



UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

*Appunti di
Sistemi Informativi Multimediali*

Anno 2021

Valentino Bocchetti

Contents

1	Informazioni	3
1.1	Organizzazione del corso e dell'esame	3
2	Struttura del corso	3
3	Analogico e digitale	4
4	Archiviazione dei dati	5
4.1	Query	5
5	Media e Multimedia	5
5.1	RDBMS (Relational DBMS)	6
5.2	SQL (Structured Query Language)	6
5.3	BLOB (Binary Large Objects)	6
5.4	IR (Information Retrieval)	6
5.5	MIRS	7
6	Tipologia e formati dei dati multimediali	8
6.1	Plain text	8
6.2	Testo strutturato	8
6.3	Testo compresso	8
6.4	Algoritmi di compressione	9
7	L'Audio	14
7.1	Rappresentazione digitale dell'audio	15
7.2	Teorema di Nyquist (importante)	17

7.3	Scelta della frequenza di campionamento	17
7.4	Companding	19
7.5	Predictive Coding	19
7.6	Mpeg Audio	20
7.7	Audio sintetico (Midi)	20
8	Le immagini	21
8.1	Percezione	21
8.2	Rappresentazione dei colori	21
8.3	Sintesi dei colori	21
8.4	Spazio RGB (3D)	23
8.5	Spazio CIE normalizzato (2D)	24
8.6	Rappresentazione SPD (Spectral Power Distribution)	25
8.7	Funzione di trasferimento	25
8.8	Correzione gamma	25
8.9	Problemi di indicizzazione e ricerca legati allo spazio di colore	27
8.10	Le origini	27
8.11	Grafica Raster e Vettoriale	27
8.12	Super-Resolution (SR)	28
8.13	Immagini GrayScale (scala di grigio)	29
8.14	Immagini a colori	30
8.15	Fondamenti della compressione	30
8.16	Compressione LOSSY	30
8.17	Serie di Fourier	32

8.18 Analisi spettrale	33
8.19 Immagini JPEG (Joint Photographic Experts Group)	33
8.20 Immagini Frattali	34
9 Il Video	35
9.1 Frame rate (velocità di scorrimento)	35
9.2 Codifiche	35
9.3 Compressione video	35
9.4 MPEG (Motion Picture Expert Group)	36
9.5 Documenti multimediali compositi (SGML)	36
9.6 PDF (Portable Document Format)	36
9.7 Requisiti di memoria e larghezza di banda	37
10 Progetto di Database multimediali	38
10.1 Architettura dei MIRS	38
10.2 Modello dei dati	39
10.3 Requisiti	40
10.3.1 Layer oggetto	41
10.3.2 Layer TIPO	41
10.3.3 Layer FORMATO	41
10.4 VIMSYS	41
10.5 Modello di un video	42
10.6 Interfaccia utente	43
10.6.1 Popolazione del DB	43
10.7 Fase di ricerca	43

10.7.1	Tipi di ricerca	43
10.7.2	Estrazione delle feature	43
10.8	Indicizzazione dei dati	44
10.8.1	Misure di similarità	44
10.9	Garanzie sulla QoS (Quality of Service)	44
10.10	Multimedia Data Compression	45
10.11	Standard di rappresentazione dei dati	46
11	Il Testo	46
11.1	Differenze tra IR e DBMS	47
11.2	Indicizzazione automatica e modello booleano per il retrieve	47
11.2.1	Struttura file	48
11.2.2	Indicizzazione automatica	49
11.3	Retrivial con modello nello spazio vettoriale	50
11.3.1	Calcolo della similarità	51
11.3.2	Tecniche basate su Relevance feedback	51
11.3.3	Altri modelli per il Retrieval	52
11.4	Misurazione delle prestazioni	53
11.5	Tecniche di IR a confronto	54
11.6	Motori di ricerca www	54
11.7	Modello di comunicazione	55
11.8	Il web	55
11.8.1	Crawler (spider)	55
12	Indicizzazione e Recupero dell'audio	55

12.1	Approccio generale	56
12.2	Proprietà e caratteristiche principali dell'audio	56
12.2.1	Caratteristiche derivabili dal Time Domain	56
12.2.2	Caratteristiche derivabili dal Dominio delle Frequenze	58
12.3	Classificazione dei segnali audio	60
12.4	Metodi di classificazione dell'audio	60
12.4.1	Classificazione Step-By-Step	60
12.4.2	Classificazione basata su caratteristiche vettoriali	61
12.5	Concetti base dell'ASR (Automatic Speech Recognition)	61
12.6	Tecniche basate sul <i>Dynamic Time Warping</i> (DTW)	61
12.7	Tecniche basate su Hidden Markov Models (HMM)	62
12.8	Tecniche basate su Reti Neurali Artificiali (ANN)	63
12.8.1	Elaborazione di UN PE	64
12.8.2	Funzioni di elaborazione	64
12.9	Tecniche di identificazione dello speaker	65
12.10	Relazioni tra audio e altri media	65

1 Informazioni

Sito di riferimento

Mail: walterbalzano@gmail.com

Registrazione al corso obbligatoria

1.1 Organizzazione del corso e dell'esame

Prova scritta + prova orale

Per un appuntamento inviare una mail

2 Struttura del corso

- Tipi di dati, formati e standard
- Analisi della percezione
- Indicizzazione e Ricerca
- Misure di efficacia e sicurezza
- Linguaggi per i metadati
- Ontologie Multimediali
- Multimedia Data Mining
- Definizioni
- Struttura di un GIS
- Aggregazione dei dati
- Cartografia digitale
- Modelli di GeoReferenziazione
- DATUM Geodetici e mappe
- Principali metodi di analisi: Voronoi, NNI, Variogrammi, Interpolazione
- La trilaterazione
- Segmenti GPS: Spazio, controllo, uso
- Sinc. Tempo/Spazio
- Il modello D.O.P.
- Errori relativistici e fonti minori
- Varianti GPS: A-GPS e D-GPS
- Tracking-Logs: file gpx
- Applicazioni ed esempi

3 Analogico e digitale

L'analogico è sostanzialmente un'onda, che riproduce esattamente una vibrazione meccanica

Il digitale invece è sostanzialmente una sequenza di numeri



Digitalizzazione

- Fase 1 → digitalizzazione dei dati (si utilizzano varie modalità e strumenti (es scanner).
- Fase 2 → Management dei dati

4 Archiviazione dei dati

Definizione ed evoluzione dei DB

	DBMS (DataBase Management Systems)	MMDBMS (MultiMedia DBMS)
DOMINIO	AlfaNumerico	MultiMediale
MATCH	Testuale esatto	Similitudine Caratteristiche
PRESTAZIONI	<input checked="" type="checkbox"/> Velocità di risposta	<input checked="" type="checkbox"/> Velocità di risposta <input checked="" type="checkbox"/> Capacità di confronto caratteristiche

M.I.R.S. → Multimedia Indexing and Retrieval Systems (sistema di memorizzazione e recupero dei dati) Gli indici sono una informazione di sintesi

4.1 Query

Richieste al sistema per una risposta precisa

Alcune tipologie di query sono quelle di tipo **spazio/temporali**, **similitudine** e **concetto** o una combinazione più o meno complesse delle 3

5 Media e Multimedia

Media: tipi di informazione o rappresentazione come dati alfanumerici, immagini, audio, video. Una classificazione può essere basata su:

- Formati fisici
- Relazioni con dimensione temporale
 - Statici
 - * Essi non possiedono una dimensione temporale e quindi il loro contenuto e significato non hanno alcuna dipendenza dal tempo (es grafici e le immagini)
 - Dinamici
 - * Possiedono una dimensione temporale ed il loro significato e correttezza dipende dalla velocità con cui vengono rappresentati (es animazioni, audio, video)

Multimedia: raccolta di tipi di media usati contemporaneamente

5.1 RDBMS (Relational DBMS)

Costituiscono la tipologia maggiormente diffusa di DBMS Presentano una struttura tabulare (Righe x Colonne):

- Righe → elementi di informazione o RECORD
- Colonne → Attributi o CAMPI

5.2 SQL (Structured Query Language)

È utilizzato per la creazione delle tabelle, per l'inserimento ed il reperimento degli elementi dal DB

5.3 BLOB (Binary Large Objects)

È una stringa binaria di lunghezza variabile che si utilizza per la memorizzazione di dati in formato binario come immagini (a differenza dell'SQL non è possibile formulare un pattern di ricerca)

5.4 IR (Information Retrieval)

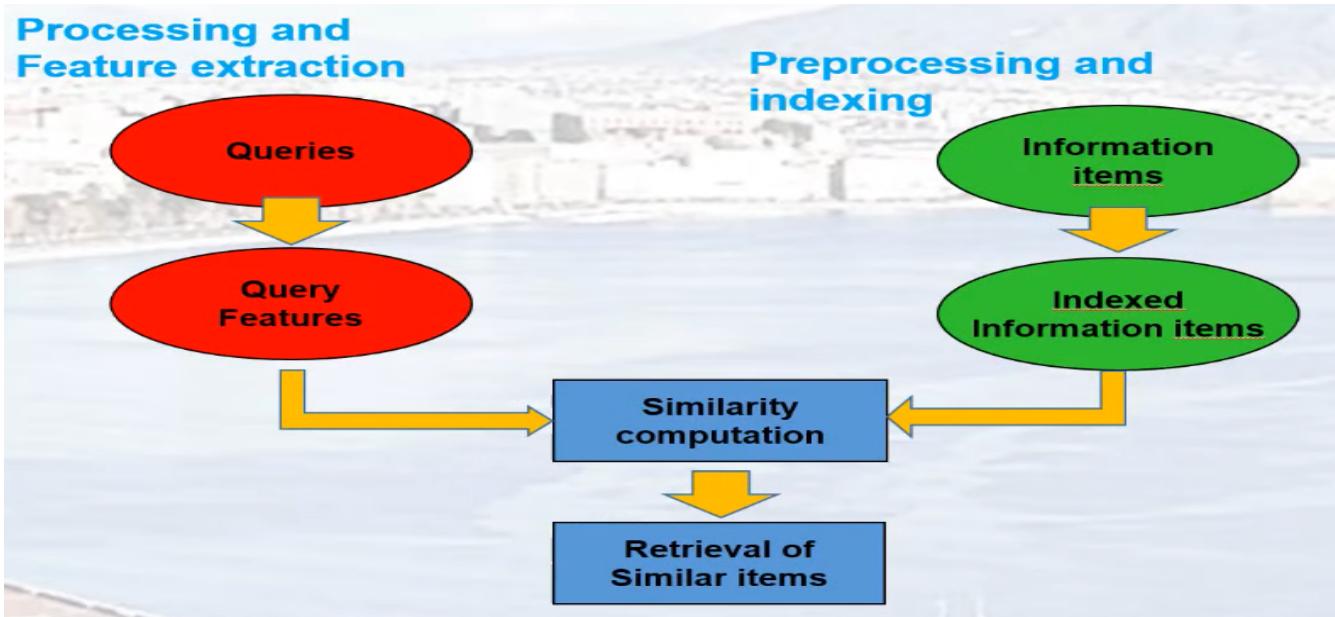
Costituiscono una parte rilevante tra i sistemi di recupero delle informazioni Operano prevalentemente in modalità testuale, ma possono essere utilizzate in ambito multimediale perchè:

- Va considerata la grande mole di documenti testuali già presenti
- Attraverso il testo è possibile fornire una descrizione ai vari media

Limiti degli IR

- L'annotazione è per la maggior parte manuale (dispendiosa nel tempo)
- L'annotazione è soggettiva e incompleta
- Le query sono esclusivamente testuali
- Molti oggetti sono difficili da descrivere

5.5 MIRS



Si prendono i media che vengono processati, trattandoli con lo scopo di indicizzarli

Proprietà:

- Query sui **metadati**
- Query su **annotazioni**
- Query su **modelli di dati o caratteristiche**
- Query su **esempi**
- Query su **applicazioni specifiche**

Nel caso dell'applicazione specifica distinguiamo anche la ricerca in:

- Medicina
- Sicurezza
- Didattica
- Editoria
- Intrattenimento
- Copyright

6 Tipologia e formati dei dati multimediali

Obiettivi	Dominio
Elaborare	Testi
Indicizzare	immagini
Memorizzare	Grafici
Trasmettere	Audio
Presentare	Video
...	...

L'obiettivo principale di un MIRS è quello di **indicizzare** e **ricercare** i dati multimediali tra cui testi, grafica, immagini, audio, video.

È pertanto molto diverso da un classico DBMS; pertanto è necessario conoscere la struttura, le caratteristiche e le peculiarità dei dati multimediali

6.1 Plain text

Testo che non contiene altri attributi (Extended ASCII, UNICODE)

6.2 Testo strutturato

Testo che presenta una formattazione (presenta quindi degli attributi) con diversi standard (.tex, .pdf, .org)

6.3 Testo compresso

Testo che sfruttando una ridondanza basata sulle ripetizioni e delle occorrenze permette di racchiudere un insieme di caratteri (con una compressione del tipo **LOSSLESS**). Con lo scopo di comprimere (e quindi occupare meno spazio) ci si trova a scegliere tra un tipo di compressione con o senza perdita

Si dovrà fare uso di algoritmi di compressione. Tra i metodi di compressione ricordiamo:

- Huffman
- Run length
- Lempel-Ziv-Welch (LZW)

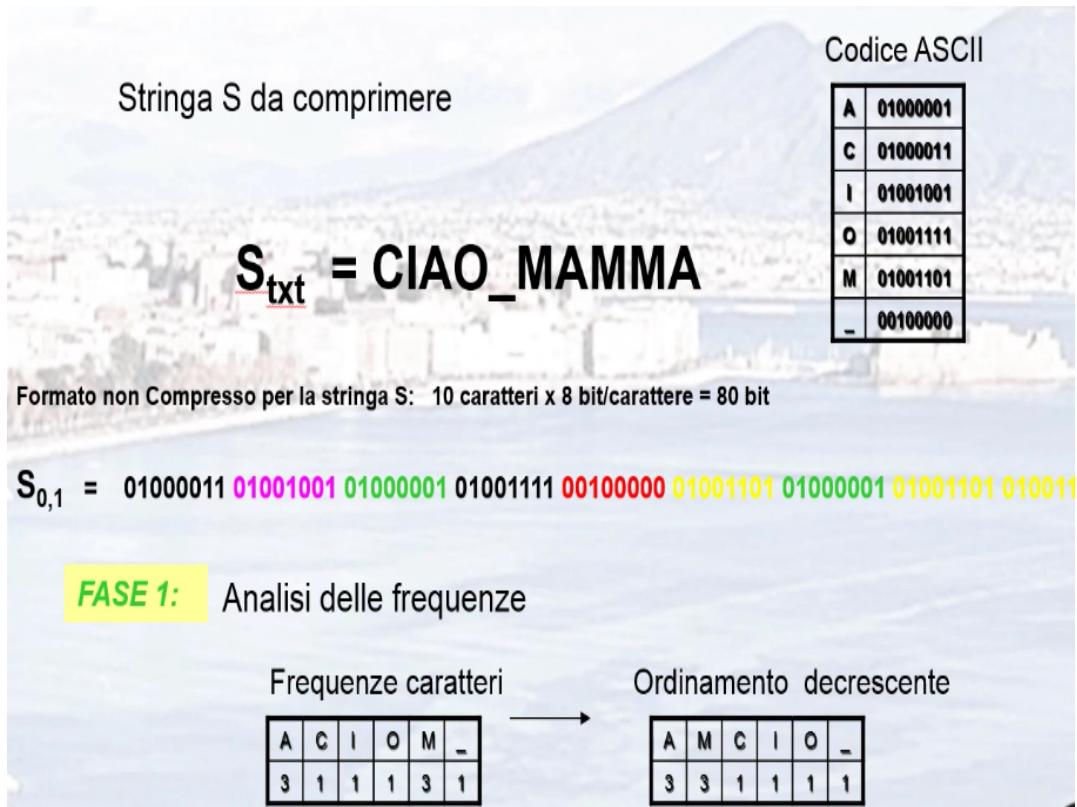
6.4 Algoritmi di compressione

Codifica di Huffman

È basato sull'analisi statistica del dato da comprimere, in particolare sulla frequenza con la quale si ripetono i suoi elementi.

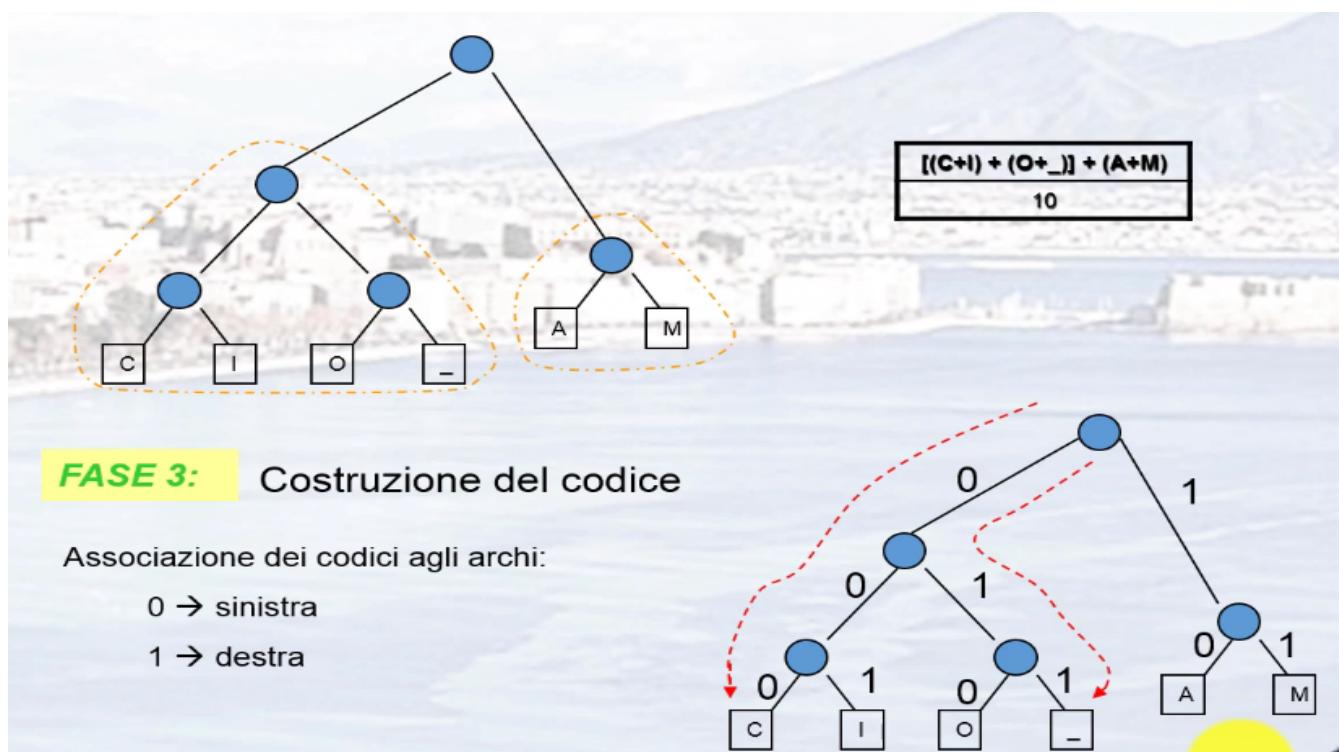
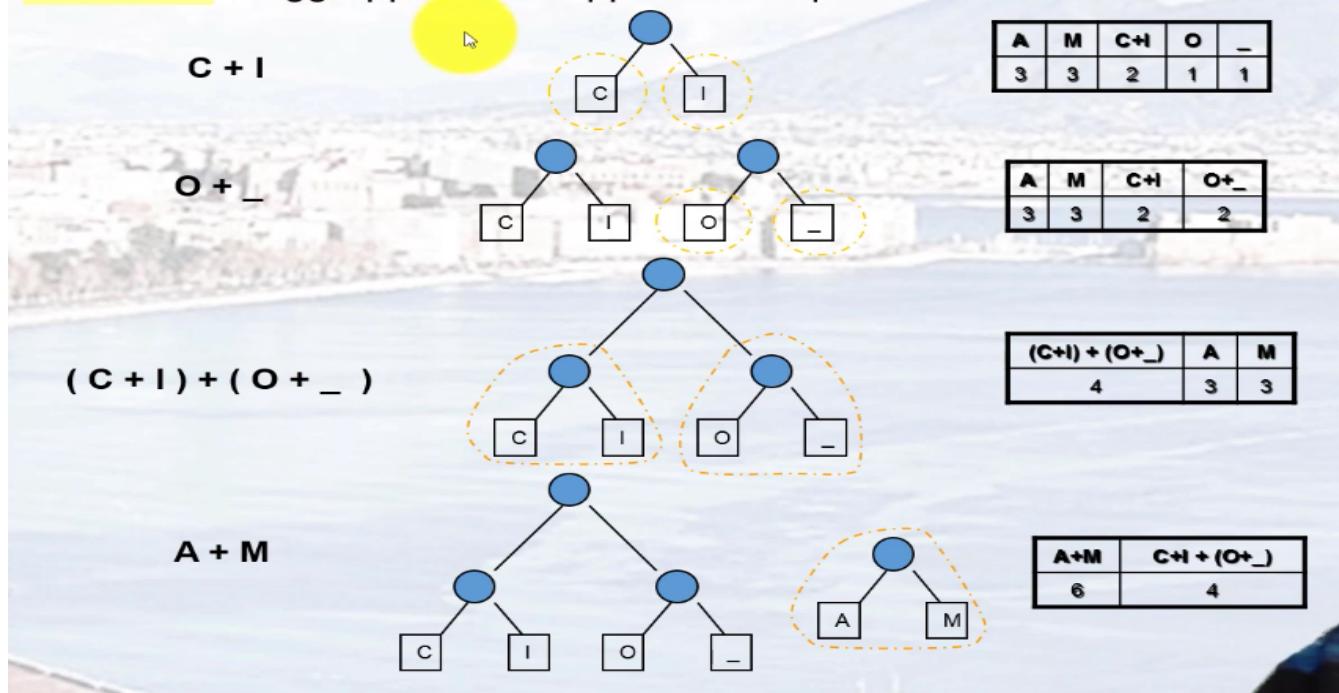
Ha una prestazione proporzionale alla variazione delle frequenze con cui compaiono i caratteri del testo da comprimere (maggiore varianza → maggiore prestazione)

es.



