



**UNIVERSITÀ
DEGLI STUDI
DI UDINE**
hic sunt futura

TECNOLOGIE WEB E LABORATORIO (a.a. 2022-2023)

RISORSE MULTIMEDIALI - PARTE 2

Daniele Salvati

Dipartimento di Scienze Matematiche, Informatiche e Fisiche
Università degli Studi di Udine

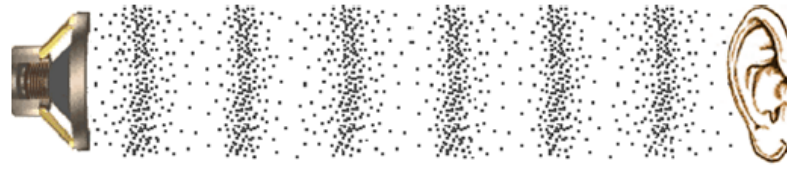
Informazioni slide

- Il materiale contenuto in queste slide è riservato esclusivamente agli studenti del corso di **Tecnologie Web e Laboratorio** del Corso di Studio in **Internet of Things, Big Data, Machine Learning** dell'Università degli Studi di Udine.
- Non è consentita la diffusione del materiale contenuto in queste slide, ma solo l'utilizzo inerente la preparazione dell'esame del suddetto corso.

Audio

Suono

- Il suono è un'onda longitudinale (chiamata onda di pressione) che viaggia attraverso l'aria o un altro mezzo.



- Il processo di registrazione e riproduzione dell'audio digitale è composto da una serie di operazioni per la digitalizzazione dell'onda di pressione sonora e per la conversione dell'audio digitale in onda di pressione:

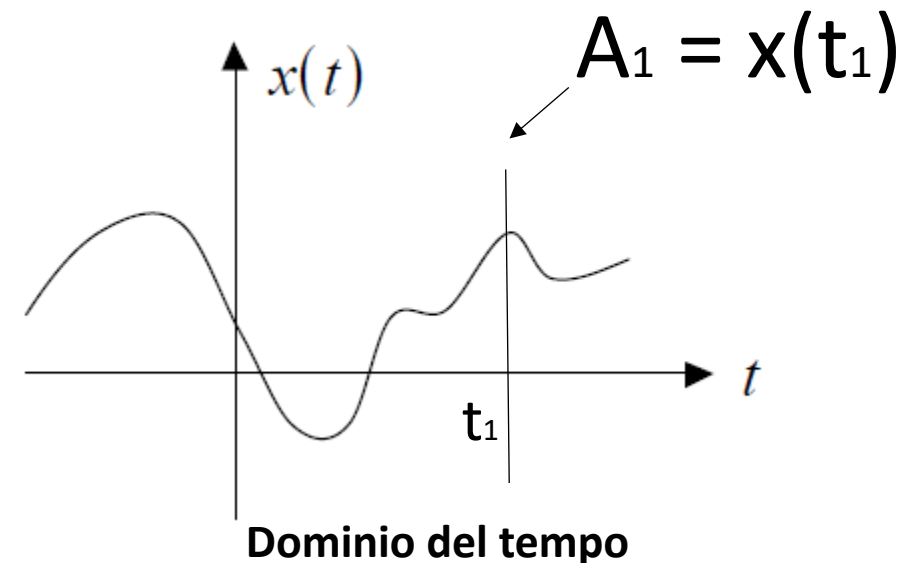


Segnale audio analogico

- Il segnale audio analogico è rappresentabile da una funzione a valori in un intervallo reale di una variabile reale, il tempo t :

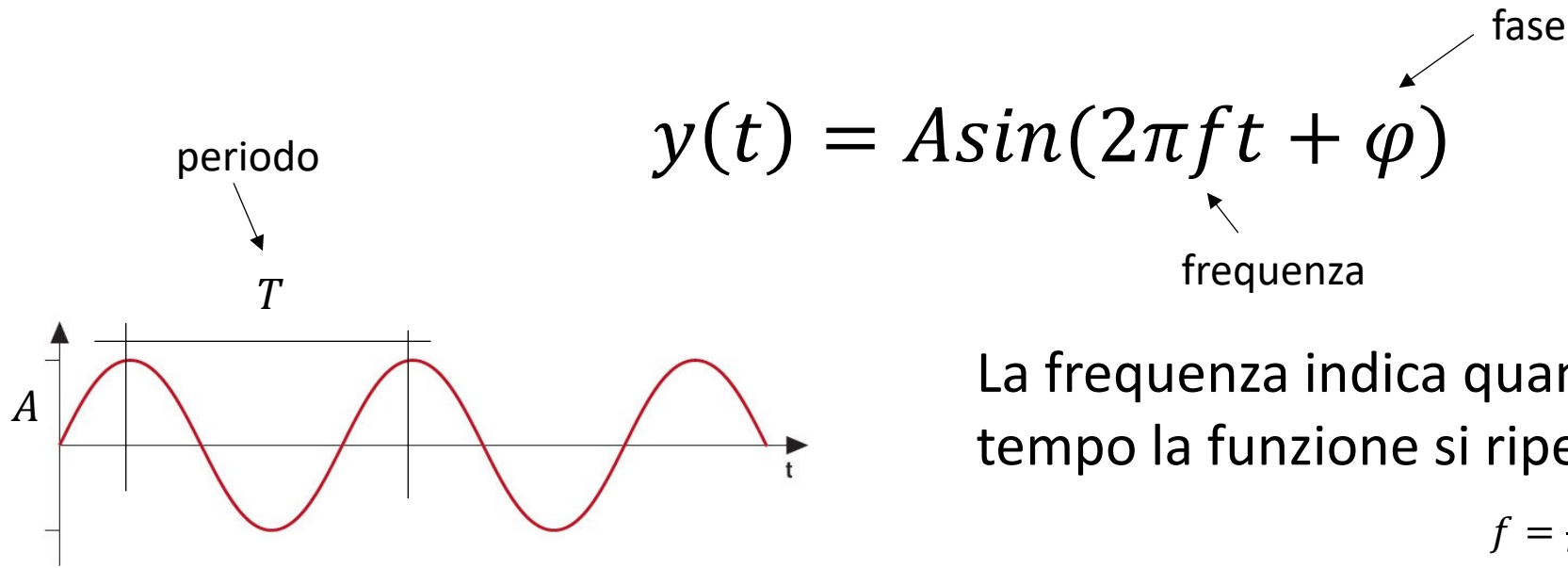
$$A = x(t)$$

con A : l'ampiezza del segnale al tempo t .



