InformaticaUmanistica

Lezione 3
Gestione di immagini ed audio

Pasquale Savino

ISTI - CNR

UNIVERSITÀ DI PISA

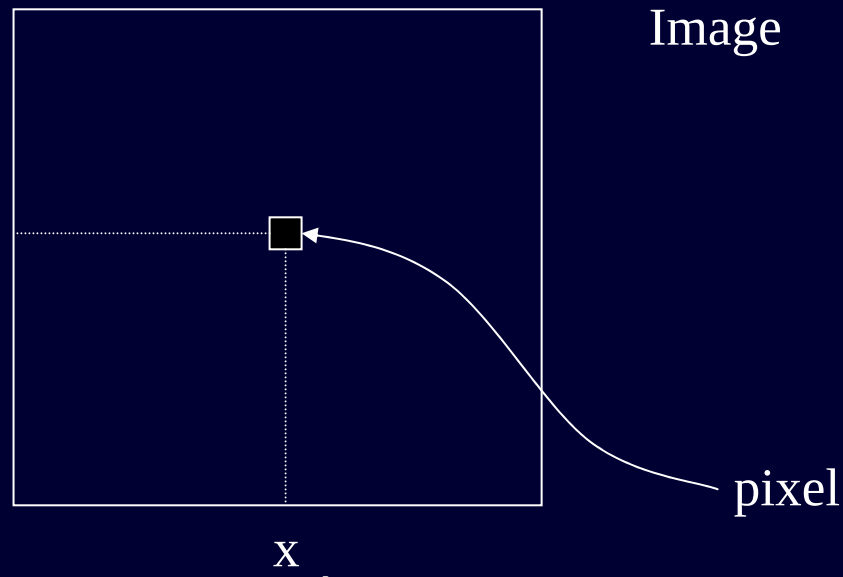
Gestione delle immagini

Acquisizione

- Le immagini possono essere acquisite utilizzando uno scanner
  - È possibile utilizzare scanner simili a quelli usati per il testo, ma con una maggiore risoluzione ed un maggior numero di colori
- Possono essere acquisite direttamente in formato digitale utilizzando una fotocamera digitale
- In casi particolari (ad es. quadri) si possono utilizzare fotocamere digitali di grande formato ed alta risoluzione

Rappresentazione delle immagini

- **Esistono due categorie di rappresentazione delle immagini:**
  - Rappresentazione vettoriale, nella quale l'immagine viene rappresentata tramite elementi grafici quali linee, curve, ecc.
  - Rappresentazione bitmap (o raster) in cui ogni immagine viene rappresentata come una sequenza di punti (pixels)
- **Nella rappresentazione bitmap, ad ogni pixel viene assegnato un valore (nero, bianco, grigio, colore), rappresentato con un codice binario**
- **Si possono applicare tecniche di compressione della codifica per ridurre l'occupazione dell'immagine**



Un'immagine è una funzione di due variabili spaziali f(x,y)

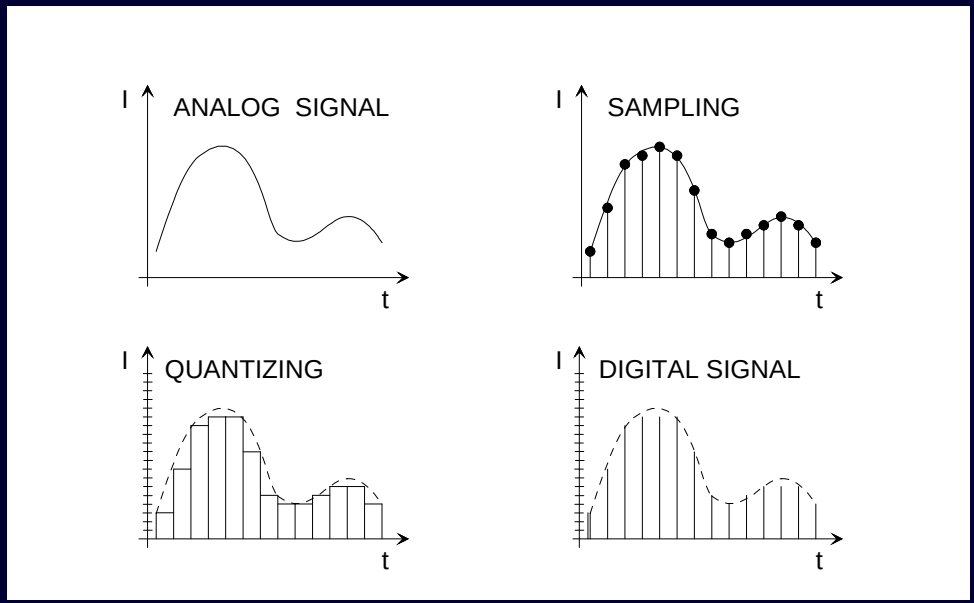
Per un'immagine a colori f(x,y) è un vettore con tre valori, uno per ognuno dei tre colori principali (rosso, verde, blu) corrispondenti all'intensità del pixel (x,y)

$f(x,y)=(f_{rosso}(x,y), f_{verde}(x,y), f_{blu}(x,y))$

Questa codifica viene indicata come RGB (red, green, blu)

Conversione Analogico/Digitale

- **Campionamento**
  - Conversione dei valori corrispondenti alla posizione dei pixel (x,y) da continui a discreti
- **Quantizzazione**
  - Conversione dell'intensità dei livelli di colori da valori continui a valori discreti



Quantizzazione

- I valori della funzione $f(x,y)=(f_{rosso}(x,y), f_{verde}(x,y), f_{blu}(x,y))$ vengono quantizzati per cui la funzione è a valori discreti.
  - Supponiamo che $f_{rosso}(x,y)$ abbia valori tra 0 ed 1
  - Possiamo imporre che i valori permessi siano 0, 0.01, 0.02, …, 0.99, 1
  - In questo caso il passo di quantizzazione è 0.01 mentre i livelli di quantizzazione sono 101
- Normalmente i livelli quantizzazione sono potenze di 2, per cui ad es. per rappresentare 256 livelli di quantizzazione sono necessari 8 bit
- Un'immagine a colori in formato RGB che richiede 8 bit per ogni colore, viene rappresentata con 24 bit per pixel

Descrizione del contenuto delle immagini

- **Per rappresentare il contenuto delle immagini si utilizzano spesso delle proprietà fisiche (dette "features").**
  - Colore
  - Tessitura
  - Forma degli oggetti