Con il codice di Huffman si potrà utilizzare una codifica variabile → associare ai valori che si presentano più frequentemente un numero di bit minori.

FASE 2: Raggruppamenti coppie con frequenza minore



Seguendo le etichette (0 ed 1) riesco a codificare le varie lettere (sequenza di etichette)

Associazione codici \leftrightarrow caratteri

Costruisco il codice leggendo le etichette degli archi dalla radice ad ogni foglia (**CodeBook**):

C	I	A	O	-	M	A	M	M	A
000	001	10	010	011	11	10	11	11	10

$$S' = 0\textcolor{yellow}{00}\textcolor{magenta}{001}\textcolor{red}{10}0\textcolor{blue}{010}\textcolor{green}{011}\textcolor{cyan}{111}\textcolor{magenta}{101}\textcolor{red}{111}\textcolor{blue}{10} \longrightarrow 24 \text{ bit}$$

Analisi:

- Compressione(S) = S'
- $|S| = 80$ bit
- $|S'| = 24$ bit
- Riduzione = $24/80 = 0,3 \rightarrow 30\%$; **Guadagno = 70%**
- Il codice ottenuto **non è univoco**.

Questo codice è univocamente decifrabile (non presenta ambiguità) \rightarrow **ogni parola non è prefissa di nessun altra parola.**

Codifica RUN-LENGHT

RUN-LENGHT è un metodo di compressione LOSSLESS che riduce le ripetizioni di caratteri, sostituendo un **RUN** (insieme di caratteri ripetuti) con il carattere che viene ripetuto e con la lunghezza del RUN:

- RUN: insieme di caratteri ripetuti
- Lunghezza RUN: lunghezza delle ripetizioni

Rappresentazione del RUN :

Sc = carattere speciale per l'identificazione della codifica

X = il carattere che viene ripetuto nel RUN

C = numero di ripetizioni del carattere.

Prestazioni buone per run-length > 3

es.

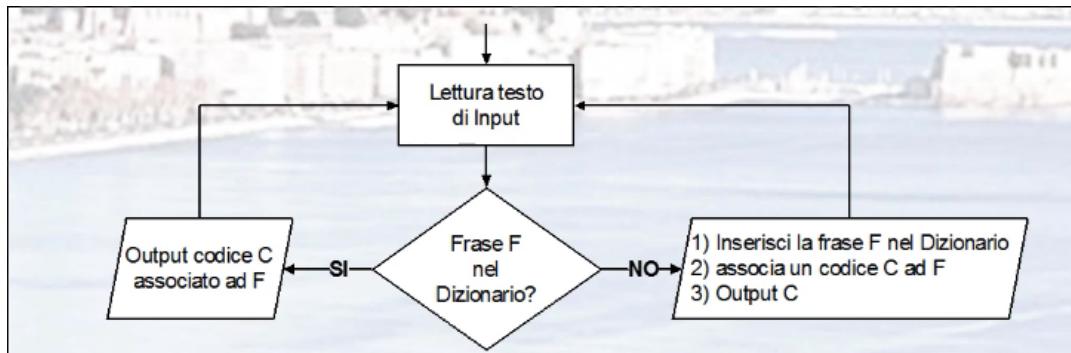
"eeeeeeetnnnnnnnn" → @e7t@n8

Applicazioni: FAX, immagini con pochi colori, ...

Codifica LZW

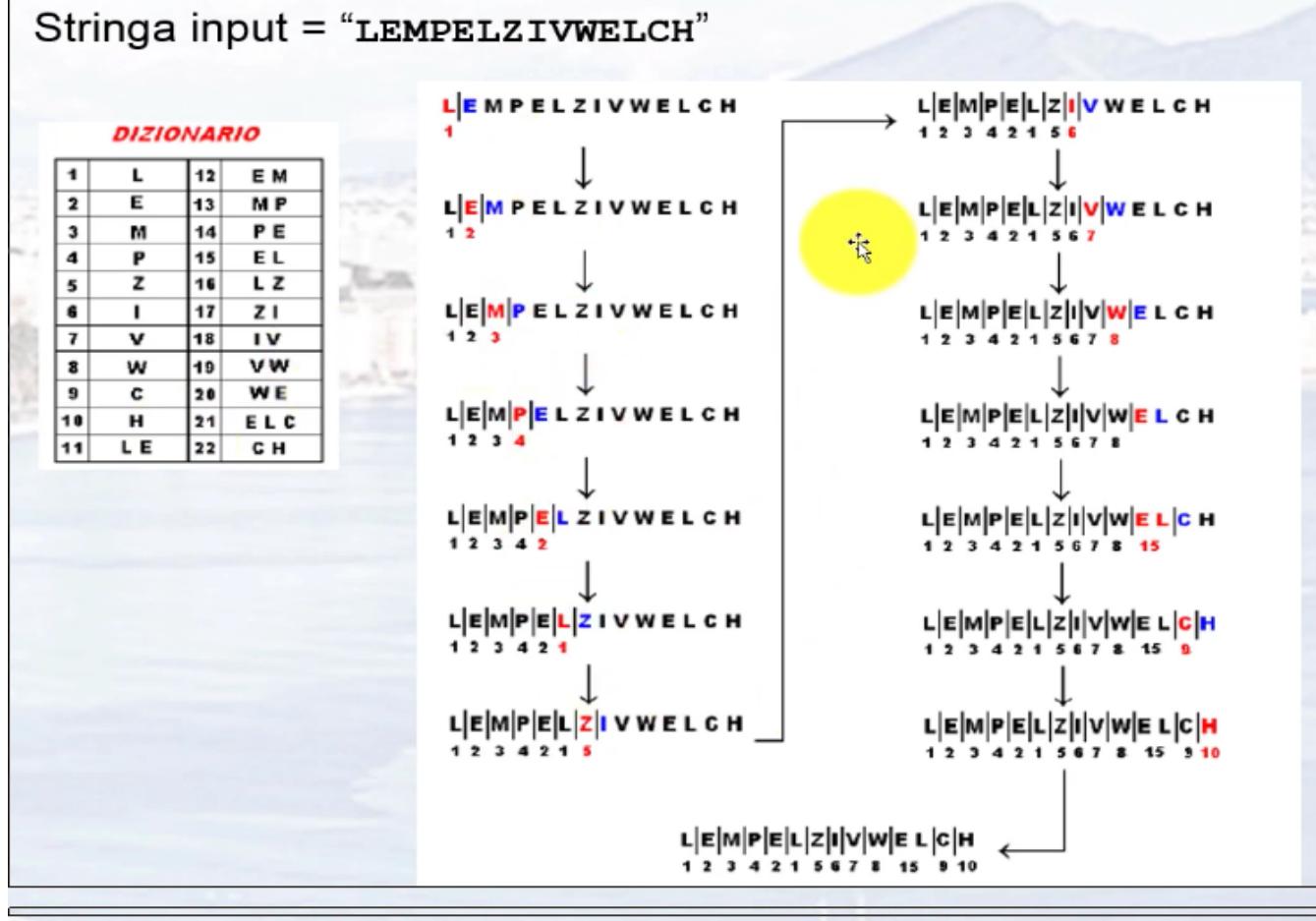
Metodo di compressione LOSSLESS che sfrutta la ripetizione di gruppi di caratteri o frasi.

Il compressore esamina la presenza delle frasi incontrate con le frasi presenti in un dizionario inizialmente vuoto



Le prestazioni sono buone per input di testo con molte ripetizioni: Linguaggio naturale

Stringa input = "LEMPPELZIVWELCH"



7 L'Audio

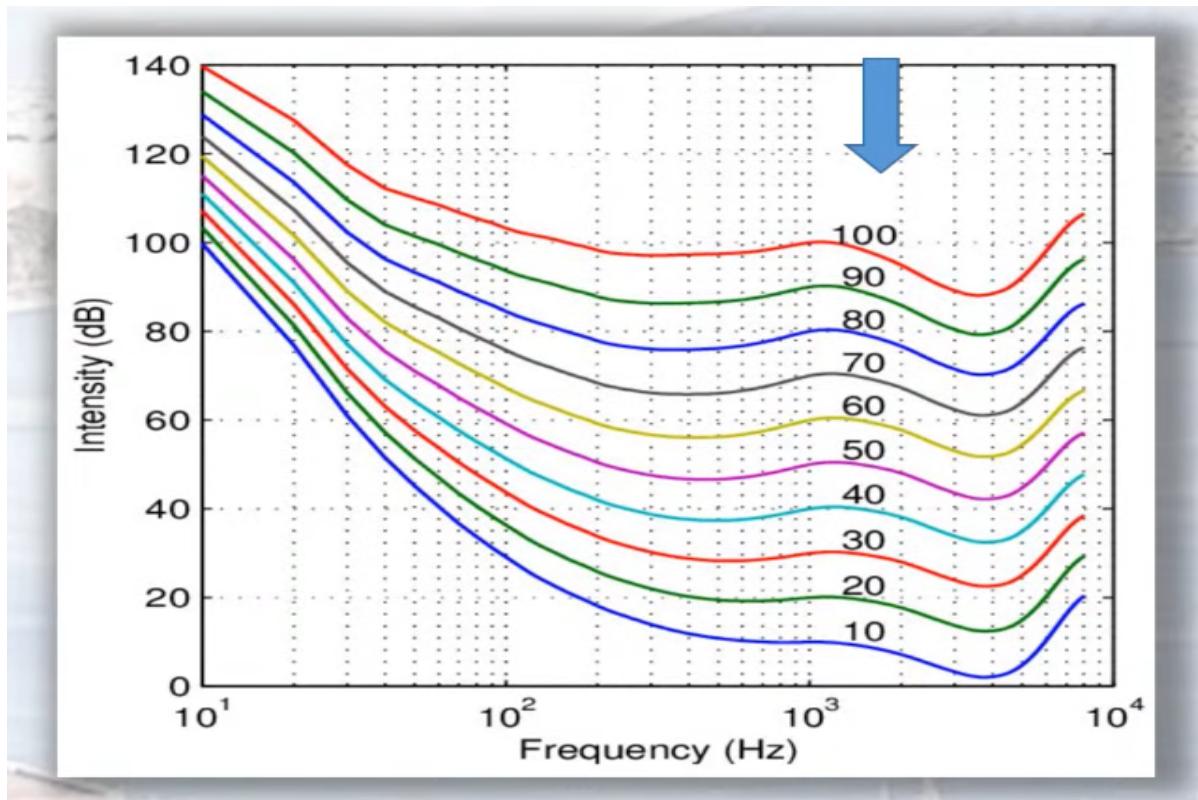
È causato da variazioni di pressioni dell'aria Le frequenze udibili dall'uomo coprono l'intervallo di 20 - 20000 Hz I descrittori fondamentali del suono sono:

- Ampiezza
- Frequenza

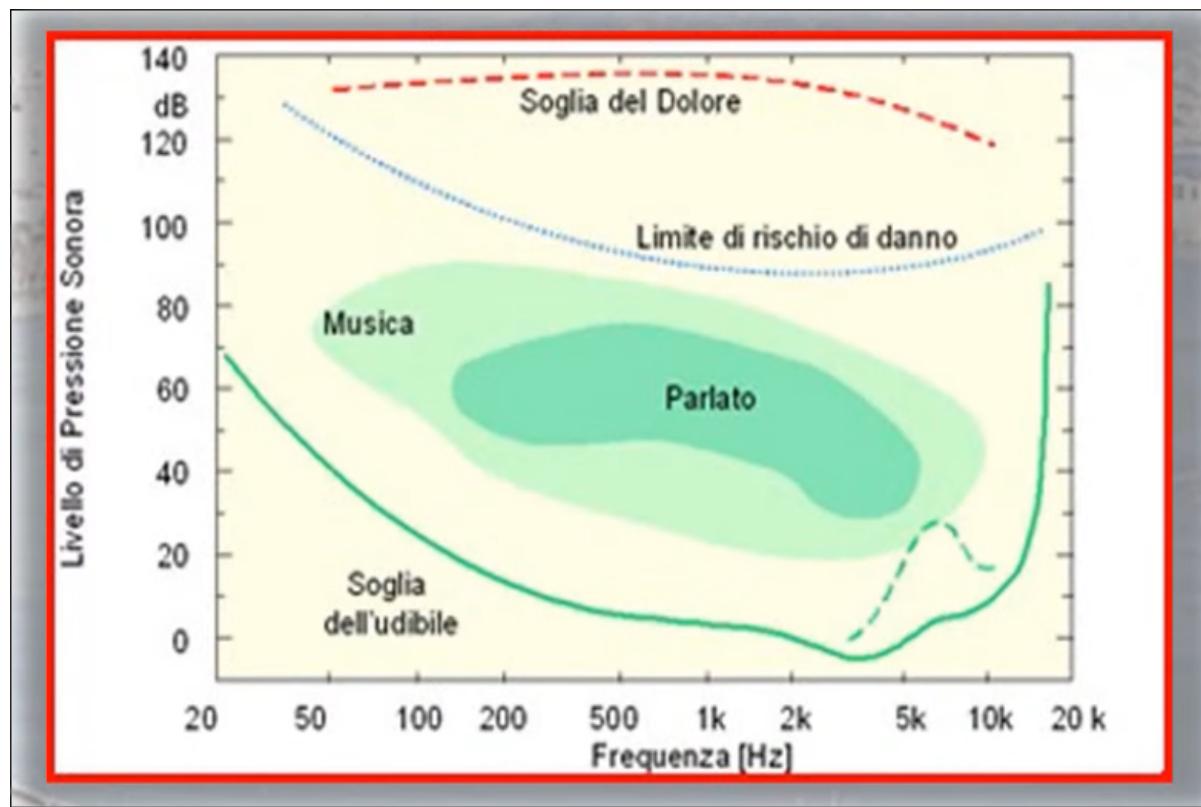
entrambe che variano nel tempo

- Le capacità uditive dell'uomo non sono lineari rispetto ad ampiezza e frequenza
- Le **curve isofoniche** descrivono quali livelli sonori percepiamo essere uguali al variare della frequenza
- La maggiore sensibilità è tra i 1000 e 5000 Hz

Curve isofoniche Fletcher e Munson

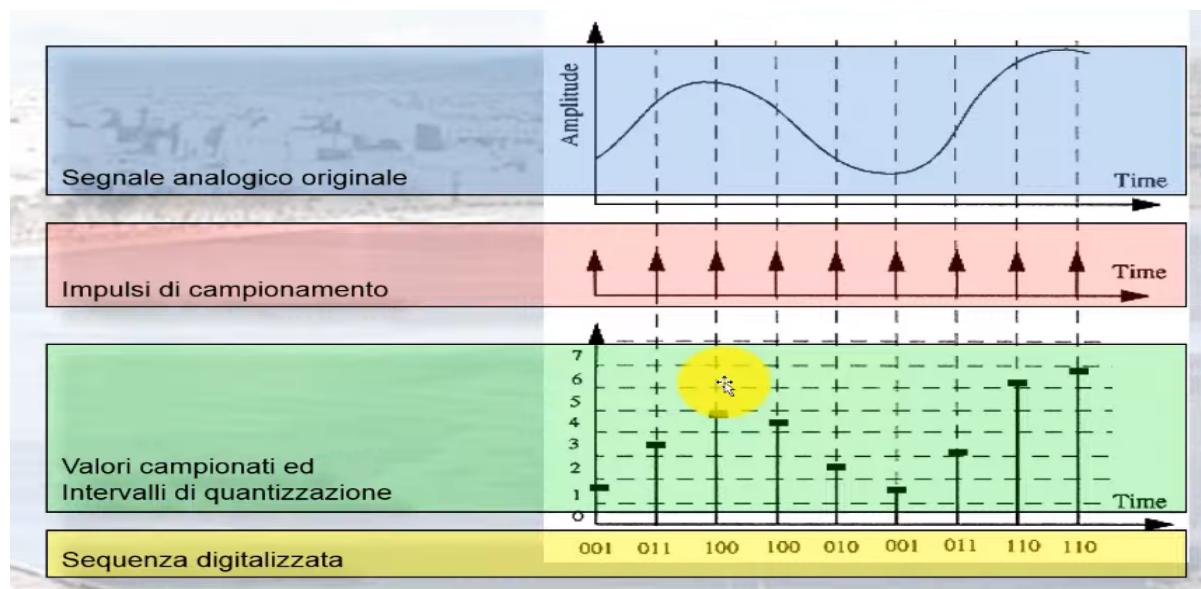


Zone di percezione



7.1 Rappresentazione digitale dell'audio

Conversione ANALOGICO → DIGITALE



Si andranno quindi a leggere dei valori intermedi (essendo impossibile fare una conversione puntiforme) attraverso degli **impulsi di campionamento** (prendiamo i valori a distanze di tempo prefissate).

