

Le Immagini, l'Audio e i Video

(a cura di R. Mascella e L. Tallini)

1. **Introduzione**
2. [La Grafica](#)
3. [L'Audio](#)
4. [I Video](#)

Introduzione

La digitalizzazione degli oggetti audio-visivi, che noi percepiamo attraverso onde elettromagnetiche con particolari frequenze (onde visibili, onde uditive), avviene sulla base di due differenti parametri:

- il numero di **campionamenti**, o di misurazioni, spazio-temporali. Queste misurazioni, effettuate con l'ausilio di appositi sensori, sono spaziali nel caso delle immagini, temporali nel caso dei suoni, sia spaziali che temporali nel caso dei video, che genericamente sono composizioni di immagini in sequenza e suoni;
- il numero di **stati fisici discreti**, o di livelli, dell'informazione "digitale". I livelli nel loro insieme costituiscono il codice di rappresentazione ed ogni singola misurazione viene così rappresentata attraverso uno di questi livelli.

Questi parametri sono tra loro complementari e interagiscono nella descrizione digitale degli oggetti visibili e/o udibili. A priori, l'accuratezza nella descrizione dell'oggetto sarà tanto maggiore quanto maggiori saranno le misurazioni e/o i livelli di rappresentazione utilizzati. Ad esempio nella prima figura sotto riportata il numero di misurazioni non è sufficiente perché alcune variazioni improvvise del segnale non vengono colte, per cui si rende necessario, come nella seconda figura riportata, eseguire misurazioni più fitte.

Fig. 1: Campionamento del segnale con 8 misurazioni.

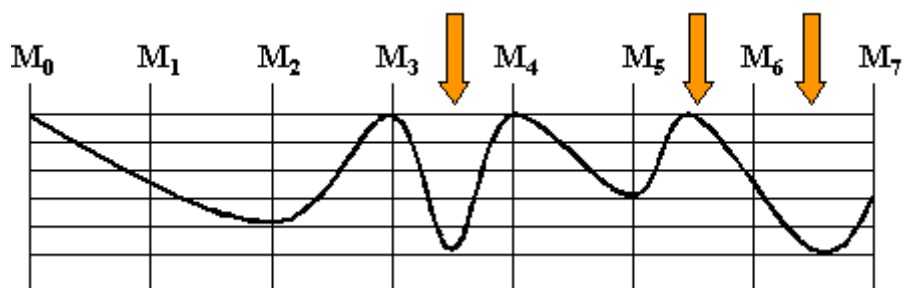
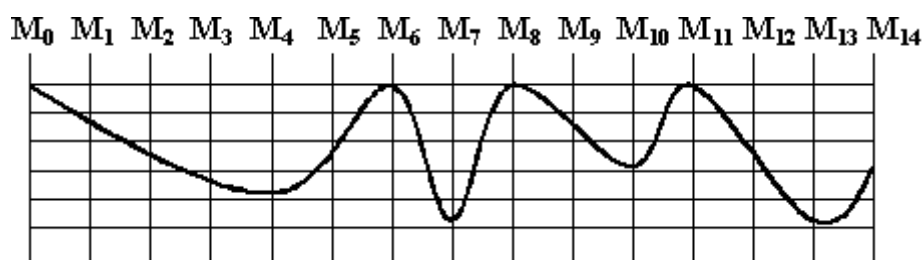


Fig. 2: Campionamento del segnale con 15 misurazioni.



Da un lato, per avere rappresentazioni di buona qualità (per i nostri sensi), è importante che questi numeri siano molto grandi; dall'altro, per avere rappresentazioni poco "pesanti" in termini di occupazione di memoria, è importante che gli stessi numeri siano molto piccoli. Queste due esigenze contrapposte non possono trovare risposta se non in una situazione di perenne compromesso. Fortunatamente, per la risoluzione di questo conflitto intervengono delle tecniche di compressione che permettono di considerare parametri più grandi a fronte dell'eliminazione (nel salvataggio degli oggetti) di informazioni "ripetitive", "ridondanti" o "inutili" secondo criteri predeterminati.

In queste dispense, dopo aver richiamato le principali proprietà fisiche delle onde, che in fondo consistono in "una serie di compressioni e decompressioni percepibili ai nostri sensi in un mezzo di trasmissione (di solito l'aria)", si descrivono i principali concetti sul loro trattamento digitale, che va dal campionamento fino all'elaborazione e al salvataggio finale in un file con formato specifico. Alla fine, non si ha più l'onda, ma una fila di numeri, che corrispondono alla digitalizzazione dell'onda. In fase di lettura e/o di output questi numeri vengono riconvertiti in analogici... per i nostri sensi.

Buona lettura.

Disclaimer. L'informazione contenuta in queste dispense è ritenuta essere accurata alla data della pubblicazione. Le presenti dispense hanno finalità e carattere esclusivamente didattico. Possono essere liberamente utilizzate, modificate e riprodotte esclusivamente per motivi scolastici, didattici e istituzionali, e non a fine di lucro.

La Grafica

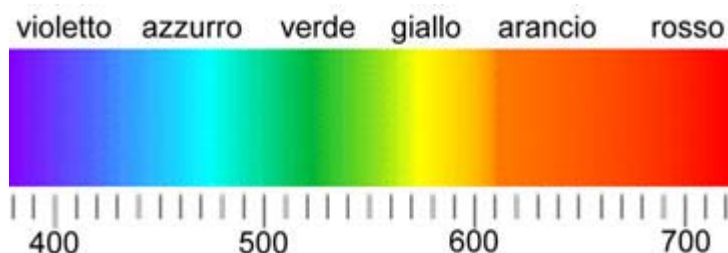
[torna all'indice](#)

1. La luce, lo spettro ed i modelli di colore

La luce è un'onda elettromagnetica percepita dai nostri occhi. Ad essa sono associabili, come d'altro canto a tutte le onde, due grandezze fisiche: la lunghezza d'onda (la lunghezza del ciclo, misurata in metri) e la frequenza (il numero di volte che il ciclo si ripete in un secondo, misurata in cicli/sec = Hertz o Hz), tra loro in relazione in quanto il loro prodotto è la velocità dell'onda (ovvero, nel caso delle onde elettromagnetiche, la velocità della luce $c = 2,99792458 \times 10^8 \text{ m/s} \sim 300000 \text{ km/s}$). Tutto ciò che i nostri occhi percepiscono è, di fatto, una combinazione di onde elettromagnetiche con lunghezze diverse, appartenenti ad uno specifico intervallo delle lunghezze che va dai 380 nanometri ($1 \text{ nm} = 10^{-9} \text{ m}$) ai 780 nm, ovvero con frequenze appartenenti all'intervallo tra i 780 GHz e i 384 GHz.

La nostra percezione del colore avviene in quanto ogni oggetto illuminato assorbe una parte delle radiazioni riflettendone la parte restante, identificabile attraverso una specifica frequenza (o lunghezza d'onda) o un insieme di esse. Le onde riflesse vengono così percepite dai nostri occhi attraverso dei sensori (i coni) disposti sulla retina ed inviate al cervello che restituisce le caratteristiche "virtuali" dell'oggetto. Ad ogni lunghezza d'onda corrisponde uno specifico colore, come si evince dalla figura seguente che rappresenta l'intero spettro dei colori:

Fig. 1: Lo spettro dei colori visibili con relative lunghezze d'onda (esprese in nanometri).



Il range così rappresentato, inoltre, è delimitato a sinistra dalle cosiddette radiazioni ultraviolette e a destra dai raggi infrarossi, che costituiscono, così, le radiazioni più vicine a quelle visibili.

I coni della retina sono suddivisi in tre categorie:

- particolarmente sensibili attorno ai 437 nm (quindi i più sensibili alla gamma dei blu),
- particolarmente sensibili attorno ai 533 nm (quindi i più sensibili alla gamma dei verdi),
- particolarmente sensibili attorno ai 564 nm (quindi i più sensibili alla gamma dei rossi).

Per questo motivo i colori rosso, verde e blu sono detti *colori primari*. Chiaramente i singoli segnali percepiti dai coni sono "sommati" e si ha una sensazione d'insieme. Il funzionamento biologico dei coni ha costituito la base per la creazione di uno dei principali modelli cromatici, il sistema RGB, che descriveremo nel seguito.

Il diagramma cromatico CIE

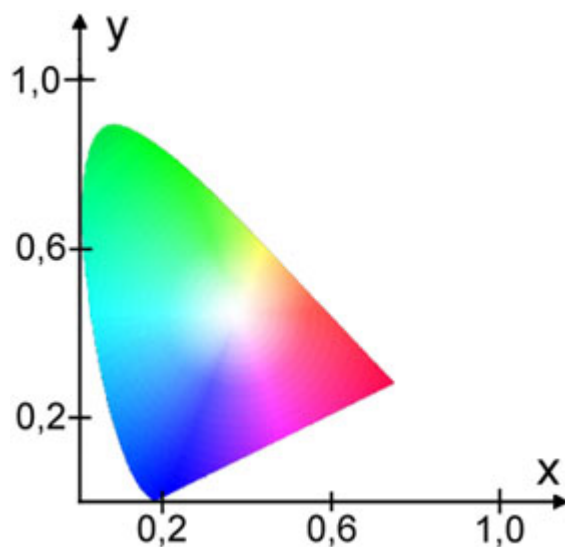
Molti scienziati si sono dedicati allo studio dei colori e della loro rappresentazione, in particolare ricordiamo I. Newton, T. Young e H. Von Helmholtz. Nel 1931 la CIE (Commission Internationale

de L'Eclairage) definì il modello riassuntivo delle precedenti ricerche, il modello CIE, che agisce attraverso tre componenti standard xyY :

- il valore x indica l'importanza relativa della componente rossa rispetto alle componenti verde e blu (valore compreso tra 0 e 1);
- il valore y indica l'importanza relativa della componente verde nei confronti delle componenti rosse e blu (valore compreso tra 0 e 1);
- il valore Y indica, infine, la luminosità (valore compreso tra 0 e 100).

Per avere un'idea di come queste tre grandezze interagiscano si può considerare lo spazio tridimensionale xyY e proiettare ortogonalmente sul piano xy l'immagine che si ottiene intersecando l'area dei colori visibili con il piano che interseca gli assi x , y , Y al valore 1 (come si vede la regione forma una specie di "ferro di cavallo"):

Fig. 2: Il diagramma cromatico CIE.



Il modello L^*a^*b

Il modello L^*a^*b costituisce un raffinamento del modello CIE. Le componenti sono ancora tre ma sono, rispettivamente:

- la luminanza L (con valore compreso tra 0 e 100), che indica la chiarezza del colore;
- la componente a^* che va dal verde al rosso (dal valore $-a$ fino al valore $+a$);
- la componente b^* che va dal blu al giallo (dal valore $-b$ al valore $+b$).

I parametri a , b possono assumere valori da zero all'infinito. Per comprendere meglio il funzionamento di questo modello cromatico si può considerare lo schema sotto riportato in cui si evincono le tre dimensioni citate.

Fig. 3: Il diagramma cromatico L*a*b.



Il passaggio tra le rappresentazioni CIE e L*a*b è possibile attraverso una trasformazione matematica non lineare, per cui è possibile, date le coordinate di colore in uno qualunque dei due sistemi, risalire alle coordinate di colore nell'altro sistema.

Il modello HSB

Il modello cromatico HSB rispecchia molto le modalità di percezione umana del colore. Le caratteristiche sono tre:

- la tonalità *H* (Hue), che indica il colore su una ruota di colori standard (valore compreso tra 0° e 360°);
- la saturazione *S* (Saturation), indica la purezza del colore, ovvero la quantità di grigio rispetto alla tonalità (valore compreso tra 0 e 100, aumenta dal centro verso l'esterno della ruota dei colori);
- la luminosità *B* (Brightness), indica la chiarezza del colore (valore compreso tra 0 e 100).

I modelli YUV e YCbCr

Il modello YUV (utilizzato in particolare nei video analogici) utilizza il seguente sistema di riferimento:

- la *luminanza* *Y*, che coincide con la componente primaria Y del modello CIE, e indica la luminosità;
- la *crominanza*, definita come differenza tra un colore ed un bianco di riferimento alla stessa luminanza. Viene rappresentata dalla coppia di valori U e V che indicano le differenze di colore o tinta.

Riportiamo, a titolo di curiosità, le relazioni che consentono la conversione con il sistema RGB:

$$Y = 0.299 * R + 0.587 * G + 0.114 * B, U = B - Y, V = R - Y.$$

Il modello YCbCr, utilizzato nelle immagini e nei video digitali (JPEG e MPEG) è strettamente legato al modello YUV, infatti è ottenuto da questo attraverso una semplice trasformazione che trasla e scala i valori

Il modello YUV, così come lo HSB, il CIE e il L^*a^*b , sono utilizzati prevalentemente per immagini analogiche (ad esempio per la TV). Gli ultimi modelli di cui parliamo, RGB e CMYK, sono, invece, assieme allo YCbCr, modelli di colore specifici per le immagini digitali.