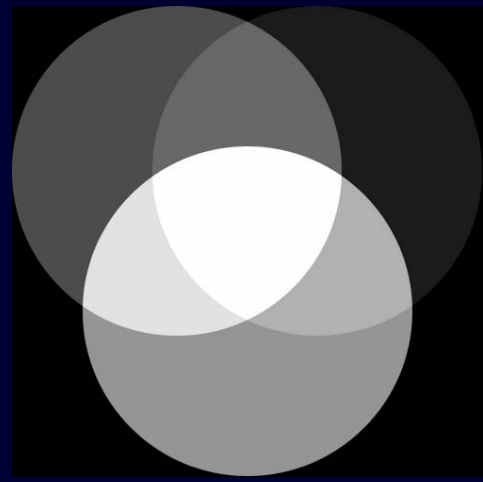
- **L'uso delle features si basa sull'ipotesi che immagini che hanno features simili sono considerate simili dagli utenti (e viceversa).**

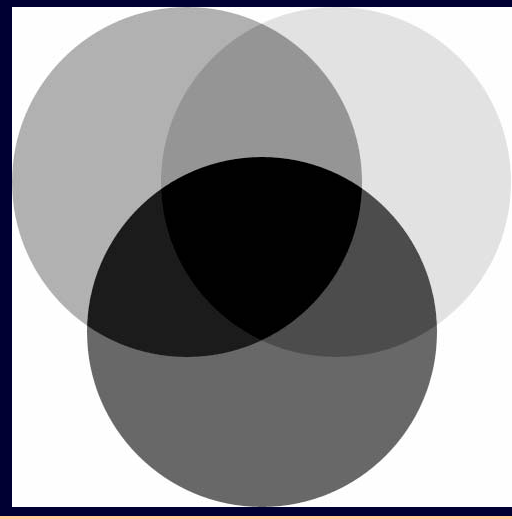
interrogazione

Immagini presenti nella BD

Visione dei colori

- In base alla teoria della visione dei colori, noi vediamo un colore, ad es. il rosso perchè una sorgente luminosa emette luce di colore rosso, oppure perché luce bianca colpisce un oggetto che assorbe il colore complementare del rosso (il ciano). In questo caso il colore si forma per sottrazione.
- Quando i colori si formano considerando sorgenti che emettono luce, diciamo che lo spazio dei colori è additivo, poiché il colore risultante è dato dalla somma dei colori componenti.
- Quando i colori visti dall'occhio sono ottenuti per assorbimento diciamo che lo spazio dei colori è sottrattivo.





Spazi RGB e CMY

- Lo spazio RGB è utilizzato principalmente per la visualizzazione delle immagini su schermo. Lo spazio RGB è additivo.
- Per la stampa di immagini si utilizza lo spazio CMY (ciano, magenta, giallo (yellow)) che è uno spazio sottrattivo.
- RGB e CMY sono dipendenti dal device di visualizzazione e non sono uniformi
- Uno spazio dei colori è uniforme quando colori vicini nello spazio dei colori sono anche simili

Altri spazi dei colori

- **CIE L*a*b e CIE L*u*v**
  - Utilizzano una componente con la luminanza (L) e due componenti cromatiche (*a* e *b*) oppure (*u* e *v*)
  - Sono entrambi uniformi
  - Il primo viene utilizzato per misture sottrattive mentre il secondo viene utilizzato per misture additive
- **HSV**
  - Hue: Tinta del colore
  - Saturation: Quantità di colore
  - Value (Brightness): Quantità di luce

Compressione delle immagini

- **Perché si comprime**
  - Dimensione delle immagini elevata sia per l'archiviazione che per la trasmissione
- **Tecniche di compressione**
  - Algoritmo di compressione e di de-compressione
  - Compressione senza perdita
    - ➔ **La compressione e successiva de-compressione non introducono alcuna modifica all'immagine.**
  - Compressione con perdita
    - ➔ **In alcuni casi si possono accettare perdite nel processo compressione – de-compressione pur di ottenere livelli di compressione maggiori**
- **Il più comune formato di compressione delle immagini è JPEG**
- **TIFF (Tag(ged) Image File Format) è il formato file comunemente più usato per immagini bitmap.**

JPEG [1/2]

- Standard ISO sviluppato alla fine degli anni ’80 (http://www.jpeg.org/). Permette la compressione di immagini fotografiche a livelli di grigio e a colori. Non è adatto per la compressione di immagini grafiche.
- Tecnica di compressione con perdita
- Permette alti livelli di compressione (fino a 20:1)
- Ripetuti processi di compressione-decompressione possono degradare la qualità dell’immagine
- JPEG2000 è l’ultima evoluzione dello standard, che permette migliori livelli di compressione e consente di operare trasmissioni progressive fino ad arrivare, se necessario, ad una compressione senza perdita.

JPEG [2/2]

- **Sequential encoding.** Ogni componente dell'immagine viene codificata in sequenza. L'ordine è quello naturale, da sinistra a destra e dall'alto verso il basso.
- **Progressive encoding.** L'immagine viene codificata in scansioni multiple con qualità sempre maggiore. In questo modo, quando si usano linee di trasmissione lente, è possibile visualizzare inizialmente l'immagine a bassa qualità che poi aumenta durante la trasmissione e decodifica.
- **Lossless encoding.** Esiste anche una modalità di compressione senza perdita, anche se il livello di compressione risulta molto più basso di quello con perdita.
- **Hierarchical encoding.** L'immagine viene codificata a risoluzioni multiple organizzate in modo gerarchico. In questo modo è possibile visualizzare l'immagine a bassa risoluzione, senza la necessità di accedere e decomprimente l'immagine alla massima risoluzione.

TIFF

- Formato file per la rappresentazione di immagini in formato bitmap. Non fornisce supporto alla memorizzazione di immagini in formato vettoriale e testo.
- Utilizzato come formato file in molte applicazioni per la gestione delle immagini (ad es. Photoshop e Paint Shop Pro), da programmi di desktop publishing quali QuarkXPress e InDesign, e da applicazioni di scanning e OCR. Permette di gestire immagini di grandi dimensioni (fino a 4 GB)
- Formato indipendente dalla piattaforma hardware e dal metodo di compressione usato
- Spazi dei colori supportati: Grayscale, Pseudocolore, RGB, YCbCr, CMYK, CIELab
- Diversi tipi di compressione, tra cui: PackBits, Lempel-Ziv-Welch (LZW), CCITT Fax 3 & 4, JPEG

