

U- 345 TESIS DOCTORAL

22-12-1998

Fínicas

T.D.

345

---

11 SET. 1998	
Nº	92120
HORA	
OFICINA AUXILIAR NÚM. 16	

**Contribuciones al soporte de transmisión  
de vídeo sobre redes ATM con servicio  
ABR, utilizando técnicas de compresión  
basadas en criterios visuales**

Santiago Felici Castell

Departamento de Informática y Electrónica  
Universitat de València

UMI Number: U607788

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U607788

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

UNIVERSITAT DE VALÈNCIA

FACULTAD DE CIENCIAS FÍSICAS

Reunit el Tribunal que subscriu, en el dia de la data,  
acordant d'atorgar, per unanimitat, a aquesta Tesi Doctoral  
d'En/ Nbr N° SANTIAGO FELICI CASTELL  
la qualificació de SOBRESALIENTE CUM LAUDE

València a 22 de DESEMBRE de 1978

El Secretari,

El President,



A large, stylized handwritten signature in blue ink, positioned to the right of the 'El Secretari,' text.

UNIVERSITAT DE VALÈNCIA  
BIBLIOTECA CIÈNCIES

Físicas

Nº Registre 13109

DATA 21-1-99

SIGNATURA T.D. 345

Nº LIBIS: j 2002725

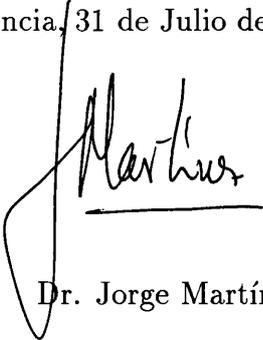
29'5 cm

D. JORGE MARTINEZ BAUSET, PROFESOR TITULAR DE ESCUELA UNIVERSITARIA DEL AREA INGENIERIA TELEMATICA EN LA UNIVERSIDAD IPOLITECNICA DE VALENCIA Y D. GREGORIO MARTIN QUETGLAS, (CATEDRATICO DEL AREA DE CIENCIAS DE LA COMPUTACION E INTELIGENCIA ARTIFICIAL EN LA UNIVERSITAT DE VALENCIA,

CERTIFICAN: que la presente TESIS DOCTORAL titulada "Contribuciones al soporte de transmisión de vídeo sobre redes ATM con servicio ABR, utilizando técnicas de compresión basadas en criterios visuales" ha sido realizada bajo nuestra dirección por D. Santiago Felici Castell, ingeniero en Telecomunicaciones

y para que conste a todos los efectos oportunos, se extiende la presente certificación, en lugar y fecha indicados.

Valencia, 31 de Julio de 1998



Fdo: Dr. Jorge Martínez Bauset



Fdo: Dr. Gregorio Martín Quetglás







# CONTRIBUCIONES AL SOPORTE DE TRANSMISIÓN DE VÍDEO SOBRE REDES ATM CON SERVICIO ABR, UTILIZANDO TÉCNICAS DE COMPRESIÓN BASADAS EN CRITERIOS VISUALES

MEMORIA PARA OPTAR AL GRADO DE DOCTOR EN INGENIERÍA DE  
TELECOMUNICACIONES  
DPTO. DE INFORMÁTICA Y ELECTRÓNICA  
FACULTAD DE CIENCIAS FÍSICAS  
UNIVERSITAT DE VALÈNCIA

Director:

Dr. Jorge Martínez Bauset  
Dr. Gregorio Martín Quetglas

©Santiago Felici Castell  
29 de Julio de 1998



*A mi hermano Vicente,  
a mis padres y mi familia.  
A Begoña.*



# Indice General

Agradecimientos . . . . .	xiii
Abreviaturas . . . . .	xv
Resumen . . . . .	xix
<b>I Introducción</b>	<b>1</b>
<b>1 Punto de partida y objetivos</b>	<b>3</b>
1.1 Marco histórico y social . . . . .	3
1.2 Limitaciones e inconvenientes actuales de las aplicaciones multimedia . . . . .	6
1.3 Punto de partida . . . . .	7
1.4 Objetivos . . . . .	8
<b>II Revisión de la situación actual</b>	<b>11</b>
<b>2 Requisitos y caracterización de las redes de comunicación para el soporte de servicios multimedia.</b>	<b>15</b>
2.1 Introducción . . . . .	15
2.2 Las redes basadas en datagramas y los servicios integrados . . . . .	16
2.3 Las redes de circuito virtual y los servicios integrados . . . . .	21
2.4 Modelo IS . . . . .	25
2.5 Conclusión . . . . .	28
<b>3 Técnicas de compresión de vídeo</b>	<b>29</b>
3.1 Proceso de compresión. Redundancia de la información . . . . .	29
3.2 Descomposición de la señal . . . . .	31
3.3 Estimación del movimiento . . . . .	33
3.4 Estándares de compresión y transporte de vídeo . . . . .	35
3.5 Conclusión . . . . .	42
<b>4 Codificación en multiresolución. Codificación subbanda</b>	<b>43</b>
4.1 Introducción . . . . .	43
4.2 Transformada Wavelet . . . . .	45
4.3 Multiresolución . . . . .	48
4.4 Transformada Wavelet Biortogonal . . . . .	53
4.5 Banco de Filtros y codificación subbanda . . . . .	54
4.6 Transformada wavelet multidimensional separable . . . . .	58
4.7 Conclusión . . . . .	61
<b>5 Sistema visual humano</b>	<b>63</b>
5.1 Descripción anatómica . . . . .	64
5.2 Modelos del SVH con descomposición subbanda . . . . .	69
5.3 Enmascaramiento del ruido o distorsión en el SVH . . . . .	71
5.4 Conclusión . . . . .	73

<b>III</b>	<b>Análisis y diseño del sistema propuesto</b>	<b>75</b>
<b>6</b>	<b>Análisis</b>	<b>79</b>
6.1	Requisitos del sistema propuesto . . . . .	79
6.2	Análisis del modelo fundamental del sistema . . . . .	82
6.3	Análisis del sistema de emisión . . . . .	83
6.4	Análisis del sistema de recepción . . . . .	89
6.5	Conclusión . . . . .	91
<b>7</b>	<b>Diseño</b>	<b>93</b>
7.1	Pautas de mínima negociación . . . . .	93
7.2	Diseño del sistema de emisión . . . . .	94
7.3	Diseño del sistema de recepción . . . . .	112
7.4	Conclusión . . . . .	113
<b>IV</b>	<b>Evaluación de prestaciones del sistema propuesto</b>	<b>115</b>
<b>8</b>	<b>Metodología de verificación</b>	<b>119</b>
8.1	Técnicas de verificación. Proceso de pruebas y ensayos . . . . .	119
8.2	Secuencias de vídeo bajo test . . . . .	120
8.3	Configuración de red: 2 conmutadores . . . . .	121
8.4	Patrones de tráfico . . . . .	123
8.5	Control de flujo: CAPAC . . . . .	125
8.6	Conclusión . . . . .	125
<b>9</b>	<b>Análisis de los algoritmos</b>	<b>127</b>
9.1	Estudio de las prioridades de las subbandas. Pérdidas de subbandas . . . . .	127
9.2	Asignación de bits. Estudio de bits por coeficiente . . . . .	128
9.3	Variación del factor de compresión . . . . .	129
9.4	Conclusión . . . . .	130
<b>10</b>	<b>Resultados experimentales</b>	<b>139</b>
10.1	Descripción de los parámetros introducidos para las simulaciones . . . . .	139
10.2	Estudio de retardos . . . . .	142
10.3	Patrón de tráfico de la fuente de vídeo . . . . .	143
10.4	Comportamiento del sistema bajo test . . . . .	144
10.5	Estadísticas de los resultados obtenidos . . . . .	157
10.6	Conclusión . . . . .	160
<b>V</b>	<b>Discusión</b>	<b>169</b>
<b>11</b>	<b>Discusión</b>	<b>171</b>
11.1	Decisiones tomadas y problemas resueltos . . . . .	171
11.2	Aportaciones . . . . .	173
11.3	Validación del sistema . . . . .	173
11.4	Discusión . . . . .	174
11.5	Conclusiones . . . . .	175
11.6	Líneas futuras . . . . .	177
<b>VI</b>	<b>Anexos</b>	<b>181</b>
<b>A</b>	<b>Estándares y sus definiciones</b>	<b>183</b>
<b>B</b>	<b>Test de secuencias codificadas en H.263</b>	<b>185</b>
B.1	Compresión con H.263 . . . . .	186
B.2	Transmisión con H.263 . . . . .	188

<b>C</b>	<b>Estudio de filtros wavelet</b>	<b>197</b>
C.1	Análisis temporal de los filtros . . . . .	197
C.2	Análisis frecuencial de los filtros . . . . .	197
C.3	Análisis multiresolución de los filtros . . . . .	197
<b>D</b>	<b>Estudio del estimador de MACR para el intervalo de descomposición</b>	<b>207</b>
<b>E</b>	<b>Histogramas de la descomposición 3D biortogonal de Miss América</b>	<b>211</b>
E.1	Histogramas de cada subbanda . . . . .	211
<b>F</b>	<b>Tabla de códigos con longitud variable</b>	<b>215</b>
F.1	Tabla ( <i>Run, Level</i> ) de códigos con longitud variable: signo 's': (0 o +) y (1 o -) . . . . .	215
<b>G</b>	<b>Configuración del sistema en el simulador DANAE</b>	<b>217</b>
<b>VII</b>	<b>Bibliografía</b>	<b>221</b>
	Publicaciones relacionadas con la tesis . . . . .	229
	<b>Indice de Materias</b>	<b>230</b>



# Indice de Figuras

1.1	<i>Mercado de la multimedia en Europa: billones de pesetas en ventas en los últimos 10 años</i>	3
1.2	<i>Esquema de aplicación de videoconferencia en red</i>	4
1.3	<i>Modelo TCP/IP</i>	5
2.1	<i>Arquitectura IP</i>	16
2.2	<i>Encapsulado de los protocolos RTP y RTCP en el protocolo de transporte UDP</i>	19
2.3	<i>Esquema de funcionamiento de los protocolos RTP y RTCP</i>	20
2.4	<i>Arquitectura de red ATM</i>	22
2.5	<i>AAL5: Capa de adaptación 5 en ATM</i>	23
2.6	<i>Protocolo de red para reserva de recursos</i>	25
2.7	<i>Servicio best effort: configuración sobre redes IP y ATM con servicio ABR</i>	27
3.1	<i>Esquema general de un codificador de vídeo</i>	30
3.2	<i>Imagen de Lisa</i>	31
3.3	<i>Fragmento de Lisa y su DCT correspondiente</i>	32
3.4	<i>Predicción con lazo cerrado: algoritmo DPCM</i>	33
3.5	<i>Secuencia de dos fotogramas en un intervalo de 160 ms y su diferencia asociada al movimiento de la escena</i>	33
3.6	<i>Análisis de correspondencia de bloques para la estimación de movimiento</i>	34
3.7	<i>Algoritmo híbrido DCT/Predicción-Compensación utilizado en compresión de vídeo</i>	35
3.8	<i>Estructura GOP de MPEG2: grupo de fotogramas</i>	38
3.9	<i>Estructura de datos en MPEG2: dependencias jerárquicas y sus relaciones</i>	38
4.1	<i>Descomposición espectral de subespacios para las técnicas de multiresolución</i>	51
4.2	<i>Descomposición con banco de filtros para 1D</i>	54
4.3	<i>Espectro del filtro paso-baja y paso-alta para multiresolución</i>	55
4.4	<i>Imagen original y su descomposición 2D con filtros biortogonales(2): residuo, detalle horizontal, vertical y diagonal</i>	59
4.5	<i>Banco de filtros para 3D y su descomposición en Fourier</i>	60
5.1	<i>El ojo humano</i>	63

5.2	<i>La retina humana . . . . .</i>	65
5.3	<i>Distribución de conos y bastones en la retina humana . . . . .</i>	65
5.4	<i>Campo receptor ON en la retina y CGL . . . . .</i>	66
5.5	<i>Campos receptores OFF en la retina y CGL . . . . .</i>	66
5.6	<i>Sistema visual humano . . . . .</i>	67
5.7	<i>Campos receptivos direccionales del cortex visual . . . . .</i>	68
5.8	<i>Función Gabor Par . . . . .</i>	70
5.9	<i>Función Gabor Impar . . . . .</i>	71
6.1	<i>Efecto bloque: imagen procesada mediante DCT con bloques de 8x8 pixels . . . . .</i>	80
6.2	<i>FFT de Lisa y ampliación del efecto de enventanado. . . . .</i>	80
6.3	<i>Lisa enventanda con una gaussiana y ampliación logarítmica de su dominio de Fourier . . . . .</i>	81
6.4	<i>Sistema de emisión . . . . .</i>	84
6.5	<i>Banco de filtros para 3D en multiresolución . . . . .</i>	85
6.6	<i>Descomposición de subbandas con 2 niveles de resolución . . . . .</i>	86
6.7	<i>Sistema de recepción . . . . .</i>	89
7.1	<i>Bloques del sistema de emisión . . . . .</i>	95
7.2	<i>Respuesta impulsional 2D biortogonal: paso baja y detalle horizontal . . . . .</i>	96
7.3	<i>Respuesta impulsional 2D biortogonal: detalle vertical y diagonal . . . . .</i>	96
7.4	<i>Ordenación en el tiempo según la obtención de las diferentes subbandas generadas en la descomposición 3D . . . . .</i>	98
7.5	<i>4 fotogramas de Miss América: unidad de descomposición para aplicar 2 niveles de resolución . . . . .</i>	98
7.6	<i>Salidas de los filtros temporales: paso baja parte 1 y 2, paso alta parte 1 y 2 . . . . .</i>	99
7.7	<i>Descomposición 2D parte 1, del paso bajo temporal 1ª resolución: residuo, detalle horizontal, vertical y diagonal . . . . .</i>	99
7.8	<i>Descomposición 2D parte 1, del paso alta temporal 1ª resolución: residuo, detalle horizontal, vertical y diagonal . . . . .</i>	99
7.9	<i>Descomposición 2D parte 2, del paso bajo temporal 1ª resolución: residuo, detalle horizontal, vertical y diagonal . . . . .</i>	100
7.10	<i>Descomposición 2D parte 2, del paso alta temporal 1ª resolución: residuo, detalle horizontal, vertical y diagonal . . . . .</i>	100
7.11	<i>Descomposición 2D del paso bajo temporal 2ª resolución: residuo, detalle horizontal, vertical y diagonal . . . . .</i>	100
7.12	<i>Descomposición 2D del paso alta temporal 2ª resolución: residuo, detalle horizontal, vertical y diagonal . . . . .</i>	101
7.13	<i>Predicción y diferencia de la subbanda paso baja final del segundo nivel de resolución . . . . .</i>	101
7.14	<i>Histograma de la predicción realizada para la subbanda paso baja final del segundo nivel de resolución . . . . .</i>	101

7.15	<i>Punto de funcionamiento para dos subbandas en las curvas de Rate-Distorsion</i> . . . . .	108
7.16	<i>Descripción de los campos de la cabecera añadida a la SDU para la identificación y proceso de cada subbanda</i> . . . . .	110
7.17	<i>Formato de paquete de adaptación AAL5+ en ATM-ABR</i> . . . . .	110
7.18	<i>Memoria FIFO para adaptación a la red</i> . . . . .	111
8.1	<i>Evolución de las varianzas asociadas a las 22 subbandas de la descomposición 3D durante 1600 ms en la secuencia de Miss América</i> . . . . .	123
8.2	<i>Configuración de red con dos conmutadores y enlace troncal cuello de botella</i> . . . . .	123
9.1	<i>Ejemplos de pérdidas de una subbanda en 2ª resolución: 1-residuo 2-detalle horizontal 3-vertical 4-diagonal</i> . . . . .	132
9.2	<i>Ejemplos de pérdidas de dos subbandas en 2ª resolución, de izquierda a derecha y de arriba a abajo: 1- residuo y detalle horizontal, 2-detalle vertical y diagonal, 3-residuo y detalle vertical, 4-detalle horizontal y diagonal, 5-residuo y detalle diagonal, 6-detalle horizontal y vertical</i> . . . . .	132
9.3	<i>Ejemplos de pérdidas de subbandas en 1ª resolución, de izquierda a derecha y de arriba a abajo: 1-detalle horizontal, 2-vertical, 3-diagonal, 4-detalle horizontal y vertical, 5-detalle vertical y diagonal, 6-detalle horizontal y diagonal</i> . . . . .	132
9.4	<i>Ejemplos de pérdidas de una subbanda en 2ª resolución: (primera fila) pérdida de detalle vertical de segunda resolución y pérdida de detalle vertical, diagonal y horizontal y detalle vertical de la 1ª, (segunda fila) pérdida de detalle horizontal de segunda resolución y pérdida de detalle vertical, diagonal y horizontal respectivamente de la primera resolución, (tercera fila) pérdida de detalle diagonal de segunda resolución y pérdida de detalle vertical, diagonal y horizontal respectivamente de la primera resolución</i> . . . . .	133
9.5	<i>Bpp para cada una de las 22 subbandas vs MACR neto con Miss América en formato QSIF</i> . . . . .	134
9.6	<i>Bpp para cada una de las 22 subbandas vs MACR neto con Miss América en formato SIF</i> . . . . .	134
9.7	<i>Celdas ATM para cada una de las 22 subbandas vs MACR neto con Miss América en formato QSIF</i> . . . . .	135
9.8	<i>Celdas ATM para cada una de las 22 subbandas vs MACR neto con Miss América en formato SIF</i> . . . . .	135
9.9	<i>Paso de cuantificación para cada una de las 22 subbandas vs MACR neto con Miss América en formato QSIF</i> . . . . .	136
9.10	<i>Paso de cuantificación para cada una de las 22 subbandas vs MACR neto con Miss América en formato SIF</i> . . . . .	136
9.11	<i>4 fotogramas de Miss América en formato SIF comprimida con 0.1278 bpp</i> . . . . .	137

9.12	<i>4 fotogramas de Miss América en formato SIF comprimida con 0.2424 bpp . . . . .</i>	137
9.13	<i>4 fotogramas de Miss América en formato SIF comprimida con 0.5303 bpp . . . . .</i>	137
9.14	<i>4 fotogramas de Miss América en formato SIF comprimida con 0.7575 bpp . . . . .</i>	137
9.15	<i>4 fotogramas de Miss América en formato SIF comprimida con 1.0606 bpp . . . . .</i>	138
9.16	<i>4 fotogramas de Miss América en formato SIF comprimida con 1.9696 bpp . . . . .</i>	138
9.17	<i>4 fotogramas de Miss América en formato SIF comprimida con 2.66 bpp</i>	138
10.1	<i>Evolución de la salida de celdas ATM para la secuencia de Miss América comprimida con 2.66 bpp para los formatos QSIF y SIF . . . . .</i>	144
10.2	<i>Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes persistentes: referencia 1 . . . . .</i>	148
10.3	<i>Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes persistentes: referencia 2 . . . . .</i>	148
10.4	<i>Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes persistentes: referencia 3 . . . . .</i>	148
10.5	<i>Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario MAN con fuentes persistentes: referencia 1 . . . . .</i>	153
10.6	<i>Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario MAN con fuentes persistentes: referencia 3 . . . . .</i>	153
10.7	<i>Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes ráfaga: referencia 1 . . . . .</i>	153
10.8	<i>Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes ráfaga: referencia 3 . . . . .</i>	161
10.9	<i>Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario MAN con fuentes ráfaga: caso 2 . . . . .</i>	161
10.10	<i>Escenario LAN con tráfico persistente: referencia 1 . . . . .</i>	162
10.11	<i>Escenario LAN con tráfico persistente: referencia 2 . . . . .</i>	163
10.12	<i>Escenario LAN con tráfico persistente: referencia 3 . . . . .</i>	163
10.13	<i>Escenario MAN con tráfico persistente: referencia 1 . . . . .</i>	164
10.14	<i>Escenario MAN con tráfico persistente: referencia 2 . . . . .</i>	165
10.15	<i>Escenario MAN con tráfico persistente: referencia 3 . . . . .</i>	165

10.16	<i>Escenario LAN con tráfico ráfaga: referencia 1</i>	166
10.17	<i>Escenario LAN con tráfico ráfaga: referencia 2</i>	166
10.18	<i>Escenario LAN con tráfico ráfaga: referencia 3</i>	167
10.19	<i>Escenario MAN con tráfico ráfaga: referencia 1</i>	167
10.20	<i>Escenario MAN con tráfico ráfaga: referencia 2</i>	167
10.21	<i>Escenario MAN con tráfico ráfaga: secuencia de vídeo de Salesman y Claire</i>	168
11.1	<i>Alternativa a AAL5+ en codificación subbanda</i>	177
11.2	<i>Ejemplos de reposición de subbanda paso baja perdida en 2ª resolución</i>	178
11.3	<i>Codificación de vídeo en subbanda para multipunto</i>	179
B.1	<i>Proyección de ensayos</i>	185
B.2	<i>Secuencia original de Salesman</i>	186
B.3	<i>Secuencia 4 de Salesman</i>	188
B.4	<i>Secuencia 5 de Salesman</i>	188
B.5	<i>Secuencia codificada H.263 de Miss América</i>	190
B.6	<i>Secuencia codificada H.263 de Claire</i>	191
B.7	<i>Secuencia Salesman con 582 errores, <math>P_e = 10^{-3}</math> BER = 0.001085</i>	191
B.8	<i>Secuencia Salesman con 321 errores, <math>P_e = 10^{-3}</math> BER = 0.000909</i>	192
B.9	<i>Secuencia Salesman con 48 errores, <math>P_e = 10^{-4}</math> BER = 0.000136</i>	192
B.10	<i>Secuencia Salesman con 11 errores, <math>P_e = 10^{-5}</math> BER = 0.000031</i>	193
B.11	<i>Secuencia Salesman con 7 errores, <math>P_e = 10^{-7}</math> BER = 0.000020</i>	193
B.12	<i>Secuencia Salesman con 14 errores, <math>P_e = 10^{-9}</math> BER = 0.000048</i>	193
B.13	<i>Secuencia Miss América con 747 errores, <math>P_e = 10^{-3}</math> BER = 0.001030</i>	194
B.14	<i>Secuencia Miss América con 26 errores, <math>P_e = 10^{-5}</math> BER = 0.000036</i>	194
B.15	<i>Secuencia Miss América con 20 errores, <math>P_e = 10^{-7}</math> BER = 0.000028</i>	194
B.16	<i>Secuencia Claire con 787 errores, <math>P_e = 10^{-3}</math> BER = 0.000989</i>	195
B.17	<i>Secuencia Claire con 103 errores, <math>P_e = 10^{-4}</math> BER = 0.000129</i>	195
B.18	<i>Secuencia Claire con 27 errores, <math>P_e = 10^{-5}</math> BER = 0.000034</i>	195
B.19	<i>Secuencia Claire con 24 errores, <math>P_e = 10^{-7}</math> BER = 0.000030</i>	196
C.1	<i>Respuesta temporal del filtro biortogonal(1)</i>	198
C.2	<i>Respuesta temporal del filtro biortogonal(2)</i>	198
C.3	<i>Respuesta frecuencial: Módulo Filtro biortogonal(1)</i>	199
C.4	<i>Respuesta frecuencial: Fase Filtro biortogonal(1)</i>	199
C.5	<i>Respuesta frecuencial: Módulo Filtro biortogonal(2)</i>	200
C.6	<i>Respuesta frecuencial: Fase Filtro biortogonal(2)</i>	200
C.7	<i>Respuesta frecuencial: Módulo Filtro Haar</i>	201
C.8	<i>Respuesta frecuencial: Fase Filtro Haar</i>	201
C.9	<i>Función de escala biortogonal(1) de 1ª resolución, análisis y síntesis</i>	202
C.10	<i>Función wavelet biortogonal(1) de 1ª resolución, análisis y síntesis</i>	202
C.11	<i>Función de escala biortogonal(1) de 2ª resolución, análisis y síntesis</i>	202
C.12	<i>Función wavelet biortogonal(1) de 2ª resolución, análisis y síntesis</i>	203
C.13	<i>Función de escala biortogonal(2) de 1ª resolución, análisis y síntesis</i>	203

C.14	<i>Función wavelet biortogonal(2) de 1ª resolución, análisis y síntesis . . .</i>	203
C.15	<i>Función de escala biortogonal(2) de 2ª resolución, análisis y síntesis . .</i>	204
C.16	<i>Función wavelet biortogonal(2) de 2ª resolución, análisis y síntesis . . .</i>	204
C.17	<i>Función de escala Haar de 1ª resolución, análisis y síntesis . . . . .</i>	204
C.18	<i>Función wavelet Haar de 1ª resolución, análisis y síntesis . . . . .</i>	205
C.19	<i>Función de escala Haar de 2ª resolución, análisis y síntesis . . . . .</i>	205
C.20	<i>Función wavelet Haar de 2ª resolución, análisis y síntesis . . . . .</i>	205
C.21	<i>Función de escala y síntesis Haar con 8 resoluciones . . . . .</i>	206
D.1	<i>Estimación de la tasa media de celdas permitida cuando el ACR es una señal cuadrada y respuesta en frecuencia del estimador para valores de <math>\alpha</math> de <math>\frac{1}{1024}</math>, 0.1, 0.2 y 0.3 . . . . .</i>	208
D.2	<i>Frecuencia digital de corte del filtro para la estimación del MACR a partir del ACR en función de <math>\alpha</math>, con ampliación de detalles para bajas frecuencias . . . . .</i>	209
E.1	<i>Histograma de la 1ª resolución, parte 1 . . . . .</i>	212
E.2	<i>Histograma de la 1ª resolución, parte 2 . . . . .</i>	213
E.3	<i>Histograma de la 2ª resolución excepto paso baja final . . . . .</i>	214
E.4	<i>Histograma del paso baja final de la 2ª resolución en DPCM . . . . .</i>	214
G.1	<i>Configuración de red con dos conmutadores y enlace troncal cuello de botella en el simulador DANAE . . . . .</i>	217
G.2	<i>Esquema de bloques del sistema de emisión en el simulador DANAE . .</i>	218
G.3	<i>Esquema de bloques del sistema de recepción en el simulador DANAE .</i>	220

# Indice de Tablas

1.1	<i>Crecimiento y previsiones de direcciones en Internet . . . . .</i>	5
2.1	<i>RSVP y ATM: correspondencia de servicios . . . . .</i>	26
5.1	<i>Organización jerárquica de células en SVH. . . . .</i>	68
7.1	<i>Formatos de imagen en vídeo . . . . .</i>	93
7.2	<i>Coefficientes de la respuesta impulsional del filtro 2D de aproximación .</i>	97
7.3	<i>Coefficientes de la respuesta impulsional del filtro 2D de detalle horizontal</i>	97
7.4	<i>Coefficientes de la respuesta impulsional del filtro 2D de detalle vertical</i>	97
7.5	<i>Coefficientes de la respuesta impulsional del filtro 2D de detalle diagonal</i>	98
8.1	<i>Varianza media de cada subbanda tras la descomposición 3D propuesta para las secuencias de Miss América, Salesman y Claire . . . . .</i>	122
8.2	<i>Distancias utilizadas para la configuración de red con dos conmutadores(2N) . . . . .</i>	124
9.1	<i>Pasos de cuantificación según los bits por pixel para cada una de las 22 subbandas de la descomposición 3D propuesta en Miss América con formato SIF . . . . .</i>	131
10.1	<i>Especificación de tasas de salida en celdas/ms para las fuentes ABR . .</i>	140
10.2	<i>Estimaciones de MACR (celdas/ms) en los instantes de la descomposición: diferentes casos con escenario LAN y fuentes persistentes . .</i>	149
10.3	<i>Probabilidades de subbandas transmitidas <math>p(T)</math> y recibidas correctamente condicionadas a ser transmitidas <math>p(A/T)</math>, para diferentes casos con escenario LAN y fuentes persistentes . . . . .</i>	150
10.4	<i>Estimaciones de MACR (celdas/ms) en los instantes de la descomposición: diferentes casos con escenario MAN y fuentes persistentes . .</i>	152
10.5	<i>Probabilidades de subbandas transmitidas <math>p(T)</math> y recibidas correctamente condicionadas a ser transmitidas <math>p(A/T)</math>, para diferentes casos con escenario MAN y fuentes persistentes . . . . .</i>	154
10.6	<i>Estimaciones de MACR (celdas/ms) en los instantes de la descomposición: diferentes casos con escenario LAN y fuentes ráfaga . . . . .</i>	155

10.7	<i>Probabilidades de subbandas transmitidas <math>p(T)</math> y recibidas correctamente condicionadas a ser transmitidas <math>p(A/T)</math>, para diferentes casos con escenario LAN y fuentes ráfaga . . . . .</i>	156
10.8	<i>Estimaciones de MACR (celdas/ms) en los instantes de la descomposición: diferentes casos con escenario MAN y fuentes ráfaga . . . . .</i>	157
10.9	<i>Probabilidades de subbandas transmitidas <math>p(T)</math> y recibidas correctamente condicionadas a ser transmitidas <math>p(A/T)</math>, para diferentes casos con escenario MAN y fuentes ráfaga . . . . .</i>	158
10.10	<i>Promediado de la probabilidad de recibir correcta una subbanda cuando ha sido transmitida (<math>p(A/T)</math>) y su intervalo de confianza . . . . .</i>	159
B.1	<i>Parámetros de entrada al compresor H.263 . . . . .</i>	187
B.2	<i>Valores medios de salida del compresor H.263 . . . . .</i>	187
B.3	<i>Resultados obtenidos de la compresión de Salesman . . . . .</i>	187
B.4	<i>Compresión con 150 fotogramas de Salesman modificando “s”y “q” . . . . .</i>	189
B.5	<i>Medios de transmisión y la probabilidad de error asociada . . . . .</i>	189
B.6	<i>Resultados de las secuencias de Salesman, Miss América y Claire comprimidas . . . . .</i>	190
F.1	<i>Tabla de códigos de longitud variable: s (0,1) signo (+,-) . . . . .</i>	215

## Agradecimientos

En primer lugar, agradezco el entusiasmo, el compañerismo, la metodología y la capacidad de trabajo de Baltasar Beferull, por continuar sin perder la ilusión, a mi amigo de batalla.

Agradezco la serenidad, motivación y coordinación de Dr. Gregorio Martín, director del LISITT e Institut de Robòtica, codirector de la tesis, sin duda una persona dedicada en cuerpo y alma al mundo universitario, a la docencia y la investigación y su proyección al mundo empresarial.

Agradezco la comprensión, la ayuda y cooperación de mi otro codirector y antiguo profesor Dr. Jorge Martínez de la Universidad Politècnica de Valencia, del Departamento de Comunicaciones, y la ayuda de sus colaboradores Vicente Traver, Enrique Ibáñez y Fernando Cifo.

Agradezco los diálogos y conversaciones siempre muy fructíferos con Rogelio Montañana del Centro de Cálculo de la Universitat, una persona excepcional.

Agradezco las ayudas y aclaraciones de Vicente Cavero, Guillermo Ayala, Ricardo Ferris y Jesús Malo, PDIs de la Universitat y a la gente que ha estado a mi alrededor, Jose Pelegrí y Juanjo Pérez compañeros de habitáculo. Quiero dar las gracias también a Dr. Fernando Pardo y Dr. Juan Domingo Esteve por su ayuda con  $\text{\LaTeX}$ .

Tengo que agradecer igualmente la colaboración de muchas personas de la Universidad de Valencia, en especial a la gente del LISITT, a Miguel Lozano, a Dr. Marcos Fernández y a Dr. Enrique Bonet, que con ellos todo ha sido más fácil.

Agradezco a Dr. Joan Pelechano, director y compañero del Departamento de Informática y Electrónica, por la buena labor, atenciones y gestión realizada.

No puedo tampoco olvidarme de mis alumnos, que al menos han sabido entender este esfuerzo, en ocasiones nada fácil. También agradezco a los administradores del sistema que mantienen la red y las estaciones de trabajo en perfecto estado de funcionamiento.

A mi hermano Vicente, a mi madre y mi padre, mi familia que han jugado también un papel fundamental para la culminación de este trabajo, por mantener siempre la sonrisa y la felicidad ante las dificultades de su hermano, su hijo, querido. Repito con todo el amor a un hermano, *gracias Vicente*.

Agradezco el tesón, ánimo, constancia de Begoña, parte de mi corazón, que ha sabido restablecer la alegría y mitigar el cansancio de mi ser.

A mis amigos de Algemés, mi ciudad natal, ese pueblo tan maravilloso de la Ribera Alta; también un agradecimiento a los amigos de Valencia.

Estoy agradecido con todos ellos, los que he nombrado y los que no he recordado al nombrar. A todos, gracias.

*Santi.*



## Abreviaturas

Siglas y abreviaturas más utilizadas:

- *AAL* ATM Adaptation Layer
- *ABR* Available Bit Rate
- *ACR* Allowed Cell rate
- *ADPCM* Adaptative Differential Pulse Code Modulation
- *ATM* Asynchronous Transfer Mode
- *BER* Bit Error Rate
- *BRM* Backward Resource Management cell
- *bpp* bits por pixel
- *CBR* Constant Bit Rate
- *CGL* Cuerpo Genuculado Lateral
- *CIF* Common Intermediate Format
- *CLP* Cell Loss Priority
- *CRC* Ciclyc Redundant Check
- *DCT* Discrete Cosine Transform
- *DPCM* Differential Pulse Code Modulation
- *EPD* Early Packet Discard
- *ER* Explicit Rate
- *FDDI* Fiber Data Distributed Interface
- *FFT* Fast Fourier Transform
- *FIFO* First In First Out
- *FRM* Forward Resource Management cell
- *GOB* Group Of Blocks
- *GOP* Group Of Pictures
- *ICR* Initial Cell Rate
- *IDCT* Inverse Discrete Cosine Transform
- *IETF* Internet Engineering Task Force
- *IP* Internet Protocol
- *IS* Integrated Services
- *ISO* International Organization for Standardization
- *ISOC* Internet Society

- *ITU* ex CCITT, International Telecommunication Union
- *ITU-T* International Telecommunication Union-Telecommunications Sector
- *HDTV* High Definition TeleVision
- *JND* Just Noticeble Distorsion
- *JPEG* Joint Pictures Experts Group
- *KL*T Karhuen Loeve Transform
- *LAN* Local Area Network
- *LOT* OverLap Transform
- *MACR* Mean Allowed Cell rate
- *MAD* Mean Absolute Difference
- *MB* Macro Bloque
- *MBone* Multicast Bone
- *MCR* Minimum Cell Rate
- *MIC* Modulación de impulsos codificados
- *MPEG* Moving Pictures Experts Group
- *MSE* Mean Square Error
- *NNI* Network to Network Interface
- *PABE* Porción de Ancho de Banda Equitativo
- *PCM* Pulse Code Modulation
- *PCR* Peak Cell Rate
- *PDU* Protocol Data Unit
- *PSNR* Peak Signal to Noise Ratio
- *QCIF* Quarter Common Intermediate Format
- *QoS* Quality Of Service
- *QSIF* Quarter Source Input Format
- *RDSI* Red Digital de Servicios Integrados; en inglés ISDN
- *RFC* Request for Comment
- *RM* Resource Management, celdas ATM
- *RTB* Red Telefónica Básica
- *RTC* Red Telefónica Conmutada
- *RTCP* Real Time Control Protocol
- *RTP* Real Time Protocol
- *RSVP* Resource reSerVation Protocol

- *SBC* Sub Band Coding
- *SDU* Service Data Unit
- *SIF* Source Input Format
- *SNR* Signal Noise Ratio
- *STR* Sistemas de Tiempo Real
- *SVH* Sistema Visual Humano
- *TCP* Transport Control Protocol
- *TIC* Tecnologías de la Información y la Comunicación
- *UDP* User Datagram Protocol
- *UNI* User Network Interface
- *VBR* Variable Bit Rate
- *VLSI* Very Large Scale Integration
- *WAN* Wide Area Network



## Resumen

Las Tecnologías de la Información y las Comunicaciones evolucionan a pasos agigantados. Es el momento en que mejores prestaciones y servicios, como demanda de una mejor calidad en las aplicaciones, son exigidas por los usuarios.

Un estudio de la evolución histórica de la telemática, evidencia el éxito de los protocolos IP (*Internet Protocol*) por su flexibilidad, simplicidad y filosofía (*'todos con todos'* y *'a nadie se le niega la entrada'*, etc.) por lo que se convierten en el medio de interconexión por excelencia. No es utópico pensar que todos los ordenadores del mundo formen una red de interconexión ofreciendo todo tipo de servicios, vistas las mejoras de IPv6.

Sin embargo la evolución de dichos protocolos y los servicios ofrecidos por estas redes basadas en datagramas, se ven frenados cuando las atraviesan diferentes tipos de tráfico, impensables en el momento de su diseño; es el caso del tráfico procedente de servicios integrados, voz y vídeo, propios de las aplicaciones multimedia y con claras restricciones tanto temporales como de pérdidas.

Las redes IP no tienen un comportamiento fiable frente a estos tipos de tráfico. Esta situación abre nuevas líneas de trabajo para superar estas deficiencias y solucionar los problemas que presentan.

Ha sido necesario la introducción de nuevos protocolos para dar soporte de tiempo real y reserva de recursos, como RTP, RTCP, RSVP etc. bajo el nuevo modelo de red para servicios integrados (modelo IS), con el fin de ofrecer en la medida de lo posible garantías de calidad, que les hace compartir y compatibilizar características básicas de las redes basadas en circuitos virtuales.

Vistas las anteriores consideraciones, el presente trabajo se centra en el estudio de la transmisión del vídeo digital proponiendo un sistema para entornos de red con servicios *best effort*, análogos a los servicios ABR en ATM, que integren controles de flujo cerrado basados en tasa explícita y negociación de mínima tasa de salida.

Una vez revisado el estado del arte en la compresión, llegamos a la conclusión de que sistemas como MPEG-2 y H.263 incluidos en la mayoría de los sistemas de videoconferencia actuales, no explotan en su totalidad el potencial de un criterio de calidad visual cuando el vídeo es degradado por los entornos de red.

A lo largo del trabajo, proponemos una alternativa diferente de compresión que permite aproximar mejor los criterios visuales en dichas situaciones de degradación de la señal. Para ello se estudia detalladamente la codificación subbanda y multiresolución, mediante transformadas localizadas tanto en espacio como en frecuencia, basadas en bancos de filtros wavelet biortogonales.

Esta decisión metodológica se basa en las propiedades aportadas por dicha transformada, que permite aproximar con mayor similitud el procesamiento del sistema visual humano frente a las técnicas tradicionales, de forma que permita mantener en la transmisión del vídeo los criterios de la calidad visual.

Presentamos las modificaciones necesarias para la consecución de los objetivos, como un sistema de transmisión de vídeo basado en colas con prioridades variables asignadas a cada subbanda en la descomposición, sin mantener entre ellas ningún tipo

de dependencia asociada por estimación de movimiento, donde el paso del cuantificador para cada subbanda pueda ser modificado introduciendo criterios de percepción, mediante algoritmos de minimización para la distorsión visual.

La modificación del paso de cuantificación, así como la descomposición en subbandas independientes, permite que sea más fácil y simple seguir el comportamiento de la red a través de los controles de flujo basados en tasa explícita.

Dichos controles de flujo aportan la información necesaria al sistema propuesto por un lado, para permitir estimar la tasa de salida de la fuente de vídeo (mediante mecanismos de autosimilitud en el tráfico multiplexado en los conmutadores) y poder asignar los bits disponibles en los coeficientes de la descomposición de forma óptima, y por otro lado, adaptar la fuente de codificación de vídeo al estado de la red a fin de evitar congestiones en los nodos intermedios para mantener su operatividad.

Finalmente se muestran resultados experimentales para evaluar el sistema conectado a una configuración de red, formada por dos conmutadores ATM con servicio ABR unidos a través de un enlace troncal, que hace de cuello de botella, y donde por el resto de conexiones a los conmutadores se introducen tráficos con comportamientos diferentes tanto persistente como ráfaga.

Terminamos el trabajo presentando la discusión y las conclusiones deducidas, además de las líneas abiertas de investigación.

# Parte I

## Introducción



# Capítulo 1

## Punto de partida y objetivos

### 1.1 Marco histórico y social

Desde la aparición del microprocesador, la evolución de la microelectrónica y el aumento exponencial en prestaciones de los computadores, ha hecho que las Tecnologías de la Información y las Comunicaciones (TIC) pasen a ser parte esencial de nuestro *modus vivendi*, en lo que se ha venido a llamar, la “sociedad de la información”.

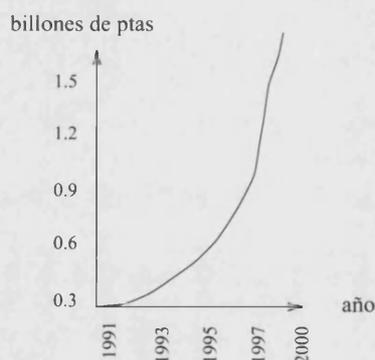


Figura 1.1: Mercado de la multimedia en Europa: billones de pesetas en ventas en los últimos 10 años

La gran oferta de productos informáticos ha permitido que la adquisición de un ordenador con excelentes prestaciones, pase a estar al alcance de una renta familiar media-baja, algo impensable en la década de los años 80 del siglo XX. La sociedad demanda cada vez servicios más avanzados e integrados, que permitan satisfacer las necesidades tanto individuales como de servicio público; se demanda tecnología, informática y telecomunicaciones, televisión por cable, televisión digital, procesadores de lenguaje, teletrabajo, telerutas, conducción con manos libres, información actualizada, telemedicina, telecompra, videoconferencia... una serie de aplicaciones sobre la infraestructura informática del usuario, que permite a distancia, gestionar e interactuar con clientes, proveedores, bancos, amistades, ocio, educación, vigilancia...

Una de las piezas clave en esta evolución son los llamados servicios multimedia y muestra de su expansión son los billones de pesetas en ventas europeas registrados en los últimos diez años, figura 1.1[IT95]. El volumen de mercado producido en los años 90 justifica el rápido crecimiento de este sector y el interés que acarrea, las expectativas que genera y las tendencias que desarrolla, como pieza clave en la evolución de las TIC.

Los servicios multimedia se distinguen por integrar en sus aplicaciones diferentes medios, texto, voz, audio, vídeo...sincronizados entre sí, permitiendo disfrutar al usuario/consumidor de un entorno de trabajo, ficticio pero más agradable, figura 1.2.

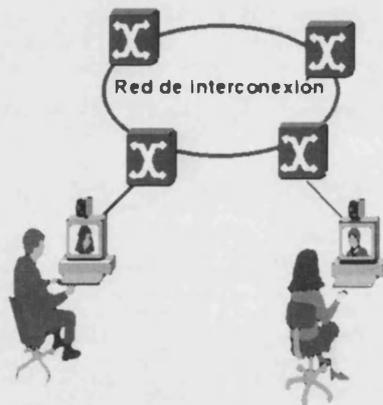


Figura 1.2: *Esquema de aplicación de videoconferencia en red*

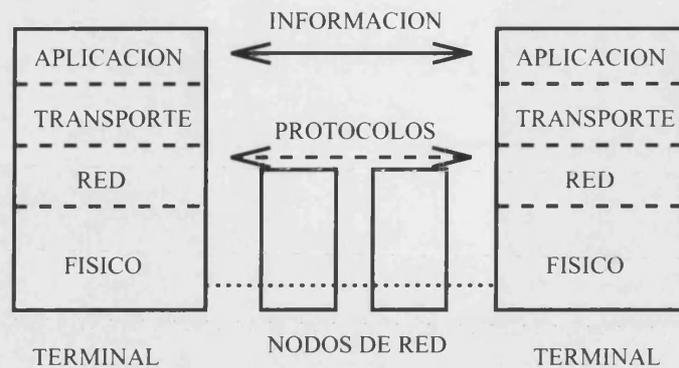
Sin embargo los requisitos telemáticos de las aplicaciones multimedia son relativamente “costosos” principalmente debido a la cantidad de información que necesitan. Cabe destacar en estos entornos que los avances en las arquitecturas de los computadores casi siempre se han visto penalizados por las redes de comunicaciones, los anchos de banda compartidos de los canales, la saturación de enlaces y de nodos.

### 1.1.1 Evolución tecnológica

La evolución de las TIC pueden y deben considerarse como un campo de conocimiento extremadamente dinámico, cuya velocidad de desarrollo es tan elevada que es difícil predecir el tiempo de vida de las diferentes tecnologías y productos.

El ámbito más estandarizado para interconexión de aplicaciones informáticas en los sistemas abiertos, viene establecido por la pila de protocolos definidos dentro del modelo TCP/IP(figura 1.3), que proporcionan la plataforma común de los servicios y las arquitecturas de red que dan la posibilidad de trabajar de forma interconectada y distribuida. El modelo TCP/IP permite abstraer detalles de implementación al considerar una arquitectura de comunicación jerarquizada, desarrollada mediante superposición de niveles de forma que cada uno de ellos ofrece sus servicios al siguiente.

La expansión de estos protocolos se basa en su simplicidad y la continua evolución para adaptar a las aplicaciones emergentes que demandan: garantías de calidad en las

Figura 1.3: *Modelo TCP/IP*

comunicaciones, seguridad, soporte de tiempo real... Una muestra de dicha expansión lo constituye el crecimiento de direcciones Internet (tabla 1.1, procedente de [Jai97]).

<i>Direcciones</i>	<i>Fecha</i>
3	1969
0.7 millones	1991
1.3 "	1993
2.2 "	1994
4.9 "	1995
9.5 "	1996
300 "(estimado)	2000
5000 "(estimado)	2003

Tabla 1.1: *Crecimiento y previsiones de direcciones en Internet*

Para hacer frente a su futuro crecimiento cabe destacar en este sentido que el campo de dirección reservado en IPv6 es de 128 bits, frente a 32 bits de IPv4. Los desarrollos actuales ayudan a superar el riesgo que toda la infraestructura, tanto hardware como software, generada por estos protocolos vaya a ser un impedimento a la inercia de la demanda social de los servicios multimedia. Así IPv6 contempla el caso de interconexión de hasta 600.000 billones de direcciones por metro cuadrado sobre la superficie terrestre[Sus96].

Esta filosofía de protocolos estandarizados en el modelo TCP/IP, ha hecho que los programadores puedan trabajar obviando niveles inferiores, consiguiendo que aplicaciones diferentes se comuniquen universalmente gracias a estos protocolos[Com91]. Con este modelo de programación se está en condiciones de por un lado generar aplicaciones no dependientes del sistema y por otro encapsular puntos críticos, como son las comunicaciones, con la posibilidad de una programación modular y orientada a objetos, obviando los detalles referidos a: requisitos de comunicación, ancho de banda, retardos, variabilidad de los tiempos de llegada (*"jitter"*), errores, sincronización, etc.

Sin embargo por las características intrínsecas a la creación de Internet, se evidencia que estas redes, aunque muy extendidas, no se adaptan en su modo nativo a las

necesidades de las aplicaciones *multimedia*, que exigen la parametrización de ciertas calidades de servicio; circunstancia ésta, especialmente llamativa si se tiene en cuenta que dicha calidad siempre ha estado presente en las redes basadas en circuito virtual como X.25, Frame Relay, ATM, etc.

El objetivo de alcanzar estos niveles de servicio sobre Internet está en el punto de mira tanto de grupos de investigación relevantes como de fabricantes de nodos de encaminamiento, tratando de introducir en la red los mecanismos necesarios para aprovecharse de las prestaciones inherentes a la filosofía de circuitos virtuales. Protocolos de reserva de recursos (RSVP), protocolos para soporte de tiempo real (RTP y RTCP), protocolos de red adaptados a las necesidades actuales (IPv6), protocolos de transporte para redes fiables (UDP) y las técnicas de multidifusión (“multicast”) integradas en los niveles de red (Multicast Bone[Kum96]), son las líneas abiertas para poder empezar a cumplir las especificaciones procedentes de los entornos multimedia en Internet.

El estudio realizado en los capítulos 2 y 6, permite observar que la evolución de ambos mundos (datagrama y circuito virtual) se aproximan entre sí a través de un modelo híbrido que tendremos ocasión de referenciar a lo largo de los diferentes capítulos, como modelo IS[She94].

Una de las clases de servicio contempladas en el modelo IS son los servicios *best effort* (expresión designada a falta de una mejor traducción de buenas intenciones), que se caracterizan por ofrecer en la medida de lo posible y de forma eficiente y equitativa los recursos disponibles en la red.

Partiendo de la hipótesis que éste es el punto de partida para el desarrollo de nuevos y futuros servicios multimedia que la sociedad demanda, el presente trabajo tiene como finalidad el análisis y diseño de un sistema propuesto para el transporte de vídeo digital.

## 1.2 Limitaciones e inconvenientes actuales de las aplicaciones multimedia

El tráfico generado por una transmisión de vídeo sin comprimir en formato QCIF para videoteléfono, es del orden de unas 40 veces mayor que el generado por una transmisión de audio en estéreo codificado en MIC; obviamente por tanto el caballo de batalla en los sistemas multimedia es la transmisión de vídeo, lo que pone sobre la mesa la necesidad de contar con una compresión-descompresión adecuada.

Otro elemento que constata las exigencias de ancho de banda en una transmisión de vídeo por red, es el hecho de que en los sistemas actuales de videoconferencia (algunos de ellos sobre Internet en experiencias piloto dentro del marco MBone [Kum96]) la calidad de audio es por lo general superior a la del vídeo.

Como es sabido, en el caso del vídeo como fuente de información, quedan superpuestas diferentes señales: los niveles de gris conocido como luminancia y el color o crominancia, informaciones éstas que se procesan en el sistema visual humano (SVH) separadamente, de forma que la percepción de cada una de ellas es diferente, como

tendremos ocasión de destacar en el capítulo 5. Así mientras el SVH muestra gran sensibilidad por los contrastes, la luminancia, el blanco y negro, no reacciona de igual forma con el color. Esta es la razón por la que el volumen de tráfico asociado a la luminancia es mucho mayor que el asociado a la crominancia.

Por tanto las dificultades encontradas para la transmisión de vídeo con criterios de calidad visual, se centran en los canales de comunicación en tanto en cuanto el ancho de banda no esté previamente negociado. Los tiempos medios de respuesta, las pérdidas de información por congestión de los enlaces o por errores en la transmisión, la variabilidad dentro de dichos retrasos, etc. son en definitiva una serie de factores que obligan a comprimir la información, protegerla y diseñar protocolos adecuados y robustos.

### 1.3 Punto de partida

Para la transmisión de información multimedia utilizando servicios de red *best effort* es necesario realizar una compresión desde el origen; con ello la información viaja comprimida para satisfacer dos criterios básicos, por un lado “*comprimir más y mejor*” para no congestionar la red y por otro lado “*negociar unos parámetros de calidad de aquello a compartir*”.

Como idea previa de la presente investigación, estaría la posibilidad de la búsqueda de un *compresor universal* que acaparase las diferentes fuentes multimedia para poder salvar los cuellos de botella de la red, ya que si todos los bits van a ser comprimidos, ¿por qué no eliminar la redundancia por igual de forma que lo que entra por un lado del sistema de compresión se descomprima en el extremo opuesto?

Si ésta fue una primera discusión, no tardó en plantearse realmente el objetivo a cumplir en la transmisión de información multimedia, las capacidades de percepción del usuario. Por tanto, si se plantea de nuevo la pregunta: ¿responde la percepción humana por igual a todos los sentidos y a diferentes excitaciones?. La respuesta tras un estudio minucioso es negativa (ver capítulo 5 y bibliografía [NJS93]). Sin embargo tras realizar una ponderación y chequeo de prioridades en la excitación del SVH, pensamos que existen elementos comunes en todo sistema de compresión: la redundancia estadística de la fuente y la codificación sin pérdidas de la información multimedia.

En particular, la posibilidad de obtener mejores métodos de compresión de vídeo, más adaptados al usuario final, más eficientes, nos hicieron plantear como punto de partida un análisis del procesamiento de la información bajo los criterios visuales [NJS93] y analizar los estándares de compresión de vídeo actuales; la idea de introducir criterios visuales en el proceso de codificación fue ya sugerida por Watson y Kunt, dando lugar a los codificadores de segunda generación [IK85].

Como se recoge en el capítulo 3, los estándares de compresión de vídeo prescinden de información redundante, sin embargo cabe preguntarse dos cuestiones, primero si realmente la información eliminada puede ser percibida o no por el SVH y segundo, si dicha información eliminada puede ser aún mayor, al objeto de conseguir mejores

tasas de compresión.

Las técnicas que desde nuestro punto de vista permiten abordar esta tarea, están basadas en las técnicas de codificación subbanda o multiresolución con transformadas localizadas en espacio y frecuencia por su analogía con el SVH (capítulo 4 y 5), donde cabe destacar las características aportadas a la descomposición de vídeo y sus ventajas (flexibilidad) ante la cuantificación, para poder adaptarse en todo momento a situaciones críticas en la transmisión tratando de mantener la calidad visual.

Aunque existen trabajos precedentes como [Vet89], [Far95] y [Che96a], éstos se centran más en las técnicas de compresión y transmisión separadamente, que en la proposición de un sistema conjunto y completo. En el capítulo 6 se entrará más en detalle en los contenidos de cada uno de los anteriores trabajos.

## 1.4 Objetivos

Como hemos señalado en nuestro punto de partida, creemos que en las tecnologías utilizadas de compresión de vídeo ante entornos de red con servicios “*best effort*” (caracterizados por sus variaciones en el ancho de banda disponible asociadas al tráfico multiplexado en los conmutadores), hacen que las propiedades del sistema visual humano puedan resultar más relevantes de lo que actualmente son utilizadas. Para constatar estas aseveraciones, nos planteamos los siguientes objetivos preliminares:

1. Revisión de la situación actual en la compresión y transmisión de vídeo
2. Estudio de las técnicas de codificación por mecanismos de multiresolución, mediante bancos de filtros wavelet y su aplicación a la compresión de vídeo buscando analogías con el sistema visual humano

concluyendo que potencialmente a la vista de los resultados, es posible ampliar, mejorar y modificar las técnicas de compresión y transmisión, ante situaciones variables y restrictivas con baja tasa de bits.

Una vez confirmada la validez de esta línea de investigación nos proponemos:

1. Obtener un algoritmo adaptativo de compresión de vídeo en multiresolución que supere los problemas inherentes al transporte por redes con servicios *best effort*, donde las redes incorporen controles de flujo cerrado basado en tasa explícita y con negociación de mínima tasa de salida para poder garantizar un mínimo de calidad visual
2. Hacer una estimación de la tasa media de salida disponible para el sistema de compresión de vídeo, basándose en las propiedades de autosimilitud del tráfico multiplexado [WW94][Ale95] y los transitorios en el comportamiento de la red
3. Asignar el reparto de bits a los coeficientes de la descomposición por mecanismos que minimicen la distorsión visual de las secuencias de vídeo, dada la restricción de la tasa media permitida de salida

4. Estudiar el valor de los parámetros de negociación con la red, para tratar de garantizar la calidad visual de las secuencias de vídeo en la transmisión
5. Evaluar y experimentar el sistema propuesto en diferentes entornos y configuraciones, con objeto de validar los análisis y diseños realizados

Con ello se aborda y se estructura la tesis en diferentes partes: la revisión de la situación actual, el análisis y diseño del sistema propuesto, la evaluación del sistema y finalmente la discusión del trabajo.



## **Parte II**

### **Revisión de la situación actual**



Los contenidos de la presente parte de revisión de la situación actual y antecedentes están estructurados en los siguientes capítulos:

- Requisitos y caracterización de las redes de comunicación para el soporte de servicios multimedia
- Técnicas de compresión de vídeo
- Codificación en multiresolución, codificación subbanda
- Sistema visual humano



## Capítulo 2

# Requisitos y caracterización de las redes de comunicación para el soporte de servicios multimedia.

Los objetivos del presente capítulo son analizar las características de las diferentes redes de comunicación basadas tanto en datagramas como en circuitos virtuales, y ver como evolucionan para dar soporte a los servicios integrados, formados por tráfico de voz, vídeo y datos.

### 2.1 Introducción

Es interesante en esta introducción analizar las particularidades y características de cada una de las fuentes de información que se integran en los servicios multimedia.

Los datos son información caracterizada por los protocolos, el sistema operativo, la arquitectura del computador y la aplicación; su comportamiento generalmente se asocia a transmisión en ráfagas.

De naturaleza diferente a los datos son la voz y el vídeo que a su vez poseen características diferentes entre sí, asociadas a las respuestas fisiológicas del hombre y por tanto sus exigencias respecto a la red son también diferentes.

En la transmisión de voz, la ordenación de las muestras es elemento clave ya que la respuesta del oído es muy crítica a estos cambios o demoras; esto es debido básicamente porque el sistema auditivo es mucho más rápido que el visual[Sus96].

Por otra parte, el sistema visual humano (SVH) analizado en detalle en el capítulo 5, está constituido por células con diferentes comportamientos que en su conjunto permiten en la percepción visual una degradación progresiva y controlada de la imagen; en el oído este proceso de degradación (asociado a la compresión con pérdidas) es mucho más crítico y su aplicación no es tan inmediata.

Además, cabe destacar que la evolución tecnológica del vídeo ha sido más lenta que la del audio. De hecho los sistemas de vídeo actuales de calidad VHS son suficientes para satisfacer las necesidades de usuario, que por lo general demanda la posibilidad

de una mayor selección de programas, más que una mayor calidad y resolución, sea el caso de la HDTV; se dice que el vídeo es más conformista que el audio. Mientras, en el audio, el usuario es más exigente como se ve en la evolución p.e. de FM mono a FM estéreo y de éste a HIFI[Sus96].

Respecto a las redes de comunicación que soportan los tráficos multimedia, las tendencias se enmarcan hacia una solución de compromiso o una elección ponderada de las prestaciones tanto de las redes basadas en datagrama como de las redes de circuito virtual. La filosofía de “acomodarse a los recién llegados aunque sea en detrimento de la velocidad” ha de seguir siendo la característica general de las redes datagrama mientras que los criterios de calidad requeridos por el tráfico multimedia han de vincular a los circuitos virtuales.

Una de las soluciones propuestas son las redes con servicios *best effort*, que se caracterizan por no garantizar en su totalidad los parámetros de calidad (retardos, pérdidas, variaciones de retardo, ...) y evidentemente hacen caracterizar o particularizar a los sistemas de compresión y transmisión, en este caso particular referente a vídeo, objetivo de estudio e investigación del presente trabajo.

## 2.2 Las redes basadas en datagramas y los servicios integrados

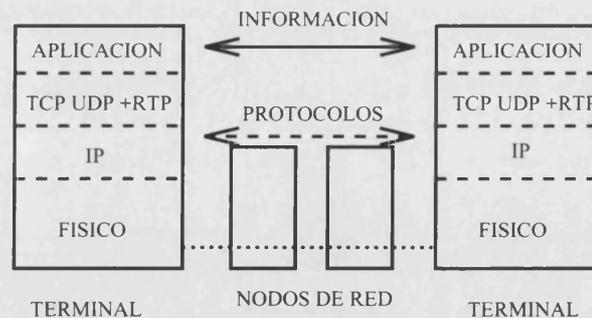


Figura 2.1: *Arquitectura IP*

La tecnología Internet, basada sobre los protocolos IP (ver figura 2.1), ha demostrado ser un posible marco para la provisión de servicios de comunicación al gran público, basado en un red universal IP sobre la que circulan paquetes de transporte con información del nivel de aplicación.

Los siguientes protocolos descritos, integrados en el modelo TCP/IP en su conjunto aportan características propias y particulares buscando dar una solución a problemas concretos y necesidades requeridas por los servicios integrados, aunque en ocasiones un poco forzadas porque estos protocolos no fueron concebidos para tal fin.

### 2.2.1 IP

Los protocolos IP (*Internet Protocol*), basados en la tecnología de conmutación de paquetes con filosofía datagrama son los encargados de interconectar los nodos a nivel de red en el modelo TCP/IP.

El éxito de esta pila de protocolos básicamente es debida a que:

- IP es una tecnología puramente de red independiente de la infraestructura física (ATM, IEEE 802.3, IEEE 802.4 ...) permitiendo multitud de medios de acceso de diferente coste, capacidad y prestaciones.
- existen numerosas aplicaciones y software que explotan su tecnología (librerías, rutinas de bajo nivel ...) que permiten fácilmente su reutilización y rápida generación de aplicaciones.
- dispone de una gestión de direcciones global con un espacio de nombres asociados que son transformables dinámicamente (DNS, "Domain Name Servers")

Los servicios que ofrece el protocolo IP a su nivel superior se caracterizan por ser:

- no fiable y sin conexión al utilizar datagramas
- con buenas intenciones para entregar cada paquete, pero sin asegurar que lo vaya a hacer bien
- con fragmentación de mensajes no transparente, según las características de las redes que atraviesan cada paquete IP

Sin embargo la versión 4 de IP (IPv4), actualmente en uso, no implementa ni separación de tráfico necesarios en las aplicaciones multimedia [She94], ni mecanismos de prioridades, para poder tratar de forma diferente a cada fuente. Estos inconvenientes han sido contemplados y se han tratado de subsanar en la nueva versión 6 de IP (IPv6) mediante la capacidad de etiquetar flujos para su proceso individualizado.

#### Direccionamiento

El direccionamiento es pieza clave en la futura evolución de aplicaciones. El direccionamiento en IP se realiza a nivel mundial, diferenciando direcciones globales (direccionan a una red) y direcciones jerárquicas (direccionan al nodo dentro de la red). A su vez, las direcciones jerárquicas se particularizan con formatos diferentes según el tamaño de la red donde se trabaje.

Independientemente del formato utilizado dentro de la dirección jerárquica, hay cuatro tipos de direcciones generales: de host o nodo, de red, de difusión y multidifusión o "multicast". Esta última es una dirección que identifica a un grupo de usuarios dentro de la red, con un formato que consta de una cabecera fija que la diferencia a las direcciones anteriores y un campo de identificación del grupo. La ventaja de este

direccionamiento estriba en que sólo se envía información a aquellos usuarios que la solicitan, evitando el transmitir información innecesaria donde no se haya requerido.

El direccionamiento multidifusión es el elemento fundamental de las aplicaciones interactivas integradas en Mbone. Mbone actualmente es una red experimental, virtual sobre IPv4 implementada a través de túneles, para poder desarrollar aplicaciones multidifusión[Kum96].

### 2.2.2 Protocolos de transporte. TCP y UDP

Por encima del nivel de red, dentro del modelo TCP/IP, se encuentran frecuentemente dos protocolos de transporte, uno orientado a conexión y otro no.

El protocolo TCP (*Transport Control Protocol*) con servicio orientado a la conexión entre los procesos extremos, subsana las deficiencias del nivel de red IP haciéndolo transparente por medio de controles de flujo; son estos controles los responsables de ralentizar las comunicaciones cuando se activan los mecanismos de recuperación en el caso de pérdidas, *multiplicando por 3 el tiempo total de transferencia*. Por tanto en aplicaciones de tiempo real es poco utilizado y en su lugar se emplea el protocolo UDP menos robusto pero más ligero y no orientado a la conexión.

El protocolo UDP (*User Datagram Protocol*) es un protocolo que no garantiza ni hace transparentes las debilidades del nivel de red y deja al nivel de aplicación encargarse de ellas; es ampliamente utilizado por aplicaciones con características de tiempo real. UDP ofrece buenos resultados debido a que la calidad de la infraestructura de las redes en la actualidad ha mejorado y las tasas de error son bajas.

### 2.2.3 Protocolos para soporte de tiempo real

El sistema de tiempo real (STR) se caracteriza sobre todo porque el tiempo de respuesta del mismo debe ser acotado y predecible.

¿Y por qué los protocolos de red y transporte clásicos TCP/IP no son apropiados para su utilización en STR? La contestación inmediata es que son protocolos orientados a la conexión pensados para funcionar sobre redes de baja o media velocidad y soportan aplicaciones sin requisitos de tiempo acotado pero que requieren transferencia fiable, como por ejemplo la transferencia de ficheros o el terminal remoto.

Nótese que los protocolos vistos de la familia TCP/IP datan de los años 70, cuando las redes ofrecían unas pobres prestaciones en cuanto a velocidad y fiabilidad, y el principal objetivo de los protocolos era obtener canales libres de errores. La tecnología de las redes ha evolucionado enormemente en los últimos 20 años y la utilización de fibra óptica como medio de transmisión ha reducido la tasa de errores a valores prácticamente inapreciables.

Ante esta situación, la ISOC (*Internet Society*) ha introducido nuevos protocolos como herramientas complementarias para dar soporte de tiempo real a las aplicaciones, ya que éstas requieren una serie de servicios que los protocolos tradicionales no son capaces de suministrar.

Cabe destacar de estas herramientas algunas características interesantes:

- *multidifusión o "multicast"*, para compartir rutas comunes y optimizar los recursos
- *resincronización*, por fechado de paquetes para reordenar y eliminar tráfico no válido por caducidad
- *control de congestión*, por bucles de realimentación para controlar la tasa de emisión con la finalidad de no saturar los conmutadores sobrecargados
- *prioridades de flujo*, para poder racionar los recursos compartidos mediante una estrategia de prioridades

Parece evidente la migración a nuevos protocolos adecuados a las características de las nuevas redes y capaces de suministrar los servicios requeridos. Con ello se analiza a continuación los protocolos estandarizados para dar soporte de tiempo real: RTP y RTCP.

## RTP

El protocolo RTP [Sch96][Jac96] (*Real Time Protocol*) fue presentado el 22-11-1995 por *Internet Engineering Steering Group* (IESG) y estandarizado por *Internet Engineering Task Force* (IETF) el 25-01-1996 y que está siendo utilizado por las mayores empresas fabricantes de software para Internet.

RTP es un protocolo del nivel de transporte adaptado para ser utilizado según las necesidades de las aplicaciones que requieran tiempo real y donde el control es realizado en los terminales extremos. No es un protocolo de transporte completo y es por ello que se monta sobre protocolos de transporte ya existentes (generalmente UDP, ver figura 2.2), integrándose en el núcleo de la aplicación para adaptarse a sus particularidades.

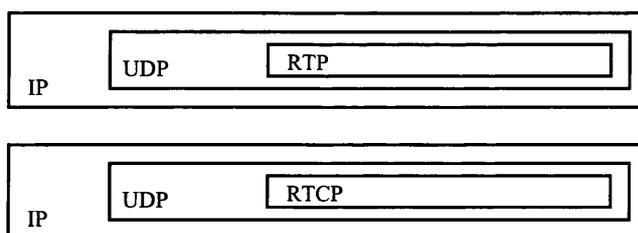


Figura 2.2: Encapsulado de los protocolos RTP y RTCP en el protocolo de transporte UDP

Fundamentalmente RTP realiza:

- reconstrucción de la base de tiempos de los diferentes flujos RTP transportados independientemente mediante marcas de sincronismo
- detección rápida de las pérdidas de paquetes a través de un protocolo adjunto para informes y estadísticas

- identificación tanto de los flujos de información transportados y de los usuarios

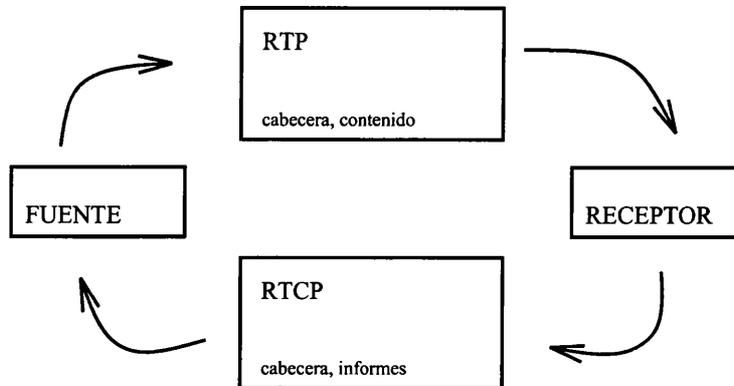


Figura 2.3: Esquema de funcionamiento de los protocolos RTP y RTCP

El protocolo RTP va acompañado de un segundo protocolo de control RTCP (*Real Time Control Protocol*) que envía al emisor una realimentación sobre la calidad de servicio de la transmisión, así como información útil definida por el usuario, figura 2.3.

