

# APUNTES TECNOLOGÍA DIGITAL

## 4º. Comunicación Audiovisual

Profesor: Raúl Hernández

**ACTUALIZADOS AL 06-5-2007**

### TEMARIO

#### I.- LA SEÑAL DIGITAL

Analógico y digital

- Definición de analógico y digital
- Digitalización
- El muestreo
- La cuantificación
- Tipos de señales digitales de video

LA NORMA 4:2:2

- Elección de la frecuencia de muestreo
- Familia de normas: 4:4:4, 4x4, 4:1:1, 4:2:0. 2:1:1

LA SEÑAL DIGITAL COMPUESTA

- Introducción
- Elección de la frecuencia de muestreo
- Muestras por línea
- Fase de muestreo
- Rango de amplitudes
- Numeración de las muestras
- Referencia temporal e identificación
- El campo digital
- Características de los interfaces

COMPRESIÓN EN VIDEO

- Teoría de la información
- Tipos de compresión
- Entropía Vs Redundancia
- Redundancia y entropía en las imágenes de televisión
- Redundancia estadística
- Redundancia perceptual
- Entropía

- Técnicas de reducción de datos
- Técnicas de compresión sin pérdidas
- Codificación de secuencias
- Codificación de longitud variable
- Introducción a la DCT
- Técnicas de compresión con pérdidas
- Codificación diferencial (DPCM)
- Compresión temporal
- Redundancia temporal
- Compensación de movimiento

## II.- MPEG: VIDEO DIGITAL

### LOS SISTEMAS DE COMPRESIÓN

- El sistema JPEG
- Aplicación de la codificación Huffman
- Características del JPEG
  
- El M-JPEG
  
- Introducción al estándar MPEG
- Introducción a la compresión de video
- Compresión
  - Codificación Intra o espacial
  - Codificación Inter o temporal
  - Codificación bidireccional
- Proyectos MPEG
  - MPEG-1
  - MPEG-2
  - MPEG-3
  - MPEG-4
  - MPEG-7
  - MPEG-21

### MPEG-1

- Compresión de audio
- Compresión de video
- Decodificación

### MPEG-2

- Perfiles y niveles MPEG-2
- Medios de predicción MPEG-2 (imágenes entrelazadas/progresivas)
- La estructura DCT-frame
- La estructura DCT-field

### MPEG-4

### MPEG 7

MPEG-21

LOS SISTEMAS DV

### III.- AUDIO DIGITAL

El audio digital.

- Características del sonido.
- Digitalización del sonido.
- Frecuencias de muestreo y cuantificación.
- Interfaces de audio digital. AES/EBU y MADI.

Compresión de audio digital

- Redundancia en sonido
- Enmascaramiento: Frecuencial, temporal. Irrelevancia estereofónica.
- Ventajas y desventajas de la compresión
- Introducción al proceso de compresión

Sistemas de compresión de audio digital

- MPEG
- AAC
- AC-3

### **IV.- SISTEMAS DIGITALES DE EMISIÓN**

- El estándar ATSC.
- El estándar DVB-T.
- El estándar ISDB-T.

### **TEMAS DE INVESTIGACIÓN**

- CÁMARAS Y TOMA DE SONIDO DIGITALES
- OPERACIÓN DIGITAL DE LA POSTPRODUCCIÓN
- CINE DIGITAL
- VIDEO Y AUDIO DIGITAL A TRAVÉS DE PLATAFORMAS DIGITALES
- VIDEO Y AUDIO DIGITAL A TRAVÉS DE REDES
- DESARROLLO DE LA TDT EN EL MUNDO OCCIDENTAL
- PRODUCCION EN LA TELEVISION DIGITAL INTERACTIVA
- PROGRAMACION EN LA TELEVISION DIGITAL

## **PRINCIPIOS FISIOLÓGICOS DE LA TELEVISIÓN**

- Los principios fisiológicos y perceptivos que permiten la ilusión de movimiento - fenómeno en el que se basa el cine, el vídeo y la televisión- son la persistencia retiniana, (aunque las teorías modernas lo descartan); el efecto phi o de ilusión de movimiento; la fusión crítica del parpadeo y el enmascaramiento visual. A estos se añaden en la televisión la integración de líneas, por el cual el ojo no aprecia la división de una imagen en líneas por encima de cierta relación de aspecto, relación que depende además de la distancia a que se encuentre el espectador de la pantalla.
- La persistencia retiniana durante mucho tiempo se supuso la responsable fisiológica de la posibilidad de lo audiovisual, aunque actualmente se discute. Responde al hecho de que las células sensitivas ópticas de la retina (conos y bastoncillos) siguen respondiendo por un tiempo (1/20 de segundo) a un estímulo visual después de que este estímulo haya cesado. Así, existe un efecto de superposición de imágenes en nuestra retina. Los críticos con la persistencia retiniana suponen que ésta haría que lo audiovisual produciría en nuestra retina un amasijo borroso de imágenes superpuestas.
- El efecto phi o ilusión de movimiento aparente, implica que vemos, como si correspondiera a un mismo objeto en movimiento, una serie de imágenes de objetos cuyas posiciones sucesivas presenten una cercanía, si ésta se da de forma sucesiva en el tiempo y si las sucesivas diferencias de posición tiene una separación determinada en el tiempo. Por encima de esa separación temporal, el ojo sólo vería dos imágenes distintas. Por debajo de esa separación, llegarían a confundirse las dos imágenes.
- Así, percibimos como un movimiento lineal lo que es una serie de imágenes separadas en cercanía espacial y temporal (efecto beta) y como aproximaciones y alejamientos si las imágenes no se desplazan en el espacio, pero guardan diferentes relación de tamaño entre ellas (efecto alfa).
- El efecto phi parece ser en las teorías actuales de la percepción el principio fisiológico que justifique mejor la imagen audiovisual.

- La fusión crítica del parpadeo, supone que dada una serie de destellos luminosos periódicos, estos tienden a ser percibidos, por encima de cierta frecuencia crítica, como un foco de luz continuo. En los albores del cine se trabajaba a 18 f.p.s, con lo que producía cierto pantalleo en la imagen proyectada. Para evitarlo el centelleo, en primer lugar, se hizo que el paso de película en el proyector aumentara hasta 24 f.p.s. Y además supuso que al elevarse la intensidad de la lámpara, y necesitarse una mayor frecuencia crítica, apareciese la obturación de la imagen proyectada en dos o tres destellos por fotograma.
- El enmascaramiento visual supone que entre dos estímulos luminosos cercanos la percepción del segundo puede afectar a la del primero. Se supone que el enmascaramiento visual es el que elimina el efecto de superposición borrosa de imágenes de la persistencia retiniana.
- Y a estos principios, comunes al cine y a la televisión, hay que sumar uno específico de este último medio, que trabaja descomponiendo la imagen en líneas consecutivas. La integración de líneas supone que el ojo no aprecia la división de una imagen en líneas consecutivas por encima de cierta relación de aspecto; así mismo, ésta dependerá de la distancia a que se encuentre el ojo de la pantalla. La agudeza visual del ojo para distinguir dos puntos colaterales es de 1 minuto de arco.

## **DESCOMPOSICIÓN Y FORMACIÓN DE LA IMAGEN EN TELEVISIÓN**

### análisis de la imagen en TV analógica

- Tanto en cine como en vídeo/TV la imagen en movimiento se transforma en una serie de imágenes fijas y discontinuas que luego se exponen en orden sucesivo y a una velocidad tal que está por debajo de esos umbrales perceptivos. El fotograma o frame es cada una de esas imágenes fijas en que se han dividido las imágenes en movimiento.
- En vídeo/TV analógica, se descompone la imagen en líneas según el barrido horizontal de un haz de electrones por el target (una superficie donde la óptica recoge y forma la imagen, y que al estar recubierta por un material sensible a la luz, convierte la información luminosa de cada punto en una información electrónica.)
- Cada punto de la imagen se traduce en un valor de voltaje según su luminosidad, correspondiendo 0 voltio al negro y 0,7 voltio a la mayor luminosidad. Para las señales de sincronismo se reserva el valor de  $-0,3$  V. El valor de la señal de video pico a pico es de 1 voltio.
- Pero así la señal que tendríamos correspondería a la luminancia, que traduciría intensidad luminosa por un valor de voltaje.
- El efecto de la televisión en color se logra descomponiendo en un prisma la imagen en color en sus componentes: rojo, verde y azul (R/G/B) y sometiendo simultáneamente a cada una de las componentes, RGB, a la exploración de un barrido electrónico que traduciría cada dato de intensidad luminosa correspondiente a cada componente a un valor independiente de voltaje. Para cada punto, en un momento concreto, habría un valor de voltaje determinado para R, otro independiente de éste para G, y otro para B. La suma de los tres valores nos daría la cantidad de intensidad luminosa total del punto en cuestión.
- Una vez barrida una línea, el haz de electrones lee la línea siguiente, y así sucesivamente hasta cubrir toda la pantalla. En cada línea, junto a la información de correspondiente a la luminancia en cada punto de luz en la línea, hay unos impulsos

fijos (que se producen a intervalos fijos de tiempo) que dan la información por una parte de cambio de línea (impulso de sincronismo horizontal, que hace que el haz vuelva al principio de la siguiente línea) y por otra de cambio de pantalla (impulso de sincronismo vertical, que hace que el haz vuelva al principio de la primera línea de lectura para captar un nuevo frame.)

- En PAL el periodo de una línea es de 64 nanosegundos ( $64 \cdot 10^{-9}$  segundos). A la parte de datos de la línea se le llama línea activa, y en esa parte, en la que se traducen los valores de intensidad luminosa a tensión eléctrica, los valores de voltaje van de 0 (negro) a 0'7 voltios (blanco). El periodo de línea activa es de 52 nanosegundos.
- A la línea activa le sigue el periodo de borrado de línea, cuyo valor es de 12 nanosegundos. En éste encontramos:
  - El pórtilo anterior (1,5 nanosegundos), a 0 voltios.
  - El impulso de sincronismo horizontal (5,8 nanosegundos), una señal cuadrada de -0,3 voltios.
  - El pórtilo posterior (5'8 nanosegundos), a 0 voltios. En este periodo se incluye el *Bursa*, que referencia en el sistema la subportadora de color en frecuencia y fase.
- En vídeo/TV, debido al barrido del haz de electrones, para evitar un efecto de parpadeo se produce el llamado efecto de entrelazamiento de líneas dividiéndose el frame en dos campos, uno compuesto por las líneas impares (campo impar) y otro por las líneas pares (campo par). La frecuencia real de la imagen de vídeo/TV en PAL es de 50 Hertzios y de 60 Hz. en NTSC.
- La integración de líneas supone que el ojo no aprecia la división de una imagen en líneas consecutivas por encima de cierta relación de aspecto; así mismo, ésta dependerá de la distancia a que se encuentre el ojo de la pantalla. La agudeza visual del ojo para distinguir dos puntos colaterales es de 1 minuto de arco.
- En PAL analógico cada campo dura 20 ms. y tiene 312,5 líneas. De este periodo, el correspondiente a campo activo es de 18,4 ms. Esto es, 287,5 líneas.

- En PAL el borrado de línea tiene 1,6 ms. Ocupa entonces 25 líneas, el tiempo que se tarda en pasar del final de campo al comienzo de un nuevo campo. Dentro de este periodo encontramos el sincronismo vertical (que ocupa 2,5 líneas, en total 160 nanosegundos, y con un valor de -0,3 voltios) flanqueado por un tren de impulsos de preigualación y otro de postigualación. En total, ocupa 7,5 líneas, 480 nanosegundos. El resto del periodo de borrado vertical es empleado para albergar otras informaciones (señales de test, teletexto, código de tiempos.)
- En el resultado final de este proceso, la imagen se transforma en una onda electromagnética.
- La conversión de la imagen en una onda electromagnética permite su grabación en formato magnético; la emisión, transmisión y recepción de la imagen a través del vacío; el montaje y conmutación elaborada de diferentes señales así como la realización electrónica (siempre que las señales estén en fase, es decir, que coincidan sus impulsos de sincronismo.)

■

#### formatos de televisión en color

- En los formatos PAL y SECAM la imagen consta de 25 frames por segundo y 625 líneas por frame. En el NTSC de 30 frames por segundo y 525 líneas por frame.
- En PAL hay que considerar que hay 25 líneas por campo que se emplean en los impulsos de sincronismo vertical y borrado, por lo que en un frame hay en PAL analógico 575 líneas activas, correspondientes a 287,5 líneas por campo.
- PAL, SECAM y NTSC son todos ellos sistemas de televisión en color compatibles con la televisión en blanco y negro. Para ello emiten no ya una información de señal por componentes separadas (R, G, B) sino una señal compuesta por la combinación de la luminancia Y junto a dos señales que dan la información de croma (dos señales en las que se cifra la información correspondiente a dos de las componentes, habitualmente correspondientes al azul, Cb o U; y al rojo Cr o V.

### luminancia y crominancia

- El valor **Y**, la luminancia, es la cantidad de luz en cada punto de la imagen. Siendo por lo tanto suma de la intensidad de los componentes

$$Y = R + G + B$$

Expresado en voltios

$$E(Y) = 0,299E(R) + 0,589E(G) + 0,114E(B)$$

- Las señales **U** y **V** se corresponden con las componentes de azul y rojo, respectivamente.

$$U = B - Y$$

$$E(U) = 0,493(E(B) - E(Y))$$

$$V = R - Y$$

$$E(V) = +/- 0,877(E(R) - E(Y))$$

- Señal en componentes es la que trabaja con la descomposición de la señal en sus componentes Rojo, Verde y Azul (R, G, B).
- La señal Y/C o S-Video trabaja con dos señales separadas:
  - Y o luminancia.
  - C o crominancia (que codifica y multiplexa las componentes U y V en una única señal).

Este sistema de señal de video se utilizó en el S-VHS y se utiliza como salida de señal de video en los ordenadores.

- En la señal en compuesta se multiplexan la señal de luminancia Y con dos señales de crominancia para obtener una única señal de video.

En PAL se multiplexa con la señal de luminancia las de componentes  $U=R-Y$  /  $V=B-Y$ , con lo que fácilmente, en la recepción, una vez decodificada la señal compuesta, se descompondrán de nuevo las señales de R, G, B.

- La señal de luminancia da lugar, espectralmente, a una señal discreta y doblemente periódica, con respecto a la frecuencia de línea y la frecuencia de cuadro, dando lugar a unos espacios entre los armónicos que permitirán acomodar ahí, en un entrelazado en frecuencia, la información de crominancia. El ancho de banda de la señal de luminancia es de unos 5,5 MHz.
- La crominancia nos indica la cantidad de color en cada punto de la imagen. Se representa por el tono del color (que se puede descomponer por la proporción de los colores básicos R, G y B) y la saturación (cantidad de color puro frente a la mezcla con el blanco en cada punto.)
- La crominancia es codificada por las señales U y V, que se modulan en cuadratura de fase con una subportadora de color.
- La suma de esas dos señales de crominancia moduladas en fase dan lugar a un vector de crominancia, de cierta amplitud y cierta fase, que dan la información relativa a cada color.
- Una variación en la fase del vector de crominancia de sólo  $5^\circ$  daría lugar a errores considerables en la apreciación de ese color. El sistema PAL intenta corregir esto modulando la señal V a cada línea alterna con un desfase de  $180^\circ$ . Si existe un error en la transmisión de esa fase en una línea, la alternancia de fase asegura que el color en cada línea tendrá una aberración discordante que será corregida en el ojo gracias a la integración espacial de las líneas consecutivas.
- Como la información de luminancia toma valores discretos, múltiplos de la frecuencia de línea y la frecuencia de cuadro, en el espectro de frecuencias de la luminancia quedan unos “huecos” que podemos cubrir con la crominancia. A eso le llamamos imbricado frecuencial.
- La frecuencia de la subportadora de color es múltiplo de la frecuencia de línea, y luego se desfasa con respecto a ésta para permitir el imbricado frecuencial de la señal de la crominancia dentro del espectro de la de luminancia. Para poder dar lugar al

imbricado frecuencial, se utiliza una frecuencia de portadora múltiplo de la frecuencia de línea y a ésta se le suma un cuarto de la frecuencia de línea, y además se le suma una frecuencia de cuadro.

- La subportadora de color tiene en PAL una frecuencia de 4,43 MHz. El ancho de banda de la crominancia es bastante más reducido que el de la luminancia.
- La señal compuesta tiene una serie de inconvenientes en cuanto a la degeneración que supone su codificación y decodificación, proceso en los que el ruido afecta a la pureza de la señal.
- Además, la diferencia entre la frecuencia de línea y la frecuencia de la salva de color da lugar en el sistema NSTC a una secuencia de 4 campos para crear señales con la misma fase. La alternancia de fase del sistema PAL, al afectar a la fase del Burst o salva de color inserto en la señal de sincronismo vertical, crea una secuencia de 8 campos en que la señal se desfasa. Esto provoca problemas al editar en algunos frames, si no están en fase con respecto a esta secuencia de 4 campos en NSTC u ocho en PAL.