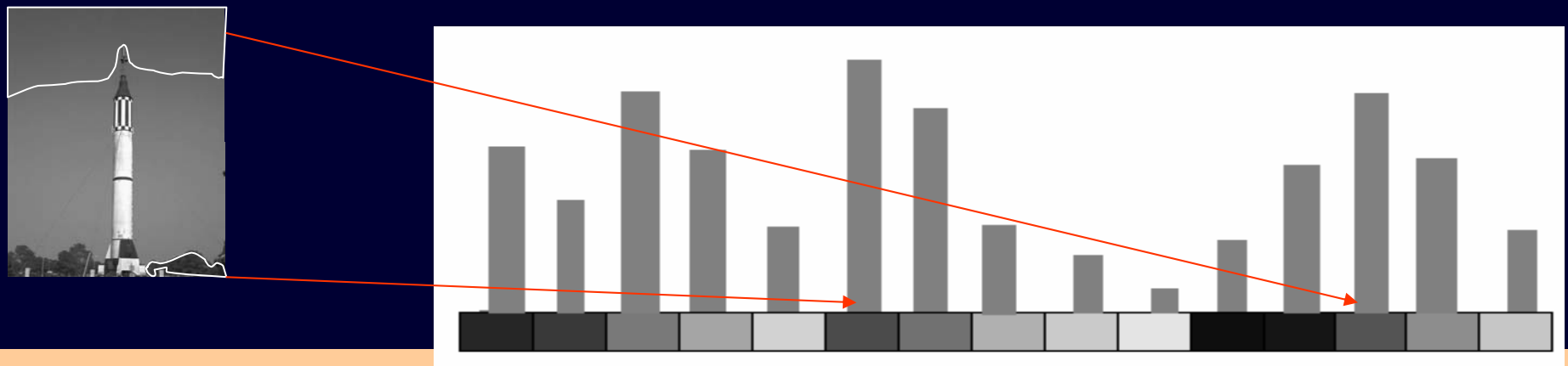
Indicizzazione delle immagini

- Indicizzazione automatica basata su un insieme di caratteristiche (“features”) delle immagini
  - Colore
  - Tessitura
  - Forma degli oggetti che compongono l’immagine
  - Organizzazione spaziale
  - In questo caso la ricerca sarà basata su una misura della similarità tra l’immagine e l’interrogazione
- Indicizzazione manuale
  - Descrizione semantica
  - Oggetti che compongono l’immagine

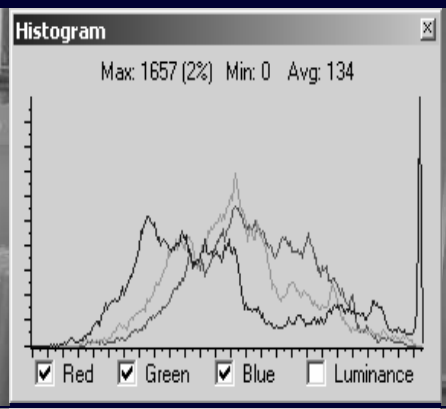
Istogrammi dei colori

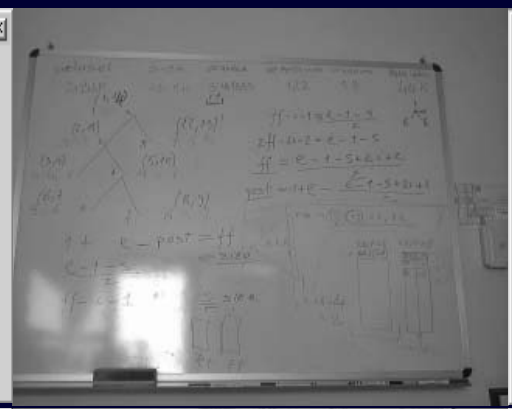
- **I colori contenuti in un'immagine sono rappresentati utilizzando un istogramma dei colori**

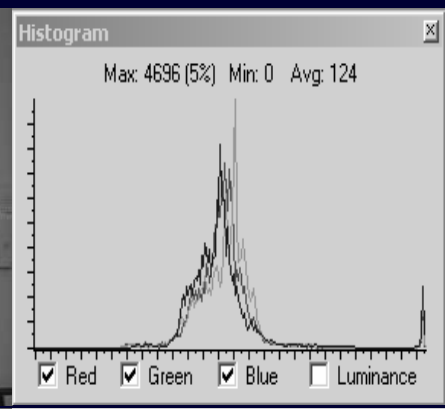
- **Istogrammi dei colori**
  - Lo spettro dei colori è suddiviso in *n* contenitori, ognuno dei quali rappresenta un colore.
  - I colori dei pixel nell'immagine sono approssimati ad uno dei colori di un contenitore.
  - Il valore contenuto in ogni contenitore è proporzionale al numero di pixel che hanno il colore di quel contenitore.

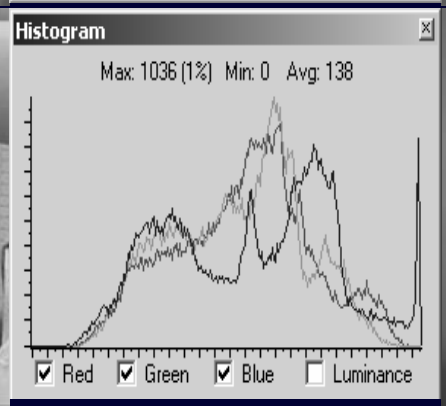


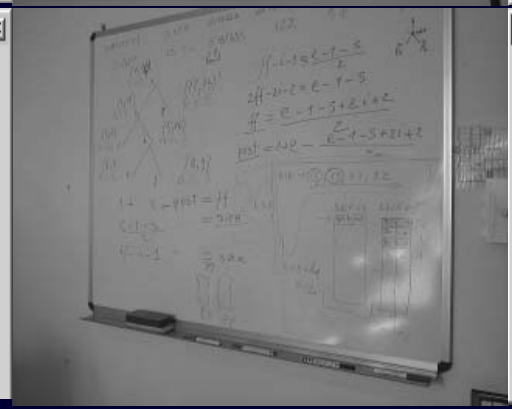
Istogrammi di colore

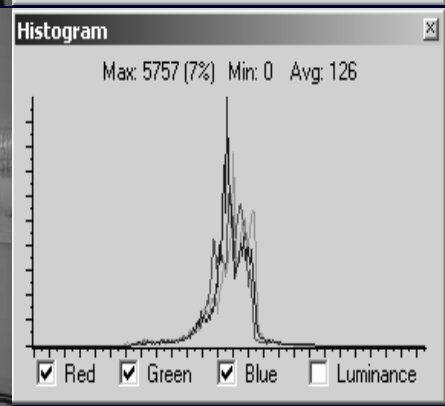












Misura della similarità tra due istogrammi di colore

- **Indichiamo con H_I l'istogramma dei colori di una immagine e con H_Q l'istogramma dei colori di una interrogazione. Quindi H_I ed H_Q sono vettori di *n* elementi**

- **La similarità tra le due immagini I e Q si calcola tramite la similarità tra i due istogrammi dei colori (intersezione degli istogrammi dei colori)**