Sin embargo la utilización de este protocolo de transporte no gestiona reserva de recursos en la red y por tanto no asegura la calidad de servicio. Además no aporta ni fiabilidad ni garantiza el tiempo de la entrega, ya que ningún protocolo sobre IP puede hacerlo, a menos que incorpore mecanismos adicionales para ello, como se ve en la sección 2.4.

Cabe destacar que en una sesión de videoconferencia típica, donde la velocidad de acceso de cada participante sea diferente, RTP permite que los flujos de audio y vídeo viajen de forma independiente para no restringir el acceso a ningún participante; en este escenario, puede decidirse por mantener la calidad del audio, que requiere menor ancho de banda, y acomodar el resto disponible, según la conexión de acceso de cada participante, para el flujo de vídeo recibiendo menor resolución aquellos con enlaces más precarios.

Para ello la cabecera RTP permite indicar el tipo de codificación realizada, de forma que se pueda cambiar la codificación para adaptarse a un acontecimiento imprevisto; p.e. en el caso que RTCP indique un estado de saturación en la red, los emisores tratarán de ajustar sus tasas de salida a la nueva situación.

## RTCP

El protocolo RTCP (*Real Time Control Protocol*) o de control se basa en la transmisión periódica de paquetes de control a todos los participantes de una sesión RTP y proporciona un retorno de información sobre la calidad de recepción de los datos transmitidos en los paquetes RTP.

El contenido de los informes para cada fuente recibida, incluye las estadísticas siguientes:

- porcentaje de paquetes RTP perdidos desde el anterior envío (8 bits)

- número acumulado de paquetes RTP perdidos (24 bits)
- número de secuencia más elevado de los paquetes recibidos (32 bits)
- cadencia o variación en los tiempos de llegada de los paquetes RTP (32 bits)
- fechado del último informe emitido (32 bits)
- demora desde la recepción del último paquete de la fuente (32 bits)

que permite reconstruir con fidelidad el estado de la red para poder planificar acciones preventivas, p.e. adaptación mediante un grado de compresión mayor, . . . .

Finalmente, la frecuencia de transmisión de paquetes RTCP recomendada por el estándar es, para emisión un 25% de su ancho de banda y para el receptor como máximo un 5% del ancho de banda global de la sesión.

## 2.3 Las redes de circuito virtual y los servicios integrados

Otro enfoque y punto de partida diferente a las redes basadas en datagramas, para satisfacer las necesidades de los servicios integrados, son las redes de circuito virtual.

Los servicios soportados por las centrales telefónicas, a medida que las centrales han ido digitalizándose, han aumentado y la cantidad de información que cruza una central telefónica hoy en día es muy diversa: voz, datos, imágenes . . . que pueden justificar diferentes redes para cada tipo de tráfico.

Sin embargo las redes específicas para servicios concretos, aunque sean la mejor solución para un problema particular, conllevan problemas de diversificación, sincronismos si la información es multimedia, incompatibilidades, costes . . . . Cabe pensar en una red óptima, que trate de solucionar y adaptarse a las necesidades demandadas por tráfico de naturaleza diferente; esta solución es simplemente óptima pero no la mejor para cada caso particular.

El objetivo de la RDSI (Red Digital de Servicios Integrados) ha sido englobar en una todas las redes especiales que han ido proliferando para dar servicios específicos y dar pie a un estándar común de interconexión. En particular, para dar servicios de banda ancha en aplicaciones con altos caudales de salida, se ha elegido la tecnología ATM basada en la transmisión asíncrona.

### 2.3.1 Arquitectura de protocolos y tecnología ATM

ATM es una técnica de conmutación, de multiplexado, incluso de transmisión, variante de la conmutación por paquetes (celdas) en cuanto recurre a los paquetes cortos y de tamaño fijo (53 bytes) una vez negociada la conexión extremo a extremo.

El tratamiento de esas celdas en los conmutadores esta limitado al análisis de la cabecera (los 5 primeros bytes de los 53 de la celda) para permitir su encaminamiento según la dirección hacia las colas de salida. La cabecera de cada celda identifica la

conexión y la carga que transporta mediante diferentes campos: dirección (VPI-VCI), identificador del tipo de carga (PTI, 3 bits), prioridad (CLP, 1 bit) y control de error de cabecera (HEC, 8 bits).

Esta ha sido la forma de adaptar la red a cualquier tipo de tráfico basándose en la conmutación rápida de celdas, con lo que el ancho de banda es aprovechado con mayor facilidad para todos los servicios; esta característica aporta una ganancia en utilización de recursos conocida como *multiplexación estadística*.

ATM ofrece dos niveles de acceso o interfaces que definen las diferentes funcionalidades para compatibilizar los equipos de diferentes fabricantes, uno para el acceso de usuario, UNI (*User Network Interface*) y otro para comunicación entre los nodos dentro de la red, NNI (*Network to Network Interface*) [AAL93]. ATM envía a través de los interfaces la información segmentada en celdas que son transportadas de forma asíncrona y reensambladas en el destino.

Respecto al tratamiento de los errores, destacar que no se efectúan dentro de la red ATM, si no que se dejan a cargo de las aplicaciones usuarias o de los equipos de acceso, de la misma manera que lo hace *Frame Relay* respecto a X.25, justificado por la baja probabilidad de error que presentan las tecnologías de red actuales.

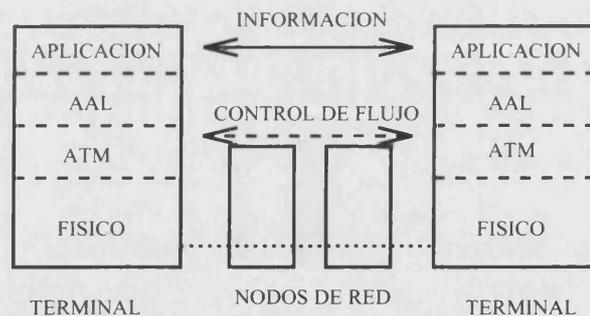


Figura 2.4: *Arquitectura de red ATM*

La arquitectura ATM está basada en la superposición de tres capas, una primera de adaptación de los datos a la estructura de las celdas, conocida como capa AAL (“ATM Adaptation Layer”), una segunda capa de conmutación y multiplexado de celdas, conocida como ATM y la última, y de más bajo nivel, conocida como capa física.

La capa física depende directamente de la tecnología utilizada en el medio físico. La capa ATM incorpora los controles de los flujos y preserva el orden en los conmutadores y la capa AAL, implementa la convergencia, segmentación y reensamblado de los datos de la aplicación con la capa ATM.

La capa más cercana a la aplicación, desde el punto de vista de programación es la capa AAL [AAL93], capa a la que haremos frecuentes referencias. Dicha capa se especifica para cada servicio y se estructura en otras dos subcapas:

- la subcapa de convergencia (CS, Convergence Sublayer) más cercana al punto de acceso de servicio, que acepta los paquetes de usuario SDU (*Service Data Unit*) y

les introduce una cabecera formando paquetes PDU (*Protocol Data Unit*); dicha cabecera puede incluir técnicas de corrección avanzadas (FEC, Forward Error Correction)

- la subcapa de segmentación (SAR, Segmentation and Reassembly), por debajo de la anterior, segmenta todos los paquetes PDU que llegan del nivel superior en celdas ATM y los reensambla en el destino

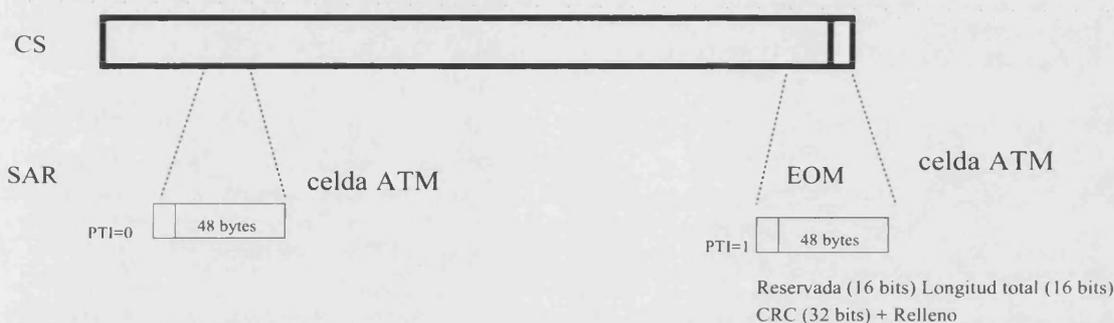


Figura 2.5: AAL5: Capa de adaptación 5 en ATM

Tiene especial interés para el presente trabajo de investigación, entre las diferentes capas de adaptación, la capa AAL5 en figura 2.5; esta capa en la subcapa CS incorpora a la SDU una cabecera (especificando la longitud, CRC y relleno de ceros) para formar la PDU, que posteriormente pasa a la subcapa SAR donde su contenido es dividido en celdas, marcadas todas ellas con el campo PTI=0 excepto la última con PTI=1 o fin de mensaje (*End Of Message*) que incorpora la cabecera introducida a la SDU.

### 2.3.2 Servicios ATM: CBR, VBR, ABR, UBR

ATM propone diferentes capas de adaptación según las necesidades y servicios requeridas para aumentar las prestaciones de la arquitectura definida.

Cada servicio incorpora mecanismos específicos para que el tráfico se conserve dentro de los márgenes estipulados, a fin de asegurar a los usuarios la calidad de servicio estipulada y que no violen los parámetros estadísticos negociados en la conexión.

El funcionamiento del conmutador se basa en tratar de predecir y estimar el comportamiento del tráfico de entrada para que dicho tráfico pueda ser acomodado dentro del conmutador por multiplexación y que le permita estimar su ocupación para negociar futuras conexiones y preservar su buen funcionamiento.

Según la categoría de servicio utilizada los servicios básicos pueden ser:

1. CBR (*constant bit rate*): para tasa constante de salida, como pueda ser transmisión de vídeo analógico digitalizado, donde el ancho de banda queda garantizado y constante según la negociación realizada; normalmente utiliza interfaz AAL1 y AAL2

2. VBR (*variable bit rate*): para tasas variables de salida, por ejemplo transmisión de vídeo digital comprimido, donde pueden haber variaciones bajo unos estadísticos determinados en la negociación a través de la interfaz AAL1 y AAL2
3. ABR (*available bit rate*): servicios *best effort*, que ofrecen el ancho de banda restante a las conexiones CBR y VBR en los conmutadores de la forma más equitativa posible a través de la interfaz AAL5 (figura 2.5), pero sin ningún tipo de calidad de servicio salvo la negociación inicial de una tasa mínima de salida (*Minimum Cell Rate*)
4. UBR (*unspecified bit rate*): servicios sin tasa negociada como pueda ser el correo electrónico que no requieren de ningún caudal específico, utilizando el ancho de banda restante de los servicios anteriores sin incorporar controles de flujo, a través de la interfaz AAL5

### 2.3.3 Controles de flujo ABR

Los servicios ABR se caracterizan por mantener a las aplicaciones extremas dentro de un bucle de realimentación o control de flujo para tratar de adaptarse a los diferentes estados de la red; en estos servicios las variaciones de los retraso están asociadas a las fluctuaciones del tamaño de las colas en los conmutadores y las pérdidas al desbordamiento en ellas.

ABR inicialmente pensado para tráfico ráfaga de datos con servicios *best effort* ha dado paso a otros tipos de tráfico más persistentes, debido a que en los datos los bucles de realimentación del control de flujo no llegan a cerrarse; las fuentes persistentes pueden adaptarse a los diferentes grados de congestión en que se encuentren los conmutadores atravesados, consiguiendo pérdidas casi nulas.

Dichos controles de flujo son realizados a través de celdas RM (*Resource Management*), celdas que tratan de llevar la información necesaria, para poder establecer el control de flujo.

Los controles de flujo se pueden clasificar en dos tipos, indicadores de congestión e indicadores de tasa explícita. Los primeros realimentan al emisor con un bit la congestión o no de alguno de los conmutadores atravesados[Jai90]. La segunda, mediante una tasa explícita de salida, generalmente la máxima posible, los conmutadores según los estadísticos negociados y de su estado interno ajustan dicha tasa a la que realmente pueden ofrecer; dicha información es pasada por todos los conmutadores, adaptando la salida de la fuente al cuello de botella de la red, al conmutador con mayor grado de congestión[Cha94].

Además, el control de flujo puede ser abierto o cerrado. En el primer caso se gestiona cada vez un intervalo de transmisión y se transmite (p.e. ATM Block Transfer, ABT-Delayed Transmission y ABT-Immediate Transmission), en el segundo caso la realimentación continua permite adaptar las fuentes dinámicamente[ITU96].

El interés del presente trabajo de investigación para compresión adaptativa de vídeo se basa en servicios ABR con controles de flujo cerrado por tasa explícita.

## 2.4 Modelo IS

Debido a la variabilidad de los retardos en las colas de los nodos de encaminamiento y a las pérdidas de paquetes por congestión, es necesario plantear un modelo para la integración de servicios sobre Internet.

El modelo IS descrito en [She94], propone una extensión a la arquitectura Internet sin pretender modificarla, por medio de dos piezas clave. La primera, la introducción de QoS por reserva de recursos de manera explícita para ofrecer servicios predecibles o de tiempo real. Y la segunda, la división del tráfico en unas pocas clases administrativas y asignarles a cada una de ellas un porcentaje de ancho de banda en cada nodo, incluso en condiciones de sobrecarga, es decir un control de enlace compartido.

Los requisitos para dar soporte a los servicios en tiempo real, estriban en determinar una cierta garantía de servicio que no puede ser alcanzada de forma general sin una cierta reserva de recursos. Es decir, que es ineludible que los nodos de encaminamiento sean capaces de reservar recursos para proporcionar una calidad específica para cada flujo de paquetes de usuario.

Para ello, el modelo IS propone cuatro elementos en los nodos de la red para introducir dichas funcionalidades:

- un clasificador de paquetes, que clasifique según a la clase que pertenece cada paquete
- un planificador de paquetes, que reordene las colas de salida relacionado con la categoría de cada clase
- un control de admisión, que acepte más conexiones en caso de no perjudicar a las ya establecidas
- una reserva de recursos, a través de una lista de parámetros, que engloben la negociación de la calidad requerida

### RSVP

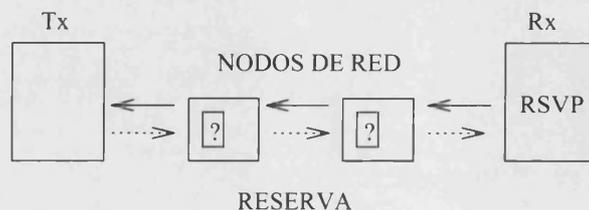


Figura 2.6: *Protocolo de red para reserva de recursos*

Siguiendo las pautas fijadas por el modelo IS aparece el protocolo RSVP (*Reservation Protocol*)[Jam96][Jam97], que funciona bajo los objetivos de:

- reservar recursos a través de una lista de parámetros especificando la calidad de servicio
- establecer *a priori* el camino por donde han de circular el resto de paquetes
- dar prioridades a los flujos de tiempo real

Para llevar a cabo la reserva de recursos, y tratando de mantener la filosofía de Internet respecto a la simplicidad y buenas intenciones, se introducen dentro de los conmutadores los estados *volátiles*, estados que permiten reservar unos QoS entre los nodos de encaminamiento a través de una especie de circuito virtual; con la restricción de que esta reserva sea periódicamente actualizada por el solicitante para que no se pierda, es por ello de donde procede el nombre de estados *volátiles*.

Como la reserva de recursos implica que algunos usuarios tengan más privilegios que otros, surge la necesidad de un control administrativo y una vigilancia sobre los recursos asignados a cada usuario y que puede ser utilizado para tarificación.

RSVP ha sido diseñado para optimizar la entrega de datos en modo multidifusión (*multicast*) siendo un caso particular el funcionamiento monopunto, y gestiona de manera particular cada punto de destino en función de sus capacidades y necesidades de calidad de servicio.

La reserva de RSVP es conducida por el receptor: el destinatario emite una petición de calidad de servicio correspondiente a sus necesidades que se encamina hacia el emisor en forma de mensaje RSVP. La petición de recursos es pues personalizada para cada destino. Ciertas peticiones de calidad de servicio pueden ser incompatibles con los recursos de uno de los nodos de la cadena, en este caso RSVP devuelve un mensaje de error a la aplicación.

RSVP funciona en modo no conectado con los nodos de encaminamiento y sólo los sistemas terminales están en modo conectado con el emisor, enviando periódicamente mensajes RSVP de control para:

- actualizar dinámicamente la reserva en la llegada o salida de un miembro en un grupo de difusión
- modificar el camino de encaminamiento
- refrescar la información del circuito virtual, ya que vencido un cierto tiempo y en ausencia de refresco la reserva desaparecerá

<i>Internet</i>	<i>ATM</i>
garantizado	CBR, tiempo real en VBR
carga controlada	sin tiempo real en VBR
servicios <i>best effort</i>	ABR y UBR

Tabla 2.1: *RSVP y ATM: correspondencia de servicios*

El protocolo RSVP es un complemento a los protocolos de transporte y que permite la fusión de las redes públicas basadas en circuito virtual con Internet[Man96] a través

de un mapeado de servicios, como referencia ver la tabla 2.1 y las publicaciones en IETF del grupo de trabajo *Integrated Services*; dicha tabla 2.1 asocia según la lista de parámetros negociada con RSVP los diferentes tipos de servicio.

La calidad garantizada de un ancho de banda y de un retardo se lleva a cabo por conformación del tráfico con un *Token Bucket* (TB), similar al *Leaky Bucket* (LB) de conexiones CBR en ATM, pero con la particularidad de poder controlar directamente el TB para modificar los retardos.

La carga controlada negocia los parámetros de un TB, similar al LB de conexiones VBR en ATM, con el fin de ofrecer al cliente un servicio para aplicaciones no restrictivas de tiempo real como si la red estuviese poco cargada, utilizando control de admisión.

### Analogías en los servicios *best effort*

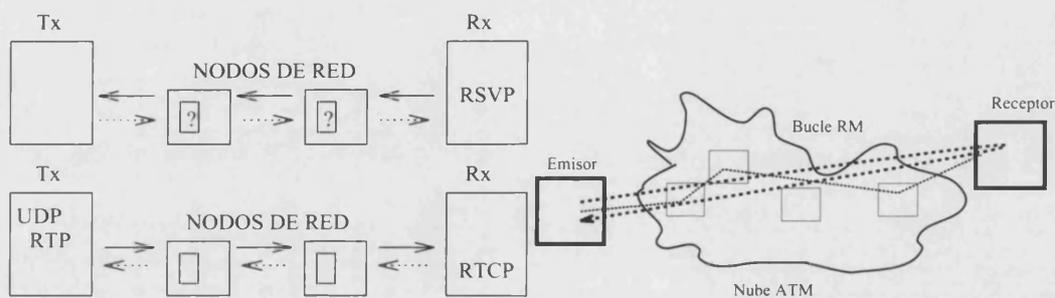


Figura 2.7: Servicio *best effort*: configuración sobre redes IP y ATM con servicio ABR

Una posible configuración de las redes IP para dar soporte de servicios integrados, siguiendo las pautas del modelo IS, se puede llevar a cabo mediante los protocolos:

- RSVP para reserva de recursos
- IP y UDP para transporte de información
- RTP y RTCP para ofrecer soporte de tiempo real

Esta configuración de red, presenta analogías con los servicios ABR de las redes ATM, como se puede ver en la figura 2.7. Dichas analogías, una vez negociado el circuito virtual, se pueden resumir en los siguientes puntos:

1. bucle de realimentación entre las aplicaciones
2. monitorización de la reserva volátil del protocolo RSVP
3. servicios *best effort*

Analicemos a continuación cada uno de dichos tres puntos.

La realimentación entre el emisor y el receptor realizada por el protocolo RTCP mantiene relación en la manera que el servicio ABR lo hace a través de las celdas RM en los controles de flujo.

El refresco de la reserva realizado en el protocolo RSVP para mantener los estados volátiles, va implícito en la realimentación del servicio ABR por una tasa mínima y variable de celdas RM de las fuentes ABR[Kim96].

Finalmente, los servicios *best effort* vienen determinados en ambos casos por un reparto equitativo del ancho de banda disponible a la salida de los enlaces[T.L97][Mar97].

A la vista de dichas analogías, cabe esperar que las aplicaciones, diseñadas para redes con dichas características, puedan ser utilizadas en su modo nativo para diferentes arquitecturas de red, tanto IP como ATM.

Sin embargo, no es evidente y queda abierto para investigaciones futuras, traducir los mecanismos de realimentación introducidos en este caso por RTCP (ver sección 2.2.3) con la tasa permitida de salida ofrecida por ABR.

## 2.5 Conclusión

La revisión de las redes de comunicación actuales para dar soporte a los servicios integrados, evidencian una clara tendencia a solucionar los problemas intrínsecos del tráfico multimedia.

Diferentes puntos de partida, tanto las redes de circuito virtual como las redes basadas en datagramas, convergen en un modelo (modelo IS) para poder ofrecer diferentes clases de servicio, entre ellos servicios *best effort*; dichos servicios se caracterizan por dar la máxima calidad a las conexiones en la medida de lo posible, tratando de repartir los recursos de la red de forma eficiente y equitativa.

La evolución de esta clase de servicios (*best effort*), queda justificada tanto por las cifras de mercado, como por el crecimiento y sus previsiones para los últimos años como se ha visto en el capítulo 1.

Sin embargo como se analiza a lo largo de los contenidos del presente trabajo, dichos servicios no son el marco ideal para el transporte de vídeo ya que no ofrecen una total garantía tanto en ancho de banda, retardos, variaciones de retardo y pérdidas.

Concluimos por tanto que es necesario revisar los sistemas de compresión de vídeo actuales, con el objetivo de mejorar aquellos aspectos que potencialmente hagan su adaptación a los servicios *best effort* de forma más simple y eficiente, manteniendo una buena calidad visual.

Por otro lado, aunque existen evidentes analogías entre las arquitecturas de red IP y ATM para dar soporte a los servicios integrados (como se ha visto en el apartado 2.4), la realimentación por tasa explícita realizada por los servicios ABR de ATM, permite de forma más directa y evidente, con mayor facilidad, adaptar la aplicación al estado de la red.

Con ello se concluye que el análisis y diseño del sistema propuesto para transmisión de vídeo se centrará en los servicios *best effort* de las redes ATM (servicios ABR).

# Capítulo 3

## Técnicas de compresión de vídeo

El presente capítulo introduce y analiza los métodos más extendidos utilizados por los estándares para la compresión de vídeo; una vez vistos éstos y considerando los servicios *best effort* ofrecidos por las redes de comunicación, los capítulos 4 y 5 introducen las herramientas necesarias para proponer una nueva alternativa que se detalla en las partes III y IV del presente trabajo.

### 3.1 Proceso de compresión. Redundancia de la información

La información visual natural que utilizamos está altamente correlacionada y contiene gran número de redundancias que se pueden extraer. Los tipos de correlación que podemos encontrar son:

- redundancia objetiva o estadística originada por la correlación existente tanto en el eje temporal como en el eje espacial y frecuencial entre las intensidades de pixels adyacentes, clasificable en:
  - redundancia espacial o correlación local entre las intensidades de los pixels vecinos
  - redundancia frecuencial o correlación espectral de las imágenes
  - redundancia temporal o correlación de un determinado fotograma con algún fotograma contiguo, debido al movimiento en la escena
- redundancia subjetiva o perceptual, debida al hecho de que el SVH es incapaz de percibir ciertas componentes visuales presentes en el vídeo; para ello hay que tener en cuenta la sensibilidad al contraste en función de las frecuencias espaciales y temporales, enmascaramiento de la información en el dominio espacial y la redundancia cromática (ver capítulo 5).

El objetivo de la compresión es por tanto eliminar gran parte de dicha redundancia mediante dos esquemas básicos con pérdidas y sin pérdidas, con el fin de disminuir la entropía de la información que deseamos guardar, transmitir o procesar.

En una compresión sin pérdidas se persigue reconstruir exactamente la información y por tanto no se realiza ningún tipo de cuantificación.

En la compresión con pérdidas se buscan factores de compresión mucho mayores, donde la cantidad y el tipo de pérdida depende de la técnica empleada y de los objetivos de compresión perseguidos.

Toda compresión de vídeo consta de la combinación de las siguientes etapas como se puede ver en la figura 3.1: descomposición de la señal de vídeo, cuantificación, y codificación sin pérdidas.



Figura 3.1: Esquema general de un codificador de vídeo

1. Descomposición, mapeado o transformación de la señal tiene como objetivo realizar una descorrelación de la señal de vídeo, para que la nueva representación obtenida sea más compacta; de esta operación depende en gran medida la compresión alcanzada. Se utilizan básicamente tres tipos de descomposiciones: predicción, transformación o ambas, esta última conocida como transformación híbrida.
2. Cuantificación de los coeficientes, exclusiva del esquema de compresión con pérdidas, consiste en discretizar los coeficientes resultantes de la descomposición de la señal.
3. Codificación sin pérdidas, tras la discretización de los coeficientes y conociendo la función de distribución de probabilidad de los símbolos, trata de eliminar la redundancia estadística mediante códigos de longitud fija o variable, p.e. codificación de Huffman, la codificación aritmética de IBM y la codificación *Run-Length*.

### 3.1.1 Medidas de calidad en imágenes

Dada una imagen reconstruida a partir de un proceso de compresión es necesario establecer criterios para medir su calidad o fijar su distorsión. Para ello, tradicionalmente han sido empleados los siguientes métodos matemáticos:

$$MSE = ||x - \hat{x}||^2 \quad (3.1)$$

$$PSNR = 10 \log \left[ \frac{I_{max}^2}{MSE} \right] \quad (3.2)$$

donde  $I_{max}$  es la máxima intensidad presente en la señal original.

La expresión 3.1 es el error cuadrático medio (MSE) entre la señal original  $x$  y la señal reconstruida  $\hat{x}$  a partir de los coeficientes cuantificados y la expresión 3.2 es la relación señal de pico a ruido (PSNR); ambas expresiones son dos criterios utilizados debido en parte a su sencillez y manejabilidad.

Estas medidas tratan a todos los pixels de una imagen o vídeo de la misma forma, y cabe destacar que desde el punto de vista perceptual esta suposición no es correcta, pues hay regiones tanto en el dominio espacial como frecuencial de la imagen, a las cuales el SVH es más sensible que a otras.

## 3.2 Descomposición de la señal

La descomposición de la señal puede ser realizada a través bien de una transformación lineal (p.e. DCT) o bien de una predicción (p.e. algoritmo DPCM) con el objetivo final de obtener una nueva fuente de información descorrelada, donde la cuantificación sea más eficiente que si se realizara sobre la fuente de información inicial[GS96].



Figura 3.2: *Imagen de Lisa*

### 3.2.1 Transformadas lineales. DCT

Las transformaciones lineales pueden ser sin solapamiento o con solapamiento, es decir todas las muestras se filtran una única vez o algunas muestras se filtran dos o más veces en el caso que se produzca solape. Las primeras se conocen como *transformadas bloque* y son ampliamente utilizadas por su facilidad computacional. Las *transformadas con solapamiento* son más complejas; un caso particular de ellas se ve en el capítulo 4.

En el caso de vídeo, el dominio espacial puede modelarse como una *fuentes de Markov* de pixels de orden  $n$  o grado de correlación de pixels, de manera que la entropía sea menor que la de una fuente de orden  $m$  si  $m < n$ , por lo que cabe esperar que las descomposiciones con solapamiento descorrelen mejor que aquellas que no lo realizan.

La transformada DCT, sin solapamiento, es la que más se aproxima a la ideal o KLT (*Karhuen-Loeve Transform*)[GG92][Jai89] y es más estable frente a truncamiento de los coeficientes que la transformada FFT. Es una transformada lineal y frecuentemente utilizada en los estándares de compresión de imagen y vídeo. Utiliza en su

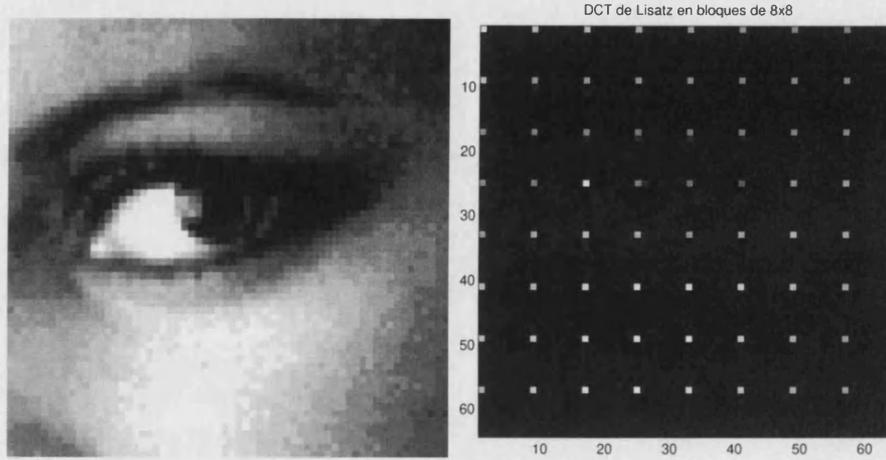


Figura 3.3: Fragmento de Lisa y su DCT correspondiente

descomposición bloques de 8x8 pixels (ver figura 3.3) que generan 64 coeficientes:

$$c(i, j) = \frac{1}{4} k(i) k(j) \left[ \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \frac{(2x+1)i\pi}{16} \cos \frac{(2y+1)j\pi}{16} \right] \quad (3.3)$$

siendo  $c(0, 0)$  y  $c(7, 7)$  los coeficientes de menor y mayor frecuencia horizontal y vertical respectivamente.

La transformada inversa de la DCT es

$$f(x, y) = \frac{1}{4} \left[ \sum_{i=0}^7 \sum_{j=0}^7 k(i) k(j) c(i, j) \cos \frac{(2x+1)i\pi}{16} \cos \frac{(2y+1)j\pi}{16} \right] \quad (3.4)$$

con  $i, j = 0 \dots 7$  para filas-columnas y siendo  $k(l) = \begin{cases} \frac{1}{\sqrt{2}}, & l = 0 \\ 1, & l \neq 0 \end{cases}$

La DCT es una transformación unitaria y en términos de notación matricial, podemos escribir  $C = T F T^t$  y  $F = T^t C T$  donde  $C = [c(i, j)]$  son las matrices 8x8 de coeficientes,  $F = [f(i, j)]$  la imagen original,  $T = [t(i, j)]$  la matriz de la transformación DCT y  $T^t$  su transpuesta.

La matriz  $T$  se caracteriza por tener la primera fila ( $i = 0$ ) de 8 elementos iguales con valor  $\frac{1}{\sqrt{N}}$  y los demás elementos de la matriz hasta completar una matriz de  $8 \times 8$  con los términos  $\sqrt{\frac{2}{N}} \cos \left[ \frac{(2j+1)\pi i}{2N} \right]$ , siendo  $N$  el número total de coeficientes.

### 3.2.2 Descomposición basada en predicción. DPCM

Este método de descomposición consiste en realizar una predicción  $\hat{x}[n]$ , normalmente lineal, de una muestra  $x[n]$  de la señal de entrada en base a otro conjunto de muestras anteriores a  $x[n]$  y posteriormente cuantificar la diferencia entre  $\hat{x}[n]$  y  $x[n]$ , es decir

$d[n] = x[n] - \hat{x}[n]$  que recibe el nombre de error de predicción [GG92][SN84], de manera que cuando la predicción es buena  $d[n]$  toma valores casi nulos. La predicción se realiza a través de un esquema por lazo cerrado (figura 3.4) para dar mayor estabilidad y evitar la divergencia introducida por la cuantificación a la predicción.

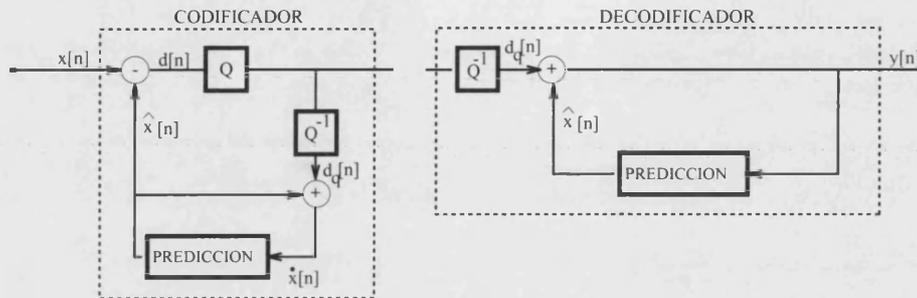


Figura 3.4: Predicción con lazo cerrado: algoritmo DPCM

En el algoritmo DPCM (*Differential Pulse Code Modulation*) la predicción del codificador se realiza a partir de valores reconstruidos después de la cuantificación  $\{\hat{x}[n-1], \dots, \hat{x}[n-L]\}$ , los mismos que en el decodificador, realizando la misma predicción tanto en el codificador como en el decodificador, permitiendo conseguir que el error cuadrático de la secuencia de salida quede controlado.

### 3.3 Estimación del movimiento

Uno de los aspectos más importantes en la calidad y capacidad de compresión de vídeo, reside en el movimiento (figura 3.5). Como se ha visto al principio de este capítulo, dentro de la redundancia objetiva o estadística, existe una redundancia temporal (*interframe*) entre el fotograma  $t$  y  $t+1$ , salvo pequeños desplazamientos.

El movimiento se puede extraer de dos formas diferentes, bien en el dominio espacial o bien en el dominio de frecuencias espacio-temporales. En el primero, ampliamente utilizado en los estándares de compresión como se ve a continuación, trata de establecer correspondencias entre fotogramas adyacentes por estimación de movimiento minimizando una función coste; en el segundo, que detallaremos en el capítulo 4, el movimiento se codifica por transformadas espacio-temporales en 3D, para los ejes vertical, horizontal y temporal.



Figura 3.5: Secuencia de dos fotogramas en un intervalo de 160 ms y su diferencia asociada al movimiento de la escena

La estimación de movimiento en el dominio espacio-temporal consiste primero en realizar una predicción de vectores de movimiento de aquellas partes de la escena que se mueven entre fotogramas adyacentes y que son nulos para partes estáticas, y segundo estimar el error de predicción producido. Esta predicción puede ser realizada de forma natural hacia delante o hacia atrás dependiendo de si el fotograma  $t$  se predice del fotograma  $t - 1$  o del fotograma  $t + 1$ .

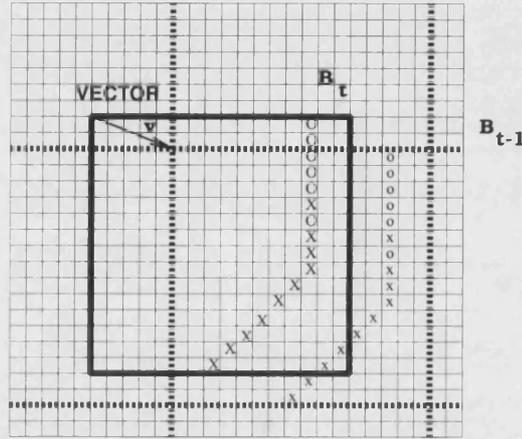


Figura 3.6: Análisis de correspondencia de bloques para la estimación de movimiento

La estimación de movimiento utilizada por los estándares de vídeo, aunque los estándares no definen su implementación, está basada en correspondencias entre bloques de forma que cada bloque está formado por un conjunto de píxeles que se mueven de la misma forma. Esta técnica supone, para realizar la estimación del movimiento, tres hipótesis: de traslación próxima, de estabilidad de la intensidad y del desplazamiento lineal.

Es por ello que presenta problemas ante rotaciones, dilataciones y/o contracciones que tratan de ser resueltos a través del error de predicción; sin embargo la imposición de que el movimiento sea traslacional facilita la implementación sobre circuitos de aplicación específica (VLSI).

La correspondencia entre bloques consiste, en dado un bloque  $B_t$  de un fotograma presente  $t$ , encontrar en el fotograma anterior un bloque  $B_{t-1}$  tal que el bloque  $B_t$  sea aproximadamente igual al bloque  $B_{t-1}$  trasladado con un vector de movimiento  $v$  que minimiza los criterios de correspondencia o función coste, ver figura 3.6. Esta función coste, según la medida del error absoluto medio (diferencia del valor absoluto) se puede expresar por

$$MAD(i, j) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N |F(x, y, t) - F(x, y, t - 1)| \quad (3.5)$$

con  $F(x, y, t)$  la intensidad perteneciente al fotograma  $t$ , fila  $y$  y columna  $x$ ,  $|i| \leq m$  y  $|j| \leq n$  y el tamaño del bloque  $M \times N$ ; en los estándares H.261 y MPEG-2 los tamaños de estos bloques son  $16 \times 16$  píxeles y en H.263 se contemplan tamaños de  $8 \times 8$  píxeles.

### 3.4 Estándares de compresión y transporte de vídeo

La compatibilidad e interoperabilidad de sistemas de compresión de diferentes fabricantes es asegurada por los organismos de estandarización. Los motivos por los cuales unos estándares prevalecen más que otros vienen determinados por las características de implementación, compatibilidad y escalabilidad.

Pese a que existen gran variedad de estándares de compresión de vídeo (ver el anexo A), éstos no suponen grandes cambios en el núcleo de compresión, consistente en versiones modificadas y ampliadas a partir del algoritmo base llamado algoritmo híbrido *DCT/Predicción-Compensación de movimiento* (figura 3.7) debido a los buenos resultados alcanzados y a la gran facilidad de implementación.

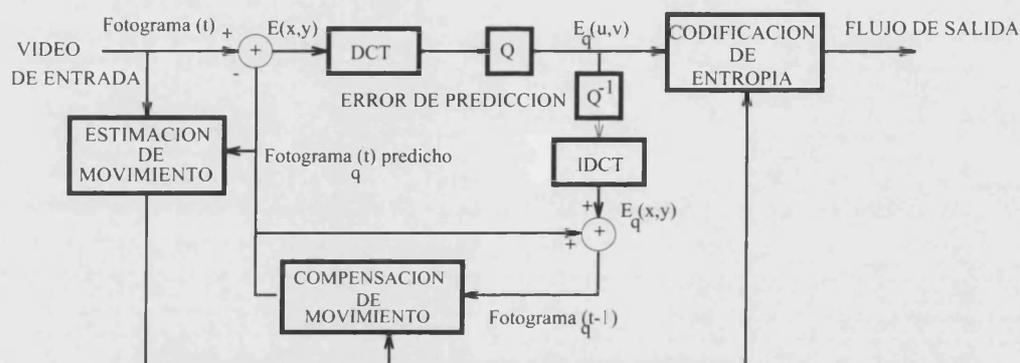


Figura 3.7: Algoritmo híbrido *DCT/Predicción-Compensación* utilizado en compresión de vídeo

El algoritmo híbrido *DCT/Predicción-Compensación de movimiento*, persigue eliminar tanto la redundancia espacial como la temporal.

La redundancia temporal es eliminada por la estimación de movimiento implementada a partir de bloques del fotograma de entrada, para obtener si es posible un error de predicción o residual de compensación de movimiento para cada uno de ellos, en el caso de no ser posible dichos bloques se codifican sin predicción.

La redundancia espacial de dicho error residual se reduce mediante la transformada de tipo bloque DCT, cuyos coeficientes son cuantificados y codificados sin pérdidas; en la realimentación del bucle interno se realiza la transformada inversa IDCT del error cuantificado al cual se le suma el bloque predicho.

La información final requerida en la transmisión, es el error residual cuantificado y codificado junto con los vectores de movimiento.

#### 3.4.1 Estándar H.261 y H.263

H.261 es un estándar de compresión de vídeo desarrollado para facilitar servicios de videoconferencia y videotelefonía en tiempo real sobre redes digitales de servicios integrados (RDSI).



H.261 está caracterizado por su escalabilidad  $p \times 64$  Kbps, siendo 64 Kbps el ancho de banda de cada canal B en RDSI y  $p = 1, \dots, 30$ . H.261 acepta dos tipos de formatos o tamaño de fotograma determinados por la resolución de los terminales: CIF (un cuarto del formato PAL) de tamaño 352x288 pixels y QCIF de tamaño 176x144 pixels.

H.261 define cuatro estructuras de datos, también utilizadas en el estándar MPEG, para componer las secuencias de compresión: la estructura mínima o bloque básico formada por 8x8 pixels como unidad de transformada, el macrobloque (MB) o la agrupación de 4 bloques básicos como unidad de movimiento, el GOB o grupo de bloques formado por 44 bloques básicos y la estructura superior autodefinida o fotograma.

El algoritmo de compresión está basado en dos modos de operación para eliminar la redundancia temporal y espacial, llamados modo *interframe* e *intraframe* respectivamente; la selección de un modo u otro se realiza automáticamente bien cuando se detecta gran información con redundancia temporal y se puede realizar una buena predicción, o bien cuando aparece un cambio de escena.

El modo *interframe* implementa el algoritmo *DCT/Predicción-Compensación* con predicción temporal hacia atrás, buscando para cada MB la compensación de movimiento si da calidad suficiente, y si no los codifica por redundancia espacial mediante transformada.

El modo *intraframe*, muy similar al estándar JPEG, realiza la compresión para todos sus MBs sin tener en cuenta los fotogramas anteriores, sin ningún tipo de predicción y por tanto divide la imagen en bloques de 8x8 pixels y realiza sobre cada uno de ellos una transformación DCT. La matriz de 64 coeficientes  $c(i, j)$  resultante es serializada mediante una lectura en zig-zag para ser transmitida, de forma que los coeficientes son ordenados de menor a mayor frecuencia y según su importancia son cuantificados no uniformemente  $c_q(i, j) = c(i, j)/Q(i, j)$ , siendo  $Q(i, j)$  el paso de cuantificación para el coeficiente  $c(i, j)$ ; dicho paso tiene en cuenta las respuesta de los estímulos visuales para cada frecuencia. Tras la cuantificación se realiza una codificación sin pérdidas, que en el decodificador se recupera, permitiendo reconstruir los coeficientes según  $c'(i, j) = c_q(i, j) Q(i, j)$ .

Los fotogramas procesados por el modo de operación *intraframe* son llamados fotogramas I y los procesados en modo *interframe*, son llamados fotogramas P o predichos.

Versiónes mejoradas de H.261 como H.263 y H.263+ están basadas en un esquema similar al descrito anteriormente, sin embargo H.261 sigue siendo el estándar para RDSI mientras que H.263 se emplea en RTC. Destacar que H.263 incluye varias mejoras frente a H.261, entre ellas la predicción avanzada (predicción a partir tanto de fotogramas anteriores como posteriores) y la estimación de movimiento tanto a nivel de macrobloque como bloque, con posibilidad de estimar incluso fuera del área visible.

En el anexo B se analizan diferentes secuencias de vídeo codificadas en H.263 y sometidas a diferentes probabilidades de error y diferentes grados de compresión.

### 3.4.2 Estándar MPEG-2

Los estándares MPEG han sido propuestos por el grupo internacional *Moving Pictures Experts Group*, creado bajo los auspicios de la Organización Internacional de Normalización (ISO) y de la Comisión Electrónica Internacional (IEC). ISO ha promulgado diversos estándares, entre ellos MPEG-1, MPEG-2, y en breve MPEG-7 y MPEG-4.

MPEG-1 ha sido desarrollado para almacenamiento de vídeo en formato CIF junto con audio simultáneamente, para capacidades de hasta 1.5 Mbps en diversos medios de almacenamiento digital, orientado a sistemas interactivos.

MPEG-2 descrito a continuación, ha sido el resultado de buscar posibles soluciones para aplicaciones de codificación de vídeo, que el estándar MPEG-1 no cubre. Las nuevas características añadidas por MPEG-2 tratan de alcanzar una funcionalidad y calidad suficientes con la mayor generalidad posible, de tal forma que se intenta en su estandarización integrar en una simple sintaxis los diversos requisitos de las distintas aplicaciones.

MPEG-3 fue diseñado originalmente para aplicaciones HDTV, sin embargo se observó que con una parametrización adecuada MPEG-2 podía cubrir estas necesidades; como consecuencia de esto el estándar MPEG-3 fue abandonado.

MPEG-4 basado en formato MPEG SIF (*Source Input Format*) busca la codificación de información audiovisual de resoluciones y objetos de una escena; además extiende el estándar para tasas de bits muy bajas (5-64 Kbps)(preve que se apruebe como estándar internacional en Noviembre de 1998).

MPEG-7 engloba el estándar para acceso a bases de datos multimedia.

### Introducción

MPEG-2 ISO/IEC13818 es un conjunto de estándares que dan servicio a un amplio rango de aplicaciones de diversas tasas de bit (2 – 20 Mbps) y configuraciones.

MPEG-2 establece una sintaxis para la representación del flujo de bits codificado y un método de decodificación, pero no de codificación. Con ello pretende obtener decodificadores económicos para usuarios, dando flexibilidad en la implementación, sin recaer en la complejidad o costes del codificador, generalmente ubicado en los centros de emisión y difusión[Tek97].

Algunas de las ventajas que posee MPEG-2 frente a H.261 y MPEG-1 son básicamente por un lado que permite la entrada de vídeo entrelazado y por otro la escalabilidad y adaptabilidad al tipo de decodificador y al ancho de banda del canal disponible, además de proporcionar algunas opciones de codificación y cuantificación mejoradas.

Uno de los inconvenientes de H.261 frente a las características propuestas por MPEG-2 es la editabilidad de la secuencia de vídeo ya que en H.261 sólo el primer fotograma está garantizado que sea *intraframe*, de forma que para extraer un fotograma de una secuencia todos los fotogramas anteriores han de ser decodificados, lo cual requiere un buffer de gran tamaño. MPEG soluciona este inconveniente introduciendo el concepto de GOP o grupo de fotogramas, ver figura 3.8, lo cual limita los requisitos de memoria.

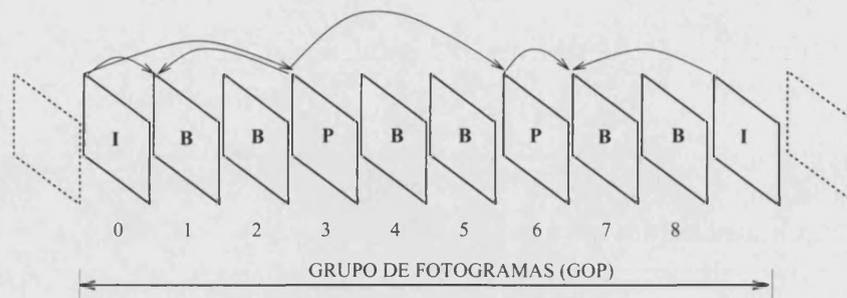


Figura 3.8: Estructura GOP de MPEG2: grupo de fotogramas

### Estructuras de datos

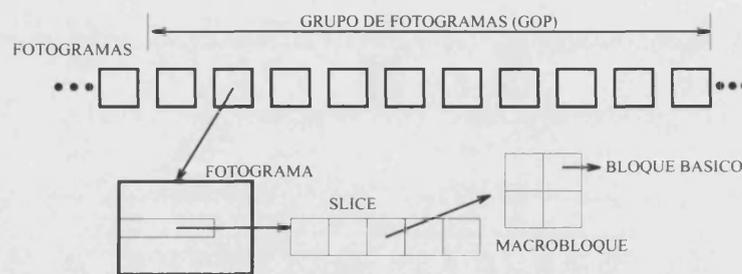


Figura 3.9: Estructura de datos en MPEG2: dependencias jerárquicas y sus relaciones

Como en el estándar H.261, el flujo de bits tras el compresor sigue una estructura de datos jerárquica. En MPEG-2 dicha estructura está formada por 6 tipos de datos (figura 3.9), donde 3 de ellas son comunes a H.261, la estructura básica, el MB y el fotograma.

Las estructuras y su funcionalidad son:

1. bloques básico de 8x8 pixels o unidad de la transformada DCT
2. macrobloques (MB) o 4 bloques básicos o unidad de estimación de movimiento
3. *slice*, estructura formada por un conjunto de MB contiguos horizontales, similar al GOB de H.261, cuya funcionalidad es la sincronización y recuperación frente a errores, de forma que su pérdida perjudique mínimamente a la estructura fotograma, evitando la propagación espacial del error
4. fotogramas, unidad primaria de pantalla, subdivididos en *slices* y que son clasificados según su codificación en tipos I, P (como en H.261) y B:
  - *Intraframes* o tipo I son codificados sin referencia a otros fotogramas y generan poca compresión, donde el coeficiente de componente continua  $c(0,0)$  tiene un tratamiento especial por su alta energía y correlación, codificando por tanto la diferencia  $\Delta_i = c_i(0,0) - c_{i-1}(0,0)$

- *Predichos* o tipo P son codificados con respecto al fotograma I o P anterior más próximo mediante compensación de movimiento
- *Bidireccionales* o tipo B son codificados a partir de fotogramas I o P más próximos, según la expresión

$$MB_t^e(x, y) = \alpha_1 \widehat{MB}_{t-k_1}(x - d1_x, y - d1_y) + \alpha_2 \widehat{MB}_{t+k_2}(x - d2_x, y - d2_y) \quad (3.6)$$

con  $\alpha_1 + \alpha_2 = 1$ , donde  $(d1_x, d1_y)$  y  $(d2_x, d2_y)$  son los vectores de movimiento para determinados MBs elegidos en los fotogramas previo  $MB_{t-k_1}$  y posterior  $MB_{t+k_2}$  dentro del GOP con referencias  $-k_1$  y  $+k_2$ , con las posibilidades:

- si  $\alpha_1 = 1$  y  $\alpha_2 = 0$  la estimación es hacia atrás o tipo P,
  - si  $\alpha_1 = 0$  y  $\alpha_2 = 1$  la estimación es hacia delante y
  - si  $\alpha_1 = 0.5$  y  $\alpha_2 = 0.5$  la estimación es hacia delante y hacia atrás
5. grupo de fotogramas (*group of pictures* o GOP), estructura formada por grupo de fotogramas considerada funcionalmente como la unidad mínima decodificable, la cual ha de incluir al menos un fotograma tipo I, ver figura 3.9
  6. secuencias de vídeo, formadas por varios GOPs, considerada funcionalmente como la secuencia completa visionable con un formato determinado

### Requisitos interactivos

MPEG-2 introduce una jerarquía de fotogramas dentro del GOP para mejorar tanto en compresión como en la edición de las secuencias.

Cabe destacar como regla general que una secuencia de vídeo codificada usando sólo fotogramas I, es decir ( I I I I I ...) similar a MJPEG, permite el más alto grado de acceso aleatorio y editabilidad pero no puede alcanzar una alta compresión, pues no incluye ningún tratamiento del movimiento. Una secuencia codificada con una inserción regular de fotogramas I y P pero no B, permite un cierto grado acceso aleatorio, al mismo tiempo una compresión moderada. Finalmente la incorporación de los tres tipos de fotogramas ( I B B P B B P B B I B B P...)(figura 3.8), permite alcanzar simultáneamente una alta compresión, un buen acceso aleatorio y una buena funcionalidad, aunque incrementa considerablemente el tiempo necesario para realizar la codificación debido a la dependencias entre los fotogramas con un retraso que no puede ser soportado en aplicaciones de videoconferencia o videotelefonía, aplicaciones de tiempo real en general.

El número de fotogramas de tipo I, P y B en un GOP depende de la aplicación, de los requisitos de la tasa de bits y del tiempo de acceso a los fotogramas. Las relaciones típicas entre los fotogramas dentro de un GOP (figura 3.8) son del tipo

$$IBBPBBPBBIBBPBBPBBI... \quad (3.7)$$

El estándar especifica que al menos uno de cada 132 fotogramas debe ser de tipo I para controlar la propagación de errores debido a posibles desacoplos entre

el codificador y decodificador. Dentro del GOP los fotogramas bidireccionales tienen especial interés, porque al no ser utilizados como referencia en el bucle de predicción (para no propagar errores) si se eliminan de la secuencia 3.7 de transmisión permiten disminuir el ancho de banda utilizado.

Por otra lado, la consecuencia de utilizar fotogramas bidireccionales hace que el orden de transmisión sea distinto de su orden natural y por ello el codificador y el decodificador deben reordenar los fotogramas. Por ejemplo si la cadena de fotogramas para visualización de un GOP es

$$I_1 B_2 B_3 P_4 B_5 B_6 P_7 \dots$$

tras el proceso de codificación queda

$$I_1 P_4 B_2 B_3 P_7 B_5 B_6 \dots$$

Además, la codificación de muchos fotogramas tipo B consecutivos hace que por una lado la distancia temporal entre dos fotogramas de referencia aumente, resultando en una pérdida de correlación temporal, y por otro lado introduce un retraso mayor, aumentando la cantidad de memoria requerida.

### Algoritmo de compresión

El algoritmo de compresión consta de los siguientes pasos:

1. decidir que fotogramas son de tipo I, P o B en un GOP.
2. en el caso de fotograma P o B, estimar un vector de movimiento, hacia atrás para cada MB en los fotogramas de tipo P y bidireccional en el caso de fotogramas de tipo B a partir de MB pertenecientes a fotogramas de tipo I o P.
3. determinar el modo de compresión para un determinado MB, bien con o sin predicción de movimiento según la calidad de la estimación

### Sincronización. Flujos de transporte y de programa

Como se ve en 3.4.2, MPEG-2 posee una estructura de datos jerárquica. Cada estructura tiene una cabecera con información y sincronización, seguida de elementos de jerarquía inferior. De esta manera, cada *secuencia* está formada por un número de GOPs y una de marca de fin de secuencia. A su vez cada GOP está formado por un determinado número de fotogramas. Cada fotograma a su vez, con *slices* y cada *slice* un número de MB...

Estas estructuras forman los flujos de transporte *Transport Stream* (TS) y de programa *Program Stream* (PS) que son los encapsulados finales en los cuales se introducen las secuencias de vídeo comprimido junto con las referencias temporales.

Los TS son flujos constantes de bits que para poder viajar por las redes se fragmentan en paquetes de longitud constante de 188 bytes; en el caso de no disponer

de información para transmitir son introducidos en el flujo paquetes vacíos. Los TS están optimizados para ser transmitidos junto con otros programas, tanto de audio como de vídeo con bases de tiempo diferentes, lo cual les permite ser transmitidos sin protocolos como RTP[Sch96].

Los PS están diseñados para el almacenamiento y procesado de un único programa con misma base de tiempos, utilizados habitualmente en los centros de producción de televisión[Tek97].

## Perfiles y niveles

MPEG-2 define cinco *perfiles* de funcionamiento: simple, principal, escalable SNR (o en frecuencia), escalable espacialmente y superior. Dentro de cada perfil de funcionamiento se distinguen 4 *niveles*: bajo, principal, alto para pantallas 1440x1152 y alto para pantallas 1920x1152; cada nivel especifica el rango dinámico que debe soportarse en los diferentes parámetros de implementación (tamaño de fotogramas, tasa de fotogramas por segundo, etc). MPEG-2 define que un decodificador con un cierto perfil y nivel determinado, ha de ser capaz de decodificar perfiles y niveles inferiores a él.

De este modo MPEG-2 permite tres tipos de *escalabilidad*, escalabilidad espacial, temporal y en frecuencia o *SNR*. Un sistema es escalable si, dado un flujo de bits comprimido, es capaz de seleccionar una parte de este flujo de forma sencilla y realizar el proceso de decodificación sin tener en cuenta el resto. Esta propiedad es interesante y muy útil sobre todo cuando el decodificador no tiene la misma velocidad de procesado que el codificador y tiene que manejar una cantidad de bits menor para mantener los mismos requisitos de tiempo real.

Los tipos básicos de escalabilidad mencionados consisten en

- *espacial y temporal* o adaptación de la resolución espacial (asociado al tamaño de fotograma) y temporal (asociada a los fotogramas por segundo) del vídeo a la potencia de procesado del decodificador y al ancho de banda del canal
- *frecuencial*, que se refiere a la capacidad de seleccionar de forma adaptativa, según la capacidad de canal o del decodificador, un determinado número de coeficientes de la transformada (asociados a las diferentes componentes de frecuencias del fotograma), de tal forma que permita obtener diferentes escalas de calidad de vídeo manteniendo la misma resolución espacial y temporal

Estas técnicas frecuenciales son conocidas como modos progresivos en la lectura de los coeficientes de cada bloque básico, implementados bien a través de métodos de selección espectral transmitiendo el mismo coeficiente  $c(i, j)$  para cada bloque, o bien por aproximaciones sucesivas transmitiendo para todos los bloques los coeficientes gradualmente desde el bit más significativo al menos significativo[RH96].

## 3.5 Conclusión

Se han analizado introduce las técnicas de compresión de vídeo frecuentemente utilizadas por los estándares, basadas fundamentalmente en la estimación de movimiento entre fotogramas para reducir la redundancia temporal y la descorrelación espacial por bloques de la imagen con la transformada discreta del coseno (DCT).

Estas técnicas generan dependencia en la estructura de datos utilizada por el compresor, que como se analiza en capítulos posteriores pueden presentar problemas para integrarse dentro de las redes con servicios *best effort*; dichos problemas están asociados a la calidad visual final de la secuencia de vídeo cuando es evaluada por el usuario.

## Capítulo 4

# Codificación en multiresolución. Codificación subbanda

El apartado 4.5 del presente capítulo ha sido realizado con la colaboración de Baltasar Beferull.

El objetivo de este capítulo es introducir un conjunto de herramientas conocidas como transformadas wavelet y codificación subbanda, para descomponer las secuencias de vídeo bajo dos criterios: que permitan introducir con mayor facilidad los criterios visuales y que se eviten las dependencias y jerarquías en la estructura de datos del sistema de transmisión de vídeo.

### 4.1 Introducción

Las técnicas de compresión y transmisión detalladas en el capítulo anterior son ampliamente utilizadas en sectores diferentes relacionados con servicios de radiodifusión o multimedia. Hay que destacar de los estándares y técnicas descritas, que existe un amplia gama de hardware que les permite aumentar en rendimiento y prestaciones.

Sin embargo los estándares tanto de compresión como las redes con servicios *best effort*, discutido a lo largo del presente trabajo, no están adaptados para trabajar conjuntamente *a priori*, ya que ambos estándares en su definición no contemplaban una posible vinculación.

Dichos inconvenientes tratan de ser resueltos por diferentes mecanismos en los nuevos estándares, incorporando técnicas mejoradas [RH96], pero manteniendo en definitiva la misma filosofía, por lo cual vuelven a tener en mayor o menor medida los mismos problemas. Obviamente estas soluciones se realizan con la intención de mantener la compatibilidad con los estándares promulgados y utilizar la infraestructura producida por los fabricantes.

Por tanto cabe destacar otra solución con un punto de partida distinto, y para ello se introducen técnicas diferentes a las vistas anteriormente con el objetivo de mantener calidad visual en dichos entornos de red. Estas técnicas utilizan transformaciones localizadas tanto en el espacio como en frecuencia.

Como es conocido, en la descomposición de una señal con funciones base perfectamente localizadas en el dominio de Fourier o de frecuencias, según el *principio de incertidumbre o de Heisenberg* estas quedan deslocalizadas en el espacio. De la misma manera que una imagen con una representación completa en el espacio no localiza en el dominio de Fourier.

El no poder localizar igual en ambos dominios genera problemas en la representación de las señales ya que no se puede capturar patrones localizados en un determinado soporte espacial con un ancho de banda también limitado. Para ello interesan bases que puedan representar la información en ambos dominios con la mínima redundancia posible, es decir, mantener dos representaciones muestreadas de la señal original pero con los mínimos coeficientes. Este límite viene fijado por el *principio de incertidumbre* y en sus proximidades se encuentran los efectos de la redundancia y el aliasing.

La relación entre el dominio espacial y frecuencial genera el concepto de textura (es importante destacar que las imágenes naturales pueden ser concebidas como suma de diferentes texturas). Se utilizan varias técnicas para la localización de texturas:

- basadas en aproximaciones globales estadísticas [R.M79]
- aproximaciones tanto en el dominio espacial como frecuencial

siendo estas últimas invariantes a los cambios de brillo, contraste o iluminación no uniforme, y por ello más interesantes.

Las técnicas para la localización espacio-frecuencia consisten en inventanar la señal, multiplicarla por una función ventana, siempre que la ventana tenga una integral finita y distinta de cero en un intervalo finito y aplicar sobre ella la transformada de Fourier.

Una de las ventanas más comunes es la ventana cuadrada como realizan los estándares del capítulo 3 al aplicar la DCT, sin embargo en procesado digital hay otros tipos de ventanas como Hanning, Hamming, Kaiser, triangular, sinc, gaussiana... con diferentes propiedades cada una de ellas [R.W96].

La transformada de Fourier de la señal inventanada, considerando la ventana invertida por temas de filtrado y convolución, es

$$TF(\tau, f) = \int_{-\infty}^{\infty} f(t)w^x(t - \tau)e^{-j2\pi ft} dt \quad (4.1)$$

presentada por [Gab46] posee dos variables  $(\tau, f)$  para la localización en el tiempo de lo que ocurre en la frecuencia, incrementando la redundancia con esta nueva representación si  $\tau$  es continuo.

Destacar que el superíndice  $x$  denota la inversión del eje temporal en el supuesto de que el inventanado se realice a través de una convolución con un filtro  $w(t)$  como se ve en el apartado 4.5.

Lo que ocurre en el dominio de Fourier es la convolución del espectro de la ventana con el espectro de la señal original. Sin embargo no es factible como se ha dicho, una perfecta localización, ya que la resolución tanto en el dominio espacial como en el de Fourier viene fijado por las características de la ventana seleccionada según su dispersión en el tiempo o en la frecuencia

$$\Delta t^2 = \frac{\int t^2 |w(t)|^2 dt}{\int |w(t)|^2 dt} \quad \text{y} \quad \Delta f^2 = \frac{\int f^2 |W(f)|^2 df}{\int |W(f)|^2 df} \quad (4.2)$$

que impiden discernir detalles en el espacio más pequeños de  $\Delta t$  y armónicos más próximos de  $\Delta f$ , que cumplen la *inecuación de Heisenberg*

$$\Delta t \Delta f \geq \frac{1}{4\pi} \quad (4.3)$$

Ambos denominadores en 4.2 son iguales por el *Teorema de Rayleigh*.

## 4.2 Transformada Wavelet

Por tanto para obtener máxima localización, la ventana ideal es aquella que tenga una fuerte caída tanto en el espacio como en el tiempo para localizar patrones y texturas en el dominio espacial, es decir, una señal pasabanda pura de energía finita.

Por otra parte es interesante analizar las propiedades allí donde se pretenden aplicar estas transformadas, en este caso las escenas naturales.

Las escenas naturales a comprimir o imágenes por lo general poseen un espectro de potencias intratable, que decae con  $1/f$ ; también poseen diferentes rangos de frecuencias, altas en detalles finos o bordes, al mismo tiempo que poseen componentes de muy bajo frecuencia en los fondos de escena, lo que obliga a pensar en ventanas adaptables al análisis y descomposición de la imagen.

El planteamiento es dada una ventana determinada que permita localización espacial y frecuencial, conseguir  $\frac{\Delta f}{f} = cte$  con el objetivo de tener una alta resolución en el tiempo para discontinuidades espaciales y que al mismo tiempo se mantenga una buena resolución en altas frecuencias.

Destacar que la descomposición de la señal en funciones del tipo anterior no admite representaciones de señales con componentes de alta frecuencia en un soporte espacial de grandes dimensiones, como pueda ser el ruido.

Para simplificar la notación 4.1 se toma

$$h(t) = w(t)e^{-j2\pi f_0 t} \quad (4.4)$$

Se puede escoger cualquier conjunto de ventanas que mantengan el ancho de banda relativo para descomponer la señal. Si se escala esa misma señal por cuestiones de coste computacional se obtiene

$$h_a(t) = \sqrt{|a|} h(at) \quad (4.5)$$

donde  $a$  es el factor de escala ( $f = f_0/a$ ), con lo que la nueva transformada de Fourier enventanada se convierte en

$$TF(\tau, a) = \sqrt{|a|} \int_{-\infty}^{\infty} f(t) h^x(a(t - \tau)) dt \quad (4.6)$$

La notación más utilizada por su sencillez es tomar  $h_{a,\tau} = \sqrt{|a|} h(a(t - \tau))$  de tal manera que con dilataciones y traslaciones se cubren ambos dominios. La expresión 4.6 queda

$$TF(\tau, a) = \int_{-\infty}^{\infty} f(t) h_{a,\tau}^x(t) dt \quad (4.7)$$

que es conocida como *transformada continua Wavelet* y su *inversa* para reconstrucción de la señal es

$$f(t) = k \int TF(\tau, a) h_{a,\tau}(t) da d\tau \quad (4.8)$$

siendo  $k$  una constante que depende de  $h(t)$ . En este caso el superíndice  $x$  no aparece por motivos que se analizan en el proceso de análisis y síntesis, apartado 4.3.

La  $h(t)$  es conocida como *función madre wavelet* si cumple las propiedades de ventana ideal vistas antes, es decir

$$\int h^2(t) dt < \infty \text{ y } \int h(t) dt = 0 \quad (4.9)$$

### 4.2.1 Transformada Wavelet discreta

Los parámetros  $a, \tau$ , correspondientes a la dilatación y traslación respectivamente, pueden ser discretizados con el fin de poder procesar computacionalmente con una base discreta de funciones y reducir la redundancia de la representación continua con valores  $a = a_0^i$  y  $\tau = jT/a_0^i$  con  $i, j$  enteros [Dau90]:

$$h_{ij}(t) = a_0^{i/2} h(a_0^i t - jT) \quad (4.10)$$

siendo los coeficientes discretos de la transformada wavelet

$$c_{ij} = \int f(t) h_{ij}^x(t) dt \quad (4.11)$$

que aproxima a la señal

$$f(t) \simeq k \sum_i \sum_j c_{i,j} h_{ij}(t) \quad (4.12)$$

de forma similar a 4.8. Si  $a_0 \rightarrow 1$  y  $T \rightarrow 0$  se obtiene el caso continuo.

Por analogías de procesamiento con el SVH, capítulo 5, el muestreo más utilizado corresponde con el *muestreo diádico* que modela la resolución en octavas, es decir las frecuencias y el espacio escaladas en factor de 2 y por tanto se fija  $a_0 = 2$  y  $T = 1$ .

