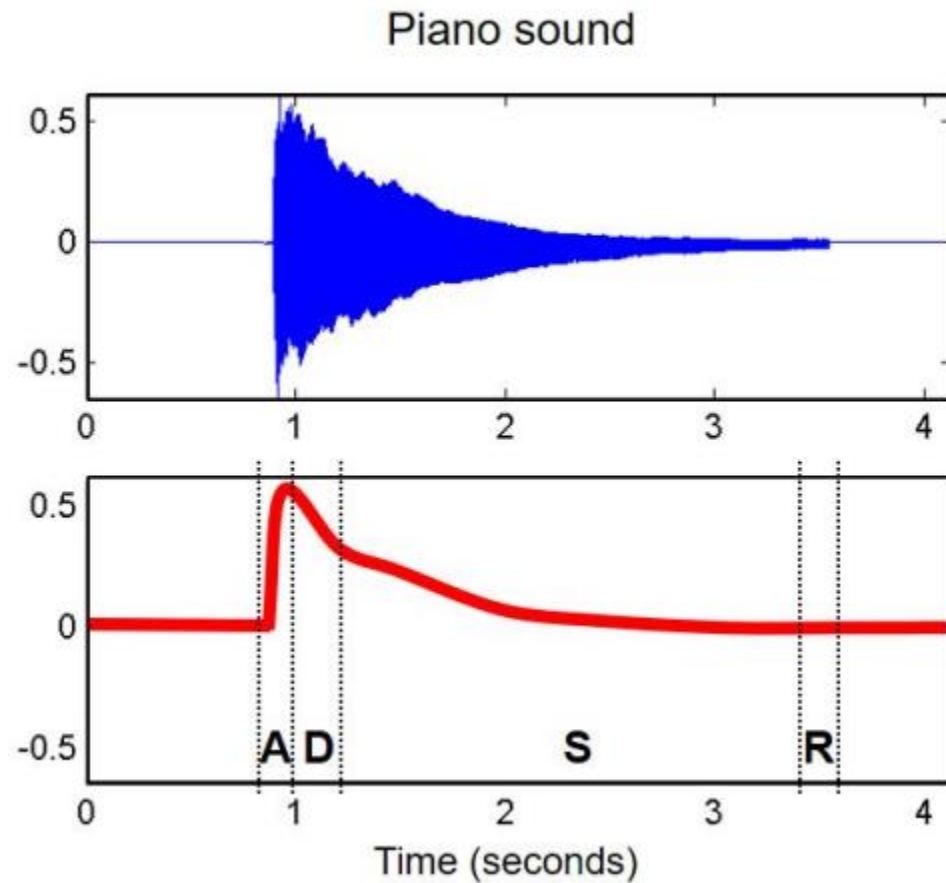
Cosa porta al riconoscimento di un suono?

INVILUPPO SONORO

Concentriamoci ora sull'inviluppo sonoro. E' esperienza comune che il suono, ad esempio, di uno strumento musicale ha un proprio inviluppo caratteristico che varia in relazione allo strumento utilizzato. L'inviluppo sonoro può essere rappresentato attraverso un modello che prende il nome di ADSR ovvero Attach, Decay, Sustain and Release.



Cosa porta al riconoscimento di un suono?



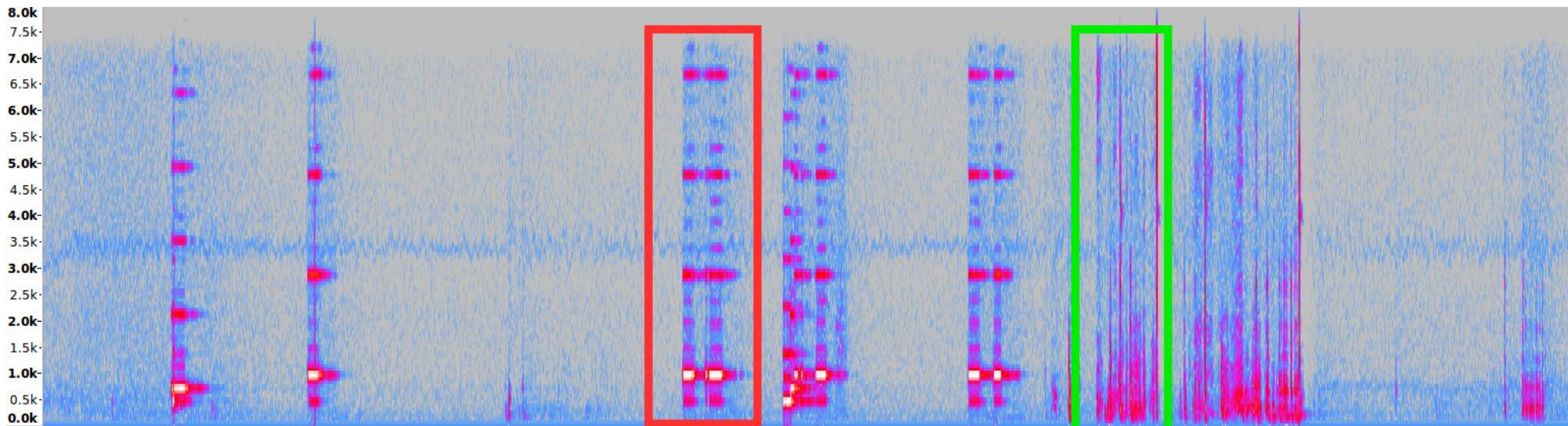
Cosa porta al riconoscimento di un suono?



Cosa porta al riconoscimento di un suono?

CONTENUTO ARMONICO

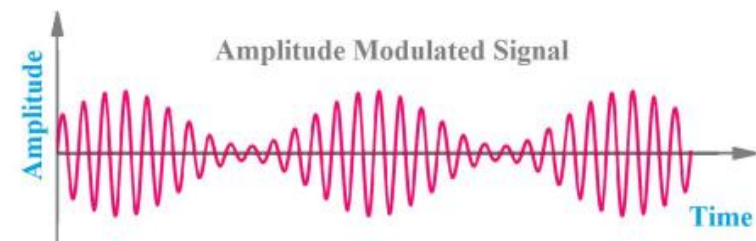
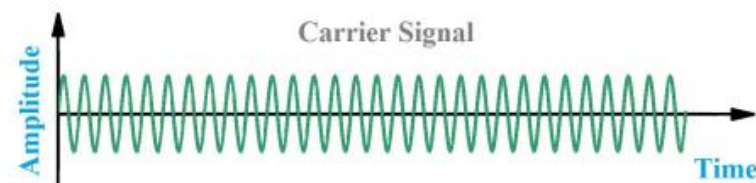
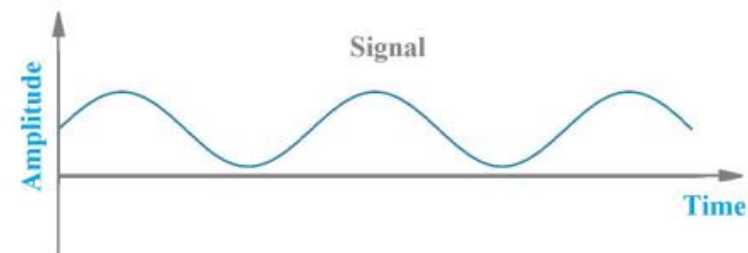
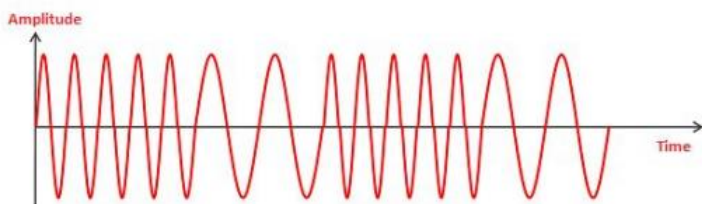
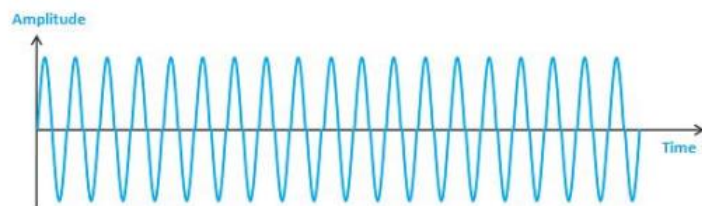
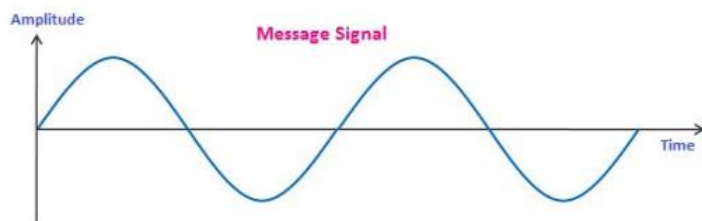
Un suono complesso è dato da una sovrapposizione di più sinusoidi semplici. Un *parziale* è una senoide che permette di descrivere un suono. Un suono armonico è un suono che ha frequenze multiple di una frequenza fondamentale



Cosa porta al riconoscimento di un suono?

