

Che cosa è, come funziona: Uno standard pervasivo (MPEG-2 video)

Parte II

ing. Marzio Barbero e
ing. Natasha Shpuza

1. Premessa

Nella prima parte, pubblicata nel numero di aprile 2003, sono stati descritti algoritmi e sintassi alla base della codifica del video secondo lo standard MPEG-2. In questa seconda parte ci si sofferma sull'evoluzione delle applicazioni dello standard, ampiamente diffuse come si desume dal titolo, e sui miglioramenti della qualità del video codificato.

2. Bit-rate e qualità

L'attenzione dei *broadcaster*, ovvero delle reti televisive quali la Rai, fu focalizzata fin dall'inizio della attività di normalizzazione dello standard MPEG-2 sulla qualità delle immagini, a garanzia del servizio fornito al telespettatore.

La codifica digitale del video si differenzia da quella analogica per il fatto che la perdita di informazione (e in genere di qualità) è concentrata soprattutto nella fase di compressione: in condizioni normali il canale non introduce degradamenti percepibili (grazie alle tecniche di protezione dagli errori) e il sistema di decodifica e visualizzazione riproduce correttamente le immagini ricevute.

MPEG-2 utilizza algoritmi di compressione *lossy* ed in particolare il degradamento della qualità è dovuto alla precisione con cui vengono trasmessi i coefficienti DCT.

Il fattore di compressione è il rapporto fra bit-rate dell'informazione video originaria (circa 166 Mbit/s nel caso di un segnale video con-

forme alla Rac. ITU-R BT.601, formato 4:2:2, e circa 125 Mbit/s nel formato 4:2:0) e il bit-rate medio del video codificato.

Nel caso del MP@ML si riesce ad ottenere una qualità dell'immagine accettabile dallo spettatore con fattori di compressione superiori a 20 (formato 4:2:0, bit-rate video medio in genere inferiore a 6 Mbit/s).

In figura 1 sono riportati i risultati di prove soggettive svolte al Centro Ricerche Rai all'epoca in cui era in corso la definizione dello standard, al fine di definire i requisiti e valutare le prestazioni degli algoritmi. Le prove misero in evidenza l'opportunità di operare con bit-rate dell'ordine dei 6 Mbit/s al fine di garantire una qualità percepita comparabile con quella delle immagini televisive codificate PAL (qualità in studio).

I sistemi di ripresa e di grafica elettronica digitale possono produrre immagini caratterizzate da una elevata definizione e ricchezza di dettagli, su cui gli algoritmi di compressione hanno difficoltà a operare, data la scarsa correlazione dell'informazione. Le sequenze più critiche da codificare sono specialmente quelle relative agli eventi sportivi e quelle in cui sono presenti elementi grafici. Alcuni esempi di quadri tratti da sequenze di test critiche sono riprodotti in figura 2.

La figura 3 illustra le caratteristiche del rumore di quantizzazione nel caso di una sequenza relativa al *basketball*, caratterizzata da molti dettagli e da movimenti rapidi (sia della telecamera che degli oggetti che compongono la scena), e dalla presenza di oggetti di tipo