Si se sigue la notación de Mallat&Zhong [MSZ92], utilizando como localización espacial la variable  $t$ , se puede considerar la función wavelet

$$\psi_i(t) = 2^{i/2} \psi(2^i t) \quad (4.13)$$

que debe cumplir la propiedad de paso-banda  $\int \psi(t)dt = 0$ , valor medio nulo.

Los coeficientes wavelet de  $f(t)$  a escala  $2^i$ ,  $i \in Z$  se expresan como se ha visto antes

$$c_i(t) = \int_{-\infty}^{\infty} f(\tau)\psi_i^x(\tau - t)d\tau = \int_{-\infty}^{\infty} f(\tau)\psi_i(t - \tau)d\tau \quad (4.14)$$

que según el *Teorema de Plancherel*, se puede expresar en el dominio de Fourier

$$TF[c_i(t)] = TF(f(t))\Psi\left(\frac{f}{2^i}\right) \quad (4.15)$$

Si se impone la condición

$$\forall f, A \leq \sum_{i=-\infty}^{\infty} |\Psi\left(\frac{f}{2^i}\right)|^2 \leq B, A > 0, B < \infty \quad (4.16)$$

es decir que en todo el espacio de frecuencias la suma de todas las dilataciones de  $\psi(t)$  tenga una norma finita, entonces bien la  $TF(f(t))$  o la misma  $f(t)$  se pueden recuperar con la función wavelet de reconstrucción  $\chi(t)$

$$f(t) = \sum_{i=-\infty}^{\infty} c_i(t)\chi_i(t) \quad (4.17)$$

si cumple

$$\sum_{i=-\infty}^{\infty} \Psi\left(\frac{f}{2^i}\right)X\left(\frac{f}{2^i}\right) = 1 \quad (4.18)$$

que puede demostrarse a partir del *Teorema de Parseval* junto con 4.16 y 4.17, que es una representación estable

$$A \|f()\|^2 \leq \sum_{i=-\infty}^{\infty} \|c_i(t)\|^2 \leq B \|f()\|^2 \quad (4.19)$$

y puede ser tan próxima a  $f()$  en  $L^2(\mathcal{R})$  si  $A/B \rightarrow 1$ .

## 4.2.2 Redundancia y muestreo crítico

El muestreo crítico es aquel que genera el mismo número de muestras en la entrada de la descomposición de la señal que en la salida, sin embargo cuando se incorpora redundancia a la nueva representación, es decir que el número de muestras es mayor, se trabaja con muestreo no crítico.

Existen muchas funciones que cumplen las condiciones 4.19 y por su construcción pueden ser redundantes, es decir que las funciones base no sean linealmente independientes. Según de la relación de los valores  $A$  y  $B$  con  $A > 0, B < \infty$ , se obtienen propiedades diferentes.

Si  $A = B = 1$  forman una base ortonormal, si  $A = B \neq 1$  forman un *tight frame* o marco hermético, si  $A \approx B$  forman un *snug frame* o marco justo, o si  $A \neq B$  forman un *single frame* o marco aislado.

Excepto el caso ortogonal, las demás representaciones son redundantes, que en algunos casos pueden ser interesantes por las propiedades de: mejor localización, disminución del aliasing, invarianza frente a traslaciones y robustez ante pérdidas; esta última propiedad se deriva de que la información perdida puede ser obtenida por combinación de otros coeficientes.

## 4.3 Multiresolución

Al introducir el concepto de multiresolución se introduce un nuevo nivel de estructuración, ya que ahora aparece una autosimilitud a diferentes escalas de resolución entre las funciones base.

La multiresolución es una de las técnicas para obtener expansiones lineales de señales en bases de tipo wavelet, localizadas en ambos dominios y con diferentes resoluciones espaciales para cada banda de frecuencia, es decir patrones de alta frecuencia en una duración espacial pequeña y grande en baja frecuencia.

Las propiedades que trata de introducir son por un lado, la suavidad en las bases utilizadas para asegurar una estabilidad numérica donde la cuantificación introduzca el menor número de artefactos posibles y por otro lado, la aproximación para conseguir una representación compacta de forma que con un pequeño número de coeficientes permita representar la señal con un error controlado.

El análisis multiresolución da lugar a subespacios de diferentes resoluciones, los cuales representan proyecciones del espacio continuo  $L^2(\mathcal{R})$  (representación wavelet continua) con coeficientes que generan una expansión lineal discreta de wavelet (transformada wavelet discreta), representando la señal a partir de varias aproximaciones sucesivas, manejando tanto aproximaciones pasa-baja como los detalles de la misma.

Desde el punto de vista de la compresión, únicamente se manejan estos coeficientes discretos, sin tener en cuenta los subespacios continuos subyacentes de diferente resolución ya que los coeficientes contienen toda la información necesaria de la representación wavelet continua.

En el apartado 4.5, se estudia la estrecha relación entre la representación multiresolución y las transformadas rápidas wavelet basadas en iteración de bancos de filtros [Mal89b], y como bajo ciertas condiciones los bancos de filtros pueden generar representaciones de este tipo [Dau92].

En primer lugar se formula la definición axiomática de un análisis multiresolución, introducida por Mallat [Mal89b], con la incorporación de subespacios complementarios ortogonales y en segundo lugar se analiza el caso correspondiente a bases wavelet biortogonales.

### 4.3.1 Definición formal

Un análisis multiresolución ortogonal para  $L^2(\mathcal{R})$  consiste en una secuencia de espacios de aproximación sucesivos  $V_j$  que satisfacen las siguientes condiciones:

- *inclusión de subespacios: mínima*  $V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \dots$  *máxima resolución*
- *completitud hacia arriba*,  $m \in \mathcal{Z}$   $V_{-m} \rightarrow L^2(\mathcal{R})$  para  $m \rightarrow \infty$  : que los subespacios contienen funciones base arbitrariamente cercanas a cualquier función de cuadrado integrable
- *completitud hacia abajo*,  $m \in \mathcal{Z}$   $V_m = \{0\}$  para  $m \rightarrow \infty$ : que los subespacios  $V_{-m}$  pueden ser tan pequeños como sea requerido
- *multiresolución*  $f(x) \in V_m \iff f(2^m x) \in V_0$ : todos los espacios son versiones escaladas del espacio central  $V_0$  y por tanto genera una relación a través de escalado entre funciones pertenecientes a subespacios de diferente resolución.
- *desplazamiento*  $f(x) \in V_0 \implies f(x - n) \in V_0 \forall n \in \mathcal{Z}$
- *función escalado*  $\varphi(x) = \varphi_{0,0} \in V_0$  : de modo que el conjunto de funciones desplazadas  $\{\varphi_{m,n}(x) = 2^{m/2} \varphi(2^m x - n) \mid n \in \mathcal{Z}\}$  es una base ortonormal de  $V_m$ .

De la definición axiomática de multiresolución puede inferirse, a partir de la propiedad de inclusión entre subespacios y la condición de escalado, la existencia de una relación lineal entre  $\varphi(x)$  y  $\varphi(2x - n)$  puesto que  $V_0$  está incluido en  $V_{-1}$ .

Si  $\varphi(x)$  pertenece a  $V_0$ , debe pertenecer también a  $V_{-1}$  y por tanto  $\varphi(x)$  ha de expresarse como una combinación lineal de funciones base pertenecientes a  $V_{-1}$ ; así la función de escalado  $\varphi$  debe cumplir la siguiente ecuación, llamada *ecuación de dilatación o refinamiento*

$$\varphi(x) = \sqrt{2} \sum_{n=-\infty}^{n=\infty} h_0[n] \varphi(2x - n) \quad (4.20)$$

Si se pasa la ecuación de dilatación al dominio de Fourier

$$\Phi(\omega) = \frac{1}{\sqrt{2}} H_0(e^{j\omega/2}) \Phi(\omega/2) \quad (4.21)$$

donde  $H_0(e^{j\omega}) = \sum_{n \in \mathcal{Z}} h_0[n] e^{-j\omega n} = TF(h_0[n])$ .

Observando la ecuación de dilatación en el dominio de Fourier, la respuesta en frecuencia de la función base correspondiente al espacio de menor resolución abarca un rango de frecuencias mitad que el correspondiente a las funciones base de mayor resolución, porque al descomponer la señal con un análisis multiresolución con un factor de escala igual a 2, se obtiene una descomposición frecuencial en octavas.

Si se aplica la condición de ortonormalidad en  $V_m$  y el *Teorema de Rayleigh*, se obtiene

$$\langle \varphi_{0,0}, \varphi_{0,n} \rangle = \frac{1}{2} \int |\Phi(\omega/2)|^2 e^{jn\omega} d\omega \quad (4.22)$$

que es nula cuando  $n \neq 0$  y dividiendo el eje de frecuencias en tramos de  $2\pi$

$$\langle \varphi_{00}, \varphi_{0n} \rangle = \frac{1}{2} \int_0^{2\pi} e^{jn\omega} \sum_k \left| \Phi\left(\frac{\omega + 2\pi k}{2}\right) \right|^2 d\omega \quad (4.23)$$

se obtiene la condición necesaria de ortonormalidad por la *fórmula de Poisson* con la expresión

$$\sum_{k=-\infty}^{k=\infty} \left| \Phi\left(\frac{\omega + 2\pi k}{2}\right) \right|^2 = 1 \quad (4.24)$$

Esta condición  $2\pi$  periódica gobierna la estabilidad y la ortonormalidad en la multiresolución de las traslaciones de  $\varphi$ .

La ortonormalidad puede relajarse a exigir únicamente que  $\{\varphi_{0,n}; n \in \mathcal{Z}\}$  sea una *base de Riesz* [Mal89b], a partir de la cual puedan definirse otras funciones  $\varphi_{0,n}^*$  las cuales sí que forman una base ortonormal; las conclusiones en este caso son las mismas.

Otra condición necesaria para obtener ortonormalidad se extrae de la ecuación 4.24 y la ecuación 4.21 haciendo uso de la periodicidad, como

$$|H_0(e^{j\omega})|^2 + |H_0(e^{j(\omega+\pi)})|^2 = 2 \quad (4.25)$$

Si las funciones de escalado según la definición de multiresolución han de representar cualquier  $L^2(\mathcal{R})$  por la inclusión de subespacios y la completitud hacia arriba, entonces

$$\int \varphi(x) dx \neq 0 \quad (4.26)$$

que además caracteriza su suavidad y continuidad, que expresado en el dominio de Fourier,  $\Phi(0) \neq 0$ , la componente continua sea no nula y por tanto considerando 4.21

$$|H_0(1)| = \sqrt{2} \quad (4.27)$$

y considerando que los detalles de la máxima resolución son pasobanda, en  $\omega = \pi$ , se obtiene

$$H_0(-1) = 0 \quad (4.28)$$

quedando así perfectamente descrito el análisis multiresolución por la función  $2\pi$  periódica  $H_0(e^{j\omega})$ .

### 4.3.2 Subespacios complementarios ortogonales

Particularizando para el *caso ortogonal*, entre un subespacio  $V_m$  y el subespacio  $V_{m-1}$ , debido a que el segundo subespacio incluye al primero y la representación es ortogonal, queda un subespacio llamado  $W_m$  que representa los detalles a añadir a las funciones pertenecientes al subespacio  $V_m$  para pasar al nivel de resolución siguiente  $V_{m-1}$  (figura 4.1). Esta idea es la que justifica la *función wavelet*, a partir de la cual se construyen por escalado y traslación las funciones que generan estos subespacios  $W_m$ .

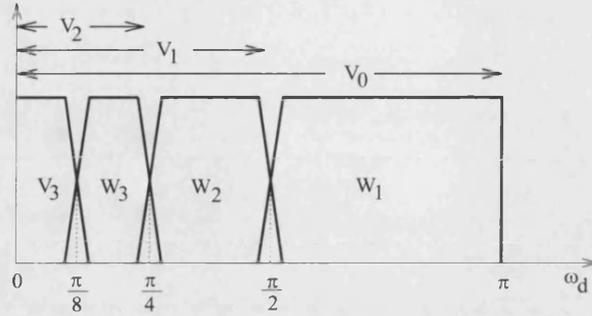


Figura 4.1: Descomposición espectral de subespacios para las técnicas de multiresolución

Al pasar de una resolución mayor a otra menor, se puede separar por una parte la versión aproximada que pasa a menor resolución y por otra parte el detalle que se pierde de esa resolución.

Por tanto se puede demostrar formalmente que para  $L^2(\mathcal{R})$ , dada una secuencia de subespacios cumpliendo los axiomas de multiresolución, existe una base ortonormal

$$\psi_{m,n}(x) = 2^{m/2} \psi(2^m x - n), \quad m, n \in \mathcal{Z} \quad (4.29)$$

tal que  $\{\psi_{m,n}\}$ ,  $n \in \mathcal{Z}$  es una base ortonormal en  $W_m$ , donde  $W_m$  con  $m \in \mathcal{Z}$  es el complemento ortogonal de  $V_m$  en  $V_{m-1}$ , de tal forma que cumple  $V_{m-1} = V_m \oplus W_m$  con  $W_m \perp W_{m'}$  si  $m \neq m'$ .

Por tanto para toda  $f \in L^2(\mathcal{R})$  donde  $P_m f$  representa la proyección ortogonal de  $f$  sobre  $V_m$ , se puede expresar como

$$P_{m-1} f = P_m f + \sum_{n \in \mathcal{Z}} \langle f, \psi_{m,n}^x \rangle \psi_{m,n} \quad (4.30)$$

donde el significado del sumatorio consiste en la incorporación de los detalles en la nueva representación a partir de una resolución menor.

Esto da lugar a que para cualquier  $V_m$ ,  $m < M$ , se exprese como la suma de un subespacio de una determinada aproximación  $V_M$ , más todos los subespacios de detalle hasta dicha resolución

$$V_m = V_M \oplus W_{M-1} \oplus W_{M-2} \oplus \dots \oplus W_{m+1} \quad (4.31)$$

que junto con las propiedades de *completitud hacia arriba* y *hacia abajo* implica que

$$L^2(\mathcal{R}) = \bigoplus_{m \in \mathcal{Z}} W_m \quad (4.32)$$

es decir, que los subespacios  $W_m$  forman una descomposición de  $L^2(\mathcal{R})$  en subespacios mutuamente ortogonales con la propiedad de escalado  $f(x) \in W_m \iff f(2^m x) \in W_0$  lo cual permite concluir que si  $\{\psi_{0,n}; n \in \mathcal{Z}\}$  es una base de  $W_0$ , entonces,  $\{\psi_{m,n}; m, n \in \mathcal{Z}\}$  es una base de  $W_m$ .

Los coeficientes wavelet  $w_{ij}(f) = \langle f, \psi_{ij} \rangle$  por tanto representan a la función  $f \in L^2(\mathcal{R})$  cuya norma puede ser expresada como

$$\|f\|^2 = \sum_{ij} |w_{ij}|^2 \quad (4.33)$$

De la misma manera que se había supuesto en 4.20, la *función madre wavelet*  $\psi(x) \in W_0 \subset V_{-1}$  se puede poner como combinación lineal

$$\psi(x) = \sqrt{2} \sum_{n=-\infty}^{n=\infty} g_0[n] \varphi(2x - n) \quad (4.34)$$

conocida como *ecuación wavelet*. Tomando la transformada de Fourier, se obtiene

$$\Psi(\omega) = \frac{1}{\sqrt{2}} G_0(e^{j\omega/2}) \Phi(\omega/2) \quad (4.35)$$

Por otra parte, el hecho de que  $\psi(x) \in W_0$ , siendo ortogonal a  $V_0$ , es decir considerando funciones complejas

$$0 = \langle \varphi_{0n}, \psi \rangle = \frac{1}{2} \int H_0(e^{j\omega/2}) G_0^*(e^{j\omega/2}) |\Phi(\omega/2)|^2 e^{jn\omega} d\omega \quad (4.36)$$

y análogamente a 4.23, independientemente de  $n$

$$0 = \frac{1}{2} \int_0^{2\pi} e^{jn\omega} \sum_k H_0(e^{j(\omega/2+\pi k)}) G_0^*(e^{j(\omega/2+\pi k)}) |\Phi(\frac{\omega}{2} + \pi k)|^2 d\omega \quad (4.37)$$

El hecho de que esta integral sea nula obliga a que el sumatorio también lo sea

$$\sum_k H_0(e^{j(\omega/2+\pi k)}) G_0^*(e^{j(\omega/2+\pi k)}) |\Phi(\frac{\omega}{2} + \pi k)|^2 = 0 \quad (4.38)$$

y agrupando los valores con  $k$  par e impar y aplicando periodicidad, se obtiene que entre  $G_0(e^{j\omega})$  y  $H_0(e^{j\omega})$  existe la siguiente relación

$$G_0(e^{j\omega}) H_0^*(e^{j\omega}) + G_0(e^{j(\omega+\pi)}) H_0^*(e^{j(\omega+\pi)}) = 0 \quad (4.39)$$

que permite expresar

$$G_0(e^{j\omega}) = -\lambda(e^{j\omega}) H_0^*(e^{j(\omega+\pi)}) \quad (4.40)$$

siendo  $\lambda(e^{-j\omega})$  una función  $2\pi$ -periódica tal que cumple la siguiente ecuación

$$\lambda(e^{j\omega}) + \lambda(e^{j(\omega+\pi)}) = 0 \quad (4.41)$$

que permite varias posibilidades, ya que cualquier función del tipo  $\lambda(e^{j\omega}) = \rho e^{\pm j(2K-1)\omega}$  con  $K \in \mathcal{Z}$  y  $\rho \in \mathcal{R}$  con una indeterminación en la fase y en la amplitud es válida. Si se elige por  $k=0$ ,  $\rho = -1$  por tanto  $\lambda = -e^{-j\omega}$ , entonces se obtiene

$$G_0(e^{j\omega}) = -e^{-j\omega} H_0^*(e^{j(\omega+\pi)}) \quad (4.42)$$

que en el dominio espacial es equivalente a

$$g_0[n] = (-1)^n h_0^*[-n+1] \quad (4.43)$$

conocidos como *mirror filter* o filtros espejo.

Por tanto las funciones wavelet  $\{\Phi_{m,n}\}$  se obtendrán de realizar desplazamientos y escalados de la *función madre wavelet*, que expresada tanto en el dominio frecuencial, como en el dominio espacial depende de las funciones de escalado

$$\begin{aligned} \Psi(\omega) &= -\frac{1}{\sqrt{2}} H_0^*(e^{j(\omega/2+\pi)}) \Phi(\omega/2) \\ \psi(x) &= \sqrt{2} \sum_{n \in \mathcal{Z}} (-1)^n h_0[1-n] \varphi(2x-n) \end{aligned} \quad (4.44)$$

Las condiciones expuestas sobre la función de escalado 4.24, son condiciones necesarias para un análisis multiresolución ortonormal pero no lo garantizan, o al menos su estabilidad; varios autores [Dau88][Mal89a] han planteado en diferentes estudios condiciones suficientes.

### 4.3.3 Inconvenientes de la ortogonalidad

La representación ortogonal en un análisis multiresolución posee ventajas como la conservación de energía y la simplicidad en la generación de las bases, pero al ser implementada mediante banco de filtros iterados acarrea graves inconvenientes como la incompatibilidad con el concepto de fase lineal y bases de longitud finita (filtros FIR) a excepción de los filtros Haar. Las respuestas impulsionales de los filtros ortogonales no pueden tener simetría par o impar, han de ser asimétricas, sin fase lineal[GS96].

Si estos filtros no pueden ser de fase lineal, entonces la distorsión de fase afectará poco si la cuantificación es fina. Por contra si la cuantificación no es tan fina (caso de altos factores de compresión), entonces aparece una distorsión fuerte debido a la no linealidad de la fase o a la asimetría de los filtros introduciendo distorsiones asimétricas que son percibidas por el SVH.

Una solución a los inconvenientes de la ortogonalidad se presenta a continuación mediante transformadas wavelet biortogonales que, aun sin conservar la energía en los coeficientes, son muy apropiadas para la compresión de imagen o vídeo.

## 4.4 Transformada Wavelet Biortogonal

Las transformadas biortogonales plantean la elección de un conjunto de funciones base discretas para la expansión lineal de una secuencia  $x[n]$  sobre dos bases  $\{\varphi_k[n]\}$  y  $\{\tilde{\varphi}_k[n]\}$ , una para análisis y otra para síntesis

$$x[n] = \sum_{k \in \mathcal{Z}} \langle x[n], \varphi_k[n] \rangle \tilde{\varphi}_k[n] \quad (4.45)$$

que cumplen

$$\langle \varphi_k[n], \tilde{\varphi}_l[n] \rangle = \delta[k - l] \quad (4.46)$$

con la propiedad característica de localización espacio-frecuencial, para generar una representación más compacta.

Esta nueva representación más general ofrece una solución a los inconvenientes de la ortogonalidad, ya que sin conservar completamente la energía permite aproximarse a una *reconstrucción perfecta*.

Esta propiedad introduce inconvenientes cuando se utiliza en la compresión de audio, puesto que en estos casos se necesita conservar la energía de las subbandas únicamente posible a través de transformadas ortogonales, ya que el oído humano es más sensible a estos artefactos; sin embargo el SVH sí que admite representaciones mediante transformadas biortogonales.

## 4.5 Banco de Filtros y codificación subbanda

La codificación subbanda consiste en descomponer la señal como suma de distintas señales con respuestas espectrales diferentes, lo que se conoce como subbandas. Diferentes técnicas han sido utilizadas para implementar dicha codificación como la transformada híbrida *HINT* o la *pirámide Laplaciana*.

La implementación computacional de las transformadas wavelets, tanto las ortogonales como las biortogonales, en un análisis multiresolución se puede realizar a través de bancos de filtros iterados, es decir realimentando una de sus salidas con la entrada.

Si se observa la función de inventanado utilizada en 4.1

$$\langle f(t), \psi_i^x(t - \tau) \rangle = \int_{-\infty}^{\infty} f(\tau) \psi_i(t - \tau) d\tau = f * \psi_i(t) \quad (4.47)$$

equivale a un filtrado espacial donde el superíndice  $x$  indica por la operación convolución, que la señal está invertida en el eje temporal.

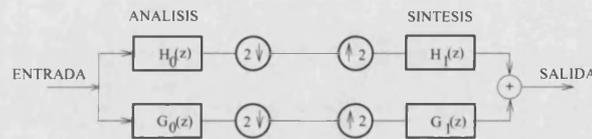


Figura 4.2: Descomposición con banco de filtros para 1D

Un banco de filtros (figura 4.2) está constituido por un conjunto de filtros junto con operadores de submuestreo o sobremuestreo, que en el dominio de Fourier descomponen la señal de entrada en diferentes subbandas de frecuencia complementarias; esto es lo que se conoce comúnmente como descomposición o codificación subbanda y cuyo número de subbandas depende de la descomposición a realizar.

En el caso más general de una representación biortogonal existen dos tipos de filtros, los filtros aplicados al analizar la señal o filtros de análisis y los filtros para reconstruir la señal a partir de las diferentes subbandas o filtros de síntesis. En el caso

particular de filtros ortogonales ambos filtros coinciden con una transformación espejo 4.43.

Si los filtros de análisis y síntesis no coinciden, han de garantizar una reconstrucción perfecta y la condición 4.46.

Denotaremos como  $h_0[n]$  y  $g_0[n]$  los filtros de análisis y como  $h_1[n]$  y  $g_1[n]$  los filtros de síntesis. A la salida de los filtros de análisis se obtiene el contenido de la señal original repartida en dos subbandas, con aproximadamente la mitad de ancho de banda de la original, realizando la función de canales de frecuencia.

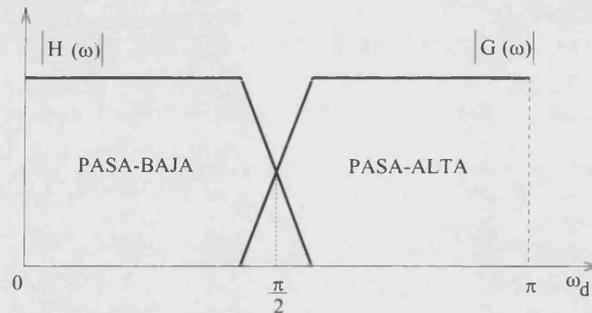


Figura 4.3: Espectro del filtro paso-baja y paso-alta para multiresolución

La operación de submuestreo por 2 es compatible con el hecho de que al dividir el espectro en dos partes, conservando cada una de ellas la mitad de la información, hace que con aproximadamente la mitad de muestras se pueda recuperar el dominio original. No son exactamente la mitad porque los filtros no son ideales y existe como se observa un aliasing (figura 4.3) entre las respuestas frecuenciales de  $h_0[n]$  y  $g_0[n]$  que debe ser eliminado en el proceso de síntesis. Como se ve a continuación, son los filtros de síntesis los encargados de invertir el proceso anulando el aliasing y dando lugar a una reconstrucción perfecta.

La relación existente, debido a la correlación, entre los filtros de análisis  $h_0[n]$  y  $g_0[n]$  y las funciones base de análisis  $\{\varphi_k[n]\}$ , y entre los filtros de síntesis  $h_1[n]$  y  $g_1[n]$  y las funciones base de síntesis  $\{\tilde{\varphi}_k[n]\}$  es

$$\begin{aligned} h_0[n] &= \varphi_0[-n] & g_0[n] &= \varphi_1[-n] \\ h_1[n] &= \tilde{\varphi}_0[n] & g_1[n] &= \tilde{\varphi}_1[n] \end{aligned} \quad (4.48)$$

que en el caso ortogonal se traduce en

$$h_1[n] = h_0[-n] = \varphi_0[n] \quad g_1[n] = g_0[-n] = \varphi_1[n] \quad (4.49)$$

donde los filtros de síntesis coinciden con las funciones base de análisis al aplicar secuencialmente la propiedad 4.47.

Considerando el efecto del submuestreo en las funciones base

$$\begin{aligned} \varphi_{2k}[n] &= h_0[2k-n] \\ \varphi_{2k+1}[n] &= g_0[2k-n] \\ \tilde{\varphi}_{2k}[n] &= h_1[n-2k] \\ \tilde{\varphi}_{2k+1}[n] &= g_1[n-2k] \end{aligned} \quad (4.50)$$

que para el caso ortogonal

$$\begin{aligned}\varphi_{2k}[n] &= h_1[n - 2k] \\ \varphi_{2k+1}[n] &= g_1[n - 2k]\end{aligned}\quad (4.51)$$

Se dice que la representación obtenida es crítica u ortogonal si se obtiene un número de coeficientes igual al número de muestras de la señal original  $x[n]$  a la salida de los bancos de filtros, después de aplicar los filtros  $h_0[n]$  y  $g_0[n]$  y submuestrear; mientras que si el número de coeficientes obtenido es mayor que el número de muestras de la señal original, se dice que es sobrecompleta y por tanto las funciones base de análisis  $\{\varphi_k[n]\}$  no son linealmente independientes entre sí y hay una redundancia implícita en ellas. En el caso de bancos de filtros biortogonales, se puede obtener tanto una representación crítica como sobrecompleta.

### 4.5.1 Reconstrucción perfecta

El proceso de convolución de los filtros se implementa matricialmente de forma que la transformada wavelet de una señal  $x$  se descompone en una aproximación paso baja  $y$  y un residuo o paso alta  $z$

$$\begin{aligned}y &= H_0x \\ z &= G_0x\end{aligned}\quad (4.52)$$

donde la reconstrucción  $x = H_1y + G_1z$ . En particular para el caso ortogonal  $H_1 = H_0^T$  y  $G_1 = G_0^T$  se puede obtener

$$x = H_0^T H_0 x + G_0^T G_0 x \quad (4.53)$$

de manera que para obtener una reconstrucción perfecta hay que exigir

$$H_0^T H_0 + G_0^T G_0 = 1 \quad (4.54)$$

Estudiando la reconstrucción perfecta en el dominio de la transformada  $z$ , los filtros  $h_0[n]$  y  $h_1[n]$  son canales pasa-baja mientras que  $g_0[n]$  y  $g_1[n]$  son canales pasa-alta.

El procesado de la señal  $x[n]$  a través del banco de filtros, matemáticamente se expresa como el producto de los filtros de análisis y la señal original. Por el *Teorema de Plancherel* y considerando el diezmado se obtiene

$$\begin{aligned}(\downarrow 2)H_0(z)X(z) &= \frac{1}{2}(H_0(z^{\frac{1}{2}})X(z^{\frac{1}{2}}) + H_0(-z^{\frac{1}{2}})X(-z^{\frac{1}{2}})) \\ (\downarrow 2)G_0(z)X(z) &= \frac{1}{2}(G_0(z^{\frac{1}{2}})X(z^{\frac{1}{2}}) + G_0(-z^{\frac{1}{2}})X(-z^{\frac{1}{2}}))\end{aligned}\quad (4.55)$$

que al sobremuestrear y filtrar con los filtros de síntesis queda

$$\begin{aligned}H_1(z)(\uparrow 2)(\downarrow 2)H_0(z)X(z) &= \frac{1}{2}H_1(z)(H_0(z)X(z) + H_0(-z)X(-z)) \\ G_1(z)(\uparrow 2)(\downarrow 2)G_0(z)X(z) &= \frac{1}{2}G_1(z)(G_0(z)X(z) + G_0(-z)X(-z))\end{aligned}\quad (4.56)$$

de manera que al recomponer la señal como suma de la aproximación ( $H$ ) y del detalle ( $G$ ), la resultante del procesado completo del banco de filtros es

$$\frac{1}{2}X(z)(H_1(z)(H_0(z) + G_1(z)(G_0(z))) + \frac{1}{2}X(-z)(H_1(z)H_0(-z) + G_1(z)G_0(-z)) \quad (4.57)$$

correspondiendo cada uno de los sumandos introducidos en la expresión anterior a las condiciones de distorsión nula, que se expresan más detalladamente como y aliasing nulo

$$H_1(z)H_0(z) + G_1(z)G_0(z) = 2 \quad (4.58)$$

$$H_1(z)H_0(-z) + G_1(z)G_0(-z) = 0 \quad (4.59)$$

similares a los obtenidos en 4.25.

Esta última condición es muy crítica ya que controla la distorsión no armónica que degrada la reconstrucción de la señal. Si se permite un retraso  $l$  de la señal original a la salida del banco de filtros, se puede obtener una condición más relajada para la condición de distorsión nula

$$H_1(z)H_0(z) + G_1(z)G_0(z) = 2z^{-l} \quad (4.60)$$

que en el caso ortogonal no puede ser satisfecha por filtros FIR a excepción de los filtros Haar.

## 4.5.2 Resultados para banco de filtros biortogonales

Las restricciones vistas anteriormente se recogen en diferentes metodologías para el cálculo de los filtros, utilizando descomposiciones de polinomios trigonométricos con términos  $e^{-ikx}$ , introduciendo diferentes grados de aproximación de los filtros y criterios de regularidad.

Una de las metodologías propuestas [Dau92], parte de aproximaciones del polinomio  $P_0(z) = H_1(z)H_0(z)$  introduciendo las condiciones 4.58 y 4.59 para reconstrucción perfecta

$$P_0(z) - P_0(-z) = 2z^{-l} \quad (4.61)$$

e insertando un número determinado  $p$  de ceros en  $z = -1$  para que el filtro  $H_1(e^{j\omega})$  (filtro paso baja) tenga una respuesta plana alrededor de  $\omega = \pi$ , que impide tener una respuesta abrupta fuera de su banda pasante, respuesta más estable, lo que implica que  $P_0(z)$  sea de la forma

$$P_0(z) = (1 + z^{-1})^{2p} Q(z) \quad (4.62)$$

Por razones de coste computacional, como el número de ceros se relaciona con la longitud de los filtros, se toma  $p = 2$  y se centra el diseño tomando  $P(z) = P_0(z) z^l$ , siendo  $l = 3$  con lo que

$$P(z) = z^2(1 + z^{-1})^2(1 + z^{-1})^2 z Q(z) = (1 + z)^2(1 + z^{-1})^2(az + b + cz^{-1}) \quad (4.63)$$

que al aplicar la condiciones de reconstrucción perfecta 4.58 y 4.59

$$P(z) + P(-z) = 2 \quad (4.64)$$

y resolviendo el sistema da como solución el polinomio

$$P(z) = z^3(1 + z^{-1})^4 \left( \frac{-1}{16} + \frac{1}{4}z^{-1} + \frac{-1}{16}z^{-2} \right) \quad (4.65)$$

y despejando  $P_0(z)$

$$P_0(z) = \frac{1}{16}(-1 + 9z^{-2} + 16z^{-3} + 9z^{-4} - z^{-6}) \quad (4.66)$$

del cual hay que despejar dos factores según el *lema de Riesz* para obtener los filtros para  $H_0(z)$  y  $H_1(z)$ . Tomando los filtros ortogonales y los de fase lineal de menor longitud (caso de filtros biortogonales), cuyos análisis temporal y frecuencial están detallados en el anexo C, se obtienen los filtros siguientes

- caso ortogonal o filtro de Daubechies D4

$$H_1(z) = \frac{1}{4\sqrt{2}}(1 - z^{-1})^2(1 + \sqrt{3} + (1 - \sqrt{3})z) \quad (4.67)$$

- primer caso biortogonal, filtro biortogonal(1)

$$\begin{aligned} H_0(z) &= \frac{1}{4}(1 + 2z^{-1} + z^{-2}) \\ H_1(z) &= \frac{1}{4}(-1 + 2z^{-1} + 6z^{-2} + 2z^{-3} - z^{-4}) \end{aligned} \quad (4.68)$$

- segundo caso biortogonal, filtro biortogonal(2)

$$\begin{aligned} H_0(z) &= \frac{1}{4}(1 + 3z^{-1} + 3z^{-2} + z^{-3}) \\ H_1(z) &= \frac{1}{4}(-1 + 3z^{-1} + 3z^{-2} - z^{-3}) \end{aligned} \quad (4.69)$$

## 4.6 Transformada wavelet multidimensional separable

Los bancos de filtros vistos antes han sido desarrollados para una dimensión, pero en el caso de compresión de imagen o vídeo se requieren transformadas waveletes de dos o tres dimensiones respectivamente.

Por cuestiones computacionales, para evitar convolucionar con matrices de varias dimensiones que ralentizarían el sistema de compresión, es necesario implementar el filtrado de manera separada para cada una de las dimensiones y que permite implementar el sistema en procesamiento *pipeline*. La transformada de varias dimensiones realizada separadamente conlleva los inconvenientes de la sección 4.3.3.

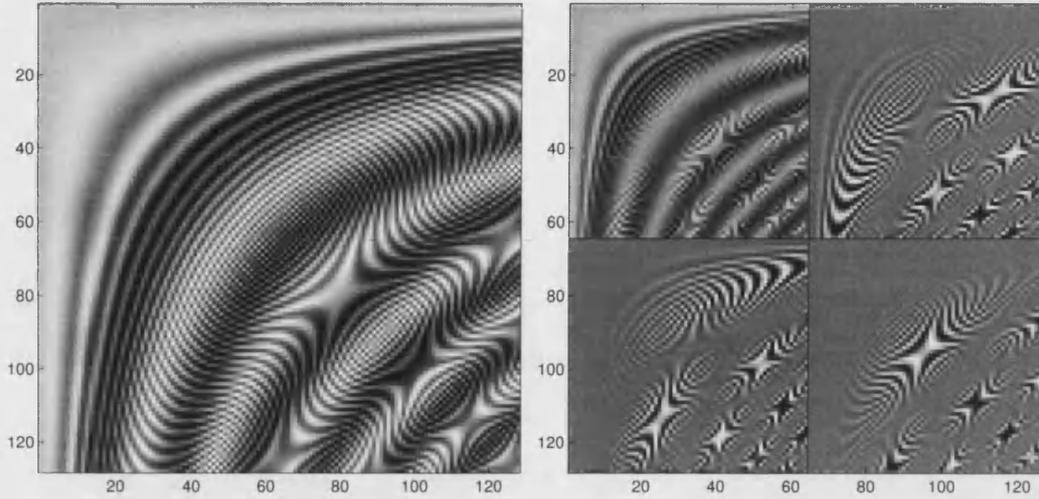


Figura 4.4: Imagen original y su descomposición 2D con filtros biortogonales(2): residuo, detalle horizontal, vertical y diagonal

Si se observa el estudio de multiresolución para dos dimensiones (2D)  $\mathbf{V}_{m-1} = \mathbf{V}_m \oplus \mathbf{W}_m$  se puede expresar como producto de los subespacios de cada una de las dimensiones  $\mathbf{V}_{m-1} = V_{m-1} \otimes V_{m-1}$  que descomponiendo en los diferentes subespacios

$$\begin{aligned} \mathbf{V}_{m-1} &= (V_m \oplus W_m) \otimes (V_m \oplus W_m) = \\ &= V_m \otimes V_m \oplus [(W_m \otimes V_m) \oplus (V_m \otimes W_m) \oplus (W_m \otimes W_m)] \end{aligned} \quad (4.70)$$

por tanto las bases que descomponen cada uno de los subespacios son

$$\begin{aligned} \varphi(x, y) &= \varphi(x)\varphi(y) \\ \psi^h(x, y) &= \varphi(x)\psi(y) \\ \psi^v(x, y) &= \psi(x)\varphi(y) \\ \psi^d(x, y) &= \psi(x)\psi(y) \end{aligned} \quad (4.71)$$

conocidos como aproximación, residuo o paso baja, detalle horizontal, detalle vertical y detalle diagonal respectivamente.

Un ejemplo de descomposición 2D diádica con el filtro 4.69 sobre una imagen de test se muestra en la figura 4.4; la parte izquierda es la señal original que se descompone en la parte derecha, en cuatro subbandas organizadas en el mismo soporte de la imagen asociadas a los filtros  $\varphi, \psi^h, \psi^v, \psi^d$ .

Por extensión al ejemplo anterior y de forma análoga se obtiene una descomposición 3D mediante un banco de filtros wavelet diádico, ver parte derecha de la figura 4.5 (destacar que el diezmo por 2 queda implícito en cada filtro) y a su izquierda se presenta la descomposición en el dominio de Fourier para dos niveles de resolución; dicho banco de filtros permite la descomposición de secuencias de vídeo con la aplicación de los filtros temporal, vertical y horizontal.

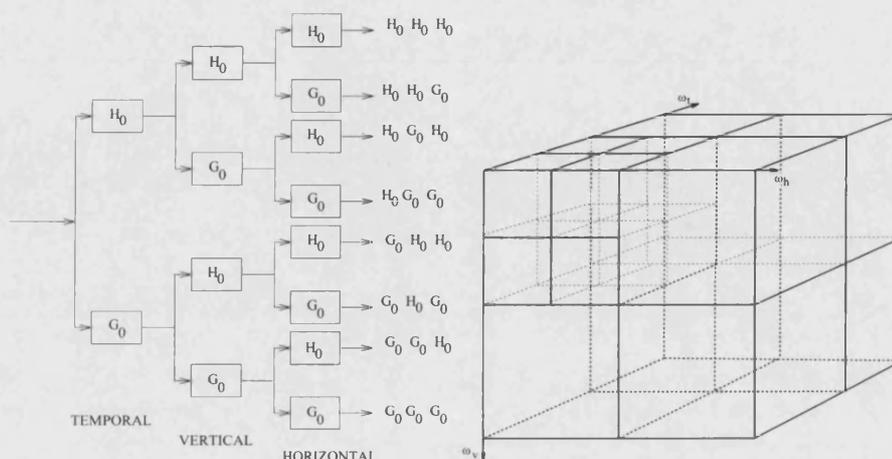


Figura 4.5: Banco de filtros para 3D y su descomposición en Fourier

$$\begin{aligned}
 \varphi(x, y, t) &= \varphi(x)\varphi(y)\varphi(t) \\
 \psi^1(x, y, t) &= \varphi(x)\varphi(y)\psi(t) \\
 \psi^2(x, y, t) &= \varphi(x)\psi(y)\varphi(t) \\
 \psi^3(x, y, t) &= \varphi(x)\psi(y)\psi(t) \\
 \psi^4(x, y, t) &= \psi(x)\varphi(y)\varphi(t) \\
 \psi^5(x, y, t) &= \psi(x)\varphi(y)\psi(t) \\
 \psi^6(x, y, t) &= \psi(x)\psi(y)\varphi(t) \\
 \psi^7(x, y, t) &= \psi(x)\psi(y)\psi(t)
 \end{aligned} \tag{4.72}$$

### 4.6.1 Correlación del eje temporal

La transformación 3D separable de una secuencia de vídeo permite reducir la redundancia del eje temporal.

Uno de los inconvenientes que posee el realizar la transformada wavelet sobre el eje temporal es el número de fotogramas a almacenar para convolucionar con los filtros tanto de análisis como de síntesis, sin embargo aporta similitudes al proceso del SVH como se ve en el capítulo 5.

El estudio que se realiza en el presente trabajo de investigación, desarrolla una descomposición wavelet sobre el eje temporal con filtros de mínima longitud y ortogonales, los filtros Haar.

Otra alternativa posible se centra en la estimación del movimiento de una transformación 2D, incorporando directamente sistemas de predicción temporal similares a las técnicas analizadas en el capítulo 3, pero en las que existen problemas de adaptabilidad, debido al solapamiento producido en la descomposición wavelet, por la existencia de discontinuidades e irregularidades en el campo de movimiento en dicha representación; se observa que vectores de movimiento de bloques vecinos apuntan a direcciones diferentes y por tanto provocan ruido considerable en la imagen reconstruida.

Para evitar dichas discontinuidades cuando se aplican los métodos tradicionales

de estimación de movimiento se utilizan correspondencias de bloques  $2N \times 2N$  con solapamiento alrededor de núcleos  $N \times N$ , por ejemplo con la función peso en el área de búsqueda  $-N < x, y < N$  de la forma

$$w(x, y) = \cos^2\left(\frac{x\pi}{N}\right) \cos^2\left(\frac{y\pi}{N}\right) \quad (4.73)$$

Destacar que este campo actualmente está en fase de estudio.

## 4.7 Conclusión

Las técnicas basadas en las transformadas wavelet permiten descomponer las secuencias de vídeo de forma separable mediante bancos de filtros en cada uno de los tres ejes temporal, vertical y horizontal. Para ello se han diseñado y analizado dos filtros biortogonales tanto para el proceso de análisis como de síntesis.

Esta descomposición localizada en el dominio espacial y de Fourier permite como se ve en el capítulo siguiente, modelar las respuestas de las neuronas que intervienen en la percepción visual.

El objetivo por tanto perseguido es aproximar a descomposiciones análogas al sistema visual humano, para poder introducir un mejor criterio de calidad en la transmisión de vídeo sobre redes con servicios *best effort*. Con ello se permite:

- controlar mejor el proceso de compresión al que se someten las secuencias para ser transportadas por la red
- introducir mejores criterios para priorizar la información a transmitir
- descomponer e independizar la información de las secuencias para evitar dependencias
- mejorar potencialmente las técnicas vistas en el capítulo anterior (discutidas en el capítulo 6)



## Capítulo 5

# Sistema visual humano

Algunos aspectos del presente capítulo han sido discutidos en colaboración con Baltasar Beferull y Jesús Malo.

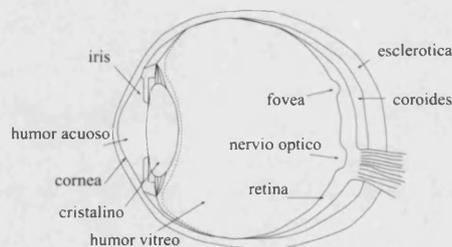


Figura 5.1: *El ojo humano*

El sistema visual humano es uno de los órganos sensoriales más completo y que más información aporta (figura 5.1) con un 80% de la información que percibimos y que exige una capacidad de proceso cerebral del 20%, parte de ella realizada inconscientemente. La capacidad de proceso requerida por el SVH puede ser comparada frente al 3% del procesamiento analítico, utilizado por ejemplo para jugar al ajedrez [Jen93].

La visión humana es compleja y en ningún caso es objetivo de esta tesis dar una explicación médica detallada ni rigurosa de los mecanismos ópticos y neuronales que mueven al SVH, mas bien hacer cita y extraer las conclusiones de las referencias más relevantes.

## 5.1 Descripción anatómica

La visión humana está formada por un sistema óptico o sistema paso-baja y diferentes capas de células y neuronas con funcionalidades diferentes. El proceso de sensación visual se divide en las siguientes etapas:

- *formación de la imagen* en la retina por transformaciones ópticas
- *codificación* a través de diferentes capas neuronales antes de llegar a la zona del *cortex cerebral*
- *representación* o procesado de diferentes *capas corticales*, con funcionalidades diferentes: detección, discriminación y reconocimiento de patrones simples
- *interpretación*, correspondiente a un nivel superior donde el cerebro asocia propiedades perceptuales a las sensaciones como color, movimiento o forma.

Las tres primeras etapas son conocidas como visión temprana (*early stage*) y a este nivel de representación es posible predecir como funciona el SVH con cierta exactitud frente a una determinada señal[Wat90].

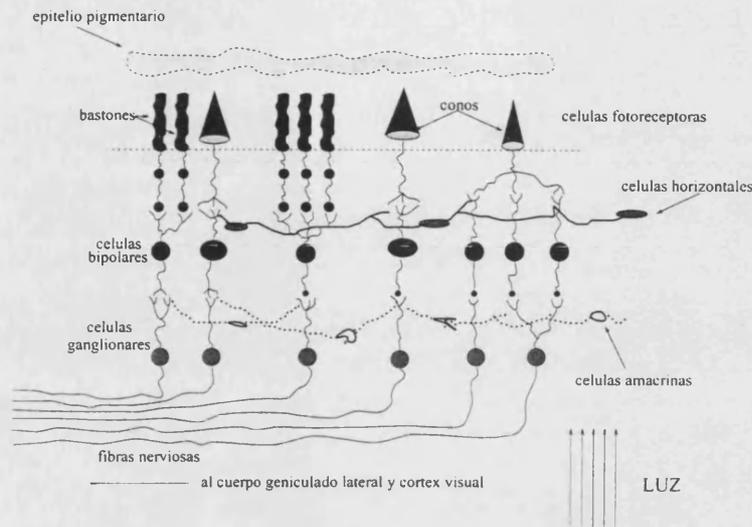
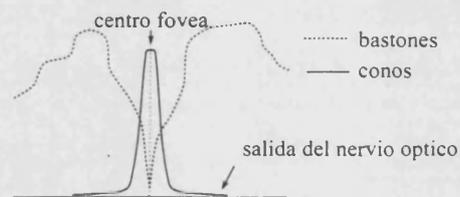
Cuando un observador se mueve y explora el entorno, la luz procedente del espectro visible se refleja en los objetos que le rodean y llega hasta los ojos. Tras atravesar una serie de capas ópticas, la córnea y el cristalino, la fuente de luz reflejada o estímulo visual se proyecta sobre el fondo del ojo o la retina.

El estímulo visual en el ojo está constantemente cambiando y de esta manera el estímulo visual es capaz de percibir la distribución espacial de la escena y sus variaciones. Este estímulo junto con la visión binocular, hace experimentar en el cerebro un sentido de movimiento a través de un espacio tridimensional.

El sistema visual procesa la información de la retina según sus proyecciones, es decir, el campo visual derecho del observador o hemicampo derecho se proyecta tanto en la retina del ojo derecho como del izquierdo, en el primer caso se conoce como hemiretina nasal del ojo derecho y en el segundo hemiretina temporal del ojo izquierdo. De la misma manera el campo visual izquierdo o hemicampo izquierdo se proyecta sobre la hemiretina temporal derecha y sobre la hemiretina nasal izquierda. El procesado que realiza el SVH es a nivel de hemicampo completo y es llevado a cabo en el hemisferio cerebral opuesto al de la proyección, es decir, el hemicampo completo derecho se procesa en el hemisferio izquierdo o viceversa.

La retina (figura 5.2) es una superficie fotosensible que transforma la imagen proyectada sobre ella en una serie de impulsos que son transmitidos por el nervio óptico hasta el cerebro. Esta superficie fotosensible está formada por un serie de células, unas sensibles a la luminancia o nivel de gris y otras al color, llamadas bastones y conos respectivamente, que sinaptan con las células bipolares. Los bastones están repartidos casi uniformemente en la retina a excepción del centro o fovea, mientras que los conos se concentran en ella (figura 5.3)[Mar94].

Los bastones están asociados en grupos a células bipolares mientras que los conos sinaptan individualmente con ellas. Por esta razón los bastones son menos sensibles,

Figura 5.2: *La retina humana*Figura 5.3: *Distribución de conos y bastones en la retina humana*

pero debido a la gran cantidad de ellos, entre 75 y 150 millones, permiten una buena agudeza visual bajo condiciones de poca luminosidad, sin percibir color, sólo brillo, claridad o sensación de luz.

Los conos, alrededor de 6 a 7 millones, son células sensibles a unas ciertas frecuencias de la radiación luminosa, distinguiéndose en sensibles al rojo, verde y azul. Al estar más concentrados en la fovea, hacen que la visión en colores sea mejor cuando se centra la mirada en un objeto determinado, mientras que la visión periférica por el contrario es más borrosa y sin distinción de colores. Los conos individualmente son menos sensibles que los bastones y por tanto tienen una respuesta temporal más rápida [Mar94].

Los bastones sinaptan a través de las amacrinas AII con las células ganglionares para transmitir sus impulsos luminosos.

Las células bipolares de los conos sinaptan con las ganglionares por vía directa; estas vías directas son moduladas por vías indirectas tanto para las células bipolares por células horizontales como para las células ganglionares por células amacrinas, ver figura 5.2.

Las vías directas e indirectas generan campos receptores tanto excitadores (ON) (figura 5.4) como inhibidores (OFF) (figura 5.5), con respuestas determinadas por la

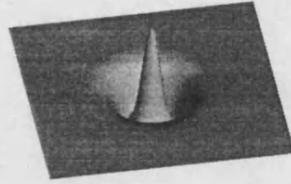


Figura 5.4: *Campo receptor ON en la retina y CGL*



Figura 5.5: *Campos receptores OFF en la retina y CGL*

distribución espacial de luz sobre ellos de forma antagonista centro-periferia.

Dichos campos receptores son invariantes a la rotación, solapados y de diferente tamaño, pequeños para los conos de la zona foveal y grandes en la periferia de la retina de manera que son selectivos a diferentes frecuencias espaciales. Los campos receptores pequeños son excitados por las altas frecuencias espaciales y los campos receptores grandes por las bajas o fondos de escena.

La sinapsis con las células horizontales regula el contraste entre las zonas antagonistas de cada campo receptor para facilitar la distinción de contornos por contraste espacial, es decir el contraste entre el objeto y el entorno o inhibición lateral.

Los campos receptores bipolares se mantienen en las conexiones sinápticas con las células ganglionares; estos campos proyectan separadamente al cuerpo geniculado lateral (CGL) (figura 5.6) con información relativa al hemicampo contralateral pero de forma independiente, es decir, en el CGL izquierdo confluyen los campos receptores de la hemiretina nasal derecha y la hemiretina temporal izquierda pero separadamente, y viceversa.

El potencial graduado de las células bipolares cuando llega a las ganglionares genera a su nivel, un potencial de acción de amplitud constante independiente de la intensidad del estímulo hasta el CGL.

La sinapsis entre las células ganglionares y el CGL se realiza a través de dos canales nerviosos, el parvosistema y el magnosistema. Un 80%, por el parvosistema con células ganglionares P que vinculan a los campos receptores pequeños con respuestas mantenidas, lentas, que intervienen en la alta resolución y análisis del color y que permiten gran agudeza visual. El resto de sinapsis se realiza por el magnosistema

con células ganglionares M que vinculan a los campos receptores grandes, más rápidos y menos agudos que los anteriores, pero con gran contraste luminoso que responden mejor a imágenes grandes espacialmente en movimiento. El CGL por tanto queda dividido en zona magnocelular y en zona parvocelular (figura 5.6)[Wat90].

Existen otros dos tipos de células ganglionares de interés, unas que responden al movimiento en una dirección con un sentido y no lo hacen al movimiento en sentido contrario, es decir con selectividad direccional asociado a la amacrina colinérgica, y otras que responden a la oponencia de color, rojo-verde (R,G) y amarillo-azul (R+G,B).

Desde el CGL las neuronas geniculadas proyectan vía radiaciones ópticas hacia la corteza visual primaria (V1 o área 17) en la región occipital del cerebro, que es el área más estudiada en el cortex de los mamíferos y sobre la que se puede hablar con mayor seguridad. Estas proyecciones conservan el sistema retinotópico, es decir con representación completa de un mapa ordenado de la proyección de la retina.

El V1 está formado por células estrelladas que sinaptan con células simples y complejas, estudiadas por Hubel y Wiesel. Las células estrelladas responden al contraste bien definido con campos receptores de sólo un ojo. Por encima y debajo de estas, conectadas por axones, se encuentran las células corticales simples y complejas, ambas más exigentes respecto al estímulo visual.

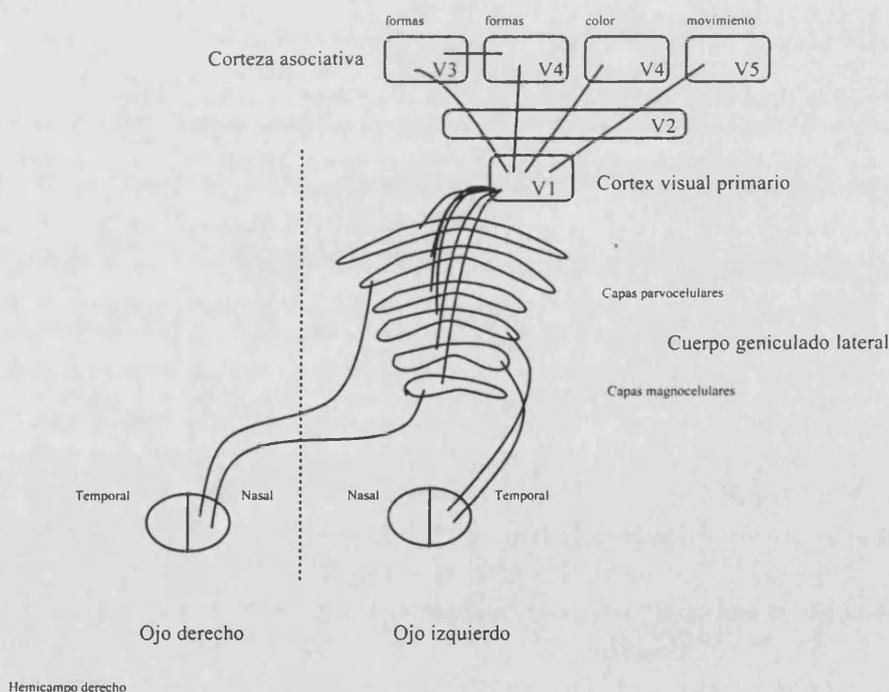


Figura 5.6: *Sistema visual humano*

Las células simples con campos receptores mayores que las estrelladas, están dispuestas en bandas ON y OFF alargadas formadas por la fusión de campos receptores convergentes de mismo centro desplazadas verticalmente en la retina, figura 5.7. Son



Figura 5.7: Campos receptivos direccionales del cortex visual

<i>funcionalidad</i>	<i>células asociadas</i>
contraste luminoso	bipolares, ganglionares
segmentos, líneas y límites	estrelladas, simples y complejas
orientación en un abanico de posiciones	complejas
cambio en los bordes	hipercomplejas

Tabla 5.1: Organización jerárquica de células en SVH.

excitadas con barras o líneas, y límites generados por contraste de colores o luz-oscuridad.

Las células complejas son similares a las simples, a excepción de que las zonas excitadoras e inhibitoras no están tan definidas y son más grandes. Reciben información de varias células simples con mismo eje desplazadas sobre una línea horizontal en la retina y fusionan la información de ambos ojos. El desplazamiento lineal es un buen estímulo para las células complejas, aunque también responden bien: a bordes y máximo si son perpendiculares al eje de simetría de las células, a diferentes orientaciones dentro de un abanico de posiciones, a la sensación de profundidad y al movimiento al fusionar la información binocular.

Tanto las células simples como complejas perciben la forma independientemente de la ubicación en la retina de manera selectiva en frecuencia, ya que las dimensiones espaciales de los campos receptores varían, en referencia al sistema retinotópico.

La capa V1 del *cortex visual* sinapta con V2 formada por células hipercomplejas que fusionan varias complejas aumentando su campo receptor y perciben información en los cambios de los bordes.

En la tabla 5.1 se muestra la relación entre las diferentes células y sus funcionalidades, observando que la información visual es generada a medida que se acerca al cerebro por estímulos más precisos y complejos.

Desde las capas V1 y V2, por encima de la organización jerárquica, existen 4 sistemas paralelos de procesamiento diferente (figura 5.6), independientes y pertenecientes a la *corteza asociativa o preestriada* (Semir Zeki)[Wat90]:

- para el color y niveles de gris (V4), con información procedente del sistema parvocelular sin poseer un nivel de resolución elevado

- para el movimiento, la localización y la organización espacial (V5 o MT) con información de células selectivas a la dirección, procedentes de capas magnocelulares con neuronas rápidas sin capacidad de distinguir colores y contrastes
- dos para la forma (V3 y V4), que responden a líneas espacialmente orientadas independientemente del color y de la forma dinámica.

La integración de los 4 sistemas se realiza por redes de enlaces o *sistema aferente-eferente*.

## 5.2 Modelos del SVH con descomposición sub-banda

Las funcionalidades vistas de las diferentes células y neuronas del SVH quedan reflejadas empíricamente por análisis, mediciones fisiológicas, estudios y efectos producidos por diferentes patologías [Mar94]. A partir de toda esta información, diferentes estudios tratan de modelizar el SVH y acotar su respuesta mediante expresiones analíticas que contemplan los diferentes comportamientos.

Históricamente Gabor [Gab46] introdujo una ventana gaussiana en la fase de descomposición de la señal, ventana que tiene un comportamiento mejor que la ventana cuadrada utilizada en JPEG, dando lugar a la llamada transformada enventanada de Fourier o transformada Gabor, óptima desde el punto de vista de localización espacio-frecuencial puesto que minimiza el *principio de incertidumbre* del capítulo 4.

Varias medidas psicofísicas [HW68] demuestran con bastante seguridad, que los campos receptivos de las células corticales (figura 5.7), tienen una forma muy similar a funciones de Gabor en 2D en el cortex visual primario y 3D en las zonas encargadas del análisis de movimiento.

### 5.2.1 Percepción de imágenes estáticas

Hubel y Wiesel estudiaron en [HW68] desde el punto de vista fisiológico las células presentes en el cortex visual primario y observaron que las células simples tenían un comportamiento aproximadamente lineal, caracterizadas por una suma localizada y pesada de los valores de intensidad de un estímulo visual, con campos receptores en diferentes niveles de resolución, con una forma muy parecida a filtros pasa-banda o función wavelet.

Watson [Wat87a][Wat90] modeló el proceso realizado por las células del cortex visual primario a través de una transformada bidimensional no separable a diferentes niveles de resolución por octavas, utilizando funciones de Gabor y realizando un submuestreo hexagonal. Esta transformada inherentemente redundante alrededor del 80%, óptima en localización espacio-frecuencia, permitía que en el dominio de frecuencia existieran canales en cada resolución, orientados 0, 45, 90 y 135 grados, consiguiendo gran similitud con las respuestas de las células simples.

En el cortex visual primario se han encontrado éstas y otras orientaciones diferentes, lo cual determina que hay una cierta redundancia que justifica la invarianza inherente en el SVH.

### 5.2.2 Percepción del movimiento

Diferentes modelos de percepción de movimiento empezaron a obtenerse posteriormente, los cuales incluían también un análisis multiresolución.

Teo y Heeger [Hee92][Hee94] obtuvieron un modelo de percepción que trataba de encontrar las respuestas en el cortex visual primario teniendo en cuenta tanto las células simples como las complejas; las células pertenecientes al área V2, sugiriendo una operación no lineal sobre la salida de V1, con el objetivo de detectar la localización y el movimiento de los contornos de las texturas. Este modelo estaba basado en una transformación subbanda tridimensional redundante utilizando pirámides *steerable*, invariantes a rotación y traslación.

### 5.2.3 Transformada de Gabor

Los planteamientos que sugieren estos trabajos quedan suficientemente completos desde nuestro punto de vista con el estudio de esta transformada.

La transformada de Gabor es un caso particular de especial interés en las representaciones wavelet, es la descomposición de la señal en bases estructuradas similares a las bases implementadas por el SVH, que permite eliminar la información redundante subjetiva en la compresión.

En la transformada Gabor la señal es filtrada por canales de varias resoluciones y orientaciones obtenidas a partir de rotaciones y escalados de la llamada función de Gabor, que consiste en un armónico de Fourier multiplicado por una gaussiana (figura 5.8 y 5.9).

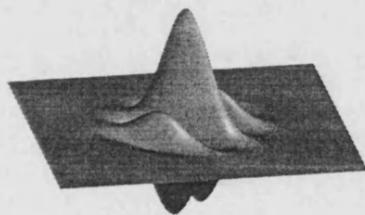
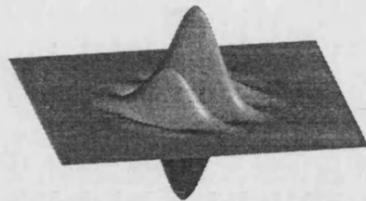


Figura 5.8: *Función Gabor Par*

Las funciones de Gabor no permiten diseñar transformadas wavelet discretas con muestreo crítico (*Teorema de Low-Balian*), debido a una falta de estabilidad numérica en la transformada wavelet inversa, necesaria para reconstruir la señal. Sin embargo,

Figura 5.9: *Función Gabor Impar*

permiten diseñar una transformación wavelet redundante que permite reconstruir la señal de manera invariante a la translación [VK95].

Las representaciones mediante funciones gabor corresponden a la igualdad en la ecuación 4.3.

La expresión analítica de la función de Gabor generalizada para 2D con localización, en coordenadas polares en el dominio de Fourier  $(f_0, \theta_0)$  y cartesianas en el espacial  $(x_0, y_0)$  es:

$$g_{x_0, y_0, f_0, \theta_0} = a e^{j(2\pi f_0(x \sin \theta_0 + y \cos \theta_0) + \phi)} e^{-\pi a^2((x-x_0) \cos \theta_0 + (y-y_0) \sin \theta_0)^2 + \gamma^2((x-x_0) \sin \theta_0 - (y-y_0) \cos \theta_0)^2} \quad (5.1)$$

donde  $a$  define el ancho de banda de la gaussiana,  $\gamma$  el grado de simetría de la gaussiana en 2D y  $\phi$  el desfase.

El valor  $a$  se puede determinar por  $a = 0.71 f_0$  para obtener un ancho de banda radial constante de una octava en una escala logarítmica y ancho de banda angular de 0.71 radianes [NRT91].

Esta transformada, aunque presenta ciertos inconvenientes para la compresión, entre ellos el no ser paso banda pura y la falta de ortogonalidad que dificulta la obtención de los coeficientes, sin embargo ha sido pieza clave en la aplicación de las transformadas wavelet a los sistemas de procesado de imagen y vídeo.

### 5.3 Enmascaramiento del ruido o distorsión en el SVH

La redundancia subjetiva o perceptual reside en las características, sensibilidad y comportamiento del SVH tanto en el dominio espacial como en el dominio frecuencial y por ello es necesario estudiar el enmascaramiento perceptual del ruido asociado a la compresión de tal forma que sea lo menos perceptible posible, es decir, sea incapaz de distinguir entre una componente perteneciente a la propia señal y la perteneciente al ruido o distorsión. Este enmascaramiento puede hacer, incluso con relación señal-ruido modesta o baja, que la percepción de dicha distorsión en la señal reconstruida después de una compresión sea nula.

Muchos experimentos psicovisuales y psicofísicos han demostrado que la percepción que tiene el SVH depende de la distribución en frecuencia de la distorsión. Dichos experimentos están basados en determinar la precisión con la cual el SVH es capaz de percibir diferentes tipos de señales.

La percepción en el SVH depende mucho de la distribución en frecuencia espacial o temporal de la distorsión. Kelly y Savoie midieron el contraste  $C = \frac{\Delta B}{B}$  y su umbral a diferentes frecuencias espaciales mediante patrones, similares a las bases de la DCT, que modulaban en amplitud  $\Delta B$  sin modificar las frecuencias temporales, fijando una valor de  $B$  luminancia promedio. Estas frecuencias espaciales venían medidas en ciclos/grado observando un comportamiento paso-banda.

Los aspectos temporales se introdujeron con señales sinusoidales del tipo  $f(x, t) = \cos(\alpha(x - vt))$ , siendo  $\omega = v \alpha$  la frecuencia temporal y  $v$  la velocidad, para evaluar los efectos transitorios. Los resultados obtenidos fueron que los aspectos espaciales y temporales en la visión humana no son independientes y son mucho más fuertes a bajas frecuencias, mientras en altas frecuencias tanto espaciales como temporales, son casi independientes.

Este contraste determina que las neuronas necesitan un contraste mínimo para poder responder, también conocido como umbral absoluto. La inversa de este valor se conoce como sensibilidad al contraste o CSF (*Contrast Sensitive Function*)[H.S89][BM93][kel79].

Por tanto, tras esta primera experimentación, si las señales que recibe el SVH son transformadas en el dominio de Fourier, cabe pensar que los coeficientes de la descomposición mantienen una fuerte relación con el umbral absoluto, dando paso a un cuantificador que determina el paso inicial en la amplitud de los coeficientes para no ser codificados en la zona muerta del cuantificador; si los coeficientes no superan su umbral forman parte de redundancia subjetiva a eliminar en un proceso de compresión[Art95].

La existencia de tolerancias a cambios en el contraste de las funciones base DCT implica que el SVH relaciona el contraste desde un dominio continuo en un dominio discreto o conjunto finito de percepciones, que determina cierta redundancia subjetiva en el SVH. La experimentación llevada a cabo por Vries-Rose, Weber y Watson en este caso, mide la capacidad que tiene el SVH para detectar una mínima variación incremental (umbrales incrementales) de contraste en una determinada señal, la cual en un principio posee un determinado contraste.

Esto permite introducir una métrica [JM97] que minimiza la distorsión subjetiva de la señal reconstruida en el proceso de compresión, que difiere de los métodos matemáticos tradicionales de medida de calidad visual con las expresiones 3.1 y 3.2.

Por otra parte, esta experimentación determina un mecanismo para la asignación de bits por coeficiente en la descomposición según el contraste y la frecuencia, determinado por la función IAF (*Information Allocation Function*)[Art95].

La expresión IAF

$$IAF(f, C) = \frac{d^2 I}{df dC} \quad (5.2)$$

donde  $I$  es la información en bits,  $f$  y  $C$  la frecuencia y el contraste respectivamente, detallada en [JM97], determina la cantidad de información asignada por el sistema

para codificar cada área del dominio frecuencia-contraste.

El proceso experimental para determinar la expresión 5.2 consiste en medir la mínima variación de frecuencia o amplitud de contraste necesaria para discriminar dos patrones sinusoidales; con ello se concluye que cuanto más información es asignada por el SVH a una cierta área de IAF, mayor van a ser percibidos los artefactos en ella.