Il modello RGB e la sintesi additiva

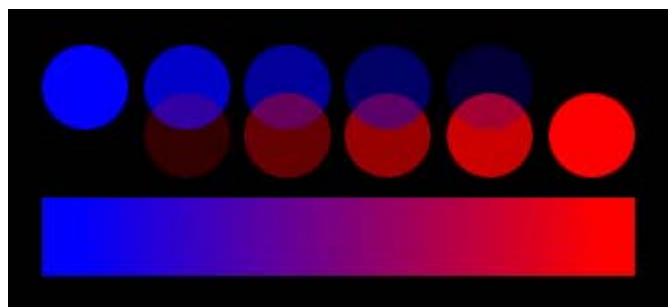
Il modello RGB si basa sull'utilizzo di tre componenti cromatiche differenti, il rosso (R), il verde (G) e il blu (B), ovvero i tre colori primari, con valori compresi tra 0 e 1. La composizione delle tre componenti in diverse proporzioni e intensità permettono di realizzare una vasta percentuale dello spettro visibile. Dallo schema seguente, che riporta i risultati delle composizioni per le quantità massime, si può evincere come la "somma" delle componenti dà luogo a colorazioni differenti: così rosso e verde producono il giallo, verde e blu restituiscono il ciano, rosso e blu restituiscono il magenta, tutti e tre insieme (con quantità massime) restituiscono il bianco, mentre l'assenza di tutte e tre le componenti cromatiche restituisce il nero.

Fig. 4: Il modello cromatico RGB.



La modalità con cui si creano tutti i colori è di tipo *additivo*, in quanto ogni colore si ottiene sommando quantità fissate dei tre colori primari. Nella figura sotto riportata, a titolo d'esempio, sono evidenziati alcuni stadi nella scala cromatica che va dal blu puro al rosso puro, diversi a seconda delle quantità utilizzate dei due colori primari.

Fig. 5: Esempio di scala cromatica.



Questo modello è utilizzato da monitor, scanner, fotocamere.

Il modello CMYK e la sintesi sottrattiva

Il modello CMYK si basa sull'utilizzo di quattro componenti cromatiche: il ciano (C), il magenta (M), il giallo (Y) ed il nero (K) con valori compresi tra 0 e 1. La logica con cui sono usati è una logica sottrattiva (come si può vedere in figura): il bianco è dato dall'assenza delle quattro componenti, il rosso si ottiene sottraendo al bianco il magenta e il giallo, e così via. La presenza delle quattro componenti, quindi, produce un'ampia percentuale della gamma cromatica visibile, fino alla mancanza di luce (nero) se sono tutte presenti con intensità massime (1). In realtà la

combinazione di ciano, magenta e giallo produce un marrone scuro ed è per questo che si aggiunge la componente del nero. La combinazione di questi quattro colori è chiamata *quadricromia*.

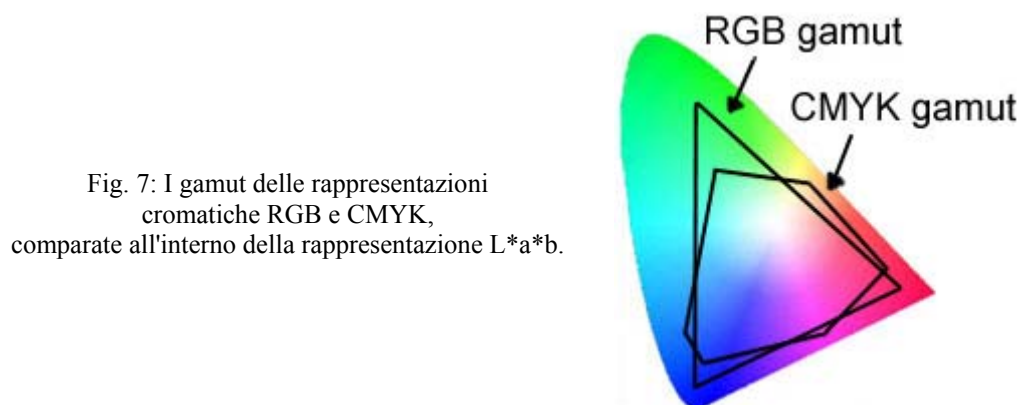
Fig. 6: Il modello cromatico CMYK.



Questo sistema, utilizzato per la stampa, si basa sulla proprietà dell'inchiostro di assorbire una parte della luminosità della carta per cui, aggiungendo inchiostro al foglio, alla sua luminosità vengono sottratte delle componenti. Per questo motivo si parla di modello *sottrattivo*.

Osserviamo, inoltre, che i colori CMY sono "complementari" ai colori RGB: il rosso è complementare al ciano (nel senso che la loro somma dà il bianco), il verde è complementare al magenta, il blu è complementare al giallo. In generale, un colore si dice complementare ad un altro se insieme danno il bianco.

Osservazione. Mentre il modello L^*a^*b riesce a coprire tutti i colori dello spettro visibile, il modello RGB non copre tutto lo spettro e pertanto alcuni colori non possono essere "visti" se si utilizza questa rappresentazione. Il modello CMYK riproduce una quantità di colori ancora inferiore, ma questi non costituiscono un sottoinsieme dei colori rappresentabili con RGB, per cui può essere preferito in certe condizioni. Nella figura sotto riportata si può evincere come i **gamut** (colori effettivamente rappresentabili) dei due modelli siano diversi con un'ulteriore importante proprietà: nelle sintesi additive il gamut coincide con il triangolo che ha tali colori come vertici, nelle sintesi sottrattive il gamut coincide con il triangolo analogo ma con i lati un po' curvi.



2. Grafica Bitmap e Vettoriale

La creazione e la rappresentazione digitale delle immagini può avvenire attraverso due tipi differenti di grafica:

1. **grafica bitmap**, o a punti, che utilizza una griglia rettangolare (raster) di punti, detti pixel (picture element). L'immagine è così una specie di mosaico con piccole tessere di stesse

dimensioni, e pertanto sono memorizzate conservando il valore dei singoli pixel. Questo tipo di grafica è utilizzato soprattutto nelle fotografie e nelle illustrazioni pittoriche.

2. **grafica vettoriale**, o ad oggetti, che utilizza disegni definiti da entità matematiche. L'immagine consiste così di oggetti grafici (punti, linee, rettangoli, curve, ecc.) realizzati attraverso equazioni matematiche, e pertanto sono memorizzate unicamente attraverso le equazioni che li definiscono. Le equazioni, d'altro canto, sono descritte con algoritmi attraverso l'uso di linguaggi specifici, come ad esempio il PostScript. Questo tipo di grafica è un'evoluzione della grafica a punti.

Confronto tra le due rappresentazioni

Il vantaggio principale della grafica a punti è che lavorando sui singoli pixel si possono ottenere effetti simili alla grafica tradizionale (pennello, matita, aerografo, ecc.), si possono cioè rendere con maggiore qualità un grosso numero di colori e texture complesse, come appunto le fotografie. Nella grafica vettoriale, invece, la qualità raggiungibile è quella dei disegni, anche se abbastanza complessi.

Il vantaggio principale della grafica vettoriale consiste nella capacità di conservare inalterata la qualità dell'immagine, anche a seguito di ingrandimenti, rotazioni o ridimensionamenti degli oggetti che la compongono. Nella grafica bitmap, invece, cambiando le dimensioni dell'immagine, la qualità diminuisce in quanto i singoli pixel devono essere riadattati; in particolare nell'ingrandimento dell'immagine, e quindi dei pixel, la composizione a griglia può diventare visibile fino a creare effetti sgradevoli (pixelizzazione). Stesso discorso avviene nelle rotazioni, a meno che non sia di 90° o multipli, che obbliga a riadattare i pixel in maniera approssimata.

Alcune differenze notevoli sono anche le seguenti:

- qualità nella rappresentazione di disegni: nella grafica vettoriale è maggiore, soprattutto per rappresentare disegni semplici, con sfumature di colore non molto elaborate, per l'alta precisione nella definizione;
- qualità della stampa: nella grafica bitmap essa dipende dalla qualità dell'immagine, nella grafica vettoriale coincide con la migliore qualità della stampante;
- visualizzazione su video: la grafica bitmap si presta meglio in quanto lo stesso monitor è formato da una griglia, le immagini vettoriali devono essere trasformate in mappe di punti (rasterizzazione);
- conversione da una grafica all'altra: è facile rasterizzare le immagini vettoriali, è difficile il procedimento inverso.
- nella grafica vettoriale più oggetti si possono trattare come se fossero indipendenti e collocati su un foglio trasparente;
- nella grafica bitmap le immagini possono essere "alleggerite" se ridotte nella definizione, anche se corrispondentemente diminuisce la qualità, nelle immagini vettoriali, invece, gli oggetti sono molto compatti, ma richiedono dei calcoli, a volte complessi, per essere ricomposti.

Un problema rilevante nella grafica vettoriale consiste nell'assenza, ad oggi, di uno standard industriale per la trasmissione di immagini su Internet. Benché alcuni formati vettoriali sono oggi di largo uso nella grafica web, per la loro visualizzazione è necessario un plug-in costruito dalla casa proprietaria del formato. Tuttavia, quei formati oggi in espansione consentono una serie notevoli

vantaggi: ad esempio ciascun oggetto può contenere informazioni aggiuntive o proprietà particolari, come l'esecuzione di un comando in seguito ad un evento (click del mouse o altro).

Nelle figure seguenti si può notare la differenza tra un'immagine vettoriale ed un'immagine bitmap, sia come tipologia che come qualità di rappresentazione. In particolare si può notare come, ingrandendo un dettaglio delle immagini, la qualità vettoriale rimane identica, mentre quella bitmap decade gravemente evidenziando la pixelizzazione o scalettatura.

Fig. 8: Immagine vettoriale semplice.

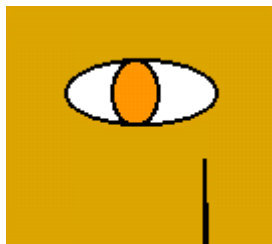
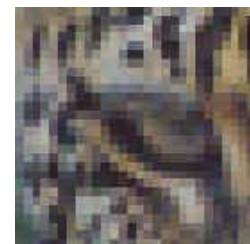


Fig. 9: Immagine vettoriale complessa.



Fig. 10: Immagine bitmap.



3. Le caratteristiche qualitative nella grafica bitmap

Le caratteristiche qualitative principali delle immagini bitmap sono:

1. la *risoluzione grafica*, o semplicemente risoluzione, che indica la quantità di informazioni distinte contenute nella stessa superficie, ovvero il numero di punti che formano la mappa dell'immagine;
2. la *risoluzione cromatica*, o profondità dei colori, che indica il numero di colori che ciascun punto può assumere.

La risoluzione grafica

La risoluzione grafica di un'immagine, misurata in **dpi** (dots per inch, ovvero punti per pollice, ricordiamo che 1 inch = 2,54 cm), è definita come il numero di punti utilizzati per rappresentare un pollice lineare dell'immagine.

Più sono i punti e maggiore è la qualità dell'immagine in quanto, a parità di dimensioni fisiche dell'immagine - larghezza e lunghezza, un numero maggiore di punti implica che gli stessi abbiano dimensioni più piccole e l'effetto di *pixelizzazione* (quando i pixel sono chiaramente visibili come in

un mosaico) è meno evidente ai nostri occhi. Contemporaneamente, però, un maggior numero di punti implica un peso maggiore dell'immagine, in quanto le informazioni da memorizzare sono più numerose. Nelle immagini seguenti si vede come all'abbassamento della risoluzione grafica corrisponde una minore dolcezza nel passaggio da un pixel all'altro.

Fig. 11: immagine a 100 dpi.



Fig. 12: immagine a 50 dpi.

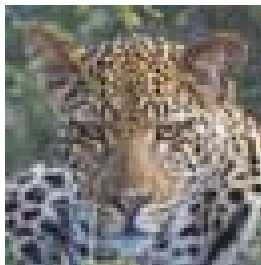


Fig. 13: immagine a 20 dpi.



Risoluzioni tipiche sono: 1800 dpi per avere la qualità tipografica, 300 dpi per la qualità della stampa laser, 60 dpi per la qualità delle immagini sui quotidiani, 72/96 dpi per avere la qualità permessa dai monitor.

Esercizio 1. Un'immagine di dimensioni fisiche 6 cm × 4,5 cm ha una risoluzione grafica di 200 dpi. Di quanti punti è composta l'immagine?

Soluzione. Innanzitutto le [conversioni](#): 6 cm = 6 / 2,54 = 2,362 inch; 4 cm = 4 / 2,54 = 1,574 inch. Dunque il numero di punti è: $(2,362 \times 200) \times (1,574 \times 200) = 148.711$.

Esercizio 2. Un'immagine di dimensioni 1,6 inch × 2,5 inch è composta da una mappa di 250.000 punti. Qual è la sua risoluzione grafica?

Soluzione. La risoluzione grafica si ottiene facendo la radice quadrata del numero $250.000 / (1,6 \times 2,5) = 62.500$ che corrisponde a 250 dpi.

La risoluzione cromatica

La risoluzione cromatica di un'immagine è definita come il numero di livelli dei colori a disposizione che si possono utilizzare per rappresentarla. Tale valore è strettamente connesso al numero di cifre binarie utilizzate: quanto più lunghe saranno le sequenze binarie (n) utilizzate per codificare i colori, tanto maggiore sarà il numero dei livelli cromatici (2^n). Per questo motivo si parla anche di *profondità di colore*, o *bit depth*.