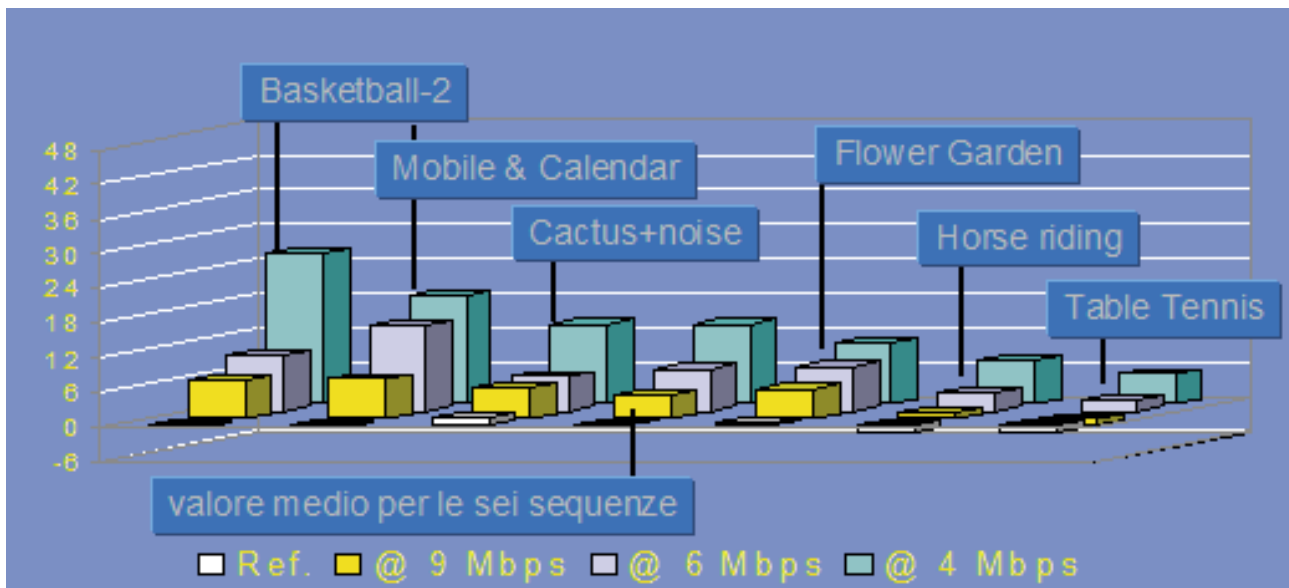


Fig. 1 - Questo istogramma riporta i risultati di prove soggettive eseguite presso il Centro Ricerche Rai per la valutazione del sistema MPEG-2 video, allora in fase di definizione.

Per valutare la qualità dell'immagine si usano metodi di valutazione soggettiva descritti nella raccomandazione ITU-R BT 500.

Il metodo DSCQS (Double Stimulus Continuous Quality Scale) è stato utilizzato per valutare la qualità nel caso dei sistemi di compressione video. Gli osservatori (almeno 15) devono indicare il livello di qualità su una scala, successivamente quantizzata da 0 a 100, sia di una sequenza di riferimento (normalmente l'originale 4:2:2) che della stessa sequenza codificata. Le due sequenze, normalmente di 10 s, sono presentate due volte una dopo l'altra, ma l'osservatore non sa quale è il riferimento (le sequenze sono indicate come A e B). Il risultato finale è costituito dal valore medio delle differenze tra i punteggi assegnati al riferimento e alla sequenza codificata. I valori differenza possono quindi variare da -100 a +100, e valori negativi indicano che la sequenza codificata è stata valutata migliore del riferimento.

In questo caso i test furono condotti su sei sequenze e per tre differenti bit-rate (4, 6 e 9 Mbit/s) ed il massimo degradamento è percepibile nel caso di una sequenza di sport (denominata *basketball-2*) al bit-rate di 4 Mbit/s.



Fig. 2 - Grazie all'esperienza acquisita nel campo della valutazione della qualità e nei sistemi di codifica con compressione, diverse sequenze selezionate dal Centro Ricerche Rai furono utilizzate nella fase di valutazione degli algoritmi e dei codificatori. Questa è una frase tratta da un documento di un esperto britannico: *"I have used the Rai MPEG test sequences on a number of MPEG codecs, and they are certainly system crackers. It is difficult to convince people that they were not recorded specially"* ("Ho usato le sequenze di test MPEG della Rai con diversi codec MPEG, e sono sicuramente in grado di mettere in difficoltà il sistema. E' difficile convincersi che non sono state realizzate appositamente").



Fig. 3 - E' riprodotta, in alto, una singola *picture* della sequenza *basketball-2*, che, come si desume dall'istogramma di figura 1, è particolarmente critica. La criticità è dovuta alla presenza di numerosi dettagli, dal movimento (sia della telecamera che dei giocatori) e dall'attenzione posta presumibilmente dall'osservatore ai dettagli (per esempio le scritte, anche pubblicitarie). Sono quindi in successione riprodotte, dall'alto verso il basso, le differenze di luminanza fra la stessa *picture* codificata rispettivamente a 9, 6 e 4 Mbit/s e l'originale in alto; le differenze sono amplificate in figura per rendere più evidente soprattutto la distribuzione del rumore di quantizzazione. E' ben visibile l'aumento di rumore di quantizzazione, al crescere del fattore di compressione. Il nostro sistema psicovisivo è meno sensibile al rumore localizzato in prossimità dei contorni, rispetto a quello che interessa le zone uniformi e il sistema di compressione sfrutta questa nostra limitazione. Ma, per alcuni dettagli "importanti", quali i loghi e le scritte, il degradamento può essere percepito e fastidioso.



grafico (caratterizzati da contorni netti fra zone uniformi e di colore differente).