MODULAZIONE IN FREQUENZA ED AMPIEZZA

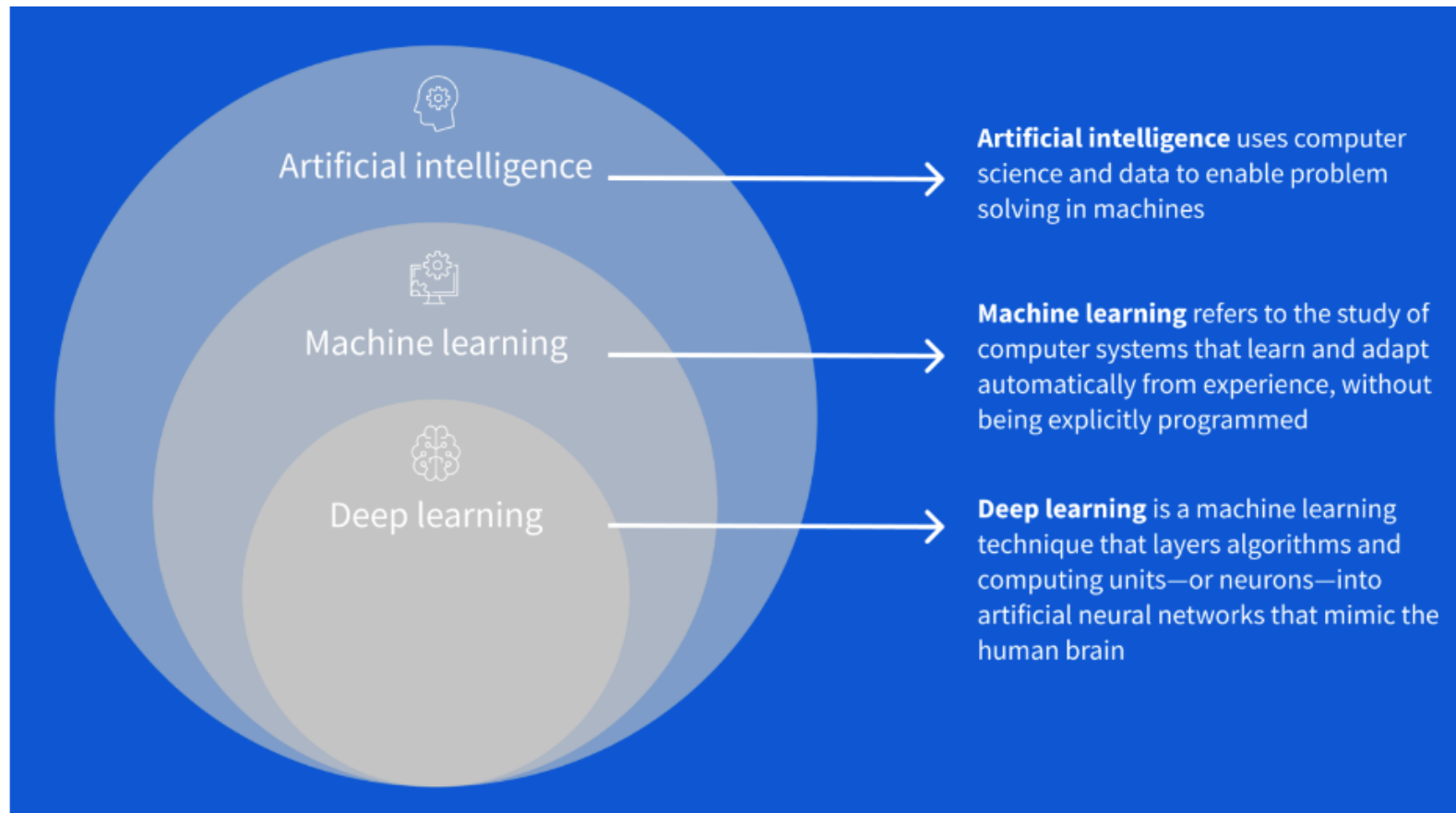
In questo caso si sta introducendo la possibilità di modulare un segnale portante in ampiezza o in frequenza. Da un punto di vista applicativo, queste due tecniche sono molto utilizzate in applicazioni musicali quali ad esempio vibrato e tremolo. Nel primo caso, si sta facendo riferimento ad una variazione periodica della frequenza, nel secondo caso, ad una variazione periodica dell'ampiezza



IA, Machine Learning e Deep Learning

INTELLIGENZA ARTIFICIALE

IA: «Disciplina che studia se e in che modo si possano riprodurre i processi mentali più complessi mediante l'uso di un computer»



IA, Machine Learning e Deep Learning

Quali sono le principali differenze?

Appare evidente che il Deep Learning si strutturi come un subset del Machine Learning, ma quali sono le principali differenze?

Machine learning	Deep learning
A subset of AI	A subset of machine learning
Can train on smaller data sets	Requires large amounts of data
Requires more human intervention to correct and learn	Learns on its own from environment and past mistakes
Shorter training and lower accuracy	Longer training and higher accuracy
Makes simple, linear correlations	Makes non-linear, complex correlations
Can train on a CPU (central processing unit)	Needs a specialized GPU (graphics processing unit) to train

IA, Machine Learning e Deep Learning

Riconoscimento di sorgenti

Nel caso specifico del riconoscimento di sorgenti o eventi di tipo acustico, la scelta ricade sul Deep Learning.

CNN ARCHITECTURES FOR LARGE-SCALE AUDIO CLASSIFICATION

Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, Kevin Wilson

Google, Inc., New York, NY, and Mountain View, CA, USA

shershey@google.com

ABSTRACT

Convolutional Neural Networks (CNNs) have proven very effective in image classification and show promise for audio. We use various CNN architectures to classify the soundtracks of a dataset of 70M training videos (5.24 million hours) with 30,871 video-level labels. We examine fully connected Deep Neural Networks (DNNs), AlexNet [1], VGG [2], Inception [3], and ResNet [4]. We investigate varying the size of both training set and label vocabulary, finding that analogs of the CNNs used in image classification do well on our audio classification task, and larger training and label sets help up to a point. A model using embeddings from these classifiers does much better than raw features on the *Audio Set* [5] Acoustic Event Detection (AED) classification task.

CNNs si sono rivelate molto efficaci nell'ambito del riconoscimento di immagini e che si rivelano molto promettenti per il riconoscimento di audio

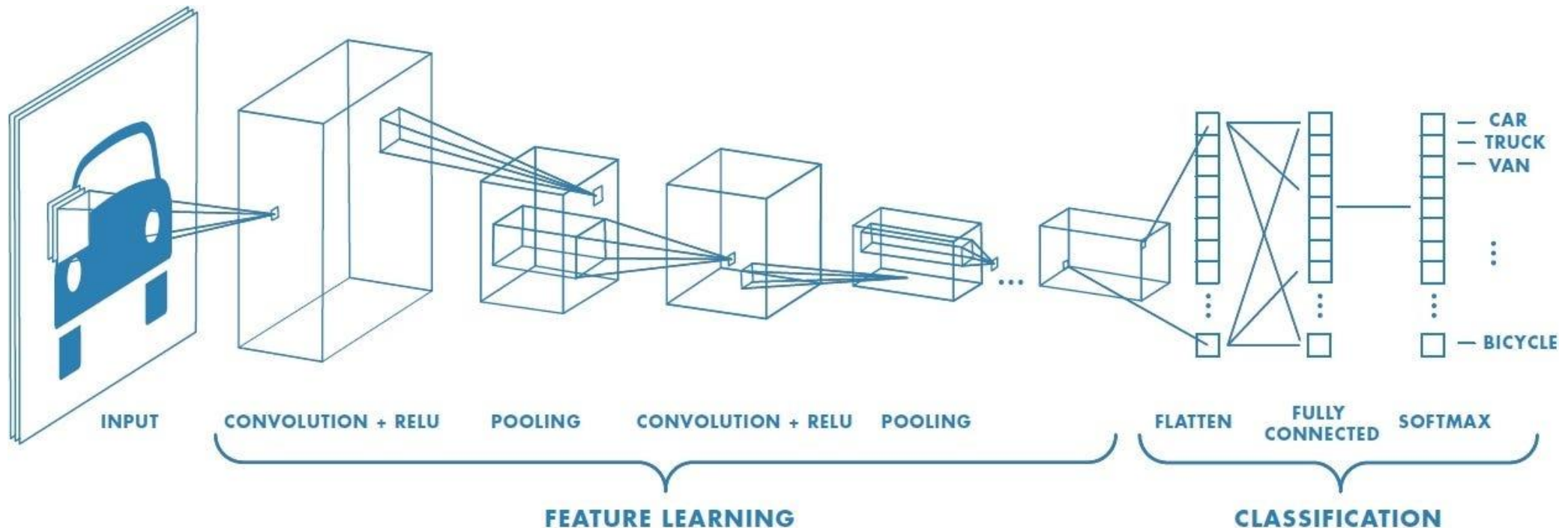
Il dataset utilizzato per addestrare la rete è stato ricavato da video

Il modello basato su questi concetti si è dimostrato migliore rispetto a quello derivante dall'estrazione delle features.