Le 3 fasi fondamentali di un processo di conversione sono:

- **Campionamento**

- Prelievo di valori assunti dal segnale analogico ad intervalli discreti di tempo (i valori in questa fase sono ancora di tipo analogico)

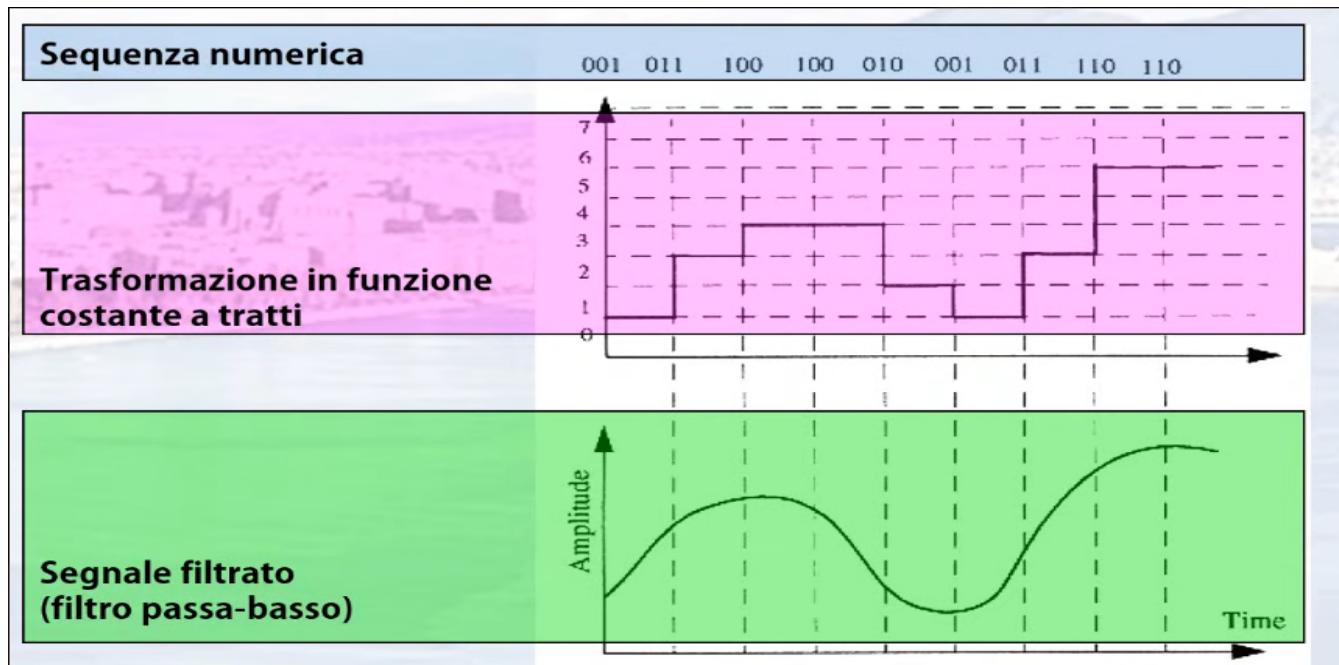
- **Quantizzazione**

- Processo di conversione dei valori continui in valori discreti. L'intervallo del segnale viene suddiviso in un numero fisso di sotto-intervalli di uguale dimensione e viene assegnato un valore (ciascun valore cade in un unico intervallo → i valori possibili sono in numero limitato). La grandezza del sotto-intervallo di quantizzazione è detto *passo di quantizzazione*

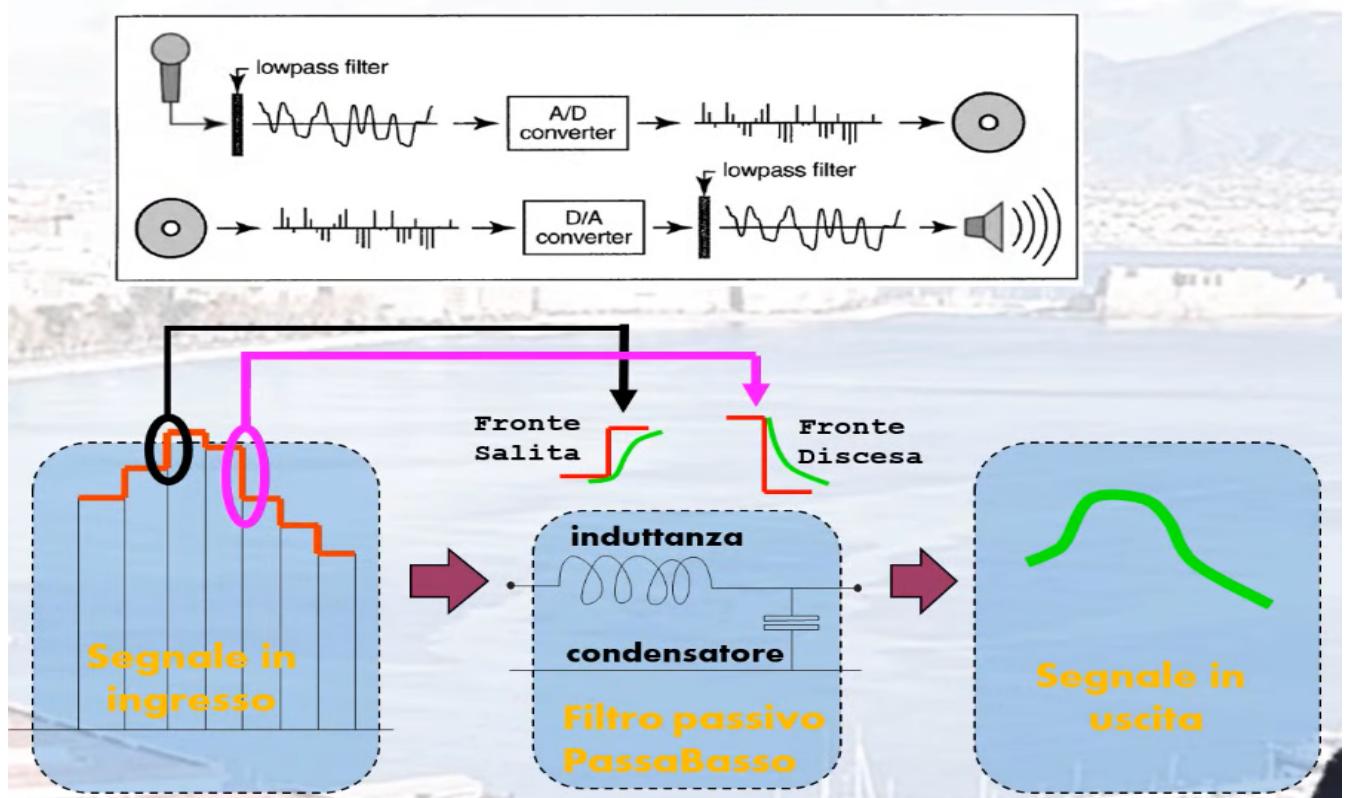
- **Codifica**

- Processo di rappresentazione numerica dei valori quantizzati. Quanto maggiore sarà la frequenza di campionamento e il numero di livelli di quantizzazione tanto maggiore sarà la fedeltà del segnale digitalizzato

Nel caso inverso avremo:



Filtro passa-basso → filtro che impedisce l'ingresso di tutte le frequenze alte (da qui il nome di passa-basso)



7.2 Teorema di Nyquist (importante)

La frequenza di campionamento è strettamente dipendente dalla frequenza massima del segnale analogico da convertire. Il Teorema di Nyquist afferma che

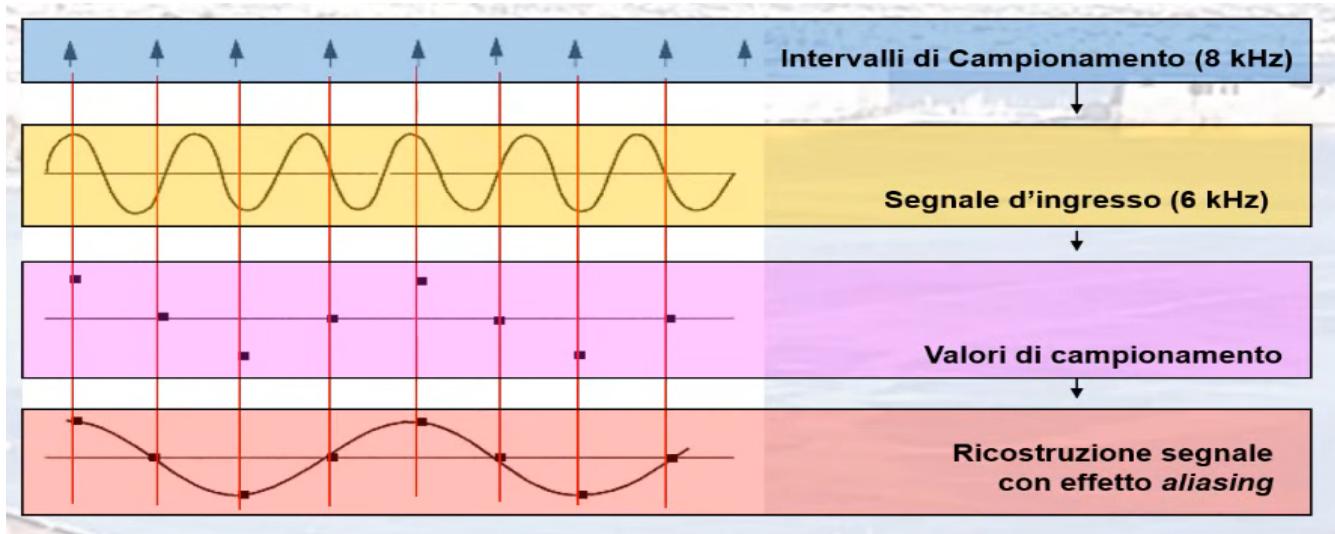
Se in un segnale analogico c'è una componente con frequenza fino a f Hz allora la frequenza di campionamento dovrebbe essere almeno $2f$ Hz

7.3 Scelta della frequenza di campionamento

Nella pratica le frequenze di campionamento sono di poco superiori al campionamento critico

Dispositivo	Freq. campionamento
CD-Audio	44.1 kHz
DAT (Digital Audio Tape)	48 kHz
Telefono	8 kHz

Esempio di **ricostruzione errata** per basso campionamento (effetto aliasing)



Errore (o rumore) di quantizzazione

$$\text{Max}\{\text{Campione_quantizzato} - \text{segnale_analogico}\}$$

Il numero **Q** dei livelli di quantizzazione determina la quantità **b** di bit necessaria per rappresentare ciascun campione:

$$b = \log_2 Q$$

La quantità di segnale digitale **SNR** (Signal Noise Ratio) viene misurata in decibel

$$\text{SNR} = 20 \log_{10} (\text{S}/\text{N}) = 20 b \log_{10} 2 = 6b$$

dove $S = \max$ ampiezza segnale;

N = errore di quantizzazione;

q = passo di quantizzazione;

$$S = 2^b q$$

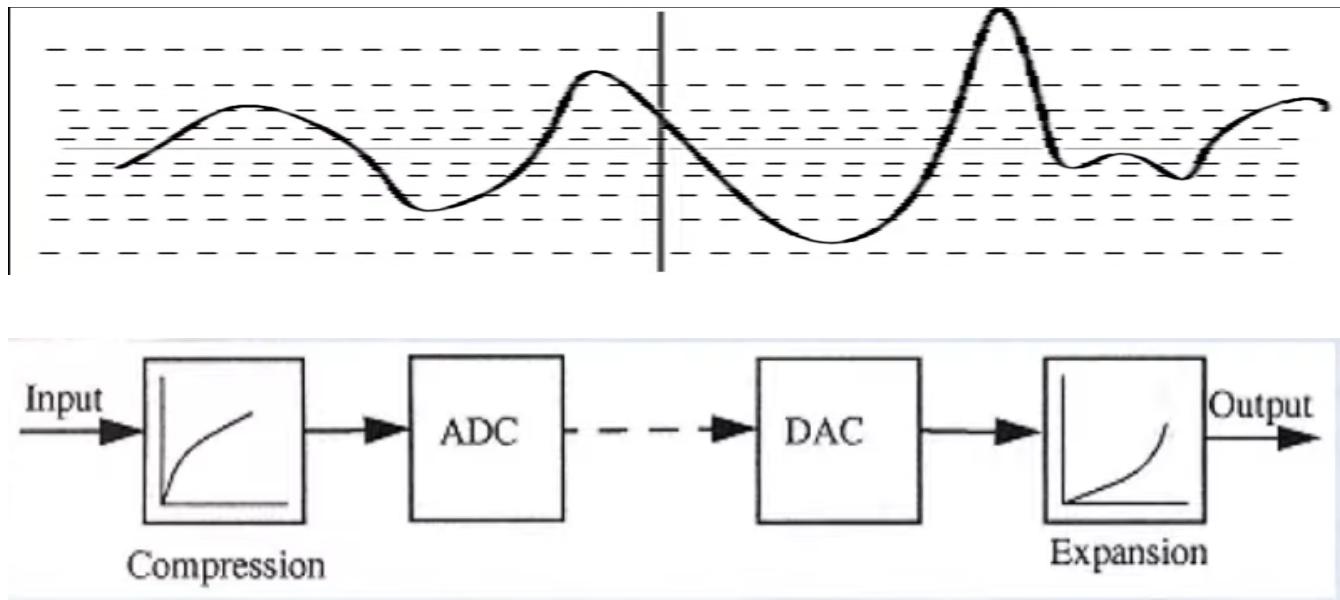
NOTE

Ogni bit in più usato per rappresentare il campione implica un aumento di 6 db del SNR

Se l'errore di quantizzazione supera il valore della soglia uditiva allora viene avvertito

7.4 Companding

Ha lo scopo di attenuare gli effetti deleteri di un canale con intervallo dinamico limitato (aumentiamo la fedeltà della traccia)



Mediante la tecnica del Companding è possibile riprodurre un segnale ad 8-bit con la stessa qualità di un segnale a 12 bit. Possiamo vederlo quindi come una sorta di compressione analogica del segnale

7.5 Predictive Coding

Con il **predictive coding** invece di codificare il valore del campione da trasmettere, si codifica la differenza tra la predizione del valore del campione ed il valore del campione attuale (*differential pulse-coded modulation*).

Il valore della predizione si ricava dai valori precedenti assunti dal segnale; tale valore è pertanto noto sia al codificatore che al decodificatore che applicano la medesima strategia

NB → Si può avere una differenza ancora minore se si sottrae alla predizione del valore del campione il valore del campione attuale

L'efficacia del predictive coding si base sul fatto che:

- Campioni vicini sono significativamente correlati;
- Per codificare una differenza occorre un numero inferiore di bit

Nel caso in cui le differenze dovessero risultare molto grandi, vengono introdotte opportuni algoritmi correttivi

7.6 Mpeg Audio

Questa tecnica è basata sul **mascheramento**: suoni di maggiore intensità "coprono" i suoni a bassa intensità, che pertanto possono essere ignorati (compressi) senza influire sulla qualità udibile dal nostro orecchio

Caratteristiche:

- Si tratta di una compressione a perdita
- Frequenze di campionamento usate → 32, 44.1, 48 kHz
- Supporto di 1 o 2 canali audio
- Presenza di famiglie di layer (1,2,3)

7.7 Audio sintetico (Midi)

Con l'acronimo **Midi** (Musical Instrument Digital Interface) si indica il protocollo standard per l'interazione degli strumenti musicali elettronici

Questo formato non contiene musica pre-registrata, ma le direttive e le specifiche per una sua riproduzione.

HW e SW ne realizzano l'esecuzione.

Le direttive quindi sono del tipo:

Esegui la nota N con una durata T e con lo strumento S

Possiamo quindi definire un file midi come una sorta di spartito musicale elettronico

Vantaggi	Svantaggi
Grandezze dei file molto ridotte	La riproduzione del suono non è univoca

Pertanto avremo il vantaggio della trasmissione, ma una variazione in base allo strumento utilizzato

8 Le immagini

8.1 Percezione

Ci si interessa del fruitore dell'immagine. Di conseguenza la percezione cambia al variare del fruitore

8.2 Rappresentazione dei colori

La luce visibile è una radiazione elettromagnetica con una lunghezza d'onda tra i 400 e i 780 nm. In base alla variazione di questa onda si avranno diverse tonalità di colore. Le 3 proprietà fisiche delle radiazioni di colore sono:

- Luminanza (illuminazione);
- Tinta (il colore);
- Saturazione (la purezza)

8.3 Sintesi dei colori

I colori sono generabili attraverso l'uso di due sintesi:

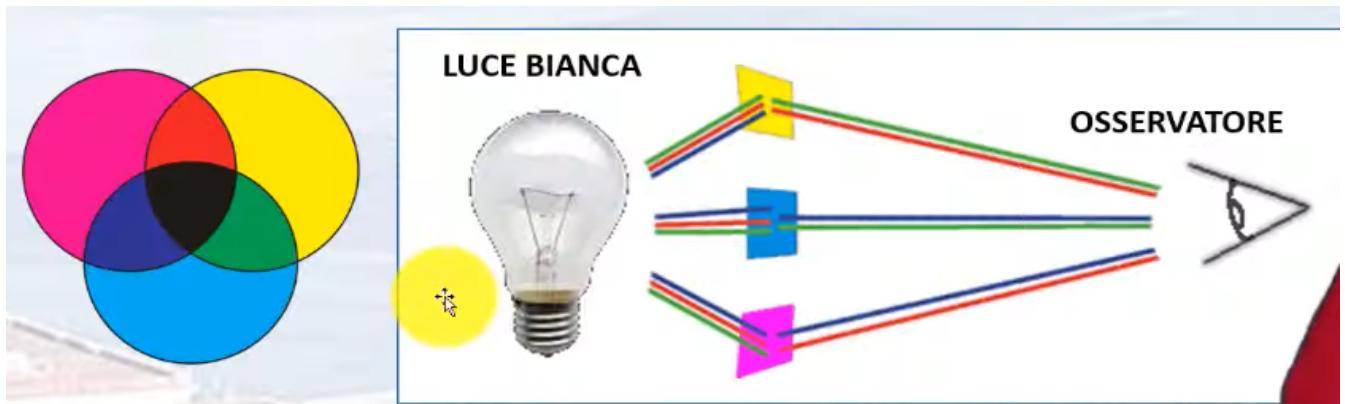
- Sintesi additiva;
- Sintesi sottrattiva.