Dominio del tempo e della frequenza

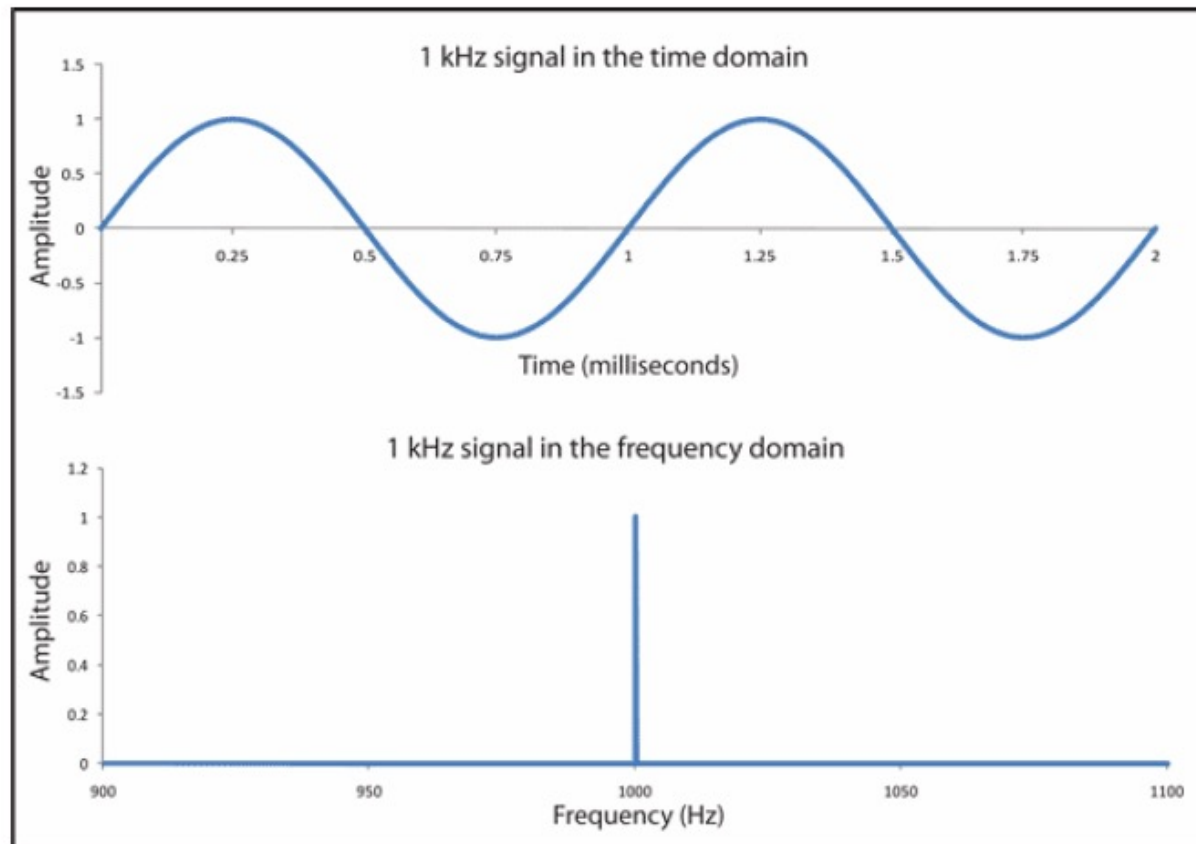
- La relazione di un suono (ma in generale di un segnale) nel **dominio del tempo** e nel **dominio della frequenza** si basa sull'idea dell'**analisi di Fourier**, e sul modello elementare di **onda sinusoidale**.
- Un segnale sinusoidale ha la seguente forma:



La frequenza indica quante volte in un'unità di tempo la funzione si ripete

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

Dominio del tempo e della frequenza (2)