IA, Machine Learning e Deep Learning

Le CNNs si sono rivelate molto efficaci nell'ambito del riconoscimento di immagini e che si rivelano molto promettenti per il riconoscimento di audio

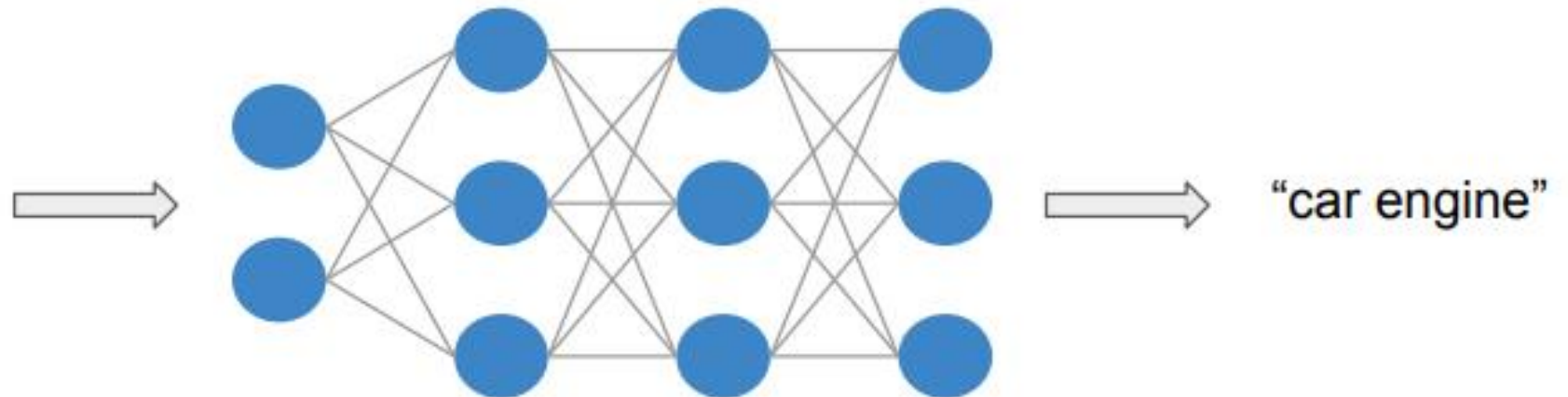
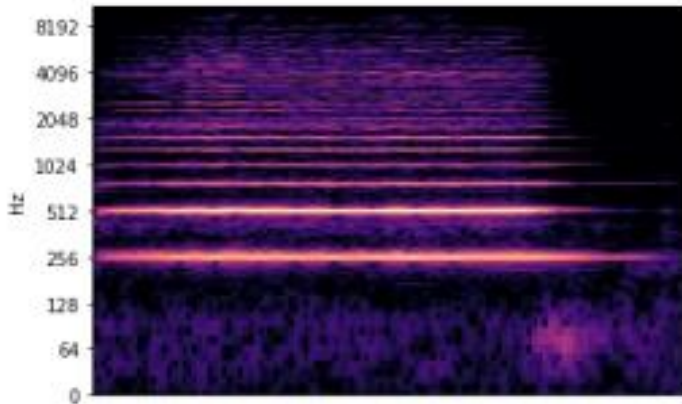
- Cosa è una CNN?
- Quale è il collegamento tra riconoscimento di immagini e di audio?



IA, Machine Learning e Deep Learning

Le CNNs si sono rivelate molto efficaci nell'ambito del riconoscimento di immagini e che si rivelano molto promettenti per il riconoscimento di audio

- Cosa è una CNN?
- Quale è il collegamento tra riconoscimento di immagini e di audio?



IA, Machine Learning e Deep Learning

Riconoscimento di sorgenti

Nel caso specifico del riconoscimento di sorgenti o eventi di tipo acustico, la scelta ricade sul Deep Learning.

CNN ARCHITECTURES FOR LARGE-SCALE AUDIO CLASSIFICATION

Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, Kevin Wilson

Google, Inc., New York, NY, and Mountain View, CA, USA

shershey@google.com

5. CONCLUSIONS

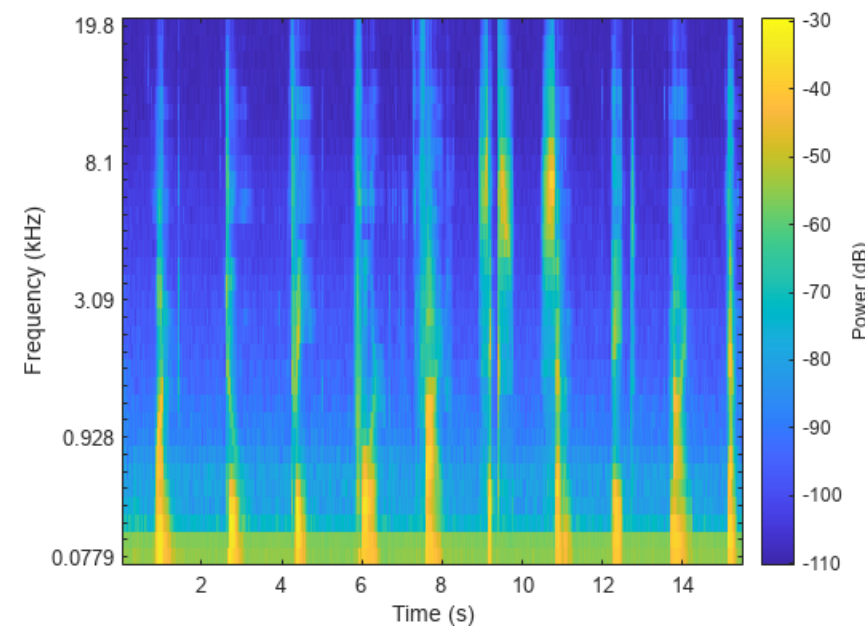
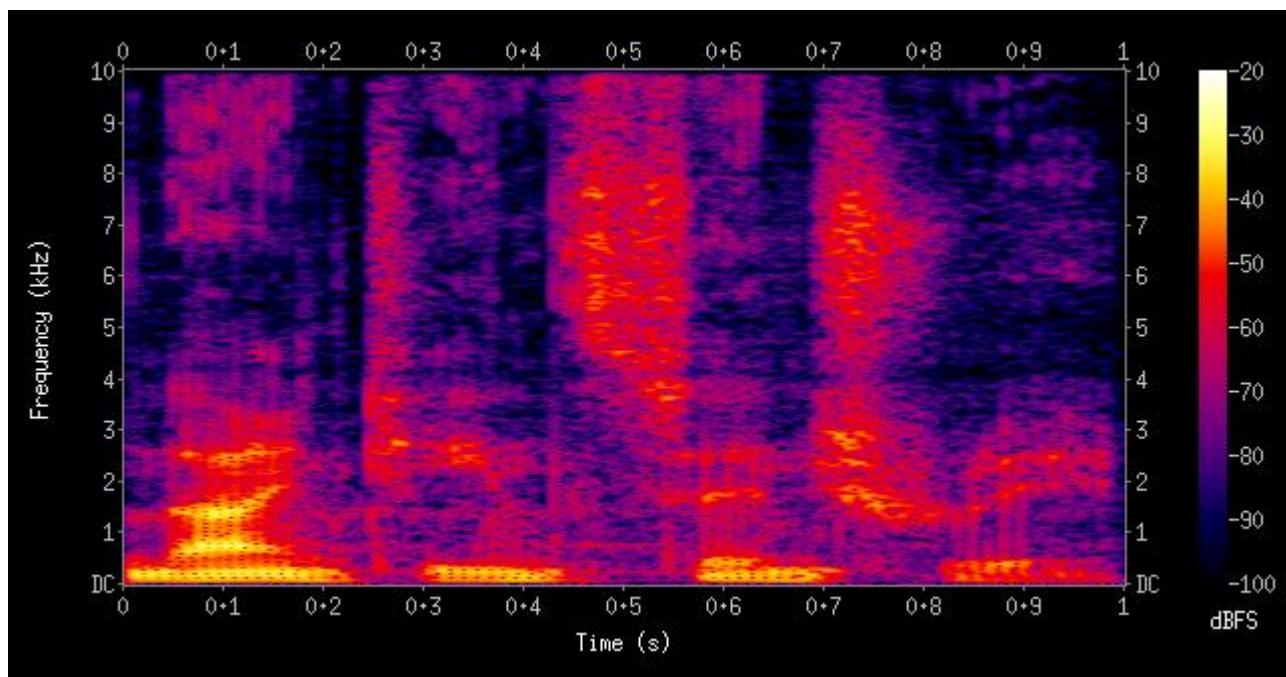
The results in Section 4.1 show that state-of-the-art image networks are capable of excellent results on audio classification when compared to a simple fully connected network or earlier image classification architectures. In Section 4.2 we saw results showing that training on larger label set vocabularies can improve performance, albeit modestly, when evaluating on smaller label sets. In Section 4.3 we saw that increasing the number of videos up to 7M improves performance for the best-performing ResNet-50 architecture. We note that regularization could have reduced the gap between the models trained on smaller datasets and the 7M and 70M datasets. In Section 4.4 we see a significant increase over our baseline when training a model for AED with ResNet embeddings on the *Audio Set* dataset.

In addition to these quantified results, we can subjectively examine the performance of the model on segments of video. Fig. 2 shows the results of running our best classifier over a video and overlaying the frame-by-frame results of the 16 classifier outputs with the greatest peak values across the entire video. The different sound sources present at different points in the video are clearly distinguished. ²

Estrazione di features: lo spettrogramma di Mel

Estrazione di features non è più strettamente necessaria, ma non è detto che sia deleteria!

Abbiamo visto come alcune reti originariamente create per il riconoscimento di immagini possano essere impiegate con successo per identificare eventi o sorgenti rumorose. La feature che risulterà maggiormente utile ai nostri scopi è lo spettrogramma Mel.