Ad una risoluzione cromatica maggiore, chiaramente, corrisponde una qualità superiore dell'immagine, in quanto il numero di sfumature dei colori (o del colore) che si possono visualizzare è più alto. Anche in questo caso, però, ad una maggiore profondità di colore corrisponde un peso maggiore dell'immagine, in quanto ciascuna informazione da memorizzare è più lunga.

Fig. 14: immagine a 256 colori (8 bit).



Fig. 15: immagine a 16 colori (4 bit).



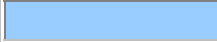
Esistono fondamentalmente due modi per rappresentare i colori utilizzati dalla risoluzione cromatica:

- A. in maniera **indicizzata**, attraverso una *paletta* con un certo numero (limitato) di colori predefiniti; solitamente si utilizzano un massimo di 256 colori differenti (utilizzando 8 bit per ogni colore), anche se questi possono essere scelti fra diversi milioni di tinte, oppure 16 colori differenti (utilizzando 4 bit per ogni colore), o anche 2 soli colori (utilizzando un solo bit) per immagini monocromatiche. In tal caso si memorizza l'indice dei colori nella paletta scelta (ovvero la posizione ordinale dei colori nella paletta) e, nel caso l'immagine contenga punti di colore non presenti nella paletta, si memorizzano i colori che più si avvicinano. La paletta, inoltre, può essere memorizzata assieme ai dati grafici oppure, nel caso di palette definite dal sistema operativo, sono files a se stanti a cui si fa riferimento.
- Un effetto molto utile, infine, è quello del *dithering*, una tecnica di abbinamento di colori diversi (eseguita attraverso algoritmi specifici) per imitare colori che non sono disponibili nella paletta. Ad esempio si può riprodurre una superficie viola alternando punti blu e rossi o una superficie grigia alternando punti bianchi e neri, e così via;
- B. in maniera **TrueColor**, specificando il valore di ogni singola componente cromatica, ad esempio utilizzando i modelli RGB, CMYK, HSB, L^*a^*b , in cui viene assegnato uno stesso numero di bit ad ognuna delle tre (o quattro) componenti. In questo caso si va da migliaia di colori fino a milioni di colori (RGB24, con 1 byte per componente cromatica) o persino molti miliardi di colori (RGB48, con 2 byte per colore).
- Nelle rappresentazioni TrueColor bisogna comunque tener presente che l'occhio umano non è in grado di apprezzare differenze se la risoluzione cromatica è superiore ad 1 byte per componente e dunque una risoluzione ad 8 bit per componente si rivela essenzialmente sufficiente per qualsiasi esigenza.

Osservazione. Nel caso, peraltro molto frequente, di risoluzione cromatica RGB24, i colori sono identificabili logicamente con una terna di numeri, ciascuno dei quali può variare:

- da 0 a 255 (ovvero $2^8 - 1$), se espressi in notazione decimale;
- da 00 a FF ($FF_{16} = 255_{10}$), se espressi in notazione esadecimale.

Nella tabella seguente sono riportati alcuni esempi.

Colore	in decimale	in esadecimale	in binario
	(255,0,0)	(FF,00,00)	111111110000000000000000
	(0,255,0)	(00,FF,00)	000000001111111100000000
	(0,0,255)	(00,00,FF)	000000000000000011111111
	(153,204,255)	(99,CC,FF)	100110011100110011111111
	(255,255,0)	(FF,FF,00)	111111111111111110000000
	(255,153,51)	(FF,99,33)	111111111001100100110011
	(255,153,255)	(FF,99,FF)	111111111001100111111111
	(153,51,102)	(99,33,66)	100110010011001101100110
	(102,102,102)	(66,66,66)	011001100110011001100110
	(0,0,0)	(00,00,00)	000000000000000000000000
	(255,255,255)	(FF,FF,FF)	111111111111111111111111

La scala di grigi

Nelle immagini cosiddette in "bianco e nero", che in questo ambito sono ribattezzate col termine "a scala di grigi", la rappresentazione avviene utilizzando una sequenza di grigi equamente distanziati tra i valori estremi bianco e nero; in pratica non si tiene conto dei colori reali dell'immagine ma solo della loro luminosità. Tale sequenza caratterizza bene una fotografia b/n se dispone di una risoluzione di almeno 40-50 livelli di grigio. La scala normalmente utilizzata prevede 64 valori, ma sono possibili anche scale con meno valori. In ogni caso, quanto più sono numerosi i passaggi che dal nero portano al bianco, tanto più l'immagine avrà colori meno contrastanti.



Fig. 16: 16 livelli di grigio.



Fig. 17: 8 livelli di grigio.

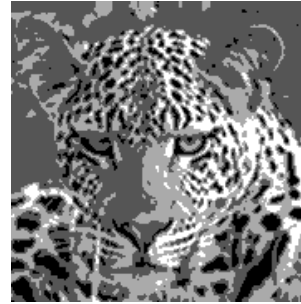


Fig. 18: 4 livelli di grigio.



Fig. 19: 2 livelli (bianco e nero).

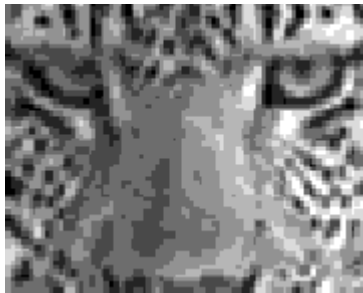


Fig. 20: questa immagine è l'ingrandimento di un'area della fig. 17. Si nota facilmente il passaggio tra i vari livelli di grigio, che in questa immagine sono solo 8. Più sono questi livelli, più il passaggio è dolce e naturale.

Oltre al numero di grigi utilizzato, sono importanti altri due aspetti, ovvero il contrasto e il bilanciamento dei toni. Aumentare (o diminuire) il contrasto significa aumentare (o diminuire) la distanza tra il tono più nero dal nero vero e proprio. Per bilanciamento dei toni, invece, si intende la distribuzione dei gradi di grigio in modo corretto tra i valori estremi bianco e nero; anche in presenza di un numero di grigi adeguato, infatti, non è detto che la situazione dei toni di un'immagine sia corretta. Pertanto, mentre il contrasto "stira" la scala dei grigi, allungandola o comprimendola, il bilanciamento adegua tali valori tra i valori estremi.

Se si passa a considerare il colore, il concetto non cambia: oltre a misurare quanto è scura ogni area dell'immagine, si attribuisce ad essa anche una certa tonalità di colore. Uno standard ragionevole è quello che comprende 256 colori. Dove possibile, oggi si preferisce usare uno standard migliore, basato su un numero di colori uguale o superiore a quello che è in grado di cogliere l'occhio umano.

Vediamo, nel seguito, un esempio di digitalizzazione nella scala dei grigi.

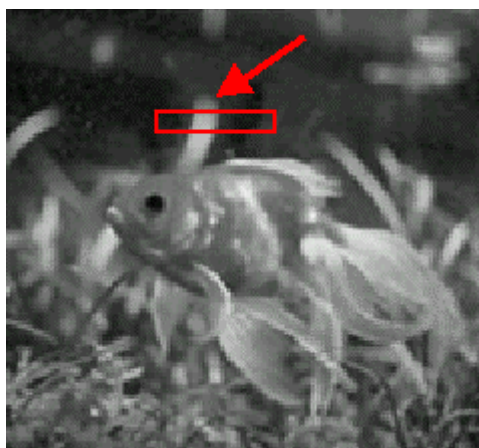


Fig. 21: la porzione di immagine indicata dalla freccia è la porzione estratta in Fig. 2.

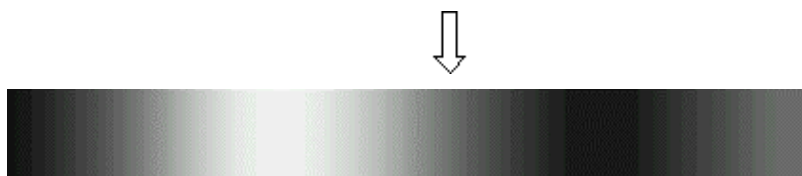


Fig. 22: porzione di immagine.

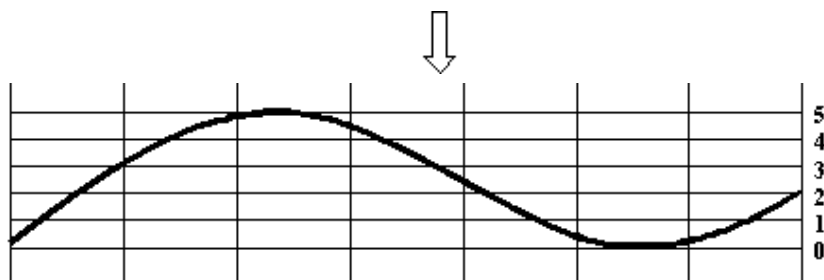


Fig. 23: curva che descrive l'andamento della luminosità. Quando nella figura si va verso il bianco la curva sale, quando si va verso il nero la curva scende (la luminosità diminuisce). La curva è una rappresentazione analogica del segnale in quanto è analoga all'andamento della luminosità. Il segnale analogico viene discretizzato fissando una scala di grigi (in questo caso da 0 a 5) e misurandolo ad intervalli spaziali regolari, ottenendo così una serie di valori numerici che rappresentano i livelli (discreti) di luminosità.

0 3 5 4 2 0 0 2

Fig. 24: La sequenza di valori numerici rappresenta il segnale digitale della porzione di immagine selezionata. Questi numeri, opportunamente codificati (in questo caso con 3 bit ciascuno), definiscono la rappresentazione digitale binaria della porzione in esame.

000 011 101 100 010 000 000 010

I canali RGB e CMYK

La risoluzione cromatica TrueColor in [RGB](#) si avvale delle tre componenti cromatiche primarie (anche detti canali), per le quali viene fissato un numero di livelli (uguale per tutti), ciascuno dei quali definisce la quantità di colore presente. Le usuali rappresentazioni utilizzano 256 valori (da 0 a 255), ovvero 8 bit, per colore primario, quindi 24 bit complessivamente, che danno luogo a $256 \times 256 \times 256 = 2^{24} = 16.777.216$ colori complessivi.

La rappresentazione cromatica RGB può avvenire anche assegnando, invece di valori numerici da 0 a 255, valori esadecimali da 00 a FF, attraverso normale conversione. Ad esempio, se le componenti RGB sono, rispettivamente, (42, 247, 132) la corrispondente terna esadecimale è (2A, 84, F7) da cui la stringa 2A84F7.

Sono però possibili anche altre rappresentazioni RGB, ad esempio RGB48 con 48 bit totali per pixel, ovvero con 16 bit per colore primario.

La risoluzione [CMYK](#) si avvale invece di quattro componenti cromatiche, per ciascuna delle quali sono fissati 256 valori, ovvero 8 bit, per un totale complessivo di 32 bit.

Poiché con profondità di colore si identifica il numero di bit con cui viene descritto il valore di luminosità del singolo primario, nel caso di RGB24 e CMYK la profondità di colore è di 8 bit, mentre nel caso di RGB48 è di 16 bit.

Esercizio 3. Un'immagine di dimensioni fisiche di 5 cm × 3,8 cm ha una risoluzione grafica di 150 dpi ed una risoluzione cromatica indicizzata con paletta da 256 colori. Quanto spazio occupa complessivamente?

Soluzione. Innanzitutto le [conversioni](#): 5 cm = 5 / 2,54 = 1,968 inch; 3,8 cm = 3,8 / 2,54 = 1,496 inch. Dunque l'immagine è costituita da $(1,968 \times 150) \times (1,496 \times 150) = 66.243$ punti. La paletta è ad 8 bit per cui ciascun colore è rappresentabile con un byte, dunque l'immagine occupa 66.243 byte = 64,6 Kb.

Esercizio 4. Un'immagine di dimensioni fisiche di 5,4 inch × 7,7 inch ha una risoluzione grafica di 75 dpi ed una risoluzione cromatica RGB24. Quanta memoria occupa complessivamente? E se avesse una risoluzione CMYK, quanta memoria occuperebbe?

Soluzione. L'immagine è costituita da $(5,4 \times 75) \times (7,7 \times 75) = 667.012$ punti. La risoluzione cromatica impiega 24 bit (3 byte) per punto, quindi l'immagine occupa complessivamente $667.012 \times 3 = 2.001.036$ byte = 1,9 Mb. Se si utilizzasse la risoluzione cromatica CMYK si impiegherebbero, invece, 32 bit (4 byte) per punto, per cui l'immagine occuperebbe $667.012 \times 4 = 2.668.048$ byte = 2,5 Mb.

Osservazione. Ricapitolando, la dimensione complessiva di un file bitmap (non compresso) si ottiene attraverso la formula:

$$\text{dimensione file} = \text{area (in inch}^2\text{)} * \text{risoluzione}^2 \text{ (in dpi)} * \text{profondità di bit}$$

dove, chiaramente, tutte le grandezze vanno espresse nelle stesse unità di misura (così, ad esempio, si potrebbe riportare tutto da inch in cm).