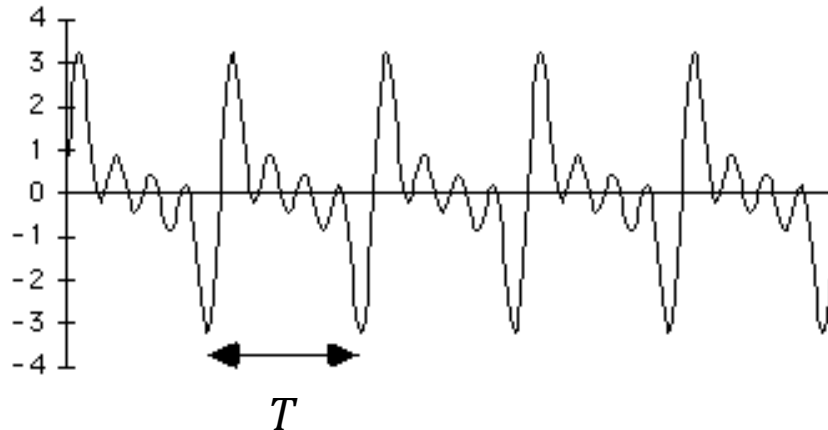


Dominio del tempo

Dominio della frequenza

Serie di Fourier

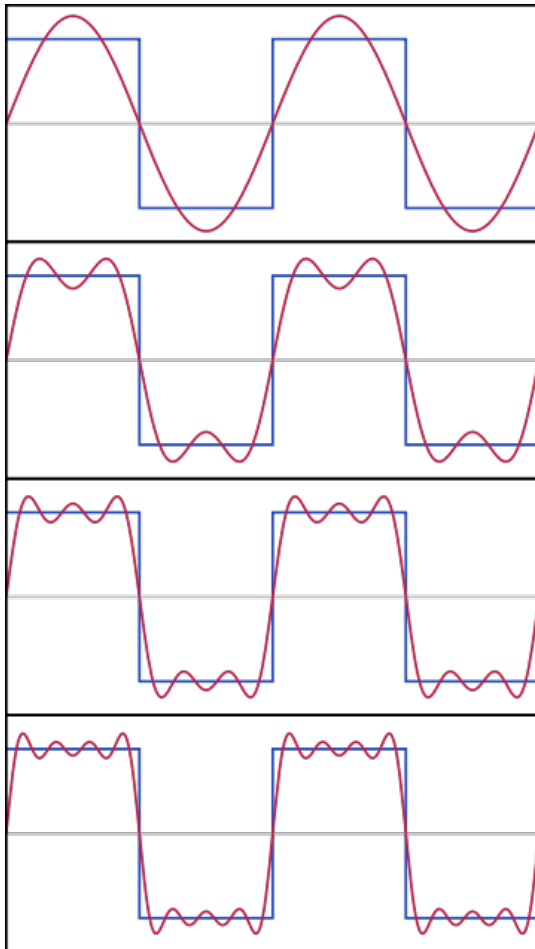
- Le **onde periodiche** possono essere viste come la **somma di più onde sinusoidali** semplici di diversa frequenza e ampiezza.
- Un suono periodico è un segnale che si ripete a intervalli regolari.



- La **serie di Fourier** ha la forma:

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{T} + b_n \sin \frac{n\pi x}{T} \right)$$

Un esempio: l'onda quadra



$$y(t) = A \sin(2\pi f t + \varphi)$$

frequenza fondamentale

+

$$y(t) = \frac{A}{3} \sin(2\pi(3f)t + \varphi)$$

prima armonica

+

$$y(t) = \frac{A}{5} \sin(2\pi(5f)t + \varphi)$$

seconda armonica

+

$$y(t) = \frac{A}{7} \sin(2\pi(7f)t + \varphi)$$

terza armonica

+

...

Trasformata di Fourier

- **La trasformata di Fourier** consente di calcolare le diverse componenti delle onde sinusoidali di un segnale.
- Permette quindi di trasformare un segnale nel dominio del tempo in un segnale nel dominio delle frequenze (in termini di ampiezze e fasi).
- Permette anche di passare dal dominio della frequenza a quello nel tempo con l'anti-trasformata di Fourier .
- La trasformata/anti-trasformata di Fourier ha la forma:

$$X(f) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-j 2\pi f t} dt$$

$$x(t) = \int_{-\infty}^{+\infty} X(f) \cdot e^{+j 2\pi f t} df$$

Trasformata di Fourier a tempo breve

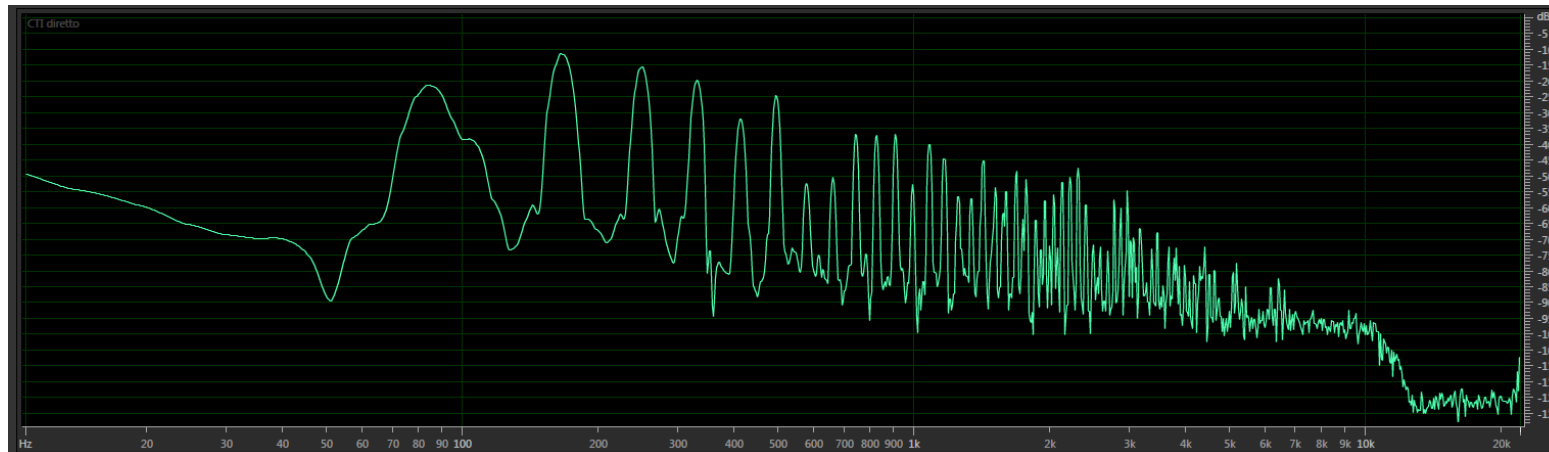
- La **trasformata di Fourier a tempo breve** (*Short Time Fourier Transform*, **STFT**) consente di determinare le componenti sinusoidale (ampiezza e fase) di un blocco di segnale a tempo breve, in modo da poter avere una rappresentazione in frequenza dei segnali di tipo **non stazionario** (cambiano nel tempo).
- La **Fast Fourier Transform** (**FFT**) è un algoritmo di elaborazione digitale, computazionalmente efficiente per il calcolo della STFT.
- **FFT** richiede che il numero di campioni del blocco del segnale sia una **potenza di 2** (tipicamente le dimensioni dei blocchi a tempo breve sono: 512, 1024, 2048 campioni).