Estrazione di features: lo spettrogramma di Mel

Un piccolo esperimento di psicoacustica

Come cambia la percezione al variare della frequenza? Consideriamo due coppie di toni puri e valutiamo quale di queste coppie sia percepita più distintamente in termini di differenza tra i suoni. Si noti che 1000 Hz corrispondono a 1000 Hz mel e che distanze uguali sul grafico riportano a caratteristiche percettive uguali

65 Hz



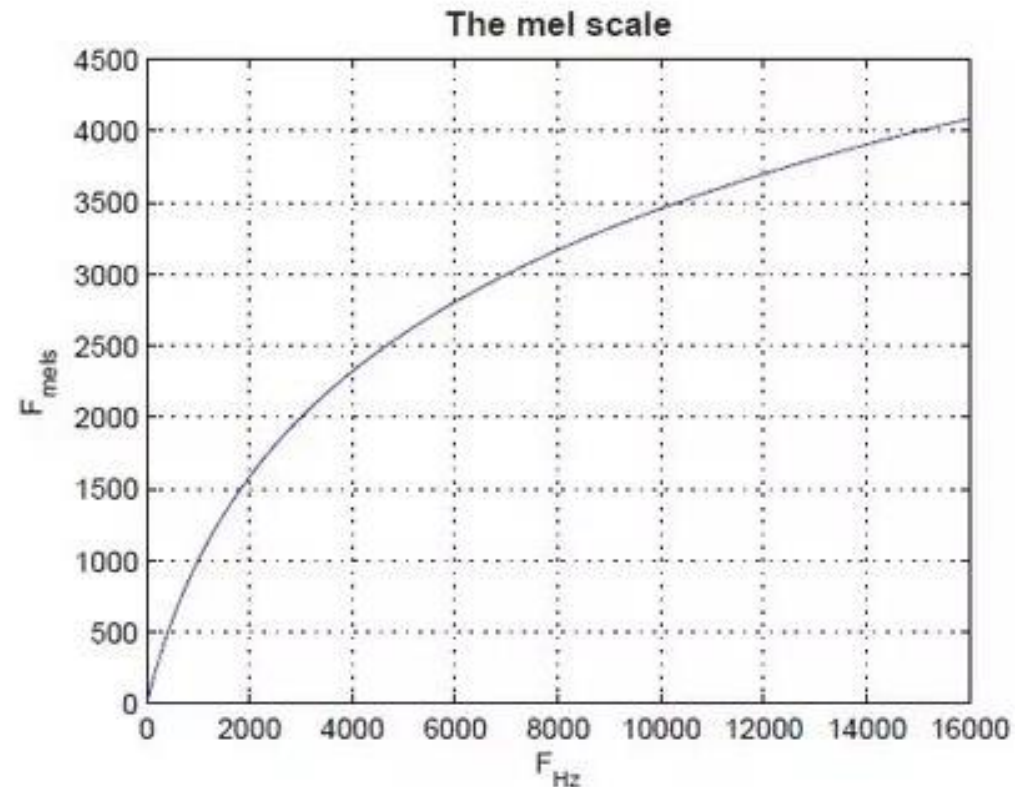
262 Hz



1568 Hz



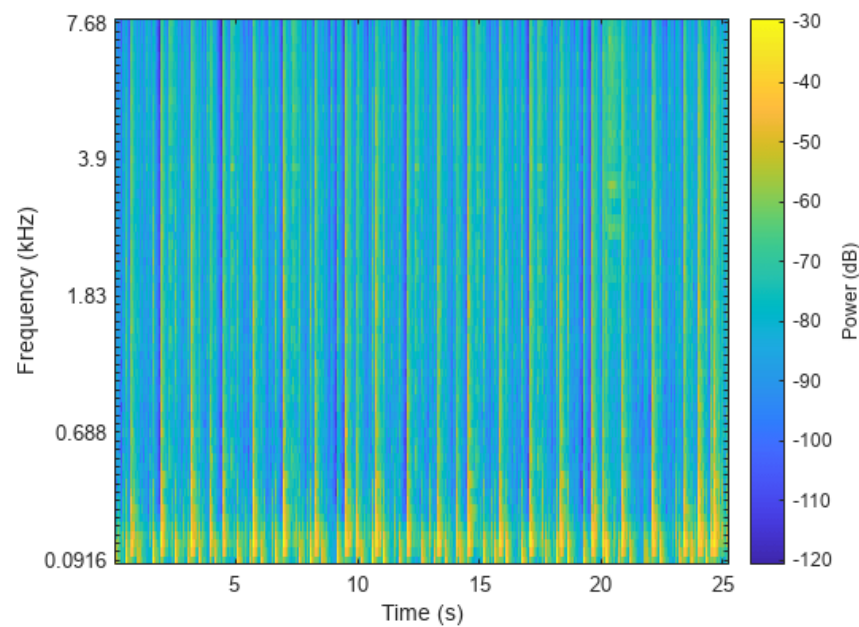
1760 Hz



Estrazione di features: lo spettrogramma di Mel

Come mai utilizziamo come feature lo spettrogramma Mel e non un classico spettrogramma

Partiamo con il definire il concetto di feature ovvero una proprietà individuale misurabile o caratteristica di un fenomeno. Cosa richiediamo dalle nostre features?



Rappresentazione nei domini del tempo e della frequenza

Vogliamo una feature che sia rilevante da un punto di vista percettivo in termini di ampiezza

Vogliamo una feature che sia rilevante da un punto di vista percettivo in termini di frequenza

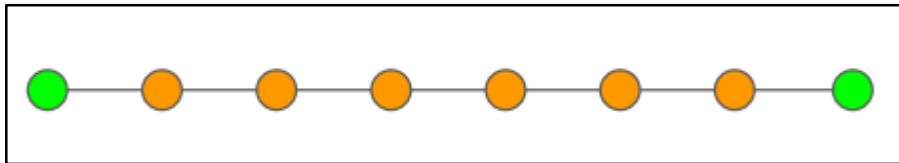
Estrazione di features: lo spettrogramma di Mel

Come ricavare uno spettrogramma Mel

Per ricavare uno spettrogramma Mel è opportuno costruire prima un banco filtri da poter poi applicare allo spettrogramma classico per tener conto di quanto illustrato in precedenza. Procediamo per step.

$$m = 2595 \cdot \log\left(1 + \frac{f}{500}\right)$$

Convertire in scala mel la minore e la maggiore frequenza considerata



Creare un numero equispaziato di bande comprese tra la frequenza massima e minima convertita in mel

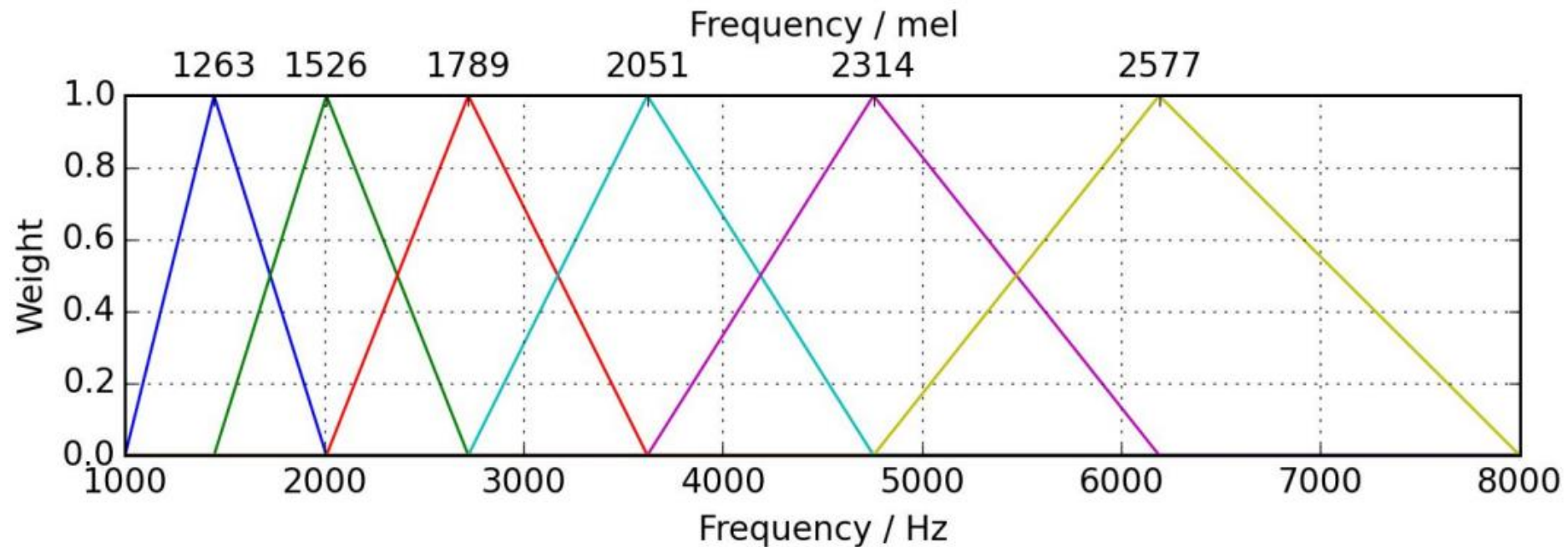
$$f = 700(10^{m/2595} - 1)$$

Riconvertire i centri banda ottenuti in Hz e arrotonda alla frequenza intera più prossima

Estrazione di features: lo spettrogramma di Mel

Come ricavare uno spettrogramma Mel

Per ricavare uno spettrogramma Mel è opportuno costruire prima un banco filtri da poter poi applicare allo spettrogramma classico per tener conto di quanto illustrato in precedenza. Procediamo per step.