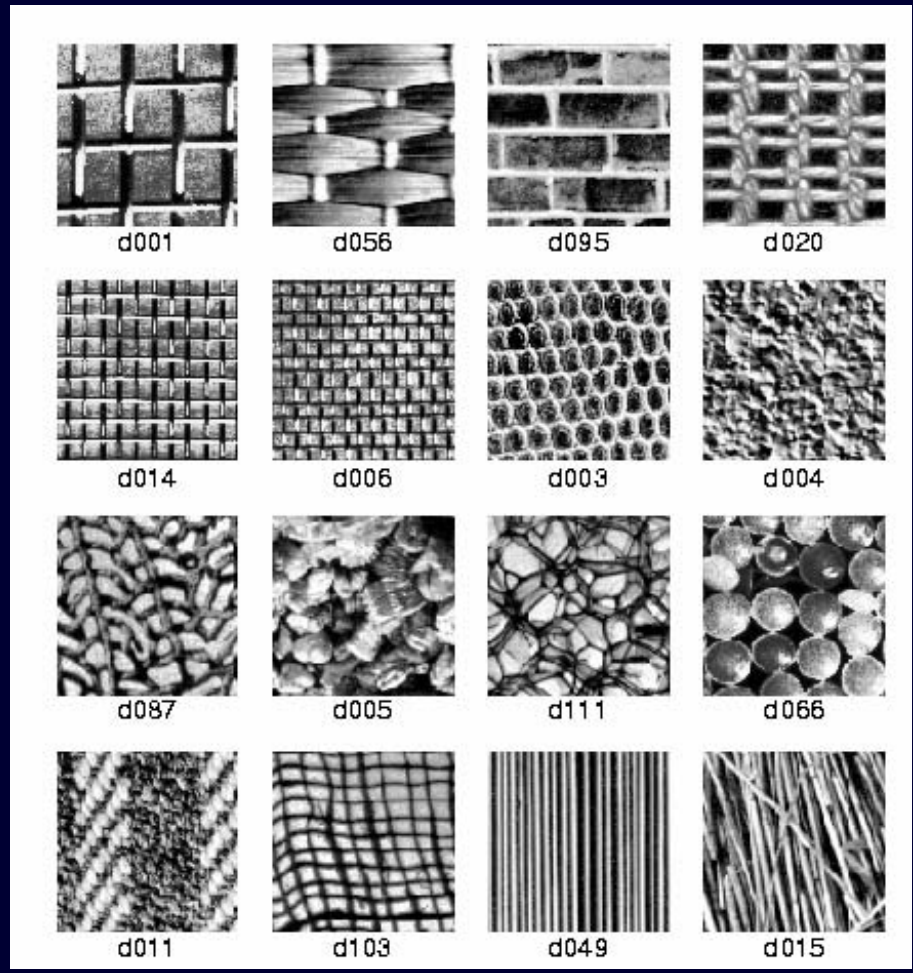
$$sim(H_I, H_Q) = \frac{\sum_{i=1}^{n} \min(H_{I_i}, H_{Q_i})}{\sum_{i=1}^{n} H_{Q_i}}$$

Proprietà dello spazio dei colori

- **Le proprietà che uno spazio dei colori deve avere per essere adatto all'indicizzazione e ricerca di immagini sono:**
  - Uniformità
    - ➔ **Colori che sono vicino nello spazio dei colori devono essere percepiti come simili.**
  - Completezza
    - ➔ **Lo spazio dei colori deve permettere di rappresentare tutti i colori che vengono percepiti.**
  - Compattezza
    - ➔ **Non vi deve essere ridondanza.**
- **Lo spazio RGB non è uniforme, per cui per la ricerca di immagini si usano altri spazi (ad es. HSV)**

Indicizzazione delle immagini

- **Textures:**
  - Pattern omogenei
  - Organizzazione spaziale dei pixel
  - Solo con i colori non è possibile descrivere tutte le caratteristiche delle immagini

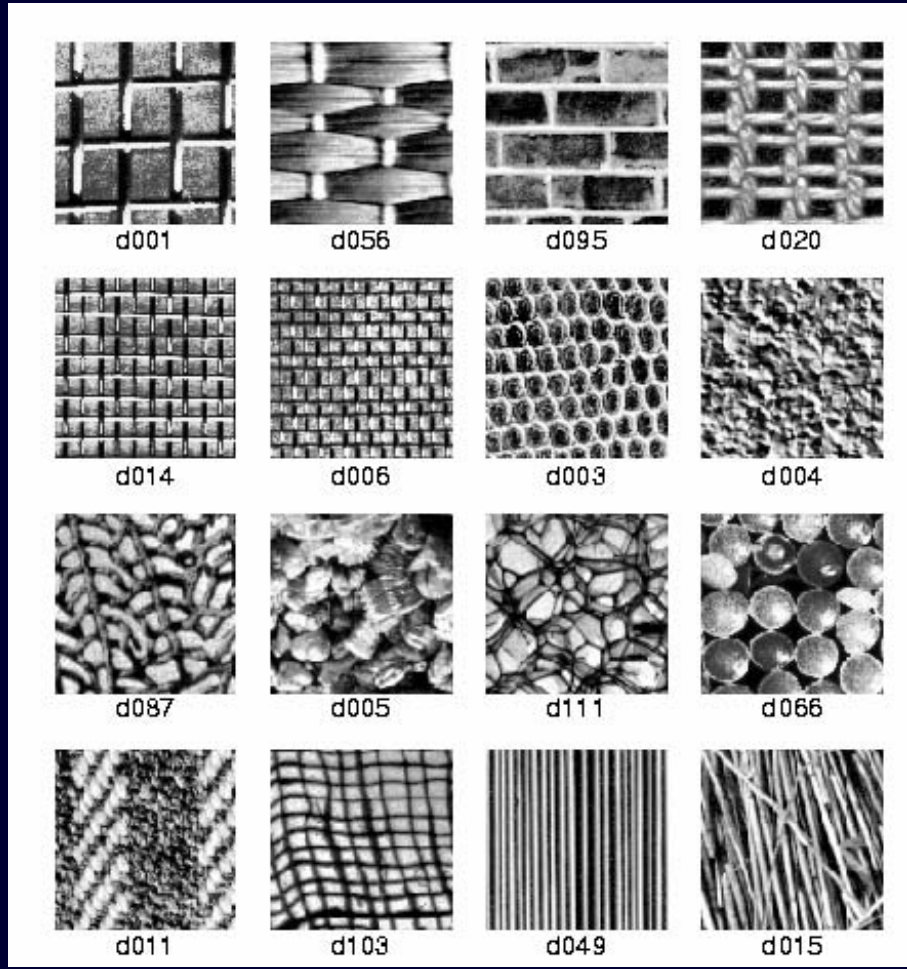


Textures

- **Le Textures si ottengono utilizzando metodi statistici per rappresentare la distribuzione spaziale dell'intesità dei pixel nell'immagine.**
- **Esistono diversi metodi per la rappresentazione delle Textures**
- **Le Texture possono essere rappresentate come istogrammi (vettori)**

