



Come funziona Shazam?



Alessandro Liotta

Miriana Martini

Roberta Macaluso

FOTO SOLO DEL VISO



FOTO INTERA CON SFONDO



FOTO DI SPALLE



NESSUNA
FOTO





Indice



■ Premessa

- Passaggio da analogico a digitale: campionamento di un segnale
- Dominio del tempo e dominio della frequenza
- Discrete Fourier Transform (DFT)

■ Tema principale

- Cattura di un suono
- Fingerprint di una canzone
- Creazione di un database di canzoni
- L'algoritmo di Shazam



Una necessaria premessa

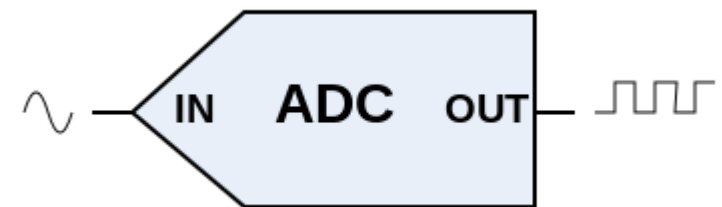


- Il suono è una variazione di pressione che si propaga come un'onda meccanica, attraverso un mezzo
- I trasduttori convertono la pressione dell'onda sonora in un segnale elettrico continuo
- Il segnale continuo non è molto utile nel mondo digitale, quindi prima che possa essere elaborato, deve essere tradotto in un segnale discreto che può essere memorizzato digitalmente



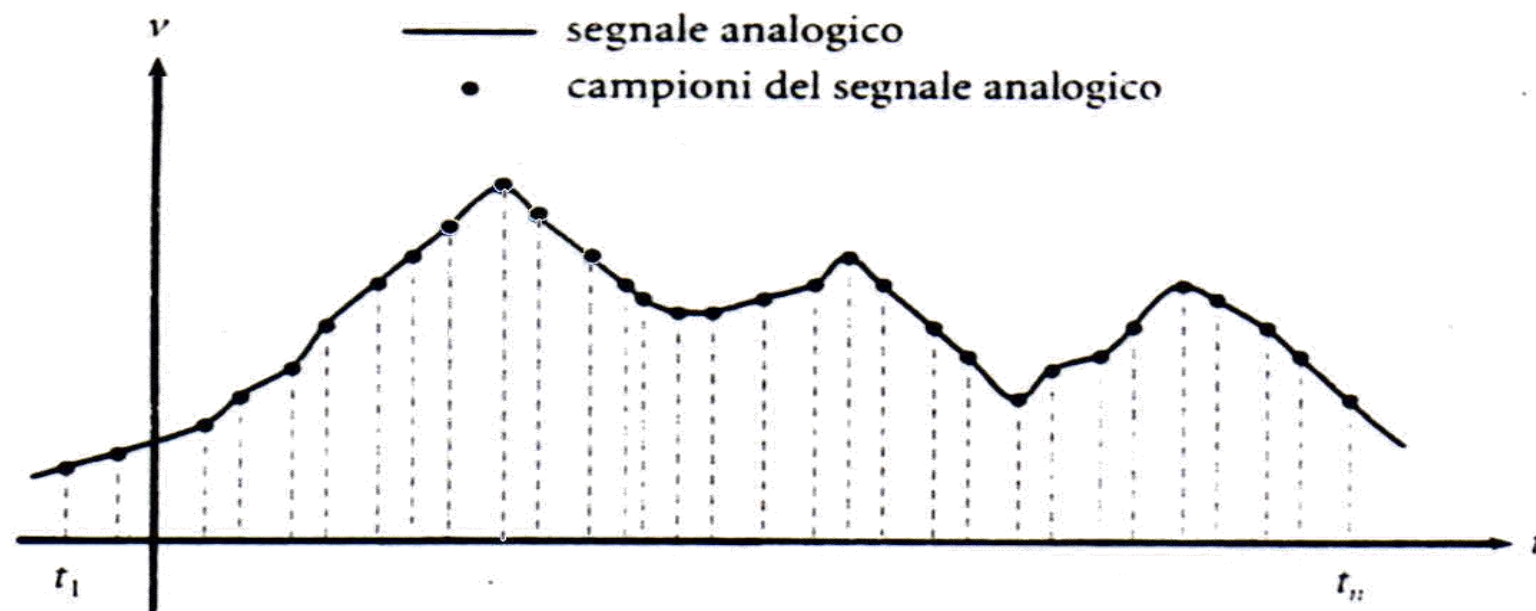
Passaggio da analogico a digitale: campionamento di un segnale