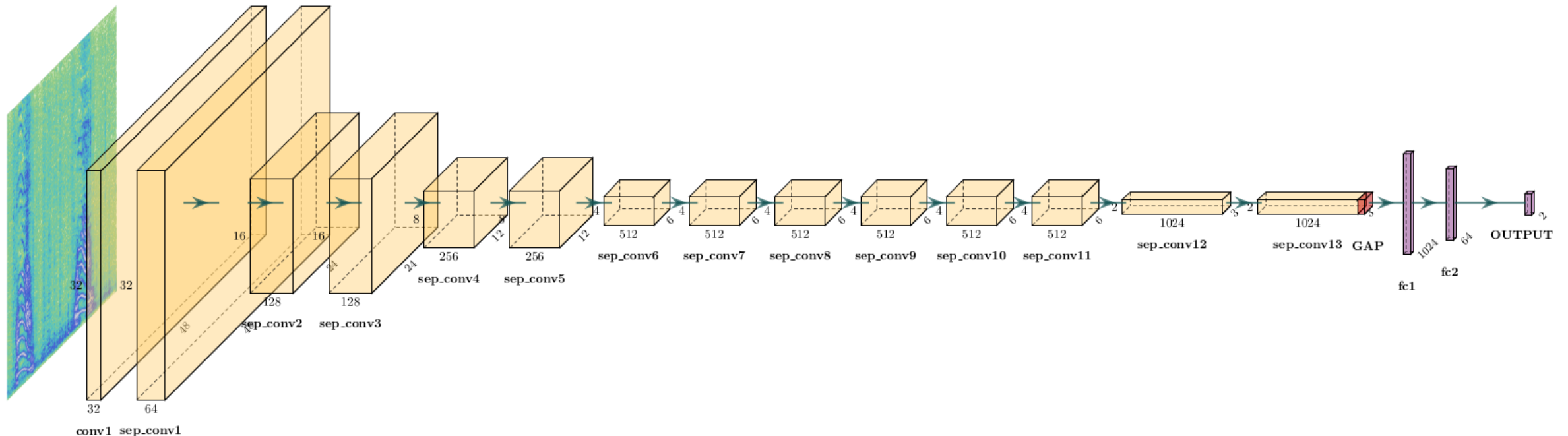


Creare un filtro triangolare di pesi da applicare al segnale di partenza

Riconoscimento di sorgenti con reti neurali convoluzionali

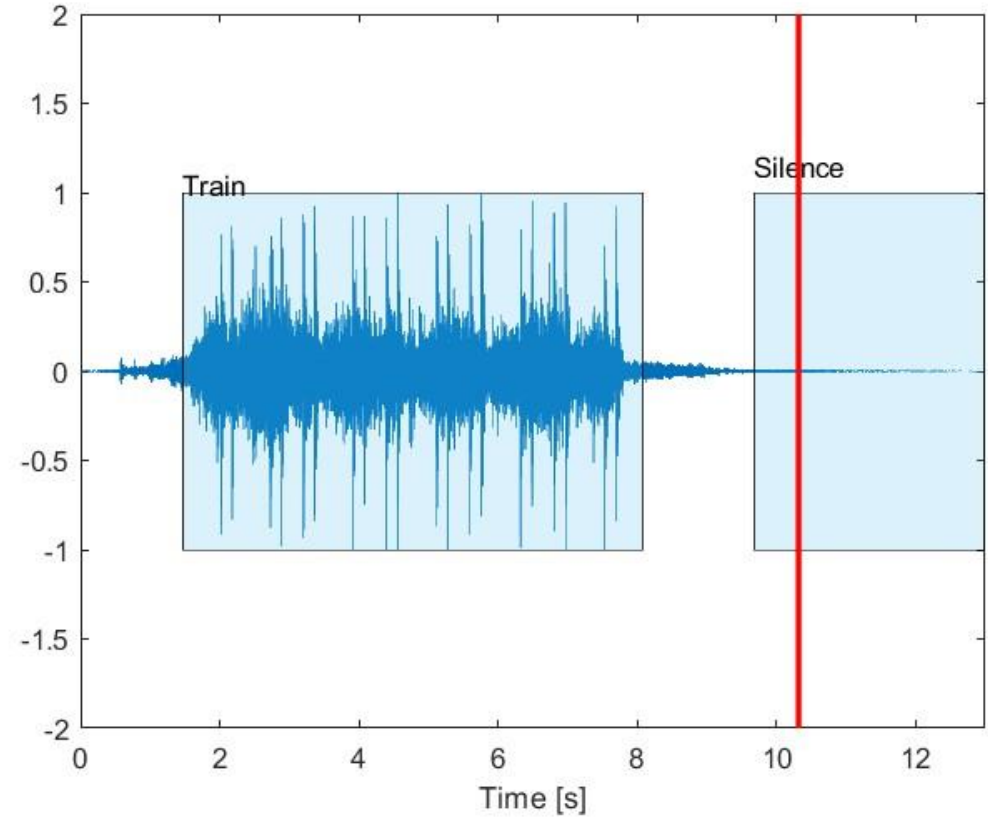
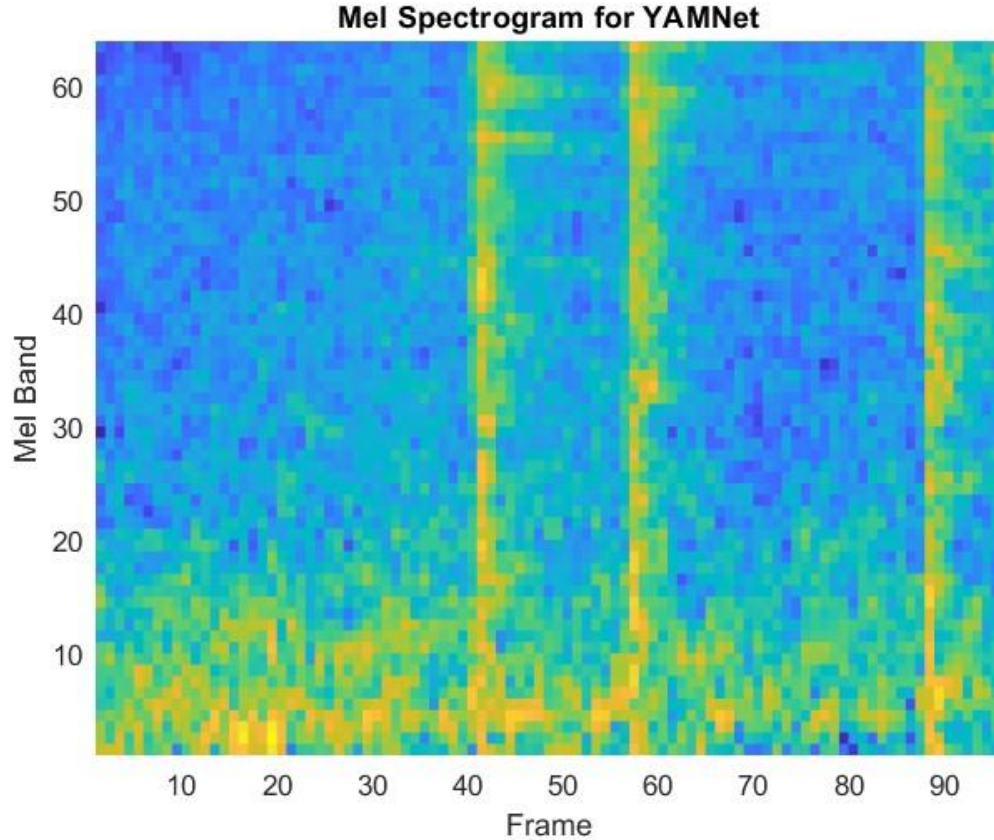
YamNet

YAMNet (Yet Another Mobile Network) – Sì, questa è la forma completa, è un modello di rilevamento acustico pre-addestrato, addestrato da Dan Ellis su un set di dati AudioSet che contiene dati etichettati da oltre 2 milioni di video di Youtube