## **TELEVISIÓN DIGITAL**

### digitalización de la señal de video

- La televisión digital supone la conversión de la señal analógica en una señal cuantificada tanto en el tiempo como en las amplitudes. Se realiza un muestreo cada cierto tiempo del valor que tiene la señal, y ésta a su vez está cuantificada en los valores que puede tomar. Sólo puede tomar ciertos valores discretos, frente a la televisión analógica en que en cada instante la señal puede tomar un valor cualquiera, ya que la señal es continua.
- La televisión digital, al estar cuantificada y convertir los valores continuos de la señal en valores discretos, permite aumentar la calidad de la imagen al separar la señal del

ruido, no generar errores acumulativos con las sucesivas generaciones, poder regenerar la información y corregir errores por aproximación. Junto a esto, la capacidad de compresión que tiene permite transmitir más información por el mismo canal. Y a esto hay que unir, al manipular datos discretos, una mayor posibilidad de postproducción.

- Los formatos más extendidos de captura y almacenamiento en cinta son el DVPRO 25 y 50 (Panasonic), el betacam DIGITAL, SX y IMX (SONY) y el DIGITAL S (JVC).
- Los formatos de emisión digital que se imponen son el DVB (Europa), el ATSC (EE.UU.) y el ISDB (Japón).
- La digitalización de la señal de video comprende los siguientes procesos: muestreo, codificación o cuantificación, compresión y archivo/transmisión.
- Muestreo: se toman muestras de forma periódica de la información. El proceso consiste en multiplicar la amplitud de la señal de video analógica por la señal de impulsos de muestreo, y la resultante es una señal modulada en amplitud:

$$PAM = A * M$$

- La cuantificación reduce el rango de datos que puede tomar una señal, de forma que trabajemos con un código binario. En éste las cifras de un número sólo pueden tomar dos valores:

$$0 \quad 1$$

- Estamos acostumbrados al código decimal, en el que se expresa una cantidad en términos de las potencias de 10 en que se puede descomponer. Por ejemplo:  $21 = 2 \times 10 + 1 = 2 \cdot 10^{**1} + 1 \cdot 10^{**0}$ .

$$546279 = 5 \cdot 10^{**5} + 4 \cdot 10^{**4} + 6 \cdot 10^{**3} + 2 \cdot 10^{**2} + 7 \cdot 10^{**1} + 9 \cdot 10^{**0}$$

En el código decimal, las cifras de un número pueden tomar diez valores diferentes:

$$0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9$$

- Pero realmente un número es independiente del código en que se expresa, aunque varíe la cifra que lo expresa, su cantidad no cambia.

- Por ejemplo, en ingeniería, en que se trabaja con números grandes, se utiliza el sistema hexadecimal, con 16 cifras.

0 1 2 3 4 5 6 7 8 9 A B C D E F

Por lo que 350 en código hexadecimal corresponde a  $1 \cdot 16^{**2} + 5 \cdot 16^{**1} + E \cdot 16^{**0}$   
= 15E

El código binario utiliza sólo dos valores, 0 y 1. Da lugar a cifras más largas de las que nos tiene acostumbrado el código decimal.

Por ejemplo, las 1001 noches de los cuentos de Sherezada, en código binario, realmente corresponde a  $1 \cdot 2^{**3} + 0 \cdot 2^{**2} + 0 \cdot 2^{**1} + 1 \cdot 2^{**0}$ , es decir, que los casi tres años de las historias árabes se reducen en código digital a ¡9 noches!

- El código binario permite expresar fácilmente un número en electrónica digital, ya que al jugar con 0 y 1 podemos asociar el 0 a la posición de apagado, o de voltaje nulo, y el 1 al de encendido.
- Las operaciones posibles en código binario se reducen a sumas de 0 y 1, pudiendo expresarse operaciones complejas de forma muy sencilla, aunque sean muy largas, como acumulaciones de simples operaciones de álgebra de Boole: operaciones lógicas, que se pueden diseñar con circuitos relativamente sencillos de electrónica digital.

#### cuantificación de la señal de video

- La **cuantificación** discretiza el rango de amplitudes: sólo se podrán tomar una serie de valores fijados de antemano, cuantificados fácilmente con valores discretos. El número de valores posibles será (en sistema binario), según n, el número de bits que elijamos para la cuantificación de:

$$Q_n = 2^n - 1$$

Si tomamos los 8 bits que forman un Byte, vemos que entonces la cuantificación puede tomar 256 valores (del 0 al 255).

- El número de bits con lo que suelen trabajar los equipos internamente es de 10, 16 y a veces 24 bits, transformándose después a la norma establecida.
- El error en los datos se reduce a la mitad de la cuantificación. Ese error produce el llamado ruido de cuantificación, que se reduce al aumentar el número n: cuanto mayor sea la cuantificación más se aproxima el valor cuantificado al valor real (Cada vez que se introduce un bit, el error se reduce por la mitad). Pero eso supone un aumento del caudal de información.
- Para saber el margen de ese ruido de cuantificación, es decir, el número de bits a utilizar en la cuantificación, tenemos que saber cuál es la relación señal/ruido adecuada, de acuerdo con la utilidad que queramos darle a esa señal.
- Esa relación se expresa como  $S/N = 20 \cdot \log(V_s/V_r)$ . Con cada bit de cuantificación el error se reduce a la mitad, con lo que cada por cada bit añadido se mejora la señal  $20 \cdot \log 2 = 6$  aprox. Podemos expresar entonces

$$S/N = 6n + 10,6 \text{ dB.}$$

(10,6 dB es una constante que aparece al considerar el valor eficaz del ruido, y no su valor pico a pico)

- Según esta fórmula, si queremos una relación mínima de 55 dB, debemos utilizar 8 bits.
- El rango dinámico del ojo humano es de 220. Con lo que con 8 bits se supera ese umbral perceptivo, y se supone que no habrá posibilidad de apreciar la diferencia entre el muestreo digital y la señal analógica, ya que 8 bits supone 256 valores posibles. Aplicando esto a cada componente de color, tenemos  $(2^8)^3 = 2^{24} = 16$  millones de colores.

conversiones analógico / digital – digital / analógico

- La variación en los valores debido a errores en transmisión, interferencias, etc. (ruido), que en el analógico no se puede separar de la señal, en digital es fácilmente rectificable, ya que sólo tenemos que ver en qué grado se ha desviado del 1 o 0 posible. Pero si el ruido aumenta se produce una ruptura de código, que anula cualquier tipo de señal.
- Antes de ser muestreada la señal se le suele introducir una señal de bajo nivel, un ruido blanco, con una amplitud de nivel de cuantificación pico a pico: el *dither*. Este ruido añadido asegurará que incluso la señal más pequeña cruzará al menos un nivel de cuantificación, de manera que los periodos adyacentes corresponderán a niveles de cuantificación diferentes.

El *dither* permite romper la correlación del error de cuantificación con la señal de entrada, logrando entonces que la transferencia digital sea más lineal, y eso mejora la cuantificación aunque produzca una pequeña pérdida en la relación señal/ ruido.

- En el proceso de conversión digital-analógico, es decir, cuando recuperamos la señal analógico a partir de la señal “escalonada” digital, se aplica un filtro de suavizado que evitará las transiciones bruscas presentes en el paso de un escalón a otro, que originará frecuencias altas. Ese suavizado se realiza con un filtro pasa-bajo que elimina las frecuencias que excedan la frecuencia de Nyquist. La respuesta de este tipo de filtro es de  $\text{sen}(x) / x$ , que asegura la recuperación de la señal original.

### señales digitales de video

- Las **señales digitales de video** son:
  - **RGB**
  - **COMPUESTA**
  - **COMPONENTES**

- En RGB sólo se trabaja en el interior de algunos equipos digitales (paletas gráficas, CAD)
- La señal compuesta aparece por digitalización de una única señal de televisión, ya sea PAL, NTC, SECAM. Son más “económicas” y además compatibles con un entorno analógico compuesto, pero arrastran los inconvenientes de la señal analógica compuesta (intermodulación luminancia-crominancia, secuencias de 4 y 8 campos), y además no son convertibles entre sí.
- En la señal por componentes se muestrean y cuantifican las señales Y, R-Y, B-Y. Esa señal es compatible para todos los sistemas de televisión, y tiene mayor calidad que la señal compuesta (mayor ancho de banda de las señales diferencias de color, menor ruido de cuantificación.) Desaparecen los efectos indeseados de intermodulación luminancia-crominancia y secuencias de 4 y 8 campos.

## Muestreo

### la norma 4:2:2

- En televisión, la **frecuencia de muestreo** no debe ser demasiado elevada, porque entonces correríamos el riesgo de muestrear el ruido. La frecuencia de muestreo debe cumplir una serie de requisitos:
  - Criterio de Nyquist: Establece que la frecuencia de muestreo mínima para cuantificar una señal sinusoidal es de dos veces la frecuencia de esta señal (si no, no se apreciaría la inversión de fase de la señal sinusoidal, por ejemplo.) Además, establece que la frecuencia de muestreo debe ser mayor que el doble del ancho de banda de la señal muestreada, para que no se produzca aliasing (solapamiento de las bandas laterales, que produciría interferencias en las zonas de solapamiento). Eso nos lleva a una frecuencia de muestreo superior a 12 MHz, ya que en SECAM hay frecuencias de hasta 6 MHz.
  - Criterio de Muestreo Ortogonal: debe ser múltiplo entero de la frecuencia de línea. Esto significa que cada línea debe contener un número entero de periodos de muestreo.
  - Norma Universal, para ser compatible con las dos frecuencias de línea existentes en el mundo (PAL-SECAM 15.625 Hz. NTSC 15734,2573 Hz.) El mínimo común múltiplo es de 2,25 MHz.
- Esto nos da una frecuencia de muestreo de 13,5 MHz, al multiplicar  $2,25 \cdot 6$  para obtener una frecuencia superior a los 12 MHz y múltiplo de 2,25 MHz.
- La Norma 4:2:2 o CCIR-601 (ITU-R BT 601) es la que se ha consolidado como norma universal de digitalización. Utiliza el muestreo de los componentes: Y, R - Y, B - Y. Supone que la luminancia se muestrea a 13,5 MHz., mientras que las señales de croma se muestrean a 6,75 MHz. En sistema PAL supone que el número de muestras de la luminancia por línea es de 864, de los cuales 720 corresponden a la línea activa, mientras que la croma es de 432 y 360, respectivamente. (En NTSC corresponde a 858 muestras por línea en la luminancia.)

- Esto significa que por cada 4 muestras de luminancia se realizan 2 de cada una de las dos señales diferencia de color.

Y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
B-Y	X		X		X		X		X		X		X		X	
R-Y	X		X		X		X		X		X		X		X	

$$13,5 : 6,75 : 6,75 \rightarrow 4 : 2 : 2$$

- Con ese muestreo de croma, que es más limitado que el de la luminancia, se asegura sin embargo que estamos trabajando por encima de la sensibilidad del ojo para apreciar alguna pérdida.
- 4:2:2 genera un flujo de 270 Mb/s.
- En cuanto a la codificación, si codificamos con 8 bits, obteniendo 255 valores posibles, el negro corresponde al 16 (00010000) y el blanco, valor máximo, al 240 (11110000). Los valores superiores se reservan para posibles sobremodulaciones.

Si codificamos con 10 bits, los 2 bits finales representan la parte fraccionaria

### señal digital de video

- La línea digital consta de 12 P -periodos de reloj de 13,5 MHz. (16 en NTSC)- antes de que arranque la línea analógica. A continuación aparecen 132 P (122 en NTSC) sin muestreo. La línea activa digital, que arranca poco antes que la analógica, consta de 720 P, tanto en la norma 625 como en la 525, para asegurar su compatibilidad. Las diferencias resultantes con la señal analógica serán eliminadas en la conversión D/A.

- La línea digital comprende 864 periodos de 13,5 MHz de los cuales 720 corresponden a la línea activa. La línea digital empieza 24 P antes que la línea analógica, es decir, tiene 24 palabras previas. A esto le sigue 4 palabras que transmiten la señal de referencia temporal (TRS) y después 280 palabras con datos de identificación y auxiliares, seguidos de otras 4 palabras de TRS. Tras esto, los datos de video, que acaban 24 periodos antes de 0H.
- No se utilizan medias líneas, con lo que la supresión de campo digital comienza en líneas enteras, y eso lleva a que el borrado de campo digital del campo impar sea de 24 líneas y el del par de 25 líneas. Como el campo par analógico comienza a mitad de la línea 313, en digital comienza a principio de ésta, con lo que el campo impar tiene 312 líneas y el par 313, con 576 líneas activas en vez de las 575 analógicas.
- En vez de impulsos de sincronismos se utilizan TRS. Y éstas no se pueden discriminar por valor de voltaje, como hacían los impulsos de sincronismo en analógico en -0'3 V., (ya que en digital sólo existen 0 y 1, y cuantificaciones de 8 y 10 bists) sino por la discriminación de grupos de ciertas secuencias binarias reservadas a los TRS, de forma que cuando aparezcan sean reconocidos como una secuencia de sincronización. Los TRS se dividen en EAV (End of Active Video) y SAV (Start of Active Video), y están colocadas al comienzo y al final del borrado de línea. Son secuencias de 4 palabras de 8 bits

1111 1111 0000 0000 0000 0000 1FVH P<sub>3</sub>P<sub>2</sub>P<sub>1</sub>P<sub>0</sub>

las tres primeras con un preámbulo fijo, identifica el TRS como tal:

1111 1111 0000 0000 0000 0000

y la última informa si estamos en el campo impar o el par, si es una línea activa o una línea de borrado vertical, y si nos encontramos al final o al principio del borrado de línea

1FVH P<sub>3</sub>P<sub>2</sub>P<sub>1</sub>P<sub>0</sub>

	0	1
F	Durante el primer campo	Durante el segundo campo
V	Durante periodo activo de campo	Durante borrado de campo
H	Al final del borrado de	Al inicio del borrado de

	línea	línea
--	-------	-------

Los cuatro últimos valores son  $P_3P_2P_1P_0$  son bits de paridad para comprobación de errores.

la norma 4:2:2 y otra normas de digitalización

- Antes de la norma 4:2:2, la UER había probado el muestreo de la luminancia a 12 MHz y de las señales de crominancia a 4 Mhz (señal 12:4:4), e incluso para postproducción se elevó el muestreo de la crominancia a 6 MHz. (12:6:6). La SMPTE propuso la codificación por componentes a 14:7:7. Finalmente, se buscó la compatibilidad llegando a la 13,5:6,75:6,75 (4:2:2).
- Si bien la norma 4:2:2 es suficiente para la apreciación de color del ojo humano, es insuficiente para operaciones de postproducción exigentes, como el croma key digital. Para ello, en que tenemos que tener una información de color alta, tenemos que jugar con una norma 4:4:4, muestreando ambas señales de diferencia de color con la misma frecuencia que la luminancia, generando un flujo binario de 405 Mb/s.

Y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
B-Y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
R-Y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

**13,5 : 13,5 : 13,5 → 4 : 4 : 4**

- Pero en algunos aparatos de postproducción se necesita además utilizar un valor de llave de transparencia por cada punto de muestreo para constituir la señal de máscara. Aparece así un cuarto valor de muestreo, K (llave) cuyo valor va desde el 0 (opaco)

hasta 255 (transparencia absoluta). Éste valor de transparencia constituye el canal alfa. Y la norma correspondiente, en que se muestrea en cada píxel cuatro valores (Y, R-Y, B-Y, K) es la llamada norma 4:4:4:4 o 4x4 y genera un flujo binario de 540 Mb/s.

- La norma 4:2:2 produce un flujo binario de 270 Mb/s., desmesurado para sistemas domésticos, así que se reduce el submuestreo de la croma a la mitad, con la norma 4:2:0. Una de las señales de croma no se muestrea durante toda una línea, pero cuando se hace es manteniendo el ancho de banda completo de la señal de croma. A la siguiente línea, se muestre la señal que no se muestreó, y la que se muestreó se ignora. La integración espacial asegura que se ve el color de forma correcta. El flujo binario es de 202,5 Mb/s.

#### Línea n

Y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
B-Y	X		X		X		X		X		X		X		X	
R-Y																

#### Línea n+1

Y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
B-Y																
R-Y	X		X		X		X		X		X		X		X	

- La norma 4:1:1 también reduce el flujo binario a 202,5 Mb/s. Lo hace muestreando ambas señales de croma 1 vez de cada 4 durante todas las líneas. Si bien eso nos parecería que sería lo más adecuado, se producen deficiencias en la apreciación de color que hacen preferible la norma 4:2:0.

Y	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

B-Y	X				X				X				X			
R-Y	X				X				X				X			

- Otra norma no broadcasting (no profesional, no emitible) es las 2:1:1, que desprecia el valor del segundo campo, con lo que su frecuencia de muestreo es 6:75:3,375: 3,375. Tiene un flujo binario de 135 Mb/s. pero con una gran merma de calidad. Otra será la 2:1:0, que funciona como la 4:2:0 pero con unas frecuencias de muestreo 6:75: 3,375:0.