Il sistema MPEG-2 ha dimostrato fin dall'inizio la sua capacità di offrire una buona qualità del video co-decodificato a fattori di compressione elevati e con immagini anche critiche. Il sistema, come si è detto nella parte prima di questa scheda, definisce la sintassi e la struttura dei dati e lascia ampia possibilità all'industria nell'ottimizzazione dei codificatori. Alcune industrie hanno pubblicato grafici in cui indicano che vi è stato un miglioramento costante dall'apparizione del primo codifica-

tore commerciale (nel 1994) ai giorni nostri. Sostengono che la stessa qualità video ottenibile all'inizio con 6 o 8 Mbit/s è ora ottenibile con circa 2 Mbit/s. Quali sono i motivi di un tale progresso?

3. L'evoluzione del MP@ML

3.1 L'integrazione

L'attenzione dell'industria fu focalizzata nella fase di definizione del sistema soprattutto nell'ottenimento del miglior compromesso fra complessità del decodificatore e qualità dell'immagine a bit-rate relativamente bassi. Infatti lo scopo principale era la definizione del MP@ML per applicazioni di tipo *consumer* (decodificatore per i STB, *set-top-box*, per la TV digitale da satellite o terrestre e per i lettori di DVD) in cui è fondamentale il costo del terminale di utente, che in questi casi si limita alla sola decodifica. In queste applicazioni il codificatore può essere molto più complesso poiché è un apparato di tipo professionale, utilizzato da chi produce o diffonde TV o DVD.

Pertanto inizialmente i processi di integrazione furono adottati soprattutto per realizzare i *chip-set* (l'insieme di microcircuiti per elaborazione e memorizzazione) necessari al decodificatore.

In questi anni l'evoluzione (in termini di velocità di elaborazione e di architettura) delle CPU alla base dei PC ha permesso la realizzazione di decodificatori software in grado di decodificare MPEG in tempo reale.

Un codificatore MPEG è conforme allo standard, cioè produce un flusso sintatticamente corretto, anche se non utilizza o utilizza in modo limitato alcuni dei *tool* previsti dallo standard. Nei primi codificatori si accettava quindi una riduzione di qualità video per contenere, costi, ingombro e consumo. La qualità ha successivamente tratto significativi vantaggi dall'incremento del livello di integrazione dei microcircuiti e dall'evoluzione dei DSP, *digital signal processor*, sia in termini di velocità e di architettura: una maggiore complessità del

codificatore permette di sfruttare tutte le possibilità offerte dallo standard.

Le tecniche per ridurre i degradamenti percepibili a più bassi bit-rate sono:

- miglioramento delle decisioni in termini di modi di scansione e codifica (*field/frame*)
- ottimizzazione della stima del movimento e della scelta della predizione
- miglioramento del controllo del *buffer* di trasmissione in funzione della struttura del GOP (*picture* di tipo I, P e B)
- analisi del contenuto dell'immagine per adattarsi alle caratteristiche psicovisive (ad esempio, è più percepibile il rumore di quantizzazione in aree a luminanza più bassa)
- pre-elaborazione delle immagini per identificare e ridurre il rumore dovuto alla sorgente, al fine di ottimizzare la qualità percepita dell'immagine co-decodificata.

Le industrie quantificano questi miglioramenti come un guadagno in termini di bit-rate, a parità di qualità percepita, di circa il 50%, ovvero la stessa qualità soggettiva che nel 1994, quando furono disponibili i primi codificatori commerciali, richiedeva 7 Mbit/s era ottenibile cinque anni dopo con circa 3,5 Mbit/s.

