



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

P.930

(08/96)

SERIE P: CALIDAD DE TRANSMISIÓN TELEFÓNICA

Calidad audiovisual en servicios multimedios

**Principios de un sistema de degradaciones de
referencia para vídeo**

Recomendación UIT-T P.930

(Anteriormente Recomendación del CCITT)

RECOMENDACIONES DE LA SERIE P DEL UIT-T
CALIDAD DE TRANSMISIÓN TELEFÓNICA

Vocabulario y efectos de los parámetros de transmisión sobre la opinión de los clientes	Serie	P.10
Líneas y aparatos de abonado	Serie	P.30 P.300
Patrones de transmisión	Serie	P.40
Aparatos para mediciones objetivas	Serie	P.50 P.500
Medidas electroacústicas objetivas	Serie	P.60
Medidas relativas a la sonoridad vocal	Serie	P.70
Métodos de evaluación objetiva y subjetiva de la calidad	Serie	P.80 P.800
Calidad audiovisual en servicios multimedia	Serie	P.900

Para más información, véase la Lista de Recomendaciones del UIT-T.

RECOMENDACIÓN UIT-T P.930

PRINCIPIOS DE UN SISTEMA DE DEGRADACIONES DE REFERENCIA PARA VÍDEO

Resumen

Esta Recomendación describe los principios de un sistema de referencia vídeo ajustable que permite generar las condiciones de referencia necesarias para caracterizar la calidad subjetiva de las imágenes vídeo producidas por los sistemas vídeo digitales con compresión. El sistema de degradaciones de referencia para vídeo (RISV, *reference impairment system for video*) puede utilizarse para simular las degradaciones debidas a la compresión de las secuencias vídeo, independientemente del esquema de compresión utilizado. Los métodos de evaluación subjetiva se describen en la Recomendación P.910.

En el Apéndice I se describe el sistema de referencia VIRIS (elaborado por Bellcore), que constituye una realización específica de un sistema de degradación de referencia ajustable. Aunque los estudios de Bellcore se basaron en el MPEG-1, el VIRIS puede emplearse también con otros esquemas de compresión, como el de la Recomendación H.261.

Se recomienda que un RISV sea capaz de producir las siguientes categorías de distorsiones, ya sea aisladamente o combinadas, con un ajuste independiente del nivel de cada degradación:

- a) Perturbaciones debidas a conversiones entre los formatos analógico y digital (por ejemplo, ruido y borrosidad).
- b) Perturbaciones debidas a la codificación y compresión (por ejemplo, trepidación, vibración de contorno y distorsión de bloque).
- c) Perturbaciones debidas a errores en el canal de transmisión (por ejemplo, bloques con errores).

En esta Recomendación se definen cinco tipos de degradación (distorsión de bloque, borrosidad, vibración de contorno, ruido y trepidación) y se indican métodos generales para reproducir estas degradaciones. En el Apéndice I se describe una realización específica de estas degradaciones llamada VIRIS. Otras degradaciones quedan en estudio.

Desde el punto de vista del espectador, las degradaciones producidas por el RISV deben ser una buena aproximación de las generadas por los sistemas de codificación y transmisión vídeo digitales.

Tres posibles aplicaciones del RISV son:

- 1) creación de condiciones de referencia en las pruebas subjetivas de sistemas vídeo digitales para que la calidad de las escenas presentadas a los espectadores abarquen toda la gama de calidades de imagen;
- 2) definición de niveles de degradación vídeo normalizados para la comparación de resultados de pruebas subjetivas; y
- 3) cuantificación de la calidad de un sistema vídeo percibida por el usuario con respecto a una referencia conocida.

Aunque esta Recomendación describe los principios de un RISV, es necesario efectuar pruebas de validación antes de recomendar una realización.

Orígenes

La Recomendación UIT-T P.930 ha sido preparada por la Comisión de Estudio 12 (1993-1996) del UIT-T y fue aprobada por el procedimiento de la Resolución N.º 1 de la CMNT el 30 de agosto de 1996.

PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) es un órgano permanente de la UIT. Este órgano estudia los aspectos técnicos, de explotación y tarifarios y publica Recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

La Conferencia Mundial de Normalización de las Telecomunicaciones (CMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución N.º 1 de la CMNT (Helsinki, 1 al 12 de marzo de 1993).

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

NOTA

En esta Recomendación, la expresión «Administración» se utiliza para designar, en forma abreviada, tanto una administración de telecomunicaciones como una empresa de explotación reconocida de telecomunicaciones.

© UIT 1996

Es propiedad. Ninguna parte de esta publicación puede reproducirse o utilizarse, de ninguna forma o por ningún medio, sea éste electrónico o mecánico, de fotocopia o de microfilm, sin previa autorización escrita por parte de la UIT.

ÍNDICE

Página

1	Alcance	1
2	Referencias.....	1
3	Abreviaturas y definiciones	1
4	Atributos de un sistema de degradaciones de referencia para vídeo (RISV).....	2
4.1	Gama de niveles de degradación vídeo ajustable.....	3
4.2	Posibles aplicaciones de un RISV.....	3
5	Degradaciones producidas por el RISV	4
5.1	Distorsión de bloque	4
5.2	Borrosidad.....	5
5.3	Vibración de contorno.....	5
5.4	Ruido.....	5
5.5	Trepidación	6
5.6	Combinación de degradaciones	6
6	Conclusiones	8
	Apéndice I.....	8
	VIRIS, una realización específica de un RISV	8
I.1	Introducción	8
I.2	Simulación de degradaciones.....	8
	I.2.1 Distorsión de bloque	9
	I.2.2 Borrosidad	10
	I.2.3 Vibración de contorno	11
	I.2.4 Ruido	13
	I.2.5 Trepidación.....	14
I.3	Cálculo de la relación señal/ruido de cresta (PSNR).....	15
I.4	Programa del VIRIS.....	15
I.5	Plan de pruebas subjetivas del VIRIS.....	17
	I.5.1 Secuencias de imágenes del VIRIS.....	17
	I.5.2 Condiciones de prueba del VIRIS	18
	I.5.3 Método de prueba del VIRIS.....	19
	I.5.4 Condiciones de observación del VIRIS.....	20
	I.5.5 Procedimientos de prueba del VIRIS.....	20
	I.5.6 Resultados de la prueba subjetiva del VIRIS.....	21
I.6	Plan de pruebas subjetivas de la vibración de contorno	25
	I.6.1 Secuencias de imágenes de vibración de contorno.....	26
	I.6.2 Condiciones de prueba del VIRIS y preparación del material de prueba.....	26