4. I formati

Le esigenze di memorizzazione e di utilizzazione delle immagini, così come per altri tipi d'informazione elettronica, possono essere diverse a seconda delle caratteristiche da esse possedute. A tal fine sono state sviluppate numerose metodologie di ordinamento dei dati, che vanno sotto il nome di "formati".

Strumenti molto utili nella memorizzazione delle immagini, in special modo per la loro utilizzazione sul web, sono le tecniche di compressione, ovvero degli algoritmi di trattamento dei dati che manipolano le informazioni e le riducono in modo da occupare un numero minore di byte. Questi algoritmi di compressione sono suddivisi in base alla perdita dei dati: si dicono "con perdita" (o lossy) se consentono la perdita dei dati, ovvero la compressione elimina definitivamente dei dati che non sono più recuperabili, e "senza perdita" (o lossless) se compattano i dati senza perderli, e quindi sono sempre recuperabili. Il processo di decompressione viene eseguito automaticamente, infine, all'apertura dell'immagine.

Vediamo quali sono i file grafici di utilizzo più comune.

Il formato Bitmap (suffisso ".bmp") è il formato standard per le immagini in ambienti compatibili DOS e Windows. Supporta profondità di colore da 1 fino a 24 bit, sia in maniera indicizzata che RGB, ma non CMYK. Nelle immagini a 4 o 8 bit è possibile usare anche una tecnica di compressione senza perdita (RLE).

Il formato WMF (Windows Metafile Format) è utilizzato per la rappresentazione di disegni vettoriali in ambiente Windows ma anche in combinazione con la grafica bitmap (questa è la caratteristica dei *metafiles*) e con il testo. Un file WMF è costituito da una lista di istruzioni a 16 bit che definiscono e controllano lo stile di questi oggetti. Esempi tipici di immagini in questo formato sono le clipart.

Il formato EMF (Enhanced Metafile Format) è un metafile avanzato che può contenere una varietà di comandi molto più ampia del metafile standard di Windows (funziona a 32 bit). utilizzato per la rappresentazione di disegni vettoriali in ambiente Windows ma anche in combinazione con la grafica bitmap (questa è la caratteristica dei "metafiles"). Un file WMF è costituito da una lista di istruzioni che definiscono e controllano lo stile di questi oggetti. Esempi tipici di immagini in questo formato sono le clipart.

Il formato SVG (Scalable Vector Graphics) è il vero metafile per il Web (raccomandato dal W3C). SVG è un linguaggio che permette di gestire la grafica (figure vettoriali, immagini e testo) in XML con molteplici strumenti come filtri, canali Alpha e interattività. Ha ancora supporto limitato sui browser, ma il suo utilizzo è in costante ascesa.

Il formato CGM (Computer Graphics Metafile) è utilizzato per rappresentazioni vettoriali, in particolare per lo scambio di immagini complesse in architettura e ingegneria. E' lo standard internazionale per la memorizzazione e lo scambio (è piattaforma-indipendente) di immagini contenenti grafica bitmap e vettoriale.

Il formato PDF (Portable Document Format) di natura proprietaria (Adobe) utilizzato per la pubblicazione elettronica di documenti in quasi tutti i sistemi operativi. Può contenere sia grafica bitmap che vettoriale, con gestione Lab, RGB, CMYK, scala di colore, scala di grigio o bitmap, supportando anche funzioni di ricerca e navigazione nel documento con collegamenti ipertestuali.

Il formato PostScript (suffisso ".ps") è utilizzato per la computer grafica vettoriale. In realtà PostScript è più precisamente un linguaggio di programmazione (il "linguaggio delle stampanti"), per cui un file .ps è un file di testo contenente un programma che viene interpretato, rasterizzato per la visualizzazione e retinato per la stampa da un processore specifico (noto col nome RIP).

Il formato Encapsulated PostScript (suffisso ".eps") è un formato per il trasferimento di disegni in linguaggio PostScript (formattato in modo particolare e con alcuni vincoli) anche su piattaforme differenti (indipendentemente dalla disponibilità di un interprete PostScript). Può contenere da singole illustrazioni fino a pagine complete con immagini sia vettoriali che bitmap. E' utilizzabile da tutti programmi di grafica, sia raster che vettoriali. Non supporta, invece, nessun canale per la trasparenza (Alpha). Per facilitare la visualizzazione durante l'impaginazione delle immagini, all'interno del file può essere salvata una rappresentazione dell'immagine in formato bitmap a bassa risoluzione.

Il formato TIFF (Tagged-Image File Format) è utilizzato per scambiare immagini bitmap tra applicativi di piattaforme diverse. E' un formato molto flessibile che può adottare sia scale di grigio

che RGB o CMYK e dunque con milioni di colori. E' il formato principale per la stampa di qualità ed ha il supporto del canale Alpha (che permette la trasparenza) e la compressione senza perdita LZW.

Il formato JPEG (Joint Photographic Experts Group, suffisso ".jpg") è uno dei formati più utilizzati, in special modo per la rappresentazione di immagini sul web (ma anche per l'archiviazione nelle fotocamere), specialmente di tipo fotografico. Supporta colori a 24 bit ed utilizza una tecnica di compressione lossy (chiamato DCS) che nasce dalla scoperta che l'occhio umano ha recettori di luminosità più precisi rispetto ai recettori dei colori. Con questa tecnica, in pratica, si separano i due canali dell'immagine (luminosità e colore) e si applicano algoritmi differenti, più precisi per la luminosità e più drastici sul colore (tipicamente vengono considerati aree di 64 pixel su cui il colore viene uniformato). La tecnica di compressione, comunque, può essere personalizzata, potendo applicarla in modo lieve, medio, fino ad essere molto forte (ovvero con rapporti di compressione differenti, fino a raggiungere l'1%). In caso di compressione drastica, l'effetto che ne consegue è la tipica quadrettatura con parti a tinta unita che si può notare nella figura 2 sotto riportata.

Supporta anche la caratteristica della *progressività* nella visualizzazione a seguito dello scaricamento. Non supporta, invece, il canale della trasparenza e tratta solo immagini statiche (a differenza del formato GIF) ma esiste uno standard correlato per filmati o immagini in movimento, noto come MPEG.

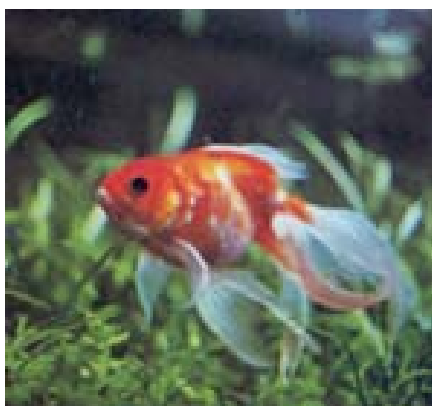


Fig. 25: immagine JPEG poco compressa, il peso finale è di 23,7 Kb.



Fig. 26: immagine JPEG molto compressa, il peso finale è di 7,86 Kb.

N.B.: l'immagine originaria è una bitmap a 75 dpi e peso 59,8 Kb.

Il formato GIF (Graphics Interchange Format) è uno dei formati più diffusi sul web, nato per l'interscambio di immagini tra una piattaforma e l'altra. E' utilizzato particolarmente per disegni o immagini senza troppe sfumature in quanto supporta fino ad un massimo di 256 colori. Questo limite si può comunque aggirare personalizzando le palette, in altre parole i 256 colori possono essere scelti.

Utilizza la tecnica di compressione LZW che è di tipo lossless e funziona nel seguente modo: invece di memorizzare, ad esempio, la stringa

"rosso-rosso-rosso-rosso-rosso-rosso-rosso-rosso-blu-blu-blu-blu-verde-verde" memorizza, con opportuna codifica, la stringa "8 rossi-4 blu-2 verdi". Tale codifica avviene sui pixel in senso orizzontale. Come si può evincere dalle figure seguenti un'immagine a tinte piatte viene maggiormente compressa se le linee sono orizzontali invece che verticali, ed effetti di "materia" diminuiscono di molto il rapporto di compressione.

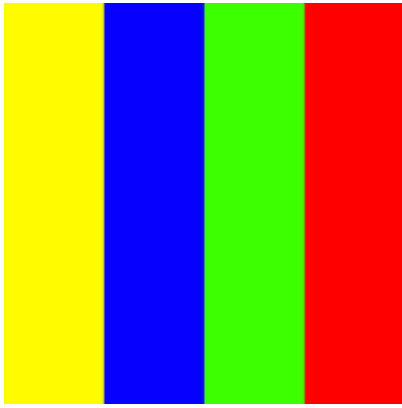


Fig. 27: immagine GIF,
peso 1,34 Kb.

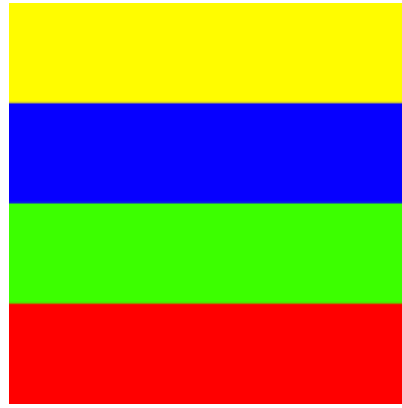


Fig. 28: immagine GIF,
peso 755 bytes.

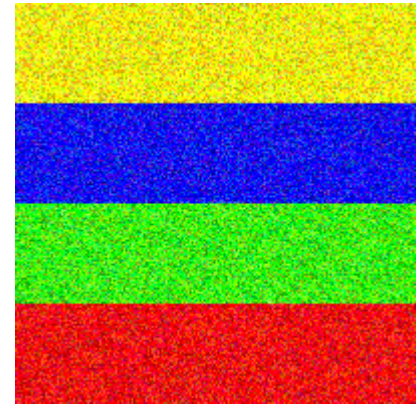


Fig. 29: immagine GIF,
peso 38,2 Kb.

L'ultimo formato (GIF89a) supporta canali per la trasparenza singoli, l'interlacciamento (per visualizzare le immagini in maniera graduale e crescente all'atto dello scaricamento), e animazioni (GIF animate, ovvero una sequenza di immagini GIF, o frame, riprodotte con intervalli temporali fissati ed eventualmente in maniera ciclica). Nelle figure che seguono, tutte GIF con sfondo trasparente, le prime cinque, opportunamente assemblate, costituiscono l'animazione della fig. 35:

Fig. 30: prima
immagine.



Fig. 31:
seconda
immagine.



Fig. 32: terza
immagine.



Fig. 33: quarta
immagine.



Fig. 34: quinta
immagine.



Fig. 35: GIF animata composta dalle 5
immagini, ripetute ciclicamente ad
intervalli di 3 centesimi di secondo.



Osservazione. Per un confronto tra i formati GIF e JPEG bisogna considerare che un'immagine a 24 bit viene fortemente penalizzata con la codifica GIF, ma questa è superiore per disegni al tratto, logo, fumetti, o se i colori utilizzati sono pochi (meno di 256). JPG è dunque superiore per fotografie, scene realistiche a colori o a scala di grigi.

Va osservato, inoltre che applicando più volte una compressione JPEG o passando da un'immagine JPEG a GIF e viceversa, le compressioni vengono rieseguite e, nel secondo caso, si trasportano i limiti di una rappresentazione sull'altra, a scapito della qualità complessiva. Per questo motivo questi formati sono destinati solo all'uso finale e non come passi intermedi di un'elaborazione.

Il formato PNG (Portable Network Graphics) è stato ideato per lo scambio e la pubblicazione di immagini sul Web e si distingue dal GIF per la sua licenza libera (è nato proprio da questo intento). La tecnica di compressione utilizzata è lossless, in alcuni casi più efficace di quella LZW, e supporta gestioni di colore RGB e scala di grigio, bitmap e scala di colore. Esiste in due diverse versioni:

- PNG-8, che usa colori a 8 bit presenta molte analogie con il formato GIF, supporta la trasparenza ma non l'animazione,
- e PNG-24 che supporta colori a 24 bit e presenta analogie con il formato JPEG, supporta una trasparenza multilivello (8 bit).

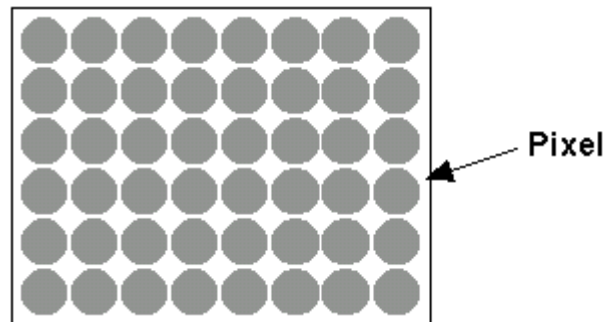
I formati DXF, DWG, DWF sono rappresentazioni di immagini vettoriali in ambiente AutoCad. L'ultimo è il formato creato per il Web ma, in generale, hanno la limitazione di essere difficilmente scambiati, se non in possesso del relativo software.

5. Le periferiche

I principali dispositivi periferici impiegati nella computer grafica sono: il monitor, la stampante, lo scanner e la fotocamera digitale. Non si può prescindere da una conoscenza di base del loro funzionamento e dei loro parametri principali, poiché questi interagiscono con le tecniche grafiche utilizzate.

Tutte queste periferiche sono di tipo raster (griglia rettangolare di pixel), ma esistono anche dispositivi non raster per applicazioni particolari.