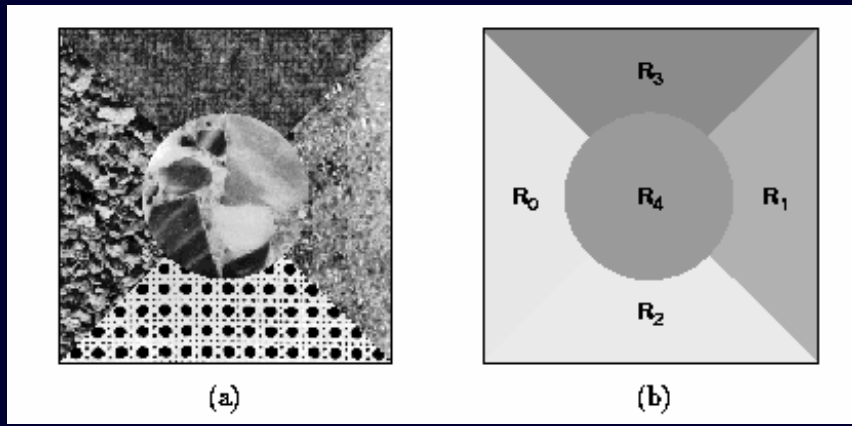
Textures

- **Le features più utilizzate sono le Tamura features:**
  - Contrasto
    - ➔ **Distribuzione dell'intensità dei pixel**
  - Coarseness
    - ➔ **Granularità della tessitura**
  - Direzionalità
    - ➔ **Direzione dominante della tessitura**



Forme

- **Forme:**
  - Estrazione di regioni
  - Segmentazione

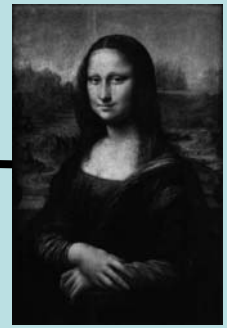


Feature globali e locali

- **Gli istogrammi di colore e le textures possono essere calcolati anche per le regioni individuate nelle immagini, oltre che per l'intera immagine**
  - Features globali vengono utilizzate per la ricerca delle immagini
  - Features locali vengono utilizzate per la ricerca di regioni contenute nelle immagini. Questo tipo di ricerca risulta più precisa perché si possono trascurare parti non significative dell'immagine.
- **Le relazioni spaziali tra le regioni permettono di specificare ulteriori condizioni nell'interrogazione.**
  - Si possono cercare gli oggetti che contengono oggetti la cui forma è di un certo tipo e che si trovano in una determinata relazione spaziale tra loro.

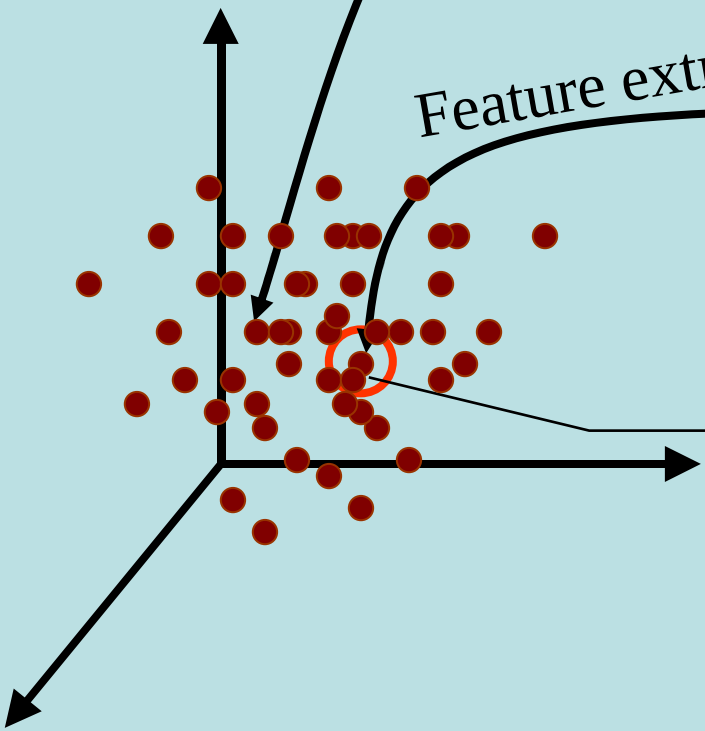
Similarity search

Feature extraction



Feature extraction

Query image



Query neighbourhood

Interrogazione

Risultato dell'interrogazione

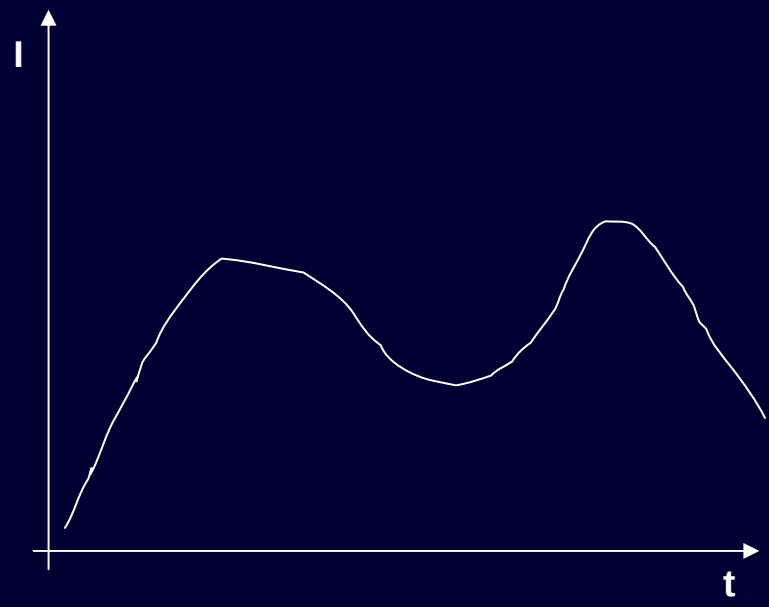
Gestione dell’audio

Vari “tipi” di audio

- **La gestione dell’audio può variare a seconda del tipo di audio che viene gestito.**
  - Parlato
    - ➔ **In questo caso è possibile considerare l’utilizzo di tecniche per la trasformazione del parlato in testo.**
    - ➔ **La qualità del riconoscimento può essere elevata per sistemi speaker-dependent.**
    - ➔ **Qualità accettabili ai fini del retrieval anche per sistemi speaker-independent**
  - Suono
    - ➔ **Un qualunque segnale audio con frequenze nel range dell’udito umano**
    - ➔ **L’indicizzazione è basata su proprietà acustiche quali *bightness, pitch, loudness***
  - Musica
    - ➔ **Si tiene conto dei diversi strumenti musicali utilizzati, dei vari tipi di suoni prodotti, degli effetti musicali, ecc.**