Sintesi additiva

Si ottiene attraverso una mescolanza dei colori

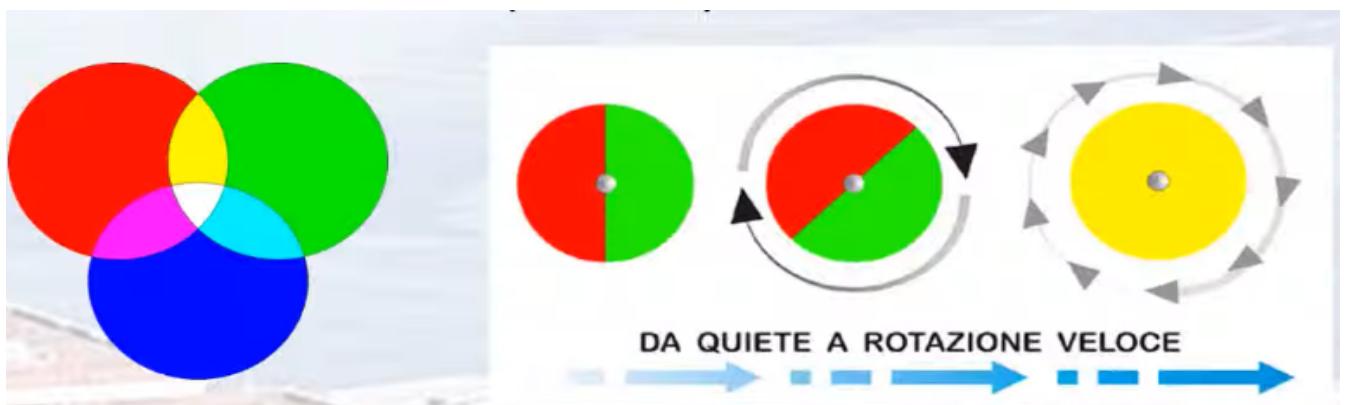
Sintesi sottrattiva

Sottraendo alcune frequenze è possibile ottenerne altre

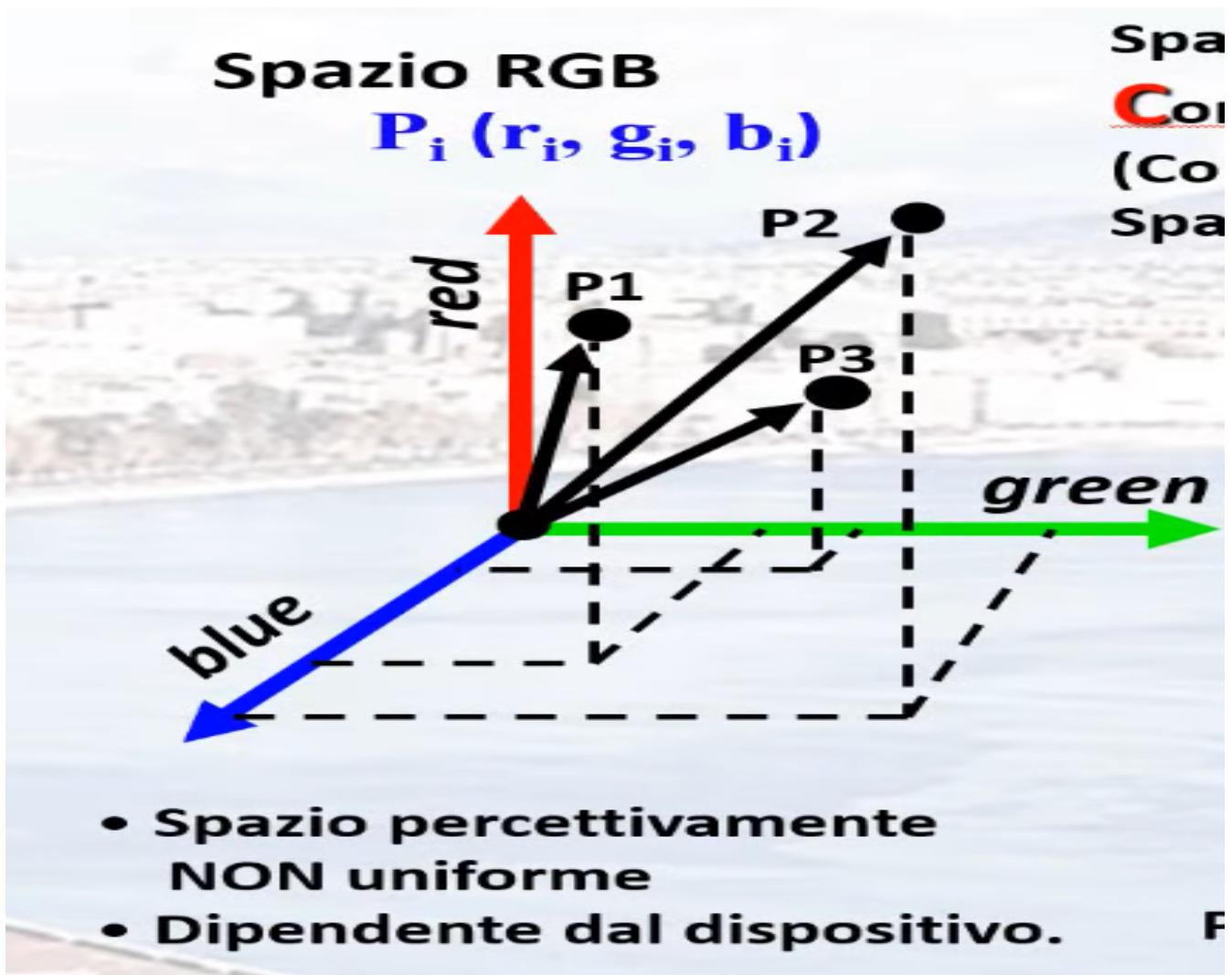


Disco di Newton

A partire dai 3 colori primari (RGB) è possibile ottenere gli altri



8.4 Spazio RGB (3D)



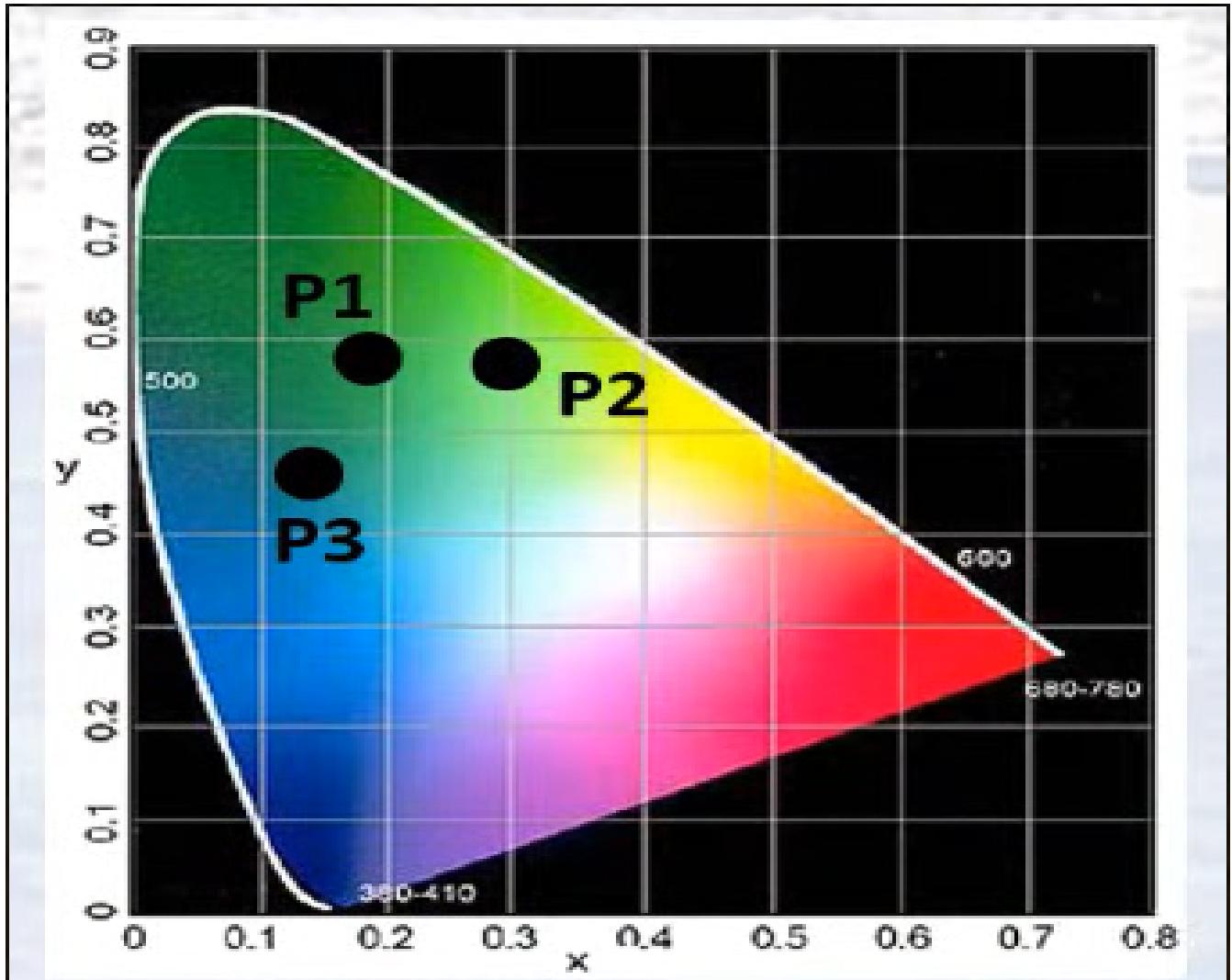
- Spazio percettivamente NON uniforme
- Dipendente dal dispositivo

8.5 Spazio CIE normalizzato (2D)

$$P_i (x_i, y_i)$$

CIE → Commission Internationale de l'Eclairage (1931) (commissione internazionale per l'illuminazione)

- Spazi (CIERGB, CIEXYZ, CIELUV, CIELAB)



Per la nostra capacità di percezione

$$\overline{(P_i P_j)}_{RGB} \neq \overline{(P_i P_j)}_{CIE}$$

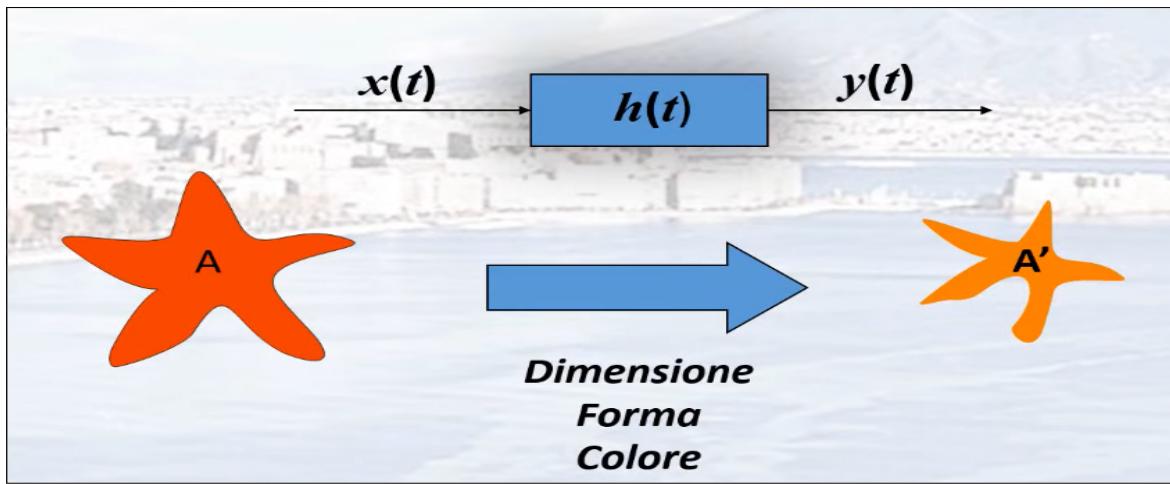
E cioè nello spazio RGB la distanza tra due punti P_i e P_j è diversa dalla capacità di percepire nello spazio CIE negli stessi due punti. Nello spazio CIE ho una rappresentazione simile a quella percepita dall'occhio umano

8.6 Rappresentazione SPD (Spectral Power Distribution)

Ogni colore "fisico" può essere rappresentato mediante la distribuzione dell'energia radiant alle varie lunghezze d'onda SPD

Vantaggi	Svantaggi
Accuratezza massima	Non descrive la relazione tra le proprietà fisiche del colore e la sua percezione visiva Complessa rappresentazione numerica

8.7 Funzione di trasferimento



8.8 Correzione gamma

Ogni strumento fisico di acquisizione o riproduzione dei colori applica una funzione non lineare alla intensità di luce catturata in relazione al segnale in **Volts** emesso

$$\text{Luminanza} = V^\gamma$$

Per neutralizzare la non linearità del dispositivo si usa la

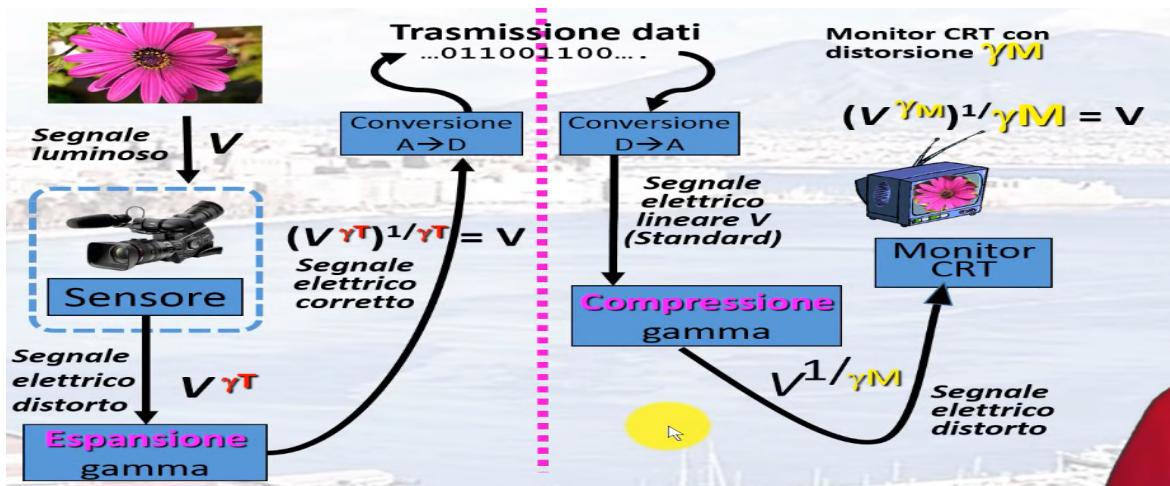
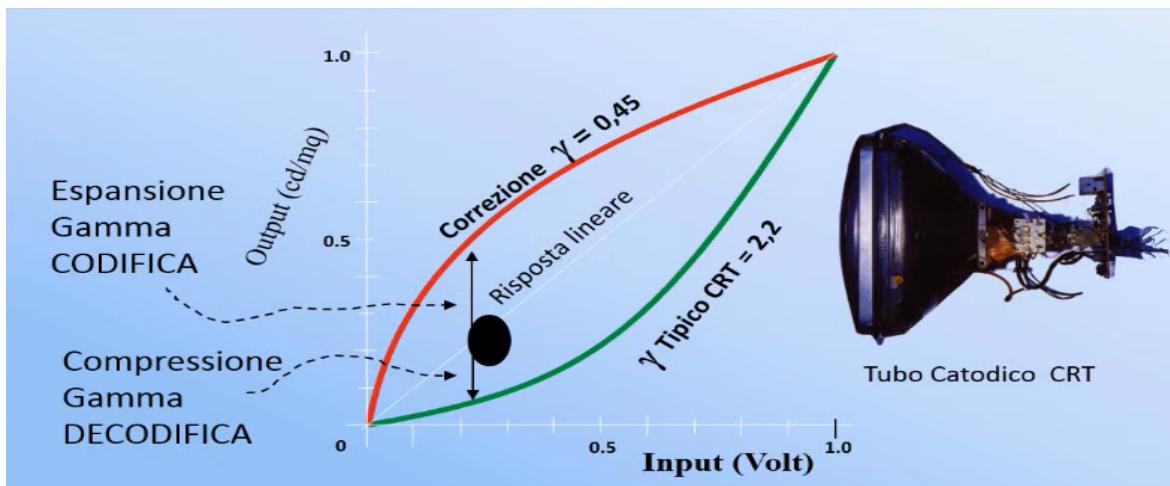
$$\text{Correzione gamma} = 1/\gamma$$

La correzione di gamma è quindi un'operazione non lineare atta a codificare e decodificare la luminanza in un sistema video/fotografico. Nel caso più semplice la Correzione di gamma è definita dalla seguente legge potenziale

$$V_{out} = V_{in}^{\gamma}$$

Questa funzione ha lo scopo di neutralizzare la non linearità del dispositivo; \mathbf{V} denota il segnale elettrico analogico (voltaggio) che il dispositivo preleva al proprio ingresso

es



$(\gamma^t$ rappresenta la gamma della telecamera)

$(\gamma^M$ rappresenta la gamma del monitor)

8.9 Problemi di indicizzazione e ricerca legati allo spazio di colore

È necessario assumere che tutte le immagini siano rappresentate **nello stesso spazio di colore** e che il valore dei loro pixel rappresentino la stessa cosa.

Data un'immagine nel formato RGB non è possibile interpretarla correttamente:

- Non vengono indicate le definizioni delle primitive RGB;
- Non viene indicato il valore della correzione gamma applicata dal device che ha prodotto l'immagine

Tutti i formati immagini più utilizzati (*GIF, JPEG, TIFF (< 6.0), etc...*) non contengono informazioni sulle primitive dello spazio di colore e sulla correzione gamma usata

8.10 Le origini

- Macchine fotografiche
- Scanner
- Fotogrammi di filmati
- Disegni elettronici

8.11 Grafica Raster e Vettoriale

Distinguiamo due tipi di classificazione delle immagini

Raster

Tecnica che descrive immagini mediante griglie di pixel colorati

Presenta due modelli di memorizzazione:

- LOSSY
- LOSSLESS

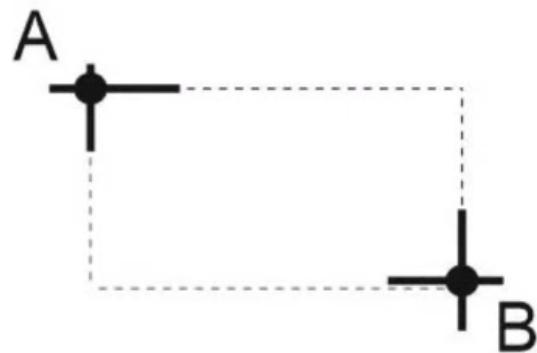
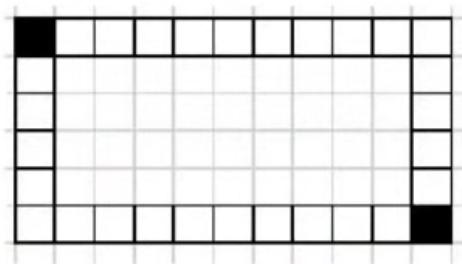
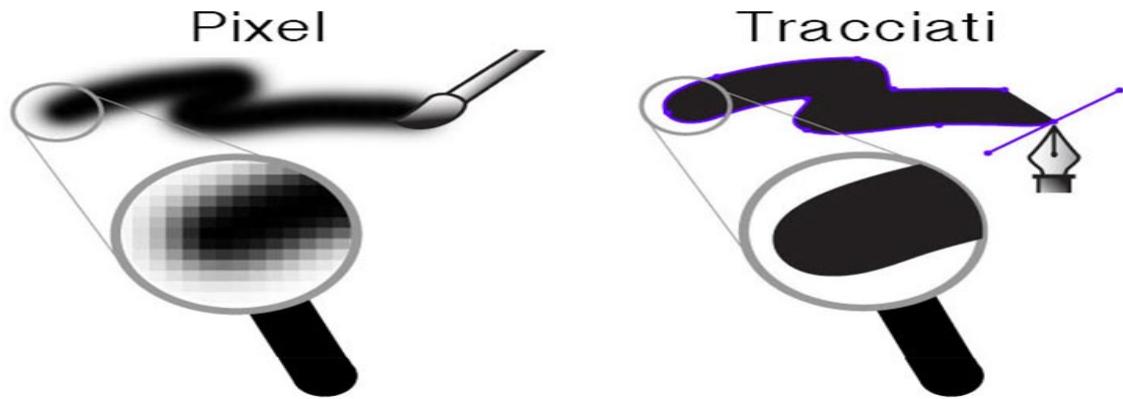
Vettoriali

Tecnica che descrive l'immagine mediante un insieme di primitive geometriche quali punti, linee, curve e poligoni ai quali sono associabili svariati attributi.

Comporta una grande qualità ed elevata compressione

L'operazione di conversione da raster a vettoriale prende il nome di **vettorializzazione**

Immagini raster e vettoriali a confronto

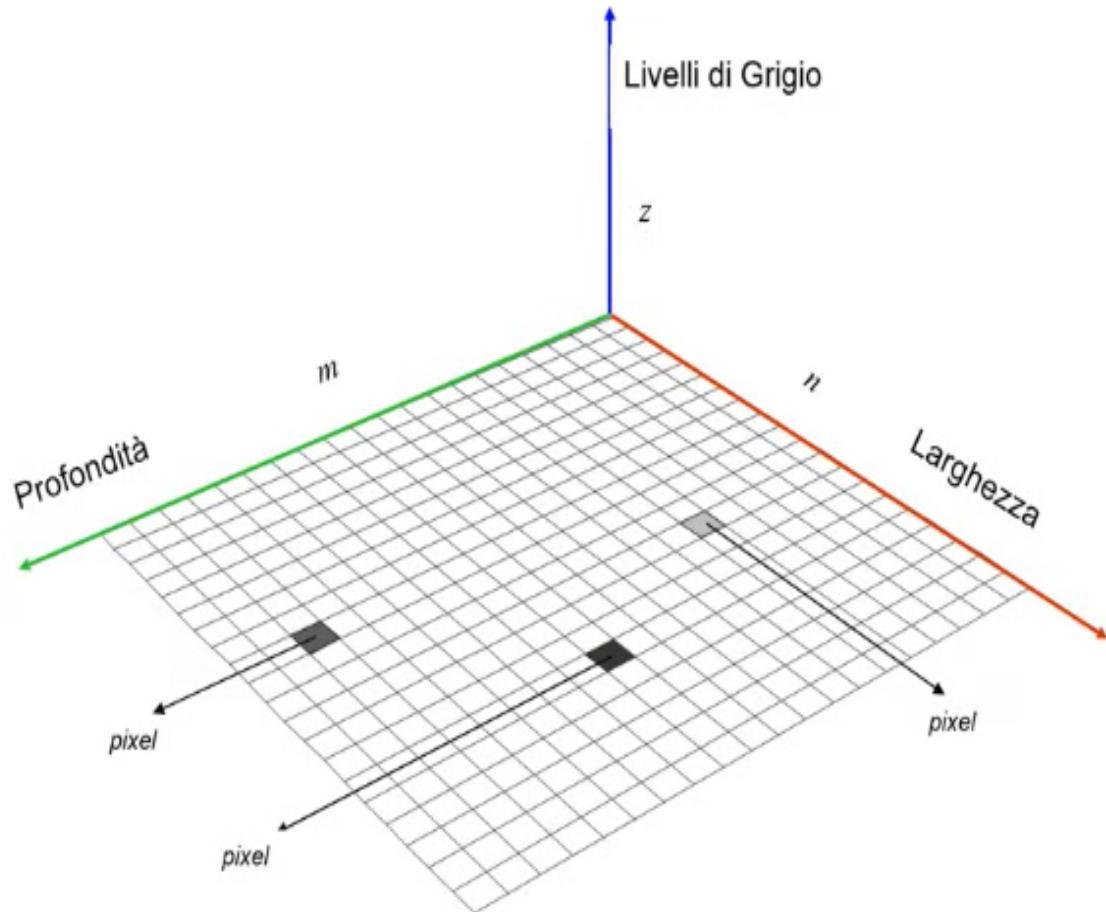


8.12 Super-Resolution (SR)

Tecnica che serve a migliorare la risoluzione delle immagini RASTER attraverso due metodi:

- Tecnica del Single-Frame;
- Tecnica del Multi-Frame

8.13 Immagini GrayScale (scala di grigio)



La grandezza di un immagine (misurata in bit):

$$m * n * [\log_2 z]$$

Ad esempio un'immagine 640 x 480 con 256 toni di grigio (1 byte per ogni pixel) richiede 307200 bytes

8.14 Immagini a colori

Un **modello di colore** è un modello matematico astratto che permette di rappresentare i colori delle immagini in forma numerica. Esistono diversi modelli:

- RGB (a sintesi additiva → simile alla rappresentazione ai toni di grigio);
- CMYK (Cyan Magenta Yellow Black, a sintesi sottrattiva);
- YUV (ambito analogico → nasce per la compatibilità della TV a colori con TV B/N ; Y = Luminanza; U,V cromie) ;
- YCbCr (equivalente della YUV → Y = Luminanza; Cb = CromaBlu; Cr = CromaRed);
- HSV (Hue Saturation Value)

8.15 Fondamenti della compressione

L'occhio umano presenta delle limitazioni su ciò che riesce a percepire.

