

# IL SISTEMA DI COMPRESSIONE AUDIO/VIDEO MPEG-2

Francesco Cisternino

Tesina realizzata per il Corso sulle nuove tecnologie di comunicazione all'interno del Corso di Teorie e Tecniche del Linguaggio Radiotelevisivo

Prof. P. L. Capucci, a.a. 1999/2000

*Che cos'è MPEG*

MPEG è l'acronimo di **Moving Picture Experts Group**, vale a dire un gruppo di lavoro dell'ISO/IEC incaricato di creare degli standard internazionali per la compressione, la decompressione, il processing e la codificazione di immagini in movimento, audio e relativa combinazione, utilizzabili su media del tipo CD-ROM, CD-i, CD Multi-sessione e Video CD. Un incontro fra l'ISO (International Standards Organisation) e l'IEC (International Electrotechnical Commission) nel 1992 ebbe come risultato lo standard di codifica audio/video denominato MPEG-1. Nel novembre di due anni dopo un meeting di cinque giorni a Singapore dette inizio ad un nuovo standard, MPEG-2, che aveva come principale novità l'introduzione di un encoder audio dotato di un modello psicoacustico multicanale. La ricerca è andata avanti, e i sistemi finora ideati e realizzati, o in fase di realizzazione, sono i seguenti:

- MPEG-1, lo standard approvato per lo scarico e il recupero di immagini animate su *storage media*;
- MPEG-2, lo standard per la televisione digitale;
- MPEG-4 versione 1 e 2, lo standard per applicazioni multimediali.

Sono attualmente in fase di realizzazione:

- MPEG-4 nelle versioni 3 e 4;
- MPEG-7, lo standard per la ricerca di informazioni multimediali (multimedia information search), filtering, management e processing;

E' infine da poco iniziato il lavoro per la realizzazione di MPEG-21, il *multimedia framework*.

Gli oltre 300 esperti di questo gruppo, provenienti da venti differenti paesi del mondo, si riuniscono periodicamente circa tre volte l'anno; in occasione di questi meeting vengono messe a

confronto le possibili innovazioni e realizzazioni dei vari progetti in corso. Per la cronaca, il prossimo incontro è previsto per il marzo di quest'anno, dal 20 al 24, a Noordwijkerhout, in Olanda.

## **MPEG-2**

Si tratta di un sistema strutturalmente suddiviso in nove parti: le prime tre hanno raggiunto lo status di Standard Internazionale, altre parti sono in fase di completamento; una di esse, la 8, è stata eliminata.

La **parte 1** si occupa di combinare uno o più *elementary stream* (flussi elementari) video e audio così come altri dati in flussi singoli o molteplici, i quali siano adatti ad essere scaricati o trasmessi.

Per far ciò si attivano due Layer (strati): il primo, denominato PES (*Packetized Elementary Stream*), viene utilizzato per assicurare una stretta sincronizzazione fra audio e video. Il secondo, strettamente dipendente dal mezzo di comunicazione destinato, è composto da due ulteriori fasi: il Program Stream (PS), il quale proviene dalla combinazione di uno o più PES, dalla base temporale comune, uniti in un unico flusso, e il Transport Stream, il quale combina a sua volta uno o più PES con una o più basi temporali indipendenti in un unico flusso. Flussi elementari condividenti una base temporale comune formano un programma. Il TS è stato progettato per utilizzi in ambienti (*environment*) nei quali è possibile che si verificano degli errori, come lo scarico o la trasmissione con media rumorosi o nei quali, in generale, possano avvenire delle dispersioni.

La **parte 2** si basa sulle potenti possibilità di compressione dello standard MPEG-1 per offrire un ampio raggio di strumenti per la codifica. Ne esistono differenti combinazioni, a seconda delle varie funzionalità. Uno di questi, in particolare, ha dato ottimi risultati per ciò che concerne le immagini aventi una risoluzione del colore di tipo 4:2:2 ed un *bitrate* (la velocità cui viaggiano i bit) più alto. Un altro, denominato MVP (*MultiView Profile*), è una sorta di applicazione combinata di strumenti già esistenti in MPEG-2 che permette di codificare con successo due sequenze video registrate con telecamere diverse poste ad angolazioni leggermente differenti ma puntate sullo stesso soggetto.

La **parte 3** è quella che riguarda l'audio: un paio di precisazioni introduttive saranno indispensabili. I codificatori MPEG, realizzati per qualsiasi tipo di segnale, sia esso musicale o vocale (il c.d. *generic audio*), sono di tipo percettuale e non del tipo chiamato *waveform*. In un codificatore di tipo percettuale il codice non ha l'obiettivo di mantenere il segnale identico dalla fase di codifica a quella di decodifica, bensì mira ad assicurare che il segnale d'arrivo appaia identico ad un orecchio umano.