Este modelo verifica que la sensibilidad del SVH disminuye ante patrones de mayor contraste, lo cual permite cuantificar mucho más, y por tanto los coeficientes  $c_i$  de la transformada sobre la imagen con valores  $c_i = \Delta B_i$  correspondientes a una determinada región de frecuencia, en la que el valor medio de luminancia es aproximadamente  $B$ , siendo  $\Delta B_i/B$  relativamente alto, se puede asignar a todos ellos un valor de cuantificación único y constante  $L$ , aunque estos valores estén bastante alejados entre sí, ya que el SVH no va a percibir los errores de cuantificación  $|L - c_i|$  [Wat87b].

## 5.4 Conclusión

El presente capítulo detalla por un lado el proceso de percepción visual humano y por otro los nexos existentes entre la descomposición realizada por el SVH y la transformada wavelet, en el caso particular de funciones de Gabor.

Con ello se valida, lo que cabía esperar tras la bibliografía, un mecanismo de descomposición localizada en espacio y frecuencia, que permite introducir con mayor facilidad los criterios visuales con los que evalúa un observador humano.

Diferentes modelos de la visión humana han sido descritos y finalmente se han recopilado diferentes técnicas para introducir criterios de enmascaramiento de ruido, muy útiles en el proceso de compresión y transmisión de vídeo.

Existen además de las técnicas vistas, otros tipos de enmascaramiento asociados con las texturas [B.98] y las no linealidades de la amplitud de los coeficientes en el SVH [Art95], pero para el análisis y diseño del sistema propuesto por decisión de compromiso no son utilizados.

Esta decisión se realiza en base a que la propuesta presentada en el presente trabajo se centra en el estudio de la transmisión sobre redes con servicios *best effort* (con posibilidad de pérdidas) en las cuales intervienen factores más importantes como la adaptabilidad de las fuentes de vídeo al grado de congestión de la red, simultáneamente con la minimización de la distorsión visual.

En definitiva, se han estudiado los elementos necesarios para poder determinar la relevancia de la información visual en el usuario final, cuando la información de vídeo vaya a ser transmitida y alterada por la red.





## **Parte III**

# **Análisis y diseño del sistema propuesto**



La presente parte se desglosa en una fase de análisis y otra de diseño para la realización del sistema propuesto, siguiendo los pasos y la notación de [Pre97]; por motivos de claridad, no se introduce la fase de implementación porque los códigos fuente de los módulos no aportan información al presente trabajo.

En la primera fase, capítulo 6, se fijan los requisitos del sistema propuesto tras haber analizado la parte de revisión de la situación actual, pasando a una descripción del modelo fundamental del sistema mediante los análisis de entidad-relación, flujo de datos y transiciones de estado.

En la segunda fase o de diseño, capítulo 7, se procede de forma similar, progresivamente, mediante las diferentes etapas de diseño: diseño de los datos, diseño arquitectónico, diseño de interfaz de los módulos y diseño procedimental.



# Capítulo 6

## Análisis

El presente capítulo introduce el entorno de trabajo para el sistema propuesto de transmisión de vídeo; con ello se plantean las funcionalidades y se analizan cada uno de los diferentes elementos que lo forman.

### 6.1 Requisitos del sistema propuesto

Se ha revisado en el capítulo 2 la evolución de las redes con servicios *best effort* y la introducción de nuevos protocolos para ofrecer servicios integrados dentro del modelo IS[She94]. En dicho capítulo se decide optar por los servicios *best effort* ofrecidos por las redes ATM con servicios ABR.

Estos servicios ofrecen continuamente los recursos de la red de forma equitativa, entre las diferentes conexiones, pero sin ofrecer ningún tipo de garantía para la calidad de servicio. Con ello, la aplicación ha de contemplar situaciones de congestión en la red y adaptarse a ellas[Hlu91].

En dichas situaciones de congestión, y para el caso de transmisión de vídeo, es fundamental mantener los criterios visuales en los sistemas de compresión, ya que el SVH es el usuario final [IK85]; los criterios visuales son en definitiva aquellos que permiten una calidad aceptable en la transmisión cuando las prestaciones ofrecidas por la red no están garantizadas.

En este tipo de escenarios, las técnicas basadas en la DCT frecuentemente utilizadas en los estándares, presentan ciertos inconvenientes que se pueden observar empíricamente en el anexo B y resumir en los siguientes puntos:

- *efecto de bloques*: el proceso de cuantificación a bloques independientes, sin ninguna condición de contorno que asegure una suavidad en las fronteras existentes, genera un efecto molesto de bloqueo en la señal reconstruida, figura 6.1. Este efecto de bloqueo es más o menos importante en función del paso de cuantificación, es decir cuanto menor sea la tasa de bits permitida la cuantificación será más fuerte, resaltando dicho efecto [RY90][RJ91]
- *efecto enventanado*: los propios bloques introducen una modulación al enventanar la señal, produciendo en el dominio espectral la aparición de armónicos que



Figura 6.1: *Efecto bloque: imagen procesada mediante DCT con bloques de 8x8 pixels*

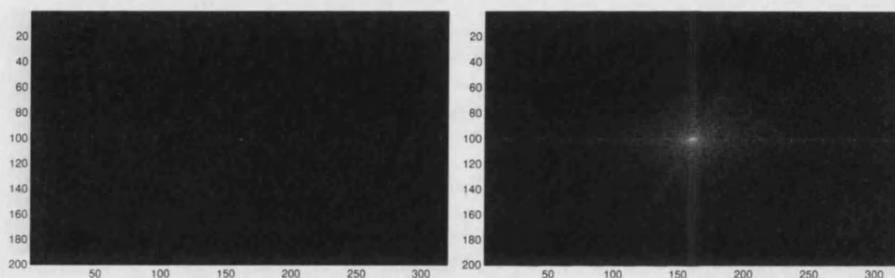


Figura 6.2: *FFT de Lisa y ampliación del efecto de enventanado.*

la distorsionan, figura 6.2, haciendo que la representación no sea tan compacta y haya más coeficientes pero con menor energía. La figura 6.3 enventana la imagen Lisa con una gaussiana (parte izquierda) y se observa un comportamiento en el dominio de Fourier mucho más natural y análogo al SVH (parte derecha) comparado con el de la figura 6.2

- *desacoplo con el SVH*: al no enventanar ni procesar como el SVH, la DCT no permite aplicar plenamente criterios de percepción visual en condiciones de *PSNR* (*Peak Signal to Noise Ratio*, señal de pico a ruido) inferiores a 20 db, donde las distorsiones pueden percibirse de forma diferente, de la misma manera que tampoco permite eliminar redundancia subjetiva
- *ineficiencia para descorrelar los pixels cercanos a la frontera entre bloques*: al separar los bloques de manera independiente sin solapamiento, los pixels de la frontera de bloques adyacentes mantienen redundancia que no les es extraída
- *enventanado espacial de tamaño fijo en la DCT*: en señales naturales reales, las componentes a altas frecuencias suelen tener una duración espacial pequeña, por ejemplo los bordes, mientras que las componentes de frecuencias bajas suelen tener una duración espacial mayor, por ejemplo el fondo, por tanto al aplicar la

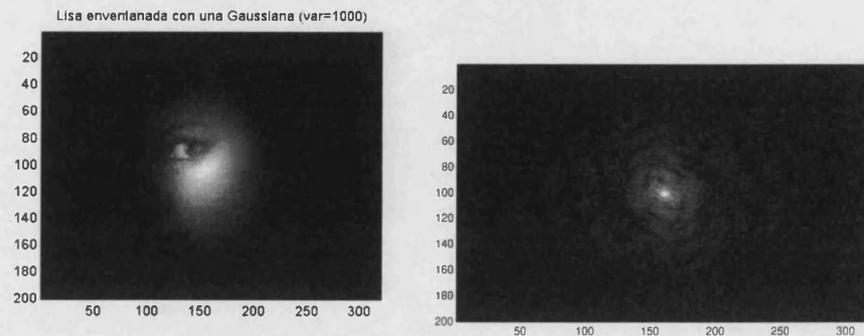


Figura 6.3: *Lisa enventanada con una gaussiana y ampliación logarítmica de su dominio de Fourier*

DCT sobre bloques de tamaño fijo todos los patrones de la señal son analizados con funciones base de la transformada que tienen el mismo soporte espacial y la representación obtenida no es tan compacta por falta de adaptabilidad en frecuencia

- *efecto Gibbs*: deriva de los inconvenientes anteriores cuando los coeficientes de altas frecuencias son cuantificados, debido a que las funciones base de la DCT no mantienen suficiente localidad espacial, y no son capaces de capturar características de alta frecuencia localizadas en los bordes, produciendo un efecto de poca definición; para poder representar dicho borde exactamente con una expansión de tipo DCT sería necesario mantener todos los coeficientes correspondientes a todas las frecuencias sin pérdidas
- *ruido mosquito*: procede directamente del efecto Gibbs cuando la imagen está en movimiento, produciendo un enturbiamiento y poca claridad
- *escalabilidad y multiresolución*: la DCT al no seguir una descomposición análoga al SVH, posee mecanismos bastantes rudimentarios para jerarquizar la información y permitir diferentes grados de resolución y escalabilidad

Trabajos que han tratado de resolver los inconvenientes de la DCT procesada por bloques han sido la transformada LOT o transformada con solapamiento [HSM89][Mal92], donde se procesa la señal de la misma manera que la DCT pero con un 50% de solapamiento. Los resultados obtenidos consiguen una compactación de la energía similar a la DCT, pues se trata de los mismas funciones base pero eliminando el efecto de bloques, aunque su utilización no ha sido muy extendida porque las mejoras frente a su complejidad no están justificadas.

Otros trabajos como [Vet89], [Bak89], [Far95] y [Che96a] han tratado de abordar el problema, pero desde una perspectiva separada tanto para la compresión como para la transmisión.

De las anteriores referencias cabe destacar [Vet89] y [Che96a] por su afinidad al presente trabajo. En la primera se estudian los métodos de codificación subbanda

con la finalidad de poder transmitir en una red de paquetes, donde se evalúa la de pérdida de subbandas (y su efecto sobre PSNR) y los anchos de banda requeridos por cada una de ellas, utilizando una descomposición 3D con 2 niveles de resolución espacial y 1 temporal. En el segundo, utilizando el mismo método de descomposición anterior, se estudia un sistema de transmisión de vídeo para telefonía móvil (con anchos de banda de 64 Kbps) con criterios visuales basado en JND (*Just Noticeable Distorsion*)[NJS93][Tek97].

En ambos casos no son abordadas las características intrínsecas a los servicios *best effort* de las redes y por tanto se concluye en la *necesidad* de buscar métodos alternativos que permitan ofrecer una calidad visual aceptable ante estos entornos.

En particular se plantean modificaciones en los sistemas de compresión y transmisión de vídeo, de manera que con una transmisión realimentada se pueda modificar el factor de compresión, respetando las pautas del sistema visual humano y tratando de adaptarse en todo momento al estado de la red para no congestionarla.

## 6.2 Análisis del modelo fundamental del sistema

Como se recoge en la sección anterior, el servicio de red con el que se ha decidido trabajar (servicio *best effort*) ha de caracterizar el sistema de transmisión de vídeo propuesto.

Para el planteamiento de los elementos del sistema, es necesario caracterizar previamente el entorno de red.

Los servicios ABR se caracterizan por ofrecer los recursos disponibles en la red de la forma más equitativa posible. Este reparto se realiza a través del control de flujo con celdas RM que realimentan extremo a extremo a la aplicación con la tasa explícita de salida (ver figura 2.7). Además, estos servicios permiten la negociación de una tasa mínima de salida previa a la conexión, pero sin ningún tipo de garantía.

La tasa de salida es continuamente ajustada a través de los diferentes conmutadores por donde pasa el circuito virtual, de manera que, mediante la aplicación de algoritmos para reparto equitativo de los recursos disponibles en cada conmutador, la tasa final obtenida y ofrecida a la fuente corresponde a la porción de ancho de banda equitativa en el enlace cuello de botella.

La indicación de la tasa permitida de salida permite llevar a un reparto de recursos óptimo en el menor tiempo posible, salvaguardando los requisitos de operación mínimos negociados previamente.

Una vez determinada la tasa explícita de salida, el siguiente paso es poder gestionar la *información* en la fuente, de la forma más acorde posible para ser transmitida y recibida. Por *información* queremos decir a la cantidad de conocimiento común que produce al observador la secuencia vista tanto “antes” como “después” de ser transmitida. Es decir, la secuencia “antes” de transmitirse es una secuencia original sin ningún tipo de procesado y que al transportarse por la red sufre alteraciones las cuales degradan siempre dicha secuencia, disminuyendo su *información* que visualizamos “después”[Sha48].

Para plantear el problema se describe el sistema global en dos sistemas, de emisión (sección 6.3) y de recepción (sección 6.4) que se ejecutan sincronizadas a través de los mecanismos de bajo nivel mencionados en la red.

El sistema de emisión ante el modelo de red o subsistema de interconexión procesa en primer lugar una descomposición natural en subbandas que reduce la redundancia y desglosa la secuencia de vídeo, sección 6.3.1, de forma más acorde al sistema visual humano (SVH) mediante prioridades, sección 6.3.2, a fin de poder controlar la *información*.

Tras ello el subsistema de estimación y asignación de bits toma una decisión de compromiso de forma adaptativa para poder ubicar la *información* entre una estimación del comportamiento de la red, sección 6.3.3, y los coeficientes de cada subbanda, sección 6.3.4, que directamente se discretizan por un proceso de cuantificación, sección 6.3.5.

En el último paso, se reduce la redundancia estadística de la fuente a través de un codificador de entropía, sección 6.3.6, y tras ello, cada subbanda independientemente se encapsula para poder ser transmitida, sección 6.3.7.

El sistema de recepción, procesa la información de forma similar pero a la inversa. En un primer paso extrayendo la información de los paquetes correctamente recibidos, sección 6.4.1, que son pasados al subsistema de decodificación de entropía, sección 6.4.2, para que tras él, una cuantificación inversa, sección 6.4.3, restaure el valor de los coeficientes con un ruido de cuantificación superpuesto y controlado, dando paso finalmente a un proceso de síntesis de las subbandas recibidas, sección 6.4.4, para ir reconstruyendo la secuencia original.

La calidad de la reconstrucción depende directamente de:

- el ancho de banda ofrecido
- la bondad del mecanismo de control de flujo y el estado de congestión de la red
- del tráfico que atraviesa el conmutador por el resto de las conexiones

## 6.3 Análisis del sistema de emisión

El sistema de emisión divide las funciones que realiza en diferentes subsistemas, independizando cada uno de ellos y describiendo los interfaces de unos con otros.

La arquitectura de emisión consta de dos módulos, ver figura 6.4. El primero de ellos, que ha sido revisado anteriormente y es descrito a continuación, llamado módulo de proceso, es objetivo de análisis y diseño como aportación del presente trabajo de investigación. El segundo módulo, en el que no se va a profundizar, lo forma la fuente ABR con el control de celdas RM, que implementa los mecanismos de control de flujo para gestionar instantáneamente la tasa permitida de salida accediendo a la UNI o interfaz de usuario.

Dicha arquitectura incorpora además, memorias FIFO para adaptación de los diferentes subsistemas.

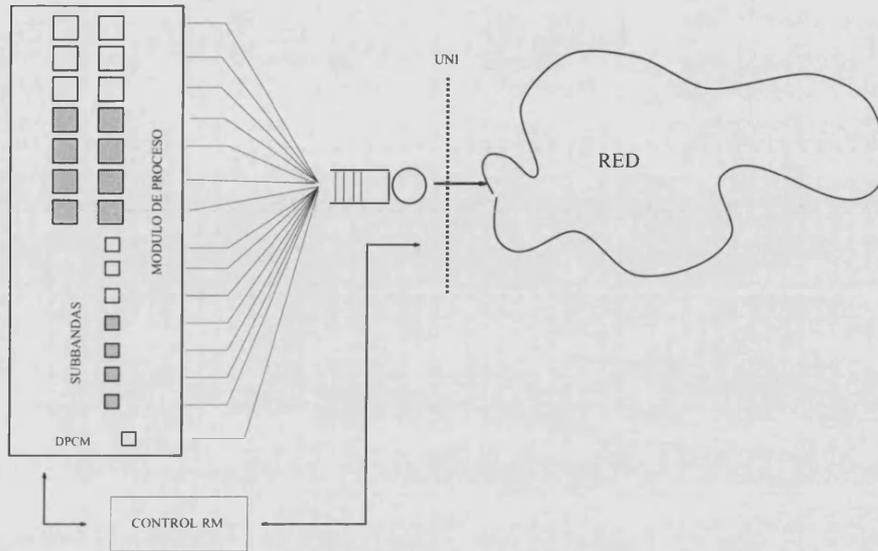


Figura 6.4: Sistema de emisión

### 6.3.1 Análisis del subsistema de descomposición

La secuencia de vídeo para poder ser adaptada a un modelo de red con servicio *best effort*, requiere de una *descomposición autónoma, autocontenida e independiente* donde el factor de compresión aplicado pueda ser variable manteniendo criterios visuales.

Para ello la opción de una descomposición subbanda y multiresolución tanto en los ejes temporal, vertical y horizontal se estima *a priori* potencialmente más acorde con los requisitos [Mal89b][Wat90], porque disminuye el grado de dependencia de unos paquetes con otros, aunque a costa de un menor grado de compresión al no introducir los métodos tradicionales de estimación de movimiento.

Por otro lado, justificando lo anterior, en compresión el interés reside en utilizar transformadas wavelet discreta crítica y con submuestreo separable, lo cual tiene como desventaja la incapacidad de distinguir frecuencias con orientación de 45 y 135 grados, que no es excesivamente crítico, ya que la mayor cantidad de información en imágenes y vídeos naturales se encuentra en direcciones horizontales y verticales [LeG88][VK95][GS96].

Esta elección se centra en la transformada wavelet biortogonal, que no conserva completamente la energía en la descomposición como se ve en el capítulo 4 pero que permite aproximar a una reconstrucción perfecta. Esta descomposición es implementada mediante bancos de filtros biortogonales por sus prestaciones ante la cuantificación, al poseer fase lineal.

Para seguir las pautas del SVH como se ve en el capítulo 5, los campos receptivos se implementan en descomposición con octavas utilizando un muestreo diádico, descomponiendo la señal en subbandas a potencias de dos. Por otro lado es importante tener un análisis multiresolución de la señal, ya que permite ver la redundancia existente en cada nivel de resolución, análogamente como se procesa en el cortex visual primario

del SVH[HW68][Wat87a].

Una decisión de compromiso entre grado de descorrelación, factor de compresión, complejidad del codificador y decodificador, cantidad de memoria requerida y retraso extremo a extremo en sistemas de tiempo real, fija a dos los niveles de resolución o iteraciones realizadas sobre el banco de filtros, ver figura 6.5.

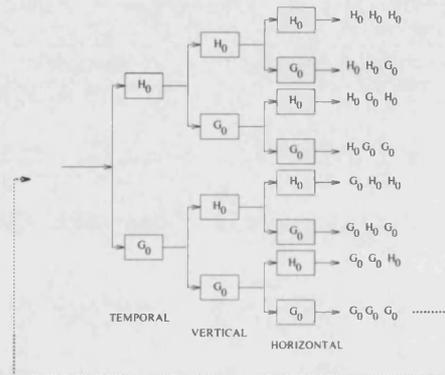


Figura 6.5: Banco de filtros para 3D en multiresolución

Como se ve en el capítulo 4, la descomposición se realiza mediante la convolución con cada filtro seguido de un diezmado por 2.

Considerando una secuencia de 25 fotogramas por segundo, en la primera iteración asociada al primer nivel de resolución, el primer filtro aplicado es el temporal y tras él, la secuencia de vídeo se descompone en bajas y altas frecuencias, en las figuras 6.4 y 6.6, bloques sin sombreadar y sombreados respectivamente, pero tomando de cada dos fotogramas sólo uno, es decir pasamos a dos secuencias, paso baja y alta temporal, de 25/2 fotogramas por segundo. Tras ello se aplica el filtro vertical que descompone cada fotograma filtrado temporalmente en dos fotogramas con resoluciones paso baja y alta vertical cada una de ellas con la mitad de altura del original. Finalmente, a cada uno de los fotogramas anteriores se aplica un filtrado horizontal, obteniendo una descomposición paso baja y alta horizontal con salida de fotogramas de anchura mitad.

Es decir, la primera iteración va tomando dos a dos fotogramas de la secuencia original y sacando, de cada dos, uno paso baja y uno paso alta temporal que a continuación son filtrados verticalmente, sacando de cada uno de ellos dos fotogramas más, y finalmente el filtrado horizontal que a su vez introduce de cada uno de los anteriores, dos más. En definitiva de cada dos fotogramas originales obtenemos  $2 \times 2 \times 2$  fotogramas que llamaremos 8 subbandas en el primer nivel de resolución, con un cuarto de tamaño de la original.

En la segunda iteración, para el segundo nivel de resolución, se procesa de forma similar, salvo que ahora la frecuencia temporal parte de 25/2 y pasa a 25/4 fotogramas por segundo, para lo que se requiere un bloque de memorias que permita retener la información en colas independientes (ver figura 6.4 del módulo de proceso).

La descomposición frecuencial en octavas del segundo nivel de resolución, parte del paso baja final de la primera iteración, es decir el resultado de aplicar los paso

baja en todos los filtros, tanto temporal, vertical y horizontal. Por tanto la necesidad en la entrada del filtro temporal de dos fotogramas requiere en el conjunto de la descomposición, el procesar 4 fotogramas de la secuencia original o lo que es lo mismo, tomar 2 subbandas paso baja de la primera iteración marcados con una X en la figura 6.6. A partir de ahí, el filtrado vertical y horizontal es realizado de la misma manera.

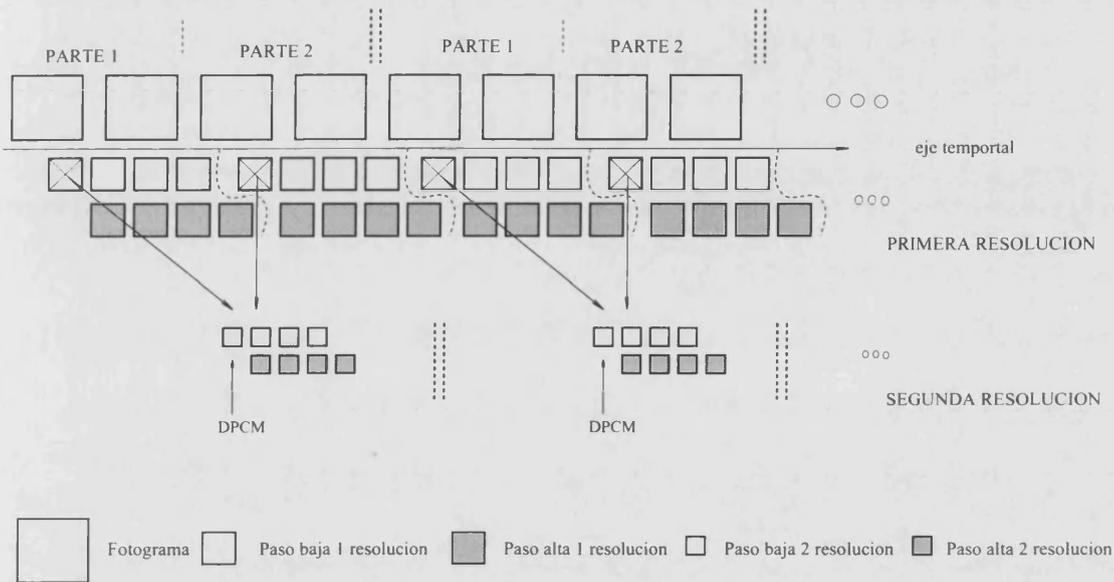


Figura 6.6: Descomposición de subbandas con 2 niveles de resolución

En resumen, las subbandas obtenidas con dos niveles de resolución para una secuencia de 25 fotogramas por segundo, ver figura 6.6, son un total de 22 detalladas a continuación:

1. cada 80 ms o dos fotogramas originales, 8 subbandas de la primera iteración primera parte
2. cada 80 ms más o los dos fotogramas originales siguientes, 8 subbandas de la primera iteración segunda parte
3. a partir de las dos anteriores o el primer nivel de resolución, es decir cada 160 ms y tomando las dos subbanda paso baja, se obtienen 8 subbandas asociadas a la segunda resolución

que para ser transmitidas, supone en el primer nivel de resolución, 7 subbandas cada 80 ms y 7 subbandas más en los siguientes 80 ms (nótese que se elimina una subbanda que pasa al segundo nivel de resolución) y en el segundo nivel de resolución, 8 subbandas. De esta forma, la transmisión de subbandas son 7 cada 80 ms y 7 + 8 cada 80 ms más, así de forma periódica.

Finalmente cabe destacar por su relevancia la subbanda paso baja del segundo nivel de resolución, que tiene un tratamiento especial debido a su alta correlación y

distribución uniforme por lo que se opta por una codificación diferencial mediante DPCM.

### 6.3.2 Análisis del subsistema de asignación de prioridades a las subbandas

El estudio de la localización espacio-frecuencia en el SVH y el desglose en diferentes texturas de las secuencias de vídeo, relaciona la cantidad de información de cada subbanda con la información perceptual obtenida de cada una de ellas.

De esta manera, la descomposición subbanda permite mediante métodos aproximados obtener respuestas neuronales que teniendo en cuenta el proceso realizado en el SVH dan más peso perceptual a unas subbandas que a otras.

Dichos pesos utilizados en el subsistema 6.3.4, permiten ponderar la información de cada subbanda, en este caso penalizar la distorsión de aquellas subbandas más visibles, donde la expresión de la distorsión está ligada además a los valores estadísticos de la varianza. Dichos pesos son estimados empíricamente por la cantidad de información visual asociada y relacionada con estudios fisiológicos realizados [Wat87b] [Jay92] [NJS93].

### 6.3.3 Análisis del subsistema de estimación del ancho disponible para la unidad de descomposición

La realimentación que se mantiene con el módulo de bajo nivel de la fuente ABR, permite progresivamente seguir el estado de la red a través de la tasa explícita de salida permitida (ACR, *allowed cell rate*) [Jai94].

Sin embargo, si se consideran variables aleatorias tanto la tasa de salida (dependiente del estado de la red, su configuración y el tráfico soportado) como la secuencia de vídeo, los órdenes de magnitud temporal de las variaciones de ambas son muy diferentes. La primera viene determinada por el ancho de banda de los enlaces, alrededor de Mbps con órdenes de magnitud temporal en las variaciones de us, mientras la segunda está determinada, bien a 25 o 30 fotogramas por segundo, es decir del orden de ms.

Por tanto, las decisiones a tomar en la descomposición y asignación de bits vienen determinadas según el estado de la red. Como la red se caracteriza por la afluencia de tráficos muy diversos que se multiplexan dentro del conmutador, para realizar una predicción en otro orden de magnitud temporal, es necesario estimar los estadísticos de dicho tráfico multiplexado, ya que directamente conforman los controles de flujo ABR.

Estudios anteriores [Ale95], han llegado a la conclusión que el tráfico multiplexado mantiene auto-similitudes con un comportamiento caótico. A partir de dicho comportamiento por autosimilitud, obtenemos una estimación del ancho de banda disponible o la tasa de bits para la unidad de descomposición (160 ms) a fin de adaptar la aplicación al comportamiento de la red.

### 6.3.4 Análisis del subsistema de asignación de bits

Tras el subsistema de estimación de bits, las pautas del comportamiento de la red han de ser la restricción del sistema de emisión; un servicio integrado sobre una red con servicio *best effort*, requiere de ambas partes buena voluntad, es decir buenas intenciones tanto en la compresión como en la transmisión.

La función por tanto de este subsistema es tratar, mediante un algoritmo de minimizar, de reducir al máximo la distorsión perceptual de la secuencia bajo las restricciones del comportamiento de la red, buscando tras la descomposición el paso de cuantificación que determina los bits por coeficiente y el número de bits totales para cada subbanda; hay que destacar que dicha asignación de bits ha de tener en cuenta el subsistema de codificación de entropía, 6.3.6.

La restricción dada en la minimización puede hacer que ciertas subbandas puedan o no ser transmitidas, este último caso es determinado cuando la asignación de bits es nula o negativa. Para aquellas que puedan transmitirse, se calcula el paso de cuantificación, para discretizar los coeficientes de cada subbanda, de forma que minimice la distorsión total.

### 6.3.5 Análisis del subsistema de cuantificación

El proceso de cuantificación es un proceso no lineal e irreversible que conlleva pérdida de información y que por tanto debe hacerse cuidadosamente para que la distorsión introducida en la señal sea lo menos perceptible posible; éste es el motivo por el cual la etapa de descomposición de la señal debe de estar relacionada de alguna forma con el SVH para que la cuantificación pueda estar basada en criterios de percepción visual.

Determinado el paso de cuantificación en la subsección 6.3.4, la división de cada coeficiente por dicho paso y el redondeo al entero más próximo, es lo que implementa el subsistema de cuantificación. Este proceso simplifica una cuantificación uniforme, como se ve en el capítulo 3.

El cuantificador es el único elemento dentro de los sistemas de emisión y recepción que introduce pérdidas de información o ruido.

### 6.3.6 Análisis del subsistema codificador de entropía

Dados los valores discretos a la salida del cuantificador, la estadística de dichos valores permite reducir la redundancia mediante dos técnicas sucesivas, llamadas “*Run-Level*” y “*Huffman*” que consisten respectivamente en:

1. simplificar las tiras de valores nulos asociando la *longitud* de valores nulos y codificando el siguiente *valor* no nulo
2. asociando a cada pareja de (*longitud, valor*) un código de menor longitud para los símbolos más probables y mayor para los menos probables según la tabla de códigos de longitud variable del anexo F.

de esta manera la información a la salida de la fuente posee la mínima redundancia posible.

### 6.3.7 Análisis del subsistema de adaptación a la red

Finalmente, el emisor conecta a la red a través del subsistema de adaptación, ver figura 6.4, que implementa dos funciones diferentes:

1. adaptar mediante una memoria FIFO (ver en figura 6.4 la memoria antes de la UNI) las velocidades de entrada y salida, una fijada por la secuencia de vídeo y la otra fijada por la tasa permitida de salida en la red
2. adaptar al formato de paquete de la red ATM, con celdas de 48 bytes de carga útil a través de la capa de adaptación para cada una de las subbandas a transmitir

Destacar que las subbandas son transportadas independientemente y ordenadas de mayor a menor prioridad. La información contenida en cada subbanda a enviar es autocontenida, es decir, la información necesaria (*"side-information"*) para poder reconstruirla. Esta información es detallada en el diseño del subsistema de adaptación, en el capítulo siguiente.

## 6.4 Análisis del sistema de recepción

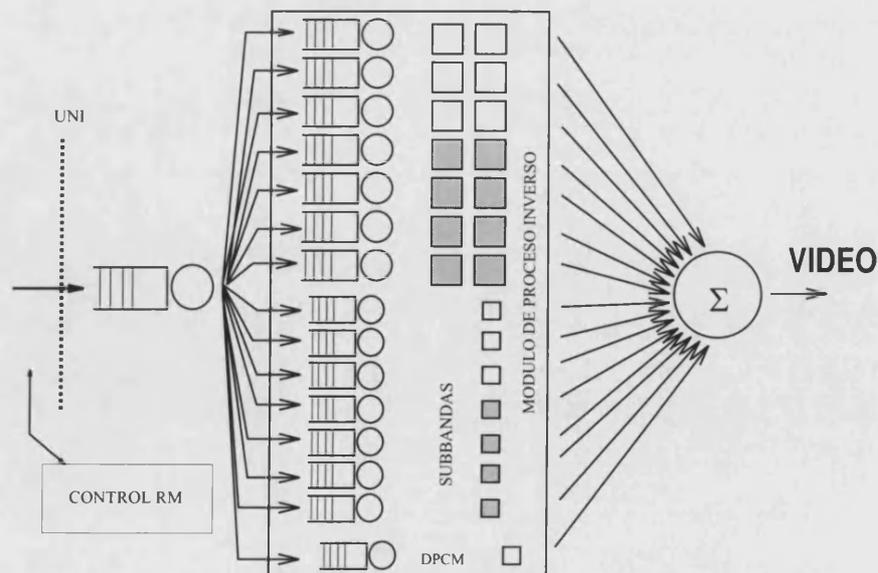


Figura 6.7: Sistema de recepción

Tras el transporte de la información por la red y sufrir pérdidas y retardos hasta cierto punto controlados por el control de flujo, el proceso de recepción se divide de forma

similar como se ha visto en la emisión, en funciones independientes realizadas en cada subsistema, figura 6.7.

De la misma manera que en el sistema de emisión, la arquitectura de recepción también consta de dos módulos, el primero de más bajo nivel, implementa el receptor ABR para la gestión del control de flujo con la red y el segundo descrito con más detalle a continuación, cada uno de los procesos intrínsecos a la aplicación, resultado del trabajo de investigación.

### 6.4.1 Análisis del subsistema de adaptación a la red

La adaptación a la red de forma similar que en 6.3.7 consta de dos funciones:

1. reducción de la variabilidad de las llegadas o “*jitter*” mediante un memoria FIFO sincronizada por la tasa de fotogramas por segundo para la reconstrucción del vídeo
2. extracción de la carga transportada por cada paquete de red a través de la capa de adaptación en la red ATM, comprobando su integridad para ser pasada la información al subsistema siguiente

### 6.4.2 Análisis del subsistema de decodificación de entropía

Extraídas las subbandas correctamente recibidas, es necesario aplicar el proceso inverso de la codificación de entropía mediante los siguientes pasos:

1. aplicar el paso inverso de la codificación por “*Huffman*” con una búsqueda en la tabla de códigos para la extracción de la pareja (*longitud,valor*)
2. deshacer la codificación de longitud variable “*Run-Level*”, introduciendo tantos coeficientes nulos como *longitud* e insertar a continuación el coeficiente *valor*

### 6.4.3 Análisis del subsistema de cuantificación inversa

Obtenidas las subbandas con los coeficientes cuantificados, el subsistema de cuantificación inversa recupera aproximadamente (no igual por el ruido de cuantificación) el coeficiente original multiplicando el valor discreto por el paso de cuantificación.

Este paso de cuantificación, así como la referencia temporal (registrada con el número de subbanda y la unidad de descomposición), están dentro de cada subbanda transportada.

### 6.4.4 Análisis del subsistema de reconstrucción

Finalmente tras todos estos subsistemas, las subbandas recuperadas son procesadas inversamente mediante el banco de filtros, interpolando y convolucionando con los filtros biortogonales de síntesis.

La salida de cada filtro forma parte de información descorrelada de la secuencia original asociada a una determinada textura, que sumando sus coeficientes reconstruye con el mayor *grado de similitud perceptual* posible la secuencia original generando la máxima *información permitida*.

Cabe destacar que dicho grado de similitud perceptual es la restricción asumida y tomada en cuenta durante todo el proceso tanto de compresión como de transmisión.

## 6.5 Conclusión

Un análisis tanto de los requisitos como de los elementos que componen el sistema propuesto ha sido desarrollado para poder pasar en el capítulo siguiente a la fase de diseño.

Se han tratado de resaltar aquellos puntos que tanto para el funcionamiento como la implementación del sistema sean relevantes, de manera descriptiva, a fin de poder encontrar posibles relaciones y vinculaciones entre ellos en su diseño.



# Capítulo 7

## Diseño

En el presente capítulo se pretende diseñar el sistema tras haber realizado el análisis en el capítulo anterior.

Para ello se especifica la arquitectura, los interfaces de los elementos que componen el sistema y los algoritmos o procedimientos de forma que permitan una directa implementación, pero sin pretender entrar en ella. Cada uno de los diferentes apartados de éste capítulo aborda su diseño con los razonamientos, estudios analíticos y referencias bibliográficas necesarias.

### 7.1 Pautas de mínima negociación

Previo al estudio del sistema tanto de emisión como de recepción, es necesario establecer las pautas de mínima negociación en la red, a partir de las cuales vaya a funcionar el sistema.

Los servicios ABR de las redes ATM se caracterizan por dos mecanismos básicos: controles de flujo cerrado basados en tasa explícita y negociación de un ancho de banda mínimo mediante un circuito virtual.

La negociación del circuito virtual se realiza previamente a la transmisión, facilitando en dicho momento al receptor el formato de imagen en emisión y la tasa de fotogramas por segundo de salida. Los formatos bajo estudio en las simulaciones realizadas son detallados en la tabla 7.1, siendo estos formatos comunes en aplicaciones de videoconferencia.

<i>Formato</i>	<i>Ancho x Alto</i>
QSIF	176 x 144
SIF	352 x 288

Tabla 7.1: *Formatos de imagen en vídeo*

En este caso el servicio ABR no permite negociar tantos parámetros de la fuente como en los servicios CBR y VBR, teniendo sólo la posibilidad de negociar la tasa mínima de salida o MCR y la de pico o PCR. Esta negociación ha de fijar el umbral

mínimo de funcionamiento de la aplicación para evitar el efecto acantilado (“*Cliff effect*”), es decir, cuando la degradación es tal que no permite recuperar la señal [Tek97].

El ancho de banda mínimo depende directamente de los métodos utilizados en la descomposición del sistema propuesto, una descomposición subbanda en octavas basada en transformadas wavelet con dos niveles de resolución. Bajo dichas condiciones, el criterio mínimo de negociación (para evitar el efecto acantilado) fija la posibilidad de transmisión de al menos una subbanda, que por criterios visuales está determinada por la subbanda paso baja del segundo nivel de resolución codificada en DPCM, que es equivalente utilizar un ancho de banda para la transmisión cada 160 ms de un cuarto de formato de la secuencia original.

Teniendo en cuenta que se trata de una negociación mínima, negociación de la cual se puede tarificar, es necesario hacer una estimación realista del ancho de banda que va a ser necesario.

Nótese que dicha negociación ha de ser independiente de la evolución de la red, de manera que la tasa permitida de salida en emisión fijada por el control de flujo ABR, ha de ser mayor que la mínima negociada.

Aunque esta negociación queda abierta en el sistema propuesto y como los bits por coeficiente para la subbanda DPCM depende tanto de la calidad con que el usuario quiera ver una determinada secuencia (pagando por ello) como de las características de la conexión, una decisión de compromiso puede fijar como valor por defecto el transmitir sólo dicha subbanda con *8 bits por pixel*, que se traduce directamente en un factor de compresión 64 a 1. Resultados de dicho factor de compresión se recogen en la sección 9.3 del capítulo 9.

Otra cota podría fijarse en transmitir sólo dicha subbanda con un bit por coeficiente para la posibilidad de codificar al menos el signo de cada uno; obviamente en este caso la calidad de la secuencia es tremendamente mala.

Por tanto el parámetro MCR (*Minimum Cell Rate*, tasa mínima de salida), tomando por defecto 25 imágenes por segundo y 8 *bpp* para un cuarto de formato de la secuencia original, supone un valor de  $\frac{\text{Ancho Alto}}{4} \cdot 8 \cdot \frac{25}{4}$  bits por segundo, siendo  $(\text{Ancho Alto})/4$  el tamaño de la subbanda que se codifica con 8 *bpp* y es transmitida a una velocidad de  $\frac{25}{4}$  subbandas por segundo. Los bits por segundo obtenidos se traducen a celdas redondeando superiormente a múltiplos de 48 bytes útiles/celda \* 8 bits/byte=384 bits útiles/celda e introduciendo la información de cabecera de 32 bits, discutida en la sección 7.2.7, es decir

$$\left[ \left( \frac{\text{Ancho Alto}}{4} \cdot 8 \cdot \frac{25}{4} + 32 \right) \frac{1}{384} \right] \text{ celdas/s} \quad (7.1)$$

## 7.2 Diseño del sistema de emisión

Como se ve en la figura 7.1, el sistema de emisión está formado por bloques independientes trabajando en un mismo proceso, codificación subbanda con prioridades por banco de filtros biortogonales, cuantificación (Q) y codificación de entropía (E).

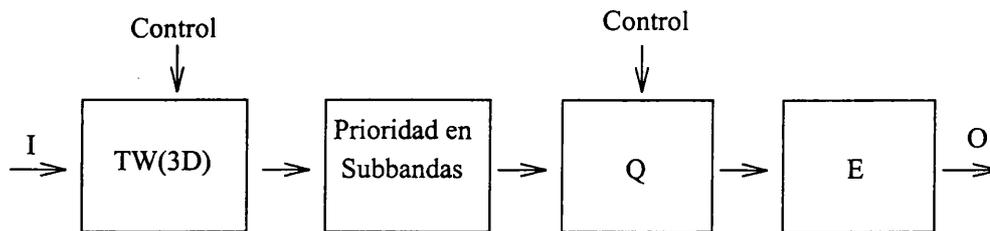


Figura 7.1: Bloques del sistema de emisión

Como el objetivo del presente trabajo es abordar el análisis y diseño de un sistema completo de compresión y transmisión de vídeo sobre redes con servicios *best effort*, concretamente redes ATM con servicio ABR donde existe posibilidad de pérdidas, es por ello que se toman varias decisiones de compromiso, entre ellas no pretender integrar de forma exhaustiva los criterios visuales como pieza clave en la medida de calidad visual de las secuencias, sino en la medida de lo posible.

### 7.2.1 Diseño del subsistema de descomposición

Se ve en el capítulo 3 y anexo B, que la estimación de movimiento degrada la transmisión cuando aparecen errores. La dependencia de datos que introduce la descomposición basada en la DCT, DPCM y los vectores de movimiento está excesivamente ligada y jerarquizada, de forma que frente a errores esporádicos y aleatorios puede destruir una secuencia de vídeo y provocar la pérdida del sincronismo.

La propuesta planteada ante dicha situación es descomponer la secuencia en diferentes fuentes de información independientes, asociadas cada una de ellas a las subbandas tras la descomposición mediante banco de filtros separables. La implementación de filtros separables facilita computacionalmente el proceso de descomposición.

Esta descomposición trata de mantener las pautas de funcionamiento del SVH. A grandes rasgos se pretende buscar comportamientos similares al SVH (visto en el capítulo 5) y que se pueden resumir en:

- la estructura de células y neuronas con conexiones directas e indirectas responden a transformadas localizadas en el espacio y frecuencia, con resoluciones por octavas
- los campos receptores de las células del cortex visual permiten realizar un análisis de texturas en diferentes orientaciones
- las funciones base que realizan la transformación en el SVH son autosimilares y escalables
- en el SVH existen mecanismos de enmascaramiento de texturas que son utilizados en los compresores de segunda generación para eliminar la redundancia subjetiva
- el SVH posee mayor resolución en las bajas frecuencias, en los detalles de resolución menor

Tras la descomposición no es posible una descorrelación total de las fuentes por la imposibilidad de utilizar filtros ortogonales, ya que generan cuando se introducen mecanismos de cuantificación discontinuidades en la fase (como se ve en el capítulo 4).

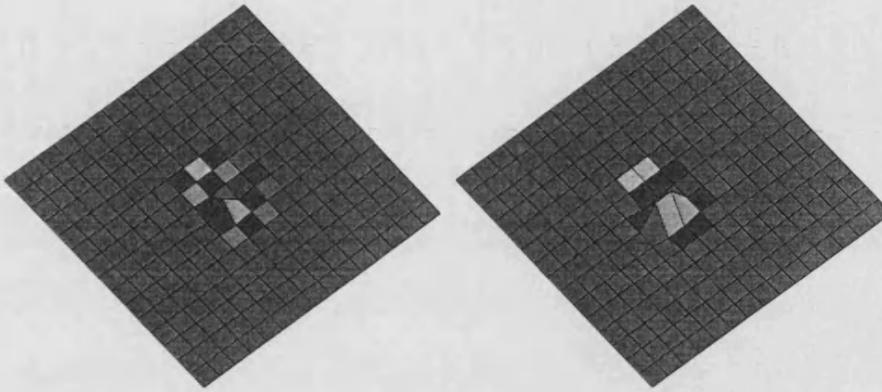


Figura 7.2: Respuesta impulsional 2D biortogonal: paso baja y detalle horizontal

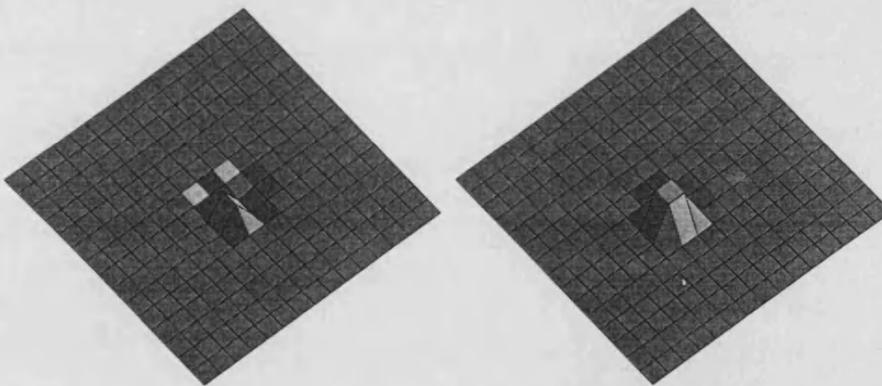


Figura 7.3: Respuesta impulsional 2D biortogonal: detalle vertical y diagonal

La utilización de filtros biortogonales genera filtros en 2D con respuestas impulsionales para detección de las texturas en diferentes orientaciones, figuras 7.2 y 7.3, y que por su analogía en el proceso de descomposición con el SVH (de forma más aproximada que la DCT) permite descomponer la imagen como suma de texturas con distintas prioridades.

La figura izquierda de 7.2 es la respuesta al impulso del filtro biortogonal paso bajo tanto horizontal como vertical, también llamado “residuo” o “aproximación” y la figura derecha la respuesta al impulso del filtro biortogonal paso bajo vertical y paso alto horizontal o “detalle horizontal”. La figura izquierda de 7.3 es la respuesta al impulso del filtro biortogonal paso bajo horizontal y paso alto vertical, también llamado “detalle vertical” y la figura derecha, la respuesta al impulso del filtro biortogonal paso alta tanto vertical como horizontal o “detalle diagonal”.

El valor de los coeficientes para cada una de las respuestas impulsionales de aproximación, detalle horizontal, vertical y diagonal se muestran en las tablas 7.2, 7.3, 7.4 y 7.5 respectivamente.

0.0625	-0.3750	0.0625
-0.3750	2.25	-0.3750
0.0625	-0.3750	0.0625

Tabla 7.2: *Coefficientes de la respuesta impulsional del filtro 2D de aproximación*

-0.0625	- 0.0625
0.3750	0.3750
-0.0625	-0.0625

Tabla 7.3: *Coefficientes de la respuesta impulsional del filtro 2D de detalle horizontal*

-0.0625	0.3750	-0.0625
-0.0625	0.3750	-0.0625

Tabla 7.4: *Coefficientes de la respuesta impulsional del filtro 2D de detalle vertical*

En el eje temporal, debido a los requisitos en tiempo real de las aplicaciones, se opta por un filtro corto, en este caso ortogonal y de dos coeficientes, conocido como filtro Haar[Che96b]. De esta manera el número de imágenes a retener para la descomposición temporal se reduce, con dos niveles de resolución, a cuatro fotogramas o 160 ms de secuencia para una tasa de 25 fotogramas de segundo. Un nivel mayor de resolución obligaría a la introducción de 8 fotogramas lo que supone un retraso de 320 ms, que por un lado quita prestaciones para tiempo real y por otro requiere aumentar las necesidades de memoria del sistema.

Un estudio más detallado, tanto en frecuencia como en espacio, de los filtros utilizados, se encuentra en el anexo C. A la vista de dichos resultados, se opta para la descomposición con los filtros de respuesta más regular, en este caso los filtros llamados biortogonal(2); dicha elección se basa en que los filtros más regulares y especialmente en el proceso de síntesis, producen menos distorsión visible en las imágenes como cabe esperar[GS96].

La figura 7.4 muestra el proceso de emisión de las diferentes subbandas tras la aplicación de 2 niveles de resolución, como se detalla en el capítulo 6.

Se analiza a continuación la evolución del procesado y filtrados a partir de 4 fotogramas de la secuencia original de Miss América, figura 7.5.

Cabe destacar que en esta sección la salida de los filtros han sido ecualizadas para poder ser visualizadas; en el anexo E se encuentran los histogramas de cada una de ellas para estimar el comportamiento estadístico de los coeficientes dentro de cada subbanda.

0.0625	0.0625
0.0625	0.0625

Tabla 7.5: Coeficientes de la respuesta impulsional del filtro 2D de detalle diagonal

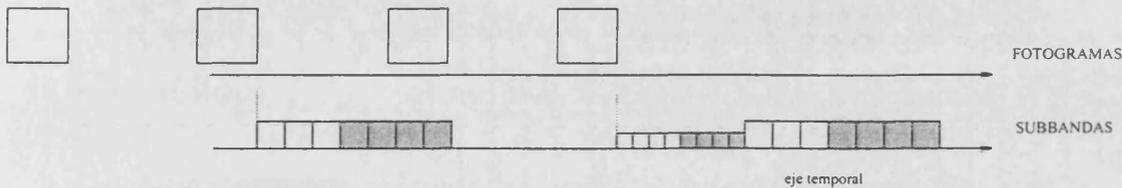


Figura 7.4: Ordenación en el tiempo según la obtención de las diferentes subbandas generadas en la descomposición 3D

Cada dos fotogramas se descompone en dos salidas asociadas al filtro temporal paso bajo y alto, por tanto para 4 fotogramas se obtiene, de izquierda a derecha en la figura 7.6, 2 paso baja temporal de los fotogramas 0-1 y 2-3 (en ellos se observa la escena con componentes de movimiento asociadas a frecuencias digitales menores de 0.25), y 2 paso alta temporal de los mismos fotogramas (en ellos se observa la escena con componentes de movimiento asociadas a frecuencias digitales entre 0.25 y 0.5).

Las salidas del filtro temporal asociadas a los fotogramas 0-1 se filtran en 2D cada una de ellas, generando las salidas de las figuras 7.7 y 7.8 respectivamente (asociadas a la primera parte de la primera resolución), con tamaño una cuarta parte del formato original de la secuencia, por los diezmados en los ejes horizontal y vertical.

Para ambas figuras 7.7 y 7.8, las diferentes subfiguras corresponden de izquierda a derecha, según la descomposición 2D vista, en residuo, detalle horizontal, vertical y diagonal. Destacar en la subfigura del detalle diagonal en 7.8, los patrones que aparecen debido al entrelazado de la cámara.

De la misma forma, con los 2 fotogramas siguientes 2-3, las salidas asociadas son respectivamente las de las figuras 7.9 y 7.10 (asociadas a la segunda parte de la segunda resolución).

Las salidas anteriores están asociadas en la figura 7.4 a los 7 cuadrados más grandes, 3 blancos y 4 sombreados de igual tamaño, donde los blancos proceden de la descomposición paso baja temporal y los sombreados a la paso alta. De los paso baja



Figura 7.5: 4 fotogramas de Miss América: unidad de descomposición para aplicar 2 niveles de resolución



Figura 7.6: Salidas de los filtros temporales: paso baja parte 1 y 2, paso alta parte 1 y 2

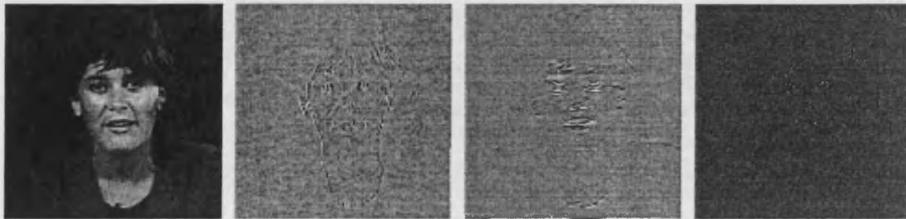


Figura 7.7: Descomposición 2D parte 1, del paso bajo temporal 1ª resolución: residuo, detalle horizontal, vertical y diagonal

temporales se ha visto que originalmente también había 4, sin embargo dicha salida pasa a descomponerse en un segundo nivel de resolución y por tanto no se transmite.

Aplicando una nueva iteración al banco de filtros 3D (figura 6.5) para la segunda resolución, utilizando como entrada las salidas paso baja final de ambas partes de la primera resolución, se obtiene de la misma manera la descomposición de las figuras 7.11 y 7.12 (dichas figuras también siguen la ordenación de residuo, detalle horizontal, vertical y diagonal).

Si se observa la subbanda paso baja de la segunda resolución (residuo del paso baja temporal), subbanda más a la izquierda en la figura 7.11, destaca su gran parecido con los 4 fotogramas originales manteniendo las componentes paso baja tanto temporal, horizontal y vertical de todos ellos. Dicha subbanda se caracteriza por la uniformidad (figura 7.14) y correlación con sus homólogas, por lo que se decide codificarla diferencialmente mediante el algoritmo DPCM, capítulo 3 sección 3.2.2, realizando la predicción a partir tanto de la subbanda de la descomposición anterior

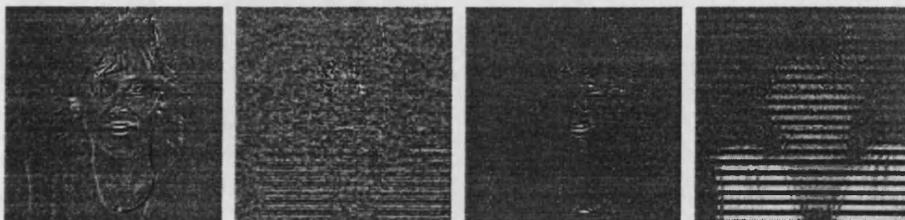


Figura 7.8: Descomposición 2D parte 1, del paso alta temporal 1ª resolución: residuo, detalle horizontal, vertical y diagonal

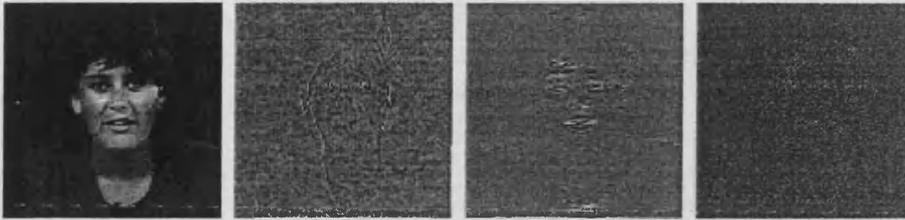


Figura 7.9: *Descomposición 2D parte 2, del paso bajo temporal 1ª resolución: residuo, detalle horizontal, vertical y diagonal*

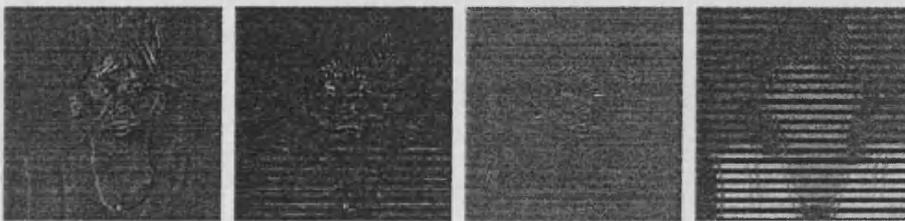


Figura 7.10: *Descomposición 2D parte 2, del paso alta temporal 1ª resolución: residuo, detalle horizontal, vertical y diagonal*

como internamente con los coeficientes vecinos.

El filtro de predicción en la figura 3.4 escogido es

$$c(x, y, t) = \frac{1.5(c(x-1, y, t) + c(x-1, y-1, t) + c(x, y-1, t))}{7} + \frac{0.5(c(x-1, y, t-1) + c(x-1, y-1, t-1) + c(x, y-1, t-1)) + c(x, y, t-1)}{7} \quad (7.2)$$

donde  $c(x, y, t)$  denota el coeficiente de la subbanda a codificar en la posición espacial y temporal  $(x, y, t)$ .

La parte izquierda de la figura 7.13 se resta a la subbanda para obtener la parte derecha de la misma figura, que finalmente será transmitida.

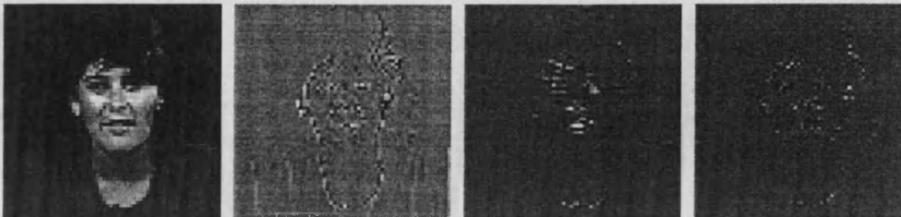


Figura 7.11: *Descomposición 2D del paso bajo temporal 2ª resolución: residuo, detalle horizontal, vertical y diagonal*

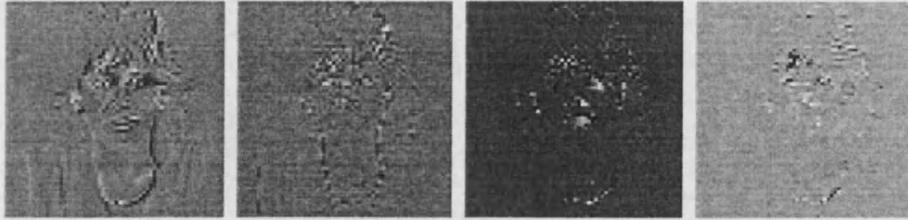


Figura 7.12: *Descomposición 2D del paso alta temporal 2ª resolución: residuo, detalle horizontal, vertical y diagonal*



Figura 7.13: *Predicción y diferencia de la subbanda paso baja final del segundo nivel de resolución*

## 7.2.2 Diseño del subsistema de asignación prioridades a las subbandas

Con una simple inspección ocular de la descomposición anterior y de los resultados aportados en el capítulo 9 ante los efectos por la pérdida de subbandas, puede observarse una posible ordenación de las subbandas según prioridad.

Dicha prioridad depende también, como se puede ver en el anexo E, de la varianza de los histogramas de los coeficientes de cada subbanda, que corresponden con una función densidad de probabilidad de una gaussiana generalizada[SLG95].

Las respuestas del SVH han sido introducidas en la bibliografía como diferentes pesos a cada uno de los coeficientes de la descomposición[Nqu91][BM93]. En este procedimiento han sido utilizados pesos obtenidos de la función de sensibilidad al

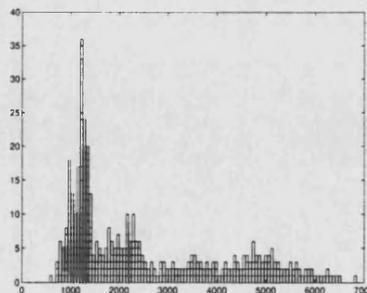


Figura 7.14: *Histograma de la predicción realizada para la subbanda paso baja final del segundo nivel de resolución*

contraste en función del contraste(C) y de la frecuencia(f) CSF(f,C) para determinar la asignación de bits en el dominio de los coeficientes de la DCT.

Otra función como la expresión 5.2 de IAF[Art95] vista en el capítulo 5, permite determinar el número de bits asignados para codificar cada coeficiente en el dominio de Fourier.

Integrando la expresión 5.2 para las diferentes amplitudes de contraste, entre  $c_m$  y  $c_M$ , la expresión obtenida CIAF *Cumulative Information Allocation Information*[Art95]

$$CAIF(f) = \frac{dI}{df} = \int_{c_m}^{c_M} IAF(f, C) dC \quad (7.3)$$

determina los pasos de cuantificación asignados a cada coeficiente.

Sin embargo como se ha venido diciendo, la descomposición realizada en el SVH se adapta mejor a una descomposición wavelet en resoluciones por octavas[Wat87b]. Por otro lado, las caracterizaciones del SVH realizadas para pesar los diferentes coeficientes [BM93] corresponden a descomposiciones con patrones sinusoidales.

Para poder modelizar dicho comportamiento en otro dominio se requiere un estudio adicional, tratando de aplicar el mismo modelo visual estudiado y experimentado mediante funciones CSF e IAF en el dominio donde se realice la descomposición. Un estudio tratando de encontrar estas funciones peso en el dominio de Gabor localizado en espacio y frecuencia, se detalla en [Art97], discutiendo la función GSF *Gabor stimuli Sensitivity Funcion*, función que recoge un proceso de degradación tanto óptico como neuronal.

El estudio [Art97] es llevado a cabo analíticamente igualando tanto un proceso con pesos basados en el dominio de Fourier como un proceso en el dominio de Gabor. Si  $i'(x)$  es la imagen final reconstruida por ambos mecanismos partiendo de una imagen inicial  $i(x)$ , se exige la igualdad en

$$i'_{CSF}(x) = FT^{-1} [CSF(k) FT [i(x)](k)](x) \quad (7.4)$$

$$i'_{GSF}(x) = GT^{-1} [CSF(x, k) GT [i(x)](x, k)](x) \quad (7.5)$$

concluyendo que

$$GSF(x_0, k_0) = GT \left[ FT^{-1} [CSF(k) FT [g(x, x_0, k_0)](k)](x) \right] (x_0, k_0) \quad (7.6)$$

donde  $x, k$  son variables en el dominio espacial y frecuencial respectivamente y  $x_0, k_0$  son puntos de localización en los dominios de Gabor con funciones base del tipo 5.1. La curva GSF tiene un comportamiento exponencial alcanzando su máximo alrededor de las frecuencias nulas. La curva GSF permite observar que existe una atenuación constante para todos los coeficientes de una subbanda de frecuencia para cada posición espacial.

En el caso de una descomposición en octavas con transformadas wavelets, se obtiene una expresión parecida, aunque de mayor pendiente, debido a la escala logarítmica en el eje de frecuencias. También de forma aproximada se puede estimar a partir de la curva CSF pero integrando a la función en el dominio de Fourier a intervalos según la descomposición wavelet.

Esta discusión nos permite estimar empíricamente un comportamiento relativo entre los coeficientes de cada subbanda de forma exponencial[Art97]. Los pesos perceptuales seleccionados para ponderar el efecto visual en la distorsión matemática introducida por cada subbanda que ponderan las varianzas en la sección 7.2.4 (ecuación 7.21) se expresan en potencias de dos para la descomposición realizada y dichos pesos son para:

1. segundo nivel de resolución, paso baja temporal, 128 para DPCM, 64 para los detalles verticales y horizontales y 32 para el detalle diagonal
2. segundo nivel de resolución, paso alta temporal, 32 para residuo, 16 para los detalles verticales y horizontales y 8 para el detalle diagonal
3. primer nivel de resolución, paso baja temporal, 4 para los detalles verticales y horizontales y 2 para el detalle diagonal
4. primer nivel de resolución, paso alta temporal, 4 para residuo, 2 para los detalles verticales y horizontales y 1 para el detalle diagonal

Estos pesos perceptuales serán utilizados en la sección 7.2.4 para la asignación de bits.

### 7.2.3 Diseño del subsistema de estimación del ancho disponible para la unidad de descomposición

Es necesario hacer una estimación presente para poder hacer una transmisión futura, ya que el procesado realizado en la descomposición vista en 7.2.1 requiere por sus dos niveles de resolución, estimaciones temporales de 160 ms siendo las transiciones producidas dentro de la red de orden muy inferior. Para ello se ha de partir de un análisis que permita estimar dicho comportamiento, lo que exige un conocimiento de las estadísticas que gobiernan dichas transiciones.

Una suposición frecuente en modelado de computadores y redes es modelar la tasa de entrada o de salida a un fuente de Poisson. Estas fuentes cuando multiplexan sus tráficos en los conmutadores, teóricamente generan un nuevo proceso de Poisson. Sin embargo empíricamente se demuestra que la suposición de fuentes de Poisson no explica el comportamiento real del tráfico ya que cuando el intervalo de observación de una fuente de Poisson crece, la distribución es mucho más suave y la desviación estándar disminuye, cosa que no ocurre en la realidad.

Se ha demostrado en [WW94][WL94] que el comportamiento del tráfico multiplexado no se suaviza y es invariante con respecto al intervalo de observación, es decir, concluyen que dicho comportamiento queda asociado a un proceso fractal, un proceso de autosimilitud de forma que los comportamientos se reproducen a diferentes escalas de tiempo.

Pretender dar un explicación a dicho comportamiento no es simple. La misma complejidad asociada a un desarrollo analítico que demuestre una función de distribución que caracterice el comportamiento del tráfico multiplexado, que tenga en cuenta todos

los factores que intervienen, desde la psicología de cada individuo, los protocolos que ejecuta, pasando por el sistema operativo y finalizando en la variabilidad del medio físico... lleva a un sistema matemático complejo e irresoluble. Dicho sistema plantea comportamientos caóticos que representados en un espacio, tienen un comportamiento fractal o de autosimilitud.

Un método de análisis del comportamiento de autosimilitud del tráfico se recoge en [Ale95] que a partir del cual permite obtener un estimador del comportamiento de la red en diferentes órdenes de magnitud temporal.

Sea un proceso estocástico estacionario

$$X = (X_1, X_2, X_3, X_4, \dots) \quad (7.7)$$

donde  $X_k$  denota el número de llegadas al conmutador producidas por unidad de tiempo. Definimos un nuevo proceso estocástico por agregación

$$X^{(m)} = (X_k^{(m)} : k = 1, 2, 3, \dots) \quad (7.8)$$

siendo

$$X^{(m)} = \frac{1}{m}(X_{km-m+1} + \dots + X_{km})(k = 1, 2, 3, \dots) \quad (7.9)$$

es decir, formado por el valor medio de la serie original  $X$  en bloques no superpuestos de tamaño  $m$ .

Si se considera  $X$  como un conjunto de variables aleatorias independientes, idénticamente distribuidas con media y varianzas finitas, entonces el proceso de agregación 7.7 mantiene su media, y la varianza decrece según  $m^{-1}$ .

Sin embargo observando el comportamiento del tráfico multiplexado real con protocolos subyacentes diversos, agregando valores de tráfico instantáneo en intervalos de tamaños diferentes se observa que la media cumple la expresión

$$E(X^{(m)}) \sim cte m \quad (7.10)$$

y la varianza sigue un comportamiento de disminución más lenta:

$$Var(X^{(m)}) \sim cte' m^{-\beta}, 0 < \beta < 1 \quad (7.11)$$

a medida que  $m \rightarrow \infty$ .

Este proceso de autosimilitud permite estimar el comportamiento del tráfico a diferentes escalas por agregación, desde escalas temporales bajas, del orden de us (relacionadas con los anchos de banda, Mbps) a escalas temporales mayores, las del vídeo, ms; este comportamiento promediador se asemeja a un filtro paso bajo.

Por tanto, partiendo de dicho comportamiento y tratando de seguir la variación del tráfico por los acontecimientos más recientes, se propone como se discute en el anexo D la expresión:

$$MACR = MACR + \alpha(ACR - MACR) \quad (7.12)$$

para la estimación de la tasa de celdas permitida, MACR (Mean Allowed Cell Rate) durante el proceso de descomposición 7.2.1 de 160 ms y que controla las variaciones de ACR (Allowed Cell Rate) respecto a la media a través del peso  $\alpha = \frac{1}{16}$ . Este peso como se recoge en el anexo anexoMACR se basa en el estudio de la frecuencia de corte del filtro 7.12 teniendo en cuenta las causas de variación del ACR.

Con ello se consigue un promediador del ACR instantáneo a diferentes intervalos de agregación, por acumulación de incrementos o decrementos respecto al MACR durante la evaluación cada vez más atenuados por potencias de  $\frac{1}{\alpha}$ , a medida que se alejen del instante de estimación. Una variación de la expresión anterior es utilizada en [Rob94].