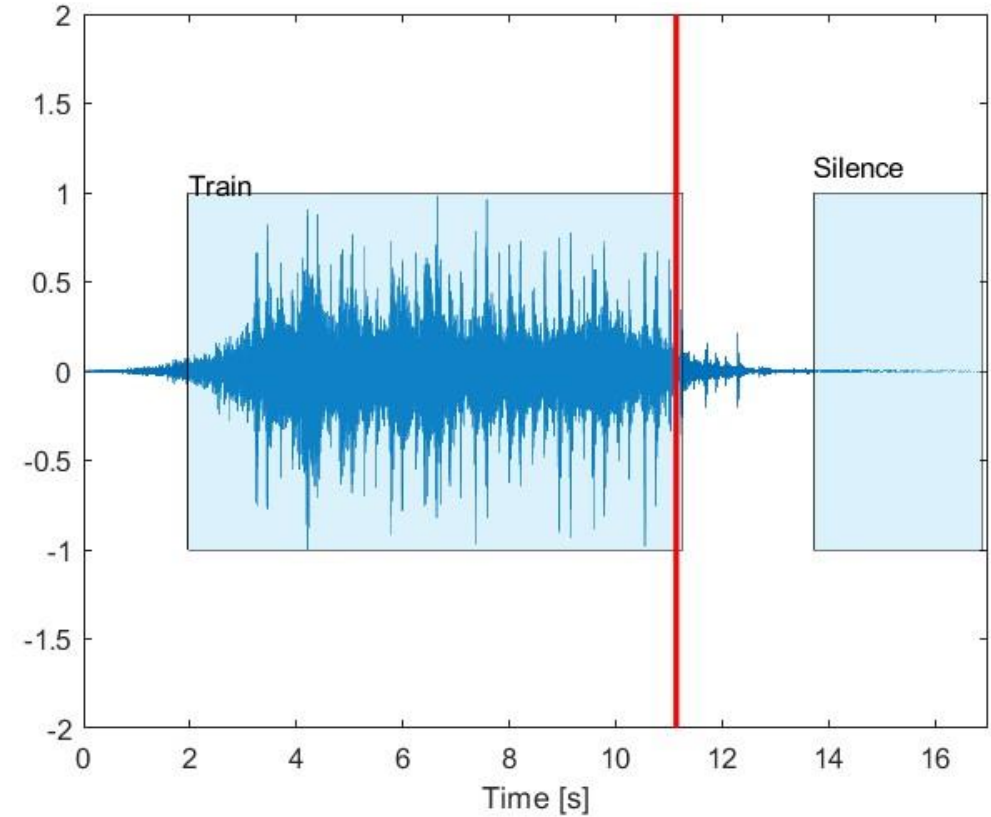
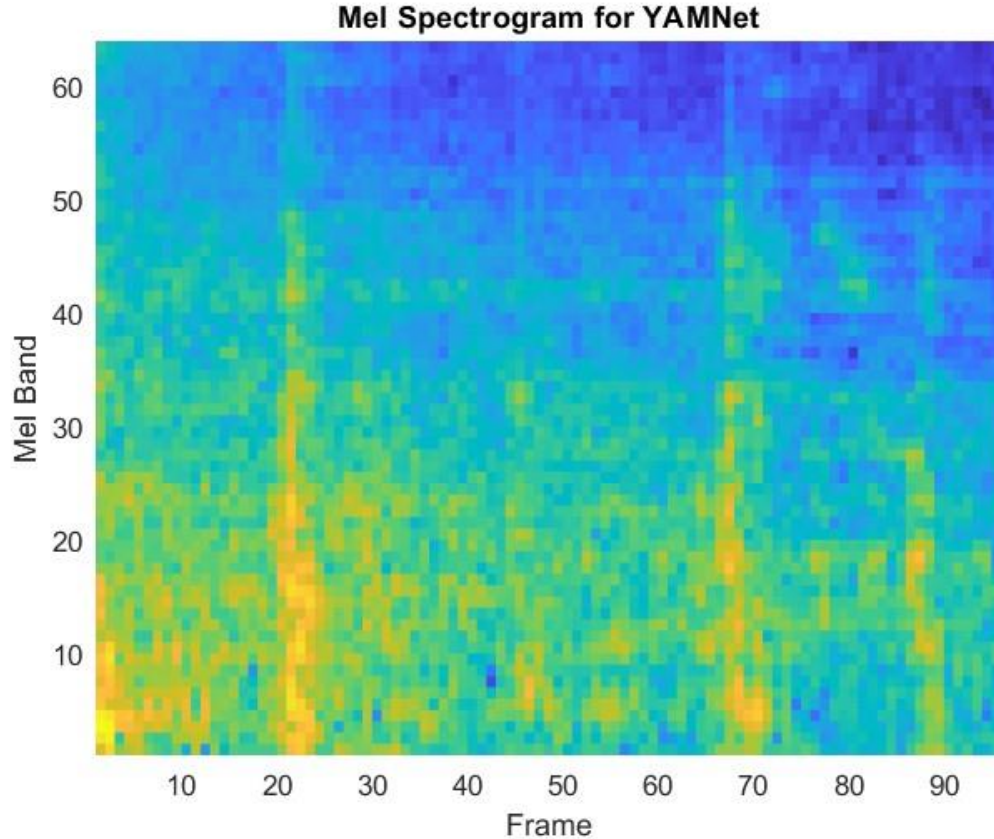


[AudioSet \(google.com\)](https://audioset.google.com)

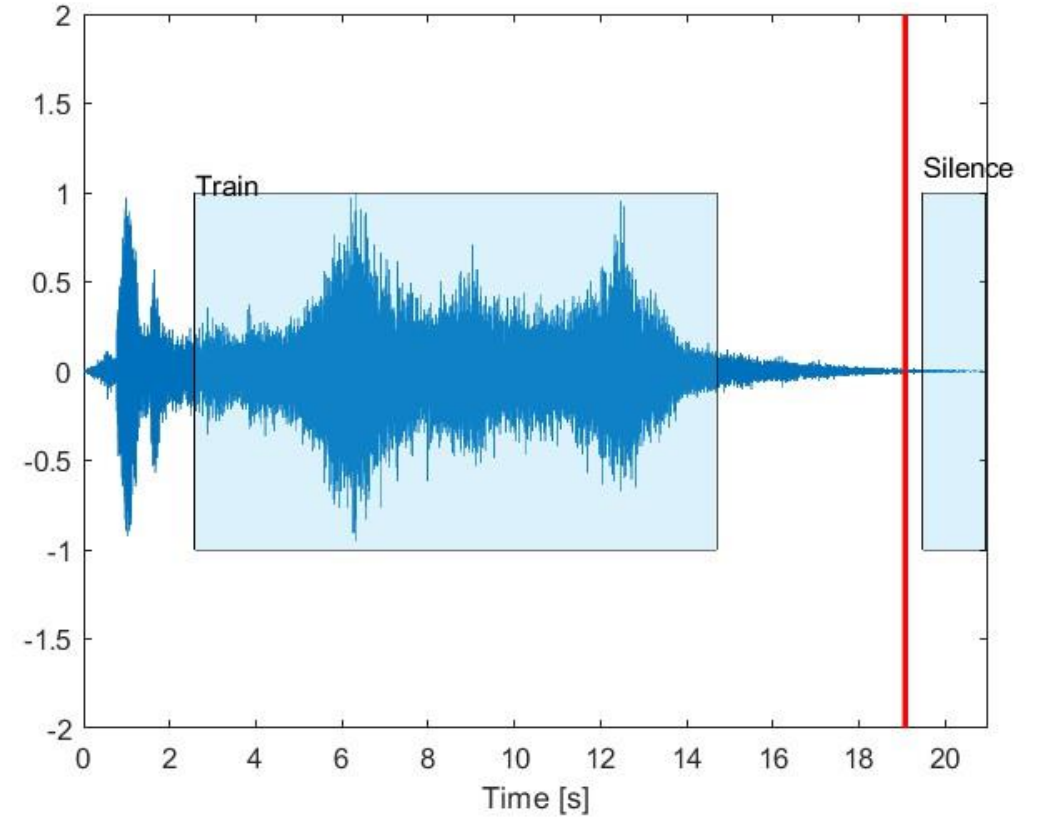
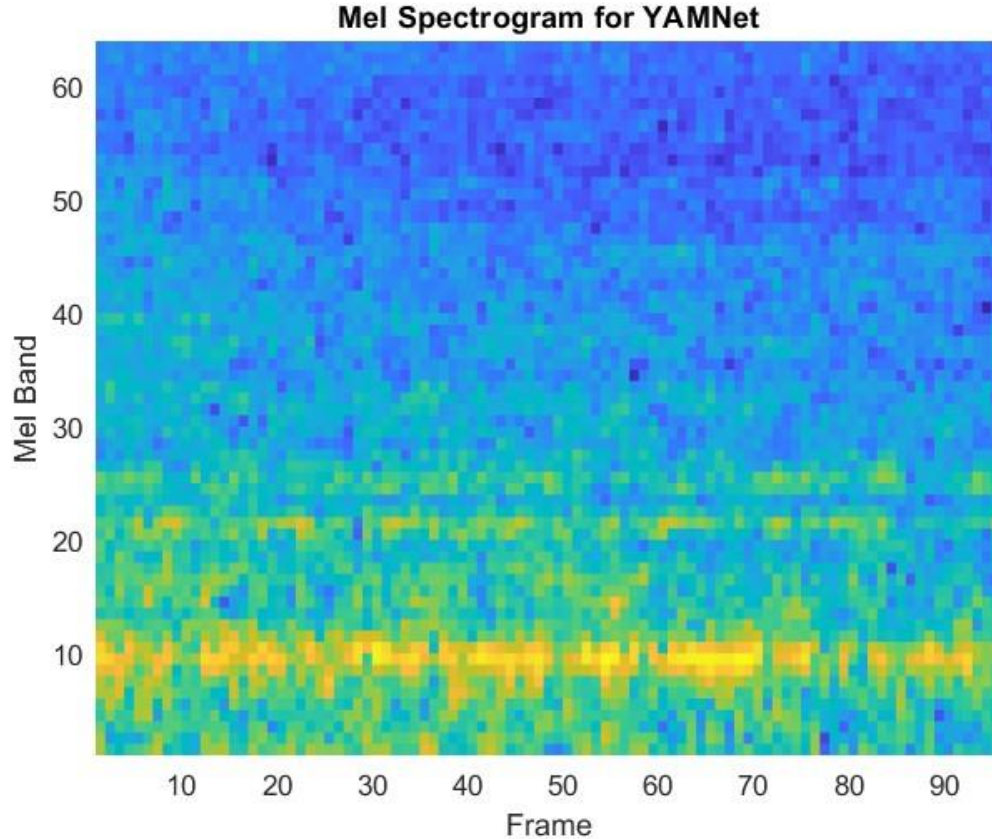
Riconoscimento di sorgenti con reti neurali convoluzionali