Esempio FFT

Corda vibrante di una chitarra



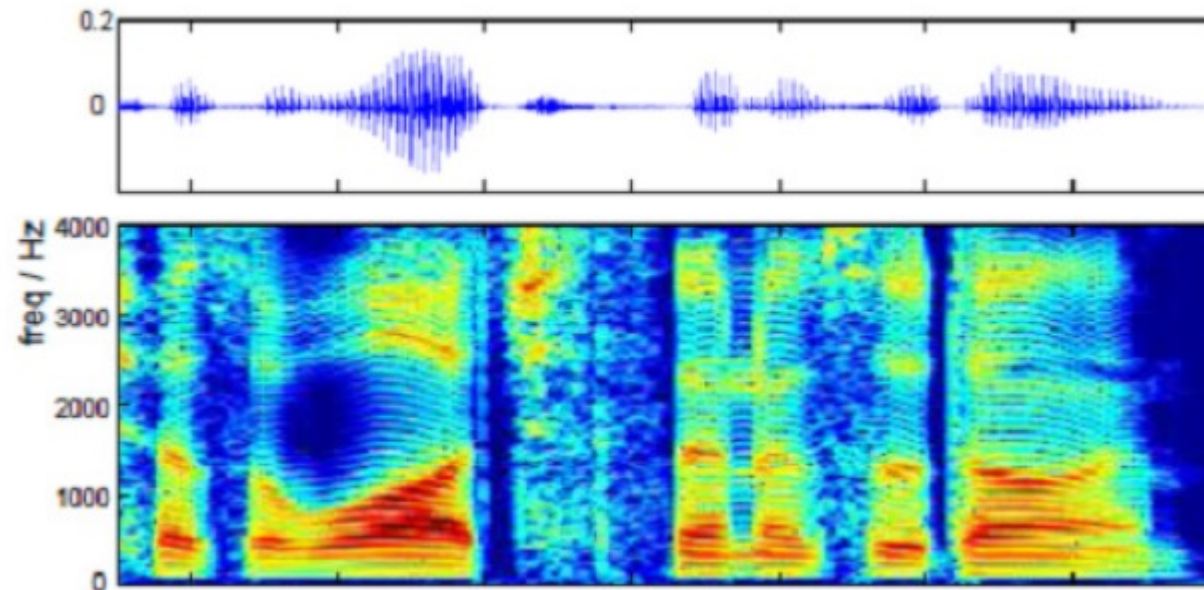
Dominio del tempo



Dominio della frequenza

Spettrogramma

- Uno **spettrogramma** è una rappresentazione tempo-frequenza di un audio digitale (rappresenta l'insieme delle FFT compiute su una sequenza di blocchi dell'intero segnale audio).

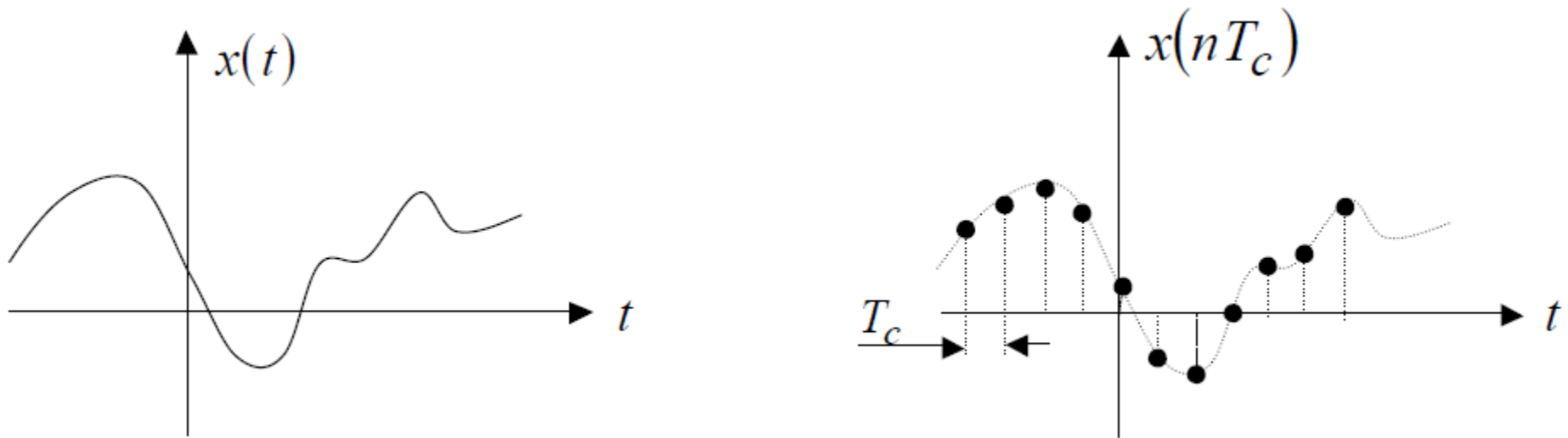


Digitalizzazione del suono

- Un audio digitale composto da campioni numerici del segnale sonoro analogico è ottenuto attraverso le fasi di:
 - campionamento
 - quantizzazione
 - codifica

Campionamento

- Il **campionamento** è la **discretizzazione** del segnale analogico **nel tempo**.