Cabe destacar, que la expresión 7.12 (en la implementación realizada del sistema) se inicializa a

$$MACR = 8 MCR \quad (7.13)$$

es decir, ocho veces la tasa mínima de salida (MCR) negociada en la sección 7.1, por decisión de compromiso entre la calidad del vídeo y los valores de ICR, MCR y PCR de la fuente ABR (correspondientes a la tasa de salida de la fuente tanto en inicio, mínima y de pico) discutidos en la sección 10.1 del capítulo 10. En este capítulo se fija  $MCR = 0.851$  celdas por ms por lo que  $MACR = 6.8$  celdas por ms, lo que supone un factor de compresión en la fuente de 8 a 1, con una calidad visual alta.

Obviamente, la estimación del MACR es el valor óptimo durante 160 ms. Pero si la red cambia de estado drásticamente y las congestiones son frecuentes llenando las colas de los conmutadores, después de haber realizado la estimación, la transmisión de vídeo naturalmente se degrada bien por pérdidas de subbandas o bien por subbandas que llegan fuera de tiempo.

Todos estos factores afectarán a la estimación del MACR para que en la siguiente evaluación, MACR recoja los cambios que se están produciendo en la red. Con estos cambios, lo que hace la fuente es adaptar el factor de compresión a la nueva situación como se ve en la próxima sección.

#### 7.2.4 Diseño del subsistema de asignación de bits

El siguiente paso una vez estimado el comportamiento de la red a través de un valor medio MACR en 7.2.3, es asignar en función de dicho ancho de banda disponible, los bits por pixel o coeficiente correspondientes a cada una de las subbandas de la descomposición wavelet.

Los bits por pixel quedan determinados con un paso de cuantificación, que introduce mayor ruido a pasos de cuantificación grandes y menor a pasos de cuantificación pequeños. Obviamente si el ancho de banda permite tener pasos de cuantificación pequeños, los coeficientes de la descomposición pueden ser transmitidos casi perfectamente. Pero si no, dicho paso de cuantificación en el sistema propuesto introduce una distorsión por cada una de las subbandas según la prioridad.

El presente subsistema, trata de minimizar la distorsión global en la secuencia como suma de las distorsiones parciales asociadas a cada una de las subbandas. Este proceso se conoce comúnmente como *Bit-Allocation* y que se basa en la *teoría de Rate-Distorsion*[GS96].

La teoría *Rate-Distorsion*  $D = D(R)$  trata de optimizar dada una distorsión máxima permitida  $D$  en la señal reconstruida, cual es la tasa de bits alcanzable  $R$  (Rate) o viceversa, dada la tasa permitida o disponible de bits ( $R$ ), estimar la distorsión generada  $D$  (ver la forma de la curva en las figuras 7.15). Esta teoría ha sido desarrollada según medidas de error basadas en errores cuadráticos y no perceptuales. Esta teoría mantiene que gran parte de la reducción de la *tasa de bits* está relacionada con un ruido presente, al que se le da una forma que las componentes de distorsión no son percibidas según el error cuadrático.

Si se tiene en cuenta la distorsión perceptual en la figura 7.15, la curva *Rate-Distorsion* obtenida quedaría por debajo de la establecida por la distorsión matemática[GS96].

Por otra parte, según [Gra90][Far92][VK95] si tras la cuantificación existe un codificador de entropía (con códigos de longitud variable), un buen funcionamiento que ajusta el 3<sup>er</sup> Teorema de Shannon[Sha48] se basa en la utilización de un cuantificador uniforme con zona muerta para cada uno de los coeficientes dentro de cada subbanda. En caso de no estar el codificador de entropía y tomando como medida de distorsión el error cuadrático medio, los cuantificadores óptimos están tabulados según el método de Lloyd-Max[LLo82][Max60].

Una medida de la distorsión ( $D$ ) es el error cuadrático medio, expresado por

$$D = E(|x - x'|) = \sum_i \int_{x_{i-1}}^{x_i} (x - y_i)^2 f_x(x) dx \quad (7.14)$$

donde  $x$  es la señal de entrada,  $x'$  es la señal cuantizada,  $y_i$  es el valor de cuantificación entre  $x_{i-1}$  y  $x_i$  y  $f_x(x)$  es la función densidad de probabilidad de la variable  $x$ .

La expresión 7.14 en el caso de una función densidad de probabilidad uniforme y para una cuantificación escalar uniforme con  $x \in [x_{min}, x_{max}]$  (con intervalo de cuantificación es  $\Delta = \frac{x_{max} - x_{min}}{2^b}$ ) siendo  $b$  el número de bits por pixel, corresponde a una distorsión

$$D = \frac{\Delta^2}{12} \quad (7.15)$$

Si la media de cada subbanda es nula, como se ha visto en 7.2.1, entonces el paso de cuantificación es  $\Delta = \frac{x_{max}}{2^{b-1}}$  y que permite expresar la distorsión en según la varianza de la variable de entrada para una distribución uniforme de la forma[VK95]

$$D = \sigma^2 2^{-2b} \quad (7.16)$$

y que para un caso más general (una distribución no uniforme) se puede poner

$$D = c 2^{-2b} \sigma^2 \quad (7.17)$$

siendo  $c$  una constante dependiente de la función de densidad de probabilidad.

La expresión de la distorsión determina que ésta es proporcional a la varianza de la variable aleatoria de entrada, en este caso por la varianza de la distribución de los coeficientes después de la descomposición, que puede ser fácilmente calculada o estimada[SLG95].

Fijada la distorsión general y visto el procesamiento de descomposición subbandas, se puede hacer la suposición que las subbandas obtenidas están descorreladas y son independientes. Por tanto la distorsión total corresponde aproximadamente a la suma de distorsiones parciales o distorsiones asociadas a cada una de las subbandas

$$D \simeq \sum_{k=1}^M D_k \quad (7.18)$$

donde  $M$  es el número total de subbandas en la descomposición, en nuestro caso 22 (ver sección 7.2.1). Por tanto la asignación de bits queda restringida al comportamiento de la red y el codificador de entropía por  $R(b)$  que fija el número de bits por pixel de la secuencia original a través de la expresión

$$R(b) \leq 4\beta \frac{MACR'}{N f_{rate}} .48.8 = R \quad (7.19)$$

siendo

$$R(b) = \sum_{k=1}^M \alpha_k b_k \quad (7.20)$$

$MACR'$  es la tasa neta media de celdas por segundo (permitida de salida), eliminando la tasa de las celdas RM de control de flujo,  $\beta$  es la ganancia de bits asociada al factor de compresión del codificador de entropía,  $N$  es el número de pixels en un total de 4 fotogramas,  $f_{rate} = 25$  fotogramas/s,  $\alpha_k$  el tamaño relativo de la subbanda respecto al formato de fotograma original,  $b_k$  el número de bits por coeficiente en la subbanda  $k$ ,  $b = (b_1, b_2, \dots, b_M)$  el vector "número de bits por coeficiente para cada subbanda" y  $\sigma_k^2$  la varianza de los coeficientes de la subbanda  $k$ .

De este modo se puede relacionar la distorsión y la restricción de la tasa de bits permitida de salida. Por tanto el siguiente paso es minimizar la distorsión por la *teoría de Rate-Distorsion* ponderada con los pesos perceptuales, dada la restricción de los bits por pixel disponibles

$$D(b) = \sum_{k=1}^M \alpha_k \omega_K c_K 2^{-2b_k} \sigma_k^2 \quad (7.21)$$

donde  $\omega_K$  es el peso perceptual asignado en 7.2.2.

El problema de minimización de la función  $D(b)$  con restricciones  $R(b)$  se resuelve por el teorema de Lagrange

$$\min(D(b) + \lambda R(b)) \quad (7.22)$$

derivando e igualando a cero la expresión anterior y fijando  $\lambda' = \frac{\lambda}{c_k}$  podemos expresar finalmente los bits por coeficiente en cada subbanda

$$b_k = 0.5 \log_2 \left( \frac{2 \ln(2) \omega_k \sigma_k^2}{\lambda'} \right) \quad (7.23)$$

e introduciendo la restricción podemos despejar  $\lambda'$  por

$$\lambda' = 2^{\frac{[\sum^{22} \alpha_k (\log_2(2 \ln 2) + \log_2(\omega_k \sigma_k^2)) - 2 * R]}{\sum^{22} \alpha_k}} \quad (7.24)$$

Este comportamiento fija los puntos de trabajo sobre la curva de *Rate-Distorsion* en puntos con misma pendiente comúnmente conocidos como *Equal-Slope* (ver figura 7.15).

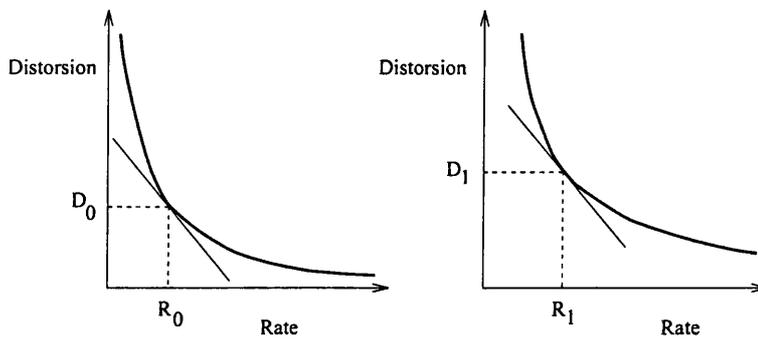


Figura 7.15: Punto de funcionamiento para dos subbandas en las curvas de *Rate-Distorsion*

Los resultados de este algoritmo se recogen en el capítulo 9, y como se puede ver las subbandas con poca energía se les asigna pocos bits y las subbandas más perceptibles, como las paso baja, con pesos y varianzas mayores consiguen mayor proporción de bits.

Sin embargo el algoritmo de minimización cuando la restricción es muy fuerte, puede hacer que haya subbandas que obtengan  $b_k$  mucho menores que 1, incluso valores negativos. En esta situación lo que se propone es no transmitir ninguna subbanda que no pueda asignar más de 1 bit por coeficiente, mínimo para al menos poder asignar el signo y reiniciar el proceso de minimización pero actualizando el valor de  $M$ .

Cabe destacar que dicho proceso de minimización por un lado tiene en cuenta la codificación de entropía (factor  $\beta$  en la expresión 7.19) y por otro lado, los bits por coeficiente pueden aumentar cuando se adapta al tamaño de paquete de la red; en el caso de ATM, cada celda puede transportar 384 bits útiles y por tanto el redondeo a una celda completa puede hacer aumentar los bits por coeficiente.

### 7.2.5 Diseño del subsistema de cuantificación

Determinados los bits por coeficiente en el subsistema anterior, se plantea en el presente subsistema obtener el paso de cuantificación y procesar cada coeficiente.

La cuantificación se puede realizar de dos formas diferentes, escalarmente o vectorialmente para cuantificar escalares o vectores respectivamente. Desde el punto de vista perceptual, no existen notables ventajas aportadas por una cuantificación vectorial respecto a una cuantificación escalar uniforme [GG92].

Para ello, conocido  $b_k$  es inmediato calcular el paso de cuantificación mediante

$$\Delta = \frac{4\sigma}{2^{b_k}-1} \quad (7.25)$$

donde se estima para las diferentes distribuciones de los coeficientes  $x_{\max} = 4\sigma$ .

Conocido el paso de cuantificación, la cuantificación uniforme consiste en dividir a cada coeficiente por el paso de cuantificador y redondear al entero más próximo

$$q = \text{round}\left(\frac{\text{coeficiente}}{\Delta}\right) \quad (7.26)$$

Aunque cuantificadores más optimizados incluyen zona muerta según el umbral de contraste en los coeficientes [Art95], en nuestra implementación por simplificación del sistema total, se ha optado por eludir el paso de cuantificación inicial. Además por tratarse de entornos de red con pérdidas, ya que los servicios *best effort* no ofrecen una garantía de calidad absoluta, las ventajas aportadas por estas técnicas de cuantificación pueden pasarse según nuestros objetivos a un segundo plano; esta discusión pasa a contenidos de trabajos mucho más específicos.

### 7.2.6 Diseño del subsistema de codificación de entropía

Tras la asignación de bits y discretizados los coeficientes a un reducido número de valores, el siguiente paso para reducir la cantidad de ceros existentes es aplicar la técnica de *Run-Level* descrita en el capítulo 6, consistente en generar parejas (*longitud, valor*) que corresponden al número de coeficientes nulos (*longitud*) hasta encontrar un coeficiente no nulo y codificar su valor.

Destacar que el gran número de coeficientes nulos en cada subbanda está relacionado con su función de distribución (ver anexo E), con media nula y desviaciones estándar muy bajas, lo que hace al aplicar el cuantificador que la mayoría de los valores sean nulos. Cuantos más valores nulos más redundancia estadística es eliminada por esta técnica.

Generadas las parejas (*longitud, valor*), el codificador de entropía propuesto está basado en el método de *Huffman*. Este método consiste en según las probabilidades de aparición de cada pareja, se va generando códigos para cada una de ellas, donde ninguno sea prefijo del otro.

De esta forma se asocian a las parejas de mayor probabilidad (las parejas más frecuentes) los códigos de menor longitud y las de menor probabilidad los códigos de mayor longitud.

La tabla de códigos de longitud variable utilizada puede verse en el anexo F, extraída de las fuentes del H.262 (ISO/IEC 13818-2). Estas tablas han sido generadas a partir de la evaluación de muchas secuencias de vídeo y recogen con carácter general la mejor codificación de entropía para códigos (*longitud, valor*).

### 7.2.7 Diseño del subsistema de adaptación a la red

Una vez obtenidas las diferentes subbandas tras la codificación de entropía, el presente subsistema realiza dos funciones:

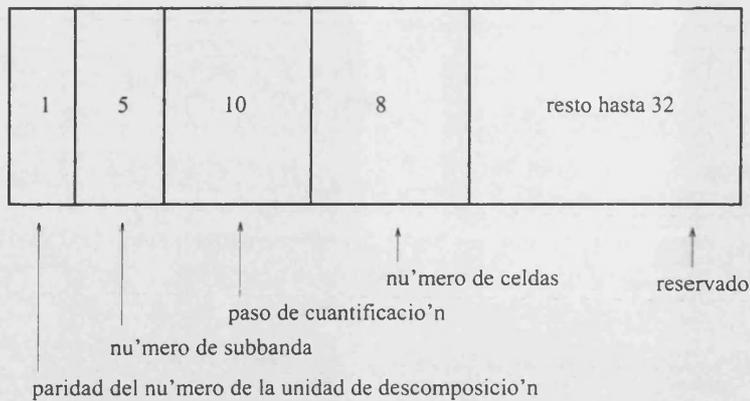


Figura 7.16: Descripción de los campos de la cabecera añadida a la SDU para la identificación y proceso de cada subbanda

- adaptar las subbandas al formato utilizado en la red y que permitan ser reconstruidas en el receptor una vez transportadas
- adaptar la fuente de vídeo al estado de la red, mediante una memoria intermedia

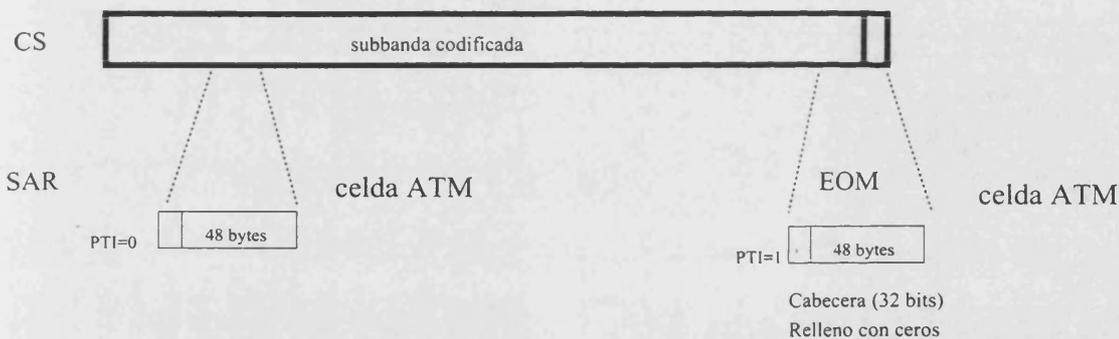


Figura 7.17: Formato de paquete de adaptación AAL5+ en ATM-ABR

A cada subbanda para su reconstrucción independiente y autónoma en el receptor, se le introduce una cabecera de 32 bits. Dicha cabecera recoge información de (ver figura 7.16):

- la paridad del número de la unidad de descomposición correspondiente con 1 bit, que permite distinguir las subbandas pertenecientes a 2 unidades consecutivas, cada una de ellas de 160 ms
- la identificación de cada una de las 22 subbandas con 5 bits
- la codificación del paso de cuantificación con 10 bits
- el número de celdas que contiene la subbanda con 8 bits

- y el resto hasta 32 bits para ampliaciones futuras o de reserva

Todo este proceso se realiza en la nueva capa AAL5 propuesta para adaptar las subbandas a la red ATM. La figura 7.17 muestra las funcionalidades de la capa. Destacar que una vez introducida la cabecera a la subbanda se introduce un relleno de ceros para completar a un número entero de celdas ATM.

La definición de una nueva interfaz AAL5+ de adaptación ATM para el servicio ABR diferente a la de la figura 2.5 del capítulo 2 se justifica porque:

- la información de CRC con 32 bits en AAL5 permite detectar errores pero no corregirlos, útil para transmisión de datos pero no para secuencias de vídeo. Sin embargo códigos correctores más potentes como Reed-Solomon, utilizados p.e. en AAL1, podrían ser utilizados, pero su misma complejidad no hace adecuado con los requisitos de tiempo real tanto en el procesamiento como en la manipulación de los datos a enviar
- la longitud total de celdas codificada en AAL5 con 16 bits, supera el tamaño de celdas a enviar para una subbanda vistos los formatos de videoconferencia más habituales; en su caso se opta por un campo de menos bits, concretamente 8

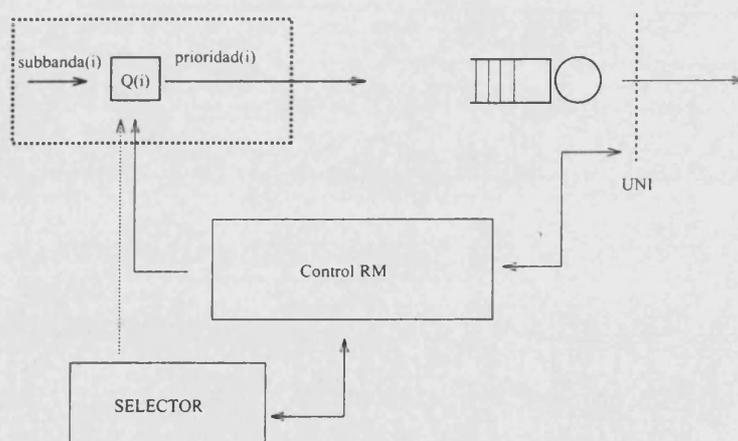


Figura 7.18: Memoria FIFO para adaptación a la red

Finalmente este subsistema al conectar con la red, figura 7.18, se adapta a las posibles variaciones de su estado a través de una cola que amortigua las posibles diferencias entre la salida y la entrada por la UNI.

La ocupación de la cola varía según la evolución de la secuencia de vídeo y de la tasa de emisión a la red. Se incorpora con el objetivo de no perder el sincronismo, un mecanismo de vaciado o inicialización de dicha cola cada 160 ms, que corresponde a la frecuencia de funcionamiento de la etapa de descomposición. La subbanda de máxima prioridad gobierna el vaciado de la FIFO, ocupando siempre la subbanda DPCM el primer lugar para la transmisión y tras ella todo el segundo nivel.

Obviamente en el proceso de inicialización de la cola ha de garantizarse que las subbandas hayan sido ordenadas con prioridad dentro de la FIFO, de forma que las más importantes siempre serán transmitidas.

Dentro de dicha ordenación además ha de tenerse en cuenta, como se ve en la figura 7.4 el proceso de descomposición en dos partes asociadas a los primeros 80 ms del primer nivel y los 80 ms posteriores del segundo nivel. La primera parte sólo genera un nivel de resolución con 7 subbandas y la segunda parte genera el segundo nivel con 8 subbandas de máxima prioridad y el primer nivel con otros 7 subbandas.

Debido que en la segunda parte van a ser transmitidas las 8 subbandas de la segunda resolución, mientras en la primera parte no, para no generar un desequilibrio en la recepción de las subbandas asociadas al primer nivel se estima el tamaño de la cola registrando las subbandas transmitidas. Conocidas éstas, en la siguiente unidad de descomposición se reparten las subbandas a transmitir entre ambas partes de forma equitativa.

## 7.3 Diseño del sistema de recepción

Una vez visto el sistema de emisión, fácilmente puede reconstruirse el sistema de recepción. Cabe destacar que con esta analogía, se genera un equilibrio de complejidad tanto en la construcción de un sistema como en otro, evitando que como en los estándares, el coste del codificador sea mucho mayor por ser más complejo, por la estimación de movimiento frente al decodificador.

### 7.3.1 Diseño del subsistema de adaptación a la red

Como ya se comenta en el capítulo anterior, la finalidad de este subsistema es adaptar el servicio ofrecido por la red de una conexión ABR al servicio esperado por la aplicación.

Por ello, el principal objetivo es eliminar los efectos en la variabilidad del tiempo de llegada de las celdas a la entrada del receptor. Este proceso se realiza activando un temporizador a la llegada de la primera celda de la sesión, de forma que no se activa el sincronismo ni el receptor, hasta pasado un tiempo suficiente para que la cola del sistema de recepción pueda absorber las posibles variaciones del tiempo de llegada de las celdas.

El tiempo de retraso queda acotado por los valores siguientes:

- superiormente, determinado por las características de la aplicación de tiempo real, la cual generalmente estima en unos 400 ms el retraso total entre los dos extremos (este retraso se calcula empíricamente como cota máxima para poder establecer una conversación sin problemas entre dos personas[Sus96])
- inferiormente, determinado por el mínimo retraso necesario para poder absorber a lo largo del circuito virtual las máximas variaciones entre los tiempos de llegada, generadas en los momentos de congestión cuando las colas de los conmutadores están llenas y además los anchos de banda disponibles de salida son mínimos (este tiempo se discute analíticamente en el capítulo 10 sección 10.2)

Esta cola de recepción está gobernada por el proceso de sincronismo que regula sus contenidos de forma que la información retardada o caduca se vacía rápidamente para no entorpecer la llegada del resto de celdas.

Este subsistema al mismo tiempo controla la posible pérdida de celdas en las subbandas, a través de la capa AAL5+ utilizando el campo de longitud de celdas transportadas; aquellas subbandas correctamente recibidas son ofrecidas al nivel superior.

### 7.3.2 Diseño del subsistema de decodificación de entropía

Recibidas las subbandas correctamente, el paso siguiente es la búsqueda de los códigos en la tabla del anexo F que permite recuperar las parejas (*longitud, valor*) que directamente reconstruyen las secuencias de coeficientes cuantificados, tanto nulos como no nulos.

### 7.3.3 Diseño del subsistema de cuantificación inversa

De la misma manera que se procesa en 7.2.5, el paso inverso de reconstrucción para la cuantificación viene fijado por

$$\text{coeficiente}' = q * \Delta \quad (7.27)$$

### 7.3.4 Diseño del subsistema de reconstrucción

Visto el capítulo 4, los filtros biortogonales implementan una reconstrucción a partir de un banco de filtros síntesis; dichos filtros son elegidos de manera que tengan una respuesta más regular que los de análisis[GS96]. Las respuestas de dichos filtros son analizadas en el anexo C.

Una vez aplicada la interpolación y el filtrado con los bancos de filtros, la suma de las diferentes subbandas recibidas ya procesadas, recupera con la “mejor intención” la secuencia original. Las subbandas que no han sido recibidas correctamente, en la implementación realizada se ha decidido por estimarlas con valor nulo, su valor medio.

Un punto discutido en el capítulo 11 es la posibilidad de recuperación de información mediante la utilización de las respuestas de los filtros paso baja temporal, que son aquellas con menor cantidad de movimiento.

## 7.4 Conclusión

El sistema propuesto presentado ha sido diseñado para tratar de aportar mejores prestaciones respecto a la calidad visual de las secuencias de vídeo, cuando éstas son transportadas por redes con servicios *best effort* que incorporen controles de flujo cerrado basados en tasa explícita.

La arquitectura tanto del sistema de emisión como de recepción está formada por diversos subsistemas de manera que cada uno de ellos cumpla su función especificada.

## **Parte IV**

# **Evaluación de prestaciones del sistema propuesto**

En la parte siguiente se tratará de evaluar las prestaciones y los comportamientos del sistema propuesto.

La presente parte de resultados está estructurada en tres capítulos.

El primero de ellos, capítulo 8, describe la metodología a seguir a lo largo de toda la evaluación.

El capítulo 9 recoge los resultados analíticos y teóricos de tres puntos relevantes del sistema: prioridades de las subbandas, asignación de bits y modificación del factor de compresión.

Finalmente, el capítulo 10 recoge los resultados experimentales, basados en simulación sobre configuraciones de red ATM con dos conmutadores ( $2N$ ) y diferentes tipos de tráfico a través de sus conexiones (persistente y ráfaga).



# Capítulo 8

## Metodología de verificación

El objetivo de este capítulo es describir las fases a realizar para verificar el sistema y especificar los elementos que van a formar parte en el proceso de evaluación: las secuencias de vídeo, la configuración de red utilizada, el tráfico que interactuará con el sistema a través del resto de conexiones(tráfico de fondo) y el algoritmo de control de flujo.

### 8.1 Técnicas de verificación. Proceso de pruebas y ensayos

El proceso de pruebas y ensayos pretende especificar el comportamiento tanto estático como dinámico del sistema propuesto. Para ello se planifica la parte de verificación y experimentación en dos fases.

La primera fase, capítulo 9, analiza y verifica el funcionamiento de los algoritmos ante situaciones teóricas, donde se suponen condiciones de entorno que engloban los posibles comportamientos a los que va a ser sometido el sistema.

En la segunda fase, capítulo 10, se realizan simulaciones en dos conmutadores ATM con servicio ABR y enlace cuello de botella (figura 8.2), interactuando con fuentes de comportamientos tanto persistente como ráfaga. En esta fase, los datos de entrada para la simulación del sistema de vídeo son ficheros de traza que recogen el comportamiento de las diferentes secuencias (tanto las secuencias como el formato de los ficheros traza se ven en la siguiente sección).

Dichas simulaciones permiten verificar los puntos críticos del sistema ante cambios de ancho de banda y situaciones de congestión y pérdidas. A través del resto de conexiones en los conmutadores ATM se introducen tráficos con comportamientos totalmente diferentes, persistente y ráfaga, que se analizan en la sección 8.4.

El simulador utilizado (DANAE [Gal97]), es un simulador de carácter general orientado a objetos que se ejecuta sobre Windows95 o NT. Los sistemas a simular son introducidos por bloques que a su vez pueden ser formados por otros bloques y/o primitivas enlazados a través de conexiones. Otros analizadores más accesibles pueden ser también utilizados[Su96].

La interacción de los bloques en el simulador puede ser tanto por memorias compartidas como por enlaces de los puertos de cada bloque. El sistema evoluciona como interacción de unos bloques con otros, a través de eventos planificados que conllevan acciones específicas: salida de una celda, llegada de una celda, etc.

## 8.2 Secuencias de vídeo bajo test

Las secuencias utilizadas en esta parte de los resultados son tres secuencias de vídeo frecuentemente utilizadas para la evaluación de diferentes técnicas de procesado realizadas sobre vídeos. El contenido de cada una de las secuencias se describe a continuación y todas ellas son secuencias del tipo busto parlante:

- *Salesman*: secuencia de un vendedor con camisa blanca y corbata gris oscura a rayas blancas, filmado en un despacho a una distancia de unos 5 metros que sostiene una caja en su mano derecha que mueve constantemente al mismo tiempo que con la cabeza y el otro brazo gesticula durante la exposición; esta secuencia es de las tres, la que más cantidad de movimiento contiene
- *Miss América*: secuencia de una señorita vestida con blusa negra y un chaleco gris, entrevistada en una sala de fondo gris oscuro, que gesticula con la cabeza, moviendo los labios y cerrando los ojos
- *Claire*: secuencia de una azafata de vuelo, vestida con una blusa blanca con chaqueta gris de cuello negro, que mueve el tórax, los labios y la cabeza rápidamente, grabada en una sala con fondo gris claro uniforme

En el anexo B dichas secuencias son visualizadas y procesadas con el sistema de codificación H.263 y transmitidas por diferentes canales de transmisión con probabilidades de error determinadas.

Estas secuencias son registradas para simulación por ficheros traza como se ha descrito previamente. Estos ficheros registran la descomposición wavelet biortogonal en dos niveles de resolución vista en el capítulo 7; esto supone procesar grupos de 22 subbandas que se registran en el fichero a intervalos de 160 ms de secuencia. Se realiza de esta manera, para evitar realizar en el computador al mismo tiempo, la descomposición de la secuencia de vídeo junto con la simulación. El formato de los ficheros traza indexa cada unidad de descomposición, de la cual identifica los siguientes campos:

- número de subbanda
- varianza de la función de distribución de los coeficientes de dicha subbanda
- peso perceptual de dicha subbanda

hasta el total de 22 subbandas.

La figura 8.1 presenta la varianza de cada uno de las 22 subbandas para 1600 ms de la secuencia de Miss América registrada en los ficheros traza, descompuesta con los filtros 4.69 vistos en el capítulo 4.

La identificación de la varianza para cada subbanda en la figura 8.1 es de izquierda a derecha y de arriba a abajo, dentro de cada subfigura:

1. dpcm (sólida), pb2rvH (punteada), pb2rVh (a trazos), pb2rVH (trazo-punto)
2. pa2rvh (sólida), pa2rvH (punteada), pa2rVh (a trazos), pa2rVH (trazo-punto)
3. pblrt1vH (sólida), pblrt1Vh (punteada), pblrt1VH (a trazos)
4. palrt1vh (sólida), palrt1vH (punteada), palrt1Vh (a trazos), palrt1VH (trazo-punto)
5. pblrt2vH (sólida), pblrt2Vh (punteada), pblrt2VH (a trazos)
6. palrt2vh (sólida), palrt2vH (punteada), palrt2Vh (a trazos), palrt2VH (trazo-punto)

donde “pa” y “pb” indica la salida paso alta y baja temporal respectivamente, “2r” y “1r” denota la resolución o iteración de los filtros (segunda y primera resolución), “v” y “V” denotan la salida del filtro paso baja y alta vertical, de la misma manera que “h” y “H”, y finalmente “t1” y “t2” denotan las subbandas del primer nivel de resolución procedente de la primera parte o segunda parte respectivamente. Destacar en dichas figuras que los valores más altos de las varianzas están asociados a las subbandas de menor frecuencia (segundo nivel de resolución y especialmente el paso baja final codificado en DPCM), como cabía esperar según el sistema visual humano.

En la tabla 8.1 se recoge la varianza media asociada a cada una de las subbandas tras la descomposición de las secuencias anteriores.

### 8.3 Configuración de red: 2 conmutadores

Una decisión de compromiso para analizar el comportamiento del sistema y poder verificar sus funcionalidades ha sido discutir el banco de pruebas o la configuración de la red donde evaluar el sistema propuesto.

Los estudios realizados en los servicios ABR de ATM utilizan configuraciones conocidas [Kim96]. De todas ellas, la más interesante para testeo del sistema propuesto es la de dos conmutadores con enlace troncal cuello de botella (ver figura 8.2), ya que las perturbaciones introducidas a la red son más controlables y directas, permitiendo evaluar los puntos críticos del sistema.

En este caso, las perturbaciones utilizadas se producen a través de técnicas de cambio en el ancho de banda del enlace troncal [Mar97] e interacción con diferentes tipos de tráfico.

<i>subbanda</i>	<i>Miss América</i>	<i>Salesman</i>	<i>Claire</i>
dpcm	104692.231	115277.5156	86174.9015
pb2rvH	2052.09	14848.268616	12766.7974
pb2rVh	2641.96	12499.5513	9064.16796
pb2rVH	56.25	335.8793	118.1300
pa2rvh	6961.31579	185979.9312	352739.5802
pa2rvH	473.0146512	822.085	551.74003
pa2rVh	274.8698726	810.43023	786.4296
pa2rVH	30.79251081	35.14584	19.5029
pb1rt1vH	21.77715556	149.014699	107.447857
pb1rt1Vh	17.97675201	95.897282	92.126884
pb1rt1VH	0.295936	0.746615	0.353321
pa1rt1vh	126.3870608	9417.239278	18208.025262
pa1rt1vH	12.73276489	16.035976	6.920156
pa1rt1Vh	3.99160441	14.959562	7.447902
pa1rt1VH	3.10323456	2.324789	0.081903
pb1rt2vH	21.62064004	148.14699	107.346833
pb1rt2Vh	18.33295489	103.37628	93.213821
pb1rt2VH	0.28826161	0.83621	0.252171
pa1rt2vh	97.93477444	9357.24682	15624.27212
pa1rt2vH	13.43295801	12.025916	7.502356
pa1rt2Vh	2.61760041	15.059361	3.46318
pa1rt2VH	3.150625	2.436119	0.17804

Tabla 8.1: Varianza media de cada subbanda tras la descomposición 3D propuesta para las secuencias de Miss América, Salesman y Claire

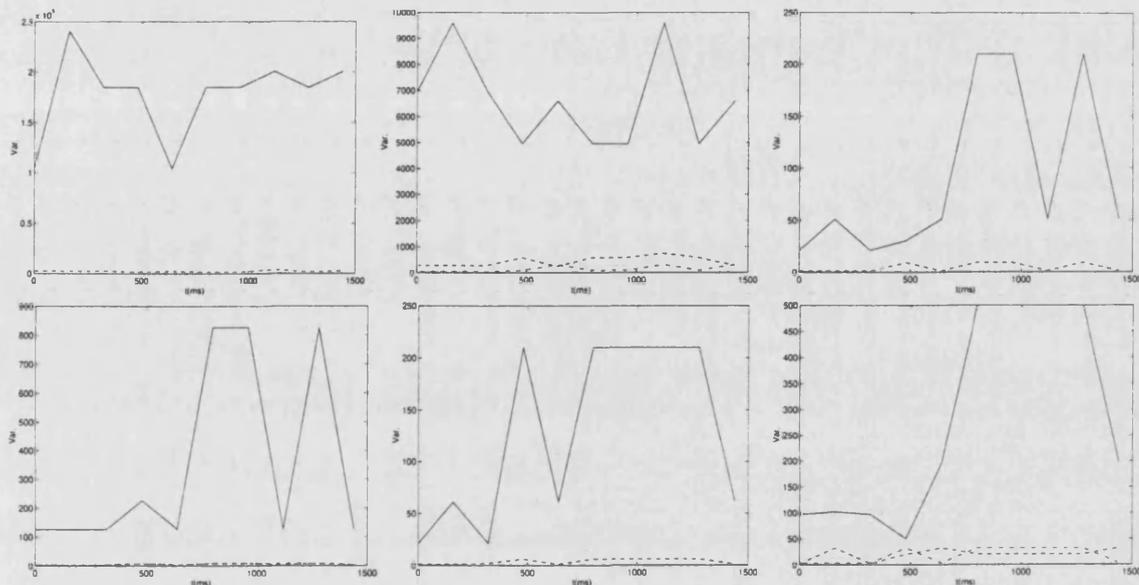


Figura 8.1: Evolución de las varianzas asociadas a las 22 subbandas de la descomposición 3D durante 1600 ms en la secuencia de Miss América

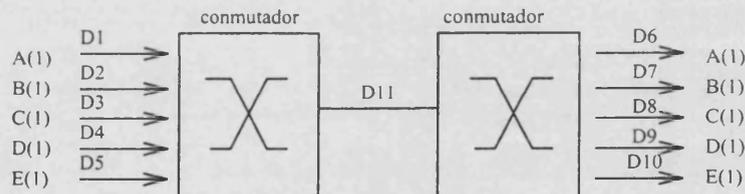


Figura 8.2: Configuración de red con dos conmutadores y enlace troncal cuello de botella

Respecto a las distancias utilizadas para los enlaces, se proponen los escenarios tanto LAN como MAN (descritas en la tabla 8.2).

No se considera el escenario WAN (con enlaces troncales de miles de Km) porque los resultados en los escenarios anteriores son suficientes para comprobar el funcionamiento del sistema, delegando en su caso al control de flujo la calidad de servicio deseada.

## 8.4 Patrones de tráfico

El comportamiento de los diferentes tipos de tráfico que intervienen en las simulaciones son un elemento fundamental que requiere un estudio previo.

Como las fuentes no presentan un perfil de tasa binario ( $s(t)$ ) constante, un parámetro que permite clasificar el tráfico generado por las fuentes es el coeficiente de ráfaga ( $B$ )

escenario	acceso (km)	troncal (km)
LAN	D1=D2=...=D10=0.2	D11=2
MAN	D1=D6=50 D2=D7=20 D3=D8=10 D4=D9=5 D5=D10=1	D11=50

Tabla 8.2: Distancias utilizadas para la configuración de red con dos conmutadores (2N)

definido por

$$B = \frac{\max [s(t)]}{E [s(t)]} \quad (8.1)$$

cociente entre la tasa pico de tráfico instantáneo y la tasa media.

Una primera clasificación a partir del coeficiente ráfaga permite dividir en tráfico en dos tipos: persistente y ráfaga.

El tráfico persistente tiene características suaves (es decir la tasa de pico es próxima a la tasa media) y por regla general es un tipo de tráfico que puede ser predicho con facilidad en la simulación. En el caso particular de las conexiones ABR en el conmutador, se asocia a aquellas fuentes que están adaptadas en todo momento al estado de la red a través de las pautas marcadas por el control de flujo. Este patrón de tráfico es generado por fuentes que siempre quieren transmitir y que en régimen permanente en el conmutador, trabajan a la porción de ancho de banda equitativo. Un ejemplo de este tipo de tráfico son los flujos de vídeo y audio en tiempo real, que pueden tener coeficientes ráfaga (B) aproximadamente entre 1 y 3.

El tráfico ráfaga se caracteriza por la aleatoriedad de las emisiones y la duración de ellas, por lo general inferiores al tiempo necesario para que se pueda cerrar el bucle de realimentación del control de flujo ABR; es tipo de tráfico es generado por fuentes que están activas y desactivas de forma aleatoria (on/off), con lo que la tasa de pico suele ser bastante mayor que la tasa media. Un ejemplo de este tipo de tráfico caracteriza a la transmisión de datos entre computadores, que pueden tener coeficientes ráfaga (B) aproximadamente alrededor de 10.

Este tráfico es el más crítico para el sistema propuesto porque al no adaptarse al estado de congestión de la red, puede desbordar las colas de los conmutadores y producir pérdidas.

El tráfico ráfaga se implementa mediante fuentes tipo ráfaga de forma similar que [Mar97]. Cabe destacar de dicha implementación que se realiza en modo unidireccional donde los patrones de tráfico ráfaga son generados a priori a la simulación mediante un fichero traza de parejas (retardo, longitud) que identifican la espera en ms hasta la siguiente emisión y la longitud de ráfaga (número de celdas en la ráfaga). Estos ficheros para evitar correlaciones en los patrones de tráfico y para generar la mayor aleatoriedad posible, se fuerza a que las diferentes fuentes lo lean en posiciones diferentes.

Las características de las ráfagas quedan analizadas en [Ada95] y [Dal94]. El primero caracteriza las ráfagas para un escenario en el que terminales remotos están

conectados a sendas redes *Ethernet* y desean intercambiar información mediante TCP/IP a través de una red ATM mediante una emulación de LAN (LAN Emulation), considerando una distribución bimodal de ráfagas de 2 y 32 celdas respectivamente (estos valores proceden de una discusión de los tamaños mínimos y máximos de las tramas de 64 y 1518 bytes respectivamente). En el segundo caso, presenta un escenario en el que parejas de terminales ATM establecen circuitos virtuales entre ellos para transferirse información utilizando TCP/IP encapsulando los paquetes IP en tramas AAL5, considerando combinaciones de longitudes 100, 300, 1000, 3000 y 10000 celdas equiprobables con retrasos de 1, 4, 10 y 40 ms. Las fuentes más usadas en simulaciones ATM se recogen en [Woj94].

En las simulaciones realizadas, cuyos resultados se recogen en el capítulo 10, se opta por las fuentes analizadas en [Ada95] porque se ajustan mejor al escenario de red con un enlace troncal de 10 Mbps, produciendo los efectos transitorios deseados en los conmutadores. En el segundo caso [Dal94], al ser las ráfagas excesivamente largas y la capacidad del enlace troncal de 10 Mbps, hace que dichas fuentes pasen a comportarse de forma persistente.

## 8.5 Control de flujo: CAPAC

Como se ha dicho en capítulos anteriores, el objetivo del algoritmo de control de flujo en el conmutador es repartir el ancho de banda disponible (sobrante a las conexiones CBR y VBR) de forma equitativa y eficiente entre las conexiones activas.

Del algoritmo CAPAC (*Congestion Avoidance with Proportional Adaptive Control*) para control de flujo utilizado en el conmutador, propuesto en [Mar97], cabe destacar los siguientes méritos:

1. complejidad de implementación reducida, sin requerir ningún tipo de parámetros por conexión, por lo que su uso en redes WAN es factible
2. una respuesta muy agresiva en estado de congestión, asegurando que los picos en la ocupación de las colas sean reducidos manteniéndose las colas por lo general vacías, que favorece para evitar la variabilidad de los retrasos
3. presenta un excelente comportamiento aunque es difícil garantizar su estabilidad con ACR muy rápidos y desmesurados por su agresividad en el control de la congestión, pudiendo causar que el punto de funcionamiento salga fuera de la región de captura del algoritmo y se vea incapaz para controlar las oscilaciones

## 8.6 Conclusión

Se ha presentado la metodología a seguir para poder evaluar el sistema ante situaciones muy diferentes que permitan estudiar sus comportamientos y prestaciones.

Por ello, en los capítulos siguientes abordamos el trabajo desde dos enfoques, tanto de manera estática, es decir evaluando los algoritmos y los elementos que forman el

sistema de forma separada, como de forma dinámica a través de simulaciones donde el sistema completo trabaja en su totalidad.

# Capítulo 9

## Análisis de los algoritmos

Previo a la evaluación experimental mediante simulaciones sobre redes ATM con servicio ABR del capítulo 10, es interesante en este capítulo validar los elementos y subsistemas que forman el sistema propuesto. El objetivo es observar el comportamiento y analizar los resultados de forma muy concreta para los tres puntos más relevantes del sistema: prioridades de las subbandas, asignación de bits y modificación del factor de compresión.

Con ello también se pretende poder prever las posibles respuestas del sistema en el capítulo siguiente y entender mejor los resultados.

### 9.1 Estudio de las prioridades de las subbandas. Pérdidas de subbandas

En el proceso de transmisión se pueden perder subbandas. En esta sección trata de reproducirse los diferentes efectos visuales ante la situación de pérdidas, de manera particular para diferentes casos.

Con la finalidad de no presentar todas las combinaciones posibles de las 22 subbandas tras la descomposición, se seleccionan las más relevantes. Para ello, se centra el estudio sobre las subbandas más visibles que según se ha visto en la sección 7.2.2 del capítulo 7 van asociadas a las subbandas con mayor resolución. Para resaltar más su efecto se acumulan en ellas la mayor parte de la información, para lo cual se realiza la descomposición sobre fotogramas iguales y por tanto ubicando toda la información en las subbandas paso baja temporales.

Estas subbandas resultantes son eliminadas tanto de forma individual como en diferentes combinaciones en el momento de la reconstrucción. Además para no confundir los resultados no se aplica el proceso de cuantificación en la reconstrucción.

Las siguientes figuras son el resultado de pérdidas generadas en diferentes subbandas procedentes de la descomposición propuesta, asociada a las bajas frecuencias temporales, en los fotogramas de Miss América (sus contenidos son leídos de izquierda a derecha y de arriba a abajo):

1. figura 9.1, pérdida de una subbanda en la segunda resolución: pérdida del residuo, detalle horizontal, vertical y diagonal
2. figura 9.2, pérdida de dos subbandas en la segunda resolución: pérdidas del residuo y detalle horizontal, detalle vertical y diagonal, residuo y detalle vertical, detalle horizontal y diagonal, residuo y detalle diagonal, detalle horizontal y vertical
3. figura 9.3, pérdida de subbandas en la primera resolución: pérdida del detalle horizontal, vertical, diagonal, detalle horizontal y vertical, detalle vertical y diagonal, detalle horizontal y diagonal
4. figura 9.4
  - (a) pérdida de detalle vertical de segunda resolución y pérdida de detalle vertical, diagonal y horizontal respectivamente de la primera resolución
  - (b) pérdida de detalle horizontal de segunda resolución y pérdida de detalle vertical, diagonal y horizontal respectivamente de la primera resolución
  - (c) pérdida de detalle diagonal de segunda resolución y pérdida de detalle vertical, diagonal y horizontal respectivamente de la primera resolución

donde la notación utilizada ha sido: residuo identifica la salida de los filtros paso baja y detalle la salida de los filtros paso alta manteniendo los demás paso baja, a excepción del detalle diagonal donde ambas salidas son paso alta.

Por tanto el detalle horizontal se refiere tanto a la salida del filtro paso alta horizontal con paso baja vertical, detalle vertical a la salida del paso alta vertical y paso baja horizontal y el detalle diagonal a la salida de los paso alta tanto vertical como horizontal.

## 9.2 Asignación de bits. Estudio de bits por coeficiente

Los siguientes resultados evalúan los algoritmos de minimización mediante Lagrange aplicado a las subbandas de la descomposición 3D (ponderando el peso perceptual y la distorsión de cada una de las subbandas) y el codificador de entropía del sistema propuesto, ante diferentes tasas de bits permitidas y para la secuencia de Miss América tanto en formato QSIF como SIF.

La asignación de bits es realizada a partir de valores linealmente crecientes de la tasa permitida de bits de salida o MACR, que teóricamente corresponde con el MACR *neto*; se refiere con *neto* a la tasa de la salida de la fuente habiendo excluido ya la tasa de celdas RM para el control de flujo.

Los contenidos de las figuras se comentan a continuación tanto para el formato SIF como QSIF respectivamente:

1. figuras 9.5 y 9.6 los bits asignados para los coeficientes (bpp) en cada subbanda
2. figuras 9.7 y 9.8 el número de celdas ATM necesarias para la transmisión de cada subbanda en servicio ABR con AAL5+
3. figuras 9.9 y 9.10 el paso de cuantificación para cada subbanda, con una cuantificación uniforme

donde cada subbanda, al igual que en la sección 8.2 del capítulo 8, se identifica en cada subfigura, de izquierda a derecha y de arriba a abajo con:

1. dpcm (sólida), pb2rvH (punteada), pb2rVh (a trazos), pb2rVH (trazo-punto)
2. pa2rvh (sólida), pa2rvH (punteada), pa2rVh (a trazos), pa2rVH (trazo-punto)
3. pblrt1vH (sólida), pblrt1Vh (punteada), pblrt1VH (a trazos)
4. palrt1vh (sólida), palrt1vH (punteada), palrt1Vh (a trazos), palrt1VH (trazo-punto)
5. pblrt2vH (sólida), pblrt2Vh (punteada), pblrt2VH (a trazos)
6. palrt2vh (sólida), palrt2vH (punteada), palrt2Vh (a trazos), palrt2VH (trazo-punto)

Cabe destacar los picos (saltos) que aparecen en algunas de las gráficas anteriores, que son asociados al mismo algoritmo de minimización cuando una nueva subbanda supera el umbral (ver capítulo 7) para ser transmitida y se disponen de suficientes bits para asignar. En dicha situación, los bits disponibles para asignar a las subbandas se reparten ahora entre, tanto las que ya se estaban transmitiendo como las nuevas que se pueden transmitir, lo que hace descender los bits de las primeras y comenzar a transmitir las segundas.

Cabe destacar que la cota máxima de asignación de bits depende de las propiedades de las imágenes de la secuencia original, en este caso 256 niveles de gris (8 bpp); por tanto la cota máxima de bits por coeficiente queda fijada cuando dado un MACR *neto* a cada pixel de la imagen original se le asignan 8 bits.

### 9.3 Variación del factor de compresión

En esta sección se presentan los resultados de 4 fotogramas consecutivos de la secuencia de Miss América (unidad de descomposición) en formato SIF a diferentes factores de compresión desde 64:1 (0.125 bpp) hasta 3:1 (2.66 bpp).

Las variaciones en el factor de compresión en el sistema final vienen determinadas por el estado de congestión de la red, a través del control de flujo ABR y el estimador de bits (sección 7.2.3 del capítulo 7).

En la tabla 9.1 se recogen los diferentes pasos de cuantificación para la descomposición de los 4 fotogramas teniendo como entradas en la tabla, los bits por pixel y

cada una de las 22 subbandas. La sintáxis utilizada para identificar cada subbanda se recoge en la sección 8.2 del capítulo 8.

Las figuras 9.11, 9.12, 9.13, 9.14, 9.15, 9.16 y 9.17 reconstruyen cada uno de los 4 fotogramas con 0.1278(62.59:1), 0.2424(33:1), 0.5305(15.08:1), 0.7575(10.56:1), 1.0606(7.54:1), 1.9696(4.06:1) y 2.66(3:1) bits por pixel (factor de compresión) respectivamente, con los pasos de cuantificación de la tabla 9.1.

Cabe destacar a la vista de la tabla 9.1, que cuando el paso de cuantificación es infinito (valor inf.) la subbanda codificada desaparece; esto como cabía esperar es más frecuente en las subbandas de la primera resolución que en las de la segunda, ya que estas últimas tiene más energía.

En la situación anterior y cuando son todas las subbandas de la primera resolución han desaparecido, se percibe un efecto de disminución a la mitad en la tasa de fotogramas por segundo, o lo que es lo mismo tener iguales cada dos fotogramas (porque las diferencias o matices entre ellos se encuentran en dicho nivel de resolución que desaparece). Son los casos de 0.1278 y 0.2424 bits por pixel en la tabla 9.1.

El caso extremo es codificar simplemente la subbanda DPCM o paso baja final, ya que en esta situación los fotogramas por segundo se reducen a una cuarta parte o lo que es lo mismo en recepción generar cada 4 fotogramas iguales.

## 9.4 Conclusión

El presente capítulo ha tratado de analizar los diferentes comportamientos del sistema ante situaciones posibles en una transmisión con servicios *best effort*; se ha tratado de constatar las hipótesis, suposiciones y aproximaciones realizadas en la fase de análisis y diseño del sistema para que éste funcione con los requisitos preestablecidos.

Para ello se ha sometido al sistema a tres estudios: el comportamiento de la pérdida de información (pérdidas de subbandas), la asignación de los bits variable con el ancho de banda disponible y la adaptación que el sistema realiza modificando su factor de compresión.

En definitiva, se ha pretendido evaluar las prioridades de las subbandas y la asignación proporcional de bits según el peso perceptual de cada una de ellas y con ello, se ha terminado estudiando como afecta el factor variable de compresión a la calidad visual.

Se concluye, que el sistema propuesto permite obtener mejores rendimientos ante entornos de red con servicio *best effort* mejorando la calidad visual frente a los codificadores actuales de vídeo. Obviamente, como contrapartida existen algunos inconvenientes y dificultades que son discutidas en el capítulo 11.

<i>subbanda vs bpp</i>	0.1278	0.2424	0.5305	0.7575	1.0606	1.9696	2.66
dpcm	20.3819	5.0656	2.5303	1.2645	0.6321	0.0790	0.0197
pb2rvH	24.1600	5.7524	2.8535	1.4212	0.7092	0.0885	0.0221
pb2rVh	27.4133	6.5270	3.2378	1.6125	0.8047	0.1004	0.0251
pb2rVH	60.0000	20.0000	4.0000	1.9355	0.9524	0.1174	0.0293
pa2rvh	44.4984	10.5949	2.6176	2.6176	1.3062	0.1630	0.0407
pa2rvH	57.9971	11.5994	5.6126	2.7618	1.3700	0.1701	0.0425
pa2rVh	44.2112	18.9477	4.2785	2.1053	2.1053	0.2596	0.0648
pa2rVH	Inf	Inf	6.3418	6.3418	2.9595	0.3495	0.0869
pblrt1vH	Inf	Inf	12.4443	12.4443	5.3333	0.5926	0.1464
pblrt1Vh	Inf	Inf	11.3064	11.3064	4.8456	0.5384	0.1330
pblrt1VH	Inf	Inf	Inf	Inf	Inf	0.6217	0.1404
palrt1vh	Inf	89.9376	12.8482	5.9958	2.9012	0.7082	0.0879
palrt1vH	Inf	Inf	Inf	9.5155	4.0781	0.9209	0.2248
palrt1Vh	Inf	Inf	Inf	15.9832	5.3277	1.0655	0.1259
palrt1VH	Inf	Inf	Inf	Inf	14.0928	0.9395	0.2237
pblrt2vH	Inf	Inf	12.3995	12.3995	5.3141	0.5905	0.1459
pblrt2Vh	Inf	Inf	11.4179	11.4179	4.8934	0.5437	0.1343
pblrt2VH	Inf	Inf	Inf	Inf	Inf	0.6136	0.1386
palrt2vh	Inf	79.1696	11.3099	5.2780	2.5539	0.6234	.1549
palrt2vH	Inf	Inf	Inf	9.7736	4.1887	0.9458	0.2309
palrt2Vh	Inf	Inf	Inf	Inf	12.9432	0.8629	0.2054
palrt2VH	Inf	Inf	Inf	Inf	14.2000	0.9467	0.2254

Tabla 9.1: Pasos de cuantificación según los bits por pixel para cada una de las 22 subbandas de la descomposición 3D propuesta en Miss América con formato SIF



Figura 9.1: *Ejemplos de pérdidas de una subbanda en 2ª resolución: 1-residuo 2-detalle horizontal 3-vertical 4-diagonal*



Figura 9.2: *Ejemplos de pérdidas de dos subbandas en 2ª resolución, de izquierda a derecha y de arriba a abajo: 1- residuo y detalle horizontal, 2-detalle vertical y diagonal, 3-residuo y detalle vertical, 4-detalle horizontal y diagonal, 5-residuo y detalle diagonal, 6-detalle horizontal y vertical*



Figura 9.3: *Ejemplos de pérdidas de subbandas en 1ª resolución, de izquierda a derecha y de arriba a abajo: 1-detalle horizontal, 2-vertical, 3-diagonal, 4-detalle horizontal y vertical, 5-detalle vertical y diagonal, 6-detalle horizontal y diagonal*



Figura 9.4: Ejemplos de pérdidas de una subbanda en 2ª resolución: (primera fila) pérdida de detalle vertical de segunda resolución y pérdida de detalle vertical, diagonal y horizontal y detalle vertical de la 1ª, (segunda fila) pérdida de detalle horizontal de segunda resolución y pérdida de detalle vertical, diagonal y horizontal respectivamente de la primera resolución, (tercera fila) pérdida de detalle diagonal de segunda resolución y pérdida de detalle vertical, diagonal y horizontal respectivamente de la primera resolución

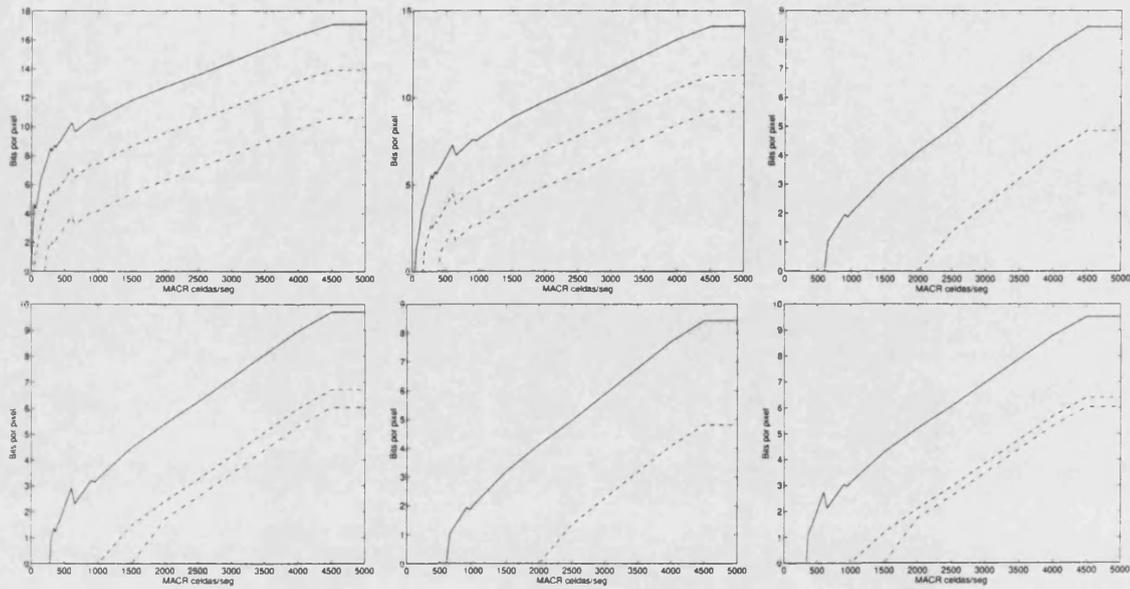


Figura 9.5:  $B_{pp}$  para cada una de las 22 subbandas vs MACR neto con Miss América en formato QSIF

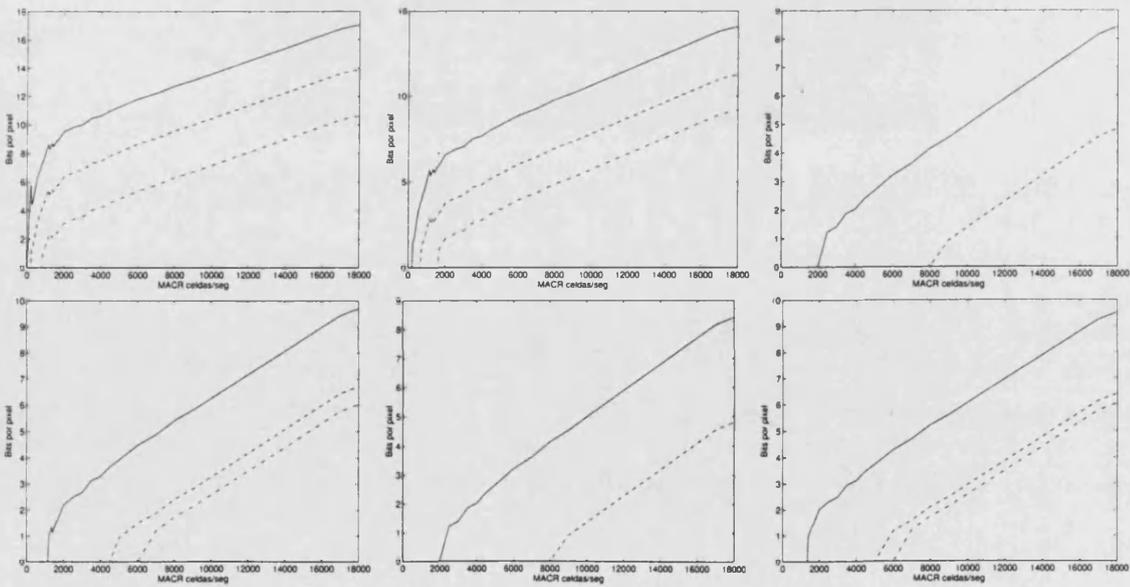


Figura 9.6:  $B_{pp}$  para cada una de las 22 subbandas vs MACR neto con Miss América en formato SIF

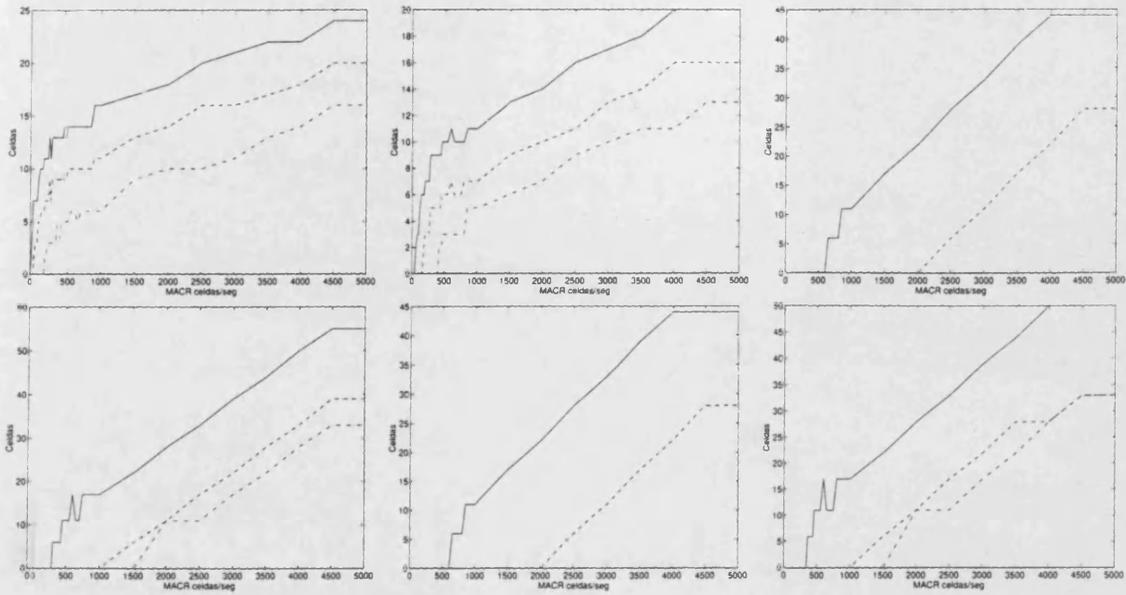


Figura 9.7: Celdas ATM para cada una de las 22 subbandas vs MACR neto con Miss América en formato QSIF

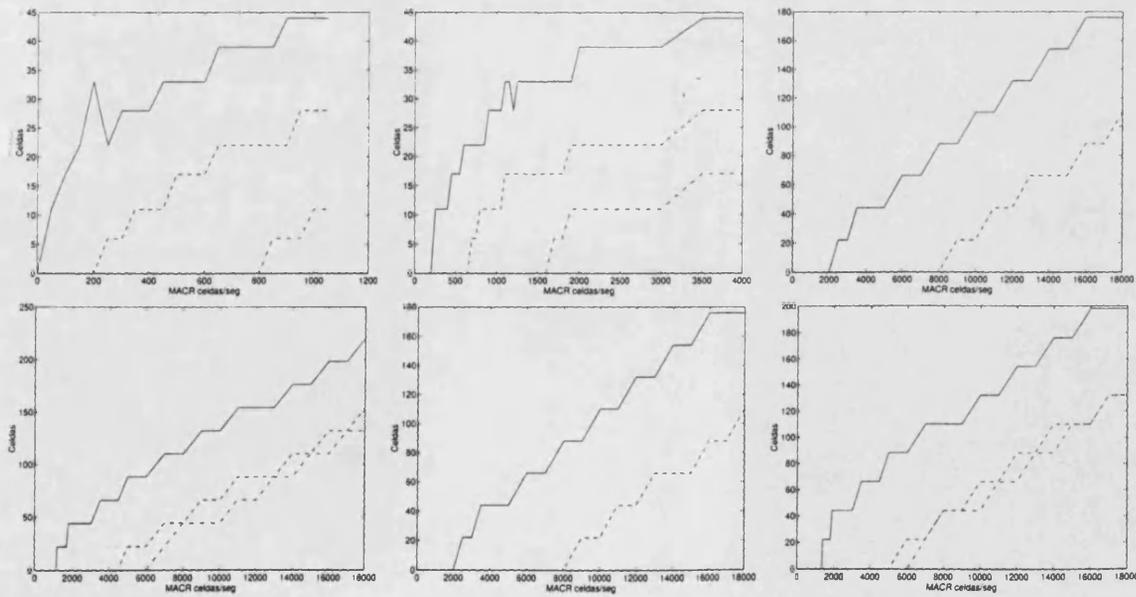


Figura 9.8: Celdas ATM para cada una de las 22 subbandas vs MACR neto con Miss América en formato SIF

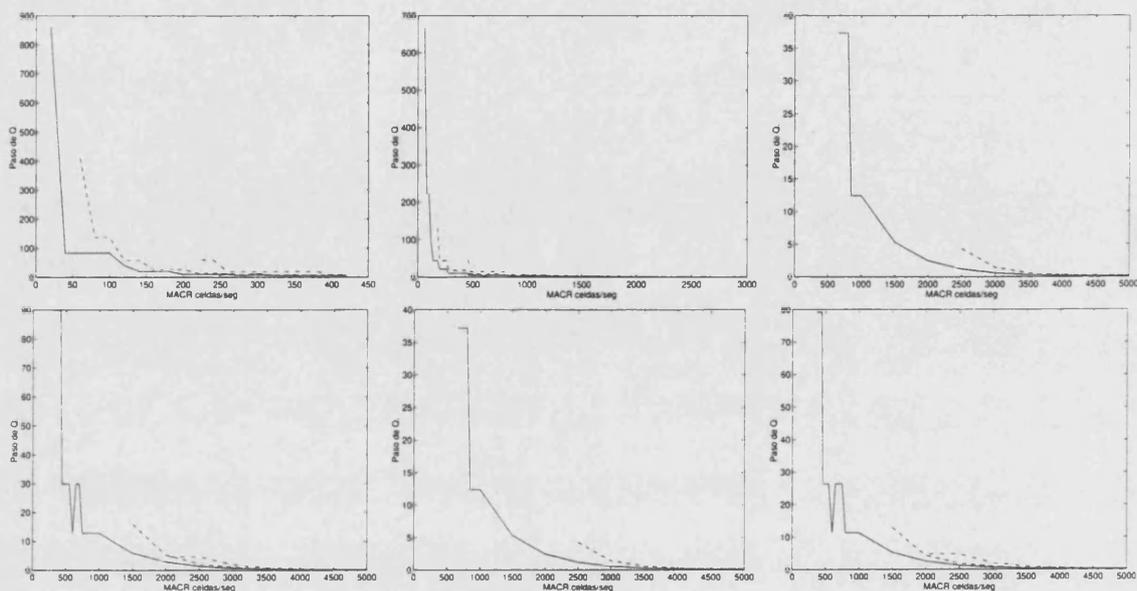


Figura 9.9: Paso de cuantificación para cada una de las 22 subbandas vs MACR neto con Miss América en formato QSIF

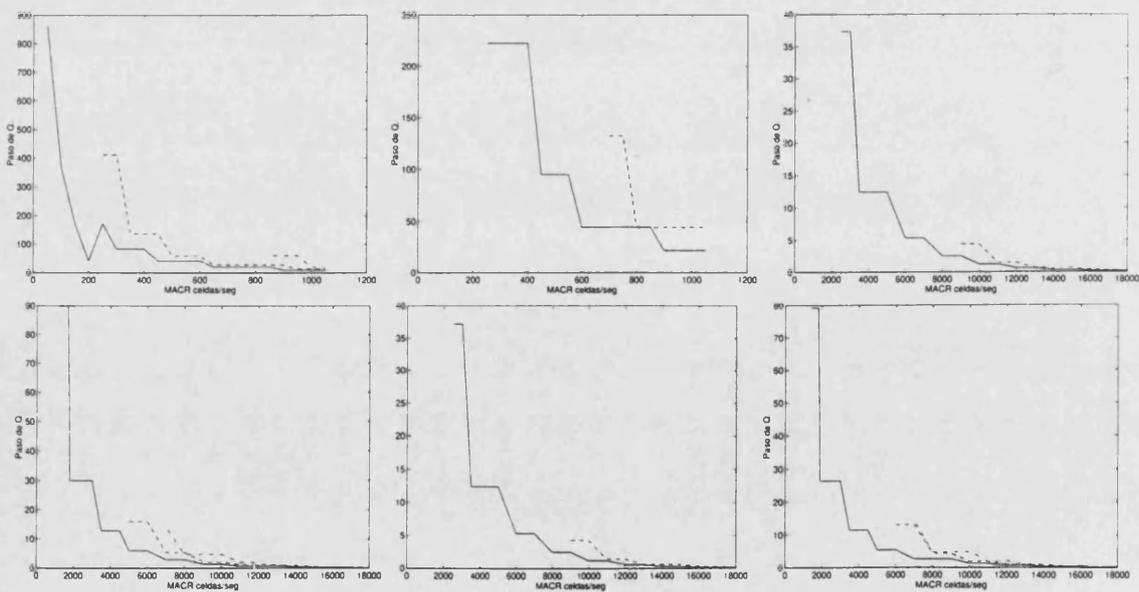


Figura 9.10: Paso de cuantificación para cada una de las 22 subbandas vs MACR neto con Miss América en formato SIF



Figura 9.11: 4 fotogramas de Miss América en formato SIF comprimida con 0.1278 *bpp*



Figura 9.12: 4 fotogramas de Miss América en formato SIF comprimida con 0.2424 *bpp*



Figura 9.13: 4 fotogramas de Miss América en formato SIF comprimida con 0.5303 *bpp*



Figura 9.14: 4 fotogramas de Miss América en formato SIF comprimida con 0.7575 *bpp*



Figura 9.15: 4 fotogramas de Miss América en formato SIF comprimida con 1.0606 *bpp*



Figura 9.16: 4 fotogramas de Miss América en formato SIF comprimida con 1.9696 *bpp*



Figura 9.17: 4 fotogramas de Miss América en formato SIF comprimida con 2.66 *bpp*

# Capítulo 10

## Resultados experimentales

En el presente capítulo se verifica el sistema mediante simulaciones sobre redes ATM con servicio ABR para la configuración de 2 conmutadores y diferentes escenarios de operación. Los módulos utilizados en el simulador tanto del sistema de emisión como de recepción se recogen en el anexo G.