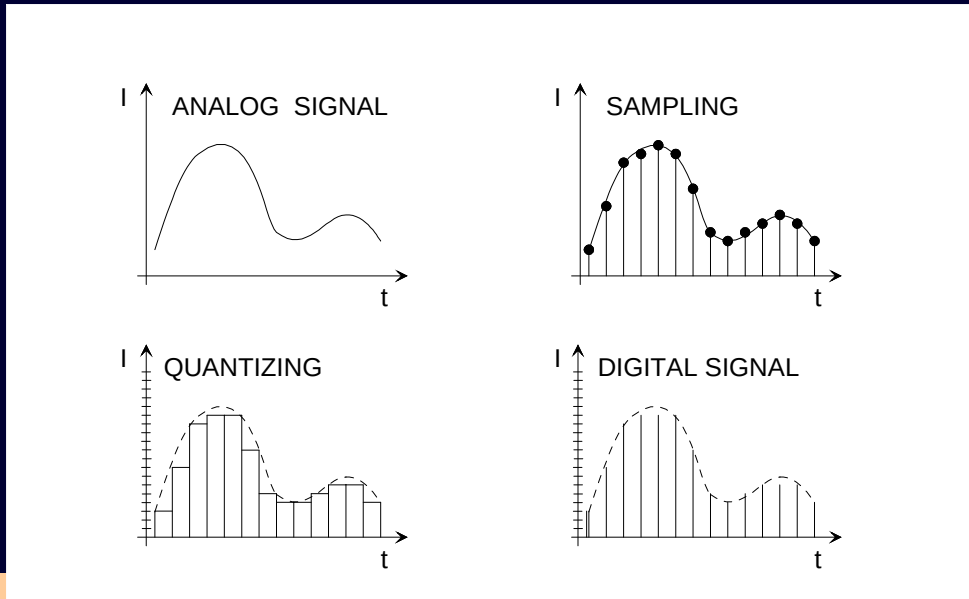
Rappresentazione dell'audio

- **Il segnale audio è costituito da vibrazioni dell'aria caratterizzate da una determinata intensità che varia nel tempo, come rappresentato in figura.**
- **Il segnale può essere rappresentato come una funzione I = f(t), dove I rappresenta l'intensità del segnale e *t* è il tempo.**
- **Analogamente alle immagini, per poter rappresentare un segnale audio in un computer dobbiamo campionarlo ad intervalli regolari di tempo e quantizzare i valori dell'intensità (si passa da un segnale continuo ad uno discreto).**
- **La conversione del segnale audio da analogico a digitale viene effettuata da componenti hardware specifici.**



Conversione Anologico/Digitale

- **Campionamento**
  - Conversione dei valori del tempo *t* da continui a discreti. Si prende un insieme limitato di campioni del segnale audio.
- **Quantizzazione**
  - Conversione dell'intensità dei livelli di intensità audio da valori continui a valori discreti



Qualità del suono

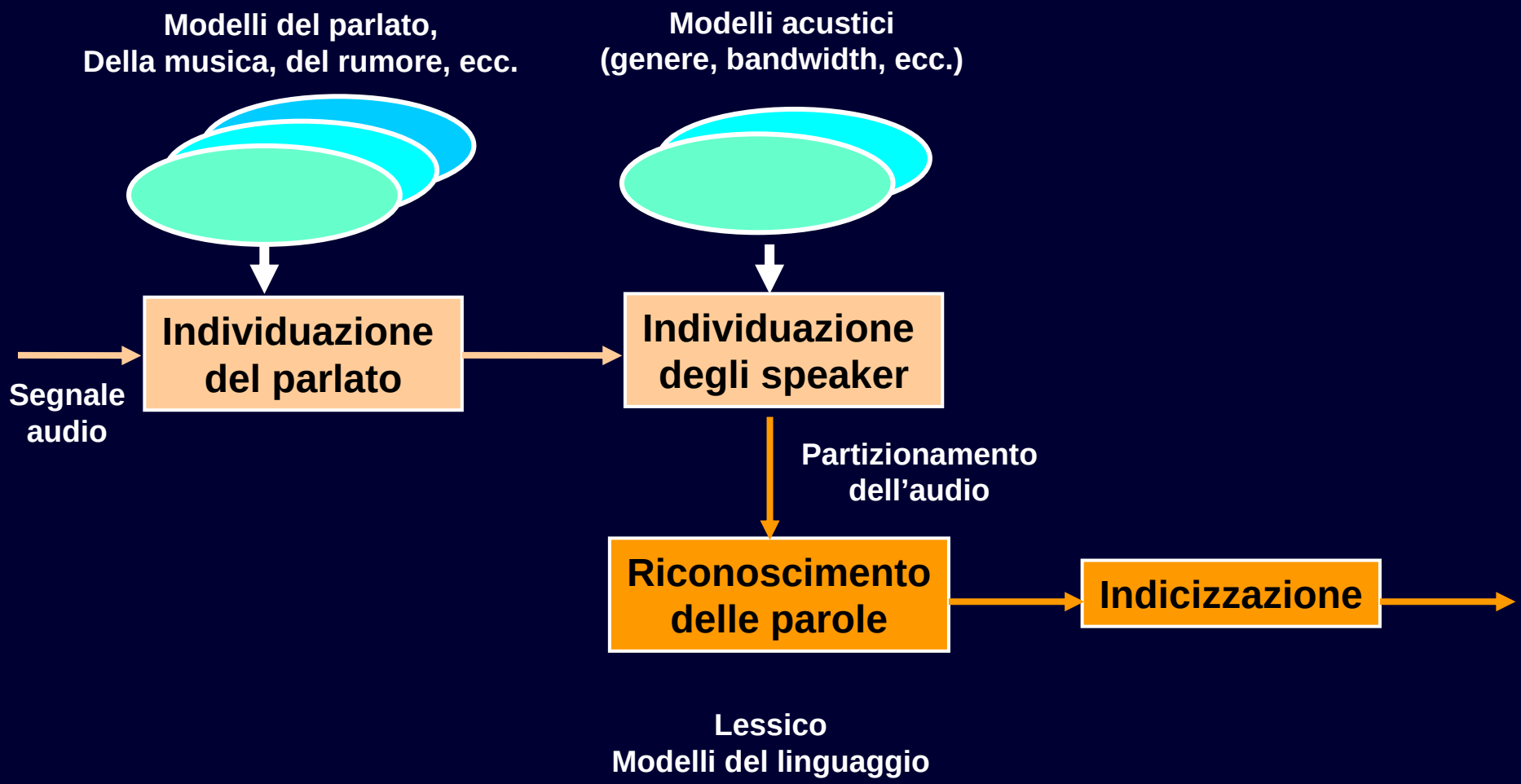
- **La qualità della rappresentazione del suono digitale dipende dal numero di campioni per unità di tempo e dal numero di bit utilizzati per rappresentare ogni campione.**

- **Teorema di Nyquist: Per rappresentare in modo preciso un segnale di frequenza *f* è necessario effettuare il campionamento a frequenza > 2*f***
  - Ad esempio, per l'orecchio umano che percepisce suoni di frequenza tra 20-20000 Hz, è necessario un campionamento a frequenze maggiori di 40000 Hz
  - Il suono dei CD è campionato a 44.1 kHz
  - Suono di elevata qualità può arrivare a compionamenti di 96kHz