Più recentemente le applicazioni di tipo *consumer* della codifica MPEG-2 si sono ampliate. Infatti si sta estendendo il mercato dei DVR, ossia i registratori di dischi ottici scrivibili e riscrivibili (Elettronica e Telecomunicazioni, n.3 dicembre 2002, "Dischi scrivibili e riscrivibili") e alcune applicazioni richiedono la codifica in tempo reale del segnale video (registratori e DVDCAM). Per questa ragione sono stati sviluppati *single-chip audio-video encoder* per la produzione di apparati *consumer* (tabella 1), in grado di operare sui multiplex DVB e DVD.

3.2 CBR, VBR e multiplex statistico

Nella prima parte (§ 5.3) si sono introdotti i termini CBR (*constant bit-rate*) e VBR (*variable bit-rate*).

Nel caso in cui si debba comprimere di un fattore costante il flusso video è fondamentale

il ruolo del *buffer*, la memoria tampone che adatta il flusso di dati a velocità variabile in uscita dal codificatore VLC alla velocità fissa del canale (di trasmissione o registrazione). Le dimensioni del *buffer* sono fissate dallo standard: nel caso di MP@ML è pari a 1835008 bit, mentre per il 422P@ML è di 9437184 bit. Se si opera a CBR, la capacità del *buffer* può essere espressa in termini di millisecondi di video: tale valore costituisce anche il maggior contributo al ritardo di co-decodifica, come si desume dalla tabella 2: tale ritardo è tanto più elevato quanto è più basso il bit-rate con cui viene codificato il video.

L'algoritmo di controllo del bit-rate che, in base allo stato di riempimento del buffer, varia la precisione della quantizzazione dei coefficienti DCT, è molto importante per ottimizzare la qualità dell'immagine. Le *picture* di tipo I generano picchi nel flusso di dati in uscita dalla codifica VLC, mentre in corrispondenza delle *picture* P e soprattutto quelle B si

Tab. 1 - A scopo esemplificativo, sono riportate alcune specifiche di un co-decodificatore single chip disponibile dalla fine del 2002 (dati tratti dal *data sheet* del uPD61051 della NEC). Il costo di questo chip era stato annunciato, per quantità superiori a 100000, pari a circa 30 \$.

codifica video	MP@ML, SP@ML, MPEG-1
preanalisi	single-pass VBR o CBR
stima del movimento	identificazione film, cambio scena +/- 128 in orizzontale, +/- 64 in verticale per le <i>P-picture</i> ; +/- 96 e +/- 48 per le <i>B-picture</i>
GOP	intervallo massimo fra due <i>P-pic- ture</i> consecutive pari a 3 (IBBP)
formato immagine	
- orizzontale (pixel)	720, 704, 544, 480, 352
- verticale (righe)	480, 240, 576, 288
- freq. quadro [Hz]	29,97 e 25
codifica audio	MPEG-1, layer 2
- lunghezza parole [bit]	16, 20 e 24
- freq. camp. [kHz]	32, 44,1, 48
multiplex, de-multiplex	MPEG-2 TS e PS, DVD-Video, DVD-VR
bit-rate	1 - 15 Mbit/s
interfacce	
- video input/output	4:2:2 ITU-R BT.656 8 bit
- audio input/output:	2 canali PCM
- host I/F	seriale, parallela, PCI
- memoria esterna	64/128 Mbit SDRAM 32 bit bus
consumo	1200 mW (tipico)
package	208-pin fine pitch QFP

ha uno svuotamento del *buffer*. Gli algoritmi di controllo sono in genere basati su modelli ottenuti a partire da sequenze video tipiche, di *training*. Quando è possibile prevedere, analizzando le *picture* prima della codifica, cambiamenti di scena o effetti speciali il codificatore può variare la struttura regolare del GOP introducendo *I-picture* nel momento più opportuno; nel caso di più quadri identici, ad esempio la trasmissione di una diapositiva o un cartello, la prima *picture* può essere di tipo *I*-, quantizzata nel miglior modo possibile, sfruttando i significativi risparmi consentiti dalle successive *P*- e *B-picture*.