	Página
I.6.3 Método de prueba subjetiva de la vibración de contorno.....	27
I.6.4 Procedimientos de prueba subjetiva de la vibración de contorno.....	28
I.6.5 Resultado de la prueba subjetiva de vibración de contorno	28
I.7 Conclusiones.....	30
Bibliografía	31

Introducción

En evaluaciones subjetivas de la calidad vídeo se utilizan a menudo condiciones de vídeo de referencia, ya que las valoraciones de calificación por categorías pueden variar de una prueba a otra o de un laboratorio a otro dependiendo de la disposición del experimento (es decir, efectos de la gama y la frecuencia, cantidad de sujetos, etc.). Las condiciones de referencia permiten comparaciones relativas entre las condiciones de vídeo de referencia y las condiciones de vídeo de interés y permiten calificar la calidad subjetiva en términos de una medida objetiva de la referencia. La inclusión de las condiciones de vídeo de referencia en una prueba hace posible la presentación de una amplia gama de condiciones de prueba y proporciona una medida de la calidad vídeo que se puede reproducir más fiablemente en otra prueba o laboratorio.

Al considerar un sistema capaz de producir ese conjunto de condiciones de referencia vídeo ajustables, es necesario analizar diversos aspectos:

- 1) se ha de determinar el tipo de degradación vídeo o de perturbaciones debidas a la compresión digital que se deben simular,
- 2) la apariencia de las degradaciones vídeo simuladas debe ser similar a la de las degradaciones reales,
- 3) se deben definir con precisión los medios para generar las degradaciones, de modo que puedan ser reproducidas en cualquier laboratorio, y
- 4) la gama de niveles de degradación debe representar una amplia gama de calidades de funcionamiento por lo que se refiere a la calidad de la imagen.

Los algoritmos de compresión vídeo pueden introducir numerosas perturbaciones de degradación en la imagen. Cinco de los tipos más comunes son: distorsión de bloque, borrosidad, vibración de contorno, trepidación y ruido. Estas perturbaciones de degradación se crean en la etapa de compresión y se deben generalmente a la pérdida de información importante por cuantización, codificación y otras técnicas de compresión. Normalmente están correlacionadas con el movimiento de la imagen.

Un sistema de referencia vídeo ajustable debe ser sencillo de realizar y debe simular las cinco degradaciones anteriores de forma similar a las producidas por codificadores vídeo digitales reales y con independencia del esquema de compresión. La gama de degradaciones simulada debe abarcar una amplia gama de calidades de imagen para evaluar la calidad de imagen de los codificadores vídeo digitales. Para ello es necesario que se efectúen pruebas subjetivas, a fin de determinar la gama de calidades de imagen a que dan lugar las degradaciones simuladas.

En la presente Recomendación se describen algunas de las degradaciones que deberían ser producidas por un sistema de degradación de referencia vídeo ajustable y los métodos generales utilizados para producirlas. El Apéndice I contiene una descripción del VIRIS, que es una realización específica del sistema de degradación de referencia de vídeo para la evaluación de codificadores vídeo digitales MPEG-1. Se señala que, antes de que el VIRIS o cualquier realización de un RISV pueda ser aceptado por todas las Administraciones, son precisas pruebas de validación.

Esta Recomendación refleja el estado actual de la investigación sobre sistemas de degradación de referencia para vídeo, que está en sus principios. A medida que avancen los trabajos mejorará sin duda la comprensión de estos métodos. La presente Recomendación será revisada cuando se adquiera un nuevo conocimiento.

PRINCIPIOS DE UN SISTEMA DE DEGRADACIONES DE REFERENCIA PARA VÍDEO

(Ginebra, 1996)

1 Alcance

Esta Recomendación describe un sistema de referencia vídeo ajustable que se puede utilizar para generar las condiciones de referencia necesarias para caracterizar la calidad de imagen subjetiva de vídeo producida por sistemas vídeo digitales comprimidos. Se puede utilizar un sistema de degradaciones de referencia para vídeo (RISV) para simular las degradaciones que se producen por la compresión de secuencias vídeo, independiente del esquema de compresión. Los métodos de evaluación subjetivos no se incluyen en esta Recomendación sino que se describen en la Recomendación P.910.

2 Referencias

Las siguientes Recomendaciones y otras referencias contienen disposiciones que, mediante su referencia en este texto, constituyen disposiciones de la presente Recomendación. Al efectuar esta publicación, estaban en vigor las ediciones indicadas. Todas las Recomendaciones y otras referencias son objeto de revisiones por lo que se preconiza que los usuarios de esta Recomendación investiguen la posibilidad de aplicar las ediciones más recientes de las Recomendaciones y otras referencias citadas a continuación. Se publica periódicamente una lista de las Recomendaciones UIT-T actualmente vigentes.

- Recomendación UIT-R BT.500-6 (1994), *Método de evaluación subjetiva de la calidad de las imágenes de televisión.*
- Recomendación UIT-R BT.601-4 (1994), *Parámetros de codificación de televisión digital para estudios.*
- Recomendación UIT-R BT.802 (1992), *Imágenes y secuencias de prueba para las evaluaciones subjetivas de códecs digitales que cursan señales producidas conforme a la Recomendación 601.*
- Recomendación UIT-T P.920 (1996), *Métodos de prueba interactivos para comunicaciones audiovisuales.*
- Recomendación UIT-T P.910 (1996), *Métodos de evaluación subjetiva de la calidad vídeo para aplicaciones multimedia.*
- Recomendación UIT-T H.261 (1993), *Códec vídeo para servicios audiovisuales a $p \times 64$ kbit/s.*

3 Abreviaturas y definiciones

A los efectos de la presente Recomendación, se aplican las siguientes definiciones.

3.1 pel (o pixel): Elemento de imagen que describe el brillo o color de un punto discreto en una imagen.

3.2 bloque: Grupo de pels. Por ejemplo, un bloque de 8×8 pels es el bloque de codificación más pequeño utilizado en algoritmos MPEG-1. En una imagen SIF hay 1320 bloques, como ése, 44 en la dirección horizontal (352 pels/8) y 30 en la dirección vertical (240 líneas/8).