Generalmente per ogni pixel di immagine i punti adiacenti sono simili; ciò costituisce la **ridondanza spaziale**

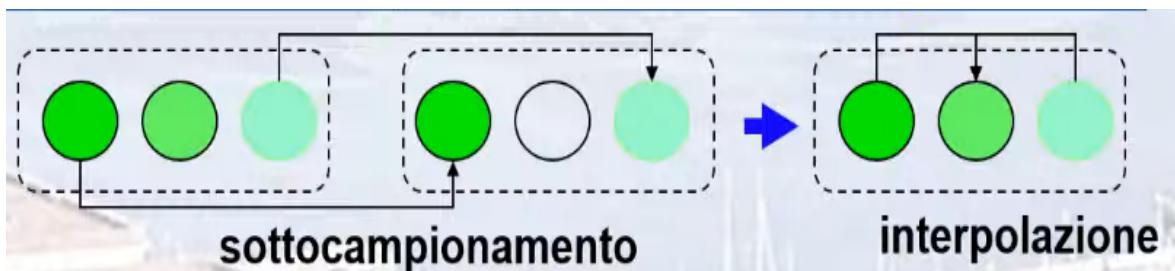
A differenza dal testo (dove non è ammessa la perdita di dati) nel caso di immagini è permessa la perdita di dati

8.16 Compressione LOSSY

Compressione con sottocampionamento

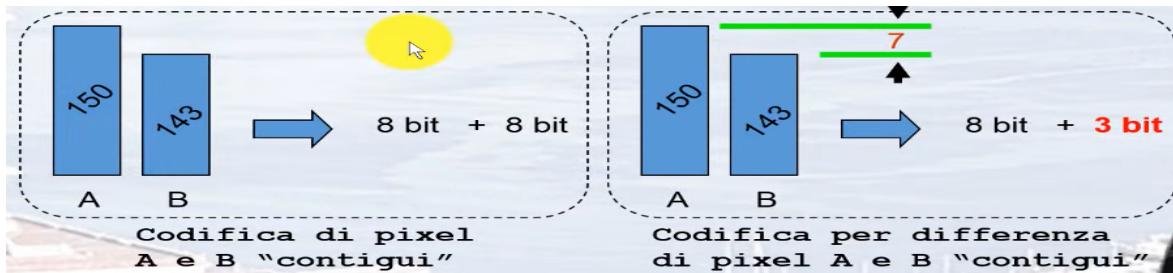
A causa della ridondanza spaziale (similitudine di pixel vicini), è possibile considerare solo alcuni pixel (da qui il campionamento). È possibile raffinare questo metodo scegliendo di sottocampionare componenti per le quali l'occhio umano è poco sensibile

Ogni immagine può essere decomposta mediante le componenti di **Luminanza** e **crominanza** (l'occhio è più sensibile al primo).



Compressione mediante Predictive Coding

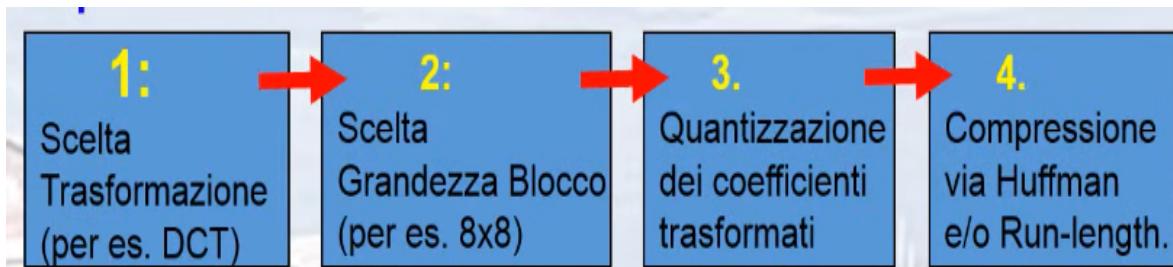
È analogo al predictive coding audio (valori spazialmente vicini sono fortemente correlati)



Codifica mediante trasformazione

L'idea è quella di suddividere un'immagine in sottoimmagini rettangolari su cui si applica una trasformazione unitaria dal dominio spaziale a quello frequenziale (si diminuisce il peso dell'immagine)

Implementazione del sistema di codifica



Le implementazioni maggiormente utilizzate sono:

- DFT (Discrete Fourier Transform)
- DCT (Discrete Cosine Transform)

8.17 Serie di Fourier

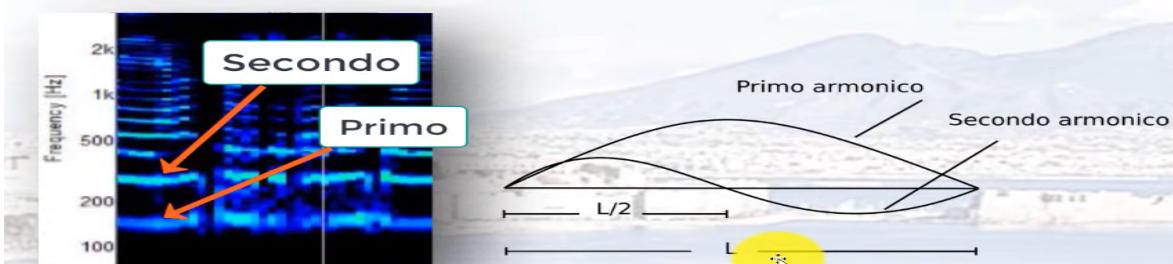
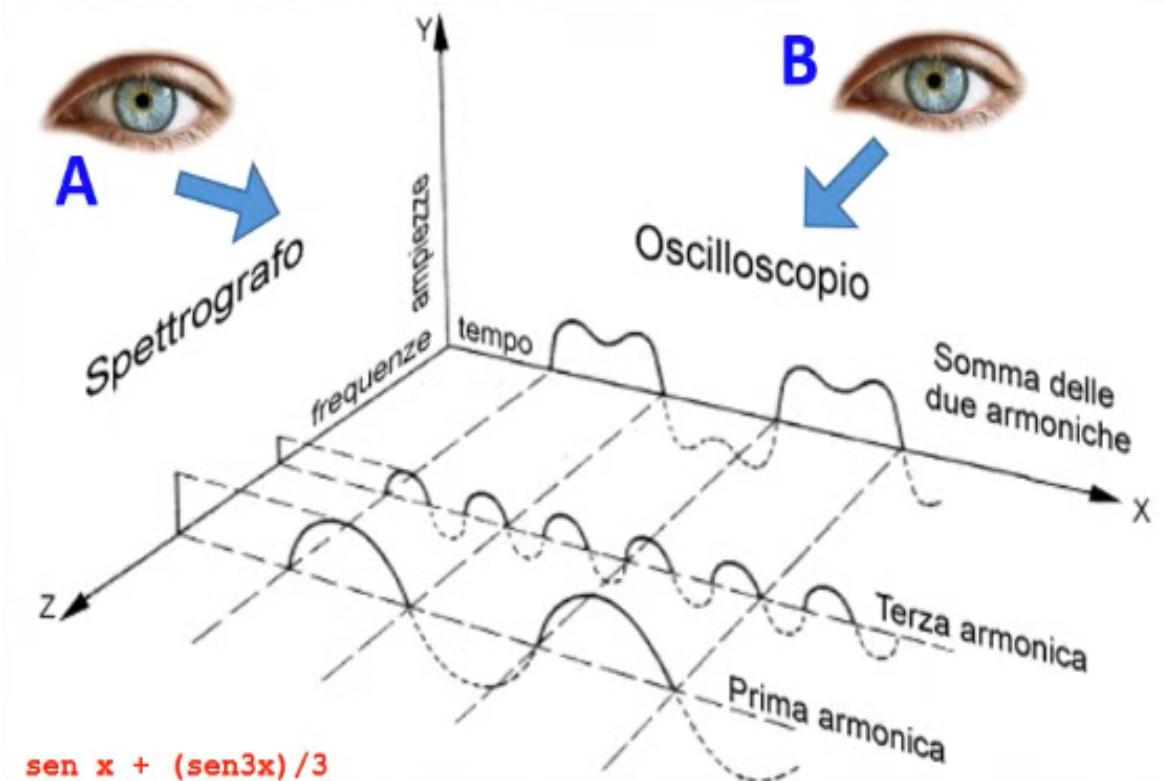
Fourier dimostrò che qualsiasi segnale periodico

$$[f(x) = f(x+T)]$$

può essere scomposto in una somma di (infiniti) segnali sinusoidali

La serie di Fourier rappresenta un segnale periodico $x(t)$ di pulsazione $\{\omega\}_0$ mediante una somma pesata di sinusodi di cui la prima (fondamentale) ha pulsazione $\{\omega\}_0$ e le successive (armoniche) hanno pulsazioni multiple di $\{\omega\}_0$

$$x(t) = a_0 + a_1 \cos(\{\omega\}_0 t + \{\Theta\}_1) + a_2 \cos(2\{\omega\}_0 t + \{\Theta\}_2) + \dots + a_N \cos(N\{\omega\}_0 t + \{\Theta\}_N)$$



Trasformata di Fourier

Decompono un segnale nelle diverse frequenze che partecipano a costruirlo, comprese le loro fasi (decomposizione unica)

$$(F f)(w) = \int_{-\infty}^{\infty} f(t) e^{-i w t} dt$$

8.18 Analisi spettrale

È possibile a partire da un segnale decomporlo nelle sue frequenze (si ottiene una vera e propria impronta digitale del segnale)

8.19 Immagini JPEG (Joint Photographic Experts Group)

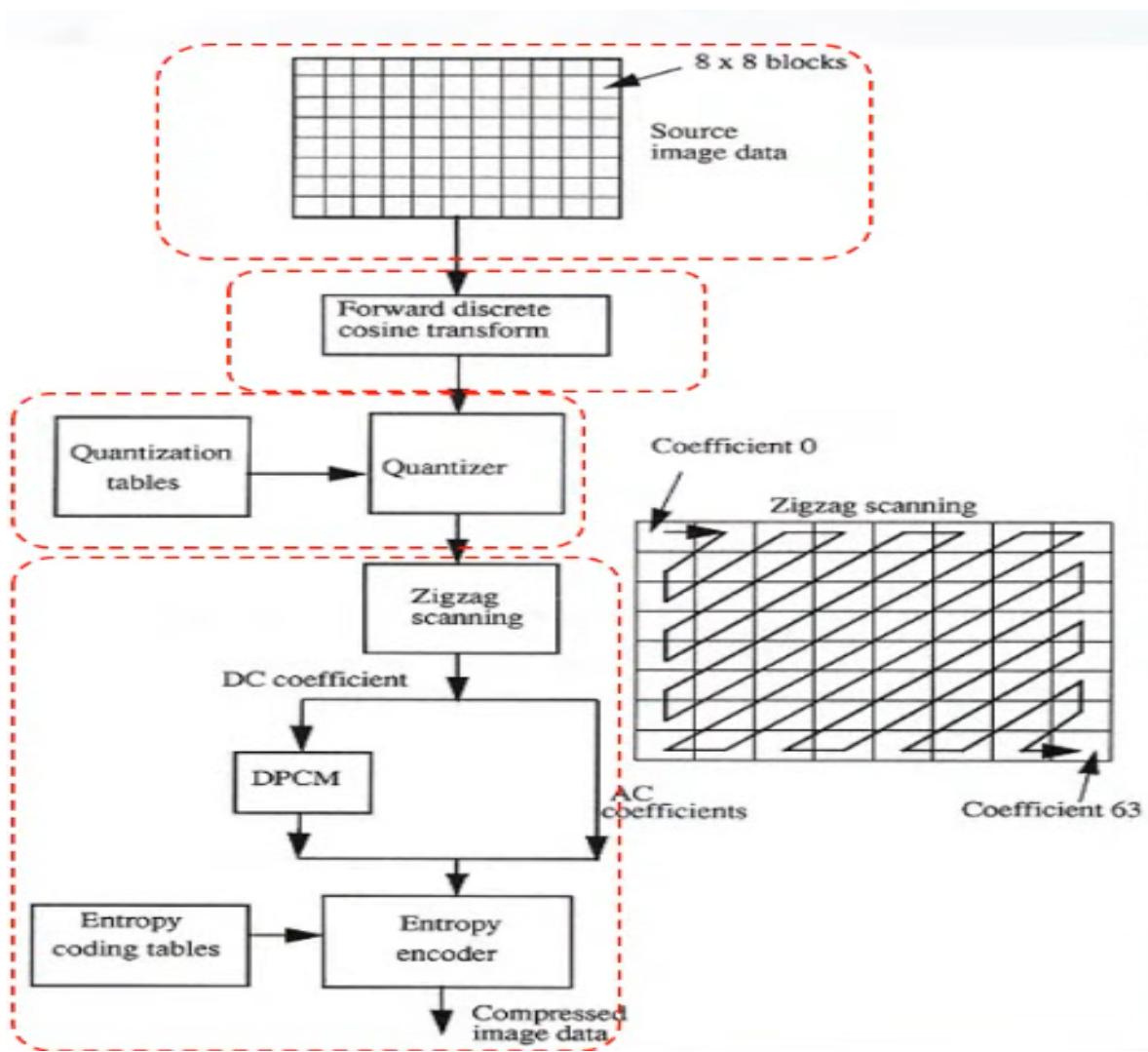
- Immagini a colori 24 bit ed immagini bianco e nero;
- Modello LOSSY;
- Modello Open Source;

Idee di base:

- Limite percezione visiva;
- Analisi spettrale (DCT)

Modello JPEG (passaggi)

1. Preparazione in blocchi
2. Passaggio al dominio frequenziale
3. Quantizzazione
4. Codifica



8.20 Immagini Frattali

Rappresentano un modo intermedio (tra Raster e Vettoriali) per la memorizzazione delle immagini → si cerca di rappresentare la mappa di bit mediante una funzione matematica

Questo metodo (di tipo LOSSY) cerca di decomporre l'immagine in parti elementari che verranno memorizzate insieme alle regole necessarie per la ricomposizione.

Questo metodo è applicabile a qualunque *scala spaziale*.

9 Il Video

Sequenza di fotogrammi o immagini visionate a frequenza cotante

9.1 Frame rate (velocità di scorrimento)

L'occhio umano riesce a percepire le immagini in maniera fluida a 25 frame al secondo (a frame inferiori si va incontro al fenomeno dello sfarfallio, con frame rate superiori è impossibile apprezzarne la differenza)

9.2 Codifiche

Codifica Intraframe

Codifica e decodifica di un flusso video descrivendo ogni singolo **fotogramma**

Codifica Interframe

Descrizione dei cambiamenti che occorrono tra un fotogramma ed il successivo partendo da un fotogramma iniziale descritto con codifica intraframe

9.3 Compressione video

- Limite percezione visiva
- Riduzione della ridondanza
 - Si sfrutta la similitudine dei fotogrammi adiacenti
 - Ogni fotogramma si divide in blocchi e si cerca la migliore corrispondenza tra blocchi di fotogrammi adiacenti
 - * Per ogni coppia di blocchi simili si determina
 - Lo spostamento del blocco (**motion vector**)
 - La differenza tra i 2 blocchi

9.4 MPEG (Motion Picture Expert Group)

	Bitrate(Mbps)	Applicazioni	SIZE	Specifiche
MPEG-1 (MP3) - 1993	1.5	VHS -> VideoCD	360x280	Basso Bitrate
MPEG-2 - 1994	10	TV digitale	720x480	Multipazionne Audio-Video Video Interlacciato
MPEG-3	40	HDTV		
MPEG-4 - 1998				Gestione VO
MPEG-7				Organizza contenuti
MPEG-21				multimediali (XML) MPEG4 +MPEG7 Protezione digitale

NB → MPEG-3 fu abbandonato poichè l'MPEG-2 riuscì a ricoprire le sue specifiche

L'MPEG definisce le specifiche per gli standards di *bitstream*

Componenti

- MPEG-Video
- MPEG-Audio
- MPEG-Systems

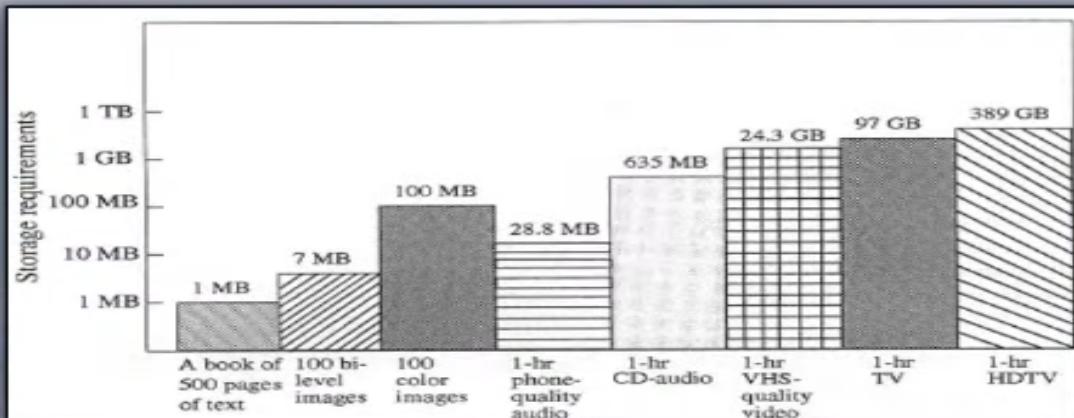
9.5 Documenti multimediali compositi (SGML)

Si va a rappresentare le parti in cui è strutturato il documento (paragrafi, capitoli) e altre particolarità del testo (note, tabelle, intestazioni)

9.6 PDF (Portable Document Format)

Formato basato su un linguaggi di descrizione completa di una pagina

9.7 Requisiti di memoria e larghezza di banda



Requisiti di Memoria di alcuni media comuni

Applicazioni	Velocità (kbps)
CD-Audio	1.411
DAT	1.536
Telefonia Digitale	64
Radio Digitale, long-play DAT	1.024
Video a Qualità Televisiva	216.000
Video a qualità VHS	54.000
HDTV	864.000

Velocità di trasmissione di alcuni media comuni

10 Progetto di Database multimediali

10.1 Architettura dei MIRS

- Modularità → si presuppone
 - flessibilità (libreria di funzioni);
 - gestione degli aggiornamenti;
- Distribuzione
 - gestione dei dati multimediali (client-server);
 - Accessi simultanei (library digitali, video on-demand);
- Presenza di moduli opzionali
 - Thesaurus manager → contiene sinonimi e altre relazioni tra le parole;
 - Integrity rule base → testa l'integrità di una data applicazione;
 - Context manager → tiene traccia del contesto dell'applicazione.



Fase 1 (inserimento)

1. L'utente specifica il tipo di input;
2. I contenuti dei dati in input vengono estratti automaticamente o semi-automaticamente;
3. Gli oggetti multimediali originali con le relative caratteristiche estratte sono inviati al (o ai) server;
4. Gli oggetti multimediali in arrivo vengono organizzati e sono dati in input al blocco seguente;
5. Esegue l'indicizzazione delle informazioni;
6. Memorizza le informazioni di indicizzazione e gli oggetti multimediali originari.