La modalità VBR è utilizzata per la produzione di film su DVD: in questo caso i vincoli sono relativi alla qualità e alla durata complessiva, espressa in minuti, del film, mentre la gestione del *buffer* diventa meno vincolante, essendo possibile variare la velocità di scrittura/lettura dei dati sul disco. La codifica può avvenire in più passaggi: durante una prima emulazione di codifica vengono memorizzati i dati relativi al riempimento del *buffer* e dei principali parametri (in particolare relativi alla quantizzazione); in base a questi dati è possibile effettuare un secondo passaggio in cui i parametri vengono forzati ottenendo un flusso VBR che ottimizza la qualità video e consente la memorizzazione del film sfruttando l'intera capacità del disco. I passaggi possono essere anche più di due e si può perfino prevedere l'intervento umano,

ovvero un esperto visiona il film e individua le scene particolarmente critiche o importanti e forza le modalità di codifica per ottimizzare la qualità soggettiva.

E' proprio l'ottimizzazione nei casi di utilizzo di VBR che permette di ottenere una codifica MP@ML video a circa 2 Mbit/s medi, pur mantenendo una buona qualità soggettiva, .

Nel caso della codifica per la diffusione in tempo reale del segnale televisivo è possibile sfruttare la variabilità statistica della criticità delle immagini solo nell'ambito di una porzione temporale del programma dell'ordine di grandezza corrispondente alla capacità del *buffer*. Per la diffusione televisiva digitale da satellite è stata però applicata la tecnica della moltiplicazione statistica. Un *transponder* da satellite consente il transito di un flusso di dati a bit-rate costante (che dipende dalla banda e dal codice di canale) che contiene un insieme di programmi televisivi: il *bouquet*. E' quindi possibile gestire i *video buffer* relativi ai singoli programmi come se costituissero un solo *buffer* comune all'intero *bouquet*, adattando così la quantizzazione alla criticità statistica mediata su più programmi: in genere non tutte le sequenze sono contemporaneamente critiche, la criticità media è quindi statisticamente inferiore a quella delle singole sequenze.

La tecnica di moltiplicazione statistica è nor-

Tab. 2 - L'operazione di co-decodifica comporta un ritardo. Il ritardo complessivo è dovuto a quattro contributi, in ordine decrescente:

- le dimensioni del buffer, sono state scelte come compromesso fra la qualità di codifica e il ritardo introdotto

- il riordino dei quadri che costituiscono il GOP, in generale è pari a 40 ms per il numero di *B-picture* costituenti il GOP

- codifica per quadro o per semiquadro, se i due semiquadri vengono codificati in modo progressivo si introduce un ritardo di 20 ms

- ritardo dovuto all'hardware, difficile da quantificare, dipende dall'apparato e dalla configurazione utilizzata.

In tabella sono riportati i valori di ritardo calcolati nel caso di codifica *frame-picture* nel caso dei due profili 422P e MP e considerando diverse combinazioni di bit-rate e struttura del GOP. Il contributo dovuto all'hardware non è considerato.

Profilo	dimensioni buffer (max) [bit]	bit rate [Mbit/s]	struttura GOP	ritardo [ms]
422P@ML	9437184	50	solo Intra	209
		20	IBIB	532
		20	IBBP	572
MP@ML	1835008	15	IPPP	142
		15	IBBP	222
		9	IBBP	303
		6	IBBP	405
		4	IBBP	557

malmente applicata nella diffusione televisiva digitale via satellite e terrestre ed è tanto più efficace quanto più è elevato il numero di programmi che costituiscono il *bouquet*.

3.2 La qualità finale

La qualità video percepita è quindi dovuta a diversi contributi.

La qualità della sorgente, soprattutto in termini di rumore, ha un ruolo importante: un'immagine rumorosa è in genere più critica da codificare perché i campioni sono meno correlati fra loro, ma le pre-elaborazioni e la stessa applicazione della DCT possono ridurre il rumore finale a tal punto che l'osservatore, talvolta, può preferire l'immagine co-decodificata, seppure meno definita, rispetto all'originale.