	<b>4:2:2</b>	<b>4:1:1 / 4:2:0</b>	<b>4:4:4</b>	<b>4x4</b>
Muestras por línea	Y: 720 B-Y: 360 R-Y: 360  TOTAL: 1440	Y: 720 B-Y: 180 R-Y: 180  TOTAL: 1080	Y: 720 B-Y: 720 R-Y: 720  TOTAL: 2160	Y: 720 B-Y: 720 R-Y: 720 K: 720 TOTAL: 2880
Muestras por frame (576 L)	829440	622080	1244160	1658880
Muestras por segundo	20736000	15552000	31104000	41472000
Bits por segundo	8 bits/muestra 165888000 10 bits/muestr 207600000	8 bits/muestra 124416000	8 bits/muestra 250 Mb/s 10 bits/muestr 311 Mb/s	8 bits/muestra 332 Mb/s 10 bits/muestr 415 Mb/s

- La señal 4:2:2 es la mínima aceptada profesionalmente, a excepción de equipos para informativos. Se emplea, sin compresión, en el D1 y D5. Con compresión en el Betacam digital, Betacam SX, Betacam MPEG-IMX, Digital S y DVC PRO.
- Para comprimir una señal de video es más adecuado submuestrear previamente alguna de las señales componentes. Una compresión 5:1 sobre una señal 4:2:0 es más satisfactoria que una compresión 7:1 sobre una señal 4:2:2.
- Pero para postproducir una señal es más adecuado trabajar sobre 4:2:2.
- La norma 4:2:0 y 4:1:1 se utiliza, respectivamente para PAL y NTSC, en DV y DVCAM, mientras el DVCPRO-25 trabaja en 4:1:1 en ambos sistemas.

- La norma 4:2:0 se utiliza también en DVD, en el sistema de compresión MPEG y en el sistema de transmisión DVB.
- Los nuevos sistemas de almacenamiento, el Blu-ray y el HD DVD, al contar con una capacidad de almacenamiento mayor, y lograr alcanzar los 36 Mb/s, pueden alcanzar muestreos de mayor calidad. El Blu-ray puede dar una salida de 4:4:4 en RGB y en YUV de 4:4:4 y 4:2:2.

#### interfaces de comunicación digital

- Para comunicar los equipos y permitir la poca degeneración de señal por multigeneración, hay que conectar de forma digital los equipos. Para ello, se debe normalizar las interconexiones, apareciendo interfaces de comunicación digital. Éstas pueden ser en paralelo o en serie.

#### interfaces en paralelo

- Las **interfaces en paralelo** transmiten cada uno de los datos de las señales Y, R-Y y B-Y por separado

Pin nº	funcion	Pin nº	Funcion
1	RELOJ +	14	RELOJ -
2	TIERRA	15	TIERRA
3	DATO 7 (MSB) +	16	DATO 7 (MSB) -
4	DATO 6 +	17	DATO 6 -
5	DATO 5 +	18	DATO 5 -
6	DATO 4 +	19	DATO 4 -
7	DATO 3 +	20	DATO 3 -
8	DATO 2 +	21	DATO 2 -
9	DATO 1 +	22	DATO 1 -
10	DATO 0 +	23	DATO 0 -
11	DATO -1 +	24	DATO -1 -
12	DATO -2 +	25	DATO -2 -
13	MALLA		

y se multiplexan para el nivel 4:2:2 con una frecuencia  $13,5+6,75+6,75=27\text{MHz}$ , de forma

CB0-Y0-CR0-Y1-CB2-Y2-CR2-Y3- CB4-Y4-CR4-Y5- CB6-Y6-CR6-Y7...

- En 4:4:4 se realiza con dos cables, cada uno a 27 MHz.

CONEXIÓN A: CB0-Y0-CR0-Y1-CB2-Y2-CR2-Y3-CB4-Y4-CR4-Y5...

CONEXIÓN B: CB1-K0-CR1-K1-CB3-K2-CR3-K3-CB5-K4-CR5-K5...

- El interface paralelo tiene problemas, que crecen para distancias largas, en cuanto a conectores grandes, cables caros, aparición de intermodulación y necesidad de apantallamientos especiales. Esto se soluciona con el **interface en serie**, que multiplexa las palabras de 10 bits en un único tren binario por un solo cable coaxial de un solo conductor.

#### interfaces en serie

- El interface en serie en una conexión unidireccional entre dos equipos en la que las palabras de 10 bits (en las que hay información de señal de video, señales de referencia temporal TRS, datos auxiliares y señales de sincronización) se transmiten multiplexados en un único tren binario a través de un solo cable coaxial de un solo conductor.
- Para evitar largas cadenas de 0 y 1 en la transmisión, que crearían problemas de propagación por el cable coaxial, se aplica a la señal un barajado pseudoaleatorizado. Luego, los 1 se convierten en transiciones, y los 0 en no transiciones o flancos.
  - Se eliminan las bajas frecuencias al eliminar las cadenas largas de 0 y 1.
  - La componente continua es 0 o muy cercana a 0, ya que la señal se encuentra en estado alto aproximadamente el mismo número de veces que en bajo.

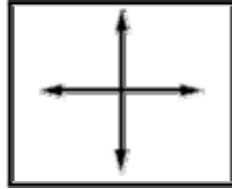
- La señal es autosincronizable, al asegurar la presencia de flancos o transiciones cada pocos periodos de bits.
- Al estar basado el estado lógico de los bits en la detención de flancos y transiciones, este interfaz es insensible a los cambios de polaridad de los cables.
- Dentro de los borrados horizontales y verticales hay un espacio que no está ocupado por ninguna información. Podemos rellenarlos con datos auxiliares, información ajena a la señal de video propiamente dicha, y que la pueden complementar: audio insertado en video, SDTI, EDH (detención y corrección de errores) y código de tiempo.

## **Compresión**

- Una vez realizado el muestreo y la cuantificación de la señal, tenemos una señal PAM cuantificada. En el caso de la televisión digital en 4:2:2, una señal que muestrea las componentes YUV con unas frecuencias 13,5:6,75:6,75 MHz. y una cuantificación a 8 o 10 bits.
- Pero esto crea unos archivos demasiado voluminosos, con unos flujos binarios de alrededor de 200 Mb/s. Hace falta reducir esto a alrededor de 10 Mb/s.
- Para ello, se debe someter a la señal a una serie de operaciones que reduzcan su tamaño y su ancho de banda. Esta serie de operaciones se conocen bajo el nombre de compresión.
- Todos los formatos de magnetoscopio (excepto el D1 y el D5, en desuso) llevan compresión: Betacam SX, IMX, Digital. La familia DV. La DVCPRO, el Digital-S. Soportes multimedia como el DVD. Sistemas de radiodifusión como el DVB o el ATSC.
- La compresión reduce así el tamaño de archivo, y con ello:
  - Podemos ubicar video y audio en formatos donde si no, no sería posible: DVD, sistemas de transmisión estrechos.
  - Podemos lograr mayor robustez y fiabilidad, en sistemas como DAT o DVTR, al grabar menos datos en el mismo espacio).

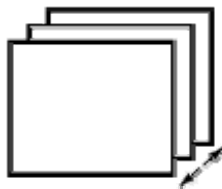
- En los magnetoscopios, podemos aumentar el tiempo de grabación de las cintas y reducir el tiempo de acceso (una misma cinta graba el doble de tiempo en Betacam SX que en SP.)
- Podemos enviar más programas sobre un mismo soporte (DVD, VOD...)
- En cualquier mensaje, podemos distinguir en los datos que lo conforman entre Entropía y Redundancia.
- Entropía es aquella parte del mensaje que no es posible reducir ni suprimir sin que se pierda algún detalle fundamental del contenido del mensaje. Constituye lo impredecible del mensaje, y es imposible comprimir esta parte de la información sin alterar el mensaje.
- Redundancia son esos datos que son susceptibles de ser resumidos o suprimidos sin que por ello se desvirtuale ni cambie el contenido del mensaje. Es aquella información que es predecible y redundante. Sobre ella es donde puede operar la compresión sin que su proceso llegue a ser irreversible.
- En video podemos considerar que hay tres tipos de información:
  - Elemento redundante, en el que hay una información que se repite o que es predecible.
  - Elemento irrelevante, información que el ojo humano es incapaz de apreciar, o que se considera que no es relevante para el mensaje.
  - Elemento básica, que constituye el núcleo del mensaje, ya que no es ni redundante ni irrelevante, y debe retenerse.
- La compresión reduce el peso de la información, eliminando la redundancia, y pudiendo afectar al elemento irrelevante, aunque la compresión no es deseable si afecta a la entropía.
- Existen tres tipos de compresión en video:

- **Compresión sin pérdidas:** se recupera la información sin pérdidas. Utilizada en sistemas informáticos. Existe precisión bit a bit, pero suele tener un bajo índice de compresión.
  - **Compresión subjetivamente sin pérdidas:** al recuperar la información las pérdidas posibles no son detectadas por el ojo humano. Existen errores de compresión, pero se sitúan por debajo del umbral de percepción. Tiene un nivel medio de compresión, y se utiliza ya en sistema broadcast.
  - **Compresión subjetivamente con pérdidas:** al recuperar la información las pérdidas posibles son detectadas por el ojo humano. Aquí tenemos errores apreciables y alta compresión. Se emplea en transmisión, donde el nivel de pérdida, aunque es alto, sigue manteniendo una calidad aceptable.
- La compresión se dirige a evitar:
- **La redundancia espacial o intracuadro:** cualquier píxel tiende a parecerse a los vecinos (en horizontal y vertical).



Espacial, en una sola imagen

- **La redundancia temporal o intercuadro:** cualquier píxel tiende a parecerse a los anteriores y posteriores en el tiempo.



Temporal, entre varias imágenes

- **La redundancia estadística:** determinados valores se repiten más que otros.

- Existen diversos métodos, herramientas y trucos de compresión, que se utilizan tanto para los sistemas de compresión JPEG (compresión intracuadro de imagen fija) y MJPEG (compresión intracuadro de imagen en movimiento) como para la compresión MPEG (compresión intercuadro de imagen en movimiento).

- **Sin pérdidas**

- Supresión de borrados
- RLC
- VLC
- DCT
- DCPM
- DCPM con predicción adaptativa

- **Trucos**

- En DCT
  - lectura en Zig-zag
  - EOF
- Compensación de movimiento

- **Con pérdidas**

- Submuestreo
- DCPM con predicción adaptativa
- Recuantificación de los coeficientes DCT



33-21-13-55-70-52 → 10001-1001-111-110111-1000110-110100

ha pasado de 192 bits a 31.

Es el tipo de codificación que emplea la transmisión por fax.

■ Codificación de longitud variable.

(VLC: Variable Length Coding)

basada en la redundancia estadística, asigna valores más pequeños a aquellos valores que más se repiten.

Es la que emplea el código morse, que asigna los valores más cortos a las letras que tienen más presencia en inglés

E → .

T → \_

Va asignando valores en longitud creciente según la menor probabilidad de su aparición en un mensaje cualquiera. Hasta llegar a los valores que estadísticamente son menos probables, a los que les da los valores más largos de combinación posible de . y \_

Q → \_ \_ . \_

Z → \_ \_ \_ . .

- Un tipo de codificación VLC es la Codificación Huffman, basada en la redundancia estadística, que asigna códigos más cortos a los valores que más probabilidad tengan de aparecer.

Para ello, en un mensaje (puede ser en un frame, por ejemplo), calcula cuáles son los valores que más se repiten. Los ordena luego por orden. Primero el que más aparece, y luego todos en orden decreciente, hasta llegar al que menos aparece.

Luego, renombra los datos, de forma que le asigna la cadena más corta (es decir, 0) al dato que más aparece. Y luego, va asignando cadenas nuevas, cada vez más largas, de forma que cada cadena nunca puede ser prefijo de otra. Formando una especie de árbol en la asignación de nuevos datos.

0

10

110

1110

1111

Etc.

Imaginemos un mensaje, por ejemplo:

10101010	11001100	11001100	00100010	00010100	11001100
00010100	11001100	11001100	00010100	11001100	11111000
11001100	00100010	11001100	00001000	00010100	11111000
11001100	11001100	00010100	00010100	11001100	00100010
11001100	11111000	00010100	11001100	11001100	00100010
00010100	11001100	11001100	00001000	00100010	11001100
11001100	00010100	11001100	00010100	10101010	11001100
11001100	11001100	11111000	11001100	00100010	11001100
00010100	11001100	11001100	00100010	11001100	00010100
11001100	11111000	10101010	11001100	11111000	00010100
11001100	00100010	00010100	10101010	00001000	

Son 65 palabras por 8 bits = 520 bits

Palabra	Aparece	Recodificamos
11001100	30 veces	0
00010100	14 veces	10
00100010	8 veces	1100
11111000	6 veces	1101
10101010	4 veces	1110
00001000	3 vez	1111

Esto, “traduciendo” de acuerdo con esta tabla, nos da

1110 0 0 1100 10  
 0 10 0 0 10  
 0 1101 0 1100 0  
 1111 10 1101 0 0  
 10 10 0 1100 0  
 1101 10 0 0 1100  
 10 0 0 1111 1100  
 0 0 10 0 10  
 1110 0 0 0 1101  
 0 1100 0 10 0  
 0 1100 0 10 0

1101 1110 0      1101 10  
 0      1100 10      1110 1111

Hemos pasado de 520 btis ¡a sólo 142 bits!

Para recomponer el mensaje, debemos transmitir junto con el mensaje la tabla de equivalencias, para que con ella el sistema pueda “retraducir”, descomprimir el mensaje original.

- Otro tipo de codificación de longitud variable es la codificación aritmética. En ella, se obtiene a partir de un mensaje un número decimal binario, al relacionar la media de aparición de una cifra con un intervalo fraccionario.

Veámoslo con un ejemplo, tomando el mismo mensaje de antes:

10101010    11001100    11001100    00100010    00010100    11001100  
 00010100    11001100    11001100    00010100    11001100    11111000  
 11001100    00100010    11001100    00001000    00010100    11111000  
 11001100    11001100    00010100    00010100    11001100    00100010  
 11001100    11111000    00010100    11001100    11001100    00100010  
 00010100    11001100    11001100    00001000    00100010    11001100  
 11001100    00010100    11001100    00010100    10101010    11001100  
 11001100    11001100    11111000    11001100    00100010    11001100  
 00010100    11001100    11001100    00100010    11001100    00010100  
 11001100    11111000    10101010    11001100    11111000    00010100  
 11001100    00100010    00010100    10101010    00001000

Son 65 palabras por 8 bits = 520 bits

Palabra	Aparece	Probabilidad	Subintervalo

		de aparición	
11001100	30 veces	0,46 %	0-0'46
00010100	14 veces	0,22 %	0'47-0'68
00100010	8 veces	0,12 %	0'69-0'80
11111000	6 veces	0,09 %	0'81-0'89
10101010	4 veces	0,06 %	0'90-0'95
00001000	3 vez	0,05 %	0'96-1'00

Y ahora, en nuestro mensaje, aplicamos de forma sucesiva un decimal, según aparece el dígito correspondiente:

0'90-0'95	0-0'46	0-0'46	0'69-0'80	0'47-0'68
0-0'46	0'47-0'68	0-0'46	0-0'46	0'47-0'68
0-0'46	0'81-0'89	0-0'46	0'69-0'80	0-0'46
0'96-1'00	0'47-0'68	0'81-0'89	0-0'46	0-0'46
0'47-0'68	0'47-0'68	0-0'46	0'69-0'80	0-0'46
0'81-0'89	0'47-0'68	0-0'46	0-0'46	0'69-0'80
0'47-0'68	0-0'46	0-0'46	0'96-1'00	0'69-0'80
0-0'46	0-0'46	0'47-0'68	0-0'46	0'47-0'68
0'90-0'95	0-0'46	0-0'46	0-0'46	0'81-0'89
0-0'46	0'69-0'80	0-0'46	0'47-0'68	0-0'46
0-0'46	0'69-0'80	0-0'46	0'47-0'68	0-0'46
0'81-0'89	0'90-0'95	0-0'46	0'81-0'89	0'47-0'68

0-0'46          0'69-0'80      0'47-0'68      0'90-0'95      0'96-1'00

Damos los primeros pasos.

Con el primer Byte, tomamos el intervalo entre 0'90-0'95.

Con el segundo Byte, tenemos que tomar el intervalo 0-0'46, aplicándolo sobre el anterior como si fuera la unidad, dividiéndolo en 100 partes y tomando sólo la parte entre 0 y 0'46. Esto es, el intervalo estará entre 0'90 y ( $< 0'90 + (0'95-0'90) \times 0'46 >$ ). Esto es, 0'90-0'923.

Con el tercer Byte, entre 0'90 y ( $0'90 + < (0'923-0'90) \times 0'46 >$ ). Esto es: 0'90-0'91058.