Fig. 36: Schema raster dei dispositivi periferici.



Il monitor

I monitor utilizzano la sintesi additiva RGB, per cui lo spazio colore è dato dalla combinazione di fasci di luce dei tre colori primari: rosso, verde e blu. Da notare, comunque, che i [gamut](#) di due monitor sono generalmente diversi ed anche lo stesso monitor può cambiare gamut col passare degli anni.

Esistono tre tipi di monitor:

- CRT (a tubi catodici), da sempre utilizzati come computer desktop, con dimensioni simili ai televisori (date dalla grandezza del tubo catodico) ma più affidabili;
- LCD (a cristalli liquidi), utilizzati sui computer portatili e sui notebook, ma in uso crescente anche nei computer desktop. Sono di piccolo spessore e con totale assenza di radiazioni elettromagnetiche;
- PDP (al plasma), tipicamente di grandi dimensioni in quanto economici e durevoli, ma non hanno alte risoluzioni.

I parametri caratteristici sono:

- le *dimensioni fisiche* dello schermo, espresse in funzione della lunghezza della diagonale (per questo sono espresse in pollici). Le dimensioni della base e dell'altezza sono facilmente ricavabili in quanto utilizzano lo stesso rapporto di 4 a 3 (4:3) utilizzato dalle TV. Tipiche dimensioni sono: 14", 15", 17", ecc.;
- il *dot pitch*, ossia la distanza (per cui è espresso in millimetri) tra fosfori di uguale colore: più questa distanza è piccola e più sono definite le immagini sul monitor. Tipicamente si va dai 0,21 mm (alta qualità) fino ai 0,28 mm (bassa qualità);
- il *refresh rate*, ovvero la frequenza di aggiornamento (misurata in Hertz) dell'immagine sul monitor: più questa frequenza è bassa e più viene avvertita dall'occhio causando affaticamento. Il valore minimo consigliato è usualmente 75 Hz;
- la *risoluzione*, ovvero il numero di punti necessari per comporre l'intera schermata: più questo numero è alto e più piccole saranno le dimensioni dei pixel visualizzati, con

conseguente migliore definizione delle immagini a schermo. Espressa come "numero di pixel della base" $\times$ "numero di pixel dell'altezza", i valori tipici sono: 640 $\times$ 480, 800 $\times$ 600, 1024 $\times$ 768, 1280 $\times$ 1024. Nei monitor meno recenti la risoluzione era fissata e non modificabile, e si avevano per costruzione 72/96 pixel per inch (ppi); i monitor multisync, invece, consentono di modificarla su una qualsiasi delle dimensioni supportate (supportate anche dalla scheda grafica) e per questo non hanno un numero di pixel per inch fissato una volta per tutte. In ogni caso la risoluzione a 72 ppi è "ideale" in quanto le dimensioni dei pixel corrispondono alle dimensioni del punto tipografico (entrambi 1/72 di pollice); in altre parole le dimensioni di visualizzazione corrispondono esattamente alle dimensioni di stampa.

La stampante

Le stampanti da scrivania, o desktop printer, possono essere in bianco e nero o a colori, ed in quest'ultimo caso utilizzano la sintesi sottrattiva CMYK. Esistono diverse tecnologie costruttive, ma le più utilizzate sono con laser a toner e a getto d'inchiostro (inkjet).

Il parametro principale di ogni stampante è la sua *risoluzione* massima di stampa, che può essere fissata o scelta tra alcune risoluzioni supportate.

Molto importante è la *retinatura* (o *screening*): un retino (screen) è formato dai punti che si trovano sull'intersezione di due fasci perpendicolari di rette parallele ed equidistanti. Ad esempio nelle stampanti in bianco e nero i livelli intermedi di grigio si ottengono per sintesi additiva spaziale, ovvero stampando puntini neri uniformemente distanziati (e uguali) sullo sfondo cartaceo. Rimanendo entro un certo limite, maggiore è questa distanza è più tenue sarà il tono di grigio visibile ai nostri occhi. La stessa tecnica è comunque utilizzata anche per il colore, attraverso la stampa di più retini (quattro) sovrapposti ad angolature diverse.

Lo scanner

Gli scanner utilizzano la sintesi additiva RGB. Le tecnologie di scansione utilizzate (attraverso un procedimento di riflessione della luce) sono fondamentalmente due: a tamburo (che utilizza i lettori ottici PMT, o fotomoltiplicatori) e a letto piano (che utilizza i sensori CCD, o dispositivi ad accoppiamento di carica), detti anche scanner desktop. Quest'ultima tecnologia, in particolare è più economica ma meno performante, anche se ancora in fase di miglioramento.

Le caratteristiche principali sono:

- la *risoluzione ottica*, ovvero il numero di sensori per pollice che nel caso di scanner CCD può arrivare a 3800 ppi ;
- la *profondità di colore*, che è generalmente a 24 o 48 bit (molta importanza ha però il software;
- l'*intervallo dinamico*, o intervallo di densità, che indica la capacità di assorbire la luce, data dalla distanza tra le densità estreme (massima e minima).

La fotocamera digitale

Le fotocamere utilizzano la sintesi additiva RGB. Questi dispositivi di sono molto simili agli scanner con la particolarità che si possono focalizzare su distanze variabili (al contrario dello scanner che può focalizzare per pochi millimetri). La cattura delle immagini avviene utilizzando la

tecnologia CCD con due possibili metodi di scansione: lineare (come gli scanner), rettangolare (lettura simultanea dei sensori).

I parametri principali, dal nostro punto di vista sono:

- le *dimensioni* in pixel, ovvero il numero complessivo di pixel che la fotocamera può utilizzare alla dimensione massima. In questi dispositivi non ha molto senso parlare di risoluzione per unità di misura in quanto la dimensione fisica dei pixel è data dal dispositivo d'uscita (così un'immagine catturata a 2048×1536 pixel, se stampata a 300 ppi ha il lato maggiore di $2048/300 = 6,83$ in = 17,34 cm, mentre se stampata a 600 ppi ha il lato maggiore di $2048/600 = 3,41$ in = 8,67 cm);
- il *rapporto di forma*, che normalmente può essere 3:2 oppure 4:3;
- la *sensibilità* nella cattura della luce, che si misura in valori ISO come nel caso delle pellicole (la maggior parte delle fotocamere ha una sensibilità compresa tra 80 e 200 ISO);
- il *frame rate*, ovvero il minimo tempo che deve intercorrere tra due scatti consecutivi (può andare da frazioni di secondo fino ai 20 secondi).

I pollici (inches). Nel sistema metrico anglosassone si hanno le seguenti relazioni:
12 inches (pollici) = 1 foot (piede), 3 feet (piedi) = 1 yard.
Inoltre il rapporto col sistema metrico decimale è dato da: 1 inch = 2,54 cm.



L'Audio

[torna all'indice](#)

1. Le onde audio

Così come le onde luminose sono percepite dai nostri occhi sotto forma di luce, le onde audio (o sonore) sono percepite sotto forma di suoni o rumori dai nostri orecchi. In maniera analoga, queste onde si sviluppano su una gamma particolare di frequenze e confinano con onde non percepibili dal nostro orecchio, chiamate infrasuoni e ultrasuoni.

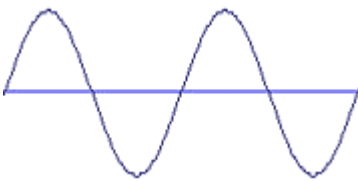
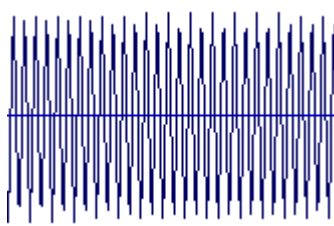
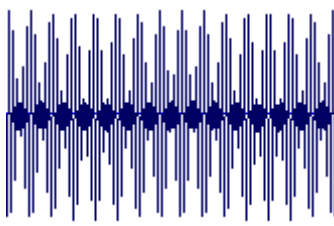
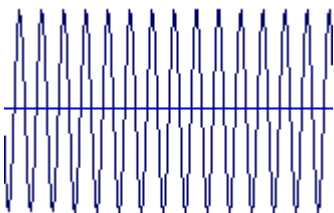
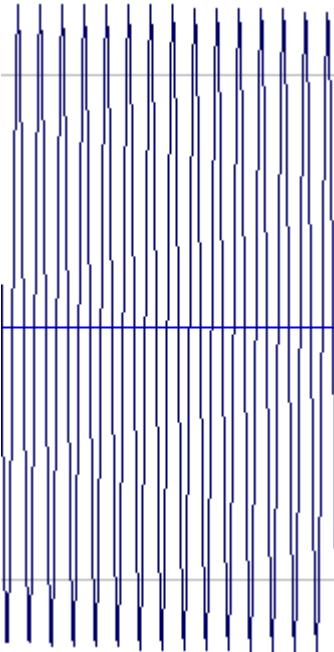
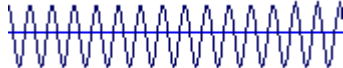
Anche se persone diverse possono avere sensibilità diverse di ascolto (ed anzi, nell'arco degli anni, ci possono essere differenze nella stessa persona), in genere si confina la nostra banda audio (per gli animali la banda è diversa) su frequenze comprese tra i 20 Hz e i 20 KHz (20.000 Hz).

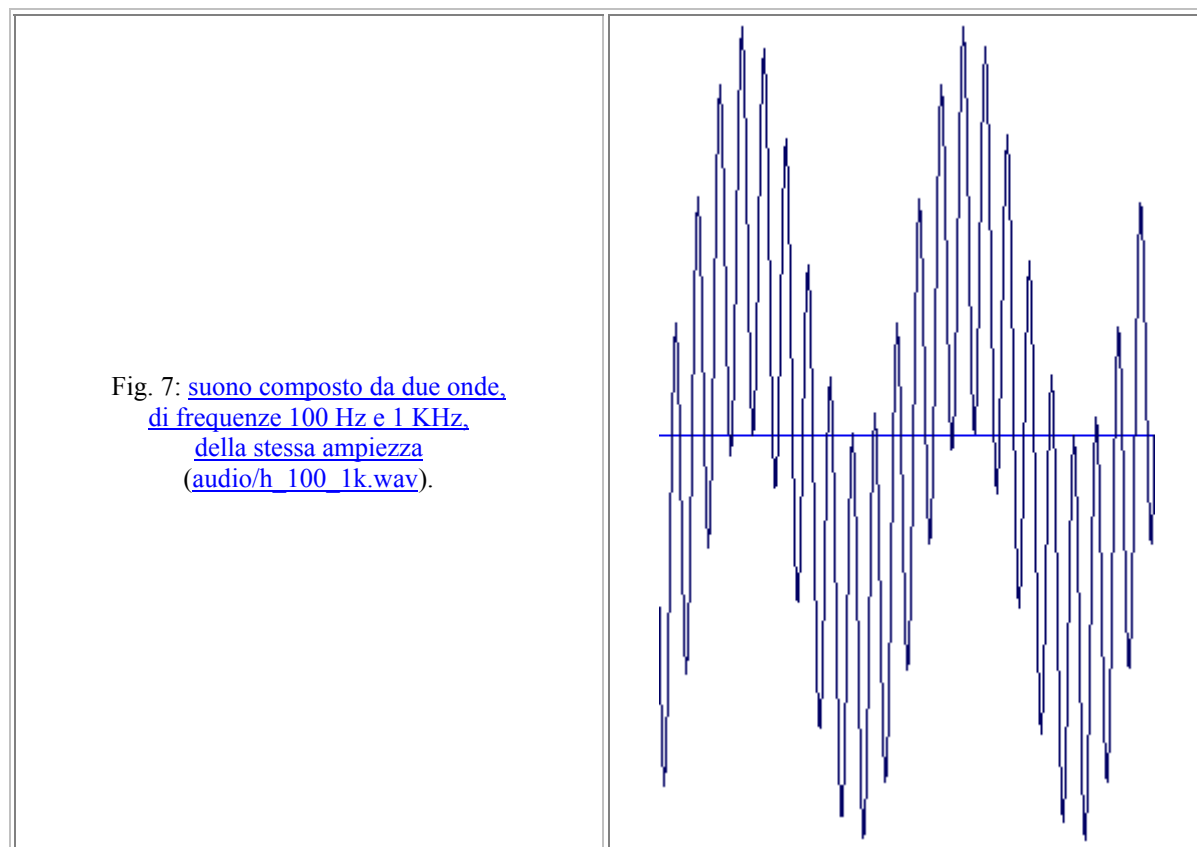
Le grandezze fisiche fondamentali da considerare nell'audio sono:

- la **frequenza** (o *altezza*), ovvero il numero di oscillazioni (o di cicli) compiute nell'unità di tempo. Ad essa è sempre legata la lunghezza d'onda, che gli è inversamente proporzionale (si ricordi che per tutte le onde vale la seguente formula: $frequenza \times lunghezza\ d'onda = velocità\ dell'onda$). I suoni a frequenza bassa (20-500 Hz) sono detti *gravi*, quelli a frequenza elevata (2-20 KHz) sono detti *acuti*;
- l'**ampiezza massima** o semplicemente **ampiezza** (o *intensità*), ovvero la differenza tra il valore massimo e quello minimo raggiunto da un'onda in un ciclo (ovvero la distanza tra l'avvallamento e la cresta). Molto semplicemente un suono con ampiezza nulla non è percepito dal nostro orecchio, mentre uno con ampiezza elevata può anche risultare fastidioso perché ad "alto volume". Nella realtà il nostro orecchio è sensibile in maniera differente alle varie frequenze, e si può rappresentare una curva di sensibilità fornendo l'ampiezza minima udibile in funzione della frequenza dell'onda. Questa curva, utilizzata anche per esprimere (in decibel: misura usata per quantizzare il rapporto tra due quantità fisiche. In acustica, "misura il rapporto tra la pressione sonora di un suono rispetto a quella dei suoni appena percettibili") l'impatto di qualsiasi sensazione sonora, rappresenta la cosiddetta *soglia di udibilità*;
- il **timbro**, che è una caratteristica qualitativa dell'onda sonora. Le sorgenti sonore complesse, come la nostra voce o gli strumenti musicali, non generano dei suoni sinusoidali "puri", nel senso che il segnale emesso è composto da una serie di due o più onde audio, ognuna con propria frequenza ed ampiezza. Il segnale nel suo complesso ha una frequenza principale e delle frequenze secondarie (aventi frequenze multiple della principale) dette armoniche.