Sequenze caratterizzate da movimenti veloci, tipicamente le riprese sportive, sono più critiche. I film in genere sono meno critici poiché la codifica in modalità *frame* dei 24 fotogrammi al secondo della pellicola è molto efficiente e le tecniche di ripresa sono tali da evitare sequenze con movimenti rapidi. Inoltre, nel caso di codifica per DVD, si adottano tecniche di codifica a più passaggi.

L'evoluzione della tecnologia, alimentata dalla forte competizione nella realizzazione di codificatori più economici ed efficienti, ha consentito di sfruttare appieno i *tool* messi a disposizione dallo standard.

D'altro canto è sempre l'innovazione tecnologica la causa di una maggiore visibilità dei difetti introdotti dalla compressione a basso bit-rate sugli schermi. Gli schermi piatti di grandi dimensioni, realizzati con tecnologia TFT, plasma e DMD, hanno un formato predefinito (dato dal numero di celle costituenti) e adatto a segnali con scansione progressiva e sono dotati di sistemi di pre-elaborazione e interpolazione delle immagini che possono enfatizzare i degradamenti (Elettronica e Telecomunicazioni, n. 2 agosto 2002, "Qualità dell'immagine percepita sugli schermi piatti").

E' quindi vero che a parità di bit-rate, la qualità è migliorata nel corso degli anni, ma è anche

vero che, per garantire buone immagini anche nella situazioni più critiche, è buona norma accettare dei costi di distribuzione più elevati e non affollare eccessivamente i *bouquet*, nel caso della diffusione televisiva, e utilizzare DVD a doppio strato (con capacità fino a 8,54 GB) nel caso di film lunghi.

La qualità del prodotto finale dipende anche dalle operazioni di produzione e post-produzione, in particolare decresce all'aumentare del numero di co-decodifiche in cascata. In genere il video viene ripreso, codificato per la registrazione o la trasmissione sulla rete di contributo verso gli studi televisivi, post-prodotto (nel caso di post-produzioni complesse il numero di co-decodifiche può essere prossimo alla decina), archiviato, codificato per la diffusione. Ciascuna operazione di co-decodifica, se effettuata con fattori di compressione elevati, introduce rumore di quantizzazione.

4. Le motivazioni per il 422P@ML

Il profilo principale (MP@ML) ha chiaramente raggiunto l'obiettivo, quello di consentire la diffusione del video digitale a casa degli utenti, a livello mondiale. Non è però adatto a soddisfare le esigenze di qualità nei processi di produzione e post-produzione televisivi.

Il raffinamento dei *tool* realizzato da MPEG permette di ottenere, per la rete di contribuzione, prestazioni in genere migliori rispetto a quelle dei codec basati sullo standard ETS 300 174 (Elettronica e Telecomunicazioni, n.1 aprile 2003, § 2).

Per applicazioni contributo occorre però soddisfare ulteriori requisiti, rispetto a quelli relativi alla qualità percepita:

- per consentire operazioni di post-produzione, anche se non particolarmente complesse, è opportuno che la definizione delle componenti di colore sia almeno quella del 4:2:2
- per garantire un'elevata qualità deve essere possibile utilizzare fattori di compressione più bassi, e quindi superare il limite in ter-

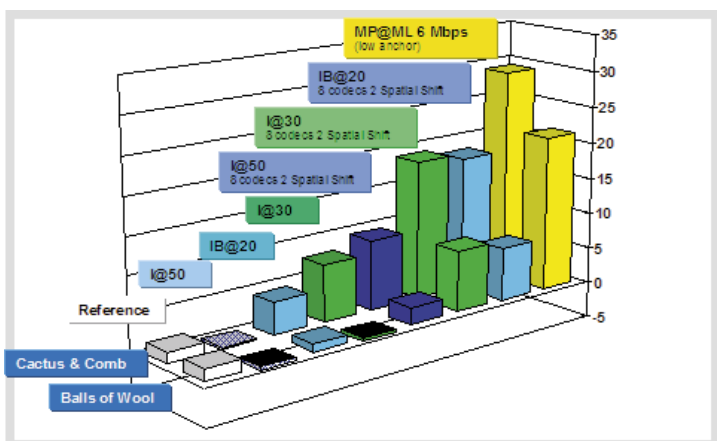


Fig. 4 - Il Centro Ricerche ha contribuito alla definizione dei requisiti per i sistemi di videoregistrazione professionale. Ad esempio, in figura sono riportati i risultati di prove soggettive effettuate per valutare le prestazioni nel caso di post-produzione, in particolare la realizzazione di un chroma-key utilizzando come foreground l'immagine co-decodificata. Sono messe a confronto varie alternative, caratterizzate da diversa struttura di GOP e bit-rate. Sono evidenti i vantaggi dell'uso di un sistema intra a 50 Mbit/s.