Con el cuarto Byte, entre  $< 0'90 + 0'01058 \times 0'68 >$  y ( $0'90 + < 0'01058 \times 0'80 >$ ). Esto es: 0'9071944 y 0'908464).

Con el quinto Byte, aplicando sobre el intervalo anterior: 0,0012696 → el fraccionario correspondiente está entre  $< 0'9071944 + 0,0012696 \times 0'47 >$  y  $< 0'9071944 + (0,0012696 \times 0'68) >$ , esto es: 0,907791112 y 0,908057728. (Lo que da un intervalo de 0,000266616).

Con el sexto Byte, entre  $< 0,907791112 >$  y ( $0,907791112 + < 0,000266616 \times 0'46 >$ ). Esto es: 0,907791112 y 0,90791375536. (El intervalo ahora es de 0,00012264336)

Con el séptimo Byte, el fraccionario correspondiente está entre  $< 0,907791112 + 0,00012264336 \times 0'47 >$  y  $< 0,907791112 + 0,00012264336 \times 0'68 >$ , esto es: 0,9078487543792 y 0,9078745094848. (Lo que da un intervalo de 0,0000257551056).

Y así continuaríamos hasta llegar a la última palabra. El número fraccionario correspondiente, que habría que expresar como número binario (es decir, en potencias negativas de 2, sabiendo que  $2^{**}(-n) = 1 / 2^{**}n$ . Es decir,  $2^{-1} = 1/2 = 0'5$ ,  $2^{-2} = 1/4 = 0'25$ , etc.)

El número resultante no sería propiamente un número, sino como hemos visto un intervalo (si bien los intervalos son cada vez más pequeños, según

codificamos un nuevo Byte). Para expresar la cadena, podemos tomar cualquier número comprendido entre los dos límites que nos dé el último Byte de la lista.

Es decir, vamos a figurarnos que este séptimo Byte fuera el último (realmente, aún queda bastante cálculo que hacer). El número que codifica esta cadena de 7 Bytes estaría entre:

0,9078487543792 - 0,9078745094848

$2^{-1}$	0,5	1	0,5
$2^{-2}$	0,25	1	0,25
$2^{-3}$	0,125	1	0,125
$2^{-4}$	0,0625	0	0
$2^{-5}$	0,03125	1	0,03125
$2^{-6}$	0,015625	0	0
$2^{-7}$	0,0078125	0	0
$2^{-8}$	0,00390625	0	0
$2^{-9}$	0,001953125	0	0
$2^{-10}$	0,0009765625	1	0,0009765625
$2^{-11}$	0,00048828125	1	0,00048828125
$2^{-12}$	0,000244140625	0	0
$2^{-13}$	0,0001220703125	1	0,0001220703125
$2^{-14}$	0,00006103515625	0	0
$2^{-15}$	0,000030517578125	0	0

$2^{-16}$	0,0000152587890625	0	0
$2^{-17}$	0,00000762939453125	1	0,00000762939453125
$2^{-18}$	0,000003814697265625	1	0,000003814697265625
$2^{-19}$	0,0000019073486328125	0	0
$2^{-20}$	0,00000095367431640625	0	0
$2^{-21}$	0,000000476837158203125	1	0,000000476837158203125
	SUMA.....		0,9078488349915

El número 0,9078488349915, que está comprendido entre 0,9078487543792 - 0,9078745094848, y que se expresa de forma binaria como 0'111010000110100011001

codifica sin ningún tipo de error la secuencia de los 7 primeros Bytes.

Es decir, quitando la parte entera, hemos obtenido el número 111010000110100011001, con 21 bites, está codificando a  $7 \times 8 = 56$  bits.

Naturalmente, que aparte de ése mensaje habría que transmitir la tabla correspondiente de valores de subintervalo.

## DPCM

- Diferential Pulse Code Modulation, que aprovecha la redundancia espacial para codificar la diferencia entre un valor y el del valor anterior. Si el PAM (o PCM) toma valores absolutos, el DPCM sólo toma el valor de lo que se ha incrementado la señal con respecto a la medida previa. No se toman valores absolutos, sólo valores relativos.

Por ejemplo, si los valores sucesivos de la señal PAM son:

214	211	210	212	209	208	210	211	214	206	208	210	211	209	211
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

los valores DPCM a partir del primer valor 214 son

214	211	210	212	209	208	210	211	214	206	208	210	211	209	211
	-3	-1	+2	-3	-1	+2	+1	+3	-8	+2	+2	+1	-2	+2

Con ello se logra que se reduzcan la magnitud de los valores que suele tomar la señal, ya que sólo interesa la variación de una muestra con respecto a la anterior. Esto significa, que podemos reducir el número de bits con que expresamos, la señal, ya que si en PAM debemos trabajar con 8 bits, en DCPM, en este caso, podemos trabajar con sólo 4 bits (el valor más alto es de +8, es decir 1000.)

Por otra parte, vemos que la DCPM logra hacernos ver valores recurrentes (ya que al trabajar en la forma en que varía la señal, más que en la forma en que la señal toma valores, podemos apreciar tendencias en la señal). En este caso, vemos que el valor +2 aparece mucho.

Sin embargo, la DCPM tiene un defecto grave en cuanto que en un dato haya un error, ya que ese error se acarrea en las muestras siguientes, alterando la señal de forma esencial y sin corrección posible. Para que esto no ocurra, hay que volver a dar el valor absoluto tras un número limitado de muestras tomadas en DCPM.

- DPCM con predicción adaptativa, una vez realizado el DCPM, codifica la desviación con respecto a un predictor del valor real. Funciona a través de los siguientes pasos:
  - Se codifica en DCPM.
  - En ciertos valores, uno de cada cuatro, se transmite el valor absoluto. En estos casos, este es el valor que se codifica.
  - Pero en otros, tres de cada cuatro, se procede a realizar el DCPM.
  - Aprovechando la redundancia ortogonal que el DCPM logra expresar tan bien, se calcula para algunas de estas últimas muestras un valor aproximado, según un predictor (algoritmo adecuado para la muestra que se toma) para cada una de estas muestras.
  - Se compara ese valor previsto con el valor real en esas muestras. Y entonces, se codifica el error de predicción.
  - Transmitimos entonces:
    - los valores reales 1 de cada 4.
    - Los valores en DCPM no sometidos a predicción.
    - El algoritmo de predicción, junto con los errores de predicción.

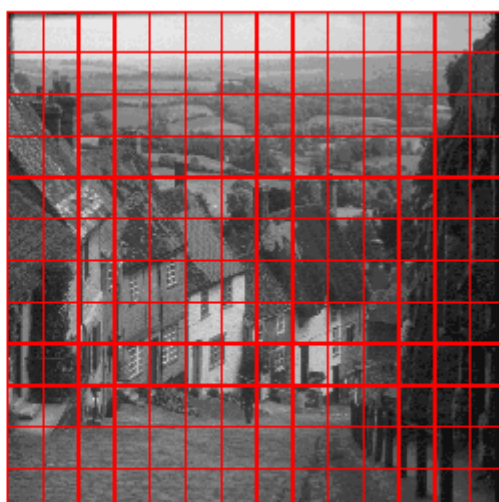
La DPCM con predicción adaptativa puede utilizarse tanto intracuadro, si la predicción es espacial, o intercuadro, si la predicción es temporal.

- Se puede combinar DCPM y VLC, basándose en esa redundancia que aparece en la DCPM. Además, si existe algún salto brusco (con valor de DCPM grande, que hasta puede ser superior al valor absoluto), ese valor no hace falta considerarlo con precisión, ya que al ser un valor poco frecuente (y por lo tanto,

con una codificación en una cadena de mayor longitud), podrá ser enmascarado en el umbral de percepción. En cambio, los valores que más se repiten, y que dan cuenta entonces de la señal, han de ser considerados con mayor precisión.

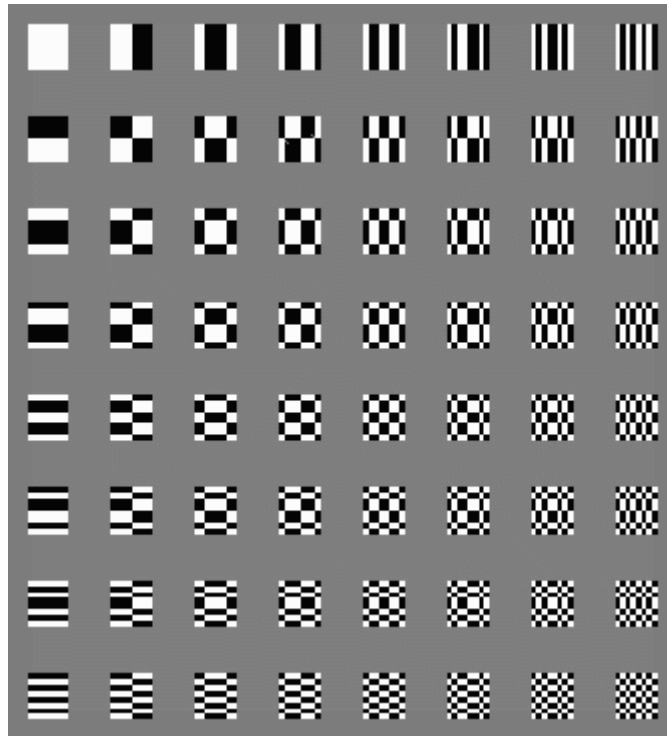
## DCT

- DCT o transformada de coseno discreto, es una operación muy utilizada en la compresión intracuadro, y está presente en las normas JPEG, MJPEG y en la familia MPEG. Realmente no efectúa compresión, sino que es una transformada de Fourier que “traduce” los datos de amplitud del espectro de tiempo a índices del espacio de las frecuencias. Esta transformada nos permite aislar los valores con mayor redundancia para eliminarlos sin que existan pérdidas apreciables en la imagen.
- La DCT divide la imagen en bloques de 8x8 píxeles. Y lo que hace es traducir cada bloque (que en cada uno de los 64 valores que toma –correspondientes a los 64 píxeles de que está formado el bloque- cuenta habitualmente con una gran dispersión de valores) a otros 64 valores, asociados a las frecuencias espaciales que conforman el bloque (una frecuencia se relaciona con la relación de detalle del bloque. Una frecuencia alta supone un grado alto de filigrana. Una frecuencia baja, que apenas existe variación en el valor del bloque a lo largo del espacio.)

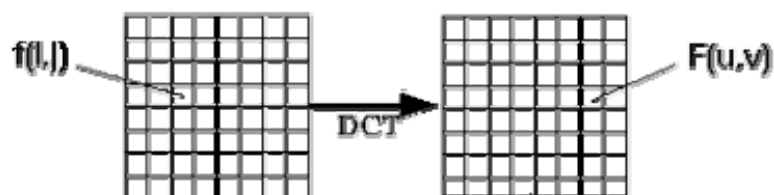


- First split the image into blocks
- Then calculate the DCT for each block

- Para ello, se considera que hay 64 patrones básicos para el bloque 8x8, por alternancia de valores 0 y 1 ( negro/blanco), en orden de frecuencia creciente y asociados a la función coseno (es decir, funciones sinusoidales con cierta simetría.)
- Los 64 valores originales (previos a la transformada), que están ligados a la amplitud de la señal en cada píxel, son traducidos por la DCT en 64 nuevos y diferentes valores que corresponden a los índices con los que participan en dicho bloque los 64 patrones básicos de frecuencias crecientes.
- Las frecuencias crecen a la derecha en horizontal, y hacia abajo en vertical. Unas y otras frecuencias crecientes se combinan en el cuadro patrón. El de abajo a la derecha, el de mayor frecuencia, en el que cada píxel varía de valor con respecto al vecino, es un pequeño damero de ajedrez de 8 pixels.
- El patrón de arriba a la izquierda representa el valor de frecuencia 0. Éste representa el valor medio de amplitud del macrobloque. Es decir, su “color” medio (esos cuadrados de color en que se descomponen los JPEG muy comprimidos). Es el valor DC, un valor uniforme para todo el bloque. Y de forma general, es el valor más predominante en los índices de frecuencia.
- El resto de los valores, relacionados con los diferentes patrones de frecuencia creciente, se llaman valores AC, y sus valores son mucho más reducidos que el de la DC. Además, muchos de estos valores AC suelen ser 0.

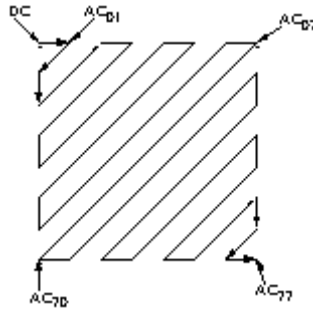


- Resumiendo, cualquier bloque de 8x8 puede descomponerse como la suma de una serie definida de algunos de esos valores patrón, y con diferente “intensidad” para cada patrón.
- La DCT en principio no comprime, simplemente traduce datos de diferente nivel de amplitud (“grises”) en cada uno de los 64 pixels (x,y), a una serie de índices relacionados con frecuencia crecientes (de izquierda arriba a derecha abajo.)



- Al traducir la DCT los bloques al espectro de frecuencia, los patrones de menores frecuencias son las que predominan en los bloques. La dispersión de valores se reduce. Suele aparecer un valor DC destacado, y una colección pequeña de índices relacionados con las frecuencias crecientes, y en los que se da un alto nivel de redundancia.
- Esos valores de frecuencia creciente expresan el grado de detalle de la imagen. Una frecuencia alta corresponde a un detalle pormenorizado de la imagen.

- Como los datos se agrupan arriba a la izquierda, y van decreciendo rápidamente hacia abajo a la derecha, se procede a leer los datos en **zig-zag**.



- Esto hace que los datos agrupen los valores más altos, y que caigan rápidamente, para encontrarnos luego con valores muy bajos o nulos.
- Como por regla general encontramos enseguida que los valores diferentes de 0 se “acaban”, podemos cuando ya no haya valores diferentes de 0 marcar una señal de EOF (fin del archivo) para que el sistema comprenda que a partir de ahí no hay más datos diferentes de 0, y que luego el sistema lo recomponga cuando descomprima.
- Por otra parte, si los valores más altos proporcionan los valores básicos de la imagen (frecuencias bajas), los valores más pequeños recaen en los valores de detalle (frecuencias altas). Podemos aplicar aquí **COMPRESIÓN PERCEPTUAL**, y suprimir esas informaciones más pequeñas correspondientes al detalle.
- Una recuantificación de los coeficientes, que aplique diferentes niveles de redondeo a los valores a medida que van creciendo las frecuencias, permite redondeos que afectan al “detalle” de la imagen. Y los valores de frecuencia más alta podrían ser redondeados a 0.
- Además, los valores de frecuencia más alta son los que suele ocupar el ruido. Por ello, al ser esos valores proclives a ser enmascarados por el ruido, no merece la pena transmitirlos en detalle.
- El proceso completo DCT supone los siguientes pasos:
  - A partir de R, G, B, se obtienen las señales compuestas: Y, R-Y, B-Y.
  - Se digitalizan según la norma 4:2:2.