Osservazione. La nostra capacità di distinguere i suoni degli strumenti, anche sulla stessa nota musicale, dipende dal fatto che ogni strumento emette il suono con le stesse armoniche aventi, però, ampiezze diverse.

Esempi di suoni. Nei grafici seguenti, sono riportati le forme d'onda di suoni puri a diversa frequenza, ampiezza e timbro.

Fig. 1: suono a 100 Hz (audio/h_100.wav).	Fig. 2: suono a 2 KHz (audio/h_2k.wav).	Fig. 3: suono a 5 KHz (audio/h_5k.wav).
		
Fig. 4: suono a 1 KHz (audio/h_1k.wav).	Fig. 5: suono a 1 KHz con 10 decibel in più (audio/h_1k_p10.wav).	Fig. 6: suono a 1 KHz con 10 decibel in meno (audio/h_1k_m10.wav).
		



Osservazione. Come tutte le onde, le onde audio si propagano a velocità che dipende in generale dal mezzo fisico (aria, acqua, ecc.) e dalla temperatura. La propagazione nell'aria delle onde sonore, ad esempio, avviene a circa 330 m/s se la temperatura è a 0° C, mentre avviene a 340 m/s se la temperatura è a 20° C.

2. Il campionamento e la codifica digitale

Per convertire un'onda sonora analogica in digitale il computer deve essere in grado di misurarne l'ampiezza in istanti successivi, ciò avviene attraverso un convertitore analogico-digitale: ognuna di queste misure si chiama *campione*. Per la riproduzione di un suono digitale è necessario effettuare la conversione inversa attraverso un convertitore digitale-analogico, che ricostruisce un'approssimazione vicina al segnale analogico originario.

I sistemi di registrazione digitale generici utilizzano la codifica PCM (*Pulse Code Modulation*) che mira ad effettuare un'accurata lettura della forma d'onda. Al riguardo, le variabili da considerare, da cui dipendono la qualità e la dimensione di un file audio, sono essenzialmente tre: la *frequenza di campionamento*, la *profondità di bit* e il *numero dei canali*.

La frequenza di campionamento

La frequenza di campionamento (*sampling rate*) è uno dei fattori che determina la rassomiglianza tra il segnale digitalizzato e quello analogico in quanto descrive il numero di volte che il segnale stesso è campionato nell'unità di tempo (secondo), ad intervalli regolari. Una maggiore frequenza di campionamento aumenta, perciò, la fedeltà nella ricostruzione della forma d'onda originaria, ma comporta anche la registrazione di un numero maggiore di campioni, per cui una maggiore occupazione di memoria.

Fig. 8: esempio di campionamento.

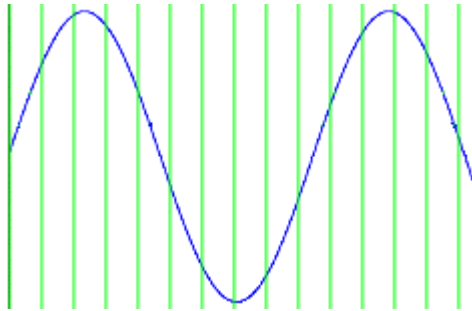
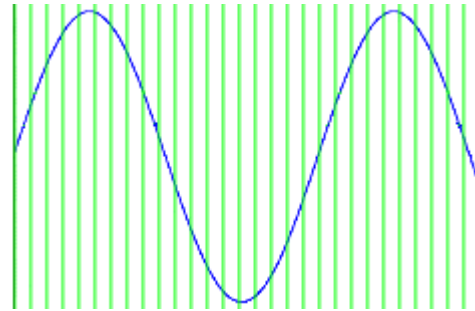


Fig. 9: esempio di campionamento più accurato, ad una frequenza più alta.



La scelta della frequenza di campionamento da utilizzare è cruciale; molto importante, a tal fine, si rivela l'apporto del teorema di Nyquist, secondo il quale "un segnale avente frequenza massima f_{max} può essere completamente ricostruito se si effettua il campionamento ad una frequenza f_c almeno doppia rispetto a f_{max} , ovvero dev'essere

$$f_c > 2 \times f_{max}.$$

Questo limite, puramente teorico poiché la frequenza massima del segnale potrebbe essere così alta da rendere praticamente impossibile la sua ricostruzione, in realtà si utilizza fissando il livello di qualità desiderata. In altre parole se ci attendiamo la ricostruzione completa del segnale fino alla frequenza f_{scelta} , basta prendere una frequenza di campionamento almeno doppia di f_{scelta} .

In campo telefonico, ad esempio, poiché non interessa la qualità audio ma solo l'intelligibilità del messaggio, viene destinata alla voce una banda di frequenza fra 300 e 3400 Hz (le massime frequenze vocali arrivano intorno ai 2 KHz) per cui si potrebbe campionare il messaggio a partire da 6,8 KHz (che è, appunto, il doppio). Un altro esempio significativo è dato dalla nota qualità dei brani musicali su CD (*CD-quality*). Nei Compact Disk la filosofia è quella di riprodurre fedelmente (al nostro orecchio) i brani musicali registrati; poiché la banda acustica del nostro orecchio arriva fino a (circa) 20 KHz, si è scelta come frequenza di campionamento 44,1 KHz (ovvero 44.100 campioni al secondo), ovvero il doppio con un po' di tolleranza.

La scelta di registrare ad una particolare frequenza di campionamento f_c implica automaticamente lo scarto delle frequenze reali del suono maggiori di $f_c / 2$; in ogni caso, nei comuni software, il numero dei campioni prelevabili può arrivare anche fino a 96.000 al secondo.

La profondità di bit

Anche la profondità di bit (*bitdepth*, detta anche risoluzione o quantizzazione) è uno dei fattori che determina la rassomiglianza tra il segnale digitalizzato e quello analogico. Essa determina il numero di livelli discreti (o valori di campionamento) che si considerano per campionare il segnale analogico, e quindi, determina l'accuratezza con la quale è effettuata ciascuna misurazione o campione, ovvero il grado di approssimazione di ogni campionamento dell'onda rispetto al suo valore reale. Esattamente, questa variabile si riferisce alla lunghezza delle parole binarie utilizzate per codificare ciascun campione del segnale.

Una profondità di bit maggiore, ovvero l'uso di parole binarie più lunghe, permette la rappresentazione di una serie più ampia di numeri e quindi una riproduzione più fedele del segnale.

Analogamente alla frequenza, però, ciò comporta anche una maggiore occupazione di memoria del file audio.

Tipicamente si utilizzano parole a 8-bit (con $2^8 = 256$ valori di campionamento distinti) oppure a 16-bit (con $2^{16} = 65.536$ valori di campionamento distinti). Gli 8 bit offrono una qualità acustica inferiore a quella di un nastro, mentre i 16 bit producono la massima qualità sonora (massima perchè il nostro orecchio non recepisce più di $2^{16} = 65.536$ suoni distinti). Più raramente esiste la possibilità di utilizzare profondità a 24 o 32 bit. ???Nelle figure e' da mettere 8 e 16 livelli???.

Fig. 8: esempio di profondità con 9 livelli
(ogni linea orizzontale è un livello).

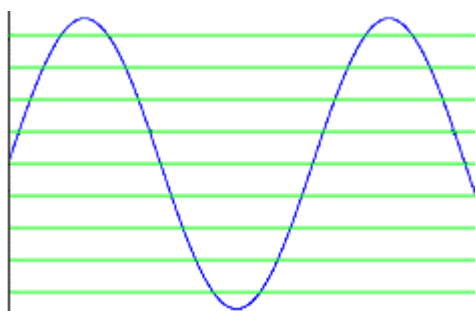
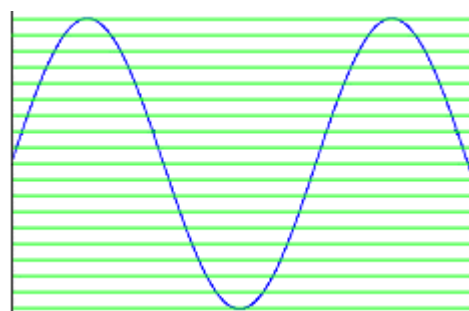


Fig. 9: esempio di profondità più accurata
con un numero (19) maggiore di livelli (di bit).



Osservazione. Un parametro molto importante a cui è legata la profondità di bit è la *dinamica*. Essa è misurata in decibel e rappresenta il rapporto fra il suono più forte e quello più debole che una stessa apparecchiatura (o un componente) può registrare e/o riprodurre. Molto semplicemente rappresenta la capacità di graduare in modo nitido l'intensità del suono (nel contesto complessivo) riproducendo nel giusto rapporto i picchi di intensità, i suoni di basso livello, e tutti i suoni la cui intensità è compresa tra i due estremi. Ad esempio, i brani di musica classica hanno una dinamica molto alta in quanto sono composti da suoni in un spettro di intensità molto ampio; dai suoni appena percettibili ai suoni molto intensi. Al contrario i brani di musica rock hanno una dinamica molto limitata.

Il rapporto con la profondità di bit consiste nel fatto che ad una maggiore profondità di bit corrisponde la possibilità di registrare e/o riprodurre una maggiore dinamica. Nella tabella seguente sono riportati alcuni valori della dinamica riproducibile in funzione del tipo di campionamento.

Bits per campione	Dinamica
4	24 dB
8	48 dB
12	72 dB
16	96 dB

Osservazione. Per quanto detto finora, la frequenza di campionamento e la risoluzione determinano insieme la fedeltà della ricostruzione analogica dell'onda audio originaria. Negli esempi sotto riportati si può evincere come ad una maggiore frequenza e/o risoluzione, il grafico dell'onda viene approssimato meglio.

Fig. 10: esempio di campionamento.

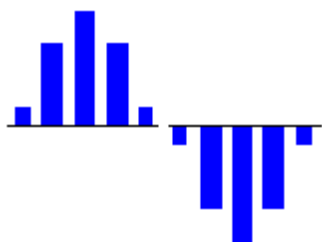


Fig. 11: esempio di campionamento più fedele.

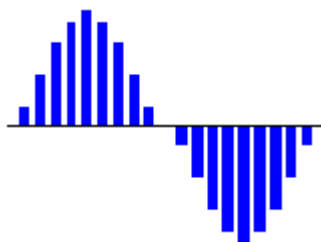
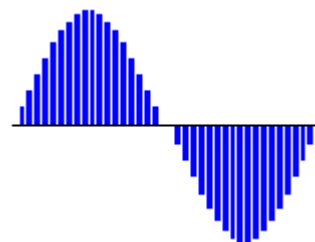


Fig. 12: esempio di campionamento ancora più fedele.



Numero di canali

Le modalità di ripartizione dei canali audio sono essenzialmente due:

- *Mono*, con un solo canale;
- *Stereo*, con due canali, sinistro e destro, separati.

Esistono, comunque, altre modalità di ripartizione che utilizzano più canali, utilizzate in special modo nel campo dei video, come ad esempio il Dolby surround.

Osservazione. Nella tabella seguente sono riepilogate le frequenze tipiche utilizzate per l'acquisizione digitale dei segnali audio.

Tipo qualità	Frequenza di campionamento	Profondità di bit	Mono/Stereo	Dimensione per un minuto
Telefono	8.000 Hz	8 bit	Mono	469 Kb
Parlato	11.025 Hz	8 bit	Mono	646 Kb
Radio mono	22.050 Hz	16 bit	Mono	2,52 Mb
Radio stereo	22.050 Hz	16 bit	Stereo	5,05 Mb
Audio cassetta	22.050 Hz	16 bit	Stereo	2,52 Mb
Compact Disk	44.100 Hz	16 bit	Stereo	10,1 Mb
Registr. DAT	48.000 Hz	16 bit	Stereo	11 Mb

Osservazione. Il prodotto delle tre variabili anzidette, (frequenza, profondità di bit e numero di canali) è definito *bitrate* (misurato in bps, bit per secondo) ed indica il numero di bit utilizzati nel campionamento, o nella memorizzazione, di un secondo di file sonoro. Il prodotto del bitrate per la durata del file sonoro definisce, invece, la dimensione complessiva del file.