mini di bit-rate che, per il MP@ML, è di 15 Mbit/s

- per certi usi, il ritardo dovuto alla co-decodifica (tabella 2) può essere eccessivo. Si consideri, ad esempio, un programma televisivo in cui due persone, ad esempio intervistatore ed intervistato, si trovano in due luoghi diversi ed il collegamento è ottenuto mediante un sistema basato su MPEG: il ritardo che interviene tra domande e risposte può essere fastidioso (si noti che da questo punto di vista il sistema ETS 300 174 risponde a requisiti più stringenti e consente di limitare i ritardi a meno di 100 ms).

Nel 1996 fu quindi definito il profilo 422P, caratterizzato dalla possibilità di codificare segnali 4:2:2, con un bit-rate massimo fino a 50 Mbit/s, la massima precisione per la rappresentazione

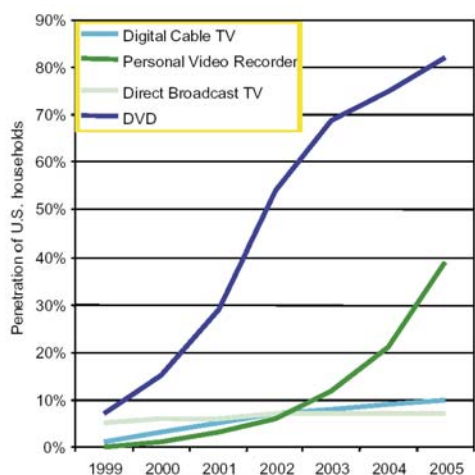


Fig. 5 - Crescita della penetrazione nel mercato USA nel periodo 1999-2005 delle piattaforme basate su MPEG (fonte: The Yankee Group).

dei coefficienti DCT (fino a 11 bit) è superiore a quella prevista per il MP (fino a 10 bit): è quindi possibile comprimere il video con un sistema quasi *loss/less*.

Il sistema IMX™ della Sony adotta il 422P@ML per la registrazione videomagnetica. Questo sistema registra il segnale televisivo a 50 Mbit/s. In questo caso si utilizza solamente la codifica intra (GOP di lunghezza 1, sono tutte *I-picture*) in modo da facilitare l'editing con la precisione al singolo quadro con prestazioni analoghe a quelle dei sistemi di videoregistrazione digitale non compressi (figura 4).

5. Il mercato

Dall'introduzione dello standard, MPEG-2 è stato il motore che ha permesso la transizione dalla distribuzione sotto forma analogica a quella numerica del segnale video, consentendo la nascita dell'attuale mercato di massa delle piattaforme digitali video. E' stato soprattutto il DVD a presentare la più rapida crescita, come illustra la figura 5, che si riferisce al mercato USA.

Riferimenti bibliografici

La MPEG Home Page è disponibile alla URL www.chiariglione.org/mpeg/index.htm in cui sono disponibili anche i link al sito www.iso.org dove sono acquistabili le versioni pdf delle varie parti dello standard, tra cui la parte video ISO/IEC13812-2-2000.

Un tutorial su MPEG edito da Laurent Boch, del Centro Ricerche Rai nel luglio 1996 è disponibile alla URL viswiz.imk.fraunhofer.de/DVP/Public/deliv/deliv.211/mpeg/mpeghome.htm.

Un tutorial di Mike Knee della Snell e Wilcox, più recente (marzo 2002), è alla URL www.broadcaspapers.com/sigdis/Snell&WilcoxMPEGVideo01.htm.

Whitepaper sono presenti sui siti di alcune industrie che producono apparati professionali basati su MPEG, ad esempio su www.harmonicinc.com.