- 3.3 distorsión de bloque (o enlosamiento):** Distorsión de la imagen caracterizada por la aparición de una estructura de codificación de bloques subyacente.
- 3.4 borrosidad:** Distorsión global de toda la imagen, caracterizada por la reducción de la nitidez de los bordes y del detalle espacial.
- 3.5 errores de color:** Distorsión de la totalidad o una parte de la imagen final caracterizada por la aparición de tonalidades no naturales o no esperadas o niveles de saturación. Estas tonalidades o niveles de saturación no estaban presentes en la imagen original.
- 3.6 vibración de contorno:** Distorsión concentrada en los contornos de los objetos y caracterizada además por sus aspectos temporales y espaciales.
- 3.7 bloques con errores:** Forma de distorsión de bloque en la que uno o más bloques de la imagen no guardan semejanza con la escena en curso o con la previa y que a menudo contrastan mucho con los bloques adyacentes.
- 3.8 trepidación (o movimiento con sacudidas):** Movimiento originalmente uniforme y continuo percibido como una serie de "instantáneas" distintas.
- 3.9 ruido mosquito:** Forma de distorsión por vibración de contorno asociada a veces con el movimiento, caracterizada por efectos parásitos en movimiento y/o diagramas de ruido con manchas superpuestos a los objetos (se parece a un mosquito que volara entorno a la cabeza y los hombros de una persona).
- 3.10 ruido de cuantización:** Efecto "nieve" o "sal y pimienta" similar a un ruido aleatorio pero no uniforme a lo largo de la imagen.
- 3.11 formato intermedio común (CIF, *common intermediate format*):** Utilizado por codificadores conformes a la Recomendación H.261, de 352 pels de luminancia \times 288 líneas.
- 3.12 un cuarto de CIF (QCIF, *quarter CIF*):** 176 pels de luminancia \times 144 líneas.
- 3.13 formato de entrada fuente (SIF, *source input format*):** Utilizado por codificadores MPEG; formato progresivo, no entrelazado de 352 pels de luminancia \times 240 líneas \times 29,97 Hz o 352 pels de luminancia \times 288 líneas \times 25 Hz.
- 3.14 formato CCIR:** Norma UIT-R (anteriormente CCIR) de vídeo digital que utiliza formatos entrelazados de 720 pels de luminancia \times 480 líneas \times 30 Hz y 720 pels de luminancia \times 576 líneas \times 25 Hz.
- 3.15 grupo de expertos de imágenes en movimiento (MPEG, *moving pictures experts group*):** Grupo de trabajo organizado por la ISO que tiene la responsabilidad de elaborar normas multimedios/sistemas.

4 Atributos de un sistema de degradaciones de referencia para vídeo (RISV)

Se recomienda que el RISV tenga la capacidad de generar (por sí sólo o en combinación las perturbaciones) debidas a:

- a) las conversiones entre formatos analógicos y digitales (por ejemplo, ruido y borrosidad);
- b) la codificación y compresión (por ejemplo, trepidación, vibración de contornos y distorsión de bloques);
- c) los errores en canal de transmisión (por ejemplo, bloques con errores).

Desde el punto de vista del observador, las degradaciones producidas por el RISV deben ser una buena aproximación a las degradaciones generadas por la codificación de vídeo digital y los sistemas

de transmisión. Más adelante se describen con mayor detalle cinco de estas degradaciones, y en el Apéndice I figuran los algoritmos para la producción de estos tipos de degradaciones.

4.1 Gama de niveles de degradación vídeo ajustable

El grado de distorsión de cada degradación debe ser ajustable independientemente. Las gamas de ajuste han de ser suficientes para abarcar la gama total de calidad de los sistemas de transmisión y codificación vídeo digitales. Además, ha de haber también una manera de combinar dos o más degradaciones y de ajustar el nivel combinado de las mismas en una gama que sea útil para probar estos sistemas. Estos dos temas son objeto de estudios continuos y en 5.6 se dan más orientaciones sobre los mismos.

4.2 Posibles aplicaciones de un RISV

Tres posibles aplicaciones de un RISV son:

- 1) creación de las condiciones de referencia en pruebas subjetivas de sistemas vídeo digitales para asegurar que la calidad de las escenas presentadas a los observadores abarquen la gama total de calidad de imagen;
- 2) definición de niveles de degradación de vídeo normales que puedan ser utilizadas para comparar resultados de pruebas subjetivas; y
- 3) cuantificación de la calidad de un sistema vídeo percibida por el usuario con respecto a una referencia conocida.

La normalización del sistema RISV es suficiente para la primera aplicación. En una prueba subjetiva, una condición de referencia es la escena de vídeo que se procesa a través del RISV para añadir degradaciones de un nivel especificado. Generalmente se agregan diversas condiciones de referencia a la lista de escenas de pruebas subjetivas y los niveles de degradación de las condiciones de referencia se eligen para abarcar una gama completa de calidad: de EXCELENTE a MALA. Esto permite que los sujetos de las pruebas subjetivas efectúen una valoración de la calidad más fiable.

Para la segunda y tercera aplicaciones, no basta con la normalización de un RISV solamente. Para estas dos aplicaciones, se han de normalizar también las escenas de vídeo utilizadas con el RISV. Esto se debe a que el nivel de degradación del vídeo digital percibida depende en gran medida del contenido de la escena de vídeo (es decir, de la información espacial y temporal de la escena).

Con la segunda aplicación, la utilización de un RISV se hace necesaria si diversos laboratorios efectúan pruebas subjetivas independientemente en el mismo sistema vídeo digital. Es posible que los laboratorios deseen comparar los resultados de las pruebas, determinar las diferencias entre laboratorios y combinar los resultados para un análisis global. Si diferentes laboratorios desean comparar los resultados de sus pruebas independientes para efectuar un análisis global y calcular las diferencias entre laboratorios, deben utilizar las mismas escenas de prueba y condiciones de referencia. Esto es posible utilizando un RISV y produciendo los niveles de degradación asociados para un conjunto específico de escenas de prueba. En general, los informes que sólo indican la fijación de degradaciones del RISV no son suficientes. Se debe indicar también cuál es la escena de vídeo utilizada para fijar los niveles de degradación del RISV (y esa escena debe ser una de las escenas de prueba normalizadas de las que se puede disponer fácilmente, como las que figuran en la Recomendación UIT-R BT.802).