Esercizio 1. Abbiamo un file sonoro in modalità stereofonica registrato con frequenza di campionamento 44.100 e con profondità di 16 bit (dunque in CD-quality). Quanto spazio occupa un solo secondo del file? A quanto ammonta il bitrate del file?

Soluzione. Ogni secondo vengono effettuati 44.100 campionamenti su due canali (in quanto stereo), per cui il numero di campioni per secondo è $44.100 \times 2 = 88.200$. Ciascun campione viene codificato con 16 bit = 2 byte, per cui la memoria occupata da un secondo di file è: $88.200 \times 2 = 176.400$ byte = 172,27 Kb. Infine il bitrate: è il calcolo che abbiamo appena fatto.

Esercizio 2. Vogliamo registrare digitalmente un file sonoro della durata di 2 minuti e mezzo, in modalità mono, con frequenza di campionamento a 22.050 Hz e con profondità di 8 bit. Sarà sufficiente un floppy disk per memorizzarlo?

Soluzione. Calcoliamo il bitrate:

$22.050 \text{ Hz (frequenza)} \times 1 \text{ (canale)} \times 1 \text{ byte (profondità)} = 22.050 \text{ Kbps (bitrate)}$

dunque la memoria necessaria per memorizzarlo è:

$22.050 \text{ Kbps (bitrate)} \times 150 \text{ secondi (durata)} = 3.307.500 \text{ bytes} = 3,15 \text{ Mb}$. Pertanto un floppy non è sufficiente.

Osservazione. Ricapitolando, la dimensione complessiva di un file audio (non compresso) si ottiene attraverso la formula:

*dimensione file = frequenza * numero canali * profondità di bit * durata in secondi del brano .*

Osservazione. Il problema principale per la diffusione dei file audio, come d'altronde di tutte le applicazioni multimediali, deriva dalle dimensioni troppo grandi dei file con campionamento PCM per essere utilizzate su Internet o su opere multimediali. A questo scopo sono state studiate alcune varianti (come DPCM, ADPCM) con l'obiettivo di alleggerire le dimensioni senza sacrificare troppo la qualità sonora. Queste tecniche permettono di gestire in maniera più efficiente la memorizzazione dei dati, ma finora non si sono ottenuti grossi risultati.

Anche l'utilizzazione di alcuni algoritmi tradizionali di compressione (come il formato Zip) non è preferita in quanto la riduzione delle dimensioni è generalmente inferiore al 30% (insufficiente per files di buona qualità) ed inoltre non è possibile riprodurre il file sonoro in tempo reale. Per risolvere questo problema è necessario un altro genere di approccio, come descriveremo nel seguito (formato MP3).

3. I formati

Il panorama dei formati per l'audio, pur non essendo ampio come nel caso della grafica, è comunque variegato. Quelli che si sono maggiormente affermati riguardano quattro principali modalità di utilizzo:

- per l'ascolto di qualità (ad esempio: WAVE, AIFF, AU), che deve necessariamente avere parametri molto alti, anche se ciò si ripercuote sulle dimensioni;
- per consentire la distribuzione/trasmmissione a dimensioni accettabili. In questo ambito alcuni formati hanno legato il loro nome principalmente allo streaming (ad esempio RA), altri lo hanno legato al download (ad esempio: MP3, VORBIS);
- per la creazione di brani come da spartito (ad esempio: MIDI).

Esamineremo, in questi contesti, i formati più diffusi. I primi tre formati riportati, WAVE, AIFF ed AU, sono tipicamente utilizzati anche per la registrazione digitale (iniziale) di suoni analogici.

Il formato WAVE (Waveform Audio File Format). E' lo standard di Windows per file sonori ed ha estensione ".wav". E' uno dei più diffusi, anche per i brani musicali presenti nei classici CD audio. Quando si registra da Windows con il "Registratore di suoni" si crea un file con questo formato, per questo è quasi divenuto uno standard assoluto. Permette di utilizzare, inoltre, diversi [codec](#) di compressione.

Osservazione. L'estrazione di audio digitale dai CD-audio, che avviene in maniera diversa dall'estrazione dei dati (con lo standard CDA) può avvenire a diverse velocità di lettura nel formato Wave. La velocità base è 176.400 byte/s (equivalente di 1 secondo a 44,1 KHz, con 16 bit su due

canali) e viene indicata con $1\times$. Tutte le altre velocità sono multiple, ad esempio $2\times$ indica il doppio, $4\times$ indica il quadruplo, $16\times$ indica sedici volte la velocità base, ecc.

Il formato AIFF (Audio Interchange File Format). E' il formato standard nei computer Macintosh, ma è anche comune per la memorizzazione e la trasmissione dei file sonori. L'estensione su PC è ".aif". Non supporta tecniche di compressione, ma esistono altri formati chiamati Aiff-compresi che supportano rapporti di compressione fino a 6:1.

Il formato AU. Molto diffuso e utilizzabile su varie piattaforme operative. E' utilizzato principalmente nei computer Unix ed ha estensione ".au". Gestisce, rispetto a Wave e Aiff, modalità più efficienti di quantizzazione che permettono una riduzione della mole di dati fino ad un rapporto di 4:1, anche se con piccola perdita di qualità.

Il formato RA (Real Audio). Di tipo compresso, con algoritmo di tipo lossy, produce un notevole abbassamento delle dimensioni ma provoca una perdita di qualità considerevole (la qualità rispetto al Wave è scadente). E' ascoltabile facendo uso del software riproduttore proprietario Real Player (il lettore è comunque distribuito gratuitamente) ed è molto utilizzato in Internet perché consente l'ascolto in streaming (in questo contesto giocano un ruolo importante i files di estensione ".rm" - creati automaticamente dal codificatore, contenenti i comandi che attivano lo streaming del file ".ra").

Il formato MP3 (MPEG 1 Layer 3). E' lo standard per la compressione di brani musicali in file di piccole dimensioni (fino ad $1:12$ delle dimensioni originarie del file) preservando la qualità sonora in fase di riproduzione. Molto utilizzato su Internet per permettere lo scaricamento completo di file sonori (operazione oggi possibile da un numero sempre crescente di siti Web), molto meno per l'ascolto immediato in streaming (che comunque è possibile). In ogni caso l'MP3 rappresenta l'algoritmo più potente in una serie di standard di codifica per l'audio sviluppati dal gruppo MPEG (Motion Picture Experts Group, a cui si devono i principali standard internazionali per la trasmissione di audio-video su Internet) e formalizzati dall'Organizzazione Internazionale ISO. L'algoritmo Mpeg 1 layer 3 (i layer 1, 2 e 3 sono livelli di compressione ordinati secondo l'evoluzione della tecnologia, ma sempre all'interno dello standard per il multimedia Mpeg1) poggia su un modello psicoacustico dell'orecchio umano, in base al quale discrimina i dati indispensabili, da quelli superflui, per la corretta ricostruzione del segnale. Ciò avviene in particolare attraverso un algoritmo di tipo lossy che opera una serie di tagli sulle frequenze "non udibili" (nel senso che non verrebbero comunque percepite) perché al di fuori della soglia di udibilità o perché "mascherate" da altre (in realtà si applicano anche altri principi come la duplicazione delle informazioni contenute sui due canali stereo). Per comprendere, almeno parzialmente, il funzionamento del modello psicoacustico, riportiamo alcune immagini con le spiegazioni dei concetti fondamentali: la soglia di udibilità ed il mascheramento.

Fig. 13: *Soglia di udibilità.*

Un suono di frequenza f (in KHz sull'asse x) che non ha una pressione sonora (in dB sull'asse y) che supera la curva di udibilità (in verde) non viene udito dal nostro orecchio. L'algoritmo MP3, pertanto, non lo prende in considerazione.

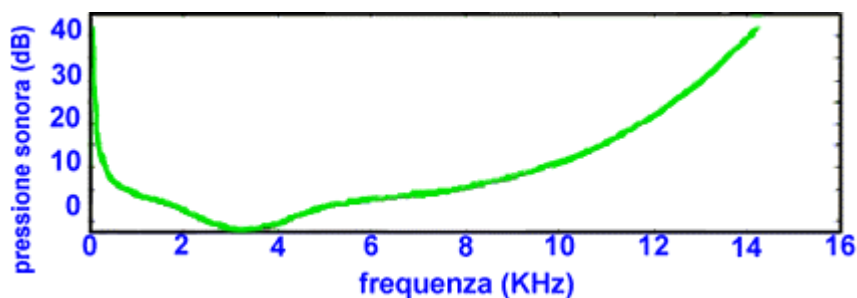


Fig. 14: *Mascheramento simultaneo.*

Un segnale di piccola ampiezza può essere non udibile (mascherato) per via della presenza di un segnale contemporaneo più intenso (mascherante), se le frequenze sono prossime tra di loro. Per ogni segnale si può definire una soglia di mascheramento (nell'immagine a fianco sono riportate le soglie corrispondenti ai suoni di frequenze .25, 1, 4 e 8 KHz tutti con stessa pressione) e i segnali sotto questa soglia non sono presi in considerazione dall'algoritmo.

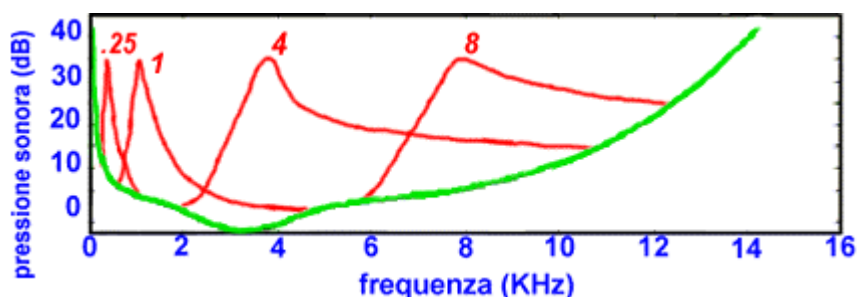


Fig. 15: *Mascheramento temporale.*

Si verifica quando due suoni si manifestano entro un breve intervallo di tempo: il segnale più forte può mascherare il più debole, anche se il mascherato precede temporalmente il suono mascherante, se quest'ultimo cade al di sotto di una curva di percettibilità (nell'immagine è illustrata la curva che indica la percettibilità a seguito dell'emissione di un segnale di 1 KHz, alla pressione acustica di 60 dB, che termina all'istante zero).

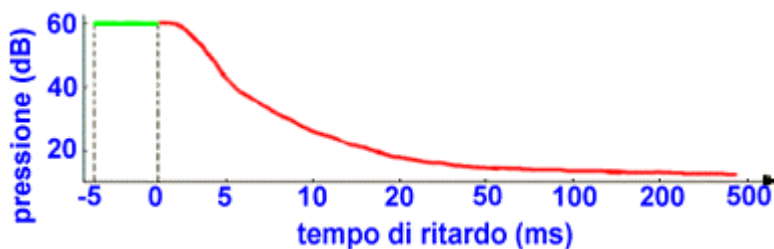
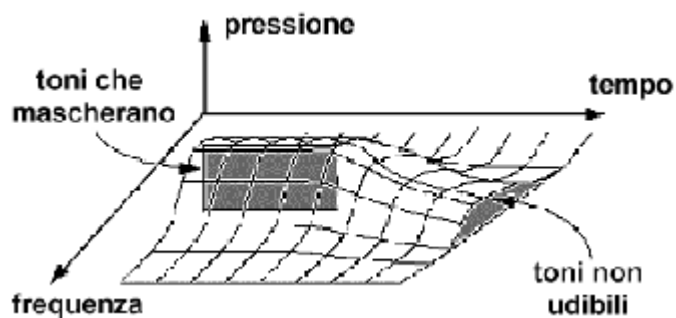


Fig. 16: *Mascheramento complessivo.*

Se in un segnale audio consideriamo un particolare tono, combinando i diversi tipi di mascheramento, si può costruire una superficie tridimensionale (tempo, frequenza e pressione come in figura) dalla quale sono identificabili i toni di conseguenza non udibili (in figura sono tutti quelli sotto la superficie quadrettata).



Con queste tecniche è possibile ottenere rapporti di compressione superiori a 1:10 (mediamente 1:12) senza una perdita apprezzabile di qualità sonora rispetto al Cd-Audio. In altre parole, un minuto di un brano in qualità CD (44,1 KHz, 16 bit, stereo) può essere ridotto da 10 Mb a circa 1 Mb (portandolo, cioè, ad un bitrate di 128 o 112 kbit per secondo) mantenendo, all'ascolto, un buon livello di qualità (generalmente si indica, come limite minimo accettabile, un bitrate di 64kbit/s). Sono comunque possibili livelli diversi di compressione che vanno da 320 kbit/s fino a 32 kbit/s. Queste tecniche, com'è facile immaginare, impegnano a fondo la CPU del PC, sia in fase di codifica che di decodifica.

Osservazione. Esistono diversi encoder (i più famosi sono Fraunhofer e Xing) per convertire files audio in MP3, basati sulle stesse tecniche ma con algoritmi diversi (ad esempio il taglio delle frequenze alte può essere fatto da partire da 16 KHz così come da 20 KHz, ecc.). Pertanto il solo bitrate non è in grado di definire la qualità, che infatti può risultare anche molto diversa a parità di bitrate.