- Un convertitore analogico-digitale esegue molte conversioni su parti molto piccole del segnale catturando di volta in volta un valore digitale che rappresenta l'ampiezza del segnale - un processo noto come campionamento
- Il *Teorema di Nyquist-Shannon* ci dice quale tasso di campionamento è necessario per catturare una certa frequenza nel segnale continuo
- L'audio viene registrato più spesso a una frequenza di campionamento di 44.100 Hz

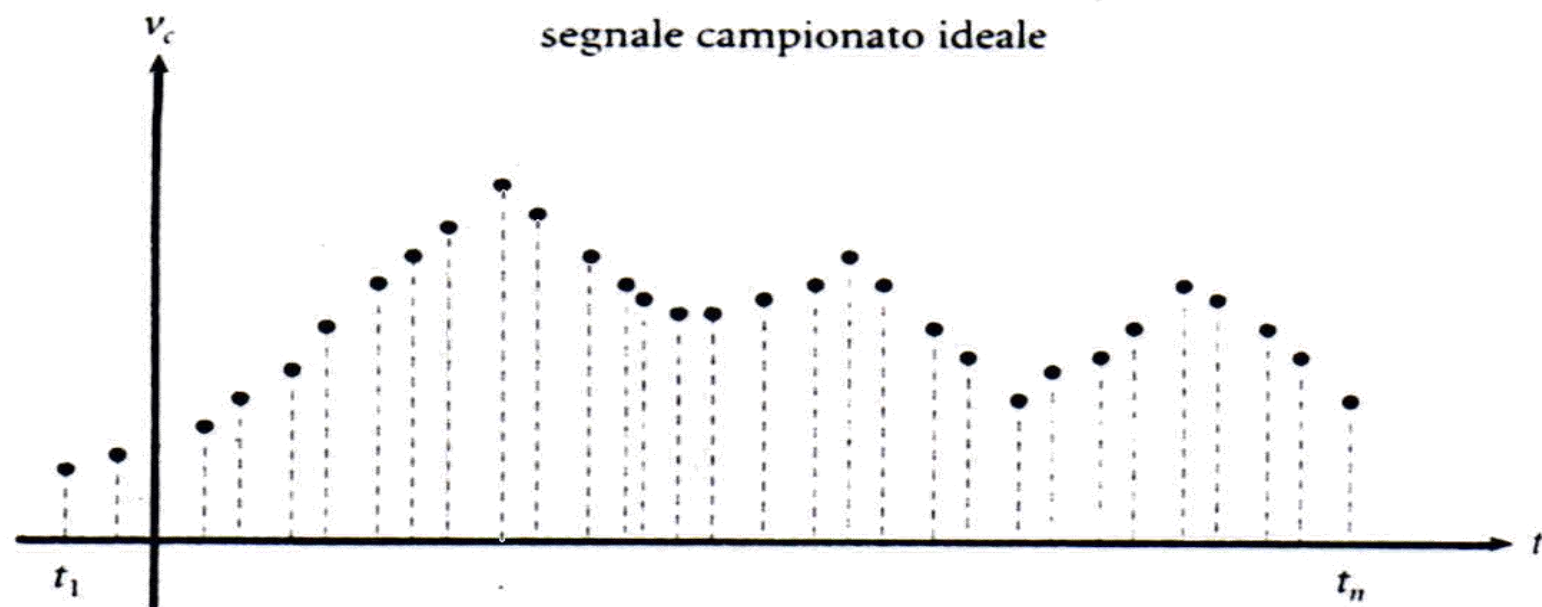




Campionamento di un segnale



(a)



(b)