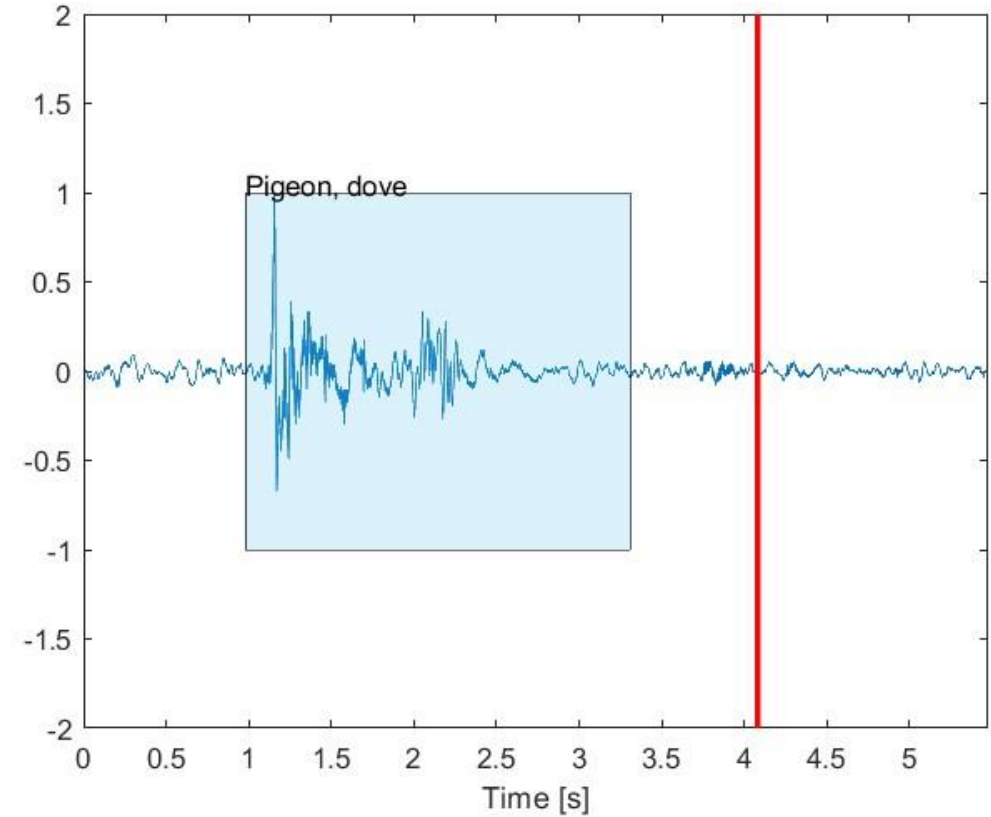
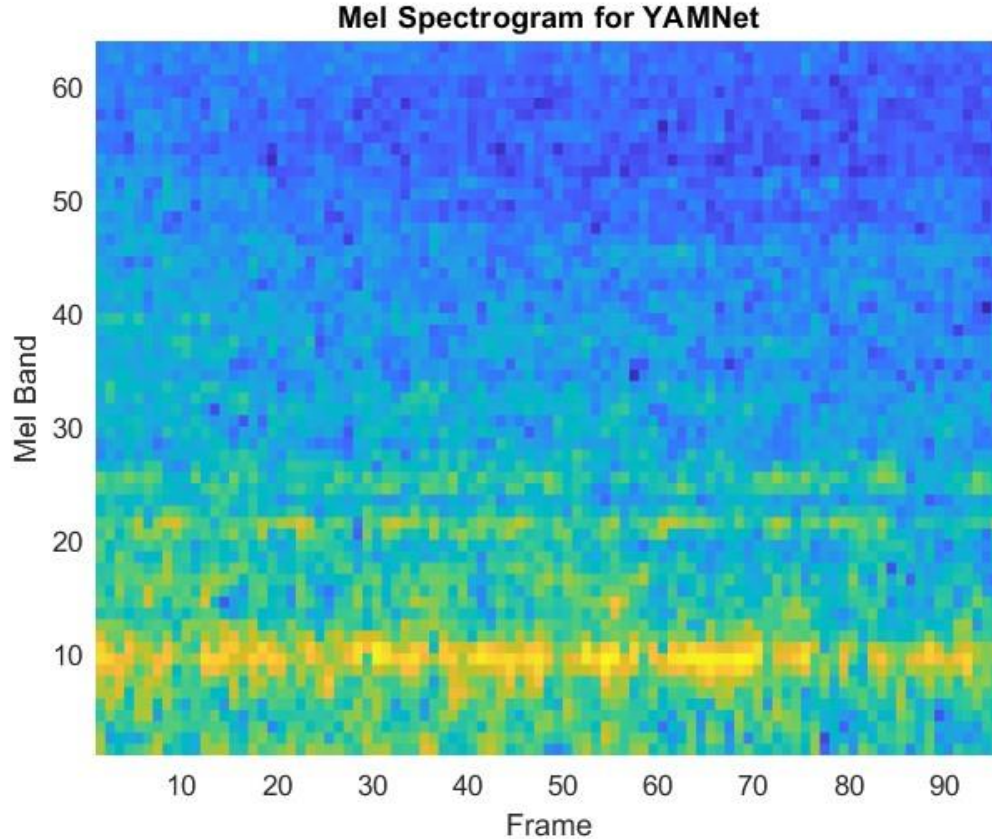
Riconoscimento di sorgenti con reti neurali convoluzionali



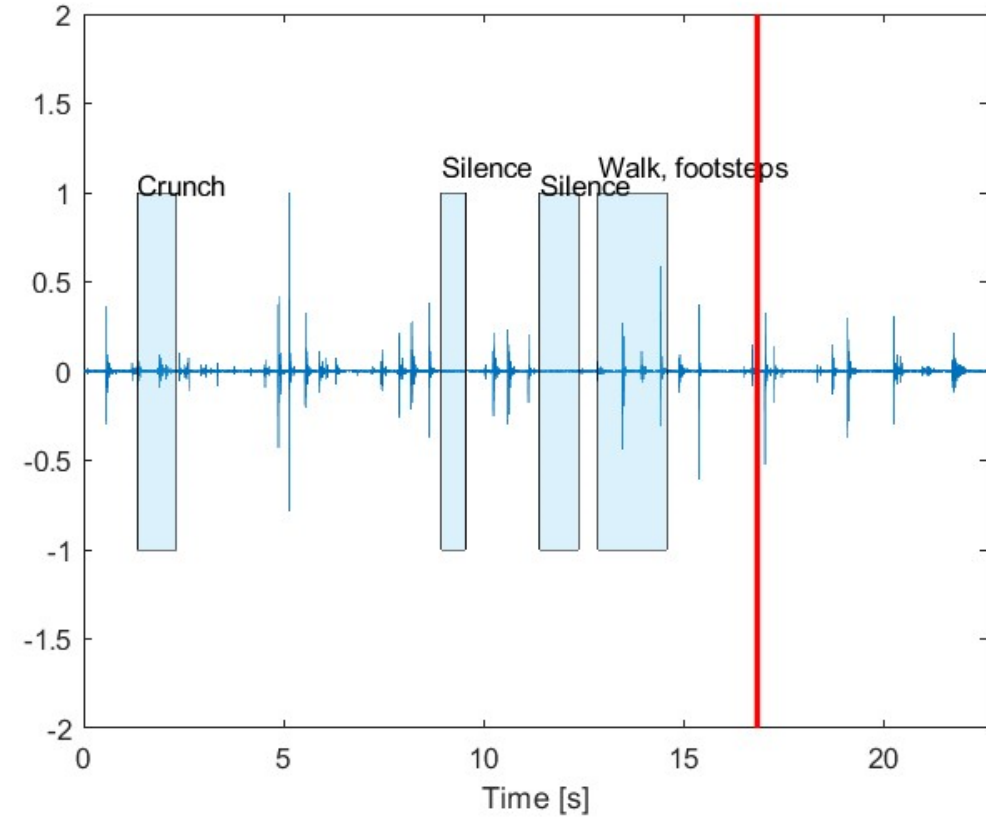
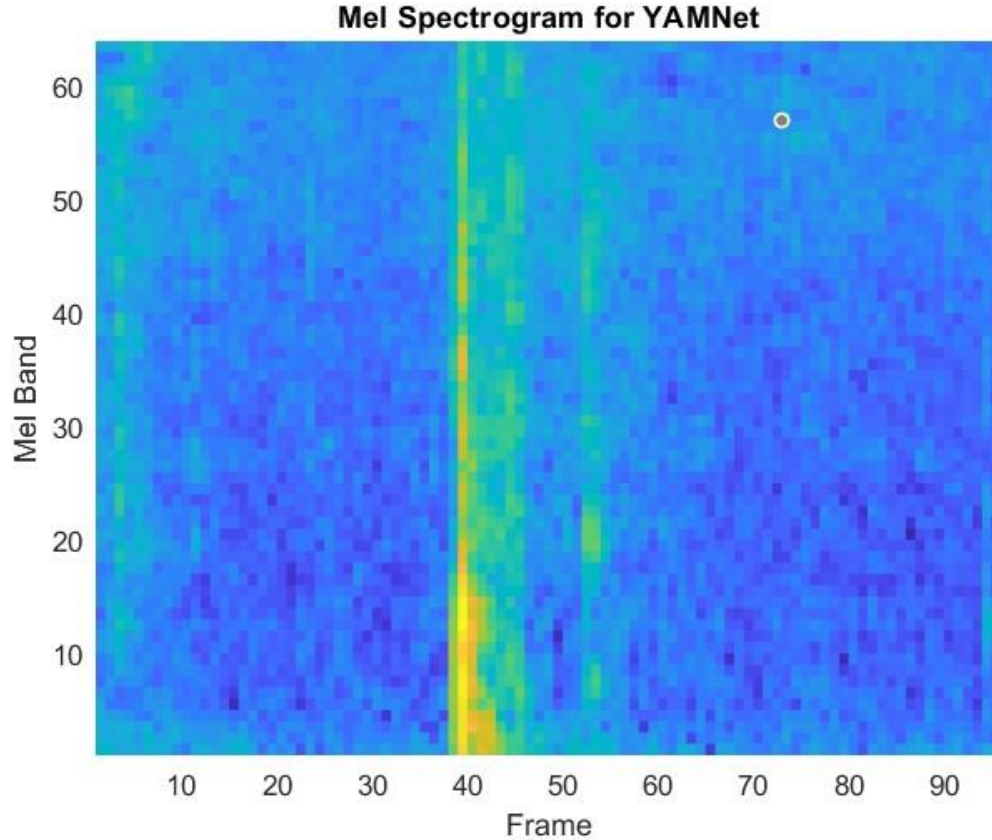
Riconoscimento di sorgenti con reti neurali convoluzionali