Qualità del suono

- **Il numero di bit utilizzato per ogni campione determina il numero di livelli di intensità rappresentabili**
- **Valori tipici sono**
  - 8 bit per campione: 256 livelli
  - 16 bit per campione: 65636 livelli
- **Qualità dei CD audio**
  - 16 bit per campione
  - 44100 campioni per secondo
  - 2 canali audio (suono sterofonico)
  - 172,3 kB per secondo, 10 MB per minuto
  - Evoluzione per il suono Dolby che richiede 6 canali audio, 24 bit per campione, 96 kHz di frequenza di campionamento

Riconoscimento del parlato



Riconoscimento del parlato

- **Partizionamento dell'audio**
  - Rimuove tutto ciò che non è "parlato" (musica, rumore, ecc.)
  - Identifica i diversi speakers ed individua il momento in cui ognuno parla
  - Utilizza modelli acustici specifici alle diverse condizioni (ad es. uomo/donna, vicino/lontano)
  - Produce un'annotazione dell'audio
- **Riconoscimento delle parole**
  - Utilizzo di modelli acustici e fonetici
  - Utilizzo di modelli linguistici
- **Il risultato del riconoscimento è un testo annotato (tempi, speaker) con un certo numero di errori (WER – Word Error Rate)**

Feature del suono

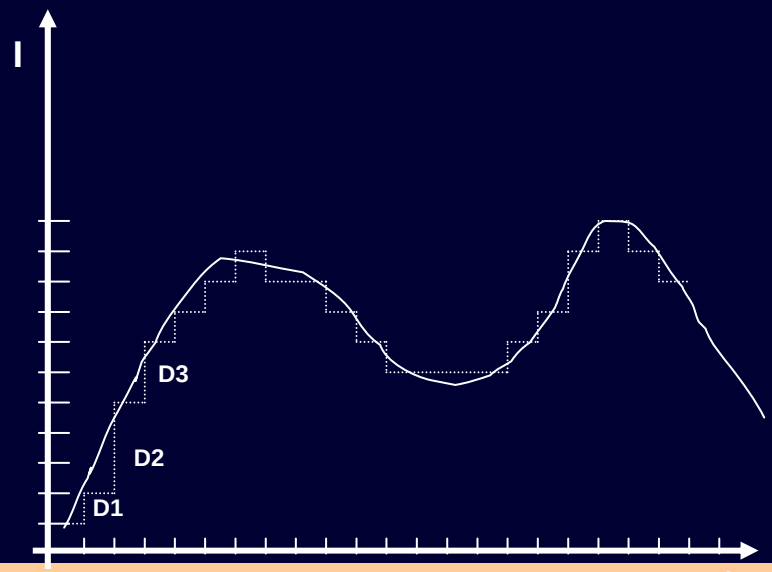
- **Feature acustiche**
  - Loudness: Intensità del suono, calcolata usando l'energia del segnale. Viene espressa in decibel.
  - Spettro di potenza: potenza del segnale per tutte le frequenze
  - Brightness: misura della quantità di alte frequenze presenti nel segnale
  - Bandwidth: larghezza della banda di frequenze
  - Tono (Pitch): legato alla percezione del suono

Feature del suono

- **Feature soggettive/semantiche**
  - Sono derivate dalle feature acustiche
  - Timbro: profilo armonico del suono. Permette di distinguere i suoni emessi da strumenti diversi.
  - Ritmo: variazioni regolari del suono.
  - Eventi: la descrizione musicale basata su spartiti è espressa sotto forma di eventi.
  - Strumenti: tipo della sorgente audio.

Tecniche di compressione dell'audio

- **DPCM (Differential PCM)**
  - Confronta campioni di segnale adiacenti e memorizza solo la differenza tra un campione ed il successivo.
  - Richiede meno bit per campione ma presenta problemi per segnali con forti variazioni



Tecniche di compressione dell'audio

- **ADPCM (Adaptive DPCM)**
  - La differenza tra due campioni può essere variabile.
  - Riduce il problema della saturazione
  - Si riduce il problema della propagazione degli errori attraverso un restart periodico della sequenza
  - Ad esempio, supponiamo di usare normalmente 16 livelli di quantizzazione, quindi 4bit. Se la differenza è maggiore di 16 (ad es. 20), si useranno 5 bit, per poi tornare a 4 bit quando le variazione del segnale si normalizzano.

MP3

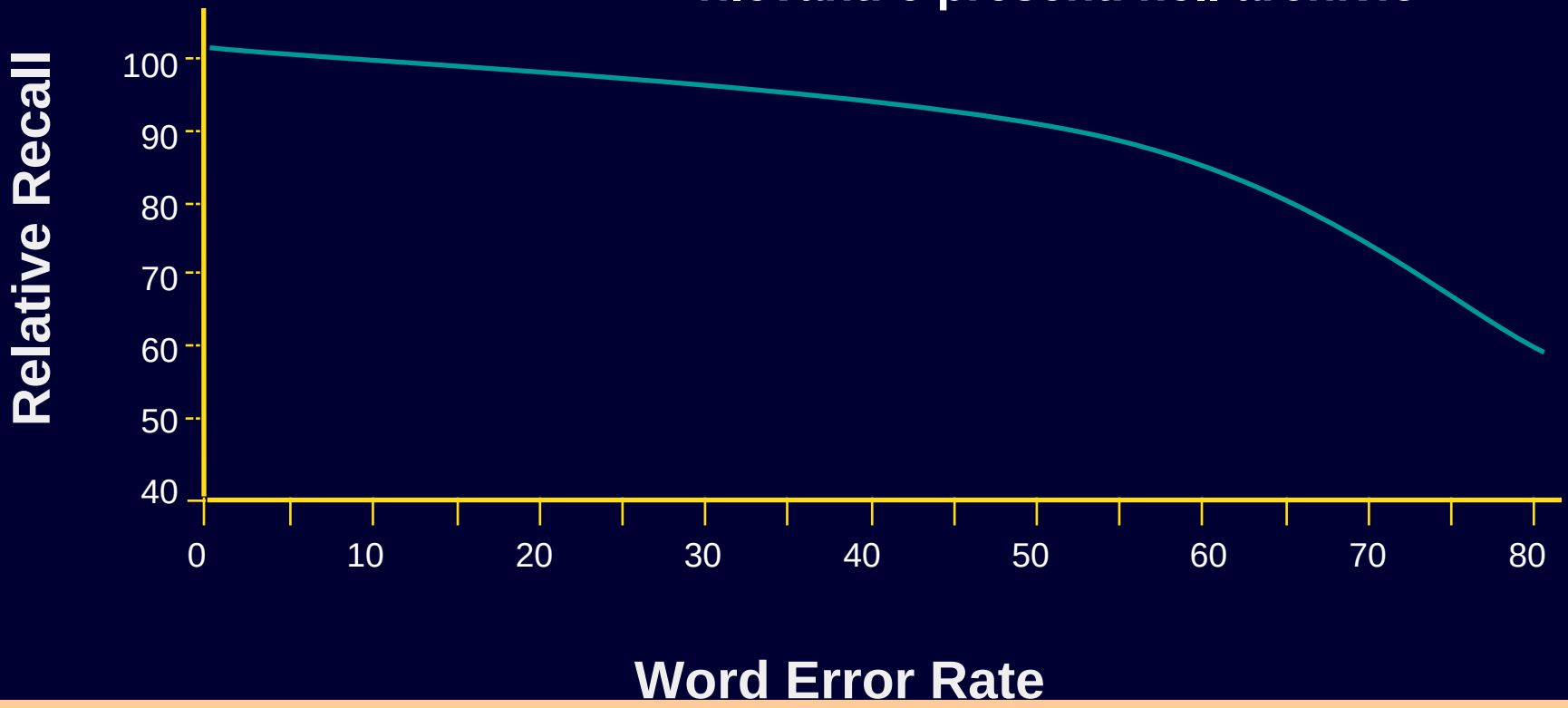
- **Formato di compressione audio (con perdita), basato sull'MPEG-1 Audio Layer 3.**
- **Basato sulla codifica PCM (Pulse Code Modulation), scarta porzioni dell'audio considerate poco importanti per la qualità della percezione umana**
- **Utilizza bit-rate (numero di bit per ogni secondo di audio) variabili tra 32 kbit/s e 320 kbit/s**
  - Per confronto il bit-rate dell'audio CD è 172 KByte/s
  - Si è verificato che tra 224-320 kbit/s si ha un eccellente qualità del suono prodotto, anche se per compressioni a 128-192 kbit/s la qualità è ancora buona
- **Possibilità di Variable Bit Rate (VBR)**
  - L'audio viene suddiviso in frame e per ognuno è possibile utilizzare un bit rate diverso a seconda del tipo di suono presente.