L'orecchio umano è un dispositivo di ricezione acustica decisamente non perfetto: sebbene la gamma di frequenze percepibili possa variare teoricamente fra i 20 Hz e i 20 kHz, accade in realtà che vi sia una soglia (la c.d. *threshold of hearing*) al di sotto della quale l'orecchio non percepisce alcun segnale. Intervengono vari fattori, soprattutto le caratteristiche d'ambiente e quelle individuali. Un esempio facile: una comune conversazione in una stanza è perfettamente udibile in condizioni d'ambiente normali; ma se in prossimità del luogo passa un aereo, è impossibile capire ciò che viene detto per via delle distorsioni introdotte alle soglie d'ascolto dei presenti. Quando l'aereo si allontana, la soglia d'ascolto ritorna normale. I suoni che diventano inascoltabili per via delle modificazioni dinamiche della soglia d'ascolto vengono definiti come "suoni mascherati". L'effetto è comune e di particolare interesse per ciò che concerne la musica. Un'orchestra che suona un *fortissimo* renderà certi strumenti inascoltabili su altri per il nostro orecchio. Quando la musica viene registrata tutte le frequenze tendono ai medi perché la risposta del dispositivo di registrazione è piatta, vale a dire non adattabile dinamicamente: il CD che andremo a suonare nel lettore non ci permetterà, così, di ascoltare certi strumenti, così come se non ci fossero. Per utilizzare al massimo un mezzo di registrazione come il CD quelle parti che contengono informazioni non percepibili dal nostro orecchio devono scomparire per far posto ad informazioni "comprensibili". In questo senso è possibile tagliare una parte considerevole delle informazioni senza avere per questo ridotto la qualità dell'audio.

Ritornando al nostro codificatore percettuale, il principale effetto psicoacustico che esso utilizza si chiama *auditory masking* (che si può tradurre come "mascherazione uditiva"): le parti del segnale che vengono "mascherate" sono chiamate "irrilevanti", in opposizione a quelle, rimosse da un codificatore alla fonte, che vengono denominate invece "ridondanti". Per rimuovere queste ultime informazioni il codificatore contiene in sé un modello che analizza costantemente il segnale e ne determina la c.d. curva di mascheramento (*masking curve*), vale a dire la soglia al di sotto della quale certi suoni o rumori non vengono percepiti dall'orecchio.

Il decodificatore non richiede, invece, un modello psicoacustico e una procedura di distribuzione dei bit. Il suo unico scopo è di ricostruire un segnale audio dai componenti spettrali codificati (*coded spectral components*) e *side information* associate.

Sia nell'MPEG-1 che nell'MPEG-2 troviamo tre differenti Layer, i quali rappresentano una famiglia di algoritmi codificanti. Tali Layer vengono indicati con dei numeri romani; ciascuno di essi ha i suoi compiti. Fondamentalmente la complessità del codificatore e del decodificatore, il ritardo fra i due momenti e l'efficienza di codifica aumenta andando rispettivamente dallo strato I attraverso il II fino al III. Il Layer I è quello relativamente meno complesso ed è specificamente adatto per applicazioni nelle quali anche la complessità del codificatore gioca un ruolo importante. E' quello utilizzato per il Solid State Audio o per il DCC, ora

non più in commercio. Il Layer II garantisce una compressione più potente e viene utilizzato in applicazioni semi-professionali, del tipo trasmissioni radio/televisive, telecomunicazioni e multimedia; è adatto poi per applicazioni del tipo "one to many", i.e. un codificatore che serve più decodificatori. Paragonato al Layer I, il II riesce a rimuovere una buona parte del segnale c.d. ridondante e ad applicare la soglia del modello psicoacustico più efficientemente. Il Layer III è più complesso ancora ed è diretto verso applicazioni con bitrate più bassi.

Va notato che mentre MPEG-1 eseguiva una codifica audio mono e due canali stereo con le frequenze di campionamento normalmente utilizzate per le applicazioni di alta qualità (48, 44.1 e 32 kHz), MPEG-2 ha incorporato un sistema più complesso in due parti. La prima consiste in un'estensione verso frequenze di campionamento più basse, fornendo una qualità del suono decisamente migliore anche a bitrate molto bassi (sotto i 64 kbit/s per un canale mono). La seconda è l'estensione del *multichannel sound*. Ma di questo se ne parlerà nella parte 7.

Ricapitolando le applicazioni, in linea generale sono quattro i campi che possiamo riconoscere: il *broadcasting*, lo *storage* (i.e. lo scarico di informazioni), l'ambito *multimedia* in generale e le telecomunicazioni. Alcune fra le più importanti sono riportate di seguito:

- DVD (Digital Video Disc, disco ottico ad alta densità sul quale sono registrati immagini e suoni di qualità elevata);
- Trasmissioni televisive digitali via satellite e via cavo (e.g. DVB, USSB, DirecTV, EchoStar);
- Digital Audio Broadcasting (trasmissione audio digitale, e.g. ADR, US-Digital Radio, World Space Radio);
- Trasmissioni radiofoniche su Internet.