Dominio della frequenza

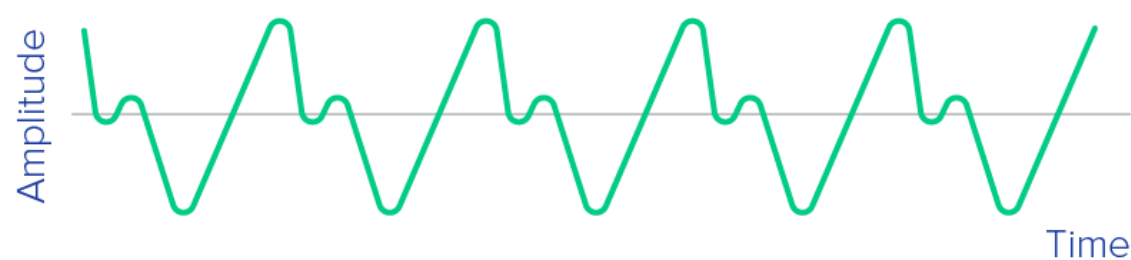
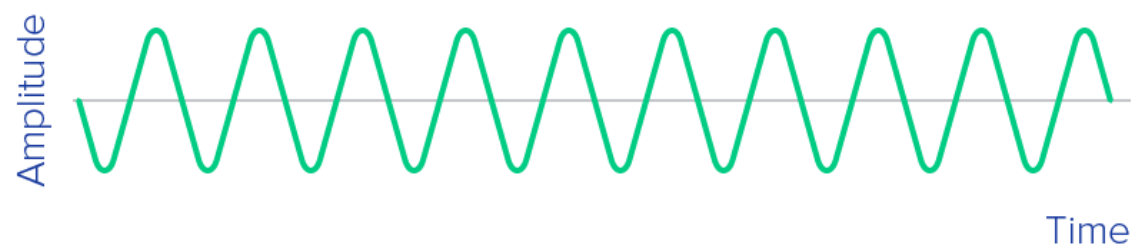
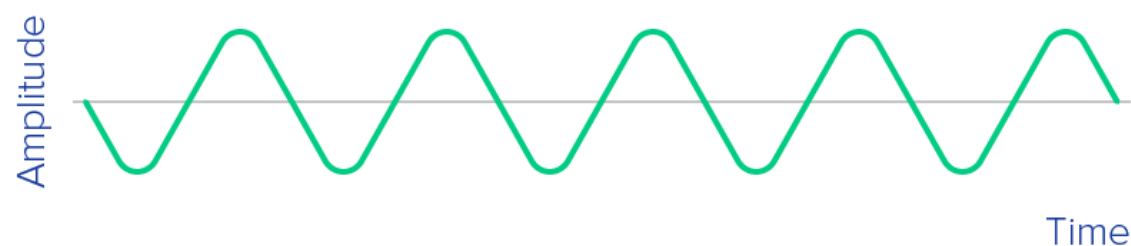


- All'inizio del 1800, Jean-Baptiste Joseph Fourier scoprì che era possibile rappresentare qualsiasi segnale del dominio del tempo considerando semplicemente l'insieme di frequenze, ampiezze e fasi corrispondenti a ciascuna sinusoide che costituisce il segnale
- Il dominio della frequenza agisce come un tipo di *impronta digitale* per il segnale del dominio del tempo, fornendo una rappresentazione statica di un segnale dinamico