T_c : **periodo di campionamento** -> la frequenza di questa temporizzazione è indicata come **frequenza di campionamento**:

$$f_c = \frac{1}{T_c} (Hz)$$

Teorema di Nyquist–Shannon

- Il **teorema di Nyquist–Shannon** o **teorema del campionamento** definisce la condizione della frequenza di campionamento per acquisire tutte le informazioni a larghezza di banda finita.
- La frequenza di campionamento è data dalla formula:

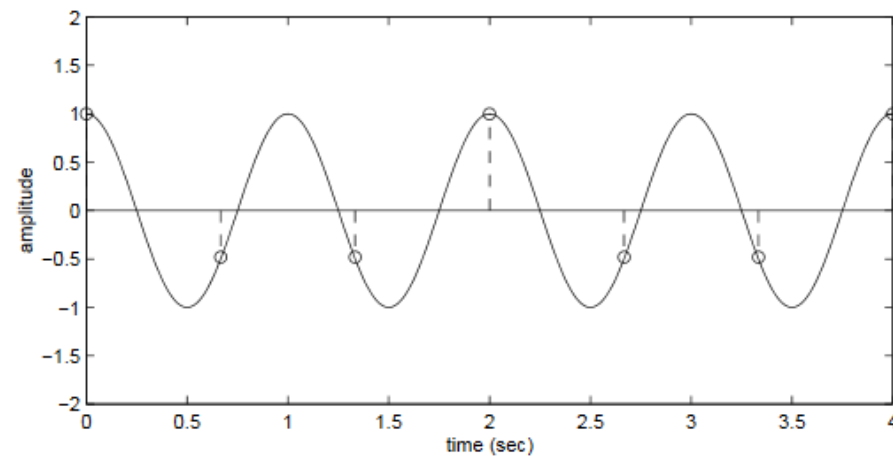
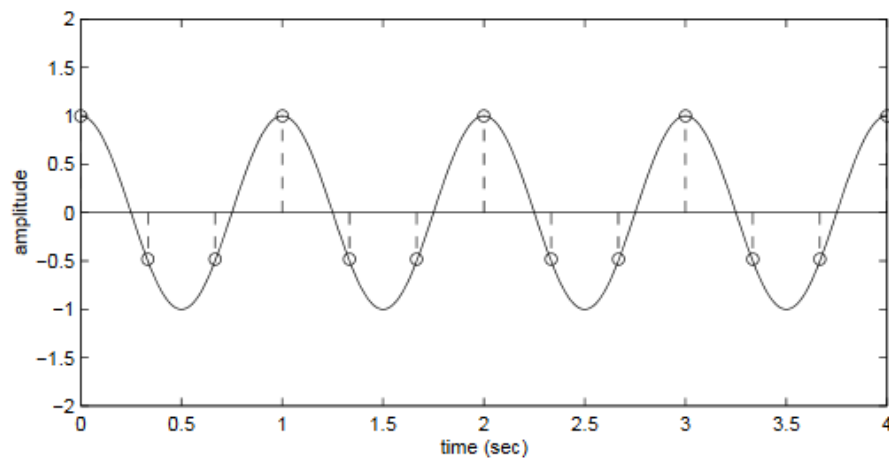
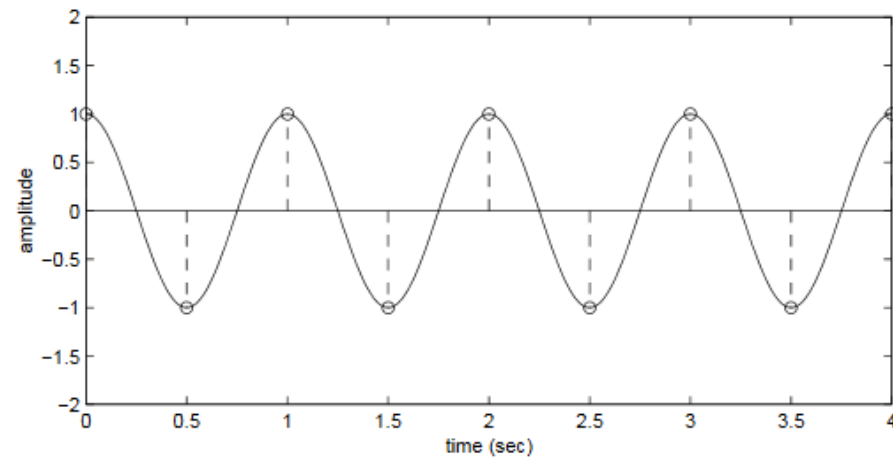
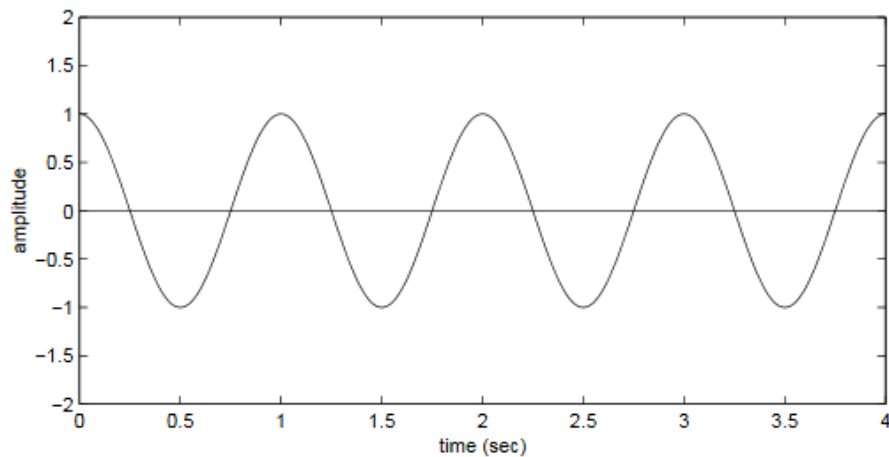
$$f_c \geq 2f_{max}$$

- La **frequenza di campionamento** deve quindi essere almeno il doppio della frequenza massima f_{max} presente nel segnale d'ingresso.