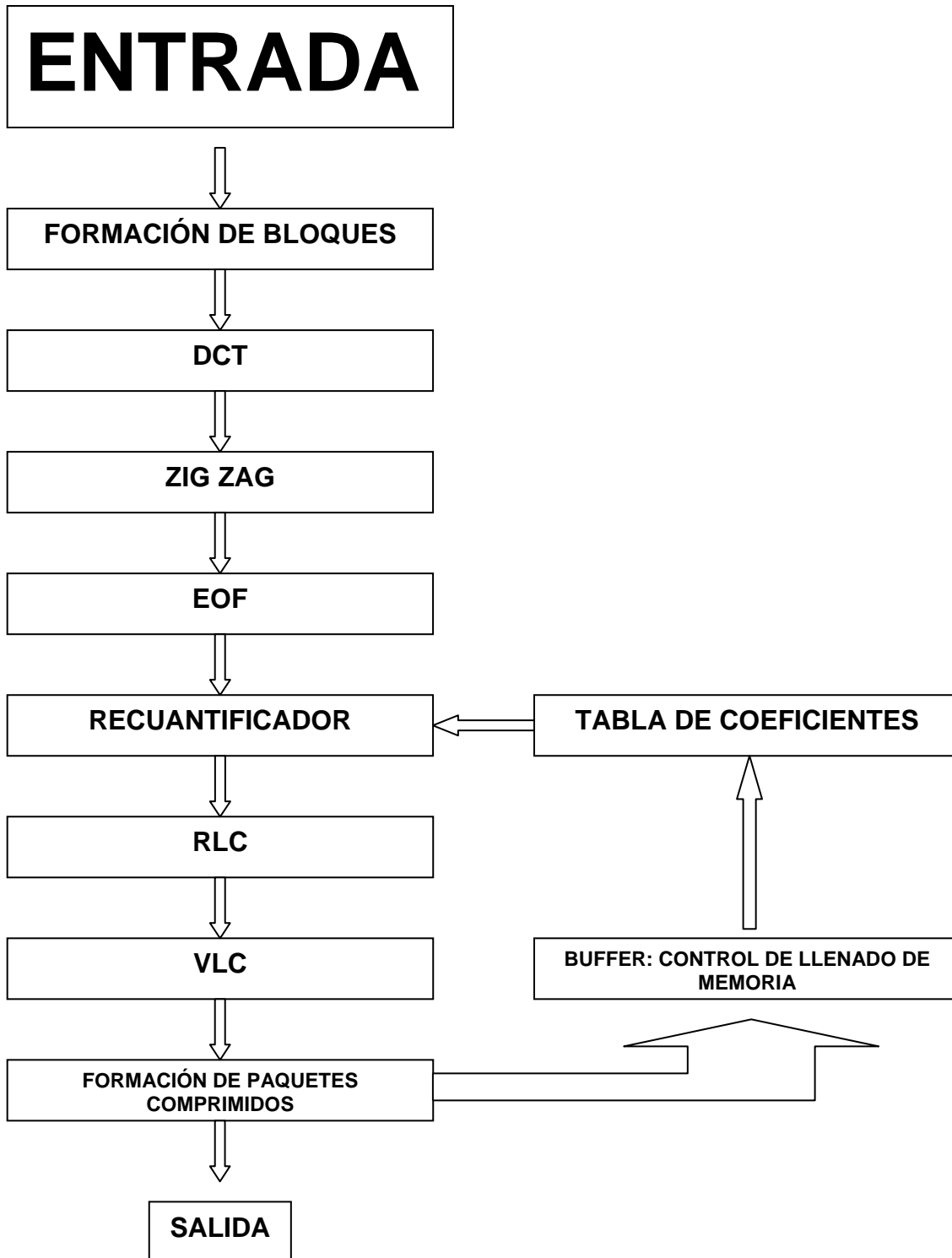
- Se realiza la transformada de coseno discreto de los valores de la amplitud sobre un bloque de 8x8 píxels.
- Se cuantifican los coeficientes frecuenciales, lo cual reducirá el número de valores que utilizamos para describir el bloque.
- Se procede a la lectura en zig-zag de los coeficientes, lo cual supone que, al estar los datos más significativos en las bajas frecuencias, se destaca estos datos por encima de los menos significativos de las frecuencias altas, que serán procesados o no según el grado de compresión que queramos.

## REMUESTREAR

- Una técnica de compresión con pérdida consiste en REMUESTREAR la señal. Pasar de 4:2:2 a 4:2:0 o 4:1:1. Eso supone un importante descenso del flujo binario.
- Sin embargo, el submuestreo, que evitaría compresiones posteriores, no es aceptado por normas de magnetoscopios, como los de la familia BETACAM, que prefiere no submuestrear aunque luego tenga que comprimir más. Mientras que la familia DV prefiere submuestrear antes para luego comprimir menos.

## BUFFER

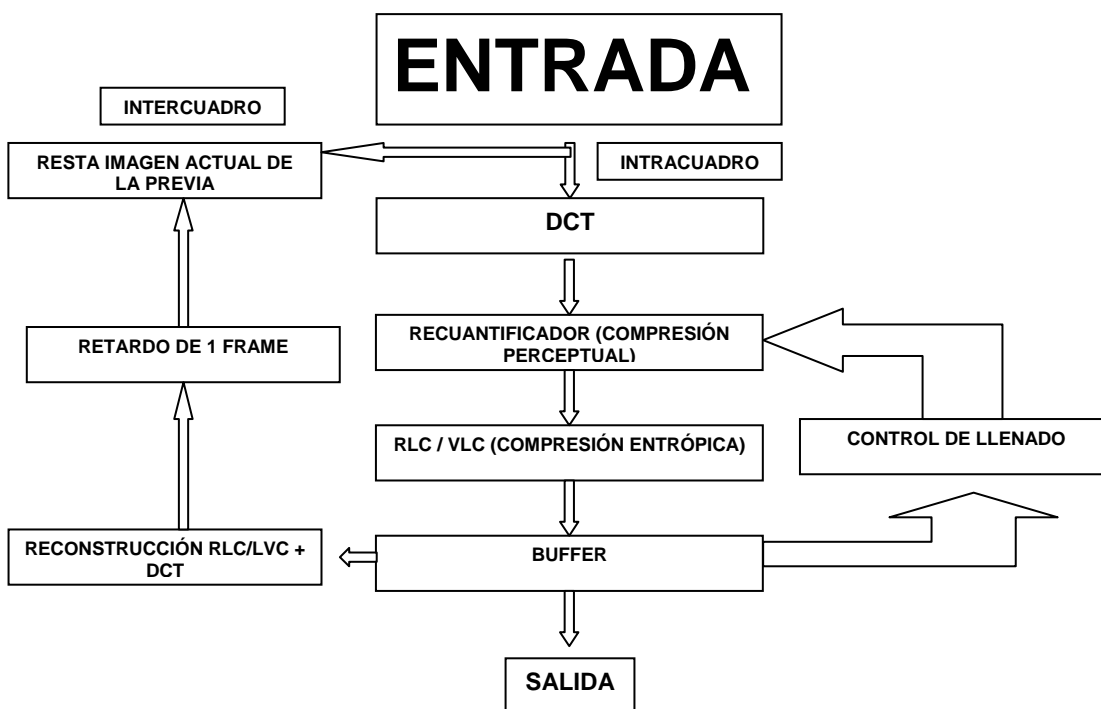
- Dado que la cantidad de información entrópica es variable, según cada imagen (podemos alternar un montaje de un primer plano, de baja frecuencias, con el de un plano general de altas frecuencias), la cantidad de información entrópica y la relación con la redundante varía continuamente. Lo que nos plantea el debate de si es conveniente variar el flujo binario para no recortar nunca la entropía o no dar en ciertos momentos información redundante.
- Sin embargo, en la grabación en cinta es necesario un flujo binario de salida estable. Para ello, se sitúa un BUFFER que regule que el flujo binario sea estable, variando en cada momento el nivel de compresión para que este flujo de salida sea siempre constante. Esto provoca que haya variaciones en cada momento de la calidad, aunque el buffer asegura que esta variación estará dentro de un nivel de calidad de imagen casi estable.



**DIAGRAMA COMPLETO DE LA COMPRESIÓN INTRACUADRO**

## Compensación de movimiento

- Temporalmente, se puede considerar que en las imágenes en movimiento, de frame a frame, apenas hay variación. Si tenemos un objeto en movimiento, apenas se habrá desplazado y su forma cambiado de un fotograma al siguiente. Por ello, podemos hablar de la COMPENSACIÓN DE MOVIMIENTO como una técnica para comprimir.
- Esto consiste en eliminar como información redundante aquellos detalles que no han sufrido variación significativa de un fotograma al siguiente, como espacialmente se hacía con el DCPM.



- Sin embargo, un objeto entero sería difícil considerarlo como un ente único, ya que al moverse suele cambiar su forma, su volumen, etc. Además, otros detalles de la imagen pueden cambiar, así como puede haber diferencias por factores fortuitos que desrelacionen la información de diferencia con relación al movimiento.
- Por ello, no se consideran objetos solidarios, sino pequeños trozos de la imagen. Macrobloques formados por 2x2 bloques de 8x8.

- Y se da un vector de desplazamiento que nos indica el movimiento de este macrobloque y a través del cuál se reconstruye su situación en el siguiente fotograma.
- La compresión intercuadro complementa la intracuarto. Para ello, reconstruimos el frame previo comprimido y lo restamos al actual para lograr la compresión temporal.

### Preparación previa a la compresión

- Antes de comprimir la señal, sería conveniente prepararla para la compresión
  - Suavizando bordes
  - Corrigiendo la apertura de la cámara
  - Limpiándola de ruidos, intermodulaciones croma-luminancia, etc.
  - Procurando tener una imagen “limpia” y sin movimientos ajenos a la imagen.

## ARTEFACTOS. Errores de compresión.

- La compresión, si no es adecuada o no se realiza con cuidado, puede dar lugar a una serie de artefactos o artificios, esto es defectos visibles de compresión. Estos artefactos afectan a la calidad de la imagen, y su efecto es irreversible y bastante molesto. Entre ellos están
  - BLOCKING o efecto mosaico: apareciendo en la imagen la estructura de bloques de DCT y de macrobloques de la compensación de movimiento.
  - BLURING, o efecto emborronado. Al reducir información esencial, aparece un emborronamiento y un efecto de corrimiento de tintas.
  - EDGE BUSYNESS (actividad en los bordes). Distorsión en los bordes de los objetos (zonas de alta frecuencia).
  - ERROR BLOCKS (bloques erróneos). Aparecen bloques que no deberían estar, desplazados temporalmente, por ejemplo, o congelados.
  - MOSQUITO NOISE (ruido nervioso). Debido al error de cuantificación entre píxeles adyacentes. A medida que la escena varía, aumentan los errores de cuantificación, que se manifiestan como puntos negros en movimiento.
  - QUANTIZATION NOISE (ruido de cuantificación), al reconstruir la señal analógica, se produce una representación inadecuada debido a los procesos de compresión.

## Artefactos (defectos) de compresión

Enciclopedia DVD

- Ver listado de términos

El vídeo, especialmente cuando es complejo, con imagen cambiante rápidamente o en escenas oscuras o de poco contraste, puede contener defectos tales como falta de trozos de imagen, imagen borrosa o ruido de vídeo, dependiendo de la calidad del procesado y de la cantidad de compresión aplicada.

Un caso típico es la falta de degradado de color, que se aprecia fácilmente en escenas nocturnas en películas mal comprimidas, donde nos aparecen cambios bruscos del color entre regiones de la imagen. Aquí va un ejemplo de "Sleepy Hollow", edición Española de TriPictures, usada habitualmente como ejemplo de mala compresión (se ha aumentado el contraste de la imagen, para acentuar más el defecto):



Es también muy común el **Efecto Mosaico** o *Macro-blocking*: Se aprecia que varios píxeles adyacentes tienen el mismo color, lo que transforma la imagen (si nos fijamos bien) en una especie de mosaico:



Otra variante es el **Ruido de Mosquito** - *Mosquito noise* o *Gibb effect*: Quedan partes borrosas y manchas (píxeles) de ruido que aparece con el color incorrecto fuera de donde deberían estar; es decir, desplazadas fuera de la parte de la imagen que está en movimiento y mezcladas con el fondo, como en este ejemplo:



También entra en esta categoría la **Posterización**: La gradación de colores de un área de la imagen en lugar de ser suave, como debería ser, es más bien abrupta como consecuencia de una pérdida de la resolución cromática durante los procesos anteriores. Es más evidente cuando esta gradación de color viene acompañada de una gradación de luminosidad, sobre todo si ésta se origina en una imagen oscura.

El resultado de todo ello es parecido al efecto que se produciría si observáramos un mapa topográfico en el que se apreciaran distintas zonas de la imagen cada una de ellas de una tonalidad uniforme, pero con cambios abruptos al pasar de una zona a la siguiente. Lo vemos mejor en esta simulación de varios fotogramas consecutivos (se observa el defecto en forma de anillos concéntricos en la puerta y la pared):



# Edge enhancement

Enciclopedia DVD

o "realce de contornos": Un [Artefacto de compresión](#) muy común al digitalizar el vídeo para su codificación en DVD, cuando se aplican procesos digitales para intentar limpiar la imagen (reducción de ruido, aumento del contraste). Estos filtros aplicados al vídeo antes de su codificación MPEG, tienden a exagerar las áreas de transición entre luz y oscuridad o entre diferentes colores, lo que produce unos halos en la imagen. Como ejemplo, fijaros en el pelo de la damisela que está a la izquierda de Elisabeth Taylor, en este fragmento de "**Cleopatra**".



# Pixelado

Enciclopedia DVD

- Ver listado de términos

Por problemas extremos de compresión en el disco o por defectos puntuales de lectura (disco sucio o rayado, lente sucia, etc.), podemos encontrarnos con la imagen en "mosaico", como en este ejemplo. Si no se soluciona limpiando el disco (que puede haber sido usado previamente como posavasos o como cenicero), puede tratarse de un defecto de la edición o de nuestra copia.



# WAT, efecto

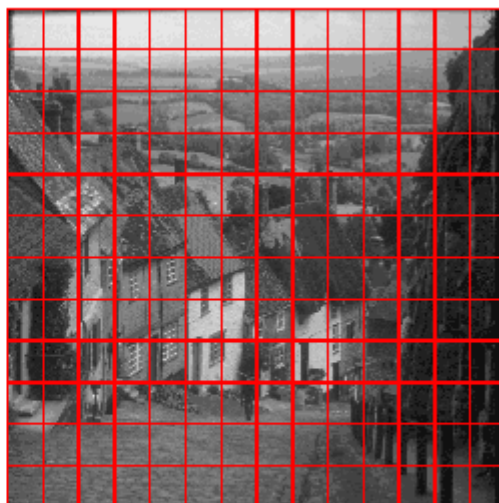
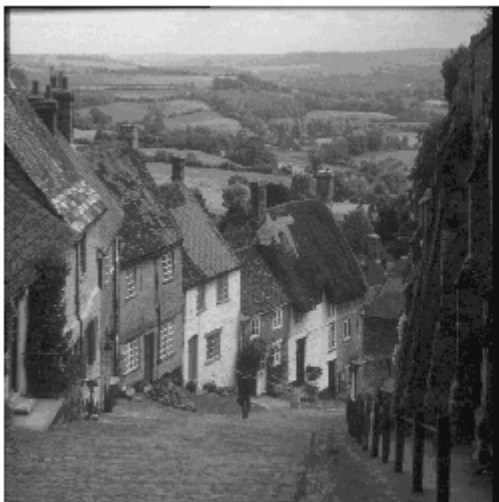
Enciclopedia DVD

- Ver listado de términos

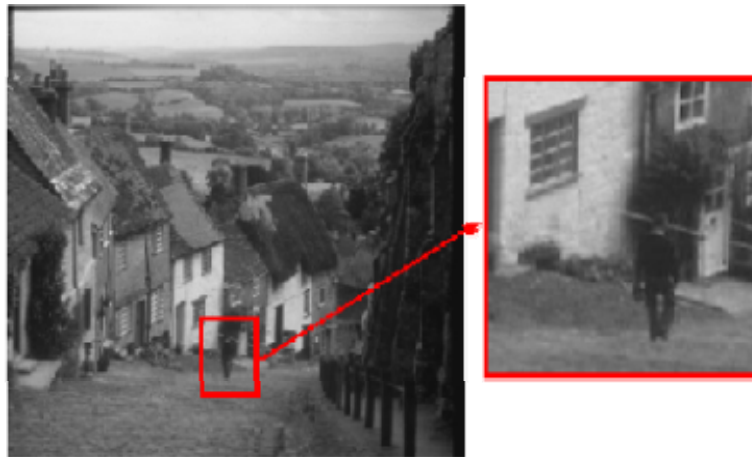
(del inglés "Women against technology" o "Mujeres contra la tecnología"): Fenómeno sociológico muy extendido (aunque con numerosas y honrosas excepciones), por el que la mayoría de esposas o compañeras de adictos al DVD se oponen sistemáticamente a la instalación de siete altavoces en el comedor familiar, a instalar este receptor A/V tan grande que no hace juego con las cortinas o al despilfarro de una parte considerable del presupuesto doméstico en la compra de componentes Hi-Fi, TV, proyectores o en películas en DVD. Claro que luego son las primeras en presumir de lo bien que han visto aquella película en la instalación de "Cine en Casa" y en invitar a toda tu familia política a ver "**Gladiator**" en la TV 16:9, con este sonido tan espectacular...

## LOS SISTEMAS DE COMPRESIÓN

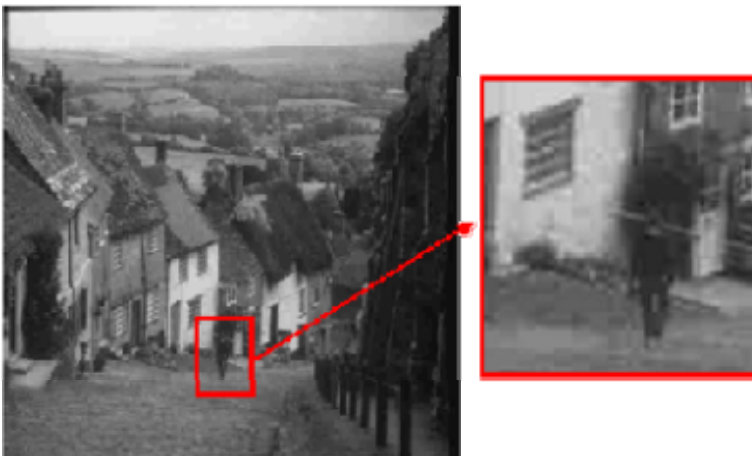
- Los sistemas de compresión para imagen digital son el JPEG (compresión intracuadro válido para compresión de imágenes fijas, y utilizado en almacenamiento y retoque fotográfico), el MJPEG (compresión intracuadro para imagen en movimiento); la familia de sistemas MPEG (compresión intercuadro de imagen en movimiento); y entre otros sistemas más, el grupo DV, utilizado en la compresión intracuadro de los magnetoscopios DV (DV, DVCAM, DVCPRO-50, DVCPRO-25 y DIGITAL-S).
  