Fase 2 (recupero)

1. L'utente esegue la query di un elemento non necessariamente già contenuto nel DB;
2. Vengono estratti i contenuti della query, così come avviene per un inserimento;
3. Il client accoglie i dati dall'estrattore e li smista al motore di indicizzazione e ricerca;
4. Il server accoglie i dati dall'estrattore e li smista al motore di indicizzazione e ricerca;
5. Il motore di indicizzazione e ricerca reperisce l'elemento contenuto nel DB che meglio si adatta alla query;
6. Recupero dell'oggetto richiesto → la destinazione di tale oggetto è l'interfaccia di visualizzazione

10.2 Modello dei dati

In un DBMS la finalità della modellizzazione è di specificare tipi e proprietà degli oggetti che dovrà contenere.

In un MMDBMS o in un MIRS le finalità della modellizzazione comprendono anche una specifica dei diversi livelli di astrazione dei dati multimediali.

Un modello di dati per un MIRS deve comprendere la descrizione di:

- Proprietà statiche → riguardanti gli oggetti stessi che costituiranno i dati multimediali, le loro relazioni e i loro attributi;
- Proprietà dinamiche → riguardanti le interazioni tra gli oggetti, operazioni disponibili sugli oggetti, interazione con gli utenti.

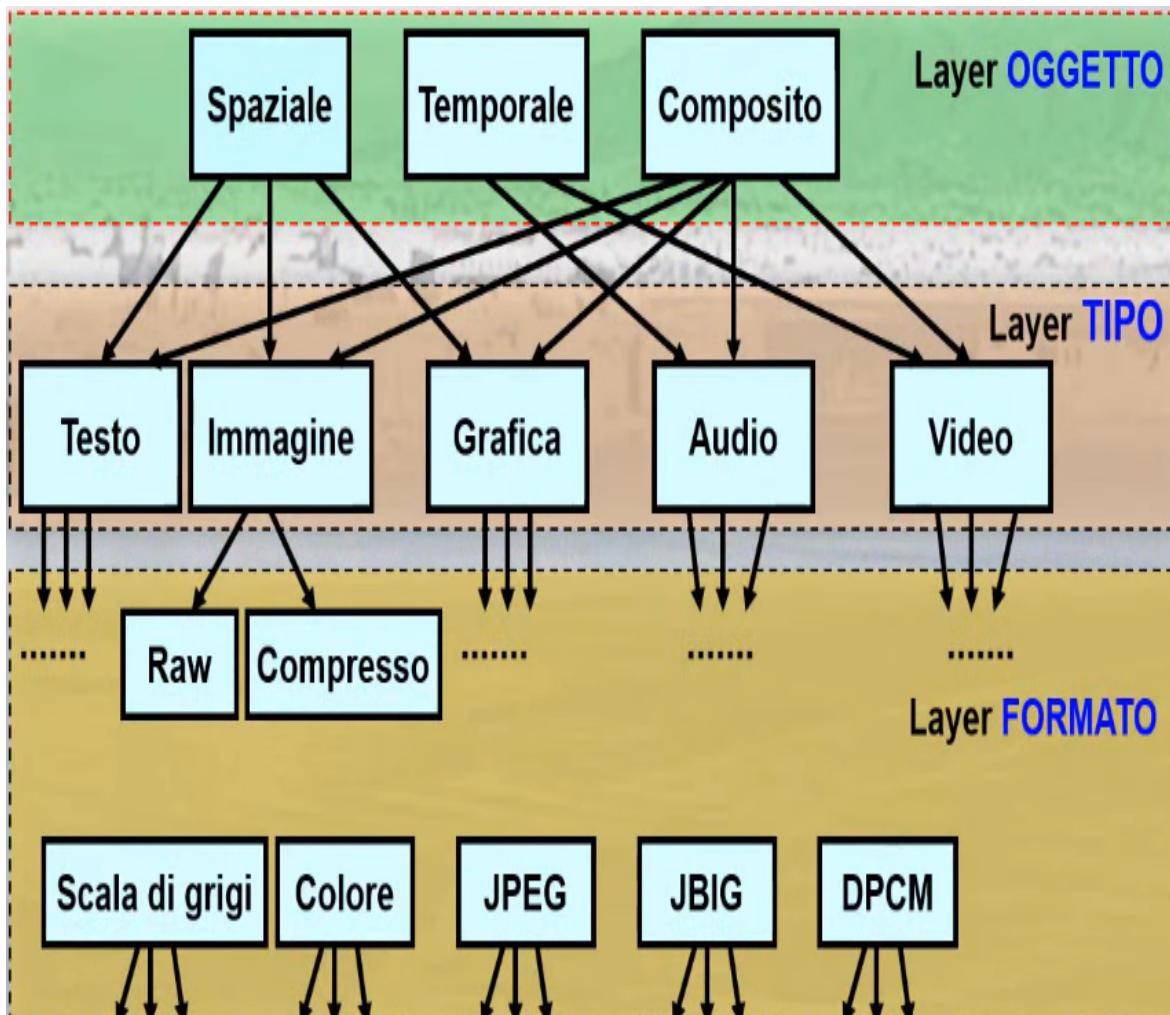
L'usabilità di un MIRS è fortemente condizionata dal modello dei dati.

l'indicizzazione di dati multimediali implica la considerazione di:

- spazi multidimensionali di caratteristiche;
- definizione di una metrica in tale spazio

10.3 Requisiti

- Estensibilità a nuovi tipi e formati di dati;
- Flessibilità per permettere l'inserimento e la ricerca a vari livelli di astrazione;
- Predisposizione per la rappresentazione dei dati multimediali semplici e composti comprese le relazioni spaziali e temporali che intercorrono tra di essi;
- Efficienza nelle strategie di memorizzazione e ricerca
 - il paradigma OO è il meglio adattabile alla modellizzazione dei dati multimediali;
- Incapsulamento di codice e dati in una singola unità chiamata oggetto;
- Il codice definisce le operazioni che possono essere effettuate sui dati
 - Struttura di tipo **Multilayer** → permette l'indicizzazione, la ricerca e l'elaborazione a diversi livelli di astrazione



10.3.1 Layer oggetto

Costituito da uno o più media con specifiche relazioni spaziali e temporali

- Relazioni spaziali
 - dimensione di ogni immagine;
 - posizione di apparizione;
- Relazioni temporali
 - tempo di inizio e tempo tra immagini;
 - sincronizzazione con il contenuto audio.

10.3.2 Layer TIPO

Contiene i tipi comuni di media. Questo tipo di informazioni sono utilizzate nella fase di ricerca e di calcolo della similarità

10.3.3 Layer FORMATO

Specifica il formato in cui il media è memorizzato:

- Raw;
- Compresso;
- Tipo di compressione usata.

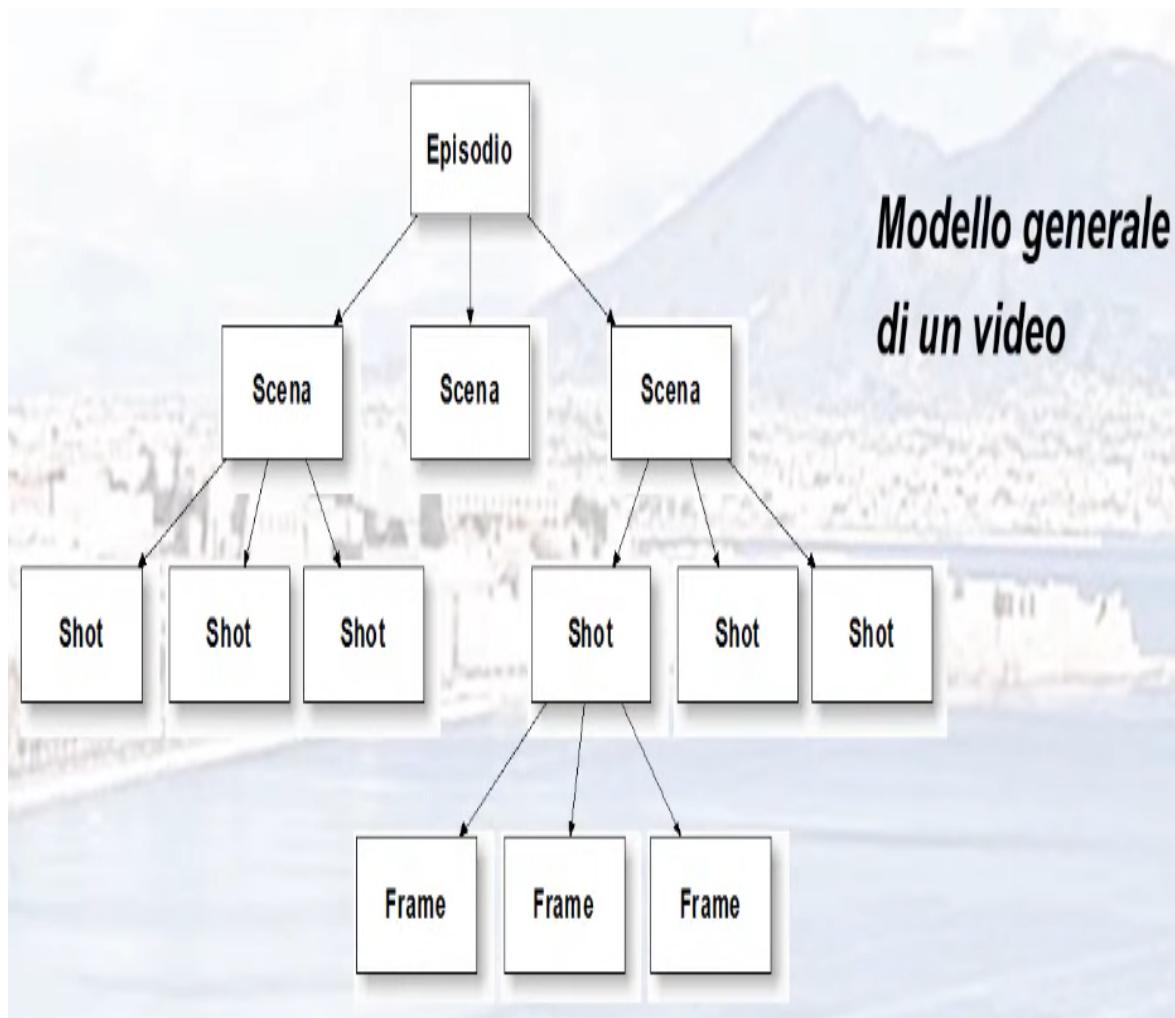
10.4 VIMSYS

Il modello VIMSYS (Visual Information Management System) per la gestione di video ed immagini è formato da 4 layer:



Tutti gli oggetti presenti in ogni layer hanno i propri attributi e metodi

10.5 Modello di un video



Ad ogni livello i dati vengono assegnati gli attributi relativi:

- Episodio;
- Scena;
- Shot;
- Frame.

10.6 Interfaccia utente

- Fornire strumenti per inserire oggetti nel DB in maniera semplice;
- Fornire strumenti per definire efficacemente le query e le esigenze di ricerca;
- Presentare i risultati delle ricerche in maniera efficiente ed efficace;
- Essere user-friendly.

10.6.1 Popolazione del DB

A differenza dei DBMS tradizionali in un MIRS i dati sono costituiti da media diversi e non hanno struttura ed attributi prefissati (gestione complessa).

L'interfaccia deve consentire l'inserimento di dati multimediali semplici e compositi e la specifica di tipologie di attributi che devono essere estratte ed indicizzate.

L'estrazione degli attributi può essere automatica o semi-automatica.

1. Tipologie delle query in un MIRS

- Multiformi → l'utente può inserirle utilizzando modi differenti e tipi di media differenti;
- Incerte → l'utente sa cosa vuole ma non sa descriverlo e riconosce il risultato corretto solo quando lo vede

10.7 Fase di ricerca

10.7.1 Tipi di ricerca

- Per specifica → attraverso attributi e feature su cui ricercare;
- Necessita di strumenti di authoring multimediali per mettere l'inserimento dell'esempio con diverse modalità

10.7.2 Estrazione delle feature

Gli oggetti multimediali gestiti dal DB sono preprocessati per estrarne feature ed attributi.

Il processo di ricerca si basa sulla ricerca e comparazione di tali feature → l'efficienza è basilare per ottenere un sistema di buona qualità.

1. Requisiti per l'estrazione delle feature

- Le feature estratte complete;
- Le feature devono essere memorizzate in maniera compatta;
- Il calcolo della distanza tra le feature deve essere veloce in modo che siano bassi i tempi di risposta del sistema.

2. Tipi di feature

- Metadata → dato che descrive un dato;
- Annotazioni testuali;
- Feature di basso livello;
- Feature di alto livello

10.8 Indicizzazione dei dati

È necessario utilizzare delle strutture di indicizzazione per organizzare la memorizzazione delle feature e fare in modo che la ricerca sia efficiente.

È necessario utilizzare strategie di indicizzazione adeguate. Questa può essere gerarchica e avvinire a più livelli:

- Può prendere in considerazione le relazioni spazio-temporali tra gli oggetti;
- Sono necessarie misure della similarità nello spaio delle feature che simulino il giudizio umano.

10.8.1 Misure di similarità

Il retrivial multimedialie è basato sulla similarità e non su matching esatto tra query ed elementi del DB;

La pertinenza dei risultati è giudicata da essere umani e quindi il maggior requisito delle misure di similarità e dei tipi di feature è che siano dei parametri adatti all'osservazione umana.

10.9 Garanzie sulla QoS (Quality of Service)

QoS specifica un insieme di parametri e requisiti richiesti in due gradi:

- Qualità preferibile;
- Qualità accettabile.

La QoS è in genere negoziata tra client e server e sottoscritt tramite un "contratto" che garantisce tali parametri in uno dei seguenti modi:

- Deterministico → la qualità è garantita pienamente;
- Statistico → la qualità è garantita con una certa probabilità;
- **Best-effort** → la qualità non è garantita.

10.10 Multimedia Data Compression

La maggior parte dei dati è salvata in formato compreso. Per l'estrazione delle caratteristiche degli oggetti multimediali occorre prima effettuare una decompressione. Esistono 3 diversi metodi

Metodo 1

- Sul server per ogni grande dato si salva anche una copia ridotta;
- La query dell'utente recupera sempre la copia ridotta;
- Se occorre anche il dettaglio allora oltre alla copia ridotta viene recuperata anche il dato originale;

Svantaggio → ridondanza dei dati sul server

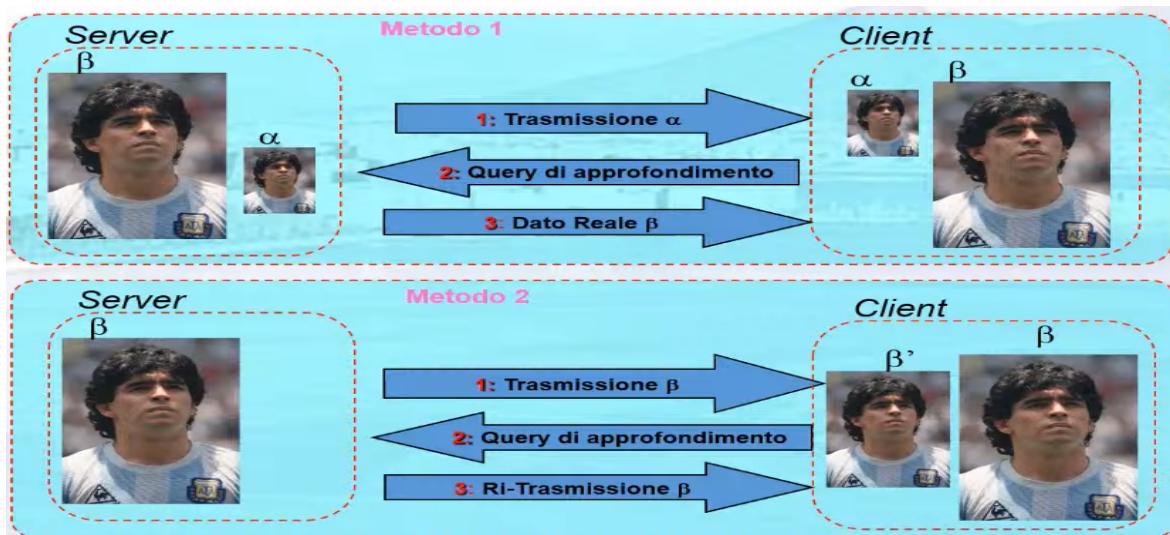
Metodo 2

- La query dell'utente recupera direttamente il dato originale;
- Il dato originale viene ridotto per poter essere rappresentata sul client
- Se occorre maggior dettaglio dettaglio allora il server ritrasmette il dato originale

Svantaggio → spreco di banda per la trasmissione.

Metodo 3

Si usano metodi di decompressione scalabili, progressivi e gerarchici



10.11 Standard di rappresentazione dei dati

L'estrazione delle caratteristiche ed il processo di confronto presume che per ogni media, il dato RAW sia lo stesso. Ciò non riflette la realtà, infatti:

- Diversi brani audio possono essere registrati a diversi livelli di amplificazione
 - il loro confronto potrebbe perdere di significato;
- Diverse immagini possono essere equalizzate in modo completamente diverso
 - il loro confronto potrebbe perdere di significato;

11 Il Testo

IR → tecniche di recupero delle informazioni (rilevanza storia).

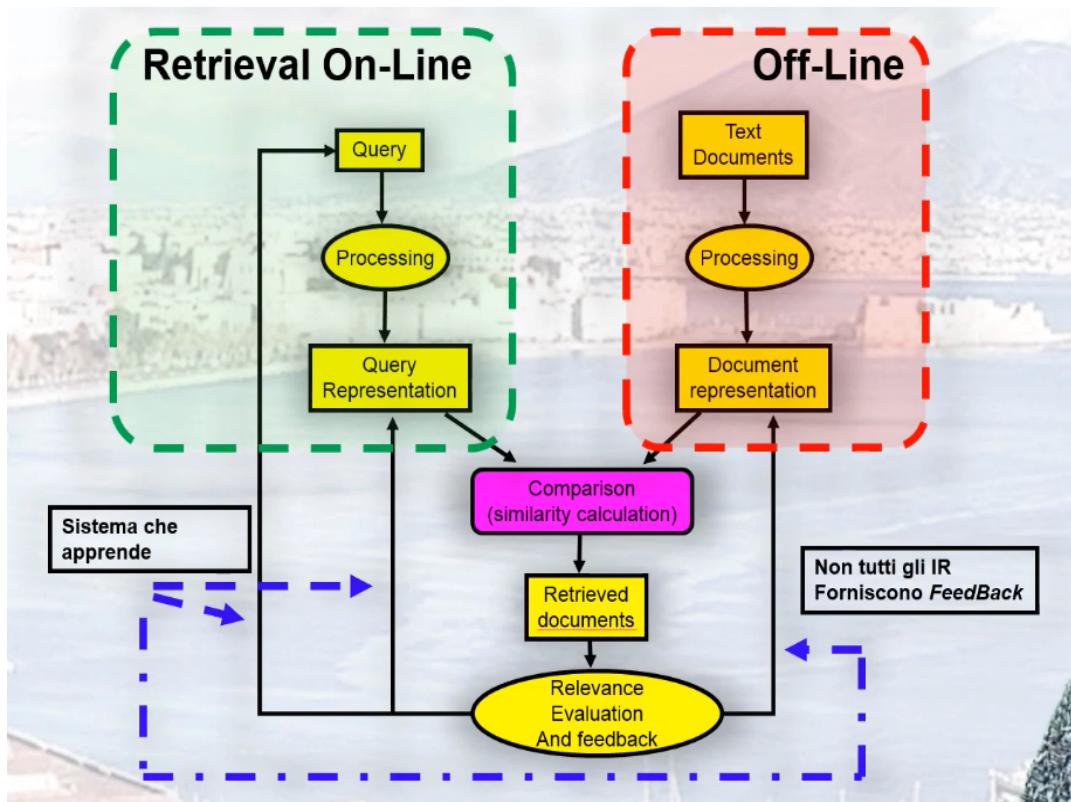
Il testo viene impiegato come strumento manuale di annotazione sfruttabile da un IR

Concetti di base

- Rappresentazione delle query e dei documenti;
- IR con diversi modelli di rappresentazione dei documenti ma simili processi di indicizzazione;
- Tecniche di confronto tra documenti e query;
- Modelli maggiormente usati
 - Match esatto;
 - Spazio Vettoriale;
 - Modello Probabilistico;
 - Modello su Cluster (raggruppamento).
- Query con linguaggio naturale;
- Tecniche di AI.