Aliasing

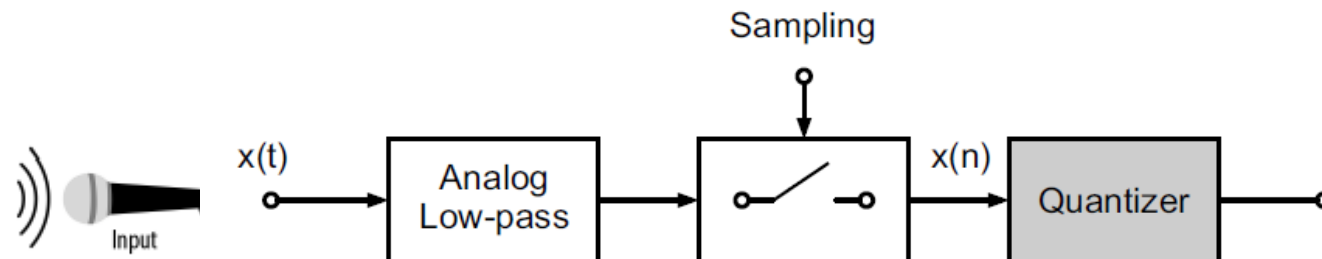
- L'**aliasing** si verifica quando un segnale campiona i dati in modo discreto a una velocità insufficiente per catturare i cambiamenti nel segnale.
- L'aliasing introduce delle distorsioni nel segnale compromettendone il contenuto.
- Queste distorsioni da sotto campionamento sono provocate dalla produzione di **frequenze non proprie del segnale originario** (dette *alias*).

Aliasing (2)



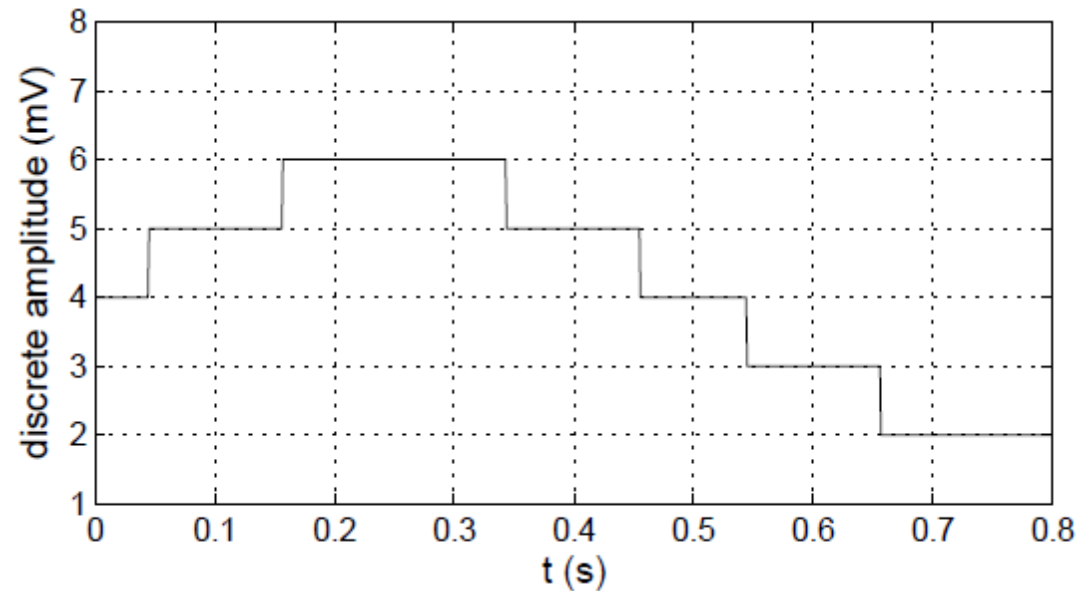
Frequenza di campionamento

- Gamma di frequenze dell'udito umano: **[20, 20000] Hz**.
- Frequenza di **campionamento ideale**: **40000 Hz**.
- Frequenza di campionamento del formato CD (*Compact Disc*): 44100 Hz.
- Il sovra campionamento può semplificare la realizzazione di filtri anti-aliasing analogici (un vero filtro non è perfetto!).
- Un anti-aliasing è un filtro passa basso che limita la frequenza massima.



Quantizzazione

- La **quantizzazione** consiste nella discretizzazione dell'ampiezza del segnale, ossia ogni numero reale è sostituito con un'approssimazione da un insieme finito di **valori discreti**.



Quantizzazione (2)

- **Quantizzazione lineare:** divide la **gamma delle ampiezze** (detta **gamma dinamica**) in 2^N valori discreti.
- Le lunghezze N delle parole sono 8 bit (256 livelli), 16 bit (65536 livelli) e 24 bit (16,8 milioni di livelli).
- **Errore di quantizzazione:** assumendo che ad ogni campione di ampiezza si assegni il valore discreto più vicino, si commette un errore che al massimo vale metà della ampiezza dell'intervallo.
- L'errore di quantizzazione si comporta come un rumore (**rumore di quantizzazione**).

Codifica

- La **codifica** dell'audio digitale consiste nel definire il formato con cui il flusso di dati audio deve essere memorizzato su un supporto o trasmesso in Internet.
- **Pulse-code modulation** (PCM): tecnica di codifica del segnale digitale in cui i campioni quantizzati vengono memorizzati senza alcuna forma di compressione (es., il **formato wav** (*Waveform Audio File Format*) sviluppato da Microsoft e IBM).
- L'audio digitale PCM è trattato come **lunghe matrici di numeri**.

Codifica con compressione

- **MP3** (*Moving Picture Expert Group-1/2 Audio Layer 3*): è un algoritmo di compressione audio di tipo lossy (perdita di informazioni).
- **AAC** (*Advanced Audio Coding*): è un formato di compressione audio lossy creato dal consorzio MPEG. Utilizzato soprattutto nello streaming.
- **FLAC** (*Free Lossless Audio Codec*): è un formato di codifica audio libero con compressione dati lossless, cioè senza perdita di qualità.
- **ALAC** (*Apple Lossless Audio Codec*): è una codifica di tipo lossless sviluppata nel 2004 come formato proprietario, ma ora open-source.