La Figura 1 ilustra la tercera aplicación posible de un RISV. Dicho sistema se puede utilizar para cuantificar la calidad de usuario de una secuencia de vídeo pasada a través de un sistema de compresión. Comparando la calidad de vídeo producida por un sistema de compresión con la producida por un RISV, es posible una medición objetiva de la calidad en términos de niveles de degradación del RISV. Según la Figura 1, el funcionamiento del RISV es como sigue: se inyecta una

escena de vídeo en el sistema vídeo sometido a prueba en el RISV. A continuación, la escena de vídeo que sale del sistema vídeo sometido a prueba y la del RISV se presentan alternativamente en un monitor. Los niveles de degradación del RISV se ajustan mediante un conjunto de "perillas" de ajuste (de hecho, parámetros de soporte lógico) hasta que un sujeto de la prueba de observación considere que las calidades vídeo son iguales.

La utilización del RISV de esta manera permite especificar objetivamente la calidad de vídeo en términos de niveles de degradación del RISV. A medida que se vayan desarrollando realizaciones de RISV se podrá ir mejorando el algoritmo especificado para evaluar la calidad vídeo, descrito por la tercera aplicación. El funcionamiento de un RISV de esta manera es, por tanto, bastante preliminar y será objeto de futuros estudios.

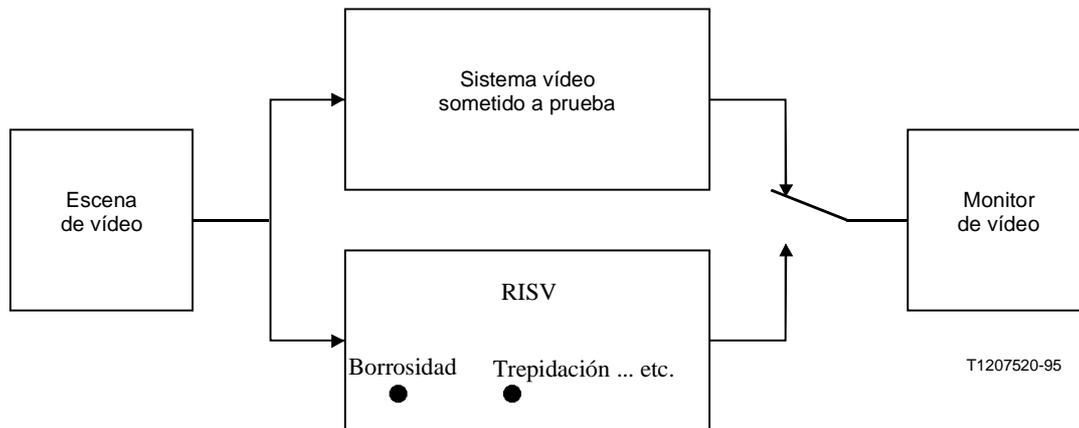


FIGURA 1/P.930

Ejemplo de RISV utilizado para cuantificar la calidad de funcionamiento de un sistema vídeo digital

NOTA – Dependiendo de la realización, es posible que el RISV no funcionen en tiempo real, ya que requiere una potencia de procesamiento de señal intensa. Quizás sea necesario, por tanto, que las escenas generadas sean almacenadas para reproducción posterior durante el proceso de adaptación.

5 Degradaciones producidas por el RISV

Son numerosas las degradaciones que debe simular el RISV. En la actualidad, las cinco degradaciones para cuya simulación se puede diseñar un RISV son: distorsión de bloque, borrosidad, vibración de contorno, ruido y trepidación. Otras degradaciones que deben ser simuladas, tales como bloques con errores y distorsión del color, quedan en estudio.

5.1 Distorsión de bloque

La distorsión de bloque (también conocida como agrupamiento en bloque) está causada a menudo por una cuantificación tosca de los componentes de frecuencia espacial durante el proceso de codificación. Para un determinado nivel de cuantificación, la distorsión de bloque es generalmente más visible en las zonas más uniformes de la pantalla. Una manera de simular la distorsión de bloque consiste en variar los valores de luminancia en un bloque de pels de tal modo que ese bloque se distinga de los pels circundantes.

Por ejemplo, en muchos codificadores que utilizan transformación discreta en coseno (DCT, *discrete cosine transform*), los pels están separados en bloques de 8×8 pels. En el codificador MPEG-1 por

ejemplo, hay 1320 bloques en una imagen SIF derivada del formato CCIR-601 de 525 líneas, 44 la dirección horizontal (352 pels/8) y 30 en la dirección vertical (240 líneas/8). Puesto que la distorsión de bloque se ve normalmente en zonas uniformes de elevado movimiento, una simulación puede implicar la degradación de un subconjunto de los bloques que se encuentran en esa zona cambiando el valor de luminancia de los bloques seleccionados según un algoritmo numérico. La posición y el número de bloques degradados junto con la intensidad del cambio de luminancia determinan la perceptibilidad y el nivel de la degradación.

Otros métodos de simulación de la distorsión de la disposición de bloque quedan en estudio.

5.2 Borrosidad

Se entiende por borrosidad la reducción de la nitidez del contorno y del detalle espacial y se produce a menudo en los algoritmos de compresión por el intercambio de bits de codificación de resolución y movimiento. Los formatos de imagen tales como SIF, CIF y QCIF tienen menor resolución que el formato de imagen CCIR 601 y, comparados con éste, presentan borrosidades incluso antes de la compresión.

La degradación por borrosidad se puede simular aplicando un filtro paso bajo de respuesta por impulsos de duración finita (FIR, *finite duration impulse response*) bidimensional y simétrico a la disposición de imagen digital. Se pueden proporcionar varios filtros con frecuencias de corte variables para permitir el control del grado de borrosidad. Se están estudiando otros métodos de simulación de la degradación por borrosidad.

5.3 Vibración de contorno

La vibración de contorno se produce a menudo por la utilización de niveles de cuantificación tosca durante el proceso de codificación dentro de un bloque que contiene contornos que delimitan zonas uniformes y pels con un nivel promedio notablemente distinto. Como resultado se produce distorsión concentrada en los contornos de objetos, caracterizada por una nitidez que varía en el tiempo o un ruido que varía en el espacio.