- El **JPEG** (Join Photographic Expert Group) es una norma para imágenes fijas que utiliza la DCT a bloques de 8x8 pixels para después aplicarles diferentes formas de compresión, reduciendo la redundancia espacial.



- First split the image into blocks
- Then calculate the DCT for each block



**Figure 2:**Original uncompressed image.



**Figure 3:**Image compressed with JPEG to 15:1.

- El JPEG comprime de 2:1 a 100:1, logrando compresiones altas con calidad muy satisfactoria (15:1). Pensado en principio para imágenes monocromas, convierte la señal RGB en señal Y Cb y Cr (luminancia más diferencias de color). Después, divide la imagen completa en Macrobloques bloques de 16x16. Para Y, lo divide en 4 (2x2) bloques de 8x8, mientras que para Cb y Cr muestrea la mitad, formando en cada caso un único bloque de 8x8.
- Luego procede a realizar DCT. Los valores de DC se comprimen aparte, como DCPM (valores diferencia) de los bloques adyacentes. Los valores AC se recuantifican según una tabla de valores que depende del índice de compresión. La información más básica se transmite sin comprimir, mientras que los detalles se transmiten de una forma aproximada, e introduciendo errores crecientes según queramos comprimir más.

- Leyendo los valores AC en zigzag, se procede a realizar codificación VLC – Huffman. Para ello, se utilizan una serie de descriptores según la diferente categoría de cada dato. Divide los datos posibles en categorías, crea una tabla con ella e indica simplemente la categoría del dato y en qué posición dentro de ésta está el dato.
- Las características del JPEG las podemos resumir entre otras, en:
  - Dos formas de codificación, entrópica (sin pérdidas) y perceptual (con pérdidas).
  - Permitir al usuario elegir el grado de compresión.
  - Permite la codificación secuencial (una sola pasada) y la codificación progresiva (múltiples pasadas).
  - Ofrece la codificación jerárquica, que permite obtener una copia de baja calidad sin descomprimir de forma completa el archivo.

## MJPEG

- **MJPEG** (Moving Join Photographic Expert Group), utilizado en procesos de edición no-lineal, es realmente la aplicación sucesiva de JPEG a los frames que componen el plano. Se elimina la redundancia espacial realizando la compresión intracuadro pero no se comprime intercuadro.
- Pero el MJPEG, como diferencia con el JPEG, mantiene un flujo binario fijo, con lo que es un sistema de calidad variable con frecuencia binaria constante (mientras el JPEG es un sistema de calidad constante con frecuencia variable).
- Se emplea en editores no-lineales. Uno de sus problemas es que no está normalizado, con lo que la compresión que se realiza en un aparato de postproducción no es decodificable por otro. La norma MJPEG está siendo desplazado por la familia MPEG.

## MPEG

- **MPEG** (Moving Picture Experts Group) se caracteriza por codificar no la imagen sino la diferencia de una imagen con respecto a la imagen anterior. Emplea codificación intercuadro con predicción y compensación de movimiento.
- Existen varias normas MPEG, que han ido desarrollándose desde 1988 con la aparición del MPEG1, y ahora siguen en desarrollo con el MPEG 21:

- **MPEG-1 (ISO/IEC 11172)**

Se estableció en 1992 en Junta de Comité Técnico de la ISO/IEC con el fin de crear un estándar en la codificación de video digital con una buena calidad de imagen y sonido a bajo bitrate. Su tasa de transmisión es mayor a 1.5Mbit/ Seg. Se compone de varias partes: IS 11172-1 que describe sincronización y multiplexación de video y audio. : IS 11172-2 que describe la compresión de señales de video no entrelazadas, : IS 11172-3 que describe una familia genérica de compresión de audio ( el archiconocido MP3), : IS 11172-4 usado para testeo e IS 11172-5 para simulación por software. Su calidad es similar a la de un video casero, las aplicaciones que mas se han aprovechado de este formato han sido el VideoCD para video y el MP3 para audio en su tercera capa.

- **MPEG-2 (ISO/IEC 13818)**

Se estableció en 1994. fue lanzado como una segunda fase de trabajo de MPEG con el fin de crear un estándar genérico que ofreciera buena calidad a altos bitrates cercanos a 5Mbit/ Seg. Para calidad NTSC / PAL. Es usado para almacenar videos en DVD y emisiones digitales por satélite, operadoras de televisión por cable etc... . A menor bitrate y resolución, es aplicado para realizar SuperVideoCD , versión mejorada de VideoCD. Según se deduce de estas aplicaciones prácticamente todos los días usamos o vemos video en tecnología MPEG 2 actualmente.

- **MPEG-3**

Fue una propuesta de estándar para la TV de alta resolución, pero como se ha demostrado que MPEG2 con mayor ancho de banda cumple perfectamente con este cometido, se ha abandonado. Este estándar no debe de confundirse con el formato de sonido MP3.

- **MPEG-4 (ISO/IEC 14496)**

Esta siendo desarrollado para aplicaciones de comunicaciones multimedia, a muy bajo caudal de datos. Tiene bastante características comunes con las antiguas versiones de MPEG, aunque su uso esta orientado a Internet, ya no presenta un esquema de codificación en si, sino que centra sus miras en conseguir nuevas funcionalidades. Esta dividido en varias versiones distintas, la primera aprobada en 1998 y la segunda un año mas tarde, en la actualidad, el grupo trabaja en el desarrollo de nuevas versiones. Presenta una eficiente compresión para el almacenamiento y transmisión de datos audiovisuales respetando calidad de video y audio con bajas tasas de transmisión. Una aplicación típica de este estándar sería la transmisión de imágenes en movimiento por Internet, a bajos caudales, y su interoperabilidad con las aplicaciones del Web. Una de las cuestiones pendientes es la comunicación eficiente de vídeo y audio. Basados en el estándar MPEG-4 están los famosos Codecs (codificador / decodificador ) DivX y el de código abierto Xvid , estos codecs se usan para comprimir video de alta calidad con muy bajo bitrate con unos resultados buenísimos.

- **MPEG-7 (ISO/IEC 15938)**

Nuevo formato estándar aprobado recientemente, que esta destinado a aplicaciones multimedia. Comúnmente llamado “Interfaz de Descripción Del Contenido”, es un estándar para describir datos de contenidos multimedia que pueda tener algún grado de interpretación del significado de esa información que puede ser accedido por cualquier dispositivo MPEG7 no esta enfocado a ninguna aplicación en concreto en vez de eso MPEG7 intentará estandarizar un amplio rango de aplicaciones. Nos permitirá en el futuro tener información sobre la propia información, se implantara en bases

de datos multimedia, guías de programación, selecciona automática de canales y todo tipo de aplicaciones visuales que con la implantación de este sistema iran mucho mas adelante, El formato MPEG7 esta asociado directamente con los contenidos audiovisuales comprimidos a través de MPEG1(almacena y descarga archivos de video ) y MPEG4 (codifica el audio y video en forma de objetos, pero ha sido diseñado para que sea independiente del formato del contenido.

- **MPEG-21 (ISO/IEC 18034)**

Proyecto a muy largo plazo en actual desarrollo que propone la creación de un “Multimedia Framework” un marco tecnológico para todo el proceso multimedia.

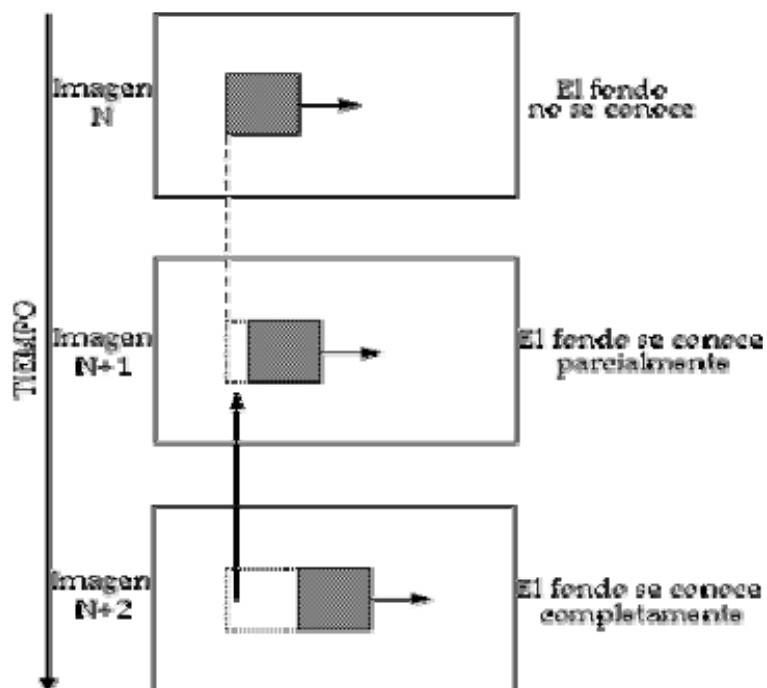
	<b>MPEG-1</b>	<b>MPEG-2</b>	<b>MPEG-4</b>
<b>estándar adquirido desde</b>	<b>1992</b>	<b>1995</b>	<b>1999</b>
<b>Max resolución de video</b>	<b>352x288</b>	<b>1920x1152</b>	<b>720x576</b>
<b>resolución PAL</b>	<b>352x288</b>	<b>720x576</b>	<b>720x576</b>
<b>resolución NTSC</b>	<b>352x288</b>	<b>640x480</b>	<b>640x480</b>
<b>Rango Máximo de Frecuencias de audio</b>	<b>48Khz</b>	<b>96Khz</b>	<b>96Khz</b>
<b>Numero máximo de canales de audio</b>	<b>2</b>	<b>8</b>	<b>8</b>
<b>Transferencia de datos máxima</b>	<b>3Mb/Seg</b>	<b>80Mb/Seg</b>	<b>5-10Mb/Seg</b>
<b>Transferencia habitual</b>	<b>1380Kbits (352x288)</b>	<b>6500Kbits (720x576)</b>	<b>880Kbits (720x576)</b>

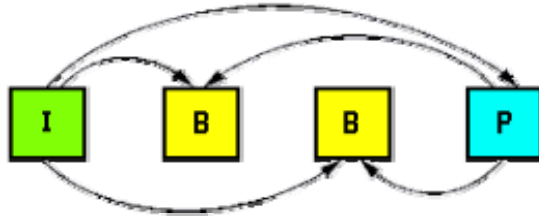
<b>Cuadros por segundo (PAL)</b>	<b>25</b>	<b>25</b>	<b>25</b>
<b>Cuadros por segundo (NTSC)</b>	<b>30</b>	<b>30</b>	<b>30</b>
<b>Calidad de video</b>	<b>Satisfactorio</b>	<b>Muy Bueno</b>	<b>Excelente</b>
<b>Requerimientos de Hardware para compresión</b>	<b>Bajo</b>	<b>Alto</b>	<b>Muy Alto</b>
<b>Requerimientos de Hardware para descompresión</b>	<b>Muy Bajo</b>	<b>Mediano</b>	<b>Alto</b>

## **MPEG-1**

- El estándar MPEG-1 se diseñó para codificar imágenes en movimiento y el audio asociado a estas para hacer uso de ellas mediante un medio de almacenamiento digital cuya velocidad de transmisión ande alrededor de 1,5 Mbit/s.
- El Mpeg-1 tiene tres partes: audio, video y sistema, que integra los otros dos. Los codificadores de audio y video funcionan independientemente, por lo tanto, se ven obligados a sincronizarse las 2 cadenas en el receptor. Esto se resuelve teniendo un reloj de sistema de 90 Khz. que proporciona el valor de tiempo a los dos codificadores. Estas marcas de tiempo se incluyen en la salida codificada y se propagan hasta el receptor, que las utiliza para sincronizar las corrientes de audio y video.
- Para comprimir las secuencias de video hace falta aprovechar las redundancias. En el estándar MPEG-1 se aprovechan la redundancia espacial y la redundancia temporal.
- La redundancia espacial puede utilizarse simplemente codificando por separado cada marco mediante JPEG.
- Para la redundancia temporal, MPEG-1 utiliza la compensación de movimiento, a partir de una imagen con compresión intracuadro (I). Calcula para la siguiente imagen P cuál es el factor de desplazamiento que la diferencia con respecto a la imagen anterior. Y entonces, se codifica el vector desplazamiento que nos indica cuánto ha variado la imagen con respecto al frame precedente.
- Sin embargo, esto no se puede aplicar a las escenas rápidas o con muchos cambios de cámara. Se necesita una manera de compensar este movimiento. Esto es precisamente lo que hace el MPEG.
- Pero MPEG no sólo actúa en una dirección temporal, sino que utiliza predicción bidireccional, prediciendo tanto las imágenes que han sucedido como las que ya han sucedido. Se despliega en sus predicciones tanto en el pasado como en el futuro. Es decir, que hay macrobloques que están predecidos por los frames que le seguirán, no sólo por los que le preceden.

- Esto permite que haya imágenes que sólo realizan predicción hacia delante, otras que sólo utilizan predicción hacia atrás y otras que resultantes de predicciones tanto hacia delante como hacia atrás.
- Para ello, MPEG define 3 tipos de imagen:
  - Imágenes I: con codificación espacial intracuadro, que supone mayor información y menos compresión. Siempre inician una secuencia y sirven de referencia para las imágenes P y B.
  - Imágenes P: con codificación intercuadro con predicción, que pueden tomar como referencia una imagen I o P. El codificador compara la imagen de referencia con la de predicción y sólo codifica el error de predicción.
  - Imágenes B: con codificación intercuadro con predicción bidireccional, que pueden tomar por referencia imágenes I o P que sean la anterior y posterior en el tiempo. Sirven para codificar elementos en movimiento y movimientos de cámara, en la que un elemento se desplaza a lo largo del cuadro. La información de un frame podría derivarse de la contenida en los frames anterior y posterior (una panorámica, por ejemplo).





- MPEG define una secuencia GOP (Group Of Pictures) de un tamaño típico, y nunca superior, a 12 imágenes. Son secuencias de imágenes I, P y B. Si existiera un cambio de plano, el codificador detectaría la falta de coincidencias con la secuencia anterior y comenzaría una secuencia nueva. Así logra evitar errores debidos a variaciones de la imagen.
- Un GOP se caracteriza por dos parámetros: M, el valor de imágenes comprendidas entre dos imágenes I. N, el valor comprendido entre una imágenes I o P y la siguiente que sea I o P.
- El error de predicción compara pequeñas áreas de la imagen: macrobloques de 16 x 16 pixels, detectando en la codificación P el desplazamiento del macrobloque de un frame a otro y calculando el correspondiente vector de desplazamiento, o codificando además la diferencia cuando exista diferencia en el macrobloque de un frame a otro. Si la diferencia es grande, el codificador considera que el macrobloque no se encuentra ya en la imagen, y entonces codifica el macrobloque con codificación espacial.
- El proceso completo MPEG consiste en:
  - Dividir la imagen en bloques (8 x 8) y macrobloques (4 bloques 2X2).
  - Procede a la codificación intracuadro (espacial). Para ello, elimina el campo par, submuestra a 4:2:0 (con lo que obtiene 4 bloques de Y y uno de cada componente Cb y Cr de color), aplica DCT, cuantifica y comprime VLC y RLC. Esto nos da las imágenes I.
  - Procede a la codificación intercuadro (temporal), aprovechando la redundancia temporal, realizando predicciones hacia delante (imágenes P) o bidireccionales (imágenes B).

- La compresión de audio Mpeg se hace muestreando la forma de onda a 32kHz, 44.1kHz o 48kHz. Se puede manejar sonido mono y estereo. El algoritmo de codificación se divide en tres capas distintas. La capa I es el esquema básico, se usa, por ejemplo, en el sistema de cinta digital (DCC). La capa II agrega asignación de bits avanzada al esquema básico. Se usa para audio de CD-ROM. La capa III agrega filtros híbridos y otras técnicas avanzadas.
- La compresión de audio se logra mediante una transformación de Fourier con la señal de audio para transformarla del dominio del tiempo al dominio de la frecuencia. El espectro resultante se divide en 32 bandas de frecuencia, que se procesan de forma independiente. Cuando existen 2 canales estéreo también puede aprovecharse la redundancia que existe por tener dos fuentes de audio muy similares. La calidad del sonido se puede ajustar desde 32Kbps hasta 448Kbps.
- MPEG-1, con un flujo binario de 1,5 Mbits./Seg. es una norma que se utiliza para el almacenamiento en CD-ROM. La norma MPEG-2, con un flujo binario de 2 a 40 Mbits/seg. permite una calidad profesional de televisión.