Dimensione di un audio digitale PCM

- Si calcola nel seguente modo:

$$\text{Spazio (bit)} = \text{n. campioni/s} * \text{n. bit/campione} * \text{n. canali} * \text{durata (s)}$$

- Esempio:
 - 30 secondi di audio stereo (2 canali) di qualità CD (44100 Hz, 16 bit)
 - Spazio = $44100 * 16 * 2 * 30 = 42.336.000$ bit

Bitrate audio

- Il **bitrate audio** è la quantità di dati utilizzati per trasmettere un **flusso audio nell'unità di tempo**.
- Il bitrate di un audio PCM (qualità CD) è 1411,2 kbps.
- Un audio mono vocale PCM, campionato a 16000 Hz (il parlato ha in genere un range in frequenza fra 80 Hz e 8000 Hz) e quantizzato a 16 bit ha un bitrate di 256 kbps.

Video

Video analogico

- Una video analogico è rappresentato da una funzione a valori in un intervallo reale di tre variabili reali:

$$I = f(x, y, t)$$

con I : l' intensità luminosa del punto P di coordinate (x, y) dell'immagine video al tempo t .

Video digitale

- Il video analogico è **campionato** (discretizzato) sia **nel tempo** sia **nello spazio**.
- I campioni di intensità luminosa sono **quantizzati** in maniera simile a quanto visto per le immagini digitali.
- I video sono ottenuti da immagini visualizzate in rapida sequenza con una frequenza sufficientemente alta da essere percepite come immagini in movimento e non come una sequenza di immagini fisse (fenomeno della persistenza della visione!).

Struttura interna di un video digitale

- L' elemento strutturale di base di un video è il **fotogramma** (*frame*), ossia ogni singola immagine del video.
- La **frequenza dei fotogrammi** (*frame rate*) è il numero di immagini presentate nell'unità di tempo (**fps**, *frame per second*).
- Per riprodurre la sensazione di movimento fluido e continuo ci devono essere almeno 10/15 immagini al secondo (si sfrutta il fenomeno di persistenza dell'immagine sulla retina: circa 1/10 - 1/16 s).
- Esempi: film al cinema 24 fps, TV 25 fps.

Struttura interna di un video digitale (2)

- **Dimensione logica** (di un frame): numero di pixel di cui è composto il frame. Esempio: 1920×1080 pixel (Full HD).
- **Profondità del colore**: numero di bit usati per codificare il colore di ciascun pixel del frame. Esempio: 8 bit per canale codifica YUV.
- **Rapporto d'aspetto del frame**: rapporto tra larghezza e altezza del frame. Esempio: 1,33 (4:3), 1,78 (16:9 widescreen), 2,35 (cinemascope).
- **Rapporto d'aspetto dei pixel**: rapporto tra la larghezza e l'altezza dei pixel di un frame (tipicamente 1,0).

Dimensione di un video (non compresso)

- Si calcola nel seguente modo:

Spazio (bit) =
dimensione logica del frame *
profondità del colore *
frame rate *
durata del video (s)

Dimensione di un video (2)

- Esempio: calcolare la dimensione associata ad un video digitale con le seguenti caratteristiche:
 - dimensione del frame: 720 x 576 elementi (SDTV)
 - frame rate: 25 fps (PAL)
 - quantizzazione: 8 bit per componente di colore (24 bit/pixel)
 - durata 1 h e 30 min

$$\text{Spazio} = 720 * 576 * 24 * 25 * 90 * 60 = 1.343.692.800.000 \text{ bit}$$

(circa 168 GB!)

Codifica video

- La codifica video può essere effettuata con due tipologie di famiglie:
 - **Codifica intra frame:** le tecniche di compressione dati sono applicate ad ogni fotogramma indipendentemente dagli altri frame del video.
 - **Codifica inter frame:** le tecniche di compressione dati sono applicate in termini di uno o più frame adiacenti.

Codifica inter frame

- Si utilizzano dei fotogrammi di riferimento decodificati in modo indipendente con codifica intra frame (**fotogrammi di tipo I**) e **fotogrammi di tipo P** (codificati con riferimento al cambiamento previsto).
- **Codifica differenziale**: ciascun fotogramma viene confrontato con quello di riferimento (ossia con un fotogramma I o P precedente) e vengono codificati solo i pixel modificati rispetto al fotogramma di riferimento.
- **Compensazione del movimento basata su blocchi**: individua la parte di un nuovo fotogramma di una sequenza video che corrisponde a quella di un fotogramma precedente, anche se riferito a una posizione diversa, dividendo successivamente il fotogramma in una serie di macro-blocchi (predizione spaziale).

H.265 e AV1

- **H.265 HEVC** (*High Efficiency Video Coding*) è uno standard di compressione video **inter frame** basato sulla codifica orientata al blocco e con compensazione del movimento.
- È uno dei formati di codifica più comunemente utilizzato per la registrazione, la compressione e la distribuzione di contenuti video (insieme alla versione precedente H.264).
- Supporta risoluzioni fino a 8K UHD (*Ultra High Definition*).
- **AV1** (AOMedia Video 1) è un formato di codifica video con licenza royalty-free e aperto per la trasmissione dei video in streaming sviluppato da Alliance for Open Media consortium.

Formati multimediali

Formati contenitori multimediali

- Un formato contenitore è utilizzato per identificare e intercalare **diversi tipi di dati**.
- I **formati contenitori multimediali** possono supportare più flussi audio e video, sottotitoli, informazioni sui capitoli e metadati, insieme alle informazioni di sincronizzazione necessarie per riprodurre i vari flussi insieme.
- Un formato può contenere diversi tipi di dati compressi con differenti codifiche per l'audio e il video.