La vibración de contorno se puede simular agregando ecos cerrados adelantados y retrasados positivos o negativos a la imagen digital en las direcciones vertical y/u horizontal para crear un halo alrededor de los contornos de los objetos. Se puede variar el desplazamiento de los ecos para controlar el grado de vibración de contorno percibido. Se puede provocar una tenue oscilación del halo modificando el desplazamiento del eco cada varias tramas.

Los ecos de la imagen digital son el resultado de ondulaciones en la respuesta en la banda de paso de la disposición de imagen. La frecuencia de la ondulación determina el desplazamiento del eco; la amplitud de la ondulación determina la amplitud del eco y la fase de la ondulación determina la polaridad del eco (positivo o negativo). Así pues, los ecos pueden producirse por filtrado de la imagen con un filtro multibanda que contiene ondulaciones de banda de paso.

5.4 Ruido

Existen muchos tipos de degradaciones por ruido, producidas normalmente por algoritmos de compresión. Dos degradaciones frecuentes son el ruido mosquito y el ruido de cuantificación. El ruido mosquito es una forma de vibración de contorno caracterizado por efectos parásitos en movimiento y/o diagramas de ruido con manchas superpuestos a los objetos. El algoritmo para simular el ruido mosquito será objeto de futuros estudios. El ruido de cuantificación se puede crear en la etapa de cuantificación, durante el proceso de codificación por compresión. Este ruido se parece a un proceso de ruido aleatorio (nieve) y puede ser un ruido gris o coloreado, pero no es uniforme a lo largo de la imagen.

En el caso de esquemas de codificación basados en bloques, la degradación por ruido de cuantificación se puede simular sustituyendo el valor de pels de luminancia en ubicaciones aleatorias por un valor aleatorio restringido. La ubicación aleatoria que se ha de degradar se puede determinar extrayendo dos números aleatorios, el primero es una posición para la ubicación fila del pel y el segundo una posición para su ubicación columna. Una vez determinada la ubicación del elemento de imagen (pel), se sustituye el valor de luminancia no degradada del pel seleccionado por un valor aleatorio escogido de la gama especificada por el nivel de luminancia más bajo y más alto (o gama ampliada) de dentro del bloque que incluye la ubicación del elemento de imagen que se ha de degradar.

Se seleccionan ubicaciones de pel adicionales hasta que se haya degradado el número deseado de pels. El número de pels no degradados es una indicación del nivel de degradación por ruido de cuantificación.

5.5 Trepidación

La trepidación o movimiento con sacudidas es un movimiento, en principio uniforme y continuo, percibido como una serie de "instantáneas" distintas. Se observa normalmente en videoteléfonos y en otros sistemas vídeo de baja velocidad binaria. Es el resultado de la omisión de tramas de vídeo para reducir la cantidad de información vídeo que se requiere que un sistema transmita o procese por unidad de tiempo. La velocidad de tramas resultante es el nivel de esta degradación.

El movimiento con sacudidas se puede simular repitiendo tramas de vídeo. Por ejemplo, cuando se utiliza un sistema de 30 tramas por segundo, para crear una velocidad de 10 tramas por segundo se utiliza una de cada tres tramas para sustituir a las dos tramas siguientes en una secuencia de vídeo. Se están estudiando otros métodos de vinculación de la trepidación que correlacionan la cantidad de movimiento en una secuencia con la gravedad de la degradación por trepidación.

5.6 Combinación de degradaciones

Un sistema RISV debe ser capaz de producir simulaciones de por lo menos los siguientes cinco tipos de degradaciones: distorsión de bloque, borrosidad, vibración de contorno, ruido y trepidación. El RISV también puede simular otras degradaciones, tales como errores de color, ruido mosquito y bloques con errores, pero estos temas quedan en estudio. Para cada una de las cinco degradaciones definidas, el RISV ha de poder ajustar el nivel de la degradación en una gama que vaya desde la imperceptibilidad de la degradación visual hasta un nivel muy molesto de degradación visual. La selección de los niveles de degradación máximo y mínimo se hace usualmente por experimentación y en base a la observación visual y apreciaciones subjetivas. Las degradaciones que se han de utilizar y el número de pasos de nivel de degradación para abarcar la gama de calidad deseada quedan a criterio del experimentador. Normalmente el proceso consiste en considerar el contenido de la escena y el codificador sometido a prueba y seleccionar apropiadamente el nivel y gama adecuados de cada degradación de manera experimental.

El RISV ha de poder introducir en una escena de vídeo más de una degradación de las cinco posibles. Al aplicar degradaciones múltiples, el orden en que se introducen las mismas adquiere importancia. Cuando se agregan dos o más degradaciones a una escena, el RISV puede funcionar de manera secuencial. Las tramas de la escena se procesan teniendo en cuenta la primera degradación elegida, y a continuación se procesan nuevamente teniendo en cuenta cada una de las otras degradaciones elegidas. Cuando se decide la secuencia de las degradaciones se deben considerar los siguientes aspectos:

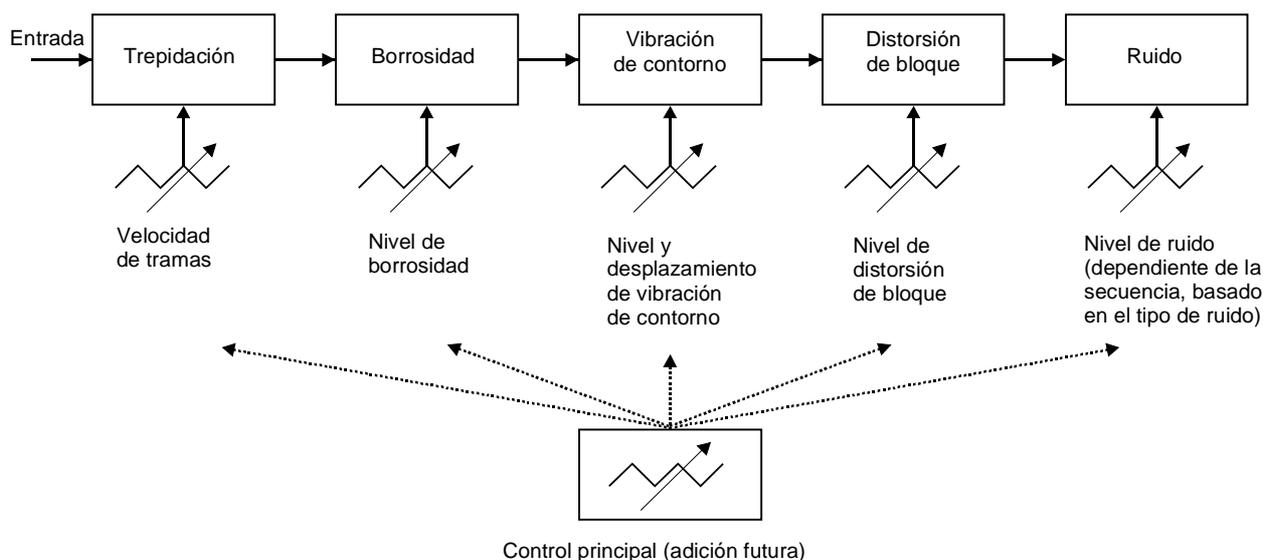
- 1) Es más eficaz, desde el punto de vista del tiempo de procesamiento, eliminar tramas para simular trepidación antes de procesar otra degradación. También es importante preparar la misma secuencia de tramas que utilizaría el codificador de compresión simulado cuando (se

selecciona un cuantificador y) se produce distorsión de bloque. En la mayoría de los casos es conveniente que la detección de zona en movimiento emplee esta secuencia con el número de tramas disminuido, mientras que el tiempo de procesamiento de la detección de contorno también se benefician del funcionamiento con menos tramas.

- 2) Todas las operaciones de detección se deben efectuar con esta secuencia de tramas reducida, y antes de aplicar las otras degradaciones (por coherencia con el apartado 1). Por ejemplo, la detección precisa de contorno se debe efectuar antes de añadir borrosidad, y las zonas de movimiento se deben identificar después de añadir trepidación.
- 3) La borrosidad se debe agregar antes que otra degradación localizada para evitar modificar sus niveles.
- 4) La eficacia del procesamiento sugiere, de nuevo, que la repetición de tramas (para completar la simulación de trepidación) se debe efectuar en último lugar, lo cual también es necesario para representar correctamente otras degradaciones, ya que no deben cambiar en tramas repetidas.

Se solicitan nuevas consideraciones al respecto como parte del estudio en curso.

La Figura 2 ilustra el mismo orden de degradaciones utilizado por el VIRIS, que es una realización específica de un RISV (véase el Apéndice I). Cada uno de los niveles de degradación se puede ajustar independientemente en la gama apropiada o bien el RISV puede tener un control principal que pueda ajustar los niveles de todas las degradaciones simultáneamente. Sin embargo, aún no se comprende bien la utilización de un control principal y este aspecto queda en estudio.



T1207530-95

FIGURA 2/P.930

Diagrama de bloques del sistema RISV que muestra el orden de las degradaciones y los controles de nivel

6 Conclusiones

En esta Recomendación se exponen los principios de un sistema de referencia vídeo ajustable que se puede utilizar para generar las condiciones de referencia necesarias para caracterizar la calidad de imagen vídeo subjetiva producida por sistemas de vídeo digital comprimidos. La utilización de las condiciones de referencia durante una prueba subjetiva permite hacer comparaciones relativas entre las condiciones de vídeo de referencia y las condiciones de vídeo que interesan y obtener de este modo una medición de la calidad de vídeo más fiable. En la presente Recomendación se describen tres aplicaciones para un sistema de referencia ajustable y las degradaciones que dicho sistema debe producir. Si bien se describe las características de un RISV, antes de poder recomendar una realización se requieren pruebas de validación.

Apéndice I

VIRIS, una realización específica de un RISV

I.1 Introducción

El propósito de este apéndice es informar de las actividades efectuadas para diseñar un sistema de degradaciones de referencia vídeo (VIRIS, *video reference impairment system*) que simula las degradaciones de distorsión de bloque, borrosidad, vibración de contorno, trepidación y ruido. Se pretende que este sistema sea una especie de herramienta general en los laboratorios de vídeo, para evaluar la calidad de funcionamiento de los codificadores vídeo digitales. VIRIS (al que a veces se hace referencia como VIRIS1, una versión más reciente del sistema de degradación de referencia vídeo) es un sistema de soporte lógico implantado en el lenguaje de programación C para manipular ficheros de vídeo digital a fin de introducir degradaciones de codificación simuladas en una imagen vídeo. Está diseñado para que funcione en imágenes SIF, pero el método se puede aplicar a otros formatos de imagen, tales como CIF, QCIF y CCIR 601. El VIRIS es una etapa preliminar y sólo es útil como herramienta de laboratorio de uso general.

En la subcláusula siguiente se analizan y definen los métodos utilizados para simular las degradaciones en el VIRIS. En I.3 se describe el cálculo de la relación señal/ruido de cresta (PSNR, *peak signal-to-noise ratio*). En I.4 se describen las características de la última realización del VIRIS, mientras que en I.5 y I.6 se describen:

- 1) el plan de pruebas subjetivas para determinar las gamas de degradaciones con las que producir gamas de calidades de imagen de amplitud apropiada; y
- 2) los resultados de las pruebas subjetivas efectuadas para caracterizar la relación entre nivel de degradación y calidad objetiva. Las conclusiones figuran en I.7.

I.2 Simulación de degradaciones

El formato de imagen de entrada para algoritmos de compresión MPEG-1 es SIF, que es un formato de 30 Hz, progresivo y no entrelazado de 240×352 pels de Y (señal de luminancia) y 120×176 pels de U y V (señales de crominancia). SIF se convierte a partir del formato CCIR 601 de 30 Hz, entrelazado, de 480×720 pels de Y y 480×360 pels de U y V . La resolución del SIF es la mitad de la de CCIR 601 en las direcciones vertical y horizontal. Se eligió como formato de entrada del VIRIS debido al posible uso de VIRIS en la evaluación de códecs vídeo de 1,5 Mbit/s utilizados en los servicios de tono de marcación de vídeo. No obstante, los métodos descritos en esta subcláusula para simular degradaciones se pueden aplicar a otros formatos de imagen con sólo pequeñas modificaciones para tener en cuenta las diferencias de los formatos de imagen.