## **MPEG-2**

- La codificación MPEG-2 es parecida en lo fundamental a la codificación MPEG-1, es más, MPEG-2 puede describirse como una “aplicación” de compresión más compleja que MPEG-1, por ello, es considerada como una unidad superior: en efecto, toma todas las herramientas anteriores añadiéndole otras. Además, la norma prevé la compatibilidad ascendente, lo que significa que un decodificador MPEG-2 deberá decodificar MPEG-1.

### **Perfiles y Niveles**

- MPEG-2 funciona definiendo una serie de Perfiles y Niveles, cuya combinación define la capacidad de un codificador para manejar un flujo binario. Eso supone diferentes combinaciones para emisión, postproducción, captura, etc. Pensado para ser utilizado en un amplio rango de aplicaciones, requiriendo cada una un nivel de complejidad distinto, según el perfil y dentro de cada perfil los niveles en que se subdividen.
  - Los Perfiles se refieren a la forma en que se comprimen y codifican las imágenes. Definen el grado de complejidad de la codificación (resolución de color y la escalabilidad de la corriente de datos).
  - Los Niveles definen máximos y mínimos de resolución de la imagen, tamaño de la imagen, muestras por segundo de la luminancia y máximo flujo binario por Perfil. Van de Alta Definición hasta calidad similar a VHS.
- En un principio, hay 24 combinaciones posibles, pero no todas están definidas. Sólo hay 11, llamadas MPEG-2 Conformance Points.
- Un codificador MPEG cuando entrega un perfil y un nivel determinado, debe además ser capaz de decodificarlo a perfiles y niveles inferiores.
- El Nivel Principal del Perfil Principal es el MP@ML, supone un estándar de calidad satisfactorio, aunque se aleja de las necesidades exigidas profesionalmente, ya que

no permite un flujo binario superior a 15 Mb/s. Esto determinará la creación de una nueva configuración, la variante MPEG 4:2:2P@ML.

		PERFILES					
		Simple	Principal	4:2:2	SNR	Espacial	Alto
NIVELES	Alto		4:2:0 1920 x 1152 80Mb/s				4:2:0 o 4:2:2 1920 x 1152 100Mb/s
	Alto 1440		4:2:0 1440 x 1152 60Mb/s			1:2:0 1440 x 1152 60Mb/s	4:2:0 o 1:2:2 1440 x 1152 80Mb/s
	Principal	4:2:0 720 x 576 15Mb/s Sin B	4:2:0 720 x 576 15Mb/s	4:2:2 720 x 608 50Mb/s	4:2:0 720 x 576 15Mb/s		4:2:0 o 1:2:2 720 x 576 20 Mb/s
	Bajo		4:2:0 352 x 288 4Mb/s		4:2:0 352 x 288 4Mb/s		

- En la configuración 4:2:2MP@ML se permite la multigeneración así como operaciones de postproducción que en MP@ML no serían posible. Para ello:
  - No se submuestra la crominancia.
  - Se aumenta el número de líneas en 32
  - Se aumenta la frecuencia binaria a 50 Mb/s. En este caso, los GOP se reducen a imágenes de tipo I o secuencias de I-B. Para 30 Mb/s los GOPs corresponden a secuencias I-B. Si bajamos a 20 Mb/s. es necesario emplear GOPs más largos (I-B-B-P, por ejemplo), con lo que se reduce la calidad si trabajamos en multigeneración. Esto significa que si podemos adquirir las imágenes a 20 Mb/s., es necesario elevar el flujo binario a 30 o 50 Mb/s. para trabajar en multigeneración.
  - Existe compatibilidad hacia atrás con el nivel principal.

## Escalabilidad

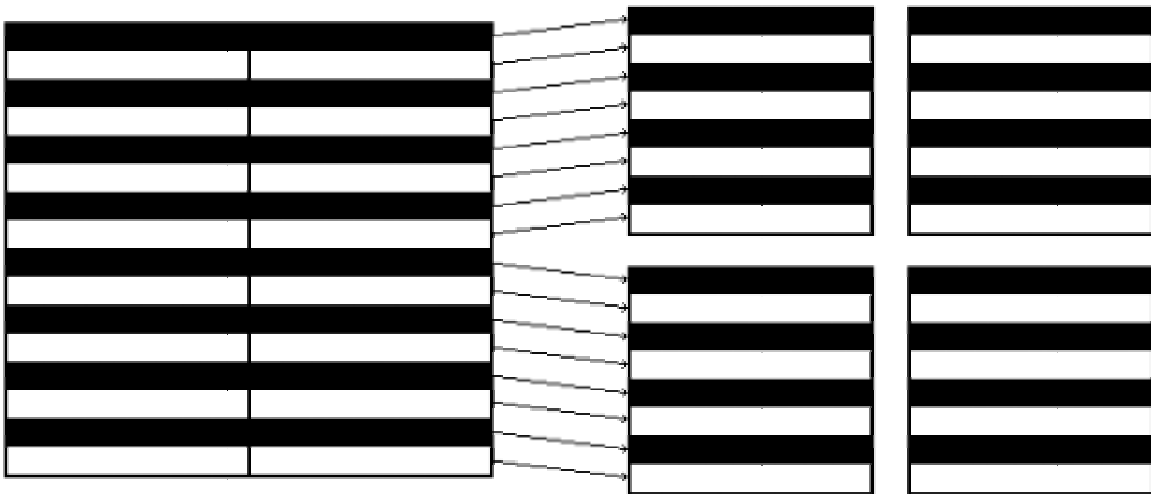
- El video bajo MPEG-2 es escalable. Es decir, es capaz de ser decodificado a baja calidad, ampliando a alta calidad, según la capacidad del receptor. Para ello, trabaja con una capa base, en la que se transmite la información básica acerca de la imagen, y con una o más capas de realce, en la que se separan los valores de detalle de la imagen. La capa base transmite la información básica de la imagen, aunque con un margen de ruido o de falta de definición, con lo que no tiene un flujo binario elevado. Las capas de realce transmiten los detalles de la imagen. Un decodificador de calidad limitada recibirá la capa base y no apreciará las capas de realce. Esta capa base se protege más que las de realce, para que la información básica de la imagen siempre llegue, aunque no lo haga con una condición de calidad máxima.
- Un decodificador con mayor potencia, será capaz de sumar la información de las capas de realce a las de la capa básica, reconstruyendo así una imagen con mayor detalle y menor ruido.
- Hay dos formas de crear esa escalabilidad: la SNR y la espacial.
- En la escalabilidad SNR (Relación Señal/Ruido) a partir de una imagen de HDTV se recuantifican de forma fuerte los valores AC, con lo que se obtiene una imagen con un gran ruido. Ésta será la capa base. Luego, se resta de la imagen HDTV esta capa de gran ruido, formándose así la capa de realce, que se codifica y transmite por separado.
- En la escalabilidad espacial, se obtiene la capa base a partir de la consideración de las frecuencias bajas de la imagen HDTV. La capa de realce se obtiene también por la diferencia entre esas dos imágenes, codificándose como señal de ayuda.
- La escalabilidad se puede referir a la relación señal/ruido, a la resolución de la luminancia, de la crominancia o la temporal.

## El multiplex MPEG-2

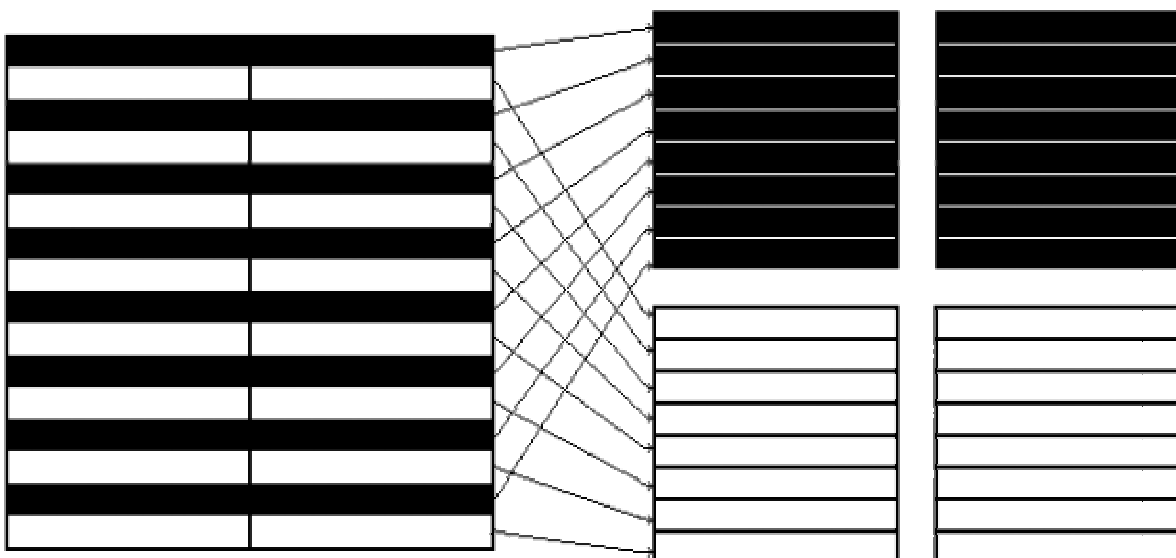
- La estructura del multiplex MPEG-2 se estructura así:
  - parte de **bloques** de 8x8.
  - Estos se agrupan en macrobloques, que serán de 4 bloques de luminancia y bloques de crominancia en relación con el grado de muestreo (4:4:4, 4 luminancia, 4 Cr y 4 Cb / 4:2:2, 4 luminancia, 2 Cr y 2 Cb / 4:2:0, 4 luminancia, 1 Cr y 1 Cb.)
  - Los macrobloques se agrupan en segmentos, que irán desde 1 macrobloque a una línea, y constituyen la información de direccionamiento intarcuadro.
  - Los segmentos en imágenes, en los que hay información temporal, y que pueden ser I, P y B.
  - Las imágenes en GOP, determinadas por los parámetros M (información de imágenes entre dos I) y N (información de imágenes entre dos I ó P). Los GOP pueden ser cerrados (si las imágenes que contienen no dependen de otro GOP) o abiertos (si alguna imagen depende de un GOP externos).
  - Finalmente, los GOP se agrupan en secuencias, cada una de las cuáles tienen su propio código de principio y final, así como datos que definen el tamaño de la imagen, las frecuencias de muestreo y las matrices de cuantificación.
- Cada uno de estos elementos dispone de su propia cabecera y datos de identificación adicionales.

## Predicción basada en campo y cuadro

- Existen dos formas de codificación, el modo DCT campo y el DCT cuadro. Una imagen puede estar codificada campo a campo o por cuadros. Al contrario que en MPEG-1, donde no existía entrelazado, el sistema codificará de una u otra manera según las características de la imagen. Una en que no haya mucho movimiento, y se requiera más detalle, se codificará como DCT-CUADRO. Mientras que si hay movimiento, será más conveniente la DCT-CAMPO.
- Naturalmente, una imagen muestreada en 4:2:0 sólo podrá ser codificada en DCT-CUADRO. Mientras que una 4:2:2 lo podrá ser de una u otra manera.



División de los macrobloques en modo CUADRO

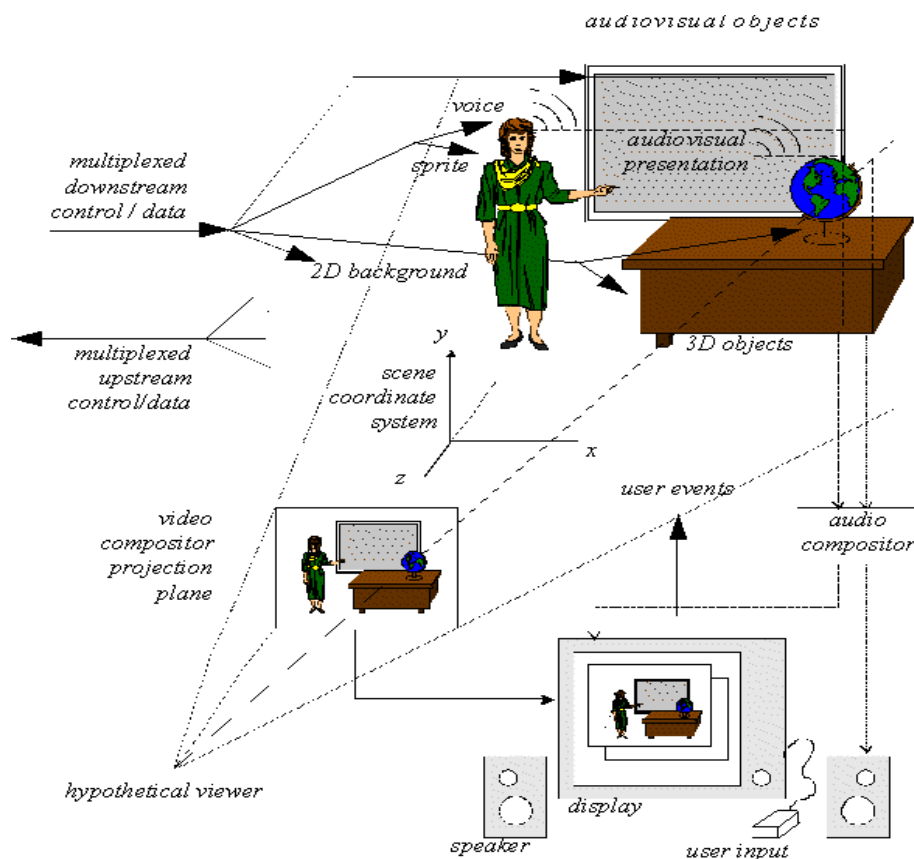


### **División de los macrobloques en bloques en modo campo (field)**

- La predicción también puede afectar a frames (progresivo) o a campos (entrelazado).
  - Una imagen I puede constar realmente de dos campos I, o bien estar compuesta de una imagen I y una P.
  - Una imagen P puede estar formada por una imagen P o por un par de campos P.
  - Una imagen B puede ser una imagen secuencial B o dos frames B.

## **MPEG-4**

- El MPEG-4 se ha pensado para la interactividad y la transmisión por Internet. En vez de basarse en el conjunto de la secuencia, el MPEG-4 se basa en el contenido. Así, mientras los estándares MPEG-1 y MPEG-2 codifican secuencias de imágenes, el MPEG-4 es capaz de crear representaciones codificadas de los datos de audio y vídeo que la forman, es decir MPEG-4 realiza una codificación basada en objetos. Define una escena audiovisual como una representación codificada de objetos audiovisuales que tienen cierta relación en el tiempo y en el espacio.
- Al igual que MPEG-1 y 2, MPEG-4 también realiza una representación por capas de la información. Pero cada cuadro de vídeo se segmenta en un número de objetos, llamados planos de objeto de vídeo (VOP). Sucesivos VOPs perteneciendo a un mismo objeto físico se denominan objetos de vídeo (VO). Un VO puede ser visto en MPEG-4 como un GOP en los estándares MPEG-1 y MPEG-2. Ejemplos de objetos audiovisuales son, objetos de Video (un jarrón, una persona hablando sin el fondo), objetos de audio( voz, música de fondo). La figura explica la manera en la cual una escena audiovisual es descrita como objetos individuales, la imagen contiene objetos compuestos que agrupan otros objetos, por ejemplo el objeto visual que corresponde a la persona que esta hablando y su correspondiente voz son unidos para formar un nuevo objeto compuesto que contiene los objetos componentes de la persona que habla.



- Los objetos tienen ciertas propiedades adjuntas como sus coordenadas espaciales, escala, localización, zoom, rotación etc... El usuario puede reconstruir la secuencia original al decodificar todas las capas de objetos y visualizándolos sin modificar los parámetros asociados a los objetos. O bien el usuario podría manipular la imagen realizando operaciones sencillas sobre los parámetros de los objetos, como moverlos, escalarlos etc... según haya planeado el autor de la secuencia. Es decir el sistema de objetos proporciona una interactividad impresionante.
- El flujo de datos de MPEG-4 permite mucha variación, incluso flujo de datos variable, lo cual lo hace interesante para la transmisión de video por Internet, puede ir desde los 9600 Bits/s hasta los 5 Mb/s. En MPEG1 y 2 se utilizan secciones cuadradas de la imagen para definir las partes en movimiento, pero MPEG4 con su sistema de objetos puede separarlos de las otras partes que definen la secuencia para realizar la codificación.
- En cuanto a la codificación de los datos MPEG-4 guarda muchas similitudes con MPEG-1 y MPEG-2, tal como la compresión basada en la DCT (Discreet Cosine

Transformation) con frames I- (keyframe), P- (predictive), y B-frames (bidireccional) aunque MPEG-4 ofrece mejores características a bajos flujos de datos.