Retrieval del parlato

- **Conversione dell'audio in testo (speech to text)**
- **La trascrizione automatica presenta errori**
  - La qualità della trascrizione si misura con il Word Error Rate – rapporto tra le parole errate rispetto a quelle riconosciute correttamente
- **Il W.E.R. può variare a seconda della qualità dell'audio e della complessità (ad es. molti speaker, rumori di fondo, ecc.)**
  - Single speaker – W.E.R. fino al 1-5%
  - Audio degradato, speaker independent – W.E.R. 40-50%

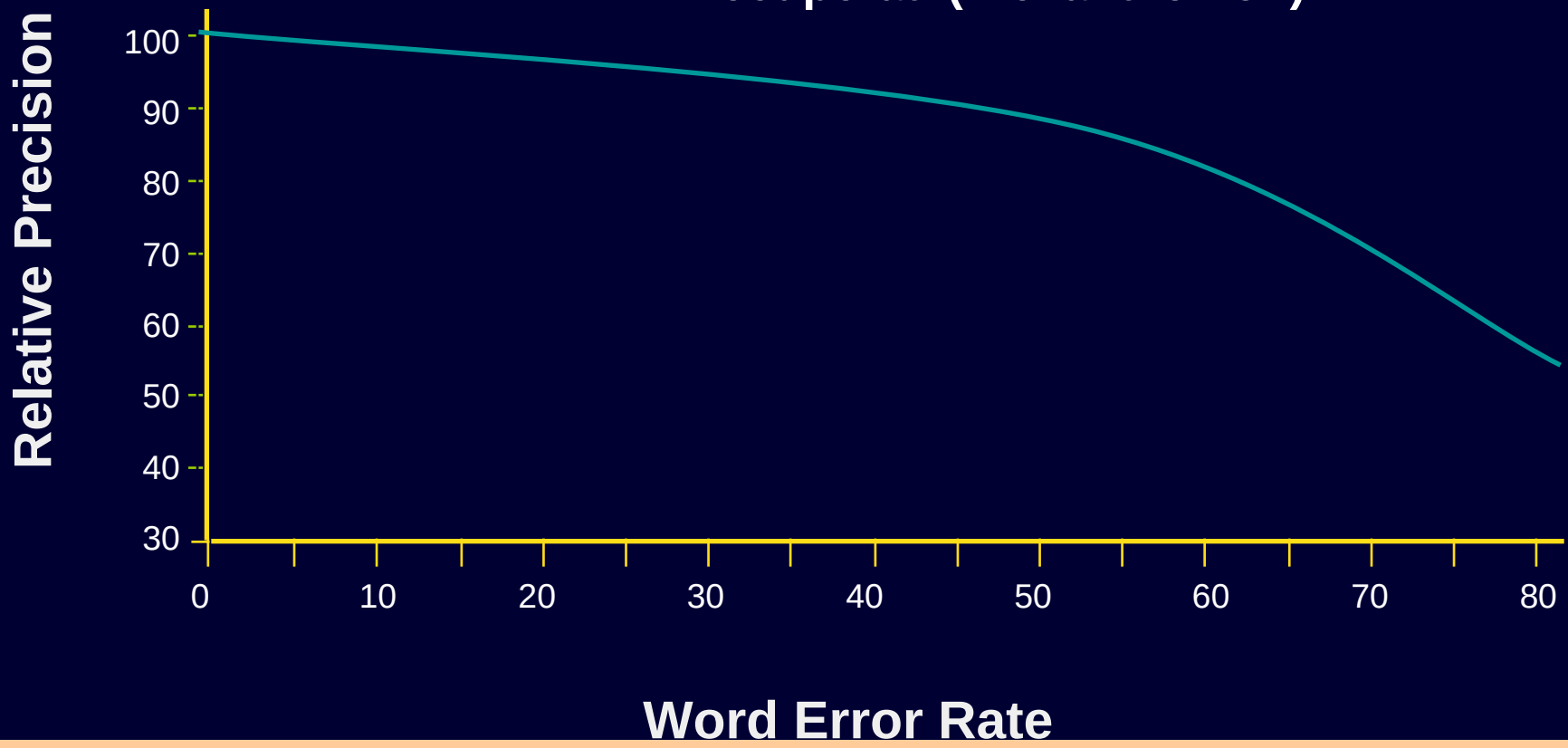
Recall in funzione del W.E.R.

Recall: rapporto tra il numero dei documenti rilevanti recuperati e quelli rilevanti e presenti nell'archivio



Precisione in funzione del W.E.R.

Precisione: rapporto tra il numero dei documenti rilevanti recuperati e quelli recuperati (rilevanti e non)



Audio retrieval

- **Feature based**
  - Sulla base delle feature estratte dall'audio si creano degli istogrammi con un procedimento analogo a quello usato per le immagini
  - La similarità (o dissimilarità) viene misurata utilizzando una distanza di tipo Euclideo
  - Le interrogazioni si possono formulare utilizzando parti di brani musicali o sintetizzatori musicali
- **Retrieval by Humming**
  - Una diversa modalità di formulazione delle interrogazioni
  - L'utente può accennare il motivo del brano richiesto