El programa C, VIRIS, funciona en una trama de vídeo SIF por vez. Sólo se manipulan los pels de luminancia para crear las degradaciones simuladas, los pels de color no se modifican.

Las cinco subcláusulas siguientes detallan los métodos utilizados en el VIRIS para simular las degradaciones por distorsión de bloque, borrosidad, vibración de contorno, ruido y trepidación en imágenes vídeo SIF.

I.2.1 Distorsión de bloque

La distorsión de bloque es causada a menudo por una cuantificación muy poco precisa durante el proceso de compresión, lo que da lugar a una distorsión o pérdida de componentes de alta frecuencia. El resultado de esta distorsión es que se hace visible la estructura de codificación de bloques subyacente. Para un determinado nivel de cuantificación, la distorsión de bloque es generalmente más visible en las zonas uniformes de una secuencia.

Un bloque de 8×8 pels es el bloque de codificación más pequeño utilizado en los algoritmos MPEG-1. En una imagen SIF hay 1320 bloques como ese, 44 en la dirección horizontal ($352 \text{ pels de luminancia}/8$) y 30 en la dirección vertical ($240 \text{ líneas}/8$).

Las ubicaciones de bloques que hay que degradar se seleccionan identificando la región de la imagen en la que cabría esperar que apareciera la distorsión de bloque. Se trata normalmente de zonas uniformes con movimiento asociado. Lo que sigue es el proceso que utiliza el sistema VIRIS para simular el efecto parásito de distorsión de bloque [4]:

- a) Los bordes de la trama en curso se identifican filtrando la trama rellena de ceros¹ mediante la utilización de la técnica de filtrado Sobel [8, 10]. Puesto que la distorsión de bloque se percibe por lo general en las zonas uniformes de una imagen, al identificar los bordes de una imagen las zonas uniformes se ubican naturalmente donde la distorsión de bloque generada por la compresión es visible normalmente.
- b) La distorsión de bloque se sitúa en las zonas uniformes de la imagen que indican movimiento.

La manera de determinar el movimiento o el cambio en la información de percepción temporal (*TI*, *temporal perceptual information*) consiste en sustraer las intensidades de luminancia de elementos de imagen de ubicación (x,y) a través de tramas sucesivas.

$$TI(x, y, t + 1) = Y(x, y, t + 1) - Y(x, y, t) \quad (\text{I.2-1})$$

donde $Y(x,y,t+1)$ es la imagen de luminancia rellena de ceros en el momento $t+1$, $Y(x,y,t)$ es la imagen de luminancia rellena de ceros en el momento t y $TI(x,y,t+1)$ es la información de percepción temporal en el momento $t+1$ y pixel (x,y) . Puesto que la distorsión de bloque se ve normalmente en zonas de movimiento, para determinar la ubicación del bloque que se ha de degradar, se acumula la *TI* de la trama en curso y de la anterior para cada bloque de la presente escena:

$$\sum_{i=-7}^0 \sum_{j=-7}^0 \text{abs}(TI(8r+i, 8s+j, t+1)) \quad (\text{I.2-2})$$

[Si $(8r+i, 8s+j)=(x,y)$ es un borde en el momento t o $t+1$, $TI(.)=0$, y $r=1$ a 44 , $s=1$ a 30 , la posición de bloque submuestreado de la imagen SIF.]

Puesto que cuanto mayor es esta suma mayor es el movimiento de bloque, se efectúa una búsqueda a través de toda la trama para ubicar el bloque (r,s) con la mayor indicación de movimiento (o de manera equivalente, el bloque con la mayor *TI*). Si este bloque también contiene dentro de él no más

¹ Trama rellena de ceros" es la adición de una banda de ceros a los lados de la trama. Esta banda de ceros tendrá una anchura mayor o igual que $(N-1)/2$ pixels, siendo N el número de derivaciones de filtro utilizadas en el filtro VIRIS más grande.

de 5 pels que han sido clasificados como bordes², el bloque se degrada y la posición se elimina en las nuevas determinaciones de ubicaciones de distorsión de bloque para esa trama. En base al número deseado de bloques a degradar, el proceso se repite para ubicar el bloque siguiente con el mayor grado de movimiento situado en una región uniforme de la trama.

Una vez seleccionados los bloques que se han de degradar, se puede simular la distorsión de bloque modificando los valores de luminancia en el bloque de pels de tal modo que ese bloque de pels sea distinguible de los pels circundantes. El proceso de degradación es como sigue: en primer lugar se determina el valor medio de luminancia de los 64 pels del bloque, P_{ave} . Cada valor de pel del bloque se fija entonces en la suma de las dos cantidades:

- 1) el valor medio entre la luminancia del pel de que se trate y la luminancia del bloque, P_{ave} ;
- 2) un número elegido al azar entre -2 y 2 . Cuando se agregan bloques a una secuencia de imagen, la ubicación de los bloques se cambia cada 15 tramas para mantener constante la posición del bloque distorsionado, de modo que el ojo perciba la degradación.

I.2.2 Borrosidad

Se entiende por borrosidad la reducción de la nitidez del contorno y del detalle espacial en una imagen. Esta degradación se produce principalmente en los algoritmos de compresión por el intercambio de bits de codificación de resolución y movimiento. Los formatos de imagen tales como SIF, CIF y QCIF tienen ya de por sí menor resolución que el formato de imagen CCIR 601 y, comparados con éste, presentan borrosidades incluso antes de la compresión. La degradación por borrosidad se crea en el VIRIS aplicando un filtro paso bajo de respuesta por impulsos de duración finitas (FIR) simétrico y unidimensional, de 15 derivaciones, a la disposición de imagen digital SIF en la dirección horizontal. Se diseñaron seis filtros con frecuencias de corte variables utilizando un conjunto de programas de diseño de filtro digital [6]. La selección de las frecuencias de corte de un diseño específico se basó en un procedimiento de aproximaciones sucesivas. Se procesaron algunas tramas de una imagen vídeo SIF con distintos filtros paso bajo y se efectuaron pruebas de comparación informales de las imágenes filtradas y no filtradas. Se seleccionaron entonces filtros específicos para cubrir una amplia gama de calidades de imagen. Dichos filtros tienen unas frecuencias de corte de diseño de 1,5; 1,0; 0,75; 0,5; 0,375 y 0,25 MHz, lo que corresponde a unos niveles de degradación por borrosidad de 1 a 6, respectivamente. En el Cuadro I.1 figuran los valores de los coeficientes de las 15 derivaciones de cada filtro.