El objetivo es observar el comportamiento del sistema (el patrón de tráfico que genera) y validar (acotar) las prestaciones en funcionamiento. Los resultados obtenidos son procesados estadísticamente para poder recoger con mayor objetividad dichas prestaciones.

Previo a ello se describe la elección de los parámetros utilizados en el simulador, tanto en la fuente como en el conmutador.

### 10.1 Descripción de los parámetros introducidos para las simulaciones

En esta sección se detallan los parámetros de configuración de las fuentes, del conmutador ABR, del tiempo de simulación y de los cambios de ancho de banda; dichos parámetros guardan relación con las pautas especificadas en [Kim96].

#### 10.1.1 Parámetros de las fuentes

Las fuentes ABR negocian con la red para establecer la conexión la tasa mínima de salida (MCR), inicial (ICR) y de pico (PCR). A continuación se detalla cada una de ellas.

Como se ha dicho en la sección 7.1 respecto a la fuente de vídeo, la negociación de la mínima tasa de salida (MCR) para una secuencia queda por defecto establecida con 8 bpp para la subbanda de mayor información (codificada en DPCM), lo que supone para el caso de un formato SIF una tasa de salida de 0.825 celdas/ms (316.8 Kbps) excluyendo las celdas RM; la tasa real de salida a negociar ha de incluir dichas celdas (cuyo número viene determinado por el parámetro  $N_{rm} = 32$  que indica cada cuantas celdas de datos sale una RM) por lo que  $MCR = 0.8516$  celdas/ms.

Para que el escenario sea más equitativo, la negociación de MCR para la fuente de vídeo se realiza de la misma forma para las demás fuentes. Con ello se evita el posible problema de que las fuentes con  $MCR=0$  (según establece [Kim96]) favorezcan a la de vídeo en momentos de congestión. Por tanto para el caso de 5 fuentes, la capacidad mínima del enlace troncal ha de ser de 4.258 celdas/ms (1.8 Mbps).

Por lo general para cada fuente, la tasa de inicio (ICR) es algo mayor que MCR y menor que la tasa pico (PCR que es fijada generalmente a la capacidad total del enlace a fin de que las fuentes cuando queden solas puedan aprovechar todos los recursos disponibles en la red).

Para el caso de 5 fuentes y viendo la tasa de salida de la fuente de vídeo según factores de compresión (con 18 celdas/ms aproximadamente en el mínimo factor de compresión) en el capítulo anterior, a fin de conseguir una utilización del enlace troncal cercana a 1 y que dicho enlace se comporte como un cuello de botella, se fija la capacidad del enlace a 10 Mbps (23.58 celdas/ms).

Como se ha dicho, la capacidad del enlace troncal determina el PCR, y por el control de flujo utilizado (CAPAC[Mar97]) el ICR se estima como la décima parte de un tanto por cien (conocido como *Target Rate* estimado en un 95%) del ancho de banda inicial, lo que supone 2.24 celdas/ms. Este valor inicial queda muy por debajo de la porción equitativa del ancho de banda asignado a cada fuente en el conmutador en régimen permanente y ausencia de perturbaciones (que se estima en 4.716 celdas/ms, la quinta parte de 10 Mbps) por lo que a través del bucle cerrado del control de flujo, el ACR irá creciendo hasta dicho valor.

Destacar que en el caso de la fuente de vídeo y según la inicialización realizada para el estimador MACR, para 160 ms independientemente de las transiciones que sufra la red en dicho intervalo, a ocho veces el MCR (ver sección 7.2.3), es decir 6.81 celdas/ms, en la configuración descrita de cuello de botella con 5 fuentes podrían transmitir 3.5 fuentes a la misma velocidad.

Los diferentes parámetros discutidos se resumen en la tabla 10.1.

MCR	0.851
ICR	2.24
PCR	23.58

Tabla 10.1: *Especificación de tasas de salida en celdas/ms para las fuentes ABR*

Los demás parámetros de la fuente ABR se recogen en [Mar97].

### 10.1.2 Parámetros del conmutador

Dentro del conmutador, es de vital importancia el intervalo de tiempo para la estimación del ancho de banda disponible para cada fuente. Debido a que las configuraciones habituales suelen referirse a accesos troncales de 150 Mbps, en este caso particular de 10 Mbps hay que modificar dicho parámetro a fin de que las variaciones del ancho de banda equitativo puedan seguirse con el ACR de las fuentes.

Las fuentes actualizan su ACR cada 32 celdas de datos (según el valor  $N_{rm}$ ) y con ello si la tasa equitativa de salida es de 4.716 celdas/ms, el periodo de modificación es de aproximadamente unos 6.78 ms.

El conmutador con control de flujo basado en el algoritmo CAPAC, estima la tasa equitativa por defecto para 150 Mbps bien a 100 celdas procesadas o bien cada ms [Mar97]. En éste caso particular de 10 Mbps vencería antes la estimación cada ms según los parámetros anteriores, lo que supone variaciones más rápidas de las que realmente la fuente puede seguir.

Por tanto, se modifica el intervalo de estimación bien a 100 celdas procesadas o bien cada 15 ms (según las pautas fijadas en [Mar97]), con lo que se fuerza a realizar una estimación (si el enlace troncal es de 23.58 celdas/ms) aproximadamente cada 4 ms.

Los demás parámetros del conmutador ABR se recogen en [Mar97].

### 10.1.3 Estudio del tráfico ráfaga

Como se ha visto en la sección 8.4 del capítulo 8, las fuentes ráfagas utilizadas siguen el modelo de [Ada95].

Estas fuentes son modelas por un fichero formado por parejas de longitud de ráfaga en celdas (ráfaga transmitida al ACR de la fuente) y retardo en ms (una vez terminada la transmisión de la ráfaga); dicho fichero es común para todas las fuentes, con la salvedad que su lectura en cada una comienza en puntos diferentes.

Para la creación de este fichero se ha supuesto el enlace troncal a 10 Mbps que se reparte entre 4 fuentes, excluyendo la fuente de vídeo. Por ser ráfaga he aproximado la simultaneidad media de éstas en un 75%, es decir que de las 4, sólo una por regla general no está activa, con lo cual permite asignar a las fuentes ráfaga activas una tasa de salida de 6.2893 celdas por ms. A su vez, la fuente cuando está activa, su tasa efectiva de salida no será inferior a 4.7169 celdas/ms.

Con las suposiciones anteriores y para el caso de un fichero formado por 100 parejas (retardo, longitud), se concluye (de forma simple) en 10 ráfagas de 2 celdas, 30 ráfagas de 16 celdas y 60 ráfagas de 32 celdas con retardos entre ellas en valor medio de 1.282 ms aproximadamente.

### 10.1.4 Duración de las simulaciones

Para un enlace troncal de 10 Mbps (23.58 celdas/ms), los transitorios producidos en las colas de los conmutadores son aproximadamente del orden de decenas de ms, por utilizar algoritmos de control de flujo reactivos a la congestión [Mar97]. Por tanto una duración de 1600 ms permite observar los diferentes comportamientos tanto en permanente como en transitorio.

Con 1600 ms de simulación, los ficheros traza de las secuencias de vídeo agrupan 10 unidades de descomposición (160 ms por unidad) y con lo cual se llegan a transmitir 40 fotogramas de cada secuencia.

### 10.1.5 Cambios de ancho de banda

El efecto de cambios de ancho de banda en el enlace troncal permite modelar el ancho de banda disponible sobrante a las conexiones CBR y VBR[Mar97].

El resultados que se pretende extraer con ello, es observar como a través de los algoritmos de control de flujo, las fuentes y en concreto la fuente de vídeo, se adapta a la nueva situación.

Los casos posibles a observar en la transmisión cuando se produzcan los cambios son de forma consecutiva:

1. cambio de ancho de banda una vez realizada la estimación de la fuente de vídeo para 160 ms; en éste caso la fuente no podrá reasignar los bits y si se trata de una disminución habrá pérdidas en las subbandas últimas transmitidas
2. adaptación de la fuente a la nueva situación
3. cambio de ancho de banda para que la fuente de vídeo trabaje cercana a la situación de mínima tasa de salida

Con los puntos anteriores los cambios de ancho de banda se producen de la siguiente forma en la simulación:

1. enlace troncal a 10 Mbps hasta 400 ms
2. desde 400 ms cambio a 5 Mbps hasta 900 ms
3. desde 900 ms cambio a 2 Mbps hasta el final

## 10.2 Estudio de retardos

En esta sección se pretende analizar los tiempos de retardo tanto a nivel de celda como a nivel de fotograma y las máximas variaciones de éstos según la ocupación de las colas en los conmutadores.

El retraso introducido por el medio físico para la transmisión de una celda se estima alrededor de 5 microsegundos por Km[Mar97].

Por tanto, el retraso introducido por la red entre la salida de una celda y la llegada, para el caso LAN es

$$(0.2 + 2 + 0.2)Km \ 5us/Km = 12us \quad (10.1)$$

y para el caso MAN, para una distancia de acceso intermedia (10 Km) es

$$(10 + 50 + 10)Km \ 5us/Km = 350us \quad (10.2)$$

En el caso de una celda RM con ida y vuelta por el circuito virtual, el tiempo anterior se multiplica por 2.

Respecto a las variaciones de retardo se observa en las simulaciones realizadas que en los casos de congestión, el número de celdas almacenadas en las colas son alrededor

de 200 aproximadamente. Con ello y suponiendo anchos de banda mínimos de hasta 2 Mbps (tasa mínima a la que por negociación de MCR puede llegar la capacidad del enlace troncal), el tiempo de vaciado y por tanto el retardo mayor introducido a las celdas últimas en la cola FIFO del conmutador es de 42.4 ms.

En el sistema propuesto y como se ve en la sección 7.3.1, existe la posibilidad de introducir un retardo para activar el receptor de vídeo llegada la primera celda. Dicho retardo se puede estimar para el caso peor visto y en configuración MAN, con una demora de 42.4 ms en la cola del conmutador más los 350 us por el medio físico.

Respecto al retardo total extremo a extremo de cada fotograma se puede estimar como la suma de los tiempos siguientes:

1. 80 ms hasta recibir los 2 primeros fotogramas en el emisor, con los cuales puede mandar 7 subbandas una vez procesadas (suponemos el tiempo de proceso de las subbandas nulo)
2. llegada la primera celda al receptor, que se puede suponer casi instantáneamente (350 us) porque el enlace está vacío, se activa el temporizador para reducir la variabilidad de las llegadas con el tiempo peor de 42.4 ms visto antes
3. vencido el tiempo anterior se arranca el sincronismo que permite aceptar celdas hasta 160 ms y con ello se cierra la ventana de los fotogramas actuales y se avanza a los siguientes (en el caso de que lleguen celdas posteriores, son descartadas por caducidad)

con lo que se obtiene un retraso de 282.75 ms; este tiempo queda dentro de los márgenes para poder ofrecer un servicio de vídeo interactivo.

Destacar que el tiempo de demora en el receptor llegada la primera celda puede ser mayor o menor al valor calculado.

Si es mayor, la cola de entrada al receptor quedaría rápidamente llena de celdas, aumentando el retraso total extremo a extremo, pero garantizando mayor número de subbandas correctas. Si es menor, por contra la cola de entrada queda casi vacía, el retraso total extremo a extremo disminuye, pero puede ocurrir en momentos de congestión que un gran número de subbandas lleguen excesivamente tarde.

Una estimación de un tiempo menor a 42.4 ms, puede ser en la suposición del enlace troncal a 10 Mbps con congestiones de unas 20 celdas en la cola del conmutador (umbral con el que el conmutador indica a las fuentes que existe congestión[Mar97]), lo cual supone un retraso de 1 ms aproximadamente.

### 10.3 Patrón de tráfico de la fuente de vídeo

Previo a las simulaciones de la siguiente sección, cabe analizar el patrón de tráfico de la fuente de vídeo. Para ello se analiza la ocupación del buffer de recepción vaciado a intervalos de 160 ms (tiempo necesitado para formar una unidad de descomposición) y que se muestra en la figura 10.1 para la secuencia de Miss América en formato QSIF

y SIF comprimida con 2.6 bpp (este factor de compresión se ha visto en el capítulo 9 fotogramas 9.17).

Las gráficas 10.1 han sido obtenidas observando el tamaño de la cola FIFO1 en el sistema de recepción (ver esquema G.3) para cada unidad de descomposición (160 ms). Se observa un patrón a intervalos de 160 ms como cabía esperar, donde los primeros 80 ms lo forman la transmisión de la primera parte del primer nivel de resolución (7 subbandas) y los 80 ms siguientes lo forma el segundo nivel de resolución (8 subbandas) seguido de la segunda parte del primer nivel (7 subbandas).

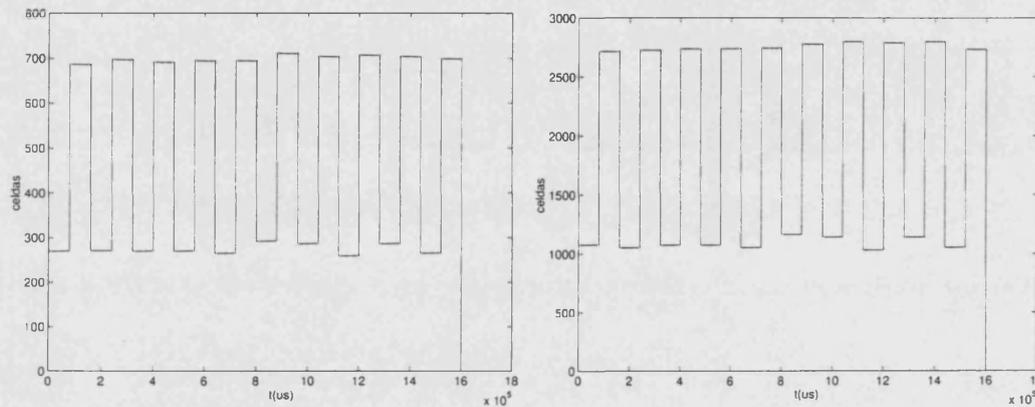


Figura 10.1: Evolución de la salida de celdas ATM para la secuencia de Miss América comprimida con 2.66 bpp para los formatos QoSIF y SIF

De forma similar para las secuencias de Claire y Salesman, y con el mismo factor de compresión se obtienen patrones casi idénticos; no son exactamente idénticos debido a los redondeos asociados a las etapas de asignación de bits y adaptación a AAL-5+.

## 10.4 Comportamiento del sistema bajo test

En esta sección se trata de analizar y evaluar el comportamiento del sistema propuesto sobre redes ATM ante escenarios descritos en el capítulo 8 (tabla 8.2) y modelados en el simulador como se recoge en el anexo G.

En las simulaciones realizadas se trabaja en:

- secuencia de *Miss América* en formato SIF (en el caso de utilizar las secuencias de *Salesman* o *Claire* se especifica)
- entornos LAN y MAN
- con tráfico de fondo tanto persistente como ráfaga
- con cambios de ancho de banda en el enlace troncal, que reflejan las variaciones asociadas al establecimiento de nuevas conexiones CBR y VBR
- con diferentes tamaños de cola del conmutador

y donde los grados de libertad negociados para la fuente antes de la conexión son la tasa mínima de salida (MCR) y el retardo de activación del receptor tras la llegada de la primera celda.

Los siguientes apartados recogen los resultados y comentarios de las diferentes simulaciones, utilizando en el simulador los parámetros calculados en la sección 10.1.

Dentro de cada apartado se estudian diferentes casos y además en aquellos que se adjuntan gráficas se utiliza una etiqueta *referencia X* para identificarlas.

Para cada caso se adjuntan dos tablas, una asociada a la estimación del MACR en los instantes de tiempo donde se realiza la descomposición (múltiplos de 160 ms) y otra asociada a las probabilidades tanto de transmisión como de recepción correcta para cada subbanda.

La transmisión o no de cada subbanda es una variable aleatoria que está relacionada con el estado de congestión de la red en el instante previo a la asignación de bits en la descomposición y las características de la propia secuencia de vídeo. Esta variable sigue una distribución de Bernuilli con probabilidad  $p(T)$  de transmisión o  $1 - p(T)$  de no transmisión[Jai92].

El estudio probabilístico de las transmisiones se calcula a partir de los casos favorables en que la subbanda vaya a ser transmitida (porque el ancho de banda de la red estimado lo permite) y los casos posibles, que son tantos como descomposiciones se realicen.

Cada subbanda a transmitir puede sufrir varios percances:

- que la red cambie de estado y no permita salir la subbanda de la cola de emisión, con lo cual en la siguiente descomposición si todavía permanece en la cola es reseteada y la subbanda es eliminada
- que salga demasiado tarde y cuando llegue al receptor esté fuera de sincronismo
- que pierda celdas o se vea dañada dentro de la red, en los conmutadores

con ello se concluye que las subbandas posibles a transmitir pueden llegar correctamente ( $p(A)$ ) o no ( $p(E)$ ), y que por el Teorema de la Probabilidad Total  $p(T) = p(AT) + p(ET)$ .

Sin embargo la probabilidad que más interesa para poder validar el sistema en su conjunto es la probabilidad condicionada de llegar correctamente cuando ha sido transmitida, es decir  $p(A/T) = p(AT)/p(T)$ . Esta probabilidad valida el sistema en los siguientes aspectos:

- si el sistema propuesto es suficientemente adaptativo (e inteligente para poder predecir el comportamiento de la red), estima las subbandas posibles a transmitir (variando el factor de compresión) según el ancho de banda disponible en la red de forma que todas las subbandas puedan llegar correctamente
- al asignar un orden de prioridades en las subbandas perceptualmente, aquellas más importantes han de recibirse correctamente con mayor probabilidad

Para poder tener un espacio muestral mayor, dentro de las diferentes configuraciones se han introducido 10 alteraciones para que los resultados de las simulaciones varíen. Estas alteraciones consisten que en el instante inicial las fuentes (excepto la de vídeo) son retrasadas apropiadamente en el margen de 0 a 10 ms.

Destacar además que para el cálculo de probabilidades de transmisión ha sido descartada la primera descomposición realizada porque la fuente de vídeo realiza la estimación sin tener conocimiento del estado de la red; en las restantes descomposiciones el MACR sí que está actualizado por el ACR. El valor del MACR en el instante inicial ha sido discutido en la sección 7.2.3 del capítulo 7.

### 10.4.1 Multiplexación con tráfico persistente. Configuración LAN

Los diferentes casos simulados en este apartado son:

1. (*referencia 1*) enlace troncal con cambios de ancho de banda, sin pérdidas en el conmutador y con un retardo inicial de 42 ms en el receptor. La figura 10.10 muestra de izquierda a derecha y de arriba a abajo:

- ACR (línea punteada) y MACR (línea sólida) de la fuente de vídeo (VFuente es la notación del sistema de emisión como se ve en el anexo G) donde se observan las variaciones en los cambios de ancho de banda en 400 ms y 900 ms
- ACR de las demás fuentes, donde se observa que todas las fuentes siguen las mismas pautas del control de flujo ABR
- la utilización del enlace troncal, donde se observa que tras el cambio a 5 Mbps la utilización baja casi al 35% debido a la respuesta agresiva del control de flujo ante la congestión quedando la cola del conmutador vacía y en el cambio a 2 Mbps ocurre de forma similar pero en este caso la velocidad de vaciado de la cola del conmutador es mucho más lenta con lo cual la utilización queda estable
- la cola del conmutador, con mayor ocupación en los momentos de cambio de ancho de banda por el retraso que conlleva la actualización del ACR en las fuentes
- la cola de la fuente de vídeo, que va llenándose a intervalos de 80 ms y vaciándose al ACR que determina la pendiente de la curva de ocupación
- la cola del receptor, que acumula celdas para eliminar la variabilidad de los retardos con la cota de 42 ms

La figura 10.2 recoge 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación.

2. misma configuración anterior pero con un retardo inicial de 1 ms; en este caso las gráficas son iguales al punto anterior, a excepción de la ocupación de la cola

del receptor que está vacía (no llegan a almacenarse celdas porque la memoria del receptor se vacía casi instantáneamente).

3. (*referencia 2*) misma configuración anterior pero con pérdidas, limitando el tamaño de la cola del conmutador a 100 celdas. La figura 10.11 muestra en su parte izquierda la ocupación de la cola del conmutador donde se observa la limitación del tamaño (la acumulación de celdas alrededor de 400 ms y 900 ms es debida al cambio de ancho de banda) y en su parte derecha la cola del receptor, que como no se le introduce un retardo inicial apreciable permanece casi vacía.

La figura 10.3 recoge 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación.

4. (*referencia 3*) misma configuración anterior pero negociando un MCR de la fuente de vídeo tres veces mayor que el calculado en la sección 10.1. La figura 10.12 muestra de izquierda a derecha y de arriba a abajo:

- ACR (línea punteada) y MACR (línea sólida) de la fuente de vídeo, donde al inicializar la estimación de MACR con  $8 \cdot \text{MCR}$  y al ser el  $\text{MCR} = 2.513$  celdas/ms, se observa que la estimación inicial es alrededor de 21 celdas/ms (evidentemente con esta estimación hay pérdidas)
- la cola de la fuente de vídeo, que de la misma manera, al sobredimensionar la estimación del MACR hace que en los primeros 160 ms la ocupación de la cola sea alta (mayor de 1000 celdas)
- la utilización del enlace, cercana a 1 especialmente en el cambio a 2 Mbps porque la cola del conmutador está llena, debido a que la fuente de vídeo mantiene valores altos de ACR y por otro lado la capacidad del enlace por el salto de ancho de banda no permite vaciarla con rapidez
- la cola del conmutador, que cuando el enlace troncal cambia a 2 Mbps entra en una situación de alta ocupación de forma continua

La figura 10.4 recoge 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación.

Cabe destacar con carácter general de las gráficas anteriores:

- los transitorios y picos que puede haber en el ACR son debidos a que la cola de la fuente de vídeo se queda vacía
- la utilización del enlace troncal está alrededor del 95%, pasando al 100% en los momentos de gran ocupación en la cola del conmutador y quedando por debajo (alrededor del 50%) cuando la cola está casi vacía y además las fuentes emiten con baja tasa de salida (por lo general después de una congestión)
- respecto a la *referencia 3*, negociar un MCR que esté por encima del valor estudiado para la conexión y en situaciones con factores de utilización altos,



Figura 10.2: Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes persistentes: referencia 1



Figura 10.3: Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes persistentes: referencia 2



Figura 10.4: Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes persistentes: referencia 3

puede hacer que las celdas de la fuente de vídeo o se pierdan en las colas o se retrasen excesivamente, concluyendo que es mejor delegar en los servicios *best effort* de la red que negociar una tasa mínima de salida alta

Las estimaciones del MACR (celdas/ms) en los instantes de la descomposición con su media y varianza (la varianza da un orden de magnitud del comportamiento variable de la red) se recogen en las tablas 10.2 para los diferentes casos. La probabilidades  $p(T)$  y  $p(A/T)$  se recogen para cada subbanda y para los diferentes casos en la tabla 10.3.

tiempo(ms) vs caso	caso1	caso2	caso3	caso4
0	6.8080	6.8080	6.8080	20.1000
160	4.4935	4.4935	4.4935	4.4953
320	4.4976	4.4976	4.4977	4.4964
480	1.5729	1.5729	1.5191	2.5131
640	2.2524	2.2524	2.2509	2.5131
800	2.2172	2.2172	2.2270	2.5131
960	1.1901	1.1901	1.0120	2.5131
1120	0.8510	0.8510	0.8510	2.5131
1280	0.8510	0.8510	0.8510	2.5131
1440ms	0.8606	0.8606	0.8510	2.5131
media	2.0874	2.0874	2.0615	2.9537
desviación	1.4695	1.4695	1.4876	0.8743

Tabla 10.2: *Estimaciones de MACR (celdas/ms) en los instantes de la descomposición: diferentes casos con escenario LAN y fuentes persistentes*

#### 10.4.2 Multiplexación con tráfico persistente. Configuración MAN

Los diferentes casos simulados en este apartado son:

1. (*referencia 1*) enlace troncal sin cambios de ancho de banda, sin pérdidas en el conmutador y con un retardo inicial de 42 ms en el receptor. La figura 10.13 muestra de izquierda a derecha y de arriba a abajo:
  - ACR (línea punteada) y MACR (línea sólida) de la fuente de vídeo, donde se observa claramente los aumentos de ACR en los momentos que la cola de la fuente de vídeo queda vacía (para el conmutador hay una fuente menos) y también cuando la fuente de vídeo deja de transmitir en 1600 ms
  - la cola del conmutador, como no hay cambios de ancho de banda tiene un máximo de 28 celdas y una ocupación casi constante de 3 celdas, mayor que en el escenario LAN porque las fuentes tardan más en reaccionar

subbanda	caso 1		caso 2		caso 3		caso 4	
	p(T)	p(A/T)	p(T)	p(A/T)	p(T)	p(A/T)	p(T)	p(A/T)
pb2rvh	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
pb2rvH	1.00	1.00	1.00	0.87	1.00	0.87	1.00	0.92
pb2rVh	1.00	0.93	1.00	0.75	1.00	0.68	1.00	1.00
pb2rVH	0.85	0.83	0.88	0.48	0.90	0.53	1.00	0.91
pa2rvh	1.00	0.62	1.00	0.58	1.00	0.67	1.00	0.76
pa2rvH	1.00	0.60	1.00	0.44	1.00	0.56	1.00	0.21
pa2rVh	0.93	0.55	0.96	0.47	0.95	0.47	1.00	0.26
pa2rVH	0.69	0.84	0.69	0.66	0.61	0.56	1.00	1.00
pb1rt1vh	0.45	1.00	0.44	1.00	0.45	1.00	0.97	0.89
pb1rt1Vh	0.45	0.84	0.47	0.76	0.50	0.95	0.87	0.93
pb1rt1VH	0.06	0.58	0.09	0.55	0.02	0.55	0.22	0.50
palrt1vh	0.78	0.37	0.77	0.33	0.74	0.35	0.78	0.38
palrt1vH	0.38	0.54	0.38	0.52	0.33	0.59	0.37	0.23
palrt1Vh	0.55	0.50	0.17	0.31	0.15	0.22	0.28	0.14
palrt1VH	0.20	0.86	0.17	0.31	0.19	0.24	0.11	0.18
pb1rt2vH	0.66	0.65	0.60	0.82	0.64	0.90	1.00	0.94
pb1rt2Vh	0.34	0.71	0.39	0.77	0.35	0.76	0.88	0.72
pb1rt2VH	0.08	0.35	0.05	0.33	0.08	0.37	0.16	0.38
palrt2vh	0.57	0.14	0.55	0.21	0.54	0.27	0.22	0.27
palrt2vH	0.37	0.52	0.31	0.55	0.32	0.55	0.12	0.55
palrt2Vh	0.30	0.34	0.27	0.35	0.30	0.30	0.21	0.13
palrt2VH	0.20	0.36	0.15	0.35	0.12	0.21	0.10	0.17

Tabla 10.3: Probabilidades de subbandas transmitidas  $p(T)$  y recibidas correctamente condicionadas a ser transmitidas  $p(A/T)$ , para diferentes casos con escenario LAN y fuentes persistentes

- la cola de la fuente de vídeo, como no hay cambios de ancho de banda y las fuentes son persistentes, hace que el ACR se mantenga estable lo que permite vaciar la cola de forma constante
- la cola del receptor, que se va llenando cerca de los múltiplos de 160 ms por los motivos del sincronismo con un desfase de tiempo inicial de 42 ms

La figura 10.5 recoge 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación.

2. (referencia 2) misma configuración anterior pero con cambios de ancho de banda. La figura 10.14 muestra de izquierda a derecha y de arriba abajo:

- ACR (línea punteada) y MACR (línea sólida) de la fuente de vídeo, donde se observa las modificaciones producidas sobre ACR por el control de flujo debido al cambio de ancho de banda en 400 y 900 ms; con 2 Mbps el ACR casi no varía respecto a MCR

- ACR de las demás fuentes y la PABE (Porción de Ancho de Banda Equitativo, línea sólida), donde se observa que cuando la PABE queda por debajo del MCR, las fuentes mantienen el  $ACR=MCR$  llenando la cola del conmutador
  - la cola de la fuente de vídeo, que varía según el ACR y por tanto cuando el enlace troncal es de 2 Mbps (a partir de 900 ms) y  $ACR=MCR$  se vacía de forma constante
  - la cola del conmutador, con un ocupación máxima de 153 celdas y por tanto en el caso de limitarla a 120 celdas se producirían pérdidas
3. misma configuración anterior pero con un retardo inicial de 2 ms; en este caso las gráficas son iguales al punto anterior, a excepción de la ocupación de la cola del receptor está vacía
  4. misma configuración anterior pero con un retraso de 42 ms y con pérdidas en el conmutador, limitando el tamaño de la cola del conmutador a 120 celdas
  5. (*referencia 3*) misma configuración anterior pero negociando un MCR de la fuente de vídeo tres veces mayor ( $MCR=2.513$  celdas/ms) que el calculado en la sección 10.1. La figura 10.15 muestra de izquierda a derecha:
    - ACR (línea punteada) y MACR (línea sólida) de la fuente de vídeo, donde se observa el alto valor inicial de  $MACR=8*MCR$  y que tanto MACR como ACR son casi constantes a MCR cuando el enlace troncal es de 2 Mbps
    - la cola del conmutador, que como se ha comentado en el escenario LAN *referencia 3* con fuentes persistentes, el aumento del MCR calculado para la fuente de vídeo satura la cola del conmutador produciéndose pérdidas por desbordamiento

La figura 10.6 recoge 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación.

Comentar que el escenario MAN mantiene por lo general una cola de conmutador mayor que el escenario LAN, debido a que las modificaciones del ACR no son tan inmediatas. Destacar también que la modificación del MCR negociado perjudica más por las pérdidas que por lo que se puede ganar en tasa de salida.

Las estimaciones del MACR (celdas/ms) con su media y varianza se recogen en las tablas 10.4 para los diferentes casos. La probabilidades  $p(T)$  y  $p(A/T)$  se recogen para cada subbanda y para los diferentes casos en la tabla 10.5.

### 10.4.3 Multiplexación con tráfico ráfaga. Configuración LAN

Los diferentes casos simulados en este apartado son:

tiempo(ms) vs caso	caso1	caso2	caso3	caso4	caso5
0	6.8080	6.8080	6.8080	6.8080	20.1000
160	4.5000	4.5000	4.5000	4.5000	4.4953
320	4.5113	4.5113	4.5113	4.5113	4.4964
480	4.4855	1.5327	1.5327	1.5201	2.5131
640	4.4880	2.2672	2.2672	2.2718	2.5131
800	4.4913	2.2123	2.2123	2.2047	2.5131
960	4.4892	1.0110	1.0110	1.0210	2.5131
1120	4.4864	0.8510	0.8510	0.8510	2.5131
1280	4.4985	0.8510	0.8510	0.8510	2.5131
1440	4.4976	0.8510	0.8510	0.8510	2.5131
media	4.4942	2.0653	2.0653	2.0647	2.9537
desviación	0.0084	1.4913	1.4913	1.4910	0.8743

Tabla 10.4: *Estimaciones de MACR (celdas/ms) en los instantes de la descomposición: diferentes casos con escenario MAN y fuentes persistentes*

1. (*referencia 1*) enlace troncal sin cambios de ancho de banda, sin pérdidas en el conmutador y con un retraso inicial de 42 ms en el receptor. La figura 10.16 muestra de izquierda a derecha y de arriba abajo:
  - ACR (línea punteada) y MACR (línea sólida) de la fuente de vídeo, donde se observa las bruscas transiciones debido a la aleatoriedad del tráfico ráfaga
  - la ocupación de la cola en la fuente de vídeo, donde se observan mayores fluctuaciones que en el caso persistente, y en ocasiones al ser la estimación del comportamiento de la red menos acertada produce situaciones tanto de alta como nula ocupación
  - la utilización del enlace, que presenta transitorios muy rápidos manteniendo una media del 85%
  - la cola del conmutador, donde se observan transiciones espúreas con una media de ocupación alrededor de 9 celdas, mayor que los casos persistentes

La figura 10.7 recoge 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación.

2. (*referencia 2*) misma configuración anterior pero con enlace troncal con cambios de ancho de banda. La figura 10.17 muestra a su derecha el ACR (línea punteada) y MACR (línea sólida) de la fuente de vídeo, donde se observan tres escalones para 10, 5 y 2 Mbps (anchos de banda del enlace) a 5.3, 2.2 y 0.85 celdas/ms. En la parte izquierda de la figura se observa la cola del conmutador donde destaca las acumulaciones de celdas justo en los cambios de ancho de banda, con ocupaciones mayores de 120 celdas
3. (*referencia 3*) misma configuración anterior pero negociando un MCR de la fuente de vídeo tres veces mayor que el calculado en la sección 10.1 y con pérdidas en



Figura 10.5: Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario MAN con fuentes persistentes: referencia 1

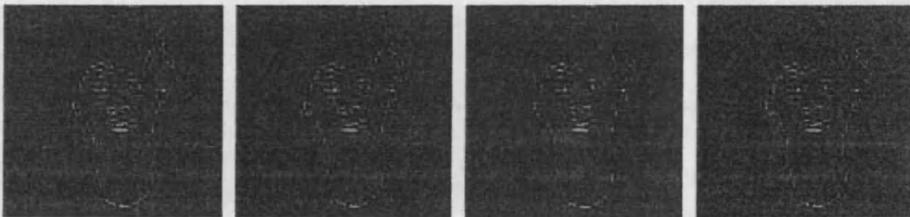


Figura 10.6: Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario MAN con fuentes persistentes: referencia 3



Figura 10.7: Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes ráfaga: referencia 1

subbanda	caso 1		caso 2		caso 3		caso 4		caso 5	
	p(T)	p(A/T)								
pb2rvh	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90
pb2rvH	1.00	1.00	1.00	1.00	1.00	0.76	1.00	0.83	1.00	0.85
pb2rVh	1.00	1.00	1.00	0.87	1.00	0.59	1.00	0.71	1.00	0.74
pb2rVH	1.00	1.00	0.81	0.82	0.82	0.50	0.83	0.51	1.00	0.82
pa2rvh	1.00	1.00	1.00	0.58	1.00	0.46	1.00	0.55	1.00	0.71
pa2rvH	1.00	1.00	1.00	0.50	1.00	0.37	1.00	0.32	1.00	0.64
pa2rVh	1.00	1.00	0.97	0.52	0.98	0.45	0.95	0.45	1.00	0.45
pa2rVH	1.00	1.00	0.68	0.52	0.66	0.31	0.67	0.51	1.00	0.43
pb1rt1vH	1.00	1.00	0.47	0.56	0.47	1.00	0.47	1.00	0.70	0.71
pb1rt1Vh	1.00	1.00	0.48	1.00	0.43	0.80	0.49	0.79	0.80	0.67
pb1rt1VH	0.02	0.54	0.07	0.81	0.01	0.58	0.05	0.54	0.12	0.28
palrt1vh	1.00	0.98	0.66	0.54	0.75	0.31	0.79	0.53	1.00	0.55
palrt1vH	0.72	0.51	0.31	0.82	0.33	0.24	0.33	0.27	0.37	0.47
palrt1Vh	0.20	0.50	0.16	0.42	0.26	0.50	0.15	0.46	0.15	0.29
palrt1VH	0.12	0.10	0.12	0.17	0.15	0.65	0.15	0.29	0.18	0.22
pb1rt2vH	1.00	1.00	0.63	1.00	0.64	1.00	0.57	0.78	0.90	0.77
pb1rt2Vh	1.00	0.85	0.37	0.87	0.33	0.84	0.34	0.38	0.62	0.53
pb1rt2VH	0.14	0.38	0.01	0.62	0.24	0.67	0.16	0.17	0.17	0.14
palrt2vh	0.76	0.19	0.58	0.10	0.24	0.52	0.50	0.54	0.87	0.51
palrt2vH	0.82	0.53	0.35	0.57	0.34	0.22	0.42	0.29	0.74	0.25
palrt2Vh	0.28	0.24	0.26	0.21	0.35	0.24	0.37	0.26	0.26	0.22
palrt2VH	0.12	0.18	0.16	0.19	0.14	0.19	0.13	0.14	0.10	0.18

Tabla 10.5: Probabilidades de subbandas transmitidas  $p(T)$  y recibidas correctamente condicionadas a ser transmitidas  $p(A/T)$ , para diferentes casos con escenario MAN y fuentes persistentes

el conmutador, limitando el tamaño de la cola del conmutador a 120 celdas. La figura 10.16 muestra a su izquierda la cola del conmutador, donde como en casos anteriores al tener un ACR mayor que la PABE en los momentos de congestión hace que cuando el enlace troncal es de 2 Mbps la celdas se retrasen excesivamente. La parte derecha muestra la ocupación de la cola del receptor que en el momento de congestión a partir de 900 ms queda vacía porque la llegada de las celdas es escasa

La figura 10.8 recoge 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación.

4. misma configuración anterior pero con las secuencias de *Salesman* y *Claire*. Las gráficas obtenidas para esta secuencia son similares al caso anterior de Miss América

Las estimaciones del MACR (celdas/ms) con su media y varianza se recogen en las tablas 10.6 para los diferentes casos. La probabilidades  $p(T)$  y  $p(A/T)$  se recogen para cada subbanda y para los diferentes casos en la tabla 10.7.

tiempo(ms) vs caso	caso1	caso2	caso3	caso4	caso5
0	6.8080	6.8080	20.1000	20.1000	20.1000
160	4.9100	4.9100	4.9953	4.9953	4.9953
320	5.3611	5.3691	5.2390	5.2390	5.2390
480	5.3250	1.6731	2.5131	2.5131	2.5131
640	5.6788	2.4322	2.5131	2.5131	2.5131
800	5.6813	2.4387	2.5131	2.5478	2.5131
960	5.5382	1.0560	2.5131	2.5131	2.5131
1120	5.5364	0.8510	2.5131	2.5131	2.5131
1280	5.3580	0.8510	2.5131	2.5131	2.5131
1440	5.6376	0.8510	2.5131	2.5131	2.5131
media	5.4474	2.2702	3.0918	3.0956	3.0918
desviación	0.2449	1.7500	1.1499	1.1478	1.1499

Tabla 10.6: Estimaciones de MACR (celdas/ms) en los instantes de la descomposición: diferentes casos con escenario LAN y fuentes ráfaga

#### 10.4.4 Multiplexación con tráfico ráfaga. Configuración MAN

Los diferentes casos simulados en este apartado son:

1. (*referencia 1*) enlace troncal sin cambios de ancho de banda, sin pérdidas en el conmutador y con un retraso inicial de 42 ms en el receptor. La figura 10.19 muestra a su izquierda la ocupación de la cola del conmutador y a su derecha la ocupación de la cola del receptor. Se observa que las cotas máximas de ocupación en la cola del conmutador son similares al escenario LAN (*referencia 1*) salvo la media de ocupación es un poco más elevada (alrededor de 13 celdas), mientras la ocupación de la cola del receptor es similar. Destacar que las gráficas del ACR de las fuentes son similares al escenario LAN con tráfico ráfaga
2. misma configuración anterior pero con cambios de ancho de banda. Las gráficas son similares al escenario LAN (*referencia 2*). La figura 10.9 recoge 4 fotografías de la secuencia de Miss América en el instante 960 ms de la simulación.
3. (*referencia 2*) misma configuración anterior pero negociando un MCR de la fuente de vídeo tres veces mayor que el calculado en la sección 10.1 y pérdidas en el conmutador, limitando el tamaño de la cola del conmutador a 120 celdas. La figura 10.20 muestra a su izquierda el ACR y PABE (línea sólida) de las fuentes y a su derecha la ocupación de la cola del conmutador. Destacar que a diferencia del escenario LAN (*referencia 3*) la cola está menos ocupada debido a que las

subbanda	caso 1		caso 2		caso 3		caso 4		caso 5	
	p(T)	p(A/T)								
pb2rvh	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00	1.00	1.00
pb2rvH	1.00	1.00	1.00	1.00	1.00	0.95	1.00	1.00	1.00	1.00
pb2rVh	1.00	1.00	1.00	0.94	1.00	0.97	1.00	1.00	1.00	1.00
pb2rVH	1.00	1.00	0.88	0.96	1.00	0.87	1.00	1.00	1.00	1.00
pa2rvh	1.00	1.00	1.00	0.82	1.00	0.74	1.00	1.00	0.99	1.00
pa2rvH	1.00	1.00	1.00	0.54	1.00	0.86	1.00	1.00	1.00	1.00
pa2rVh	1.00	1.00	0.99	0.77	1.00	0.87	1.00	1.00	1.00	1.00
pa2rVH	1.00	1.00	0.62	0.88	1.00	0.77	0.92	0.92	0.96	0.75
pblrt1vH	1.00	1.00	0.52	1.00	0.82	1.00	0.70	0.77	0.82	0.82
pblrt1Vh	1.00	1.00	0.51	1.00	0.75	0.46	0.23	0.53	0.30	0.77
pblrt1VH	0.12	0.05	0.17	0.09	0.24	0.89	0.81	0.43	0.17	0.46
palrt1vh	1.00	1.00	0.97	0.58	0.84	0.68	0.91	0.96	0.98	0.18
palrt1vH	1.00	1.00	0.35	0.56	0.40	0.54	0.15	0.44	0.17	0.44
palrt1Vh	0.13	0.96	0.13	0.99	0.20	0.11	0.20	0.36	0.13	0.34
palrt1VH	0.13	0.97	0.17	0.39	0.19	0.15	0.26	0.44	0.26	0.44
pblrt2vH	1.00	1.00	1.00	0.85	0.99	0.65	0.89	0.74	0.84	0.73
pblrt2Vh	1.00	1.00	0.52	1.00	0.53	0.55	0.55	0.64	0.69	0.67
pblrt2VH	0.08	0.12	0.42	0.24	1.00	0.55	0.37	0.46	0.34	0.41
palrt2vh	1.00	0.97	0.14	0.64	0.20	0.67	0.37	0.94	0.35	0.93
palrt2vH	0.93	0.77	0.35	0.57	0.36	0.29	0.27	0.21	0.28	0.27
palrt2Vh	0.28	0.59	0.29	0.51	0.25	0.23	0.20	0.27	0.29	0.29
palrt2VH	0.36	0.31	0.20	0.31	0.19	0.21	0.12	0.40	0.16	0.30

Tabla 10.7: Probabilidades de subbandas transmitidas  $p(T)$  y recibidas correctamente condicionadas a ser transmitidas  $p(A/T)$ , para diferentes casos con escenario LAN y fuentes ráfaga

fuentes más cercanas al conmutador asumen de forma más rápida la congestión producida por la fuente de vídeo, mientras que en el escenario LAN las distancias al ser iguales impedía este comportamiento de las fuentes

4. misma configuración anterior pero con las secuencias de *Salesman* y *Claire*. La figura 10.21 muestra la ocupación de la cola del conmutador tanto para la secuencia de *Salesman* (izquierda) como *Claire* (derecha). Los comportamientos de la cola siguen las mismas pautas comentadas en el punto anterior

Las estimaciones del MACR (celdas/ms) con su media y varianza se recogen en las tablas 10.8 para los diferentes casos. La probabilidades  $p(T)$  y  $p(A/T)$  se recogen para cada subbanda y para los diferentes casos en la tabla 10.9.

tiempo(ms) vs caso	caso1	caso2	caso3	caso4	caso5
0	6.8080	6.8080	20.1000	20.1000	20.1000
160	4.7800	4.7880	4.9953	4.9953	4.9953
320	5.1611	5.1612	5.0180	5.0180	5.0180
480	5.2800	1.7579	2.5131	2.5131	2.5131
640	5.5643	2.4420	2.5131	2.5131	2.5131
800	5.5968	2.4767	2.5131	2.5131	2.5131
960	5.5382	1.0450	2.5131	2.5131	2.5131
1120	5.8354	0.9432	2.5131	2.5131	2.5131
1280	5.2876	0.9427	2.5131	2.5131	2.5131
1440	5.6434	0.9300	2.5131	2.5131	2.5131
media	5.4096	2.2763	3.0672	3.0672	3.0672
desviación	0.3165	1.6527	1.0996	1.0996	1.0996

Tabla 10.8: *Estimaciones de MACR (celdas/ms) en los instantes de la descomposición: diferentes casos con escenario MAN y fuentes ráfaga*

## 10.5 Estadísticas de los resultados obtenidos

El objetivo de esta sección es acotar las estadísticas obtenidas en el apartado anterior, fijando el intervalo de confianza de las estimaciones realizadas.

El estimador más interesante del sistema es  $p(A/T)$  para cada subbanda, el cual ha sido estudiado en entornos restrictivos de red como se ha visto en los apartados anteriores. Las muestras seleccionadas son por tanto un subconjunto del espacio total de características de fuerte restricción.

Sin embargo como el número de muestras obtenido no es infinito, hay que calcular el intervalo de confianza del promediado de  $\hat{p}(A/T)$ , cuyo significado es la probabilidad media promediada para los diferentes casos de los apartados anteriores y que evalúa la probabilidad de que una subbanda llegue correctamente cuando el sistema suponía que podía llegar (por eso se pensaba en transmitirla).

El estudio del intervalo de confianza asociado al promediado de los diferentes  $\hat{p}(A/T)$  viene determinado por el número de muestras (en este caso 190 evaluaciones: 40, 50, 50 y 50 respectivamente en los apartados anteriores) y de la varianza de los valores a promediar. La expresión analítica para un caso general de promediado  $\hat{p}$  es

$$\left[ \hat{p} - N_{Z_{\alpha/2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{N_t}, \hat{p} + N_{Z_{\alpha/2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{N_t} \right] \quad (10.3)$$

siendo  $N_t$  el número de repeticiones del experimento, en este caso el número de transmisiones realizadas para la subbanda promediada con  $\hat{p}$  [Run96].

El intervalo de confianza determina en este caso el margen tanto por arriba como por abajo donde la probabilidad obtenida por el promediado puede ser válida en el  $1 - \alpha$  por cien de los experimentos que se realicen; o lo que es lo mismo, que dado un nuevo experimento sólo en un  $\alpha$  por cien el valor obtenido quedaría fuera del intervalo de confianza.

subbanda	caso 1		caso 2		caso 3		caso 4		caso 5	
	p(T)	p(A/T)								
pb2rvh	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
pb2rvH	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
pb2rVh	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00
pb2rVH	1.00	1.00	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00
pa2rvh	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.92	1.00
pa2rvH	1.00	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00
pa2rVh	1.00	1.00	1.00	0.85	1.00	1.00	1.00	1.00	1.00	1.00
pa2rVH	1.00	1.00	0.76	0.70	1.00	1.00	0.91	0.95	0.91	0.72
pb1rt1vH	0.97	1.00	0.76	1.00	1.00	1.00	1.00	0.77	0.85	0.89
pb1rt1Vh	0.97	1.00	0.56	1.00	0.59	0.51	0.71	0.53	0.46	0.72
pb1rt1VH	0.11	0.01	0.19	0.02	0.37	0.07	0.34	0.49	0.29	0.42
palrt1vh	0.95	1.00	0.88	0.86	0.48	0.78	0.84	0.74	0.34	0.28
palrt1vH	0.86	1.00	0.45	1.00	0.25	0.66	0.43	0.69	0.85	0.56
palrt1Vh	0.19	0.45	0.17	0.84	0.14	0.38	0.38	0.34	0.57	0.37
palrt1VH	0.25	0.44	0.17	0.99	0.11	0.41	0.23	0.51	0.23	0.49
pb1rt2vH	0.96	1.00	0.87	1.00	0.99	0.90	0.93	0.91	0.91	0.87
pb1rt2Vh	0.91	1.00	0.44	0.92	0.73	0.79	0.71	0.74	0.78	0.74
pb1rt2VH	0.24	0.07	0.15	0.04	0.17	0.22	0.56	0.28	0.59	0.49
palrt2vh	0.32	0.90	0.45	0.99	0.50	0.80	0.45	0.50	0.49	0.97
palrt2vH	0.75	0.78	0.37	0.76	0.39	0.55	0.32	0.53	0.39	0.51
palrt2Vh	0.27	0.56	0.25	0.54	0.23	0.33	0.21	0.35	0.28	0.30
palrt2VH	0.10	0.33	0.13	0.35	0.19	0.27	0.20	0.25	0.12	0.27

Tabla 10.9: Probabilidades de subbandas transmitidas  $p(T)$  y recibidas correctamente condicionadas a ser transmitidas  $p(A/T)$ , para diferentes casos con escenario MAN y fuentes ráfaga

Siguiendo la notación de [Run96], se toma  $1 - \alpha$  del 90% lo cual determina el valor  $N_{Z_{\alpha/2}} = 1.96$  (este valor supone el valor del eje de las x donde la normal cubre el 90% del área entre  $-N_{Z_{\alpha/2}}$  y  $N_{Z_{\alpha/2}}$ ).

Para la aplicación de la fórmula 10.3 ha de cumplirse la condición  $\hat{p} N_t > 5$ [Run96] y que en nuestro caso se cumple para todas las 22 subbandas.

La tabla 10.10 recoge el valor promediado de  $p(A/T)$  para cada una de las subbandas con su intervalo de confianza.

Comentar que en aquellas donde el intervalo de confianza es grande (sea el caso de palrt1VH) es debido a que son subbandas con poca prioridad y por tanto el sistema no asegura tanto la llegada (dependiendo fuertemente del comportamiento de la red) como es el caso de pb2rvh (subbanda de máxima prioridad).

En general, se observa como cabía esperar que a subbandas con mayor prioridad, la probabilidad de llegar correctamente cuando han sido transmitidas es mayor, lo cual significa que el sistema ante diferentes estados de congestión de la red con servicio *best effort* trata de mantener los criterios visuales (llegada de las subbandas más

<i>subbanda</i>	$\hat{p}(A/T)$	Intervalo
pb2rvh	0.9900	0.0141
pb2rvH	0.9495	0.0311
pb2rVh	0.9011	0.0425
pb2rVH	0.8542	0.0519
pa2rvh	0.8142	0.0554
pa2rvH	0.7347	0.0628
pa2rVh	0.7426	0.0626
pa2rVH	0.7663	0.0654
pblrt1vH	0.9163	0.0461
pblrt1Vh	0.8032	0.0725
pblrt1VH	0.4137	0.1639
palrt1vh	0.6000	0.0772
palrt1vH	0.5832	0.1052
palrt1Vh	0.4463	0.1484
palrt1VH	0.4342	0.1669
pblrt2vH	0.8689	0.0523
pblrt2Vh	0.7621	0.0779
pblrt2VH	0.3311	0.1303
palrt2vh	0.5821	0.1013
palrt2vH	0.4879	0.1082
palrt2Vh	0.3295	0.1284
palrt2VH	0.2616	0.1575

Tabla 10.10: Promediado de la probabilidad de recibir correcta una subbanda cuando ha sido transmitida ( $p(A/T)$ ) y su intervalo de confianza



importantes según el SVH).

## 10.6 Conclusión

Este capítulo ha tratado de evaluar el sistema propuesto ante diferentes escenarios de transmisión en redes ATM con servicios ABR (*best effort*).

Se observa respecto a los resultados recogidos en el anexo B para los codificadores estudiados en el capítulo 3 una mejora en la reducción de artefactos que dañan las secuencias y la calidad visual de las imágenes.

Los efectos negativos derivados por la estimación de movimiento y la DCT han sido superados.



Figura 10.8: Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario LAN con fuentes ráfaga: referencia 3



Figura 10.9: Reproducción de 4 fotogramas de la secuencia de Miss América en el instante 960 ms de la simulación en escenario MAN con fuentes ráfaga: caso 2