Osservazione. Per creare un file MP3 a partire da un brano su CD sono necessari due tipi di software: un "ripper" per l'estrazione della traccia audio, e un "encoder" per la codifica effettiva in MP3; l'operazione è dunque relativamente semplice. Poiché la copia e la successiva redistribuzione di musica è protetta dalle leggi sul copyright, i siti Web che semplicemente promuovono questo formato sono a volte "accusate" di incoraggiare le violazioni del copyright.

Il formato OGG VORBIS. Questo formato, con estensione ".ogg", è un formato di compressione lossy per l'audio di tipo open source e con licenza gratuita. E' stato sviluppato come alternativa ai formati proprietari per la codifica digitale dell'audio, come ad esempio MP3. Pur fornendo la possibilità di specifiche audio migliori, come ad esempio il numero di canali audio (fino a 255) e la migliore qualità, non è molto diffuso, soprattutto a causa dei software per la lettura audio (è supportato solo da alcuni). Altri formati del progetto OGG sono: Flac per compressione lossless, Theora per il multimedia.

I formati più recenti: WMA e MP3Pro. Il formato WMA (Windows Media Audio) è l'ultimo formato della Microsoft per la compressione audio. Realizzato come alternativa ad MP3, di cui migliora la compressione di circa il 20%, in realtà è più di un'alternativa in quanto la stessa casa lo ha "imposto" come nuovo formato sul lettore integrato Media Player di Windows XP (non supportando più, contemporaneamente, il formato MP3).

Il formato MP3Pro, invece, è un'evoluzione del formato omonimo. Esso mantiene inalterata la qualità di riproduzione ma dimezza la grandezza dei files (mp3) attraverso algoritmi più potenti specialmente sulle frequenze alte. Pur essendo compatibile con il vecchio formato, la lettura (sui

vecchi lettori) fornisce una qualità di riproduzione più bassa, ma soprattutto la codifica non è ancora supportata da un numero significativo di encoder.

Il formato MIDI (Musical Instrument Digital Interface). E' un tipo di file audio completamente diverso in cui, piuttosto che i campioni di un'onda sonora, sono registrate le istruzioni per suonare le note, con i relativi tempi e con diversi strumenti musicali, come da spartito. Ideati per codificare i gesti compiuti da un musicista, gli standard MIDI stabiliscono un linguaggio musicale comune, che consente lo scambio di informazioni tra i diversi software.

I files, creati attraverso sequencer, hanno estensione ".mid" e dimensioni notevolmente inferiori, anche rispetto ai formati compressi. Pur avendo il vantaggio di essere modificabili nota per nota, è quasi impossibile riprodurre un brano come l'originale, anche se ciò dipende dall'abilità del creatore midi (lo svantaggio principale consiste nell'impossibilità di inserire la voce). La loro riproduzione, infine, può variare anche di molto su hardware diversi, in quanto ciascun produttore, pur rispettando la tipologia dei timbri, esprime il potenziale sonoro delle proprie macchine in maniera diversa.

Osservazione. Una tipologia particolare di file midi sono i cosiddetti "Midi con testo", o anche i file con formato ".kar" (karaoke), che incorporano delle parti testuali da visualizzare su schermo durante la riproduzione audio.

I moduli musicali. Con alcuni formati (come Mod, Xm, It, S3m) è possibile coniugare alcuni aspetti del campionamento con la modulazione. Con questi formati, oltre alle informazioni per l'esecuzione audio, di solito molto leggere, si possono memorizzare veri campioni dei suoni originali. In tal modo si può acquisire il timbro di uno strumento e riprodurlo sopra uno spartito, producendo, di conseguenza, un suono reale e non imitato da una scheda sonora.

Cos'è un codec. Un codec è un algoritmo di codifica/decodifica (compressione/decompressione) utilizzato, appunto, per codificare e decodificare vari tipi di dati, particolarmente quelli che altrimenti richiederebbero una notevole quantità di memoria, come i files audio e video. Fra i codec più comuni vi sono quelli che permettono di convertire un segnale analogico in suono digitalizzato. Possono essere usati sia per l'invio di dati in streaming in tempo reale che per la compressione di file da memorizzare.

Cos'è lo streaming. E' una modalità di trasmissione di un flusso di dati (stream) (filmati, brani musicali) da una sorgente che consente lo scaricamento ed il loro utilizzo in tempo reale, senza attendere che lo scaricamento sul computer sia terminato. Esistono due diverse modalità: in diretta (live streaming, che richiede protocolli di trasmissione appositi) e in differita.





[torna all'indice](#)

1. Cenni sul Video Digitale

Un video si compone, tipicamente, di immagini in sequenza e di audio (sincronizzato). Molti dei concetti introdotti nelle pagine dedicate alla [grafica](#) e all'[audio](#) trovano fondamento nel video digitale; a questi va aggiunto il concetto delle immagini in sequenza.

Il nostro occhio, ed in particolare la retina, quando viene colpito da una sensazione luminosa la ritiene per un certo lasso di tempo. Questa caratteristica, meglio sintetizzata come la capacità di permanenza dell'immagine sulla retina, viene sfruttata nel video, la cui riproduzione consiste in una serie di immagini fisse in sequenza. La discretizzazione, in questo caso, consiste nell'identificazione della sequenza di immagini statiche, chiamate fotogrammi (o frame), che compongono il video.

Dunque un concetto fondamentale è dato dal numero di fotogrammi (o *framerate*, indicato spesso con fps) che vengono trasmessi da un segnale video. Se questo numero è alto, il nostro occhio non sarà in grado di distinguere le immagini statiche e si avrà la sensazione di un movimento fluido, se questo numero è piccolo, invece, la sensazione è quella di "scattosità" delle immagini. Il nostro occhio percepisce un movimento fluido delle immagini se queste sono almeno 24 al secondo (24 fps) e fino a 30 fps si notano miglioramenti; oltre questo limite, generalmente, la differenza è minima. Ad esempio, nel sistema televisivo europeo PAL sono previsti 25 fps.

Nei video digitali, in particolare, il framerate concorre a formare il concetto principe, in grado di individuare la qualità e/o le dimensioni di un video, ovvero il *bitrate* (o *data rate*, che consiste nel numero di bit utilizzati nella codifica di un secondo di video). Il bitrate di un video si compone pertanto di:

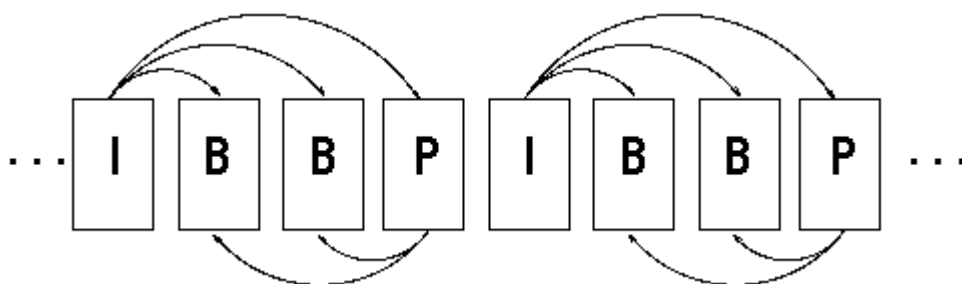
1. *bitrate del video*, in cui sono importanti
 - la dimensione delle immagini statiche, data dal prodotto tra la risoluzione (espressa come punti complessivi del quadro, ad esempio 320×240) e la profondità di colore;
 - la quantità di immagini in sequenza, ovvero il framerate;
2. e *bitrate dell'audio*, dato dal prodotto tra frequenza di campionamento, modalità di riproduzione e profondità di bit (come già visto nella sezione [audio](#)).

2. Cenni sulla compressione video

Le caratteristiche principali nella compressione video, successivamente seguite anche da altri standard di fatto, sono quelle tracciate nel 1991 dal comitato MPEG 1, che stabiliscono la struttura degli audio e video digitali compressi che rientrino all'interno del bitrate di 1,5 Mbit/s. Le caratteristiche specificano come comprimere i segnali audio e i segnali video e come integrarli consentendo la sincronizzazione.

Le tecniche utilizzate derivano in gran parte da quelle proposte dallo standard JPEG per la compressione di immagini fisse. Esse vanno dalla decimazione della crominanza (riduzione delle informazioni sulla crominanza delle immagini), alla trasformata discreta coseno, alla quantizzazione dei coefficienti, alla codifica di Huffman.

Una tecnica caratteristica nella compressione delle immagini in movimento è la cosiddetta "compensazione di moto" (*motion compensation*). Poiché molti fotogrammi sono simili, con questa tecnica si possono memorizzare interamente alcuni fotogrammi, che fanno da riferimento, ed indicare soltanto le variazioni per quelli successivi (da qui il nome della tecnica). Ad esempio nello standard Mpeg sono individuati tre tipi di fotogrammi: fotogrammi (I) codificati singolarmente ed indipendenti, fotogrammi (P) predetti sulla base di quelli indipendenti, fotogrammi (B) ottenuti attraverso un'interpolazione fra i fotogrammi dei due tipi descritti. Ovviamente l'ordine del fotogramma nel video non coincide, generalmente, con l'ordine di invio al lettore, che deve interpolare tra fotogrammi non consecutivi. La figura seguente illustra schematicamente la tecnica.



2. I formati

Vediamo in rapida rassegna in principali formati con qualche cenno alle caratteristiche principali.

Il formato AVI (Audio Video Interlaced). E' il classico formato video proprietario della Microsoft, il più conosciuto ed utilizzato dagli utenti PC. Le dimensioni dei files in questo formato sono altissime, ma sono supportati moltissimi codec di compressione (quali SuperMac Cinepak, DivX, MS Video1, MS RLE, Intel Indeo, M-Jpeg, ecc.). Utilizzato soprattutto per l'acquisizione e l'editing dei video. Non permette lo streaming.

Il formato MPEG 1 (MPG, MPE, MPA, MPEG, M1V, DAT). Si avvale dell'algoritmo di compressione MPEG 1, è utilizzato in particolare nei Video CD (VCD). Offre una buona qualità video e può contenere audio stereo, video e immagini con una compressione di 30:1. Questo compressore utilizza un data rate fisso di 1150 Kbit/s per il video e di 224 Kbit/s per l'audio. Le estensioni possono essere diverse.

Il formato MPEG 2 (MPG, M2V, M2P, MPV2, MP2V). Si avvale dell'algoritmo di compressione MPEG 2 che ha la peculiarità del bitrate variabile, richiede hardware molto potente per la cattura e la compressione. Garantisce una qualità audio/video elevata ed include anche il Surround su diversi canali. Una delle principali applicazioni è il Super Video CD (SVCD).

Osservazione. Le ultime versioni, MPEG 4 e MPEG 21, sono standard molto più ambiziosi che utilizzano tecniche che arrivano fino ad un approccio dell'intelligenza artificiale per ricostruire le immagini. In ogni caso forniscono possibilità superiori di gestire gli oggetti multimediali all'interno dei filmati. Qualche problema è però legato alla lettura, in quanto i vecchi lettori DVD non supportano questi nuovi formati.

Osservazione. Menzione particolare merita il codec "DivX ;-)", o semplicemente DivX, che ha raggiunto larga diffusione. Esso utilizza la compressione secondo lo standard MPEG 4, ma ne migliora le prestazioni, e può applicare una riduzione con rapporto di 10:1 sui video basati su MPEG 2 (ad esempio DVD) con qualità accettabile (anche l'audio è generalmente in formato Mp3).

Il formato VOB. Anche questo formato si avvale dell'algoritmo MPEG2 ma contiene anche audio a 6 canali Dolby con sottotitoli, punti di vista multipli e menu interattivi. L'applicazione principale è il DVD.

Il formato Microsoft Netshow (ASF). Formato che succede all'AVI per permettere lo streaming. In realtà il rapporto qualità/compressione è ritenuto generalmente buono sia per il Web che per riproduzioni off line.

Il formato Windows Media (WMA, WMV). Utilizzato particolarmente in rete per lo streaming in quanto utilizza una tecnologia scalabile che crea varie versioni dello stesso filmato in MPEG4 a bitrate diversi.

Il formato Apple Quick Time (MOV). Formato proprietario della Apple e principale estensione multimediale del sistema operativo Macintosh, passato in seguito su PC per la grande diffusione. Gli algoritmi di compressione sono di varia natura e scopo (dal Jpeg a quelli proprietari Apple). Utilizzato su Cd-rom, video e applicazioni interattive on line.

Il formato Real Media (RAM). Utilizzato principalmente per lo streaming audio/video, è quasi diventato lo standard per la creazione e distribuzione di multimedia sul Web. In fase di streaming la qualità sonora è molto alta, quella video è inferiore (ma qualità e fluidità delle immagini dipendono anche dalle capacità della macchina locale che riproduce il filmato). Da notare che l'encoder è scaricabile gratuitamente on line.

Il formato Digital Video (DV). E' il formato di acquisizione delle nuove videocamere, con compressione Motion Jpeg. Le specifiche tecniche variano per ogni azienda produttrice.