Le **parti 4 e 5** precisano il modo in cui possano essere progettati i test che devono verificare se i bitstream e i decodificatori soddisfano le esigenze già indicate nelle parti precedenti. Questi test possono essere utilizzati:

- Dai costruttori dei codificatori e relativi clienti per verificare se il codificatore produce bitstream validi;
- Dai costruttori dei decodificatori e relativi clienti per verificare se il decodificatore fornisca effettivamente le caratteristiche dichiarate;
- Da applicazioni volte a verificare se le caratteristiche di un dato bitstream soddisfano le richieste; controllare, e.g., se la dimensione delle immagini codificate non supera il valore massimo concesso per quella determinata applicazione.

La **parte 5** dà una completo software di attuazione (full software implementation) delle prime tre parti dello standard MPEG-2.

La **parte 6** è il c.d. DSM-CC (Digital Storage Media Command and Control), che consiste in un ampio kit di accessori che organizzano le funzioni di controllo e le operazioni in particolare per gestire MPEG-1 e 2. Nel modello DSM-CC viene creato un paradigma del tipo Cliente- Network soggiacente- Server per la selezione, l'accesso e il controllo di fonti video e audio; il network è trasportato dall'*MPEG-2 transport multiplex* o in modo indipendente, se necessario.

DSM-CC può essere usato per controllare la ricezione video, per fornire funzioni normalmente reperibili su di un videoregistratore (fast forward, rewind, pausa ecc.) o per scopi di altro tipo, tra i quali il trasporto di pacchetti di dati. Non va dimenticato che può inoltre operare unitamente a *packet network* della prossima generazione, come RSVP, RTSP, RTP e SCP.

La **parte 7** concerne l'audio c.d. multicanale, che risulta essere il miglioramento in senso evolutivo dei vecchi sistemi. Sebbene i primi esperimenti sull'audio multicanale risalgano ai primi anni trenta, si è dovuti passare dallo stereo, nel quale venivano utilizzati i due canali Sinistro e Destro, e gradualmente attraverso le versioni del *surround* prima ad un singolo canale aggiunto e poi a due canali indipendenti per arrivare al 1996 quando, nel corso del CeBit, si è realizzato un test per l'Home Theatre con attuazioni differenti. Il sistema surround utilizzato da MPEG-2 era basato su di un set-up di cinque canali, ai quali ne veniva poi aggiunto un sesto, denominato LFE (*Low Frequency Enhancement*), consistente sostanzialmente in una specie di subwoofer dedicato alle frequenze più gravi. Gli studi e le applicazioni più recenti hanno poi raggiunto l'obiettivo di creare un sistema a ben sette canali, perfettamente compatibile con il precedente. La compatibilità, obiettivo primario per i progettisti, è garantita anche per il Dolby Pro Logic e per altri formati con minor numero di canali. E' stato possibile realizzare risultati di questo livello grazie all'utilizzo di algoritmi particolarmente sofisticati e al Variable Bit Rate (velocità variabile cui viaggiano i bit). Urge qualche spiegazione sull'importanza della velocità dei bit: nel caso di un codificatore CBR (*Constant Bit Rate*), la velocità d'uscita è costante, e.g. 4 Mb/s, ma ciò non significa che il numero di bit per sequenza sia costante. Ponendo il caso di un concerto registrato su DVD, noteremo che vi sono dei *fortissimo*, delle pause, e così via. I bit vengono distribuiti secondo la complessità del frammento, ma la velocità di bit al secondo deve rimanere identica: questo può portare ad una codifica inefficiente. L'utilizzo, invece, di un flusso variabile risolve il problema, garantendo delle performance ben più elevate.

La **parte 8** era originariamente stata progettata per codifiche video nel caso in cui gli input samples superassero i 10 bit. Le ricerche su questa parte sono cessate per scarso interesse.

La **parte 9** consiste nella c.d. RTI (Real Time Interface), la quale riveste il ruolo di adattatore nei confronti di qualsiasi network DSM sul quale viaggia il Transport Stream.

#### Fonti consultate

ISO ed MPEG in generale:

<http://www.iso.ch/>

[http://www.sv.philips.com/mpeg/mpeg\\_info/html/mpeg\\_general.html](http://www.sv.philips.com/mpeg/mpeg_info/html/mpeg_general.html)

<http://drogo.cselt.it/mpeg/>

Su DSM-CC:

[http://www.erg.abdn.ac.uk/public\\_html/research/future-net/digital-video/ima.../dsm-cc.htm](http://www.erg.abdn.ac.uk/public_html/research/future-net/digital-video/ima.../dsm-cc.htm)

<http://www.cselt.it/leonardo/mpeg/documents/dsmcc/dsmcc.htm>

Audio:

<http://www.philips.com/sv/newtech/mpeg>

<http://www.mpeg-empowered.com>

Per consigli, chiarimenti e altro:

[f\\_cisternino@hotmail.com](mailto:f_cisternino@hotmail.com)

Francesco Cisternino

N° matr. 1240 - 35057

Dams ind. Musica

#### Ringraziamenti:

Questo lavoro non sarebbe stato possibile senza la collaborazione del dott. Massimiliano Neri, che ha fornito una parte dei materiali, e dell'ing. Leonardo Chiariglione, membro di MPEG, i cui chiarimenti durante la stesura sono stati indispensabili.