Dominio del tempo e dominio della frequenza

Time Domain



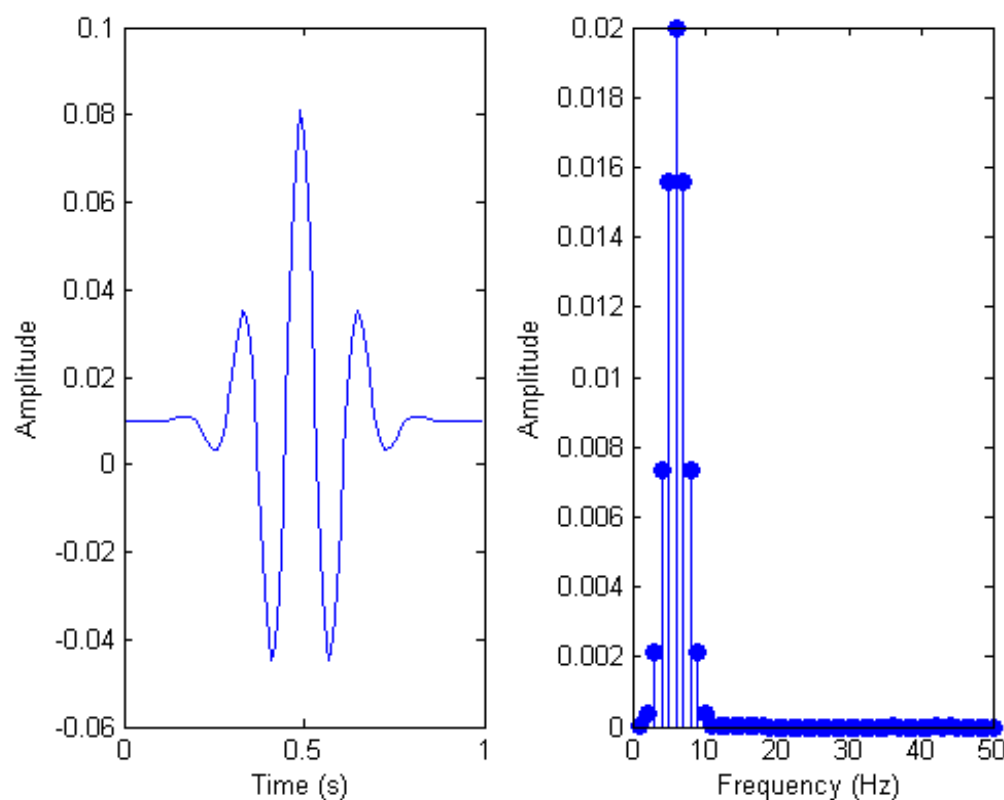
Frequency Domain





Discrete Fourier Transform (DFT)

- Un modo per convertire il nostro segnale dal dominio del tempo al dominio della frequenza è utilizzare la *Discrete Fourier Transform* (DFT)





Cattura di un suono



- Uno sfortunato effetto collaterale della FFT è che perdiamo una grande quantità di informazioni sui tempi, ma per descrivere una canzone abbiamo bisogno di sapere quando è apparsa ogni frequenza
- Ecco perché introduciamo una sorta di *finestra scorrevole*, o un chunk di dati, e trasformiamo solo questa parte delle informazioni
- Una volta che abbiamo informazioni sulla composizione delle frequenze del segnale, possiamo iniziare a determinare il *fingerprint* della canzone



Fingerprint di una canzone



- La sfida principale adesso è come distinguere, nell'oceano di frequenze catturate, quali frequenze sono le più importanti
- Individuiamo le frequenze con la massima ampiezza in diversi intervalli più piccoli che analizziamo separatamente
- Questa informazione costituisce una firma per questo intervallo della canzone, e questa firma diventa parte del *fingerprint* nel suo complesso



Creazione di un database di canzoni



- Per facilitare la ricerca audio, questa firma diventa la chiave in una tabella hash. Il valore corrispondente è il tempo in cui questo insieme di frequenze è apparso nella canzone, insieme all'ID della canzone (titolo del brano e artista)

Hash Tag	Time in Seconds	Song
30 51 99 121 195	53.52	Song A by artist A
33 56 92 151 185	12.32	Song B by artist B
39 26 89 141 251	15.34	Song C by artist C
32 67 100 128 270	78.43	Song D by artist D
30 51 99 121 195	10.89	Song E by artist E
34 57 95 111 200	54.52	Song A by artist A
34 41 93 161 202	11.89	Song E by artist E



L'algoritmo di Shazam



- Molti dei tag hash corrisponderanno all'identificatore musicale di più brani (potrebbe capitare che qualche parte della canzone A suoni esattamente come un pezzo della canzone E)
- Dobbiamo controllare anche la tempistica
- Consideriamo i_1 e i_2 , momenti nel brano registrato, e j_1 e j_2 momenti nel brano dal database. Possiamo dire che abbiamo due combinazioni tra le differenze di tempo se:

$RecordedHash(i_1) = SongInDBHash(j_1) \text{ AND } RecordedHash(i_2) = SongInDBHash(j_2)$

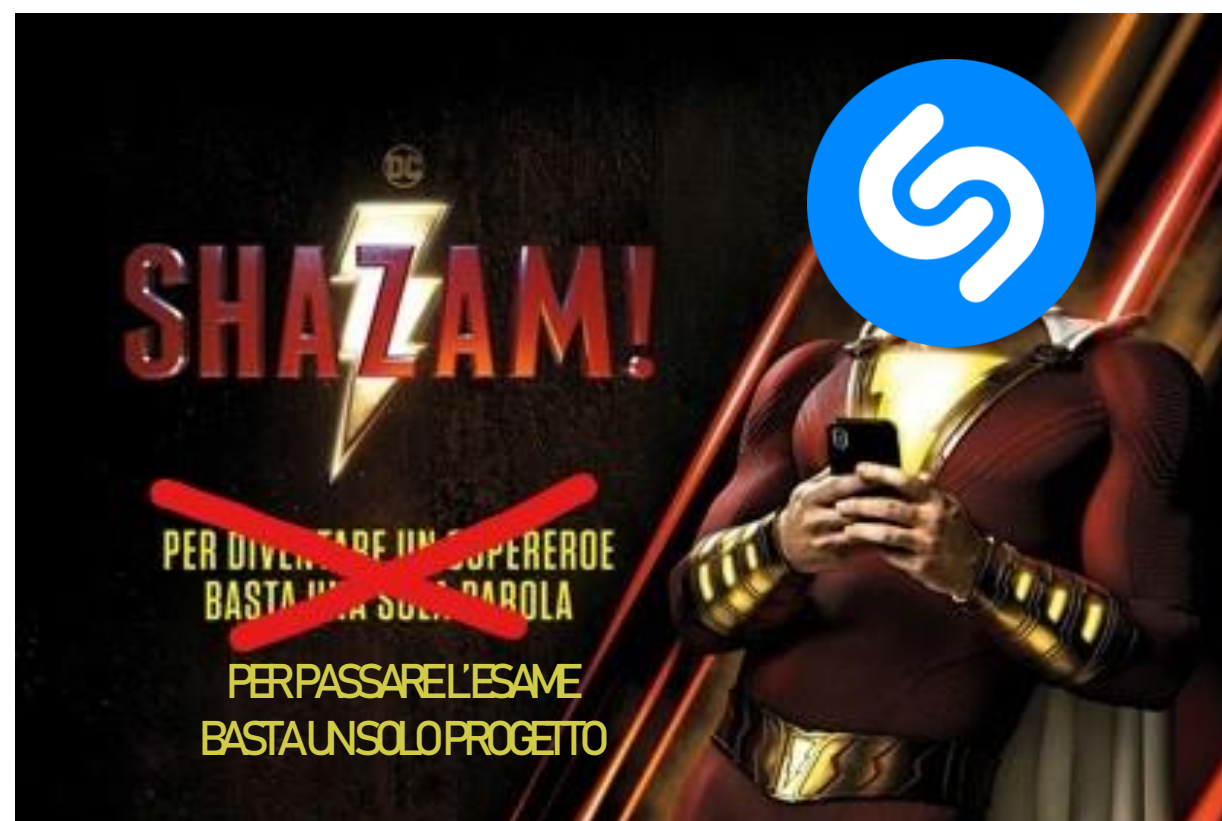
AND

$abs(i_1 - i_2) = abs(j_1 - j_2)$



Conclusioni

- Adesso che abbiamo imparato come funziona Shazam possiamo andare tutti al cinema a vedere il nostro documentario





Alessandro Liotta	O4500923	Ing. Elettronica
Miriana Martini	O46001185	Ing. Informatica
Roberta Macaluso	O46001205	Ing. Informatica

GRAZIE PER L'ATTENZIONE