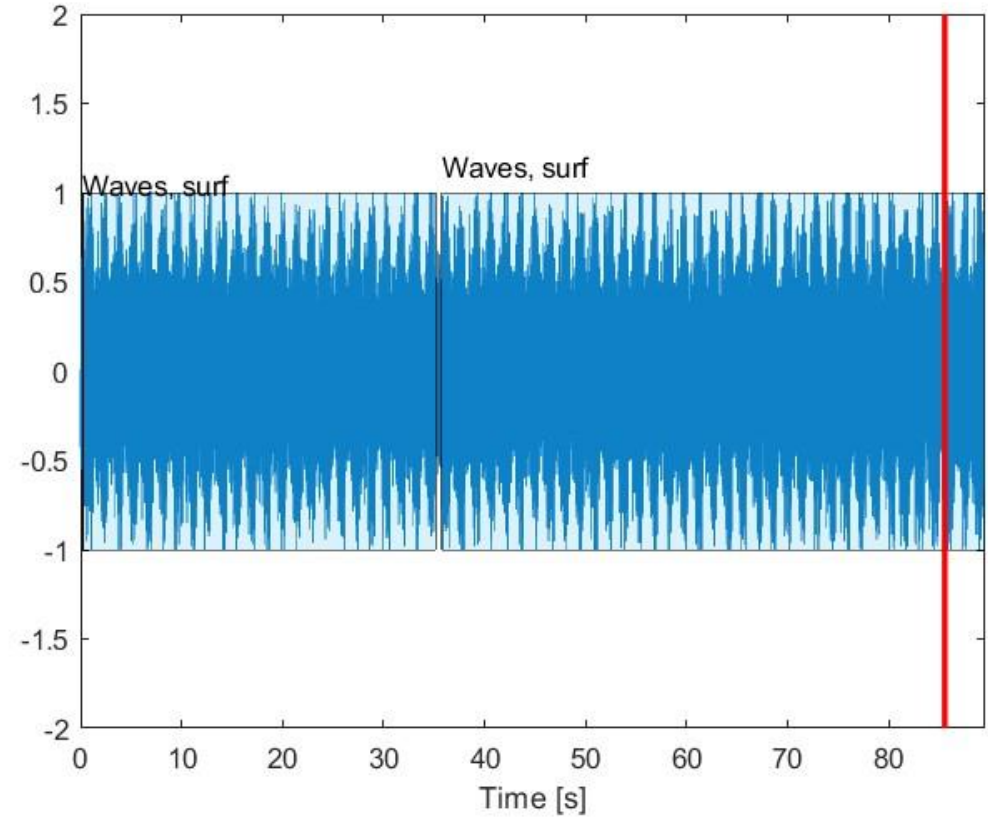
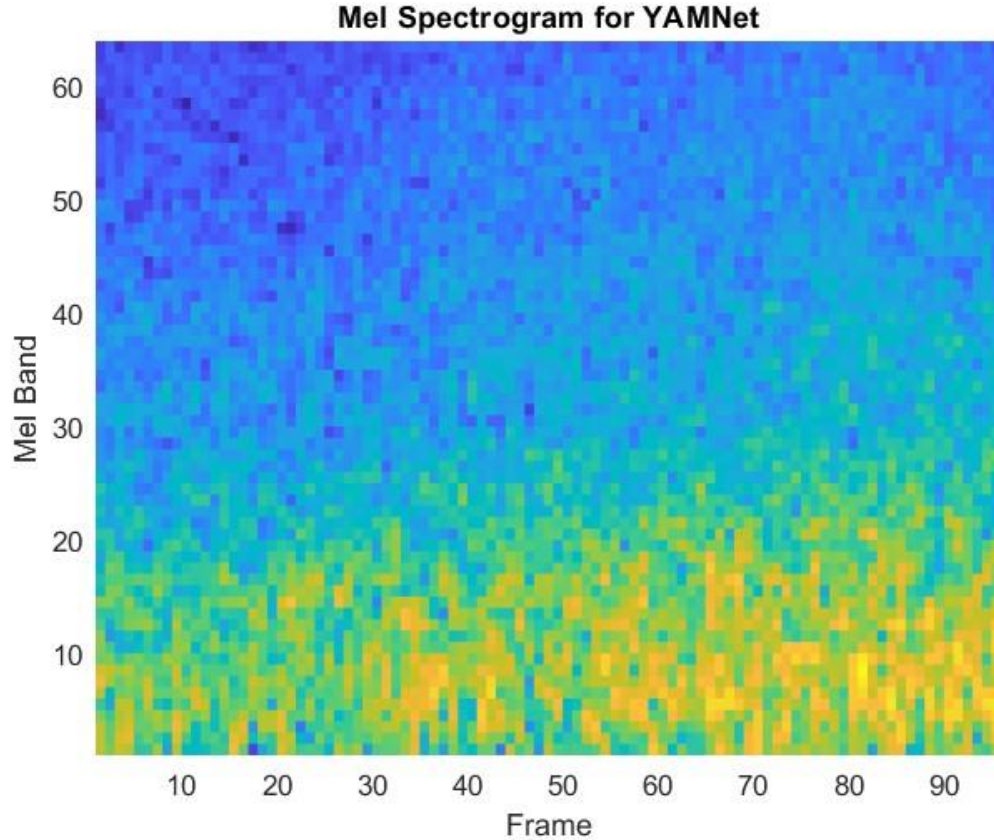
Riconoscimento di sorgenti con reti neurali convoluzionali



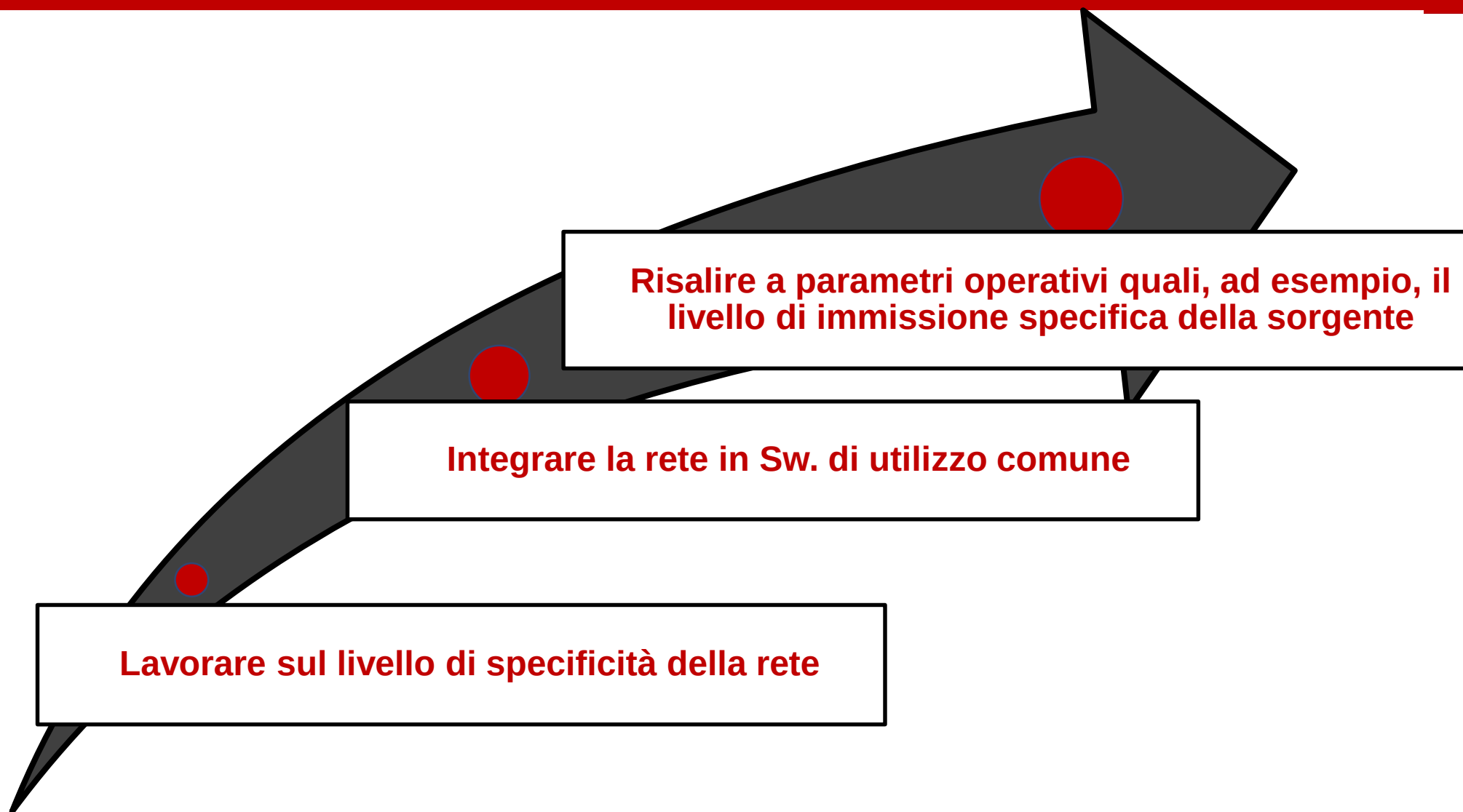
Riconoscimento di sorgenti con reti neurali convoluzionali



Riconoscimento di sorgenti con reti neurali convoluzionali



Sviluppi futuri



A decorative graphic consisting of numerous vertical bars of varying heights, arranged in a wave-like pattern across the center of the slide. The bars are light gray and have a slight 3D effect with shadows.

Grazie per l'attenzione!