Formati contenitori multimediali (2)

- Alcuni esempi di contenitori multimediali:
 - **AVI**: standard Microsoft Windows
 - **MKV** (Matroska): è un formato contenitore standard aperto non limitato a nessun formato di codifica specifico
 - **MOV**: standard QuickTime Apple
 - **MP4**: contenitore standard audio e video per MPEG-4
 - **Ogg**: formato contenitore libero per il trasporto di flussi di bit, supportato nell'HTML5
 - **WebM**: formato libero e royalty-free di Google sviluppato per l'uso con HTML5

Compressione dei dati

Compressione dei dati

- La **compressione dei dati** produce una rappresentazione più compatta delle informazioni, andando a ridurre il numero dei bit necessari per riprodurre i dati digitali.
- Esistono diversi meccanismi di compressione:
 - **Lossless**: senza perdite di informazione
 - **Lossy**: con perdita di informazione

Contenuto informativo

- Una misura del contenuto informativo di un messaggio è fornita dalla **teoria dell'informazione**.
- **Shannon** introdusse il concetto di **auto-informazione** (*self-information*) con cui si può definire l'informazione associata all'accadere di un evento.
- Dato un evento A e la sua probabilità di accadere $p(A)$, l'auto-informazione di A è:

$$I(A) = \log \frac{1}{p(A)} = -\log p(A)$$

- Scegliendo 2 come base del logaritmo la misura dell'informazione è in bit.

Entropia

- La definizione di auto-informazione può essere estesa per misurare il contenuto informativo di un messaggio formato da un insieme di N simboli.
- Si definisce **Entropia** la seguente quantità H:

$$H = \sum_i I(A_i) p(A_i)$$

- Per massimizzare l'informazione di un messaggio in codice binario è necessario trovare un metodo di codifica che descriva il messaggio con un numero medio di bit pari all'entropia della sorgente (es. codifica di Huffman e codifica aritmetica).

Metodi di compressione lossless

- Uno degli algoritmi più efficienti di tipo lossless è dovuto a **Lempel, Ziv e Walsh (LZW)**.
- **LZW** è un metodo a **sostituzione a dizionario dinamico**, perché **costruisce un dizionario dei dati**.
- L'informazione può essere vista come una stringa di dati (**sequenza di dati**), all'interno del quale sono identificati dei pattern (**sotto-sequenze**) e ricercati nel dizionario.
- Se non li si trova, viene costruito un codice per quel pattern e viene aggiunto al dizionario.
- Se un pattern è già presente nel dizionario il suo codice viene scritto nell'output del file compresso.

LZW

- Esempio: La stringa di input è **ababbcbc**
- Prima fase: identificazione dei pattern (costruzione dinamica del dizionario, si cerca il pattern più corto non ancora presente nel dizionario):

a b ab bc bc

- Seconda fase: codifica della stringa:

12344

- Il rapporto o tasso di compressione sarà

$$\text{RC} = 100 \times \frac{5}{8} = 62,5 \%$$

- Il rapporto di compressione si identifica anche come 8:5.

Dizionario (<i>codebook</i>)	
a	1
b	2
ab	3
bc	4

Run length coding (RLC)

- **RLC** è un **metodo a predizione** di compressione lossless, ossia si basa sulla storia della sequenza dei dati.
- Una sequenza di caratteri formata da più caratteri uguali è detta **run**.
- Un run è “codificato” con due byte:
 - **primo byte** = **numero di caratteri**
 - **secondo byte** = **carattere**

Esempio RLC

- La sequenza **AAAAAAAAAAAAAAAAA** (15 caratteri di tipo A) è codificata come **15A**.
- La sequenza **AAAAAbbbXXXXt** è codificata come **6A3b5X1t**.
- I file di testo (es, ASCII) raramente contengono run lunghi (spesso contengono run di uno o due caratteri). RLC quindi non è adatto in questi casi.
- Un'immagine che contiene molti dati contigui dello stesso colore (nero) può invece essere compressa bene con RLC.

Metodi di compressione lossy

- I metodi di tipo lossy codificano i dati andando a rimuovere l'**informazione ridondante** che risulta poco rilevante, ottenendo **rapporti di compressione più elevati rispetto ai metodi lossless**.
- L'informazione ridondante è spesso quella non rilevabile dal **sistema percettivo** (uditivo e visivo) umano.
- I dati decodificati saranno quindi solo in parte la rappresentazione del dato originale, parte dell'**informazione è persa in modo irreversibile**.
- La qualità del dato compresso sarà in generale inferiore all'originale, ma l'informazione essenziale non è comunque modificata in modo significativo.

Discrete Cosine Transform (DCT)

- Le **tecniche orientate alla frequenza** che fanno uso della **Discrete Cosine Transform (DCT)** sono tra le più utilizzate per la compressione loss.
- La DCT è usata per esempio per JPEG (bitmap), MP3 e AAC (audio), H.265 (video).
- La DCT è una classe di operazioni matematiche (simile alla FFT) in grado di trasformare un segnale nel dominio del tempo in una rappresentazione nel dominio delle frequenze, in questo caso con **funzioni coseno reali**.

Discrete Cosine Transform (DCT) (2)

- La DCT a una dimensione ha la forma:

$$X_k = \sum_{n=0}^{N-1} x_n \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) k \right] \quad \text{for } k = 0, \dots, N-1$$

- Nel dominio delle frequenze DCT, gli algoritmi di compressione relativi ai diversi tipi di risorsa multimediale (immagine, audio, video) riescono ad ottenere livelli di compressione molto alti, sfruttando caratteristiche percettive umane (es., il mascheramento del suono per MP3).

Esempi

Non compresso 48,3 kB



JPEG 9,3 kB



JPEG 3,1 kB



MP3

Freq. Campion.	Bitrate	Qualità
8000 - 16000 Hz	16 - 96 kbps	bassa
16000 - 32000 Hz	96 - 256 kbps	sufficiente
44100 Hz	256 - 320 kbps	buona
48000 Hz	320 kbps	ottima

Esercizio

- Una presentazione multimediale è costituita da un video e da una traccia musicale della durata di **1 minuto**.
- La traccia audio è di qualità CD (freq. campion. = **44100, 16 bit, stereo**) e il video ha le seguenti caratteristiche:
 - frame rate: **25 fps**
 - dimensione frame: **320 x 240 pixel**
 - gamma tonale: **256 colori**
 - rapporto di compressione **5:1**
- **Si calcoli:**
 - l'occupazione di memoria (in MB) dell'intera presentazione multimediale.
 - il tempo (in s) necessario per scaricarla supponendo di usare un collegamento con una velocità di trasmissione di 128 kbps.