11.1 Differenze tra IR e DBMS

DBMS	IR
Struttura omogenea dei record	Record non strutturati
Componenti prefissati (record)	Attributi non prefissati
Record definito completamente e univocamente dai propri attributi	Indicizzazione attraverso keyword, descrittori, indici
Retrivial con match esatto (Query ↔ valore dei campi dei records)	Retrivial con match approssimato o parziale



11.2 Indicizzazione automatica e modello booleano per il retrieve

Retrieval con modello booleano

- Text-pattern search system;
- Pattern

- stringhe;
- espressioni regolari.

Strumenti → grep, egrep, awk.

es

```
grep root /etc/passwd
```

11.2.1 Struttura file

Il problema della scelta della rappresentazione della conoscenza:

- Flat file
 - ASCII;
 - Ricerca lineare;
- Inverted files;
- Signature files
 - Generazione di signature;
 - Metodo hash;
 - Confronto tra signature della query e del documento;
- Alberi;
- Grafi.

1. Inverted files Contiene un insieme di righe di testo. Ogni riga contiene:

- Il termine che si vuole cercare;
- una sequenza di puntatori a documenti e/o records che contengono quel termine

Spiega quindi l'inversione del verso di ricerca → prima la chiave e poi il documento che contiene la chiave

Inverted file:

Term ₁	Record ₁ ; Record ₃
Term ₂	Record ₁ ; Record ₂
Term ₃	Record ₂ ; Record ₃ ; Record ₄
Term ₄	Record ₁ ; Record ₂ ; Record ₃ ; Record ₄

Query	Output
Term ₁ AND Term ₃	Record ₃

Il processo di ricerca è più efficiente rispetto al flat-file → non si analizzano i documenti interi ma solo l'inverted file da cui si ricavano i collegamenti ai documenti che contengono la chiave (o soddisfano la query)

Inverted file con operazioni estese

È possibile raffinare la ricerca supponendo che sia significativa:

- Differenza di peso tra un termine e un altro;
- La posizione del termine;
- La frequenza del termine.

Definiamo quindi 2 operatori di prossimità:

- WITHIN SENTENCE
 - presenza dei 2 termini cercati nello stesso paragrafo del record recuperato;
- ADJACENT
 - presenza dei 2 termini confinanti nel record recuperato;

Struttura generale dell'inverted file esteso

Term i : Record_n°, Paragrafo_n°, Frase_n°, Parola_n°

Inverted file:

Information	R99, 10, 8, 3; R155, 15, 3, 6; R166, 2, 3, 1
retrivial	R77, 9, 7, 2; R99, 10, 8, 3; R166, 10, 2, 5

Query	Output
Information WITHIN SENTENCE retrivial	R99

11.2.2 Indicizzazione automatica

Il processo di indicizzazione del file di testo prevede diverse fasi. Lo scopo è filtrare il testo in modo da ottimizzare la significatività delle informazioni da considerare per le ricerche:

1. Stop Word
 - si escludono elementi insignificanti (dipendono dal linguaggio);
2. Stemming
 - si considerano solo i termini comuni di parole analoghe;
3. Thesaurus

- possibilità di sostituire diversi termini simili che compaiono nel testo con un unico termine (usando un vocabolario);

4. Weighting

- i termini che compaiono nel testo hanno diversa importanza

1. Operazioni Booleane con i pesi dei termini Termine → Record1,Peso1; Record3,Peso3; ...

Term1	R1,0.3; R3,0.5; R6,0.8; R7,0.2; R11,1
Term2	R2,0.7; R3,0.6; R7,0.5; R9,0.5
Term3	R1,0.8; R2,0.4; R9,0.7

Operatore	Descrizione	Esempio di Query	Output
OR	Si considera il peso maggiore dei records che contengono il termine della query	Term2 OR Term3	R1, R2, R9, R2, R7
AND	Si considera il peso minore dei records che contengono il termine della query	Term2 AND Term3	R2, R9
NOT	Si considera la differenza dei pesi dei termini che contengono il termine della query	Term2 AND NOT Term3	R3, R7

2. Calcolo dei pesi Il peso di un termine deve considerare il numero di volte in cui il termine compare sia nel documento sia nell'insieme complessivo dei documenti. Per calcolare i pesi:

$$W_{ij} = tf_{ij} \log\left(\frac{N}{df_j}\right)$$

Dove:

- W_{ij} rappresenta il peso del termine j nel documento i ;
- tf_{ij} rappresenta la frequenza del termine j nel documento i ;
- N rappresenta il numero totale dei documenti del DB;
- df_j rappresenta il numero dei documenti del DB che contengono il termine j

11.3 Retrial con modello nello spazio vettoriale

Si sfrutta il prodotto scalare tra vettori

Il modello spazio vettoriale presume l'esistenza di un determinato insieme di termini che rappresentano i Documenti e le Query. Un documento D_i e una query Q_j sono definite nel seguente modo:

$$D_i = [T_{i1}, T_{i2}, \dots, T_{ik}, \dots, T_{iN}]$$

$$Q_j = [Q_{j1}, Q_{j2}, \dots, Q_{jk}, \dots, Q_{jN}]$$

Dove:

- T_{ik} è il peso del k-esimo termine relativo al documento i;
- Q_{jk} è il peso del k-esimo termine relativo alla query j;
- N è il numero totale di termini usati nel documento e nella Query

11.3.1 Calcolo della similarità

$$S(D_i, Q_j) = \sum_{K=1}^N T_{ik} Q_{jk}$$

$$S(D_i, Q_j) = \frac{\sum_{K=1}^N T_{ik} Q_{jk}}{\sqrt{\sum_{K=1}^N T_{ik}^2 \sum_{K=1}^N Q_{jk}^2}}$$

(prodotto dei moduli dei vettori T_{ik} e Q_{jk})

CONTINUA A 1 ora e 06

11.3.2 Tecniche basate su Relevance feedback

L'idea è quella di sfruttare le valutazioni degli utenti permettendo di raffinare i pesi dei termini sia della query sia dei documenti. La regola applicata è la Formula di Rocchio:

$$\vec{Q}_m = (a \vec{Q}_o) + (b \frac{1}{|D_r|} \sum_{\vec{D}_j \in D_r} \vec{D}_j) - (c \frac{1}{|D_{nr}|} \sum_{\vec{D}_k \in D_{nr}} \vec{D}_k)$$

Variabile	Valore
\vec{Q}_m	Vettore della Query modificato
\vec{Q}_o	Vettore della Query originale
\vec{D}_j	Vettore del documento relativo
\vec{D}_k	Vettore del documento non relativo
a	Peso originale della query
b	Peso del documento relativo
c	Peso del documento non relativo
D_r	Set di documenti relativi
D_{nr}	Set di documenti non relativi

Si passa quindi alla modifica dei documenti con importanti vantaggi:

- La modifica di un documento beneficia anche altri utenti;
- I valori dei pesi dei documenti vengono aggiornati considerando la query stessa;
- Vengono aumentati i pesi dei termini che si trovano sia nella query che nell'insieme dei documenti rilevanti;
- Vengono diminuiti i pesi dei termini che si trovano nell'insieme dei documenti rilevanti ma non nella query.

11.3.3 Altri modelli per il Retrieval

Modello probabilistico

- Considera la dipendenza dei termini e le loro principali relazioni;
- Scarso successo a causa dell'elevata difficoltà di calcolo delle probabilità su cui esso è fondato;

Modello basati su cluster

Si basa sull'idea di definire un insieme di gruppi in cui ciascun gruppo contiene elementi "simili". La query poi cercherà all'interno del cluster

- Similarità per coppie
 1. Ogni documento è rappresentato come un vettore;
 2. Si calcola la similitudine per ogni coppia di documenti e si popola una matrice delle distanze;
 3. Inizialmente ogni documento viene posizionato in un cluster;
 4. La coppia meno distante formerà un agglomerato (si rimuove la coppia dalla matrice sostituendola con l'oggetto congiunto);
 5. Si ripete fino ad ottenere una matrice da un solo elemento.
- Clustering Euristico
 1. Si considera ul primo documento (costituisce il 1° cluster);
 2. Ogni documento viene confrontato con il cluster generato e viene quindi collocato nel cluster più "vicino";
 - Nel caso la distanza sia eccessiva si fonda un nuovo cluster;
 3. Si ripete il processo fino ad esaurimento dei documenti.

Il primo metodo ha una complessità temporale elevata ma la generazione prodotta è unica.

Il secondo metodo non ha una complessità temporale elevata, ma la generazione prodotta non gode di univocità

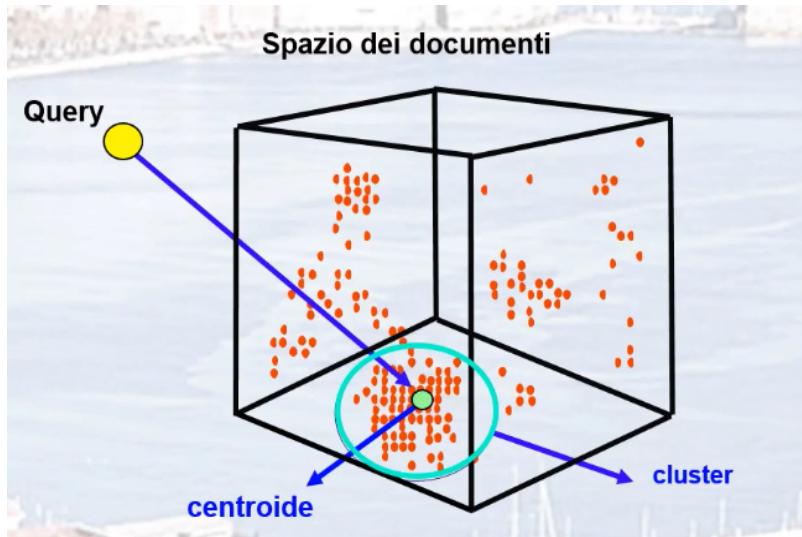
Cluste-Based Retrieval

Dopo aver generato i cluster sia le ricerche che il recupero dei documenti risulteranno efficienti ed efficaci.

Ognuno di questi cluster è caratterizzato dal proprio centroide che è calcolato mediando tutti i vettori del cluster.

Il processo di retrieval confronta la query con i centroidi dei cluster e selezionerà il cluster più simile.

L'output sarà costituito dai documenti del cluster individuato (tutti i documenti del cluster o i documenti del cluster con maggior grado di similitudine)



11.4 Misurazione delle prestazioni

Si basa su:

- Vecolità di ricerca;
- Recall → capacità di recuperare informazioni rilevanti dal DB
 - rapporto tra il numero di documenti rilevanti recuperati ed il numero totale di elementi rilevati nel DB;
- Precisione → misura l'accuratezza dei documenti recuperati
 - rapporto tra il numero di documenti rilevanti recuperati ed il numero totale di documenti recuperati.

$$\text{precisione} = \frac{|docR \cap docP|}{|docR|}$$

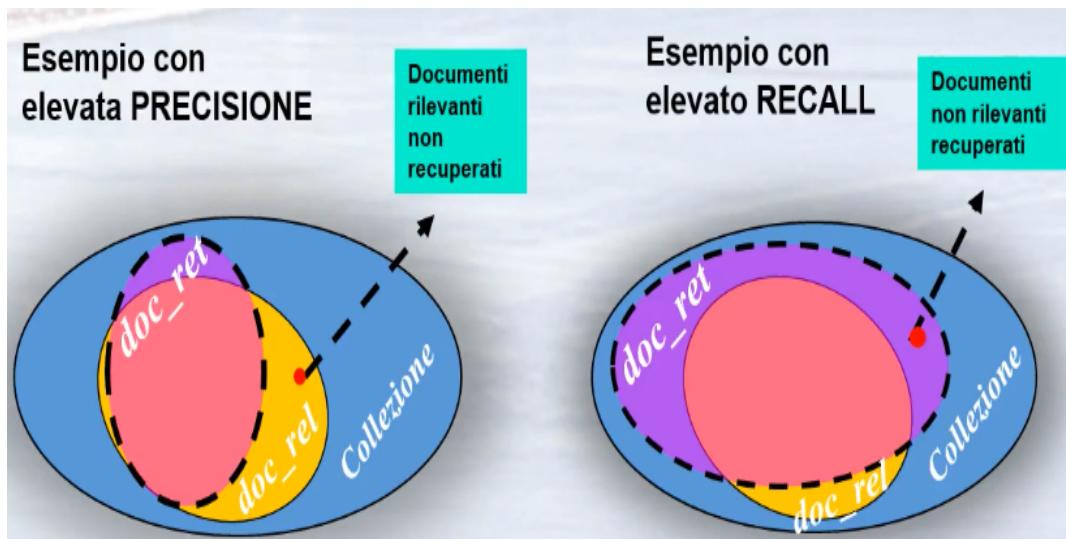
$$\text{recall} = \frac{|docR \cap docP|}{|docP|}$$

Dove:

- docR → documenti recuperati;
- docP → documenti pertinenti.

Nella pratica, i parametri di valutazione recall e precisione vengono valutati in modo congiunto, ed una buona prestazione è indice di un oculato compromesso.

Tipicamente quanto maggiore è la recall tanto minore risulta la precisione (e viceversa)



11.5 Tecniche di IR a confronto

- L'indicizzazione automatica ha una prestazione simile a quella manuale
 - si ottiene un reale miglioramento mixandole;
- Impiegando un insieme di query simili la performance del recupero mediante confronti parziali risulta migliore del match booleano esatto;
- Modello probabilistico e spazio vettoriali hanno performance analoghe;
- Tecniche di recupero basate cluster e sul modello Probabilistico hanno performance analoghe;
- L'uso del Relevance feedback consente un reale miglioramento della prestazione;
- L'uso di uno specifico dominio di conoscenza e del profilo utente che effettua la query, produce un significativo miglioramento della prestazione

11.6 Motori di ricerca www

I motori di ricerca sul www costituiscono il tipo di applicazione più utilizzato sul Web.

I documenti web sono memorizzati come Ipertesti in formati quali HTML. Questi sono strutturati in:

- Nodi;
- Link;
- Anchor

11.7 Modello di comunicazione

Presenti diverse regole di base che governano la comunicazione (protocolli), come:

- HTTP;
- HTTPS;
- FTP;
- FTPS.

11.8 Il web

- URL (uniform resource locator) → indirizzo che univocamente identifica ogni documento della rete
 - Protocoll://Servername[:port]/Path/Document-name.

11.8.1 Crawler (spider)

Tipo di bot che analizza i contenuti di una rete (o di un DB) in un modo metodico e automatizzato, in generale per conto di un motore di ricerca

Nome Spider	Motore di ricerca
googlebot	Google - Yahoo!
fast	Fast-Alltheweb
slurp	Yahoo!
ia_archiver	Alexa

Poichè i documenti selezionati sono molteplici per poterli mostrare all'utente è necessario organizzarli (Ranking).

Un modo per organizzare i documenti, assegnando loro dei pesi, è passato sulla formattazione del testo HTML

12 Indicizzazione e Recupero dell'audio

Il modo più utilizzato per classificare vari brani audio è basato sul titolo o sul nome del file → è chiara la soggettività di tale metodologia e sia l'incapacità di poter supportare query per esempio.

Ulteriore complicazione nasce dal fatto che non esiste uno standard di memorizzazione del file audio e ciò comporta grosse problematiche per il loro confronto → È necessario sviluppare tecniche e metodi di retrieving

12.1 Approccio generale

Si realizza un primo livello di distinzione:

- Parlato;
- Musica;
- Rumori.

La gestione è quindi diversificata per tipologie.

Ad es lo studio del parlato consiste nel convertire il file audio in parole di testo (**speech recognition**) su cui si potrà effettuare una query tradizionale

12.2 Proprietà e caratteristiche principali dell'audio

I segnali audio vengono rappresentati:

- Nel dominio temporale;
- Nel dominio delle frequenze.

Ciascun tipo di rappresentazione è particolarmente idonea per l'estrazione di determinate caratteristiche.

Oltre alle caratteristiche estraibili dall'audio rappresentato nei due precendi domini, è possibile estrarre caratteristiche che possono essere soggettive

12.2.1 Caratteristiche derivabili dal Time Domain

Tecnica più immediata ed intuitiva per la rappresentazione di un segnale la cui ampiezza varia nel tempo.

Il silenzio è rappresentato dallo 0. I valori di segnali possono essere positivi e negativi a seconda che la pressione d'aria provocata dall'onda sonora sia superiore o inferiore alla pressione atmosferica in condizioni di silenzio

Average Energy (Energia media)

Indica la *rumorosità* del segnale audio ed è calcolabile mediante la relazione:

$$E = \frac{\sum_{n=0}^{N-1} x(n)^2}{N}$$

Dove:

- $E \rightarrow$ energia media del brano audio;
- $N \rightarrow$ numero totale dei campioni valutati;
- $x(n) \rightarrow$ valore del campione n -esimo

Zero Crossing Rate (Frequenza di passaggio per lo 0)

Indica con quale frequenza cambia segno l'ampiezza del segnale ed è calcolabile mediante la relazione:

$$\text{ZCR} = \frac{\sum_{n=1}^N |\operatorname{sgn} x(n) - \operatorname{sgn} x(n-1)|}{2N}$$

Dove:

- $\operatorname{sgn} x(n) \rightarrow$ segno di $x(n)$ e assume valori
 - 1 se $x(n) > 0$;
 - -1 se $x(n) < 0$;

Silence Ratio (Quantità di silenzio)

Indica la proporzione di silenzio nel brano musicale; è il periodo entro il quale i valori assoluti di ampieza di un certo numero di campioni (e non solo un singolo valore) e per un "certo" tempo siano prossimi ad una soglia specifica

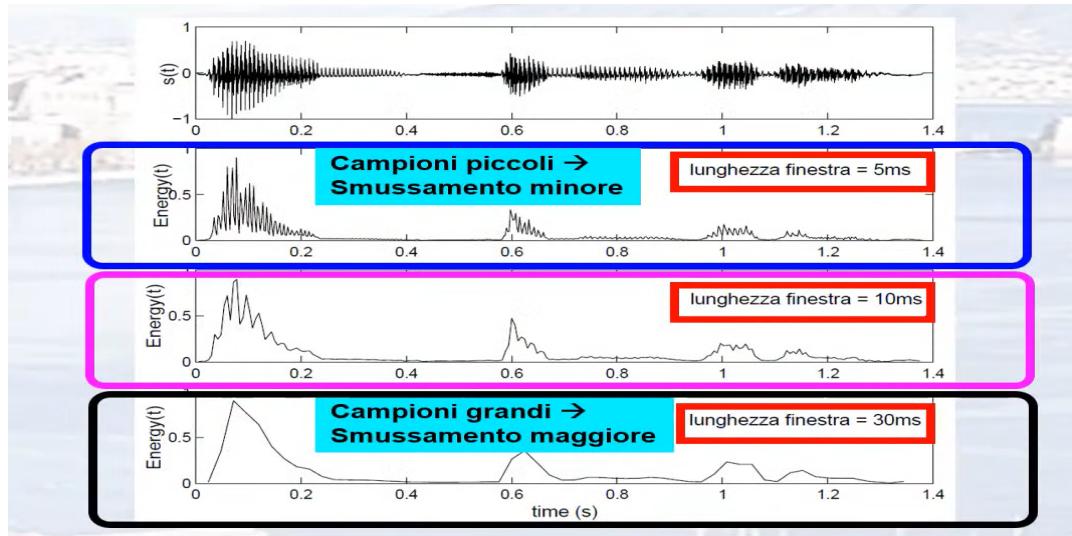
$$\text{Silence Ratio} = \frac{\text{Somma dei periodi di silenzio}}{\text{Lunghezza totale del brano}}$$

Average Magnitude

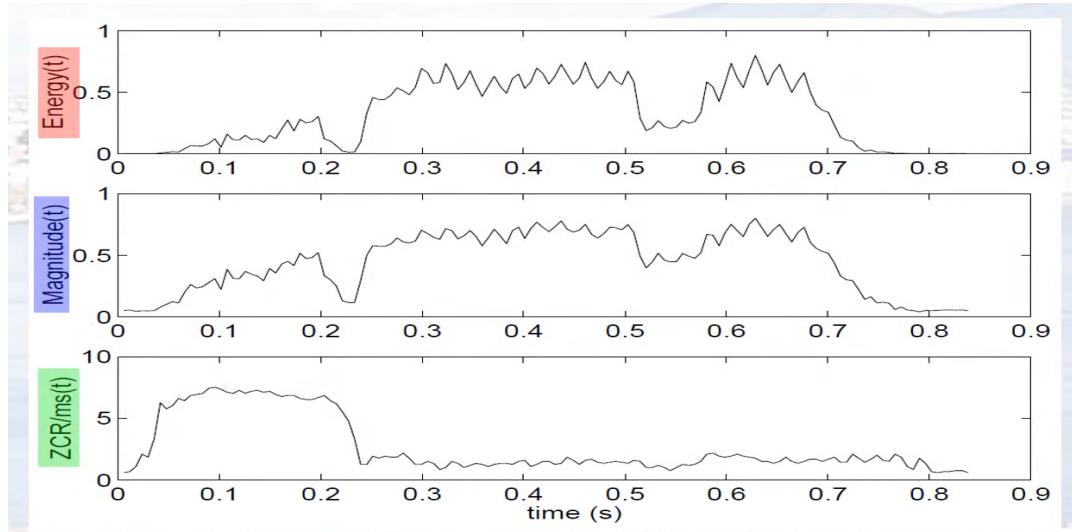
Poichè l'Average Energy aumenta fortemente per grandi ampiezze di segnale (i campioni compaiono al quadrato) si introduce il concetto di *Average Magnitude*:

$$M = \frac{\sum_{n=0}^{N-1} |x(n)|}{N}$$

Fissata la frequenza di campionamento e facendo invece variare la grandezza del campione (la lunghezza della finestra) la funzione Energia assume un andamento più smussato



ZCR , Magnitude ed Energy a confronto



12.2.2 Caratteristiche derivabili dal Dominio delle Frequenze

La rappresentazione nel dominio delle frequenze deriva dalla rappresentazione nel dominio temporale applicando la trasformata di Fourier

Dominio Temporale → Fourier → Dominio delle frequenze.