- MPEG-4 tiene varias versiones. La versión 1, cuya especificación fue aprobada en octubre de 1.998; y la versión 2, aprobada en diciembre de 1.999. La versión 2 añade una serie de mejoras y ampliaciones al MPEG-4, pero la mayoría de las implementaciones soportan sólo la versión 1.
- Los famosos codecs de video basados en MPEG-4 que se usan en la actualidad como por ejemplo Divx, Xvid , QuickTime solo utilizan los perfiles mas básicos.
- Otra característica sorprendente de MPEG-4 es el MPEG-J, una librería de Java para controlar MPEG-4. Combinando ambos, es posible crear aplicaciones de la complejidad de Java dentro del entorno del MPEG-4 y, aún más complejo, el propio applet de Java puede estar “incrustado” dentro del flujo MPEG.
- La repercusión de MPEG-4 puede apreciarse teniendo en cuenta la cantidad de aplicaciones que se han anunciado basadas en él. Las industrias de TV digital planean sustituir la actual implementación del MPEG-2, por el MPEG-4 en un futuro no muy lejano ya que el MPEG-4 permitirá a la industria duplicar o triplicar el número de canales disponibles, y permitir la televisión interactiva (ITV).
- Como se puede ver, gracias al MPEG-4 se abre un gran abanico de posibilidades que irán creciendo a medida que se vayan creando nuevas herramientas para la creación y distribución de contenidos ya que actualmente las implementaciones de MPEG-4 son muy básicas y generalmente relacionadas con el perfil mas bajo del video, omitiendo sus grandes posibilidades de interacción.

## **MPEG-7**

- MPEG-7 se conoce también como MCDI (multimedia Content Description Interface), es un sistema de descripción y catalogación de los contenidos multimedia.
- Quiere construir un conjunto normalizado de descriptores, esquemas y un lenguaje estándar que pueda ser utilizado para describir los contenidos multimedia. Y no lo hará a través de la utilización exclusiva de texto, como lo hacen los sistemas de catalogación actuales, sino a través de parámetros multimedia.

## **MPEG-21**

- Nacido para afrontar el problema de la distribución de contenidos multimedia. MPEG-21 espera relacionar la forma en que los contenidos multimedia se relacionan entre sí, y cubrir los huecos que puedan quedar en la infraestructura según se desarrollen nuevas normas.
- Explora muchos aspectos muy diferentes, algunos de los cuáles se solapan entre sí, e intentará afrontar temas como el de la autenticidad, la calidad, los derechos de autor, del usuario u la privacidad.

## **Sistema de compresión DV**

- Se utiliza en los magnetoscopios DV, DVCAM, DVCPRO y DIGITAL-S. El sistema DV trabaja con la DCT y a nivel intracuadro. Se realiza un submuestreo 4:1:1 (en NSTC) y 4:2:0 (para PAL), excepto el DVCPRO, que trabaja en 4:1:1 en ambos sistemas.
  - DV, DVCAM, DVCPRO-25 trabajan a 25 Mb/s.
  - DVCPRO-50 y DIGITAL-S trabajan a 50 Mb/s, con lo que comprime la mitad menos que los anteriores.
  - El magnetoscopio en desarrollo DVCPRO-100 trabajará a 100 Mb/s, para ser utilizado en HDTV.
- Cuando trabajamos con imágenes entrelazadas, es posible procesar en campo (que trabaja en 2 bloques de 4x8, por lo que se llama 2-4-8) o en cuadro (que trabaja en 8x8, por lo que se llama 8-8), según predomine en la imagen información de movimiento o detalle espacial, respectivamente.
- Uno de los aspectos más destacables del DV es el proceso de “barajado de bloques intracuadro”, que mejora la calidad obtenible para cierta frecuencia binaria, y facilita la obtención de trenes binarios de frecuencia estable. Para ello, se forman macrobloques, de 4 bloques de luminancia, 1 de Cr y 1 de Cb. Se divide la imagen en superbloques, de forma que en PAL una frame está formado por un ancho horizontal de 5 SB x una altura vertical de 12 SB (en NTSC, 5 x 10).
- Y de ahí, se forman segmentos de video, formados por 5 macrobloques de diferentes zonas de la imagen (cada macrobloque tiene  $5 \times (4+1+1) \times 8 \times 8 = 1920$  Bytes), con lo que se espera igualar la dificultad de los segmentos. Todo esto es previo a la compresión.
- De esta manera, se puede hacer un cálculo previo del flujo de la imagen, antes de la compresión (al contrario que en MPEG, en que la tasa de compresión, la mayor o menor recuantificación de los valores AC, dependía del nivel alcanzado por el buffer en la compresión).

- Para ello, el análisis se realiza sobre los segmentos de video de 5 macrobloques (5 x (4+1+1) x 8x8 = 1920 Bytes) de manera que tras la compresión no excedan los segmentos un tamaño de 385 Bytes (con lo que se comprime en 5:1).
  - En DV existen 64 grados o tablas de recuantificación. Estos se dividen en 4 niveles, cada uno con 16 tablas. El nivel 1 corresponde a las frecuencias bajas, el 2 y 3 para las medias y el 4 para las altas frecuencias.
  
- Los pasos de la compresión DV son:
  - Se realiza el barajado de segmentos.
  - En cada zona de imagen se calcula la media de coeficientes AC (sin contar los DC), ya que estos son el grado de detalle de la imagen en esa zona. Según ese valor obtenido, se elige una de los 4 niveles para el segmento.
  - Una vez dentro del nivel, se elige una de las 16 tablas calculando con qué recuantificación el segmento es igual o no excede los 385 Bytes. Con ello, se garantiza una tasa binaria por cuadro estable.
  - Una vez realizada la compresión por recuantificación de los coeficientes AC de la DCT, se desbarajan los macrobloques. Se realiza la codificación VLC.
  - Tras el desbarajado y VLC, todos los macrobloques de una misma "fila" de SB se graban en una misma pista de la cinta magnética. Un frame tiene 12 pistas en PAL (10 en NTSC).

## AUDIO DIGITAL

### El sonido

- El sonido, físicamente, es un fenómeno debido a la transmisión de un frente de ondas en un medio elástico. La propagación de este frente se realiza a través de variaciones en la posición de las moléculas que conforman ese medio elástico debido a una variación de presión, humedad o temperatura. Cada molécula transmite esa variación de posición, esa oscilación, a las moléculas vecinas, creándose una onda sonora que se transmite de forma longitudinal con respecto al movimiento de las partículas.
- En términos sensoriales, el sonido es la respuesta a esas variaciones de presión cuando estimulan el oído.
- El sonido se caracteriza por una intensidad, un tono y un timbre.
- La intensidad, relacionada con la amplitud de la onda sonora, nos permite distinguir entre sonidos fuertes, medios o débiles, y nos permite percibir el volumen o potencia del sonido. Para el oído humano, el volumen se mide como comparación de un sonido en relación a otro, que se toma de referencia. Es una medida logarítmica: si se doblan los decibelios, eso significa que la potencia se ha elevado al cuadrado.
- El tono nos dice la frecuencia de ese sonido. Se mide como Hercios (ciclos por segundo). El oído humano percibe frecuencias entre 15 y 20.000 Hz, lo que da un ancho de banda de 20 KHz. Las frecuencias altas son percibidas como agudos, las bajas como sonidos graves.
- El timbre es la respuesta sensorial al número e intensidad de armónicos de la frecuencia fundamental. Cada fuente emisora tiene un timbre característico.
- Perceptivamente, la sonoridad de un sonido es algo subjetivo, que depende tanto de su intensidad como de su frecuencia. Para equilibrar eso, hablamos de niveles sonoros, líneas en que para obtener una misma respuesta sonora necesitamos diferentes decibelios para diferentes frecuencias. Cada nivel sonoro se mide en X fonos, y esta X corresponde a la cantidad de dB necesaria para producir la misma respuesta a 1 KHz.

- El rango dinámico del oído humano se limita en la parte alta por un nivel de dolor (habitualmente, 120 dB) y en la baja por el umbral de audición. El oído humano divide el espectro en una serie de 24 franjas independientes de resolución, de manera que la presencia de una frecuencia predominante dentro de una de estas franjas produce un fenómeno de enmascaramiento en el resto de las frecuencias de esta misma franja.

## Digitalización de la señal de audio

- La señal de sonido analógico se muestrea y cuantifica para obtener la PAM, la señal digitalizada.
- Las frecuencias digitales, según el criterio de Nysquist, deben superiores al doble del ancho de banda, es decir, a  $2 \times 20$  KHz. Se dan dos, 44,1 KHz, que es la primera que se utilizó, a partir de la grabación de sonido digital en el sistema de video analógico Umatic, a razón de 3 muestras de sonido por cada canal en cada una de las líneas de vídeo (empleándose 588 líneas en PAL y 490 en NTSC). Luego  $3 \times 588 \times 25 = 44.100$  KHz. (en NTSC  $3 \times 490 \times 30 = 41.100$ )
- Profesionalmente se prefiere la que introdujo el uso del CD: 48 KHz., no tanto por calidad, sino por compatibilidad con la frecuencia de campo de TV y por mantener una relación sencilla con 32 KHz., que fue la primera frecuencia de muestreo propuesta para radiodifusión digital.
- La cuantificación afecta al rango dinámico de la señal, y lo expresa como una palabra de  $n$  bites. Los valores posibles que puede tomar la señal serán  $2^n$ . Según el número de bits de cuantificación, la relación señal ruido SNR =  $6,02 n + 1,76$  dB. Para SNR  $\geq 90$  dB, que es una medida considerada aceptable de respuesta de sonido, vemos que para  $n = 16$ ,  $16 \times 6,02 + 1,76 = 98,1$  dB. 16 bits es una codificación adecuada.
- El error de cuantificación corresponde a  $\pm q/2$ , en que  $q$  es el intervalo mínimo de cuantificación. Este ruido es independiente del rango de amplitudes de la señal de entrada, sino del nivel de cuantificación, y su valor no es aleatorio, como ocurre con el ruido analógico. Es una forma de distorsión, que será tanto más molesto cuanto más bajo sea el nivel de la señal a codificar.

- Para compensar esta correlación entre el ruido de cuantificación y la señal, que hace que el ruido afecte más a las amplitudes más bajas, existe una forma diferente de cuantificación, la cuantificación diferencial y no lineal. Ésta es una cuantificación que es más sensible para valores de amplitud mayor (con lo que se reduce el ruido de cuantificación), mientras que para amplitudes mayores es menos sensible, ya al ser de mayor amplitud la señal la relación señal/ruido será mucho mayor, y el ruido inapreciable.
- Antes de digitalizar la señal analógica, se le suma un dither o ruido blanco, con una amplitud inferior a un nivel de cuantificación. Sirve para descorrelacionar de entrada de la frecuencia de muestreo y el error de cuantificación, y evitar que al muestrear la señal no exista en la señal PAM variación suficiente como para reconstruir la señal de entrada.
- Como forma de comunicación digital estandar entre equipos está establecido el interfaz AES/EBU, que permite también codificar a 20 y 24 bits, lo cuál da una SNR de 122,18 y 146,26 dB respectivamente (la de 16 bits es de 98 dB).
- El interfaz MADI (Multichannel Audio Digital Interface) sirve para proporcionar una norma de conexión entre los equipos de audio digital multicanal. Transporta hasta 56 canales de audio AES/EBU en serie por un cable coaxial a 125 Mb/s.

### **Compresión en audio digital**

- Una señal de calidad tiene un flujo binario de entre 0,75 y 1,15 Mb/s por canal.
- Profesionalmente, la compresión de audio digital tiene poco sentido, ya que, en relación con las altas tasas de flujo binario del video digital, la cantidad de información que mueve el audio digital es fácilmente soportable. Además, el audio comprimido daría muchos problemas para trabajos de postproducción.
- Para audio no profesional, la compresión permitirá aumentar la capacidad de almacenamiento de un soporte determinado y reducir la velocidad de transmisión necesaria. Además, podemos gracias a la compresión transmitir otra serie de datos.

Por ejemplo, transmitir o almacenar una señal multicanal 5.1 con el mismo flujo binario que una única señal de audio.

- Como problemas, la compresión puede afectar a la calidad. Requiere equipos de descompresión complejos y puede provocar latencias, retardos en el tiempo de acceso a la señal.
- La compresión elimina la redundancia de la señal. Y ésta puede ser:
  - Redundancia entrópica. La compresión la elimina sin que luego al descomprimirse se pierda información.
  - Redundancia perceptual, la compresión ocasionará una pérdida en la señal, pero ésta no será apreciable perceptualmente. Y ésta puede operar:
    - De forma que el oído humano sea insensible a la compresión.
    - De forma que la pérdida no sea notada por el oído humano ya que está enmascarada por el contexto.

## **Enmascaramientos**

- Existen una serie de fenómenos de enmascaramiento que pueden ser aprovechados para comprimir la señal, suprimiendo información enmascarada, por una parte; y por otra, aprovechar ese enmascaramiento para hacer no perceptible el ruido.
- Enmascaramiento frecuencial o simultáneo, aprovecha la división sensorial para el oído humano del espectro en una serie de 24 franjas independientes de resolución.
  - La presencia de una frecuencia predominante dentro de una de estas franjas produce un fenómeno de enmascaramiento en el resto de las frecuencias de esta misma franja, siempre que el nivel sonoro del sonido enmascarador en relación al enmascarado sea superior a cierto umbral.
  - Un ruido, no coherente y con un ancho de banda amplio, necesita menor proporción para enmascarar. Un tono puro en cambio para enmascarar necesita una mayor proporción.

- El enmascaramiento aumenta cuanto mayor sea el nivel de presión absoluta para frecuencias mayores a la enmascaradota, mientras que para las frecuencias inferiores se mantiene constante.
- El enmascaramiento aumenta en cuanto a espectro de frecuencias enmascaradas cuanto mayor sea la frecuencia enmascaradota.
- Enmascaramiento temporal. Somos incapaces de oír sonidos que se produzcan un poco antes y un poco después de un sonido intenso (preenmascaramiento y postenmascaramiento, respectivamente). El intervalo de postenmascaramiento (para tonos puros, entre 50 y 200 ms.) es muy superior al de preenmascaramiento (unas diez veces más pequeño).
- Entre otros efectos que son aprovechados por los sistemas de compresión destaca la irrelevancia estereofónica, que permite no considerar aquella información transmitida por el sistema de sonido estereofónico que es irrelevante para la localización espacial del sonido, con lo que no contribuye a la estereofonía, y puede ser transmitida como una única información, reduciendo la información a la mitad.

## **Compresión de audio**

- La forma en que trabaja la compresión de audio es la siguiente:
  - Se divide la señal en el tiempo, en una serie de frames, que duran unos pocos milisegundos.
  - Estos frames se dividen en subbandas de frecuencia mediante su transformación del espacio de tiempos al de frecuencias.
  - Se determina en cada subbanda el valor mayor, y se escalan los datos relativos a esa subbanda, de forma que el máximo corresponda a ese valor mayor.

- Se aprovecha el efecto de enmascaramiento, de forma que de acuerdo con éste recuantificamos los datos al número de bits correspondiente al umbral creado por ese enmascaramiento.
- Transformar el espectro de audio desde el dominio del tiempo al de las frecuencias, con técnicas similares a las utilizadas en video, nos permite trabajar con no las 32 bandas de audio sino con una cantidad superior, hasta 1024, y eso nos permitirá cuantificar con menos datos una banda de audio.

### **Sistemas de Compresión de Audio Digital**

- No parece que haya posibilidad de unificar los sistemas de compresión. Desde hace ya años se puede decir que hay dos predominantes: los relacionados con el sistema MPEG, de desarrollo predominantemente europeo, y que son la base del DAB (sistema de radio digital implantado en Europa), y los AC-3 de DOLBY, extendidos en la ATSC.
- Los sistemas de compresión MPEG, tanto en la norma MPEG-1 y MPEG-2, definen una serie de capas, Capa 1, 2 y 3 (análogas a los perfiles, de orden creciente y compatibles entre sí). Cada capa supone una capacidad de compresión mayor, una mayor sensibilidad a la hora de aplicar compresión perceptual y una mayor capacidad para lograr compresiones de mayor calidad. La capa 3 de MPEG-1 es la base del sistema MP3, que logra gran calidad con una compresión casi transparente. El sistema MPEG-2 aporta aún mayores garantías en la compresión para lograr mayor calidad. Sin embargo, al ser superado por sistemas como el AC-3 quedó desbancado dentro de la norma MPEG-2 por el sistema AAC, un sistema que sin embargo no era compatible con las normas previas.
  - El sistema AAC (Advance Audio Coding), mal llamado MP4, supone un avance aún mayor sobre la norma MPEG-2, ya que proporciona alta calidad a frecuencias de 96 Kb/S.. Utiliza para ello un sistema de perfiles, igual que en video: Main Profile, Low Complexity Profile y Scalable Simple Rate

Profile. Utiliza los mismos principios básicos de codificación que el MPEG-1, añadiendo herramientas que permiten mayor compresión y calidad.

- El sistema AC-3 supone un avance sobre los sistemas AC-1 y AC-2 creados por Dolby para dar sonido estéreo a la HDTV, obtenido por matrización de un sistema multicanal (4-2-4). En AC-3 se evitan las limitaciones impuestas por el matrizado de canal, y se envían varios canales discretos. Codifica 5 canales de espectro completo más uno más de realce de bajas frecuencias.