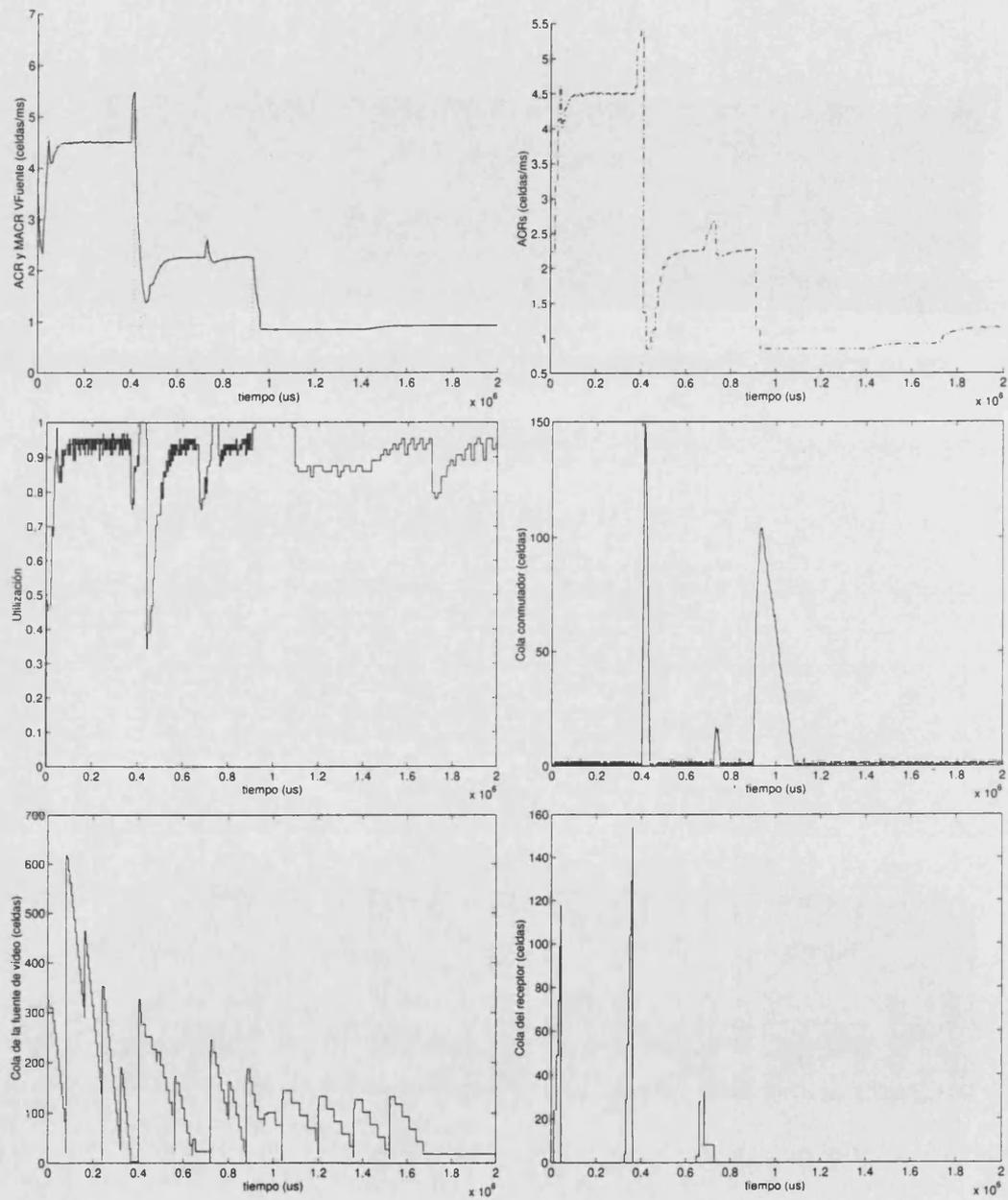


Figura 10.10: Escenario LAN con tráfico persistente: referencia 1

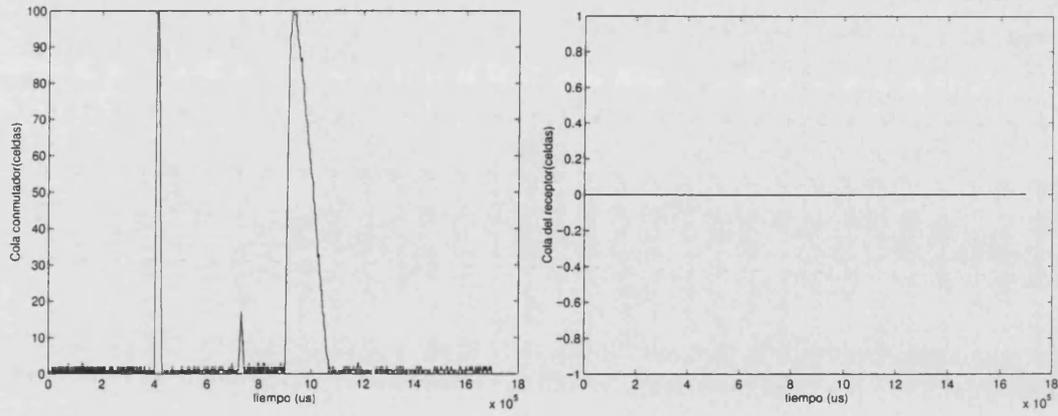


Figura 10.11: *Escenario LAN con tráfico persistente: referencia 2*

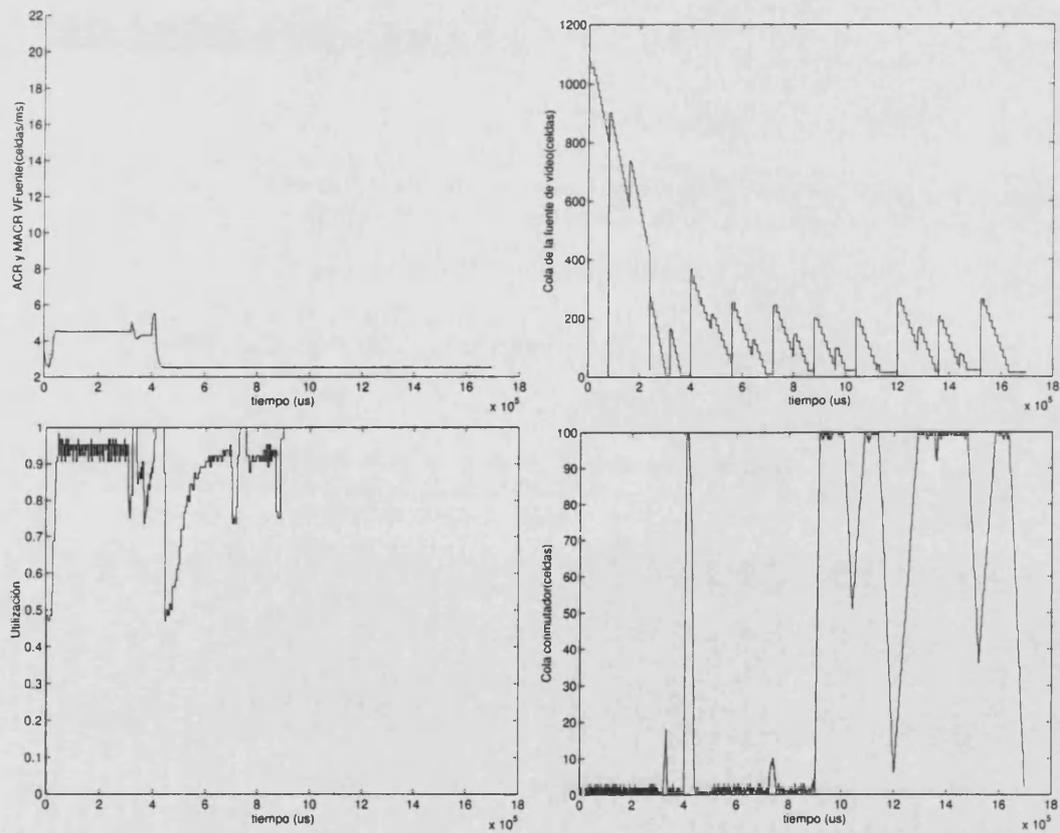


Figura 10.12: *Escenario LAN con tráfico persistente: referencia 3*

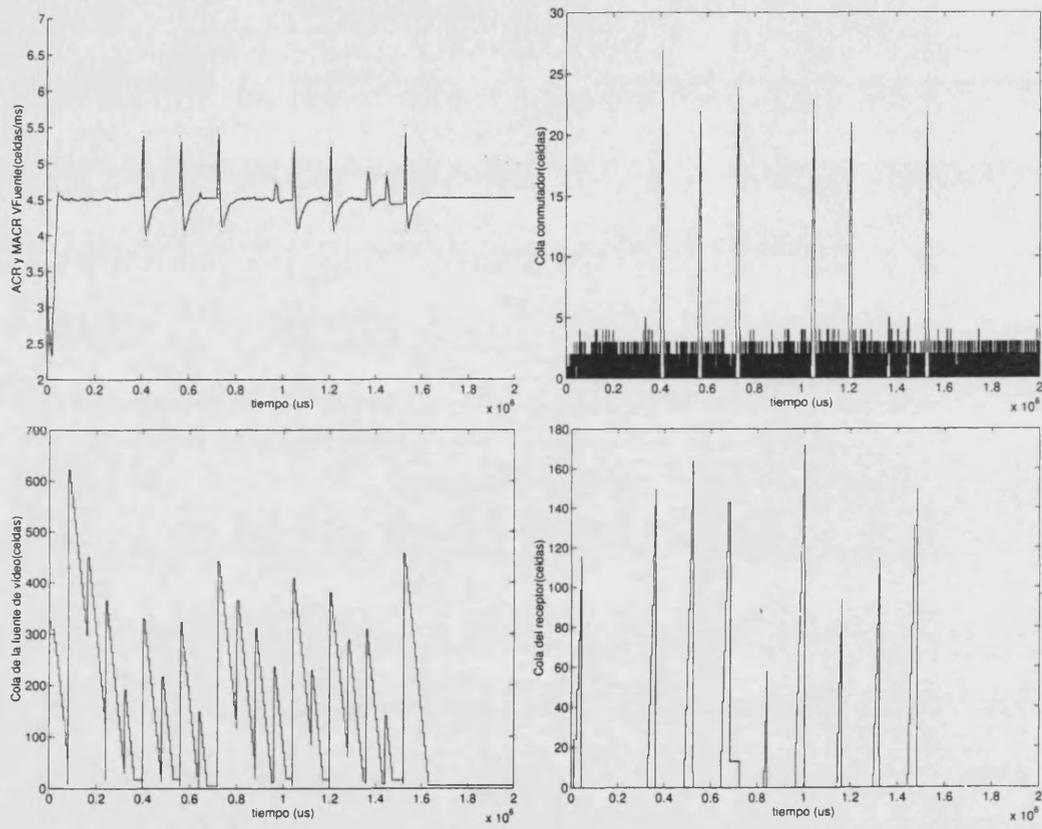


Figura 10.13: *Escenario MAN con tráfico persistente: referencia 1*

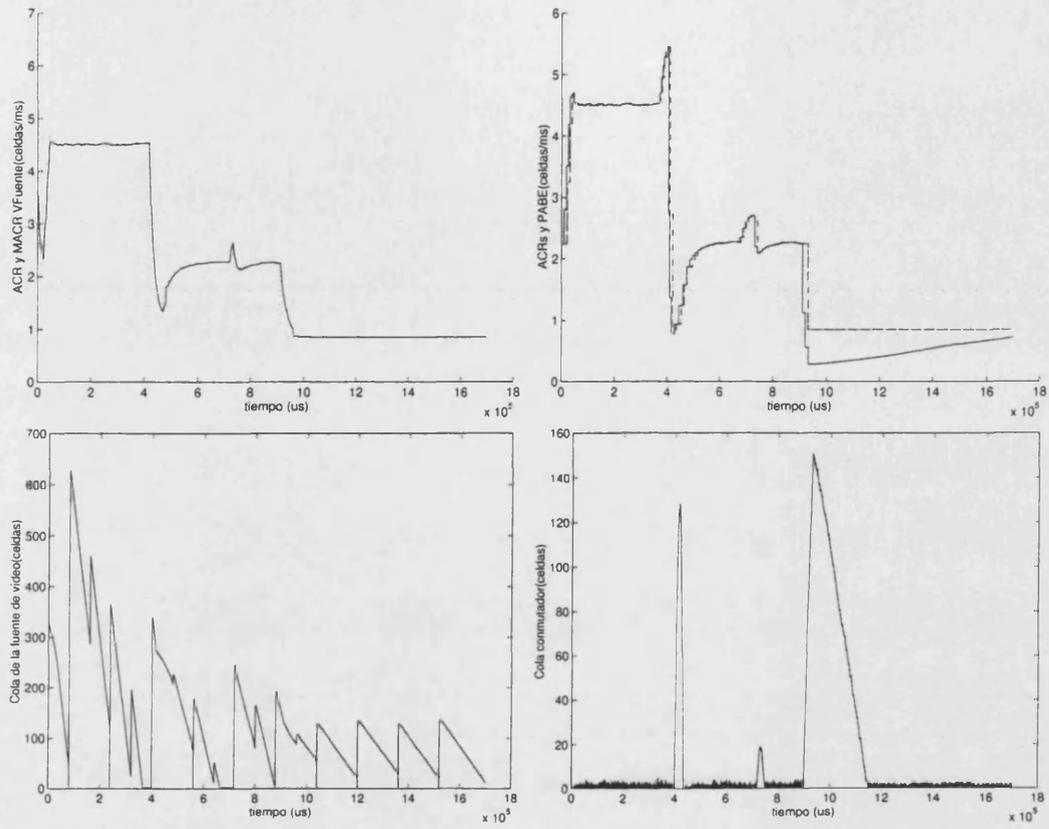


Figura 10.14: Escenario MAN con tráfico persistente: referencia 2

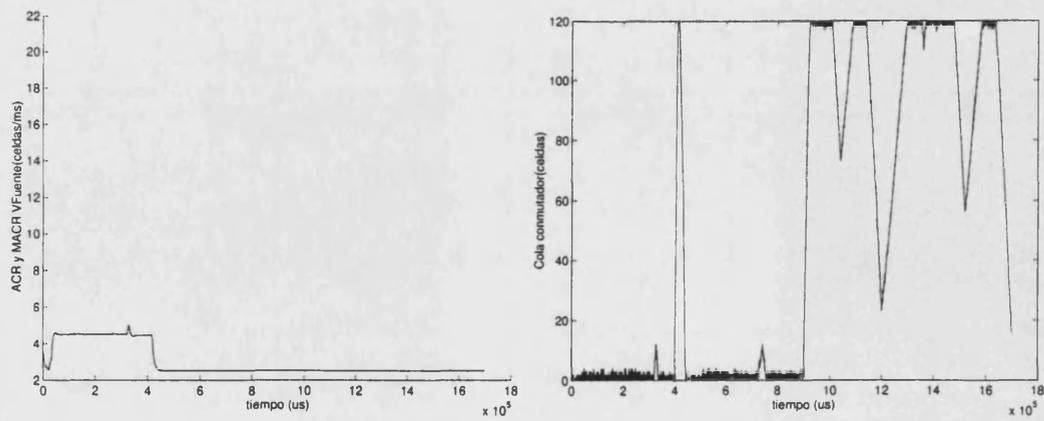


Figura 10.15: Escenario MAN con tráfico persistente: referencia 3

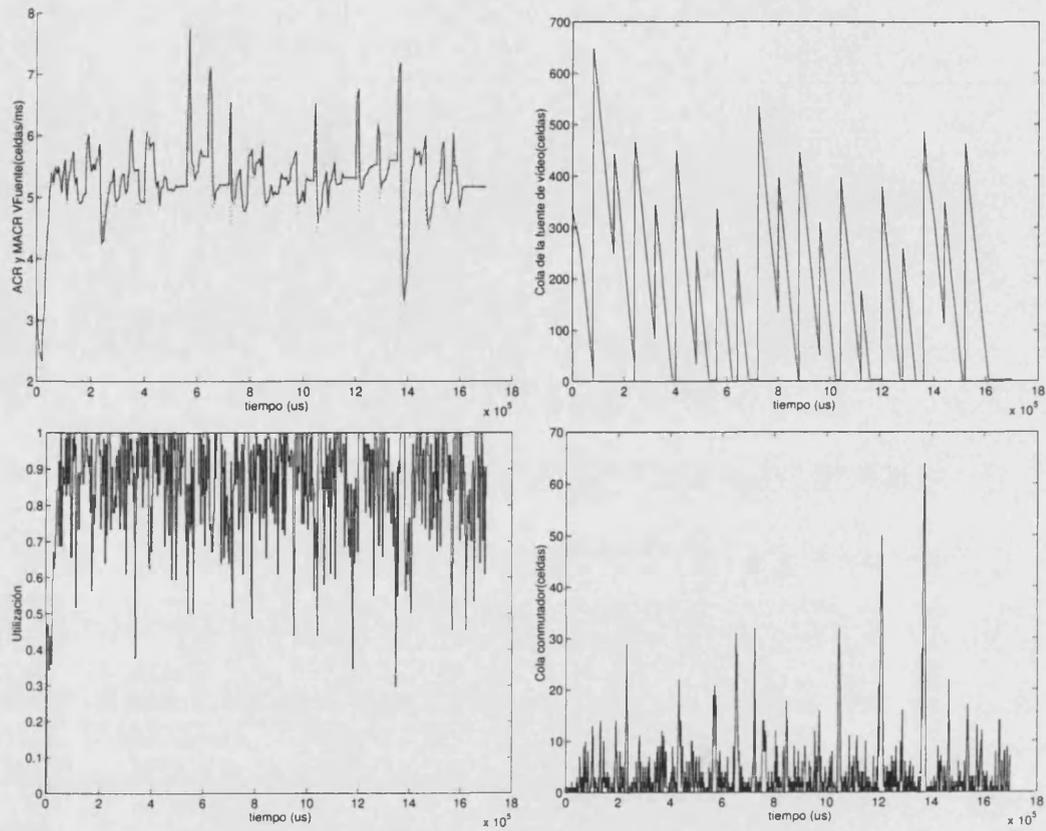


Figura 10.16: *Escenario LAN con tráfico ráfaga: referencia 1*

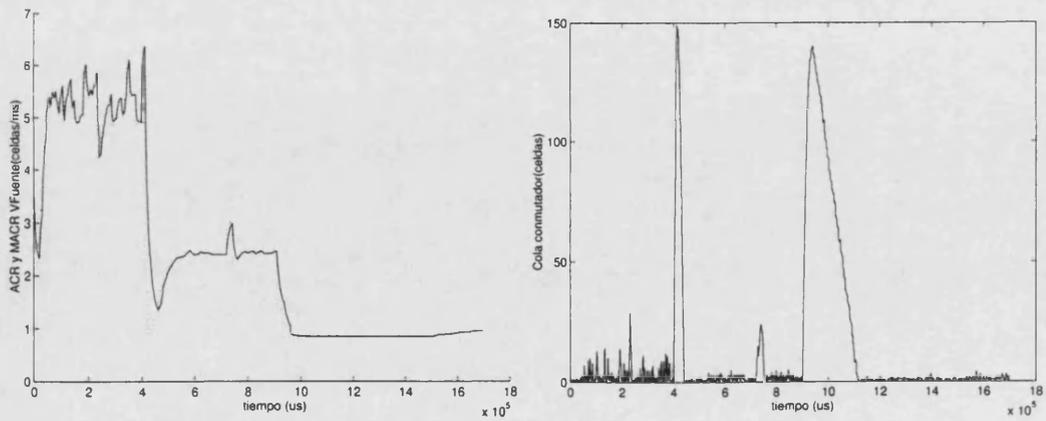


Figura 10.17: *Escenario LAN con tráfico ráfaga: referencia 2*

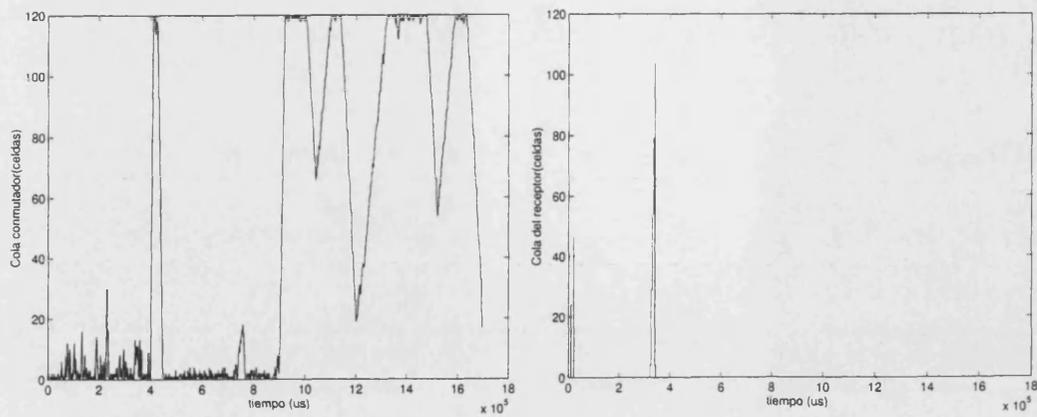


Figura 10.18: Escenario LAN con tráfico ráfaga: referencia 3

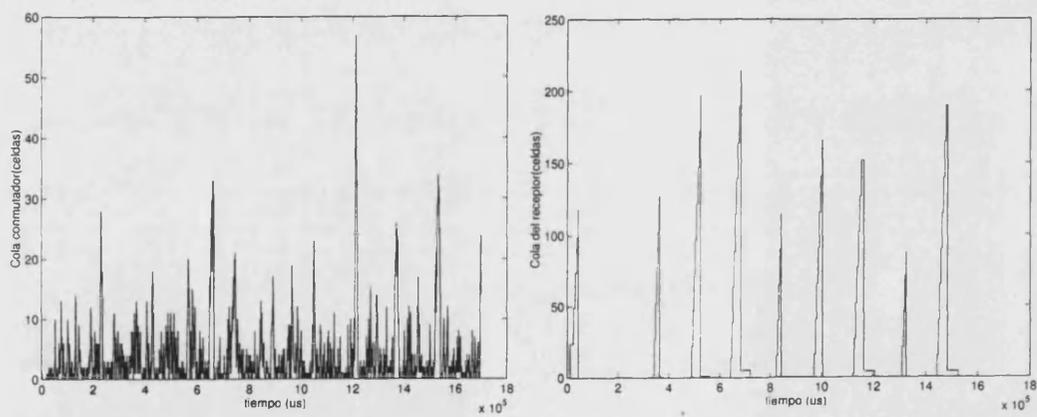


Figura 10.19: Escenario MAN con tráfico ráfaga: referencia 1

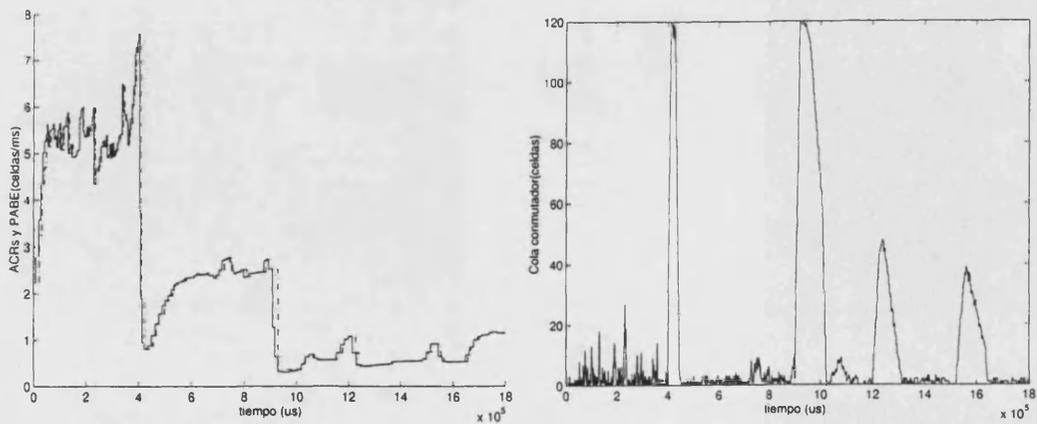


Figura 10.20: Escenario MAN con tráfico ráfaga: referencia 2

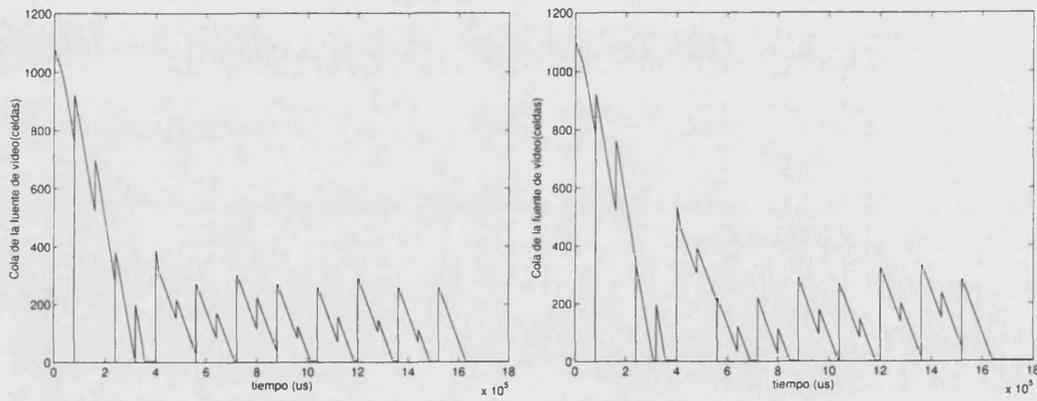


Figura 10.21: *Escenario MAN con tráfico ráfaga: secuencia de vídeo de Salesman y Claire*

**Parte V**  
**Discusión**

# Capítulo 11

## Discusión

Tras haber analizado y diseñado el sistema propuesto para compresión y transmisión de vídeo sobre redes con servicio *best effort*, en los capítulos 8, 9 y 10 se recogen los resultados que permiten evaluarlo. En ellos han sido analizadas las prestaciones tanto de los diferentes elementos del sistema como del sistema completo.

Dichos resultados como cabía esperar, verifican que los objetivos iniciales marcados en el comienzo de la tesis, es decir introducir los criterios visuales en el proceso de compresión y transmisión, han sido cumplidos, satisfaciendo los requisitos establecidos en el capítulo 6.

Estos resultados son discutidos en la sección 11.4 del presente capítulo en la sección 11.4. Además éste capítulo incluye la discusión de las decisiones tomadas, las aportaciones realizadas y las conclusiones obtenidas, así como las líneas futuras abiertas.

### 11.1 Decisiones tomadas y problemas resueltos

Respecto a la red de interconexión, la decisión de una implementación sobre ATM con servicio ABR estriba básicamente en las dificultades que entraña actualmente, según los medios de que disponemos experimentar sobre Internet porque los protocolos para un modelo IS, principalmente RSVP no están ampliamente difundidos entre los diferentes nodos de encaminamiento.

Por otra parte los protocolos RTCP de realimentación de informes (como servicio para aplicaciones de tiempo real), no aportan la información suficiente y apropiada como para poder estimar la tasa permitida de salida de las fuentes (necesaria en la asignación de bits para los coeficientes de la transformada) a fin de adaptarse a los diferentes situaciones de congestión en la red. Una línea que queda abierta, es el estudio que permita dentro del modelo IS relacionar y estimar los diferentes parámetros y variables entre servicios equivalentes tanto sobre redes ATM como sobre redes IP.

Respecto a los controles de flujo ABR utilizados, el comportamiento que pueda tener el control de flujo no es un objetivo marcado en la tesis por dos razones. Primero porque las aportaciones de unos controles de flujo frente a otros varían y dependen de la configuración utilizada, sin presentarse de momento un mecanismo

que supere a los demás en todas las configuraciones, y en definitiva lo que unos ganan en un sitio, pierden en otro; obviamente es una línea de investigación muy particular y específica[Mar97]. Y segundo, que respecto al sistema propuesto, el requisito de mantener colas con ocupaciones pequeñas para que la variabilidad de retardo sea reducida y la adaptación progresiva al estado de la red en configuración de cuello de botella, por lo general es bien soportado por los controles de flujo habituales con realimentación de tasa explícita. Estos controles de flujo suelen llevar a las fuentes a su punto de óptimo de funcionamiento y adaptación a la congestión de forma más rápida[Cha94].

Otra decisión tomada reside en la manera de enviar la información. Cabe destacar que este punto ha sido crítico porque la decisión de compromiso es relativamente complicada. Inicialmente pensé en la multiplexación de los flujos [Sal97][Mor97], unos con mayor prioridad y otros con menor prioridad. El funcionamiento se basaba en la emisión de dos flujos de información, uno de ellos con tasa MCR y el otro con tasa ACR-MCR, de tal forma que la suma de ambos generaba la tasa permitida ACR.

Esto suponía mandar a intervalos constantes de tiempo  $\frac{1}{MACR}$  celdas de alta prioridad y a intervalos  $\frac{1}{ACR-MACR}$  el resto. La multiplexación ya está implementada en la capa AAL3-4 (pero no en AAL-5) aunque sin embargo no se implementa multiplexación dentro de la misma PDU. En definitiva, aunque permita transmitir a MCR con mayor seguridad y mínimas pérdidas, esto complica en exceso la capa de adaptación AAL.

Además también pensé en métodos para marcar prioridades utilizando el CLP para distinguir el tráfico de mayor prioridad respecto al de menor, pero al ser modificable dicho bit en los conmutadores no era fidedigna su utilización.

Respecto al proceso de descomposición, en la selección de los filtros (sección 4.5) se permite cierto grado de libertad, buscando implementaciones separables con fase lineal, aliasing nulo y reconstrucción perfecta.

El desarrollo de nuevas bases son objetivo de la mayoría de los trabajos de investigación[Dau92]. Las técnicas de compresión basadas en la transformada Wavelet se centran en la búsqueda de bases con propiedades interesantes para compactar la energía en pocos coeficientes, de manera que puedan implementarse aproximaciones sucesivas, es decir se puede desechar un apropiado subconjunto de funciones base con mínima pérdida de información.

Se ha decidido en la implementación realizada que la distinción de diferentes orientaciones quede relegada a las orientaciones más predominantes[LeG88][VK95][GS96] como son la vertical, horizontal y diagonal, sin pretender buscar respuestas mucho más complejas y mejor adaptadas al SVH; dicha decisión se realiza por tratarse tanto de compresión como de transmisión de vídeo. Además que un estudio de compresión mucho más adaptado al SVH en entornos de red con servicios *best effort* sin calidad de servicio garantizada y con pérdidas no aporta ventajas suficientes que justifiquen su complejidad.

La selección del filtro Haar para el eje temporal, radica tanto en la capacidad de memoria necesaria como en el retardo introducido al sistema al tener que retener fotogramas; de la misma manera fijar dos niveles de resolución permite suficiente decorrelación sin aumentar la complejidad y el coste computacional.

## 11.2 Aportaciones

La aportación principal realizada es un algoritmo de compresión de vídeo basado en criterios visuales, mediante la descomposición wavelet 3D en octavas, adaptado para la transmisión sobre redes con servicios *best effort* que incorporen controles de flujo cerrado por tasa explícita y negociación de mínimo ancho de banda de salida. El objetivo perseguido ha sido tratar de degradar visualmente las secuencias mediante criterios visuales para adaptarlas a la tasa de salida de la fuente.

Con la descomposición 3D realizada, mediante subbandas independientes y con diferentes prioridades, se han evitado los inconvenientes de la estimación de movimiento debido a las dependencias y relaciones asociadas entre las diferentes estructuras de datos de los compresores actuales, especialmente entre fotogramas.

El algoritmo anterior, está formado por pequeñas aportaciones, que tratan de tomar una decisión de compromiso para la implementación final. Dichas aportaciones son:

- introducir una descomposición mediante banco de filtros biortogonales separables basado en diferentes estudios neuronales[Wat87a], para que haya permitido introducir los criterios visuales con mayor facilidad
- incluir mecanismos de minimización para la distorsión perceptual, utilizando funciones que caracterizan las respuestas visuales humanas[Art97] [BM93]
- adaptar la tasa de bits de la fuente de vídeo comprimido a la red, haciendo un seguimiento de la evolución de la congestión a través de la realimentación del control de flujo, y estimando el comportamiento mediante estudios de autosimilitud de los estadísticos del tráfico multiplexado[Ale95][WW94]
- ordenar las subbandas independientes en la transmisión, para que los mecanismos de tiempo real inherentes a las secuencias de vídeo perjudiquen lo menos posible a la información transportada

La introducción de la arquitectura propuesta basada en sistemas de colas con prioridades según el usuario final, con la realimentación de la tasa permitida de salida y con una predicción del comportamiento de la red, permite extender e introducirse como herramienta de experimentación de nuevos sistemas telemáticos.

## 11.3 Validación del sistema

El sistema ha sido probado analíticamente y en entornos de simulación (ver capítulos 9 y 10), tratando de analizar sus prestaciones respecto a los requisitos deseados.

Cabe destacar que no han sido presentados resultados MSE o PSNR (ver expresiones 3.1 y 3.2) de las secuencias debido a que al alterar los pesos de la distorsión de cada una de las subbandas, en la teoría *Rate-Distorsion* al aplicar el método de minimización de Lagrange (ver capítulo 7), estos valores no serían los óptimos matemáticamente aunque sí perceptualmente.

Respecto a los costes computacionales destacar, como se ha visto en el capítulo 4, que las transformadas wavelets requieren procesar la imagen de forma completa, lo cual es una desventaja frente a los mecanismos utilizados por los estándares de compresión basada en la DCT y estimación de movimiento. Estos costes no han sido introducidos en la memoria porque quedan fuertemente ligados a la fase de implementación y a la arquitectura del computador sobre la que se ejecute.

## 11.4 Discusión

Se ha visto dentro del marco de trabajo establecido (las redes con servicios *best effort*), que los resultados obtenidos cumplen los requisitos establecidos en el capítulo 6. Por tanto la descomposición 3D utilizada y estudiada en el capítulo 4, incorporando los mecanismos propuestos de adaptabilidad al estado de congestión de la red, permite abordar los problemas planteados en los estándares actuales de compresión de vídeo cuando son utilizadas estas redes como medio de transporte.

Pasemos a discutir detenidamente dichos resultados.

Los resultados del capítulo 9 se centran en validar los algoritmos propuestos para las prioridades de subbanda, la asignación de bits y las variaciones del factor de compresión. Por tanto cabe destacar a la vista de los resultados obtenidos que el sistema da una buena calidad visual, cuando han sido aplicados los estudios de minimización de la distorsión perceptual total propuesta.

En el capítulo 10 han sido presentados los resultados finales tras la transmisión sobre redes ATM con servicio ABR. En dichos resultados viene a reproducirse de forma dinámica los resultados anteriores y se observa que, en el caso de tráfico persistente por el resto de conexiones cuando en el enlace troncal hay cambios de ancho de banda, es la configuración más crítica.

Ello es debido a que la utilización del enlace es alta por ser fuentes persistentes, por tanto en los momentos de transición fuerza al conmutador a retener gran número de celdas, con una alta probabilidad de desbordamiento. En el caso de fuentes ráfaga, este efecto es menos importante porque la utilización del enlace es menor, sin embargo introduce un problema adicional que son las fuertes variaciones del MACR, lo cual puede hacer sobredimensionar en ocasiones el ancho de banda permitido. Una posible solución para éste último caso, puede ser modificar el valor  $\alpha$  del filtro de estimación, en función del coeficiente ráfaga del ACR.

En general, tanto para fuentes persistentes como ráfaga, la mejor opción para dar un servicio de vídeo es mantener un sistema de realimentación rápido (en nuestro caso la diferencia LAN y MAN a 10 Mbps no ha sido crítica) para evitar dichos transitorios (con lo cual las colas permanecen casi vacías sin introducir excesiva variabilidad en el retardo de llegadas de celda), y delegar en la red y sus buenas intenciones el reparto equitativo del ancho de banda disponible. Con estas premisas se permite mantener la red en un grado de congestión bajo y por tanto la probabilidad de pérdidas en valores casi nulos.

Como ha podido también comprobarse, negociar valores de tasa mínima de salida

MCR excesivamente altos (por encima de la porción de ancho de banda equitativa PABE en el conmutador cuello de botella), aunque en un principio puede llegar a pensar en una mayor calidad, en situaciones de congestión puede llegar a producir excesivas pérdidas tanto por desbordamiento como por retraso en la llegada al receptor.

Otro elemento que cabe discutir, es el tiempo de retardo inicial tras la llegada de la primera celda en el receptor (mecanismo para reducir la variabilidad de las llegadas). Este tiempo cuanto más alto, genera mayor probabilidad de subbandas correctas a cambio de un mayor retraso total extremo a extremo y una mayor ocupación en la cola de llegada del receptor.

Se ha observado en las simulaciones realizadas que la elección de éste valor es importante respecto al funcionamiento del sistema global y que sería interesante poderlo modificar dinámicamente. Sin embargo por las características de la realimentación ABR por tasa explícita, no puede obtenerse información del dejitter introducido por la red. Para poder controlar el dejitter sería necesario introducir en el nivel de aplicación, protocolos similares a RTP y RTCP.

Alternativas y posibles soluciones a las variaciones de los tiempos de retraso, tal y como ha sido propuesto el sistema, sería saltar o repetir fotogramas en la secuencia reproducida con la finalidad de poder controlar el retraso introducido por la cola del receptor.

Por último, cabe destacar respecto a la implementación final, que el sistema propuesto permite la viabilidad de ésta, de la manera en que ha sido analizado y diseñado. Por otro lado, el hecho de no realizar estimación de movimiento (sino una transformada 3D tanto de análisis como de síntesis), equilibra el coste tanto del emisor como del receptor, a diferencia de los codificadores y decodificadores de vídeo estándar (como p.e. puede verse en MPEG-2).

## 11.5 Conclusiones

Las investigaciones hechas nos han confirmado que las técnicas de compresión y transmisión de vídeo presentan una difícil migración hacia los servicios *best effort* si los criterios visuales no son explotados adecuadamente, sin que ello suponga en ningún momento tratar de poner en entredicho la operatividad y las prestaciones de los estándares estudiados.

La estimación de movimiento incorporada, de la misma manera que permite aumentar el factor de compresión (reducir la entropía), introduce dependencias en las estructuras de datos que perjudican la información visual ante pérdidas no previstas en la transmisión sobre redes con servicios *best effort*, p.e. un fotograma tipo B será visible siempre y cuando los P e I que lo generan hayan sido recibidos correctamente.

Esta dependencia ha sido abordada a través de una transformada 3D con la finalidad de descomponer los fotogramas como suma de texturas independientes.

Por otro lado la búsqueda de una descomposición similar a la del SVH, permite que dicha transformada, incorpore ciertas ventajas para hacer más eficiente visualmente la asignación de bits sobre los coeficientes. Esta descomposición está más adaptada al

SVH que las técnicas basadas en métodos progresivos sobre los coeficientes de la DCT según ha sido visto en el capítulo 5.

En el caso de secuencias dañadas, se puede optar bien por mecanismos de reconstrucción del error por estimación del comportamiento de los pixels vecinos, bien a través de la introducción de ceros (por tener histogramas con media nula) en aquellas subbandas perdidas; esta última opción es equivalente a no hacer nada.

En el sistema propuesto cabe destacar el equilibrio entre la complejidad del emisor y del receptor frente a los estándares actuales en los que el decodificador es más simple (el que compra el usuario) que el codificador. Este resultado permite compensar los costes tanto del sistema de emisión como el de recepción.

Por otro lado respecto a la transmisión, si realmente la simplicidad ha de ser la herramienta para participar en servicios *best effort*, la mejor opción es utilizar protocolos simples y relegar en el control de flujo para que adapte la fuente a la red, siempre y cuando la fuente pueda seguir las pautas marcadas por dicha tasa permitida; en caso contrario si se tratan de introducir protocolos más complejos u otros controles de flujo por encima de ABR para subsanar las deficiencias de los servicios *best effort* aparecen problemas, p.e. caso de TCP sobre ABR[Goy98].

Finalmente como ha podido comprobarse en las simulaciones realizadas en el capítulo 10, tratar de alterar el valor del MCR para tener una tasa mínima de salida elevada en la fuente de vídeo, en situaciones de alta congestión perjudica más que beneficia. Se concluye que es mejor dejar dar libertad a la red y que los controles de flujo operen los servicios *best effort* de la forma más equitativa posible.

Vistas las conclusiones, podemos pensar en posibles mejoras, como se discute a continuación.

### 11.5.1 Posibles mejoras

Respecto a la descomposición pueden introducirse implementaciones más eficientes en el proceso de filtrado (convolución) mediante métodos polifásicos[Vai93].

Respecto a la transmisión, en situaciones como las planteadas en el presente trabajo, cabe la posibilidad de aplicar técnicas más robustas a base de redundancia (algunas de ellas actualmente aplicadas en audio) consistentes en incorporar por cada PDU tanto la subbanda a transportar como información de las subbandas previas con el fin de poder extraer dicha información en el caso de pérdida; un claro ejemplo puede verse en la figura 11.1.

Por otro lado, también puede pensarse en utilizar códigos correctores para poder recuperar las subbandas dañadas. Si esta mejora no ha sido contemplada, ha sido por la lentitud en la manipulación de la información, que hace que dichos códigos sean ineficientes en técnicas de tiempo real.

Respecto a la reconstrucción de las pérdidas, una de las mejoras incorporadas en la transmisión de MPEG2, es la enumeración de las celdas ATM enviadas en cada PDU a fin de poder detectar el error y que no se propague espacialmente. Esta técnica es utilizada indirectamente en la implementación del sistema propuesto considerando la subbanda perdida como una subbanda nula (con ello se conserva la media nula de la

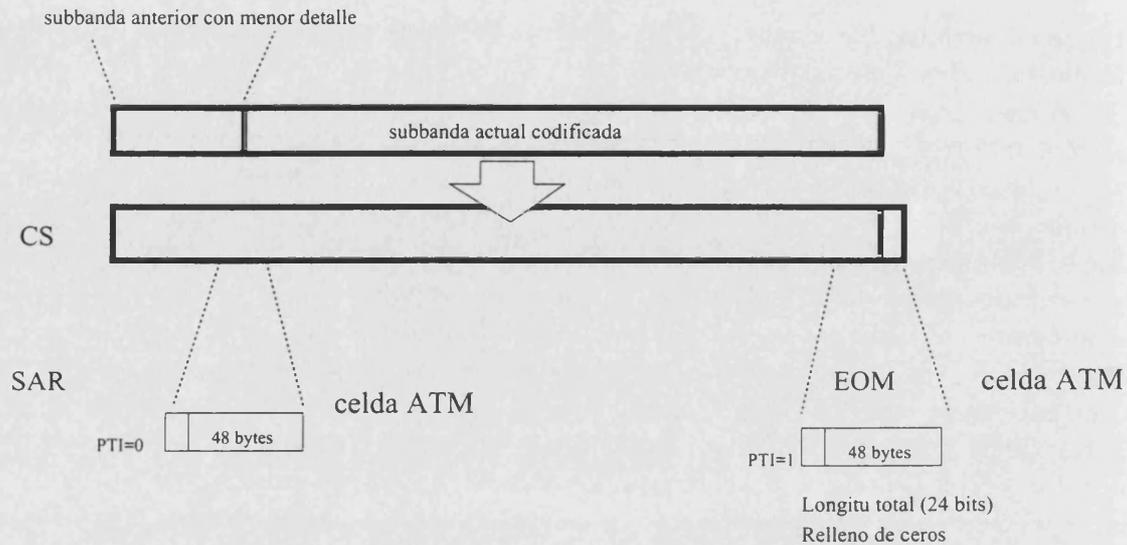


Figura 11.1: Alternativa a AAL5+ en codificación subbanda

subbanda).

Respecto a los controles de flujo, como se ha discutido anteriormente las variaciones de unos con otros son mínimas; sin embargo dadas las características de la aplicación, transmisión con prioridades de las diferentes subbandas, cabe pensar en la utilización de métodos reactivos frente a las congestiones basados en la incorporación a los conmutadores de técnicas EPD *Early Packet Discard* [Flo94] consistentes en descartar completamente las PDU cuando de antemano se sepan que vayan a perderse en el conmutador (cabe recordar que el servicio sigue siendo *best effort*). Esta mejora se puede llevar a cabo porque la información viaja compacta dentro de la subbanda y de forma independiente.

También como se recoge en [T.L97] puede ser interesante una renegociación del MCR cuando se utilizan modelos autoregresivos para estimar el comportamiento de la fuente de vídeo. En nuestro caso y como se ha visto en los resultados obtenidos, dado el principio de prioridades en las subbandas por criterios visuales, cabría pensar en modificar los algoritmos de reparto de ancho de banda en los conmutadores, teniendo en cuenta el valor del MCR u otro parámetro que permita introducir los criterios suficientes para poder etiquetar los diferentes flujos según su prioridad.

## 11.6 Líneas futuras

La implementación completa del sistema es necesaria para una experimentación con prototipos reales que permita una validación "real". Se desarrolla en paralelo a la tesis con diferentes proyectos, implementaciones sobre IP a través de *sockets* UDP y protocolos RTP y RTCP.

Respecto a los contenidos abordados a lo largo del trabajo realizado, se abre un foro

de debate entre los mecanismos tradicionales y aspectos tan diferentes abordados desde perspectivas de la transmisión y la compresión. Aplicar los estudios ópticos y cerebrales como motor de nuevos desarrollos junto con técnicas avanzadas y la tecnología de red, es navegar en un ente dinámico con muchas posibilidades y multitud de puertas abiertas.

Surgen por los grados de libertad que plantea el marco del problema y la versatilidad del entorno, posibles campos de desarrollo que paso a describir:

- estudio de nuevos filtros y mecanismos para adaptar el sistema visual humano ante los entornos de transmisión propuestos
- aplicar métodos de estimación de movimiento y métodos de codificación más eficientes explotando las propiedades de la codificación subbanda[VK95]



Figura 11.2: Ejemplos de reposición de subbanda paso baja perdida en 2ª resolución

- aplicar técnicas para recuperación de subbanda frente pérdidas y errores. Un ejemplo fácil consiste en la sustitución de subbandas de bajas frecuencias de unos fotogramas en otros donde se hayan producido pérdidas, ver figura 11.2 donde la primera fila representa un fotograma de 80 ms de retraso con sustitución de la subbanda paso baja y el error cometido, y en la segunda fila se reproduce exactamente pero con un retraso de 160 ms
- estudio de mejores algoritmos para la estimación del comportamiento de la red (analizando su comportamiento de autosimilitud por multiplexación de tráfico[WW94][WL94]). De manera inmediata cabe pensar en estudiar la posibilidad de modificar el valor  $\alpha$  del estimador MACR (expresión 7.12) en función del estado de la red y la frecuencia de corte de las variaciones (anexo D)
- proponer mecanismos para eliminar los efectos de *de jitter* en momentos de congestión de la red

- transmisión multiresolución para aplicaciones en domótica, ya que los anchos de banda en estos entornos son muy reducidos y variables, según la instalación
- plantear controles de flujo en servicios *best effort* más adaptados a la compresión de vídeo

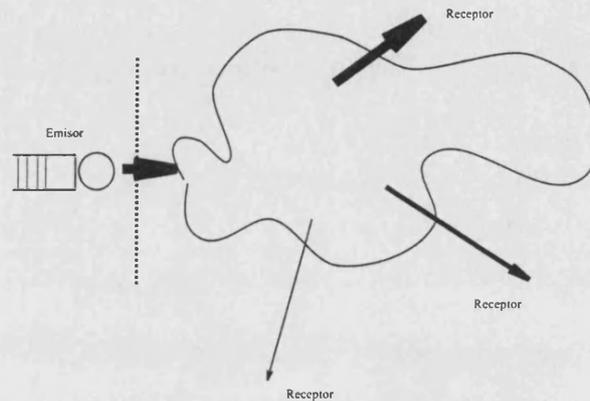


Figura 11.3: Codificación de vídeo en subbanda para multipunto

- ampliar y estudiar métodos multipunto para transmisión de vídeo sobre redes con servicios *best effort*, figura 11.3

Sin embargo podemos hablar de líneas cuando realmente existe un grupo detrás que las respalde. Concluyo mencionando las tres líneas que sí bien no han aparecido ahora, si que se puede colaborar en una misma dirección:

- aplicaciones multimedia en el marco de Internet, IPv6, ATM y la plataforma MBone
- evolución a nuevos controles de flujo para ATM en los servicios ABR y UBR
- compresión con herramientas más complejas e integración de las tecnologías de la información y codificación



# Parte VI

## Anexos

# Apéndice A

## Estándares y sus definiciones

ITU-T: Normas H.:

- *H.221*: Sincronización de múltiples canales a 64 Kbps
- *H.230*: Control y estado de equipos
- *H.243* y *H.231*: Modo de funcionamiento para videoconferencia multidifusión (*multicast*)
- *H.245*: Protocolo de control para la comunicación multimedia, frecuentemente utilizado en H.320 que facilita la compatibilidad entre productos previsto para redes ATM (H.310), RDSI, RTC o LAN
- *H.261*: Codificador de vídeo en H.320 utilizado en formato CIF a 352x288 o QCIF a 176x144
- *H.263*: Codificador de vídeo utilizado en H.323
- *H.312*: completa a la norma H.261 permitiendo transmitir imágenes en múltiplos de 64 Kbps hasta 2Mbps
- *H.320*: Conocida como px64 basada en H.261, con p de 1 hasta 30, constituye un conjunto de normas propuestas por la ITU-T para videoconferencia para RDSI-BE
- *H.321*: Adaptación de H.320 del terminal de videotelefonía en entornos RDSI-BA
- *H.322*: Sistemas y terminales de videotelefonía para redes locales con calidad de servicio garantizada
- *H.323*: Adaptación de H.320 para redes de calidad de servicio no garantizada (LAN)
- *H.324*: Adaptación de H.320 a la transmisión multimedia sobre red telefónica a menos de 20 Kbps

ITU-T: Control de conferencia. Normas T.:

- *T.120*: Conjunto de recomendaciones para las conferencias multidifusión y compartición de aplicaciones ofimáticas de vídeo, audio y datos
- *T.121*: Presentación de aplicaciones tipo
- *T.122*: Definición de servicios audiovisuales y de comunicación multidifusión audiográfica
- *T.123*: Capas de protocolo para aplicaciones de teleconferencia audiovisual y audiográfica
- *T.124*: Control global de la conferencia
- *T.125*: Especificación del protocolo de servicio de comunicación multidifusión
- *T.126*: Especificación del protocolo para imagen fija, compartir la ‘pizarra’ así como imágenes con y sin anotaciones
- *T.127*: Protocolo de transferencia multidifusión de archivos binarios

ITU-T: Normas de audio:

- *G.711*: PCM a 64 Kbps (8 bits)
- *G.721*: AD-PCM a 32 Kbits
- *G.722*: AD-PCM a 48, 56 o 64 Kbps, recomendado para codificación de señales de audio comprendidas entre 50 y 7000 Hz
- *G.723.1*: Compresión de la palabra, desarrollado por DSP Group a 5.3 y 6.3 Kbps
- *G.726*: reemplaza al G.721 AD-PCM en 16, 24,32 o 40 Kbps
- *G.727*: AD-PCM codificado en 2, 3, 4 o 5 bits con velocidades 16, 24, 32 o 40 Kbps como complemento a G.726 que especifica la manera de cambiar la velocidad dinámicamente para adaptarse a la saturación de la red de forma similar al G.764 y RTP
- *G.728*: compresión de la palabra a 16 Kbps utilizando LD-CELP
- *G.764*: Recomendación para la puesta en paquetes de la voz comprimida según las recomendaciones G.726 y G.727 similar a RTP de IETF

Estándares ISO:

- *MPEG-1 ISO 11172-x*
- *MPEG-2 ISO 13818-x*
- *MPEG-4 ISO 14496-x* en elaboración
- *MPEG-7 ISO 15938-x* en elaboración

## Apéndice B

### Test de secuencias codificadas en H.263

El presente anexo expone los resultados experimentales sobre las tres secuencias de vídeo utilizadas en el capítulo 8.

Dichas secuencias han sido grabadas en formato QCIF (176x144) para poder ser comprimidas con el estándar H.263 de videotelefonía visto en el capítulo 3. Se ha escogido este estándar por cuestiones computacionales y por la disponibilidad de las fuentes en *TELENOR* de libre disposición, aunque los resultados básicos son similares en todos los estándares vistos, basados en la transformada discreta del coseno.

Por la dificultad que acarrea el plasmar las secuencias completas de varios segundos de duración, con la finalidad de no llenar con demasiados fotogramas este trabajo, de las diferentes secuencias, han sido extraídos fotogramas aislados.



Figura B.1: *Proyección de ensayos*

Los experimentos se realizan con dos objetivos, evaluar la robustez tanto en la compresión como en la transmisión sin remarcar en costes temporales del proceso, ya que varían enormemente en función de la máquina o del hardware que se disponga. La proyección de ensayos se puede ver en la figura B.1.

Respecto a la compresión, los experimentos realizados se basan en el análisis de las secuencias codificadas en el estándar H.263, sometido a diferentes criterios y parámetros.

ros de compresión, como el área de búsqueda en la estimación de movimiento, pasos de cuantificación tanto en los fotogramas  $I$  como  $P$  e interpolación de fotogramas. Las conclusiones obtenidas se recogen visualmente en las diferentes secuencias codificadas sobre *salesman*, así como en las tablas adjuntas para cada una de las secuencias comprimidas.

Respecto a la transmisión, las tres secuencias anteriores, comprimidas de forma análoga, son sometidas a la transmisión por canales donde son dañadas y alteradas con diferentes probabilidades de error. Se visualizan los mismos resultados que en el caso anterior además del número de errores producidos.

## B.1 Compresión con H.263

Las tablas B.3 y B.4 son el resultado de la compresión de la secuencia *salesman* (figura B.2) con estimación de movimiento por búsqueda exhaustiva, codificación de línea con códigos de longitud variable por *Huffman Modificado* y muestreo umbral. Las secuencias 4 y 5 de la tabla B.4, correspondientes a la mejor y peor calidad respectivamente, son mostradas en las figuras B.3 y B.4. Los parámetros más relevantes de entrada al compresor H.263 modificado, son descritos en la tabla B.1 y los valores medios de salida en la tabla B.2. Aunque el objetivo perseguido en la presente tesis se centra en la compresión de la luminancia, dado que H.263 incluye la crominancia, se tendrá en cuenta en el análisis los coeficientes  $YCbCr$  a partir del cambio de formato de  $RGB$ , donde

$$\begin{aligned} Y &= +0.299R + 0.587G + 0.114B \\ Cr &= +0.701R - 0.587G - 0.114B \\ Cb &= -0.299R - 0.989G + 0.886B \end{aligned} \quad (B.1)$$



Figura B.2: *Secuencia original de Salesman*

s	tamaño de ventana de búsqueda
q	paso de cuantificación en los fotogramas <i>I</i>
I	paso de cuantificación en los fotogramas <i>P</i>
S	número de imágenes sin codificar

Tabla B.1: *Parámetros de entrada al compresor H.263*

SNR(Y)	señal a ruido de Y
SNR(Cb)	señal a ruido de Cb
SNR(Cr)	señal a ruido de Cr
intra	número de macrobloques codificados <i>I</i>
inter	número de macrobloques codificados <i>P</i>
Coeff(Y)	número de coeficientes de Y codificados
Coeff(C)	número de coeficientes de Cb y Cr codificados
Vector	número de vectores de movimiento codificados
CBPY	bloques de Y que se codifican
MCMPC	bloques de Cb y Cr que se codifican
Header	número de bits de la cabecera
Bits	número total de bits por fotograma
nF	número de fotogramas codificados
BR	Tasa de salida en Kbps

Tabla B.2: *Valores medios de salida del compresor H.263*

<i>n secuencia</i>	1	2	3	4	5	6	7	8	9
s	16	16	16	0	0	16	8	8	15
q	31	2	20	20	31	20	20	2	5
I	31	2	8	8	31	8	8	2	5
S	0	0	0	0	0	2	2	2	3
SNR(Y)	26.65	42.50	31.99	31.89	26.61	31.82	31.82	42.39	36.07
SNR(Cb)	35.64	44.69	38.50	38.49	35.63	38.55	38.55	44.70	40.23
SNR(Cr)	36.43	45.23	39.02	39.02	36.44	39.10	39.10	45.10	40.74
intra	0	0	0	0	0	0	0	0	0
inter	17	82	14	14	17	20	20	92	42
Coeff(Y)	54	10590	178	218	88	326	326	16403	5870
Coeff(C)	0	732	0	0	1	0	20	1066	138
Vector	106	202	72	58	76	142	142	278	174
CBPY	45	342	50	52	48	67	67	397	181
MCMPC	17	226	14	14	17	20	20	308	80
Header	53	54	54	52	53	54	54	53	53
Bits	376	12247	469	496	384	709	709	18605	6596
nF	10	10	10	10	10	4	4	4	37
BR	11.30	367.42	14.09	14.88	11.54	7.09	7.09	186.05	49.48

Tabla B.3: *Resultados obtenidos de la compresión de Salesman*

Figura B.3: *Secuencia 4 de Salesman*Figura B.4: *Secuencia 5 de Salesman*

## B.2 Transmisión con H.263

En esta sección el objetivo es observar el comportamiento de la trama de bits del estándar H.263 frente a entornos restrictivos y ruidosos. Las tablas B.6) son los resultados de las secuencias que van a ser transmitidas. Las figuras B.5 y B.6 son fotogramas de las secuencias comprimidas.

Los errores a introducir caracterizan a los medios de comunicación siguientes donde cabe destacar que en el caso de la *FDDI* o transmisión por fibra óptica, la tasa de errores es despreciable y la imagen es casi exactamente recuperada.

En la captura de los fotogramas de las secuencias dañadas, se especifica la probabilidad de error y la tasa de bits erróneos (*BER, bit error rate*).

<i>n secuencia</i>	10	11	12	13	14	15	16
s	0	15	15	15	0	15	15
q	20	20	28	2	31	10	5
I	8	8	8	8	8	8	8
S	0	0	0	0	0	0	0
SNR(Y)	30.16	30.19	29.33	42.44	30.05	34.75	35.65
SNR(C'b)	37.23	37.26	36.90	44.63	38.95	41.40	40.06
SNR(Cr)	37.54	37.52	37.13	45.18	39.12	41.65	40.64
intra	0	0	0	0	0	0	0
inter	15	15	11	70	10	25	39
Coeff(Y)	200	179	70	9384	67	886	3087
Coeff(C)	2	1	0	496	1	9	70
Vector	53	61	55	161	43	71	96
CBPY	53	58	37	292	35	108	166
MCMPC	16	16	11	168	11	29	62
Header	53	53	53	53	53	53	53
Bits	484	469	328	10655	311	1258	3635
nF	150	150	150	150	150	150	150
BR	14.54	14.09	9.85	319.68	9.36	37.77	109.07

Tabla B.4: *Compresión con 150 fotogramas de Salesman modificando "s" y "q"*

GSM	$10^{-4}$
PTSN	$10^{-5}$
FDDI	$10^{-9}$

Tabla B.5: *Medios de transmisión y la probabilidad de error asociada*

<i>secuencia</i>	Salesman	Miss América	Claire
s	8	8	8
q	10	10	10
I	10	10	10
S	0	2	2
SNR(Y)	31.68	36.82	35.99
SNR(Cb)	37.73	37.28	36.32
SNR(Cr)	38.15	36.13	38.90
intra	0	0	0
inter	23	39	11
Coeff(Y)	743	577	789
Coeff(C)	9	134	48
Vector	69	203	132
CBPY	100	133	100
MCMPC	27	83	42
Header	53	54	53
Bits	1102	1285	1265
nF	300	46	51
BR	33.08	12.85	12.66

Tabla B.6: *Resultados de las secuencias de Salesman, Miss América y Claire comprimidas*



Figura B.5: *Secuencia codificada H.263 de Miss América*



Figura B.6: *Secuencia codificada H.263 de Claire*

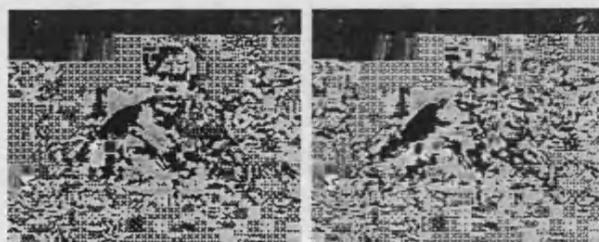


Figura B.7: *Secuencia Salesman con 582 errores,  $P_e = 10^{-3}$  BER = 0.001085*

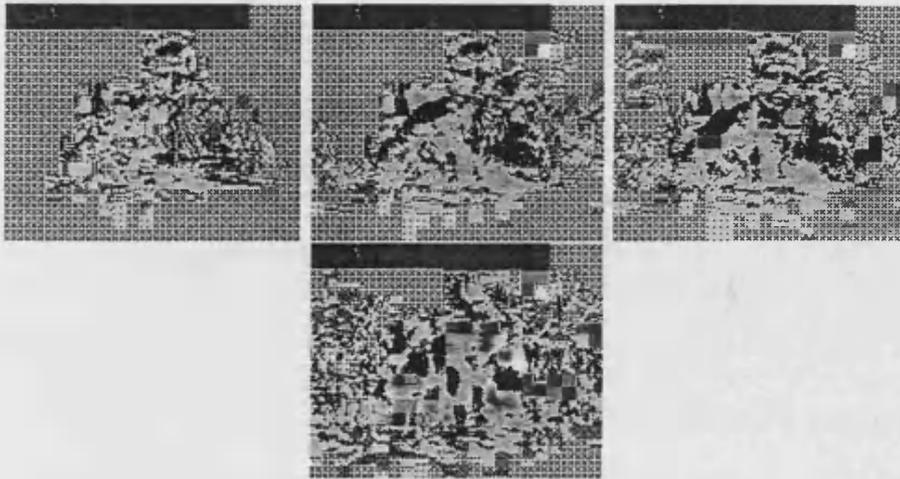


Figura B.8: *Secuencia Salesman con 321 errores,  $P_e = 10^{-3}$  BER = 0.000909*



Figura B.9: *Secuencia Salesman con 48 errores,  $P_e = 10^{-4}$  BER = 0.000136*



Figura B.10: *Secuencia Salesman con 11 errores,  $P_e = 10^{-5}$  BER = 0.000031*



Figura B.11: *Secuencia Salesman con 7 errores,  $P_e = 10^{-7}$  BER = 0.000020*



Figura B.12: *Secuencia Salesman con 14 errores,  $P_e = 10^{-9}$  BER = 0.000048*

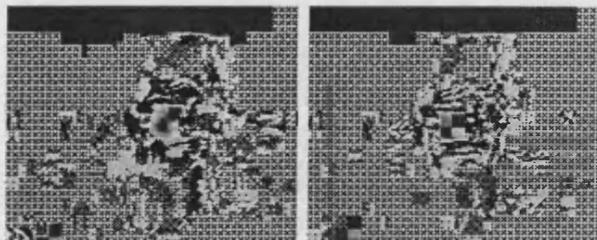


Figura B.13: *Secuencia Miss América con 747 errores,  $P_e = 10^{-3}$  BER = 0.001030*



Figura B.14: *Secuencia Miss América con 26 errores,  $P_e = 10^{-5}$  BER = 0.000036*



Figura B.15: *Secuencia Miss América con 20 errores,  $P_e = 10^{-7}$  BER = 0.000028*

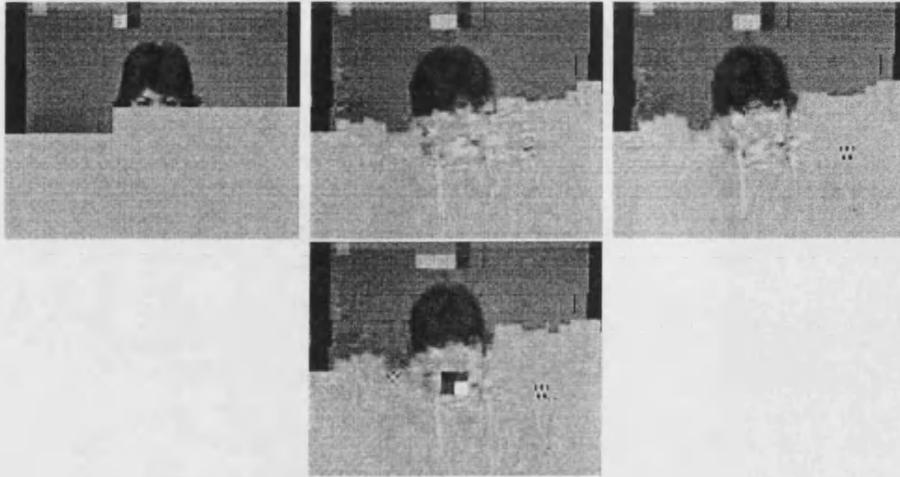


Figura B.16: *Secuencia Claire con 787 errores,  $P_e = 10^{-3}$  BER = 0.000989*



Figura B.17: *Secuencia Claire con 103 errores,  $P_e = 10^{-4}$  BER = 0.000129*



Figura B.18: *Secuencia Claire con 27 errores,  $P_e = 10^{-5}$  BER = 0.000034*



Figura B.19: *Secuencia Claire con 24 errores,  $P_e = 10^{-7}$  BER = 0.000030*

# Apéndice C

## Estudio de filtros wavelet

### C.1 Análisis temporal de los filtros

Expresión de los filtros en el eje de tiempos. Filtro biortogonal(1) para el eje tanto vertical como horizontal: figuras C.1. Filtro biortogonal(2) para el eje tanto vertical como horizontal: figuras C.2.

### C.2 Análisis frecuencial de los filtros

Expresión de los filtros en el eje de frecuencias digitales  $[0..\pi]$ . Filtro biortogonal(1): respuesta frecuencial en módulo y fase (figuras C.3 C.4). Filtro biortogonal(2): respuesta frecuencial en módulo y fase (figuras C.5 C.6). Filtro ortogonal(Haar): respuesta frecuencial en módulo y fase (figuras C.7 C.8).

### C.3 Análisis multiresolución de los filtros

Filtro biortogonal 1: primera resolución (escala y wavelet, figuras C.9 C.10), segunda resolución (escala y wavelet, figuras C.11 C.12). Filtro biortogonal 2: primera resolución (escala y wavelet, figuras C.13 C.14), segunda resolución (escala y wavelet, figuras C.15 C.16). Filtro Haar: primera resolución (escala y wavelet, figuras C.17 C.18), segunda resolución (escala y wavelet, figuras C.19 C.20), octava resolución (escala y wavelet, figura C.21).



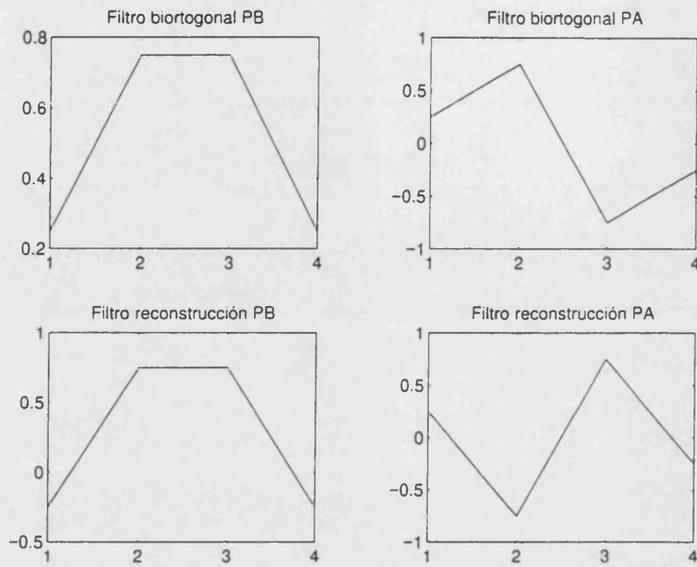


Figura C.1: Respuesta temporal del filtro biortogonal(1)

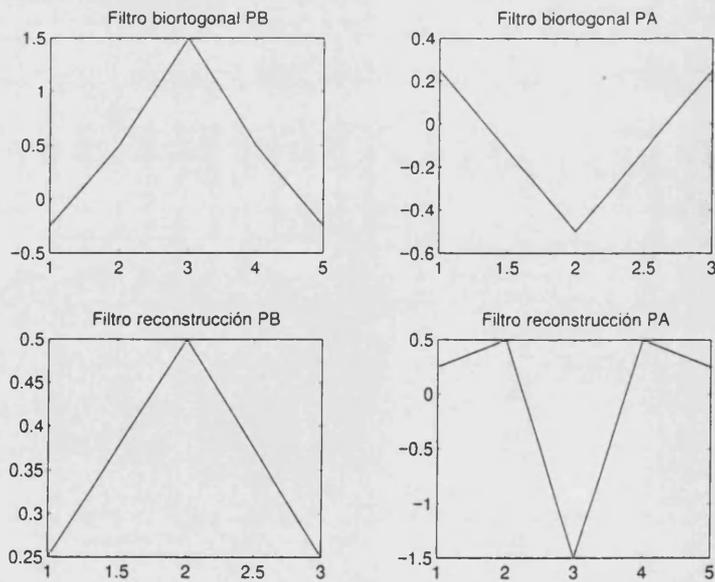


Figura C.2: Respuesta temporal del filtro biortogonal(2)

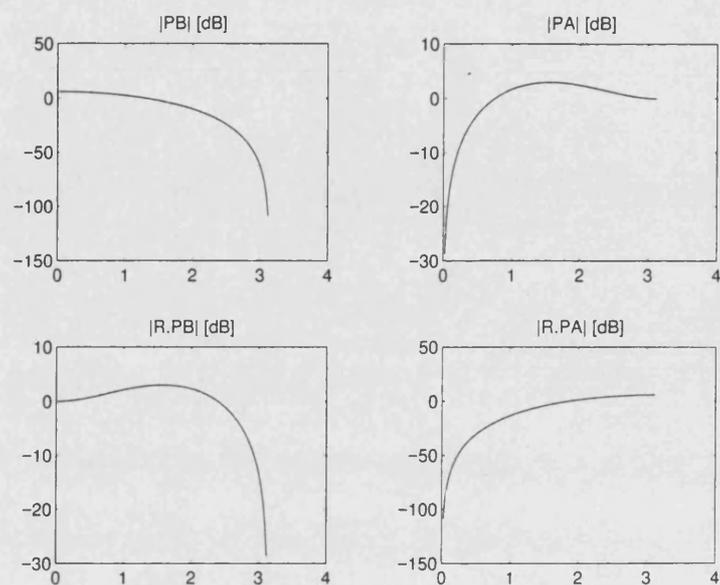


Figura C.3: Respuesta frecuencial: Módulo Filtro biortogonal(1)

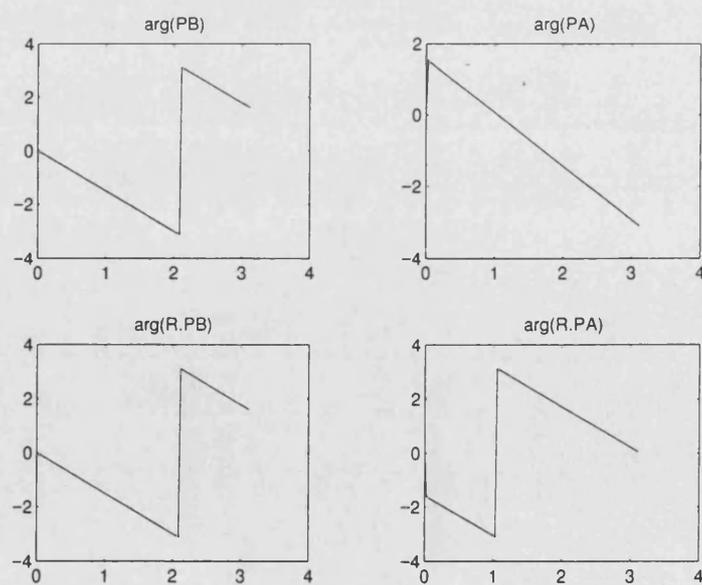


Figura C.4: Respuesta frecuencial: Fase Filtro biortogonal(1)

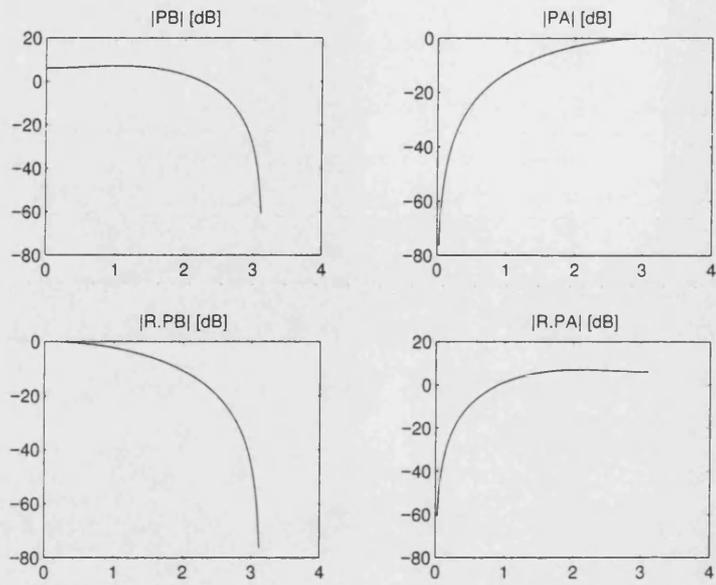


Figura C.5: Respuesta frecuencial: Módulo Filtro biortogonal(2)

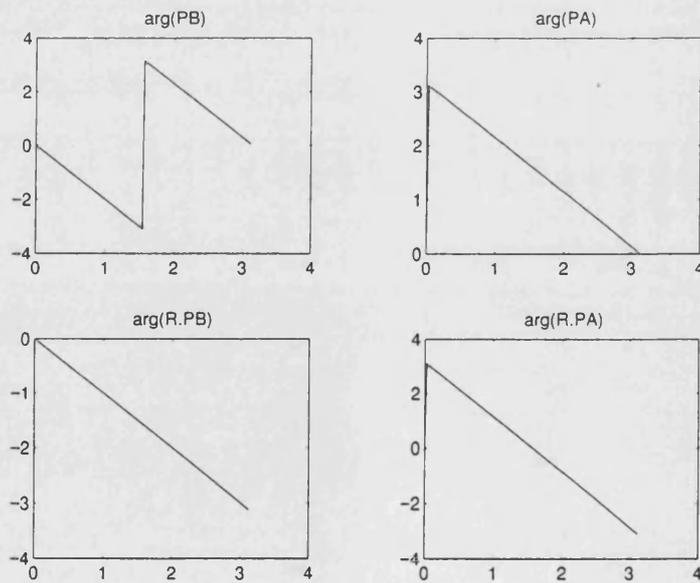


Figura C.6: Respuesta frecuencial: Fase Filtro biortogonal(2)

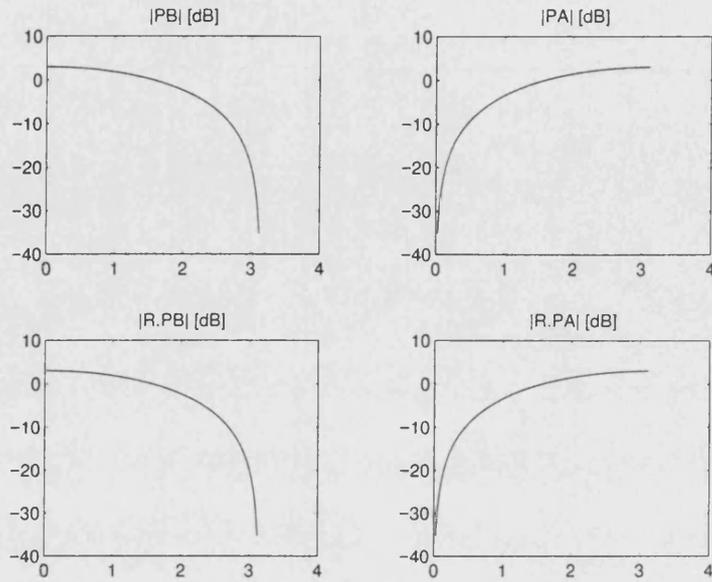


Figura C.7: Respuesta frecuencial: Módulo Filtro Haar

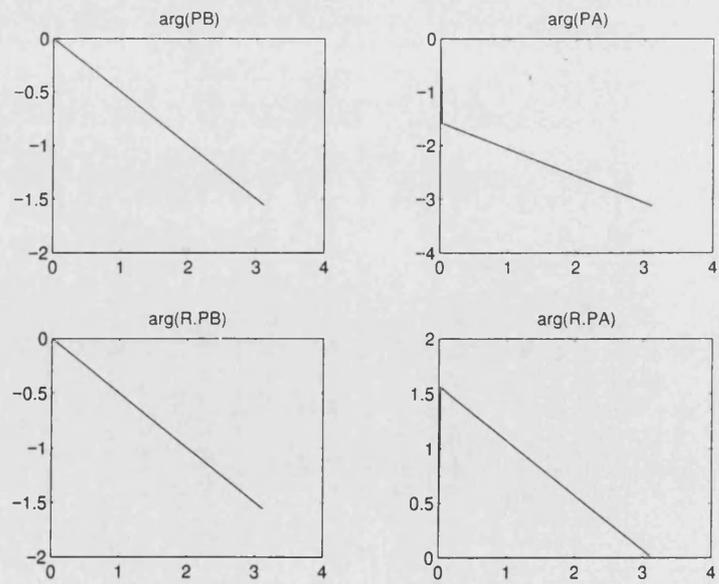


Figura C.8: Respuesta frecuencial: Fase Filtro Haar

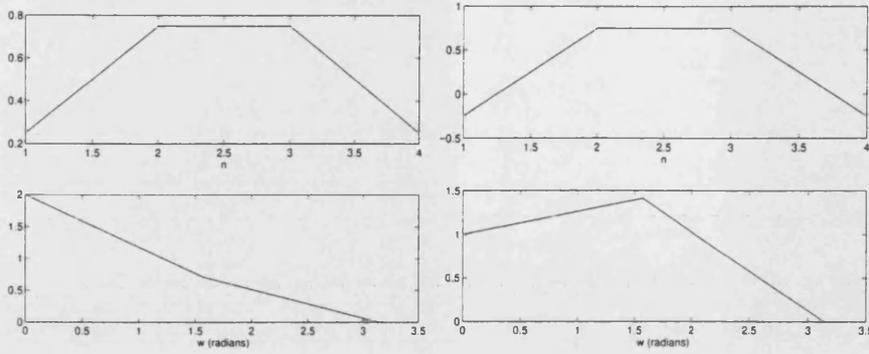


Figura C.9: Función de escala biortogonal(1) de 1ª resolución, análisis y síntesis

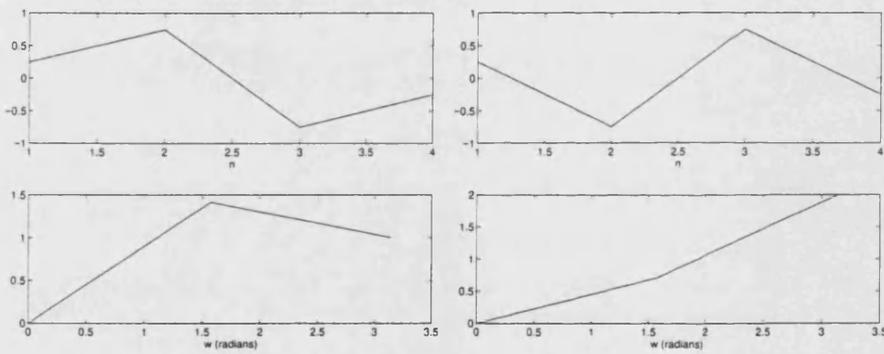


Figura C.10: Función wavelet biortogonal(1) de 1ª resolución, análisis y síntesis