Nel Dominio delle frequenze il segnale viene rappresentato come ampiezza che varia in dipendenza della frequenza
→ tale rappresentazione mostra in che modo è distribuita l'energia alle varie frequenze.

La rappresentazione nel dominio delle Frequenze è comunemente detta **Spettro del Segnale**

Attraverso una cromia è possibile conoscere l'ampiezza del segnale.

Bandwith

Gamma (o range) delle frequenze di un suono. Per calcolarlo:

$$\text{Range_frequenze} = \text{Massima_frequenza} - \text{Minima_frequenza}$$

Lo spettro del segnale facilità la valutazione della distribuzione delle frequenze componenti.

La presenza di alte frequenze nel segnale audio comporta che con alta probabilità il segnale in oggetto contiene musica.

7 kHz rappresenta un buon valore nello spettro per la soglia che determina se un file audio contiene parlato o musica.

La classificazione di frequenze *alte* o *basse* dipendono dall'applicazione che dovrà trattare tali segnali.

L'energia complessiva di ciascuna banda è la somma dei suoi componenti.

Centroide Spettrale

Individua il punto medio della distribuzione di energia di un suono.

Il centroide del parlato è inferiore al centroide della musica

Armoniche:

Un suono prodotto da un corpo vibrante non è mai puro, ma è costituito da un amalgama in cui il suono fondamentale se ne aggiungono altri più acuti e meno intensi → gli armonici pertanto hanno una importanza fondamentale nella determinazione del timbro di uno strumento e nella determinazione degli intervalli musicali

Le armoniche di un suono sono multiple in frequenza rispetto a una frequenza più bassa detta **frequenza fondamentale**.

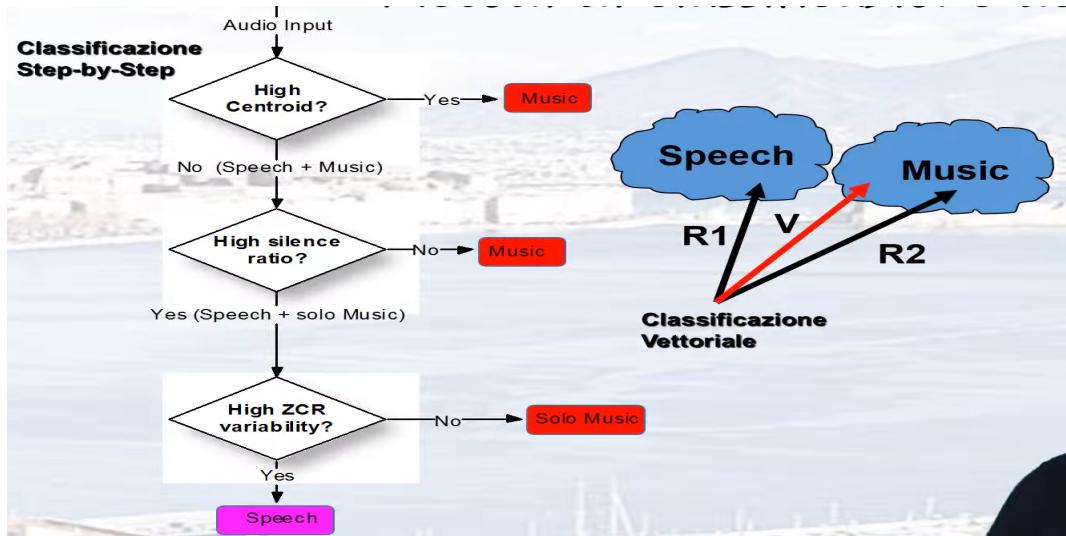
12.3 Classificazione dei segnali audio

Le principali categorie oggetto di studi di classificazione dei suoni sono la Musica e il Parlato. Le principali caratteristiche/differenze sono riassunte nella seguente tabella:

Caratteristiche	Parlato	Musica
Larghezza di banda	0-7 kHz	0-20 kHz
Centroide Spettrale	Basso	Alto
Quantità di silenzio	Alto	Basso
ZCR	Molto variabile	Meno variabile
Ritmo regolare	No	Si

12.4 Metodi di classificazione dell'audio

12.4.1 Classificazione Step-By-Step



Riconoscimento del parlato

L'approccio fondamentale per l'indicizzazione ed il recupero del parlato è basato sulla conversione dei segnali audio vocali in testo su cui successivamente applicare tecniche di IR.

Il problema del riconoscimento del parlato (ASR) viene ricondotto ad un problema di pattern matching.

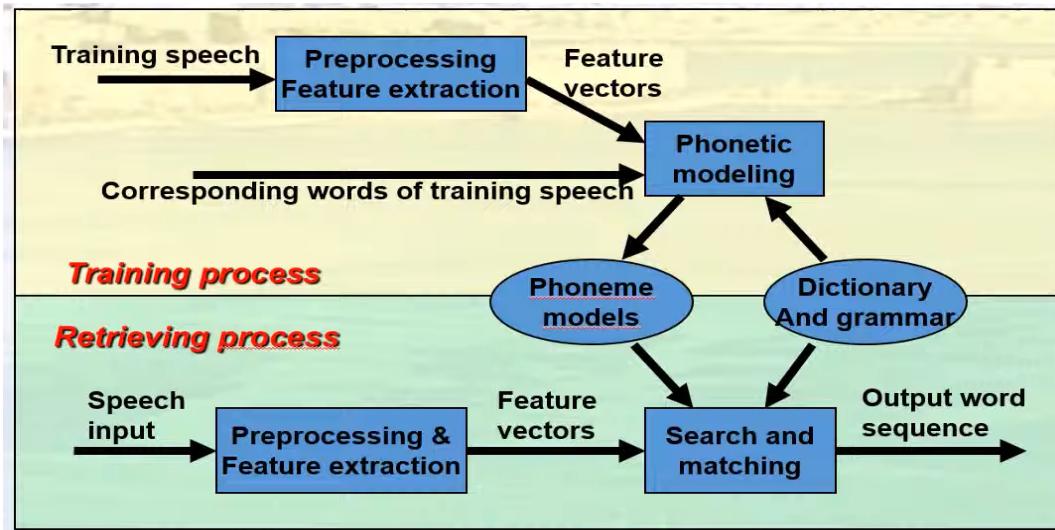
In generale si suddivide il parlato in singole unità ognuna delle quali viene confrontata con i vettori di feature raccolti nella fase di training → in tal modo viene trovato il matching migliore utilizzando la distanza euclidea tra i vettori di feature

12.4.2 Classificazione basata su caratteristiche vettoriali

I valori di un insieme di caratteristiche del suono considerato costituiscono le componenti di un vettore V che verrà confrontato con un altro vettore di caratteristiche R che rappresenta il vettore di riferimento di ciascuna classe di pezzi audio

12.5 Concetti base dell'ASR (Automatic Speech Recognition)

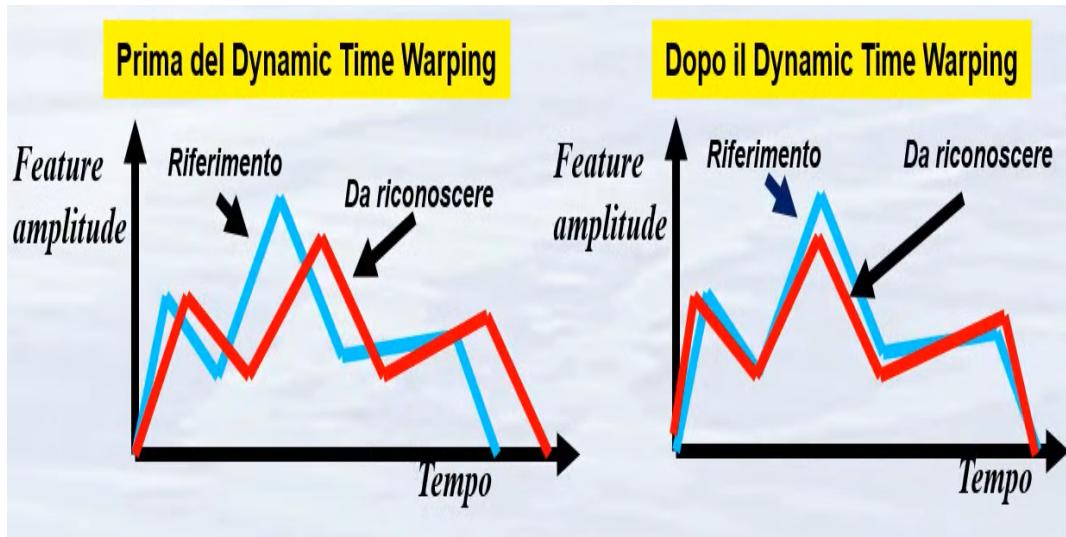
Il processo di riconoscimento è quindi un processo **statistico** che può essere migliorato fornendo una conoscenza del linguaggio utilizzato.



12.6 Tecniche basate sul *Dynamic Time Warping* (DTW)

Questa tecnica cerca di normalizzare la durata dei frame del parlato da riconoscere con quella dei frame memorizzati durante la fase di training → tale tecnica si basa sul considerare le variazioni temporali in modo non lineare:

- Si *dilata* o si *contrae* l'asse dei tempi in modo da far coincidere picchi di segnale:



Il problema risulta complesso perchè:

- Persone diverse possono impiegare tempi diversi per pronunciare lo stesso fonema;
- La stessa persona può pronunciare lo stesso fonema in maniera differente

L'obiettivo primario del DTW consiste nel confrontare 2 sequenze temporali $X=(x_1, x_2, \dots, x_N)$ e $Y=(y_1, y_2, \dots, y_M)$ e cercare il minimo costo complessivo di una serie scelta di coppie (x_i, y_j)

Dal punto di vista intuitivo, un allineamento ottimale tra le sequenze X ed Y percorre un sentiero di avvallamenti più profondi possibile (valori bassi) all'interno della matrice C

12.7 Tecniche basate su Hidden Markov Models (HMM)

La parola *nascosto* indica che per un osservatore non è visibile la sequenza di *stati*, ma solo la sequenza di output dei simboli.

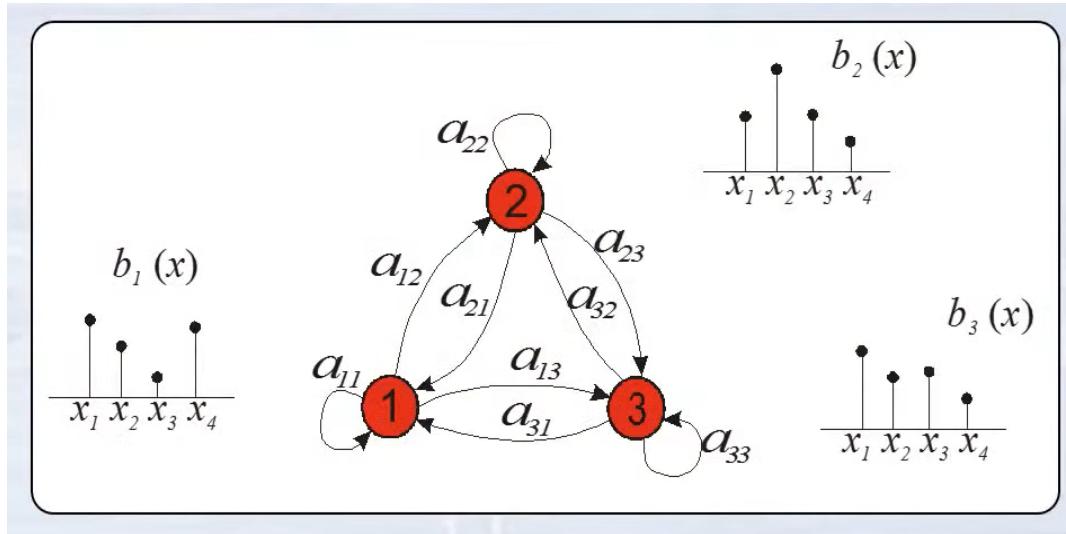
Questa tecnica è molto usata e porta a buoni risultati sia per il riconoscimento dello scritto che del parlato.

Una HMM è costituita da:

- Stati;
- Probabilità di transizione;
- Probabilità di generazione dei simboli.

Durante la fase di training viene costruita una HMM che per ogni fenomeno considera (sottoforma di probabilità) la variabilità degli speaker, il rumore di fondo e le differenze temporali.

Durante il riconoscimento occorre trovare quale HMM è più probabile che generi la sequenza di feature vector di input



Dove:

- Simboli $\rightarrow x_1, x_2, x_3, x_4$;
- Probabilità dei simboli $\rightarrow b_i(x)$;
- Probabilità di transazione $\rightarrow a_{ij}$;

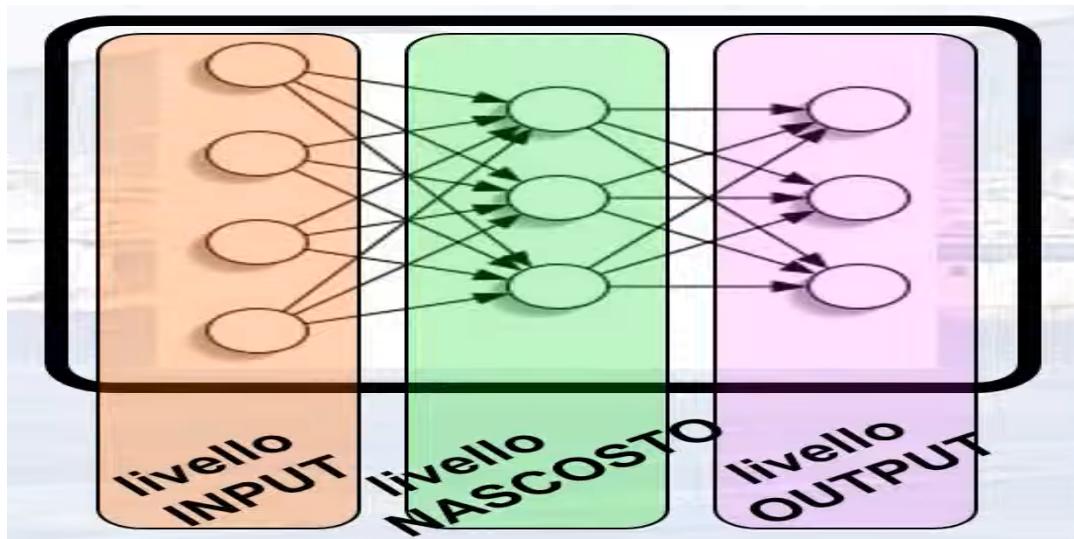
12.8 Tecniche basate su Reti Neurali Artificiali (ANN)

Simula una rete interconnessa da link con peso. Costituita da 2 fasi:

- Training \rightarrow i vettori di caratteristiche ottenuti durante l'addestramento di parlato servono per tarare i pesi dei link della rete;
- Recognition \rightarrow L'ANN seleziona il fonema più verosimile basandosi sulle caratteristiche dei vettori.

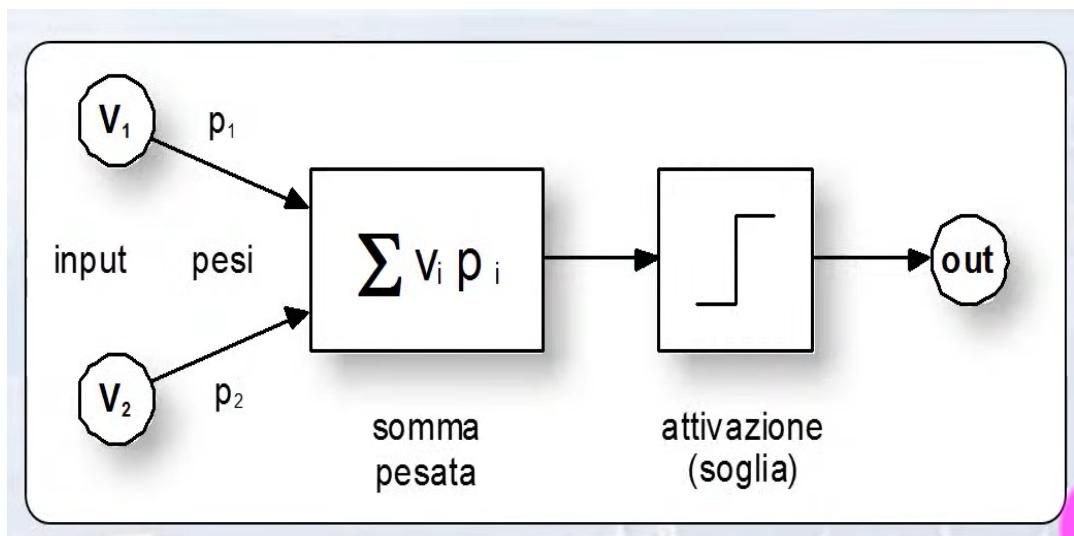
Lo scopo della fase di training è quella di fornire un risultato più probabile per un input assegnato.

Le ANN sono basate su implementazioni che prevedono l'utilizzo di un grande numero di elementi di calcolo (PE \rightarrow Processing Elements) molto semplici e interconnessi tra loro. Ogni PE implementa una semplice funzione matematica di tipo non lineare e rappresenta un **NEURONE**.



12.8.1 Elaborazione di UN PE

- Ogni neurone della rete effettua una somma pesata INTEGRAZIONE degli input derivanti dalle connessioni con gli altri neuroni;
- L'input pesato viene poi valutato da una funzione detta TRASFORMAZIONE che determina l'output del singolo PE;
- Normalmente le funzioni di trasformazione sono funzioni molto semplici e non lineari (FUNZIONI SOGLIA).



12.8.2 Funzioni di elaborazione

- Hanno il compito di restituire l'output di relazione all'input totale ricevuto sul neurone;

- Si utilizzano delle funzioni a soglia che danno luogo ad un'attivazione del neurone solo nel caso in cui l'input su tale neurone superi una soglia pre-determinata;
- Questo simula il comportamento dei neuroni reali i quali reagiscono solo se stimolato sopra una certa soglia;

Tra le più utilizzate abbiamo:

- Lineari;
- Gradino;
- Sigmoide;
- Semi-Lineare.

12.9 Tecniche di identificazione dello speaker

Cercano di estrarre informazioni su chi sta parlando

12.10 Relazioni tra audio e altri media

L'audio molto spesso è parte di un oggetto multimediale composito dove esistono delle forti relazioni temporali tra video e audio.

Possiamo quindi utilizzare la conoscenza su uno dei media per migliorare l'indicizzazione e la comprensione del contenuto dell'altro media