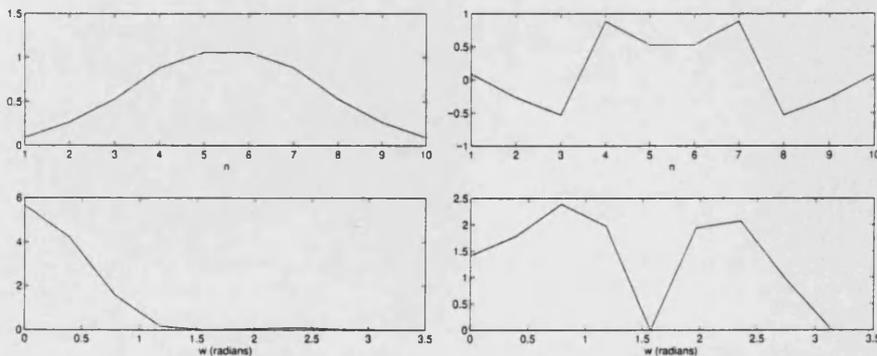


Figura C.11: Función de escala biortogonal(1) de 2ª resolución, análisis y síntesis

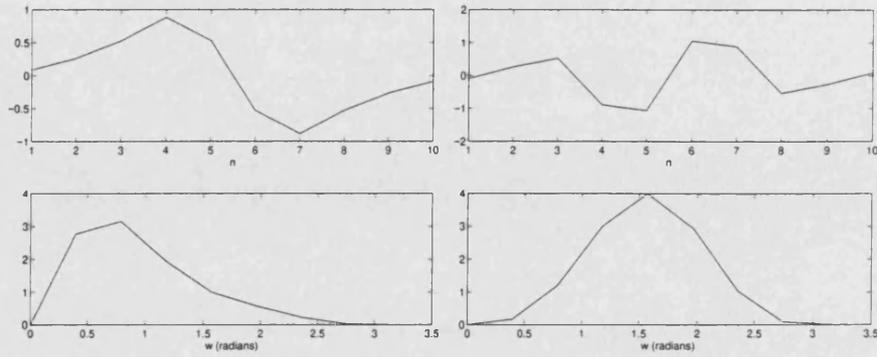


Figura C.12: Función wavelet biortogonal(1) de 2ª resolución, análisis y síntesis

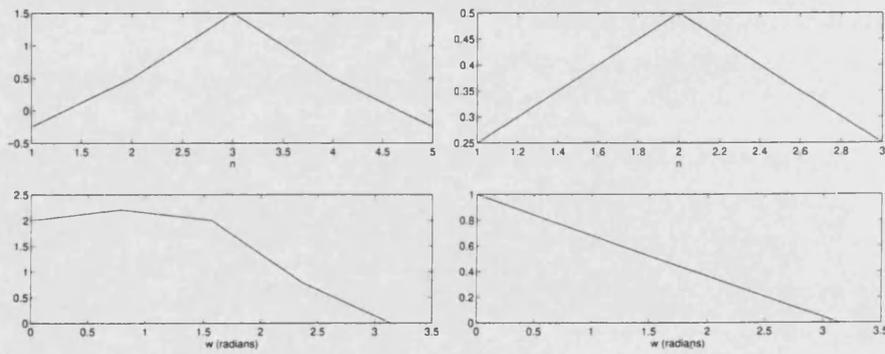


Figura C.13: Función de escala biortogonal(2) de 1ª resolución, análisis y síntesis

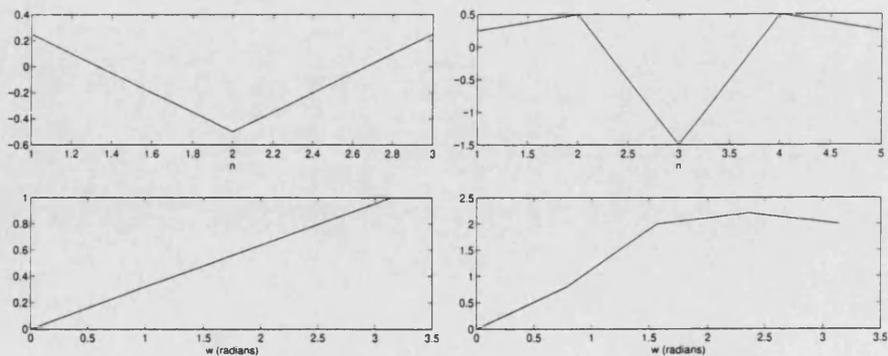


Figura C.14: Función wavelet biortogonal(2) de 1ª resolución, análisis y síntesis

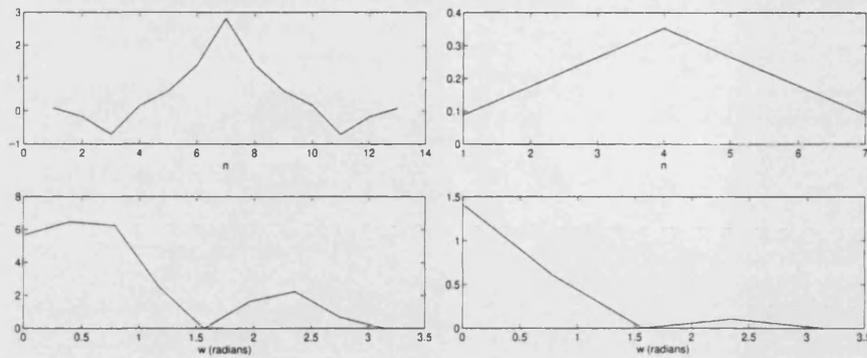


Figura C.15: Función de escala biortogonal(2) de 2ª resolución, análisis y síntesis

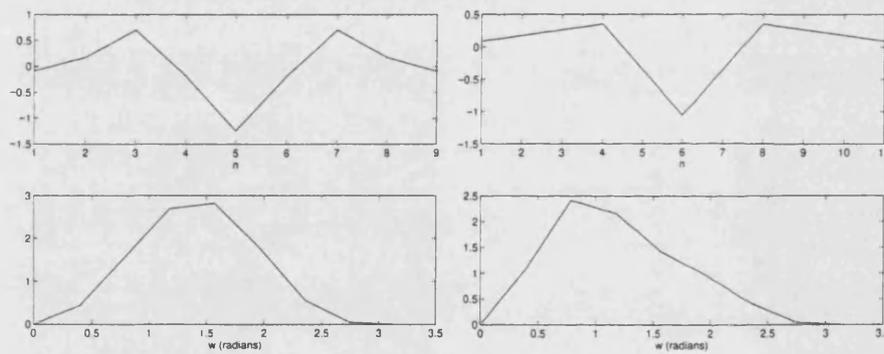


Figura C.16: Función wavelet biortogonal(2) de 2ª resolución, análisis y síntesis

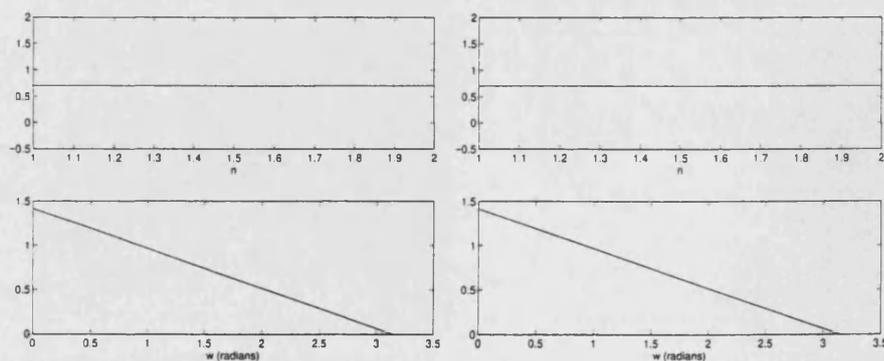


Figura C.17: Función de escala Haar de 1ª resolución, análisis y síntesis

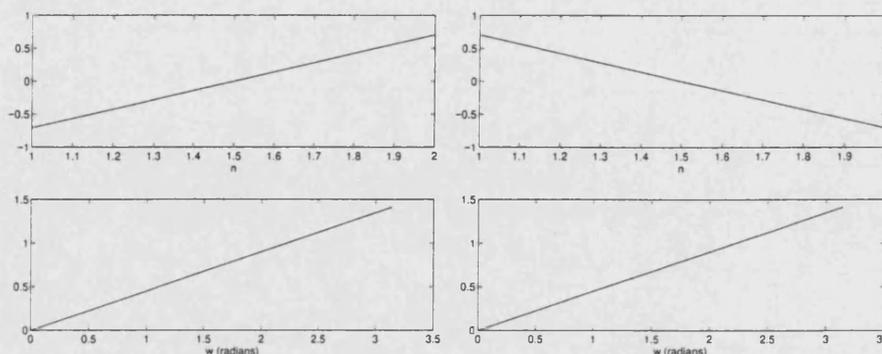


Figura C.18: Función wavelet Haar de 1ª resolución, análisis y síntesis

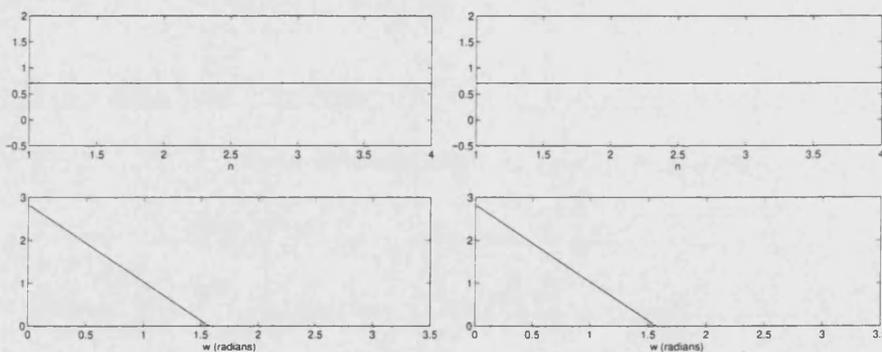


Figura C.19: Función de escala Haar de 2ª resolución, análisis y síntesis

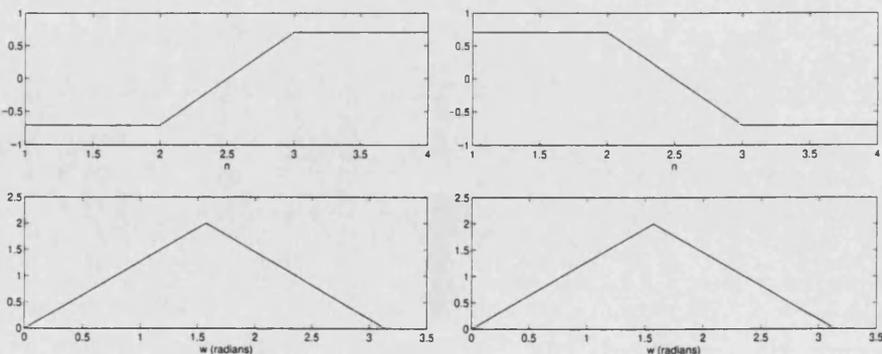


Figura C.20: Función wavelet Haar de 2ª resolución, análisis y síntesis

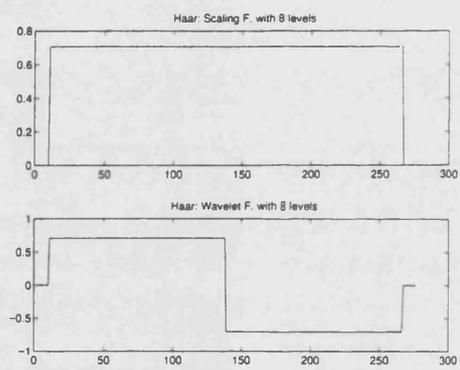


Figura C.21: *Función de escala y síntesis Haar con 8 resoluciones*

## Apéndice D

# Estudio del estimador de MACR para el intervalo de descomposición

Las propiedades de autosimilitud del tráfico multiplexado en los conmutadores, discutido en [Ale95] [WW94] y [WL94], permite estimar el comportamiento del tráfico a diferentes escalas de tiempo; es decir que los estadísticos del tráfico en intervalos de observación pequeños permiten estimar los estadísticos en intervalos de observación mayores o viceversa.

Desde el punto de vista de la transmisión de vídeo en el presente trabajo, el intervalo de interés corresponde al tiempo necesario para la captura de los fotogramas necesarios para realizar una descomposición, que como se analiza en la sección 6.3.1 se fija a 160 ms.

La búsqueda de un estimador con dichas propiedades para poder determinar una aproximación de la tasa media de salida de la fuente (MACR) en un intervalo de 160 ms a partir de transiciones instantáneas de la tasa permitida de salida (ACR), introduce la siguiente expresión

$$MACR = MACR + \alpha(ACR - MACR) \quad (D.1)$$

visto en 7.12 del capítulo 6, y que controla las variaciones de ACR (Allowed Cell Rate) respecto a la media a través del peso  $\alpha$  ( $\alpha < 1$ ) según un filtro paso bajo; con ello se consigue un promediador del ACR instantáneo a diferentes intervalos de agregación, por acumulación de incrementos o decrementos, cada vez más atenuados por potencias de  $\frac{1}{\alpha}$  a medida que se alejen del instante de estimación, respecto al MACR durante su evaluación (cada 160 ms).

Podemos ver en la figura D.1 la estimación realizada a medida que evoluciona el agregado para tres valores diferentes de  $\alpha$ , en este caso  $\frac{1}{8}$ ,  $\frac{1}{16}$ ,  $\frac{1}{128}$  frente a una señal cuadrada. Se observa que a  $\alpha$  pequeño, las estimaciones son más próximas a la media.

La expresión 7.12 analizada en el dominio de Fourier permite determinar la frecuencia de corte del filtro a fin de poder controlar la rapidez de las variaciones de MACR según las fluctuaciones de la entrada ACR. La función de transferencia del

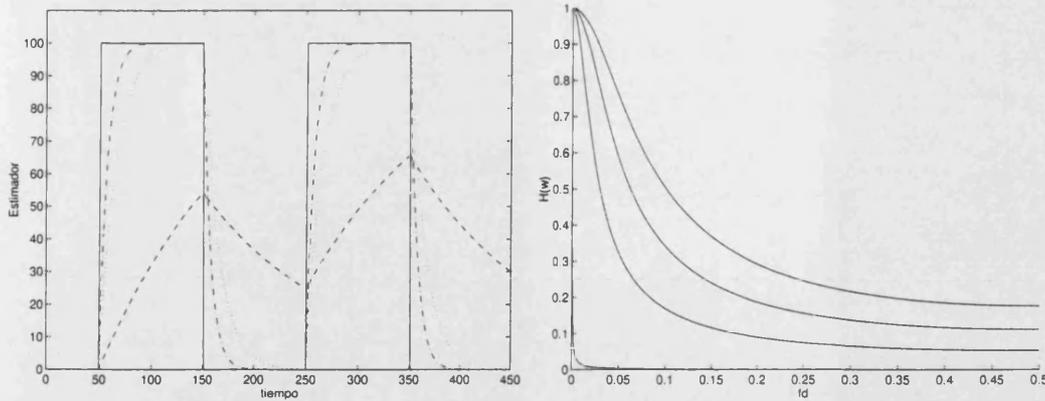


Figura D.1: *Estimación de la tasa media de celdas permitida cuando el ACR es una señal cuadrada y respuesta en frecuencia del estimador para valores de  $\alpha$  de  $\frac{1}{1024}$ , 0.1, 0.2 y 0.3*

filtro 7.12 (ver parte derecha de figura D.1) es

$$H(z) = \frac{\alpha z}{z - 1 + \alpha} \quad (\text{D.2})$$

donde  $z = r e^{jw}$  con  $w = 2\pi f_d$  siendo  $f_d = \frac{f_a}{f_s}$  la frecuencia digital relacionada con la frecuencia de la señal de entrada que es muestreada a  $f_s$  (frecuencia de muestreo, en este caso la tasa de celdas por ms de salida de la fuente).

La representación de la frecuencia de corte digital (cuando la función de transferencia es  $\frac{1}{\sqrt{2}}$ ) para diferentes valores de  $\alpha$  se muestra en la gráfica D.2 y analíticamente se expresa

$$\cos w_c = \frac{\alpha^2 + 2\alpha - 2}{2(\alpha - 1)} \quad (\text{D.3})$$

Para estimar el valor de  $\alpha$  a utilizar en la expresión 7.12 se requiere un estudio del motivo de las transiciones del ACR de entrada, que viene determinado por un conjunto de factores, entre ellos los más relevantes son:

1. el intervalo de tiempo que utiliza el conmutador ATM para actualizar el ancho de banda equivalente asignado a las conexiones activas, que determina el ACR de las fuentes
2. la frecuencia de actualización del ACR según la tasa de llegada de celdas BRM (esta tasa se configura a través del parámetro  $N_{rm}$  en el transmisor ABR [Kim96] y que fija a cada 32 celdas de datos la salida de una RM)
3. los posibles transitorios que se producen con las conexiones y que dependen del tamaño de buffer del conmutador
4. la tasa o velocidad media de salida de la fuente, que se puede estimar por el parámetro MCR (tasa mínima de salida) o ICR (tasa inicial) de la fuente;

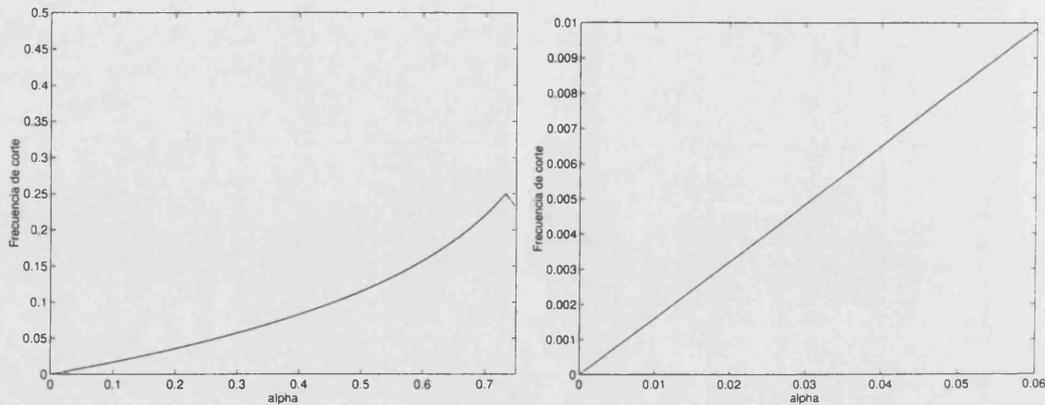


Figura D.2: Frecuencia digital de corte del filtro para la estimación del MACR a partir del ACR en función de  $\alpha$ , con ampliación de detalles para bajas frecuencias

destacar que el valor PCR (tasa de pico de la fuente) es un valor utópico, sólo alcanzable en el momento que la fuente utilice sola los recursos del conmutador

La discusión por tanto del valor  $\alpha$  queda en función de los puntos anteriores. A continuación se describe cada uno de ellos.

Según [Mar97] y como se detalla en el capítulo 10, en el conmutador los periodos de refresco o intervalos de estimación del ancho de banda equivalente se realizan cada 100 celdas procesadas, que para una configuración de cuello de botella con enlace troncal de 10 Mbps (23.58 celdas por ms) se aproxima a 4 ms. Estas modificaciones se llevan a cabo en las fuentes a intervalos mayores, fijados por la tasa de salida de la fuente, aproximadamente 4.716 celdas por ms, y el valor  $N_{rm}$  (32 celdas), con ello cada 6 ms la fuente puede cambiar el ACR.

Por otro lado la duración de los transitorios asociados a la congestión del conmutador están directamente relacionados con el tiempo de vaciado de su buffer, que como se puede ver en el capítulo 10 según el control de flujo utilizado no supera en dichos momentos las 200 celdas, lo cual conlleva un tiempo de 8.48 ms vaciando a la tasa de salida del enlace troncal.

Por tanto, para variaciones de ACR según  $N_{rm}$  y para evitar los transitorios anteriores que no son muy representativos para el comportamiento de la fuente de vídeo, se estima un valor de frecuencia de corte digital de  $f_c = \frac{1}{3 \cdot 32}$  tomando  $f_s = 4.716$  celdas por ms, valor que procede de asumir variaciones más estables o permanentes relativas a tres modificaciones consecutivas del ACR en la fuente.

Para valores pequeños de la frecuencia de corte (en frecuencias digitales menores de 0.01) se puede obtener una expresión aproximada para la frecuencia digital de corte ( $f_c$ ) en función de  $\alpha$

$$f_c = \frac{0.0099}{0.06} \alpha \quad f_c < 0.01 \quad (D.4)$$

y que sustituyendo para  $f_c = \frac{1}{3 \cdot 32}$  permite aproximar  $\alpha = \frac{1}{16}$ ; se escoge potencia de dos porque las multiplicaciones y divisiones en potencias de dos son por lo general más

rápidas computacionalmente.

Además, es estos casos, la mejor opción de  $\alpha$  por tratarse de servicios *best effort* es una estimación optimista del MACR, permitiendo confiar en la buena voluntad de los protocolos subyacentes y los controles de flujo.

# Apéndice E

## Histogramas de la descomposición 3D biortogonal de Miss América

### E.1 Histogramas de cada subbanda

Este anexo recoge los histogramas de la descomposición con filtros biortogonales realizada en el capítulo 7. Como cabe esperar todas las subbandas poseen media nula y una varianza proporcional a la cantidad de información visual aportada. Además las suposiciones realizadas sobre su función de densidad de probabilidad se aproxima con exactitud.

La descripción de las figuras siguientes, de arriba a abajo y de izquierda a derecha, es siguiendo la notación del capítulo 8:

1. figura E.1:  $p_{b1rt1vH}$ ,  $p_{b1rt1Vh}$ ,  $p_{b1rt1VH}$ ;  $p_{a1rt1vh}$ ,  $p_{a1rt1vH}$ ,  $p_{a1rt1Vh}$ ,  $p_{a1rt1VH}$
2. figura E.2:  $p_{b1rt2vH}$ ,  $p_{b1rt2Vh}$ ,  $p_{b1rt2VH}$ ;  $p_{a1rt2vh}$ ,  $p_{a1rt2vH}$ ,  $p_{a1rt2Vh}$ ,  $p_{a1rt2VH}$ , que poseen un comportamiento similar a las de la figura E.1 porque los filtros poseen las mismas propiedades para el mismo nivel de resolución
3. figura E.3  $p_{b2rvH}$ ,  $p_{b2rVh}$ ,  $p_{b2rVH}$ ;  $p_{a2rvh}$ ,  $p_{a2rvH}$ ,  $p_{a2rVh}$ ,  $p_{a2rVH}$
4. figura E.4 dpcm

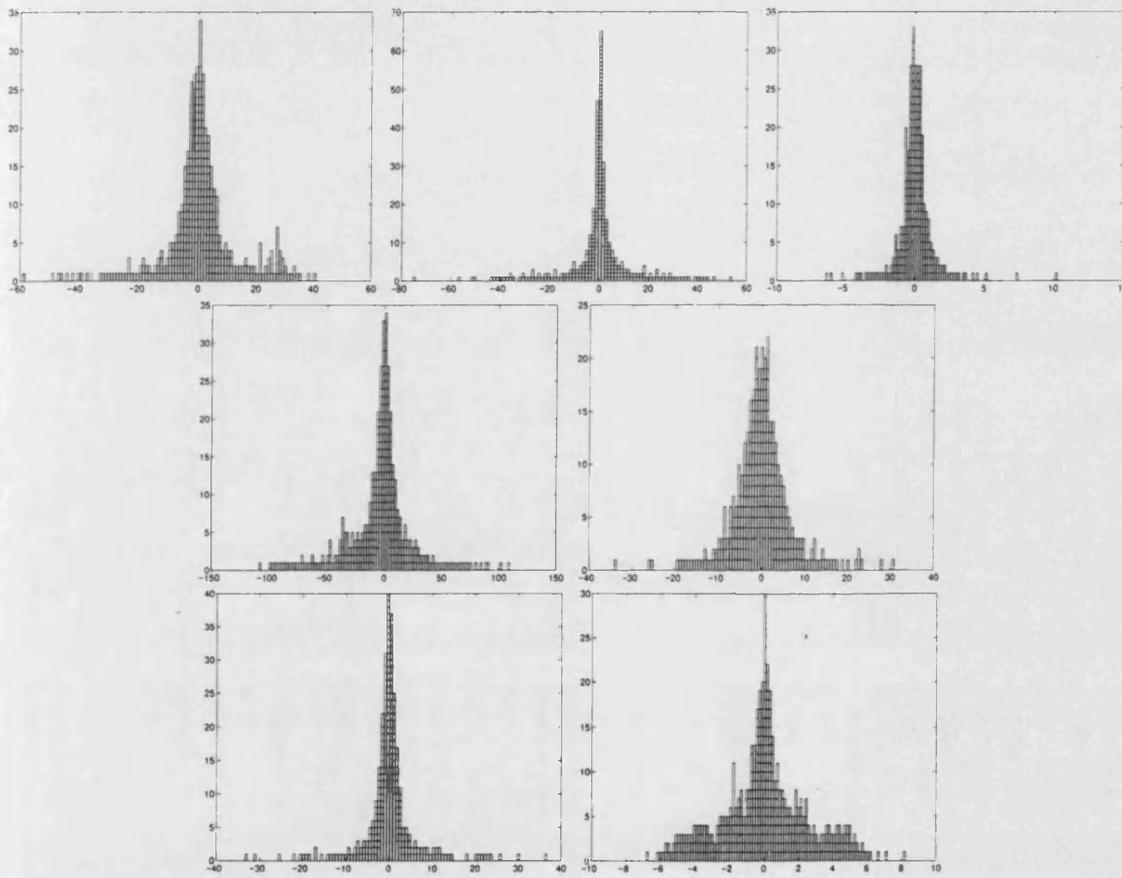


Figura E.1: Histograma de la 1<sup>a</sup> resolución, parte 1

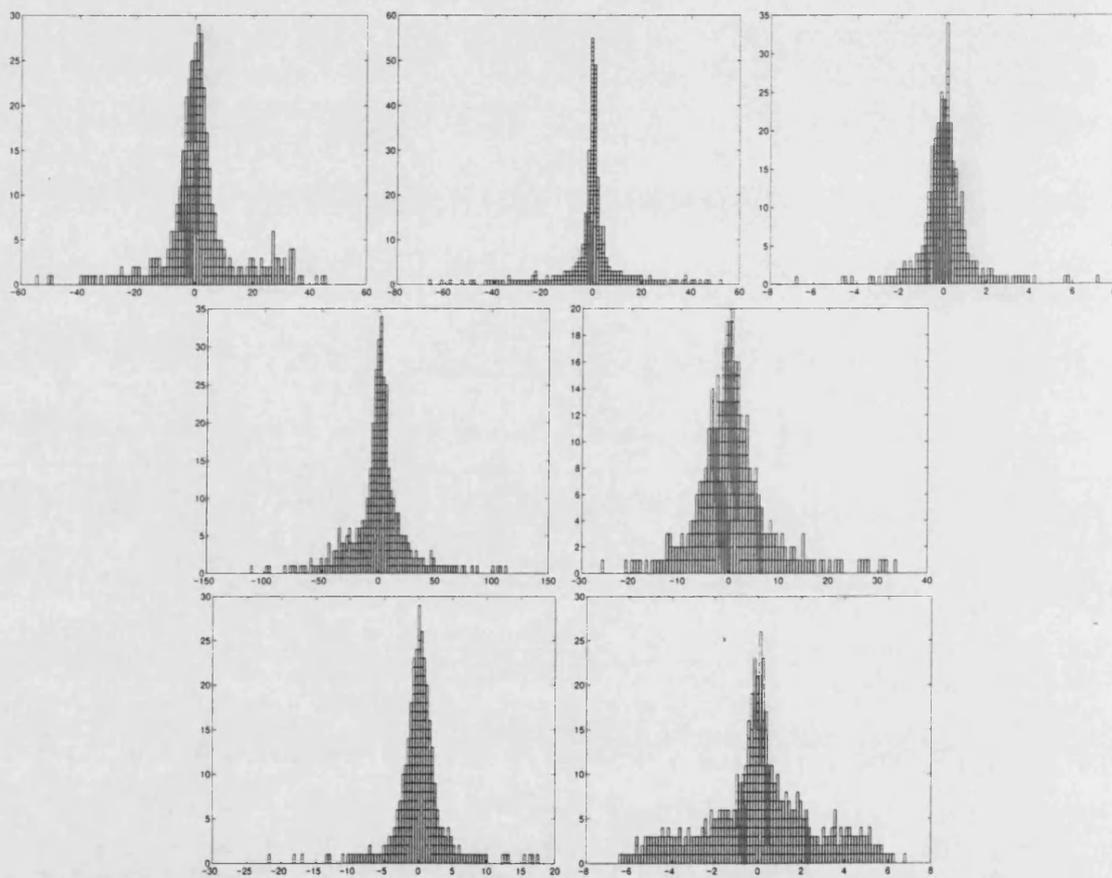


Figura E.2: Histograma de la 1ª resolución, parte 2

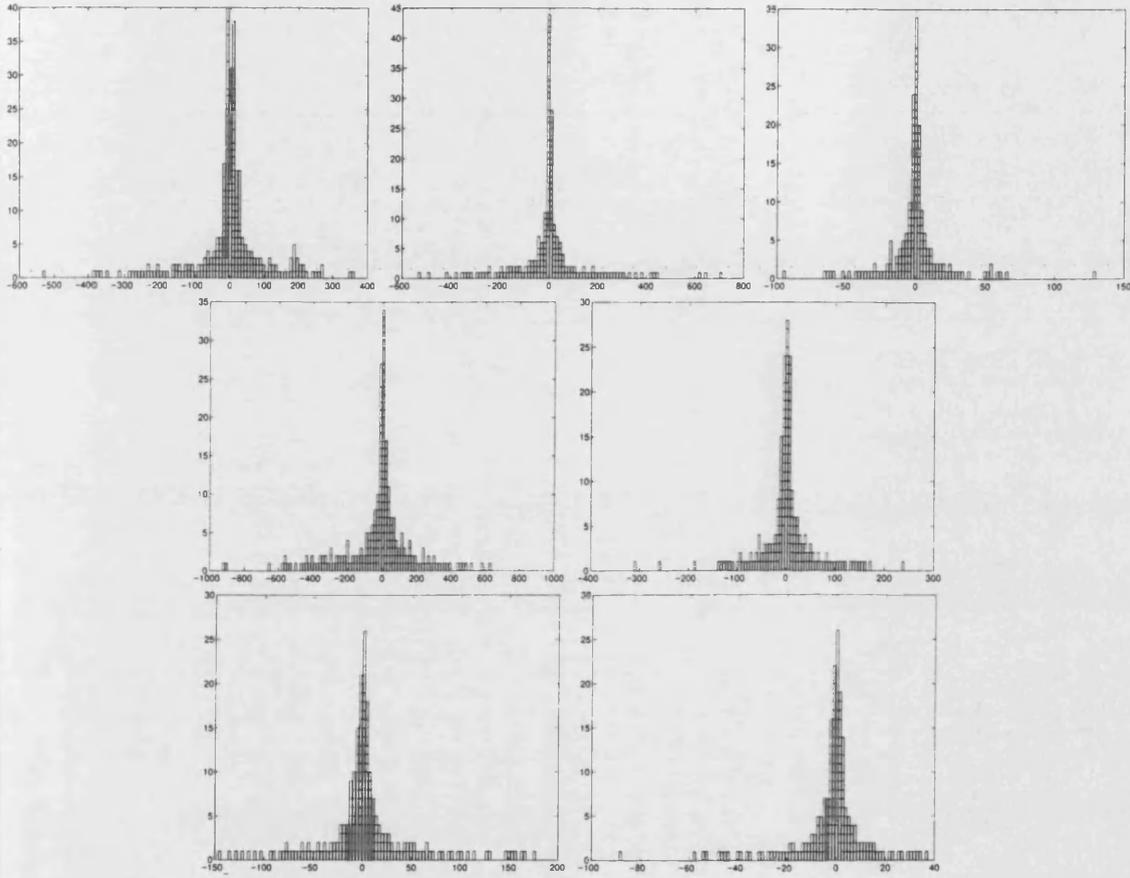


Figura E.3: *Histograma de la 2ª resolución excepto paso baja final*

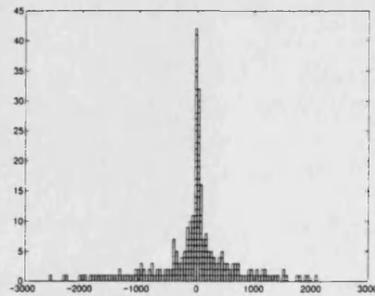


Figura E.4: *Histograma del paso baja final de la 2ª resolución en DPCM*

## Apéndice F

### Tabla de códigos con longitud variable

F.1 Tabla (*Run, Level*) de códigos con longitud variable: signo 's': (0 o +) y (1 o -)

<i>Códigos de longitud variable</i>	<i>run</i>	<i>level</i>	<i>Códigos de longitud variable</i>	<i>run</i>	<i>level</i>
10	Final de bloque		0000 101 s	9	1
11 s	0	1	0000 01	Salida	
011 s	1	1	0010 0110 s	0	5
0100 s	0	2	0010 0001 s	0	6
0101 s	2	1	0010 0101 s	1	3
0010 1 s	0	3	0010 0100 s	3	2
0011 1 s	3	1	0010 0111 s	10	1
0011 0 s	4	1	0010 0011 s	11	1
0001 10 s	1	2	0010 0010 s	12	1
0001 11 s	5	1	0010 0000 s	13	1
0001 01 s	6	1	0000 0010 10 s	0	7
0001 00 s	7	1	0000 0011 00 s	1	4
0000 110 s	0	4	0000 0010 11 s	2	3
0000 100 s	2	2	0000 0011 11 s	4	2
0000 111 s	8	1	0000 0010 01 s	5	2
			0000 0011 10 s	14	1

Tabla F.1: Tabla de códigos de longitud variable: s (0,1) signo (+,-)

<i>Códigos de longitud variable</i>	<i>run</i>	<i>level</i>	<i>Códigos de longitud variable</i>	<i>run</i>	<i>level</i>
0000 0011 01 s	15	1	0000 0000 0110 00 s	0	23
0000 0010 00 s	16	1	0000 0000 0101 11 s	0	24
0000 0001 1101 s	0	8	0000 0000 0101 10 s	0	25
0000 0001 1000 s	0	9	0000 0000 0101 01 s	0	26
0000 0001 0011 s	0	10	0000 0000 0101 00 s	0	27
0000 0001 0000 s	0	11	0000 0000 0100 11 s	0	28
0000 0001 1011 s	1	5	0000 0000 0100 10 s	0	29
0000 0001 0100 s	2	4	0000 0000 0100 01 s	0	30
0000 0001 1100 s	3	3	0000 0000 0100 00 s	0	31
0000 0001 0010 s	4	3	0000 0000 0011 000 s	0	32
0000 0001 1110 s	6	2	0000 0000 0010 111 s	0	33
0000 0001 0101 s	7	2	0000 0000 0010 110 s	0	34
0000 0001 0001 s	8	2	0000 0000 0010 101 s	0	35
0000 0001 1111 s	17	1	0000 0000 0010 100 s	0	36
0000 0001 1010 s	18	1	0000 0000 0010 011 s	0	37
0000 0001 1001 s	19	1	0000 0000 0010 010 s	0	38
0000 0001 0111 s	20	1	0000 0000 0010 001 s	0	39
0000 0001 0110 s	21	1	0000 0000 0010 000 s	0	40
0000 0000 1101 0 s	0	12	0000 0000 0011 111 s	1	8
0000 0000 1100 1 s	0	13	0000 0000 0011 110 s	1	9
0000 0000 1100 0 s	0	14	0000 0000 0011 101 s	1	10
0000 0000 1011 1 s	0	15	0000 0000 0011 100 s	1	11
0000 0000 1011 0 s	1	6	0000 0000 0011 011 s	1	12
0000 0000 1010 1 s	1	7	0000 0000 0011 010 s	1	13
0000 0000 1010 0 s	2	5	0000 0000 0011 001 s	1	14
0000 0000 1001 1 s	3	4	0000 0000 0001 0011 s	1	15
0000 0000 1001 0 s	5	3	0000 0000 0001 0010 s	1	16
0000 0000 1000 1 s	9	2	0000 0000 0001 0001 s	1	17
0000 0000 1000 0 s	10	2	0000 0000 0001 0000 s	1	18
0000 0000 1111 1 s	22	1	0000 0000 0001 0100 s	6	3
0000 0000 1111 0 s	23	1	0000 0000 0001 1010 s	11	2
0000 0000 1110 1 s	24	1	0000 0000 0001 1001 s	12	2
0000 0000 1110 0 s	25	1	0000 0000 0001 1000 s	13	2
0000 0000 1101 1 s	26	1	0000 0000 0001 0111 s	14	2
0000 0000 0111 11 s	0	16	0000 0000 0001 0110 s	15	2
0000 0000 0111 10 s	0	17	0000 0000 0001 0101 s	16	2
0000 0000 0111 01 s	0	18	0000 0000 0001 1111 s	27	1
0000 0000 0111 00 s	0	19	0000 0000 0001 1110 s	28	1
0000 0000 0110 11 s	0	20	0000 0000 0001 1101 s	29	1
0000 0000 0110 10 s	0	21	0000 0000 0001 1100 s	30	1
0000 0000 0110 01 s	0	22	0000 0000 0001 1011 s	31	1



## Apéndice G

# Configuración del sistema en el simulador DANAE

Los diferentes bloques que forma tanto el sistema de emisión como de recepción se muestran en las figuras G.2 y G.3.

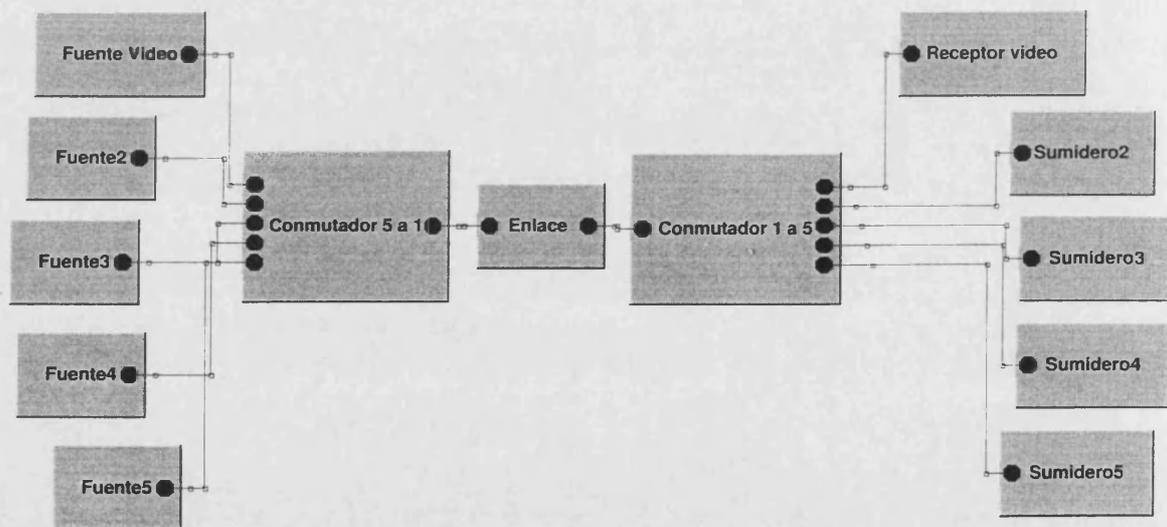


Figura G.1: Configuración de red con dos conmutadores y enlace troncal cuello de botella en el simulador DANAE

La configuración de cuello de botella (figura G.1) está formada por:

- dos conmutadores que implementan el control de flujo basado en tasa explícita
- los enlaces tanto de acceso (desde las fuentes al conmutador) como troncales (entre conmutadores)

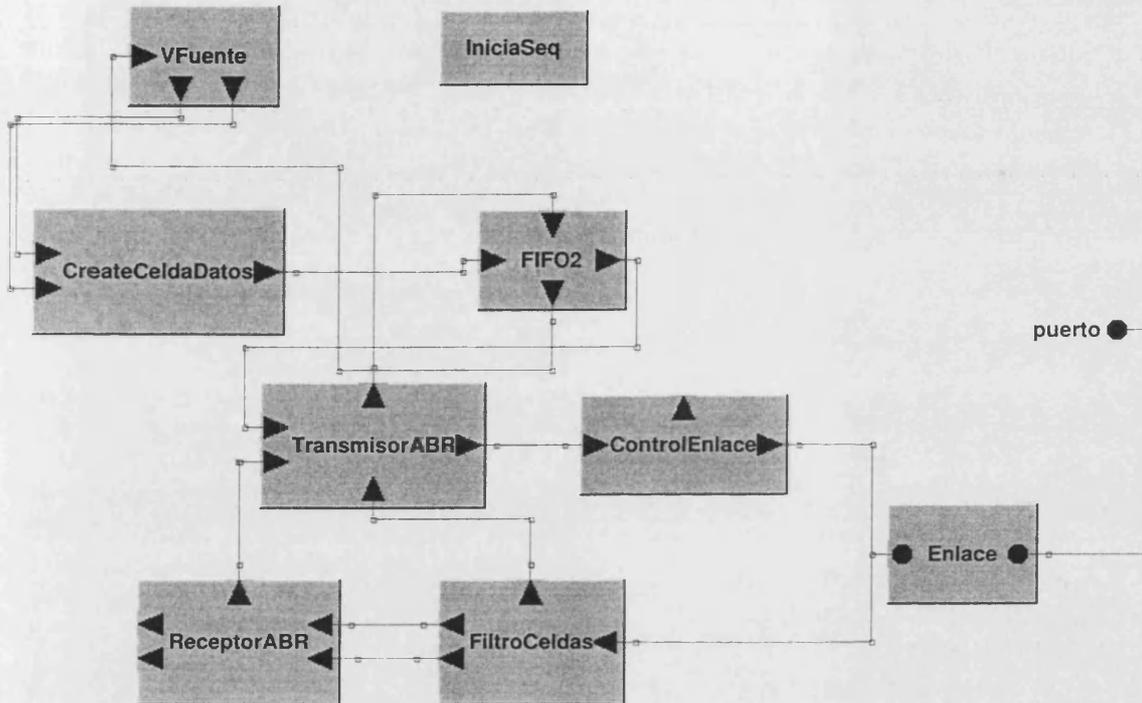


Figura G.2: Esquema de bloques del sistema de emisión en el simulador DANAE

A los conmutadores se conectan 5 fuentes a través de conexiones independientes, entre ellas el sistema de emisión y de recepción de vídeo, y el resto un conjunto de fuentes generadoras de tráfico de fondo conocido.

El sistema de emisión (figura G.2) está formado por 8 bloques que a continuación se describen:

- IniciaSeq, que abre el fichero de trazas descrito en la sección 8.1
- VFuente o módulo de proceso como se describe en la sección 6.3 que tiene como entrada el tamaño de ocupación de la cola FIFO2 (para saber la porción de subbandas transmitidas) y como salida los puertos para generar las diferentes celdas que forman una PDU, un puerto para celdas con PTI=0 y el otro puerto para celdas con PTI=1 o final de PDU; este bloque controla la inicialización del bloque FIFO2 a través de un evento enlazable
- CreateCeldaDatos es un bloque que configura la cabecera de la celda con el identificador de la conexión

- FIFO2 es una memoria de tamaño fijo con la propiedad de ser inicializada por el módulo VFuente y cuya salida se conecta al bloque TransmisorABR, el cual va señalizando los instantes de salida por el puerto de entrada a la FIFO2 en la parte superior
- TransmisorABR es un bloque que implementa la máquina de estado del transmisor ABR según la tasa explícita realimentada por el control de flujo. Tiene tres entradas: una de la FIFO2 con las celdas a transmitir, otra para el ACR calculado en el bloque ReceptorABR y otra para la entrada de las celdas FRM procedente del bloque FiltroCeldas para generar las BRM. Tiene dos salidas, una para indicar permisos de salida de celda a la FIFO2 y otro para salida al bloque Enlace por el bloque ControlEnlace
- ControlEnlace es un bloque que ranura en slots de tiempo de duración de celda según el ancho de banda del enlace de acceso
- Enlace es un bloque que conecta directamente con el conmutador y que implementa el enlace de acceso, retrasando cada celda de salida 5 us por Km
- FiltroCeldas es un bloque que recibe todas las celdas de vuelta al emisor, de manera que tanto las BRM como de Datos las pasa al bloque ReceptorABR (en este caso como la conexión es unidireccional el puerto de Datos no lleva información) y las FRM las pasa al bloque TransmisorABR para su retorno en formato BRM
- ReceptorABR es un bloque que en este caso sólo se utiliza para calcular el ACR según el control de flujo

El sistema de recepción (figura G.3), de forma similar al sistema de emisión, está formado por 9 bloques, 5 de ellos comunes: Enlace, ControlEnlace, ReceptorABR, TransmisorABR y FiltroCeldas. Los 4 bloques diferentes se describen a continuación:

- Mux2a1 es un bloque que multiplexa en un mismo flujo las celdas tanto PTI=0 como PTI=1 y como las celdas llegan ordenadas a través de la conexión, la multiplexación almacena en la FIFO1 las PDU completas (si no han habido pérdidas por la red)
- FIFO1 es un bloque de memoria, que a diferencia de FIFO2 no incorpora ninguna inicialización, cuya funcionalidad es la de reducir la variabilidad entre los tiempo de llegada de las celdas
- Mux1a2 es el bloque inverso a Mux2a1 descrito antes
- VDestino es el módulo de proceso en el receptor descrito en 6.4, que tiene dos puertos como entradas, tanto para las celdas con PTI 0 y 1, y como salida un puerto para la señal de sincronismo en la FIFO1 para ir vaciando

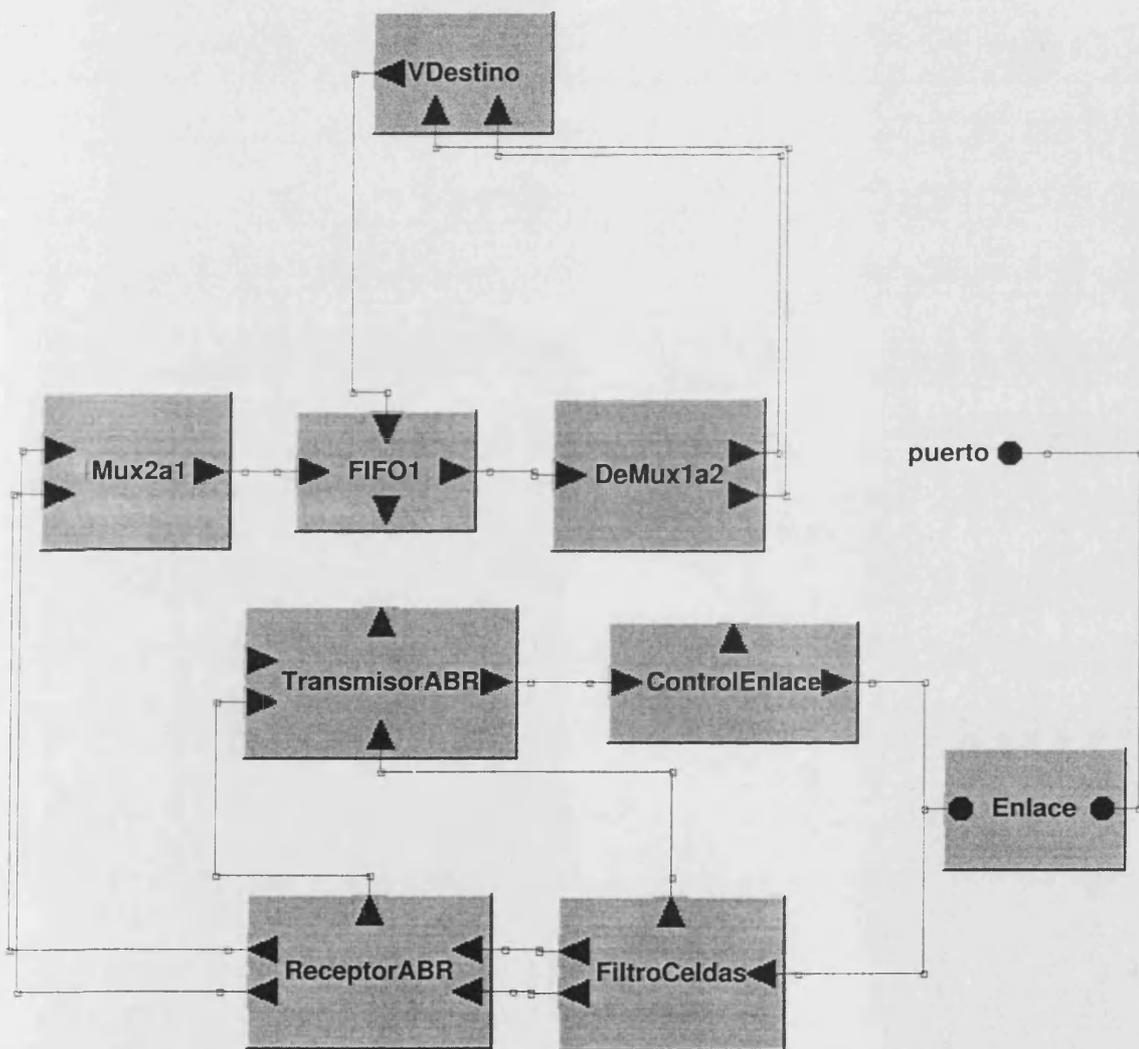


Figura G.3: Esquema de bloques del sistema de recepción en el simulador DANAE

**Parte VII**  
**Bibliografía**

# Bibliografía

- [AAL93] Rec. i363 b-isdn atm adaptation layer spec. *ITU-T*, 1993.
- [Ada95] G.J. Armitage K.M. Adams. How efficient is ip over atm anyway? *IEEE Network*, pages 18–26, January 1995.
- [Ale95] Ross Alexander. Monitoring, analysis and simulation of packet switched network traffic, January 1995. Thesis Master of Science.
- [Art95] J. Malo A. Pons J. Artigas. Bit allocation algorithm for codebook design in vector quantization fully based on hvs nonlinearities for suprathreshold contrasts. *Electronic Letters*, 31(15):1222–1224, 1995.
- [Art97] J. Malo A. Pons A. Felipe J. Artigas. Characterization of the human visual system threshold performance by a weighting function in the gabor domain. *Journal of Modern Optics*, 44(1):127–148, 1997.
- [B.98] Beferull B. Compresión de vídeo: Codificación subbanda. Technical report, Dpto. Informática y electrónica. Universitat de Valencia, 1998.
- [Bak89] J.C. Darragh R. Baker. Fixed distorsion subband coding of images for packet-swithced networks. *IEEE Journal on Selected Topics in Commun.*, 7:789–800, 1989.
- [BM93] H Shi B Macq. Perceptually weighted vector quantization in the dct domain. *Electronic Letters*, 29(15):1382–1384, 1993.
- [Cha94] A. Charny. An algorithm for rate allocation in packet switching networks with feedback, May 1994. Master's Thesis.
- [Che96a] C Chou C Chen. A perceptually optimized 3d subband codec for video communicatio over wireless channels. *ieeecsvt*, 1996.
- [Che96b] Chun-Hsie Chou Chi-Wei Chen. A perceptually optimized 3d subband codec for video compression over wireless channels. In *IEEE. Transactions on circuits and systems for video tech.*, pages 143–156, 1996.
- [Com91] D.L. Comer, D. Stevens, editor. *Internetworking with TCP/IP Vol II*. Prentice Hall, USA, 1991.

- [Dal94] D Hughes P Daley. Limitations of credit based flow control. *ATM Forum*, September 1994. 94-0776.
- [Dau88] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure and Applied Mathematics*, 41:909–996, 1988.
- [Dau90] I. Daubechies. The wavelet transform, time-freq. localization and signal analysis. *IEEE Trans. Information Theory*, 36,5:961–1005, 1990.
- [Dau92] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.
- [Far92] N. TanabeÑ. Farvadin. Subband image coding using entropy-coded quantization over noisy channels. *IEEE Journal on Selected Areas in Commun.*, 10(5):926–942, June 1992.
- [Far95] C. Podilchuck N. JayantÑ. Farvadin. Three dimensional subband coding of video. *ieeep*, 1995.
- [Flo94] A. Romanov S. Floyd. Dynamics of tcp traffic over atm networks. *ACM SIGCOMM*, September 1994.
- [Gab46] D. Gabor. Theory of communication. *Journ. IEE*, 1946.
- [Gal97] Javier Pérez J.C. Galiana. P.f.c. simulador de redes orientado a objetos para evaluación de prestaciones en entornos atm. Technical report, E.T.S.I.Telecomunicación U.P. Valencia, Noviembre 1997.
- [GG92] A. Gersho and R. M. Gray. *Vector quantization and Signal Compression*. Norwell. MA: Kluwer Academic, 1992.
- [Goy98] S. Kalyanaraman B. Vandalore R. Jain R. Goyal. Performance of tcp over abr with long-range dependent vbr background traffic over terrestrial and satellite atm networks. *INFOCOM'98*, 1998.
- [Gra90] R. M. Gray. *Source Coding Theory*. Kluwer Academic Publishers, 1990.
- [GS96] Nguyen Gilbert Strang, T, editor. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, USA, 1996.
- [Hee92] D. J. Heeger. Normalization of cell responses in cat visual cortex visual neuroscience. *Visual Neuroscience*, 9:181–192, 1992.
- [Hee94] D. J. Heeger. The representation of the visual stimuli in primary visual cortex. *Current Directions in Psychological Science*, 1994.
- [Hlu91] N. Yin M.G. Hluchyj. A dynamic rate control mechanism for coded traffic in a fast packet network. *Selected Areas Communication*, 9-2:1003–1012, September 1991.

- [H.S89] K.N. Ngan K.S. Leong H.Singh. Adaptive cosine transform coding of images in perceptual domain. *IEEE Trans. Acoust., Speech and Signal Proc.*, 37(1):1743-1749, November 1989.
- [HSM89] D. H. Staelin H. S. Malvar. The lot: Transform coding without blocking effects. *IEEE Trans. Acoust., Speech and Signal Proc.*, 37:553-559, 1989.
- [HW68] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 1968.
- [IK85] M. Kunt A. Ikonomopoulos and M. Kocher. Second-generation image coding techniques. *Proceedings of the IEEE*, 1985.
- [IT95] ITU-T. World telecommunication development report. Technical report, <http://www.itu.ch.WTDR95/toc.html>, 1995. Table of Contents.
- [ITU96] Rec. i371 traffic control and congestion control in b-isdn atm. *ITU-T*, May 1996.
- [Jac96] H. Schulzrinne S. Casner R. Frederick V. Jacobson. Rfc1890. Technical report, Audio-Video Transport Working Group, January 1996.
- [Jai89] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.
- [Jai90] K. Ramakrishnana R. Jain. A binary feedback scheme for congestion avoidance in computer networks. *ACM Transactions on Computer Systems*, 8-2:158-181, 1990.
- [Jai92] R Jain. *The art of computer systems performance analysis*. John Willey and sons, 1992.
- [Jai94] A. Charny D. Clarck R. Jain. Congestion control with explicit rate indication. *ATM Forum*, July 1994. 94-0692.
- [Jai97] Raj Jain. Networking trends and their impact. Technical report, The Ohio State University, 1997. <http://www.cis.ohio-state.edu/ALT126jain/cis788-97>.
- [Jam96] R. Braden L.Zhang S.Berson S. Herzog S. Jamin. Resource reservation protocol. Technical report, Internet Draft, November 1996.
- [Jam97] R. Braden L.Zhang S.Berson S. Herzog S. Jamin. Resource reservation protocol. Technical report, Internet Draft, RFC 2205, September 1997.
- [Jay92] N. Jayant. Signal compression: technology targets and research directions. *IEEE Journal on Selected Areas in Commun.*, 1992.
- [Jen93] L Harris M Jenkin, editor. *Spatial Vision in humans and robots*. Cambridge University Press, USA, 1993.

- [JM97] J Artigas J Malo, A Pons. Subjective image fidelity metric based on bit allocation of the hvs in the dctdomain. *Image and vision computing*, 31(15):535–548, 1997.
- [kel79] D. H. Kelly. Motion and vision ii. stabilized spatio-temporal threshold surface. *Journal of the Opt. Soc. of Am.*, 69(10):1340–1349, 1979.
- [Kim96] R. Jain S. Kalyanaraman S. Fahmy R. Goyal S. Kim. Source behavior for atm abr traffic management: An explanation. *IEEE Communications Magazine*, pages 50–57, November 1996.
- [Kum96] Vinay Kumar, editor. *MBone. Interactive Multimedia on the Internet*. New Riders Publishing, España, 1996.
- [LeG88] R Ansari H Gaggioni D LeGall. HDTV coding using a nonrectangular subband decomposition. *Proc. of the SPIE*, pages 821–824, November 1988.
- [Llo82] S P Lloyd. Least square quantization in pcm. *IEEEIT*, 1982.
- [Mal89a] S Mallat. Multiresolution approximations and wavelet orthonormal bases of  $l_2(\mathbb{R}^n)$ . *Trans. Amer. Math. Soc.*, 315,1:69–87, 1989.
- [Mal89b] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 1989.
- [Mal92] H. S. Malvar. Extended lapped transforms: Properties, applications, and fast algorithms. *IEEE Trans. on Signal Process.*, 1992.
- [Man96] A. Dermirhjis B. Edwards B. Braden S. Berson M. Perez A. Mankin. Rsvp and atm signalling. *ATM Forum*, January 1996. 96-0258.
- [Mar94] Maria Cinta Puell Marin. Codificación de la señal visual. Technical report, Escuela Universitaria de Optica y Optometria de Madrid, 1994.
- [Mar97] Jorge Martínez. Provisión de servicios de datos sin conexión en la rdsi-ba, March 1997. Tesis doctoral.
- [Max60] J. Max. Quantizing for minimum distortion. *IRE Trans. on Inform. Theory*, 1960.
- [Mor97] D. Campillo X. Hesselbach O. Morillo. Estudio de la calidad de servicio de un alisador de tráfico destinado a aplicaciones de audio y vídeo. *URSI*, 1997.
- [MSZ92] S. Mallat S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14,7:710–732, 92.
- [NJS93] J. Johnston N. Jayant and R. Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 1993.

- [Nqu91] D.L. McLaren D.T. Nguyen. Removal of subjective redundancy from decoded images. *Proc. Inst. Elec. Eng.*, 138(5):345-350, October 1991.
- [NRT91] A. Navarro R. Taberbero. Gaussian wavelet transform: two alternatives fast implementations for images. *Mult. Systm. and Sig. Process*, 2:421-436, 1991.
- [Pre97] R Pressman. *Ingeniería del software*. Mc Graw Hill, 1997.
- [RH96] K. R. Rao and J. J. Hwang. *Techniques and Standards for Image , Video and Audio Coding*. Signal processing series. Prentice Hall, New Jersey, 1996.
- [RJ91] M. Rabbani and P. Jones. *Digital Image Compression Techniques*. Bellingham, WA: SPIE Press, 1991.
- [R.M79] Haralick R.M. Statistical an structural approaches to texture. In *Proc. of the IEEE*, pages 786-804, 1979.
- [Rob94] L Roberts. Enhanced prca proportional rate-control algorithm. *ATM FORUM*, August 1994. 94-0735 R1.
- [Run96] D. Montgomery G. Runger. *Probabilidad y estadística aplicada a la ingeniería*. Mc Graw Hill, 1996.
- [R.W96] Oppenheim A. Shafer. R.W., editor. *Discrete-Time signal processing*. Prentice Hall International, USA, 1996.
- [RY90] K.R. Rao and P. Yip. *Discrete cosine transform: algorithms, advantages, applications*. San Diego, CA: Academic Press, 1990.
- [Sal97] X. Hesselbach S. Sallent. Estudio del control de tasa de la fuente en servicio abr para aplicaciones de audio y vídeo. *Jornadas de Ingeniería Telemática, Bilbao*, 1997.
- [Sch96] H. Schulzrinne. Rfc1890. Technical report, Audio-Video Transport Working Group, January 1996.
- [Sha48] C. E. Shannon. A mathematical theory of communication. *Comm. ACM*, 27:398-403, 1948.
- [She94] B. Braden D. Clark S. Shenjer. Rfc1633. Technical report, NetWorking Group, June 1994.
- [SLG95] Kamran Shafri and Alberto Leon-Garcia. Estimation of shape parameter for generalized gaussian distributions in subband decomposition of video. *IEEE Trans. on Circuits and Systems for video tech.*, 5(1):52-56, 1995.
- [SN84] N. S. Jayant and P. Noll. *Digital Coding of Waveforms*. Prentice-Hall, Englewood-Cliffs NJ, 1984.

- [Su96] N. Golmie A. Koeing D. Su. The nist atmnetwork simulator. operation and programming. Technical report, NIST Internal Report, August 1996.
- [Sus96] J. Susbielle, editor. *Telefonía en Internet*. Gestión 2000, España, 1996.
- [Tek97] Tektronix. Convergence seminar tour 97. Technical report, Tektronix, 1997. Seminar.
- [T.L97] K.Ramakrishnan T.Lakshaman, P. Mishra. Transporting compressed video over atm networks with abr feedback control. *ATM Forum*, 1997.
- [Vai93] P. P. Vaidyanathan. *Multirate systems and filter banks*. Prentice Hall, 1993.
- [Vet89] G. Karlsson M. Vetterli. Packet video and its integration into the network architecture. *ieeejosaic*, 1989.
- [VK95] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Signal processing series. Prentice Hall, 1995.
- [Wat87a] A. B. Watson. The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics and Image Processing*, 1987.
- [Wat87b] A. B. Watson. Efficiency of a model human image code. *Journal of the Opt. Soc. of Am.*, 1987.
- [Wat90] A. B. Watson. Perceptual components architecture for digital video. *josa*, 1990.
- [WL94] W. Willinger W.E. Leland. Statistical analysis and stochastic modeling of self-similar data traffic. *International Tele-Traffic Conference*, 14:319–328, 1994.
- [Woj94] L Wojnaroski. Base-line text for traffic management sub-working group. *ATM Forum*, October 1994. 94-0394r5.
- [WW94] James Gordon Walter Willinger, Ashok Erramilli. Applications of fractals in engineering for realistic traffic processes. *International Tele-Traffic Conference*, 14:35–44, 1994.

## Publicaciones relacionadas con la tesis

- [FM98a] S. Felici J. Martínez B. Beferull G. Martín Estudio del comportamiento dinámico de redes de paquetes con control de flujo cerrado para transmisión de vídeo digital comprimido. In *URSI XIII Symposium Nacional*, Pamplona, Spain 1998.
- [FM98b] S. Felici J. Martínez B. Beferull G. Martín Métodos adaptativos de compresión de vídeo para sistemas de transmisión dinámicos con controles de flujo cerrado. In *URSI XIII Symposium Nacional*, Pamplona, Spain 1998.
- [FB98] S. Felici B. Beferull J. Martínez G. Martín An adaptive video compression algorithm based on wavelet filter banks for best effort networks with closed flow control. In *SoftCOM IEEE Conference on Software in Telecommunications and Computer Networks*, Bari, Italy 1998. *Submitted*.

# Indice de Materias

## — A —

aliasing nulo, 57  
asignación de bits, 105, 128  
ATM  
  AAL, 22  
  AAL5+, 89  
  ICR, 139  
  MACR, 104, 207  
  MCR, 93, 139  
  NNI, 22  
  PCR, 139  
  PDU, 23  
  SDU, 22  
  UNI, 22

## — B —

bastones, 64  
Bit Allocation, 105

## — C —

células  
  amacrinas, 65  
  bipolares, 64  
  complejas, 67  
  estrelladas, 67  
  ganglionares, 65  
  hipercomplejas, 68  
  horizontales, 65  
  selectivas direccionales, 67  
  simples, 67  
caótico, 103  
campos receptores, ON-OFF, 65

capas corticales, 64  
codificación  
  aritmética, 30  
  de entropía, 88  
  Huffman, 88, 90  
  run level, 88, 90  
  sin pérdidas, 30  
coeficiente  
  ráfaga, 123  
comportamiento  
  de autosimilitud, 103  
conos, 64  
cortex cerebral, 64  
corteza asociativa, 68  
cuantificación, 30  
cuatificación  
  Lloyd-Max, 106  
cuerpo geniculado, 66

## — D —

descomposición, 30  
dirección  
  broadcast, 17  
  multicast, 17  
  multidifusión, 17  
distorsión nula, 57  
Domain Name Servers, 17  
DPCM, 32

## — E —

efecto  
  acantilado, 94  
  bloque, 79

enventanado, 79  
 Gibbs, 81  
 equal slope, 108  
 error cuadrático, MSE, 30  
 estimación de movimiento, 70, 84, 95,  
 112  
 estimación de movimiento, 33

## — F —

factor de compresión, 129  
 FDDI, 188  
 filtro  
   análisis, 55  
   biortogonal, 53  
   de predicción en DPCM, 100  
   espejo, 53  
   Haar, 53, 60  
   síntesis, 55  
 fotograma, 36  
 fotogramas  
   interframe P, 39  
   intraframes I, 38  
 fractal, 103  
 fuente  
   de Markov, 31  
 función  
   madre wavelet, 46

## — G —

Gabor, función, 71

## — H —

H261  
   GOB, 36  
   H.263, 36  
 HDTV, 16  
 Heisenberg, principio de, 45  
 hemisferio, 64  
 hemiretina, 64

Huffman Modificado, 186

## — I —

información, 82  
 IP, 17

## — J —

jitter, 90  
 JND, Just Noticeable Distorsion, 82

## — M —

magnosistema, 66  
 mapeado, 30  
 minimización por Lagrange, 107  
 movimiento  
   macro bloque, 36  
   MAD, 34  
 MPEG, 37  
   MPEG1, 37  
   MPEG2, 37  
   MPEG3, 37  
   MPEG4, 37  
   MPEG7, 37  
 MPEG2  
   entrelazado, 37  
   escalabilidad, 41  
   estructura de datos, 38  
   modo progresivo, 41  
   niveles, 41  
   ordenación de fotogramas, 40  
   perfiles, 41  
   program stream, 40  
   SNR, 41  
   transport stream, 40  
 muestreo diádico, 46  
 multiplexación estadística, 22  
 multiresolución, 48

## — P —

parvosistema, 66  
 pico de señal a ruido, PSNR, 30  
 pirámide steerable, 70  
 proceso de sensación visual, 64

## — Q —

QSIF, 93

## — R —

Rate Distorsion, 106  
 RDSI, 21, 35  
 reconstrucción perfecta, 54  
 redundancia, 29  
 retina, 64  
 retinotópico, 67  
 RSVP, 25  
 RTCP, 20  
 RTP, 19, 41  
 ruido mosquito, 81

## — S —

SIF, 93  
 sistema aferente-eferente, 69  
 subbanda, codificación en, 54  
 SVH, 63  
   prioridades, 87, 127

## — T —

TCP, 18  
 texturas, localización de, 44  
 tráfico  
   multiplexación, 103  
   persistente, 124  
   ráfaga, 124  
 transformación

  unitaria, 32  
 transformada  
   con solapamiento, 31  
   DCT, 31  
   DCT/predicción, 35  
   FFT, 31  
   Gabor, 70  
   KLT, 31  
   wavelet, 45  
   wavelet 3D, 84, 95  
   wavelet 3D inversa, 90

## — U —

UDP, 18

## — V —

V2, 68  
 V3, 69  
 V4, 68  
 V5 o MT, 69