² Se ha determinado mediante experimentación que si la cantidad devuelta por aplicación del filtro Sobel a bloques de 3×3 pels es superior a 500, el pel se clasifica como un borde para los fines de la degradación de ruido correlacionado de señal.

CUADRO I.1/P.930

Valores de coeficientes para filtro paso bajo

	Coeficiente de derivación del filtro					
Derivación del filtro	Nivel de degradación por borrosidad (frec. corte de filtro paso bajo, MHz)					
	1 (1,5)	2 (1,0)	3 (0,75)	4 (0,5)	5 (0,375)	6 (0,25)
[-7],[7]	-1	-2	-3	0	2	-2
[-6],[6]	1	-1	-3	-3	1	2
[-5],[5]	3	1	-1	-5	-4	4
[-4],[4]	6	5	3	-3	-6	-3
[-3],[3]	10	9	8	5	-1	-9
[-2],[2]	13	14	15	15	13	3
[-1],[1]	15	17	20	24	28	31
[0]	16	19	22	28	34	47

El filtrado se lleva a cabo en el VIRIS aplicando el filtro paso bajo a la disposición de imagen de luminancia SIF a lo largo de cada una de las 240 filas. La señal de salida filtrada, y_n , viene dada por,

$$y_n = \frac{1}{S} \sum_{i=-7}^7 h_i x_{n-i} \quad (I.2-3)$$

donde:

x_n Disposición de imagen SIF de entrada

h_i Coeficientes del filtro (Cuadro I.1)

$$S = \sum_{i=-7}^7 h_i$$

$$0 \leq n < 352$$

En la ecuación (I.2-3) la señal de entrada que identifica los límites de tramas verticales, x_n , se amplía en siete muestras al comienzo y al final de cada línea para permitir el cálculo de la señal de salida borrosa cerca de los límites de la trama. El valor asignado a estas muestras de ampliación es igual al valor del pixel de luminancia que linda con el borde.

I.2.3 Vibración de contorno

Se han desarrollado dos algoritmos distintos para efectuar degradaciones por vibración de contorno, a saber:

- 1) degradación por vibración de contorno aplicada sólo a los bordes verticales de los objetos;
- 2) degradación por vibración de contorno aplicada a los bordes verticales y horizontales.

El algoritmo para la simulación de la degradación por vibración de contorno unidireccional ya no se recomienda. En las siguientes subcláusulas se describe la simulación de la degradación por vibración de contorno de la última versión del sistema VIRIS.

I.2.3.1 Realización del VIRIS de vibración de contorno

La distorsión por vibración de contorno se produce por un nivel de cuantificación demasiado alto en un bloque que contiene una zona uniforme y algunos elementos de imagen (pels) con un nivel promedio notablemente diferente. La vibración de contorno se simula en el sistema VIRIS agregando ecos cerrados adelantados y retrasados negativos a la imagen SIF en las direcciones vertical y horizontal para crear un halo alrededor de los contornos de los objetos. Los ecos se han diseñado con desplazamientos de 0,375; 0,5 y 0,75 microsegundos y la oscilación del halo se crea cambiando el desplazamiento del eco cada cinco tramas.

Puesto que el sistema VIRIS se diseñó originalmente para evaluar la calidad de funcionamiento de los codificadores MPEG-1, la simulación de degradación por vibración de contorno se concibió de modo que se simulara mejor dicha degradación tal como se observa en un vídeo MPEG-1. Se determinó por observación experimental que el efecto parásito vibración de contorno en videos MPEG-1 se produce tanto en los lados de los objetos (es decir, en los bordes verticales) como en la parte superior e inferior de los mismos (es decir, en los bordes horizontales); por esta razón, el algoritmo de simulación emplea un algoritmo de filtrado bidimensional. Además, aunque MPEG-1 introduce perturbaciones de borrosidad por el filtrado de la imagen para simular el efecto de halo, no fue necesario introducir borrosidad adicional en la imagen para simular el efecto parásito de borrosidad en el MPEG-1. Los ecos utilizados para simular la degradación por vibración de contorno tendían también a difuminar los bordes de la imagen.

I.2.3.2 Método de simulación de vibración de contorno del VIRIS

Como se indicó anteriormente, la degradación por vibración de contorno se simula en el sistema VIRIS utilizando un filtro con ondulaciones en la respuesta de amplitud en la banda de paso para crear una degradación de eco. La frecuencia de la ondulación determina el desplazamiento del eco; la amplitud de la ondulación determina la amplitud del eco y la fase de la ondulación determina la polaridad del eco (positiva o negativa). Las ondulaciones en la respuesta de frecuencia se producen aplicando filtros multibanda digitales de 13 derivaciones [6] a la disposición de imagen SIF, primero en la dirección horizontal y después en la dirección vertical. El Cuadro I.2 muestra los valores de los coeficientes para los tres conjuntos de filtros multibanda que producen los tres desplazamientos de eco. El desplazamiento de eco que, en concreto, se utiliza en una trama se determina mediante la entrada del usuario en el programa VIRIS.

El valor de coeficiente de la derivación 0 de cada filtro se iguala a un valor de 175. Con la excepción de un par de ponderaciones de derivación de cada uno de los tres filtros de desplazamiento de eco, todas las demás ponderaciones de derivación de filtro son 0. La ubicación de la derivación de filtro que tiene un valor de coeficiente distinto de cero con respecto a la derivación central determina el desplazamiento de eco, mientras que el valor del coeficiente determina la amplitud del eco. El valor de coeficiente real lo introduce el usuario en el programa VIRIS y puede variar de -1 a -30, en pasos enteros. El mismo valor de coeficientes se aplica a cada uno de los tres filtros.

El filtrado se lleva a cabo en el VIRIS aplicando los filtros multibanda digitales a la disposición de imagen de luminancia SIF, primero a lo largo de cada una de las 240 filas y después descendiendo por cada una de las 352 columnas. La señal de salida filtrada, z_n , para cada caso viene dada por la siguiente expresión:

$$z_n = \frac{1}{S} \sum_{i=-6}^6 h_i x_{n-i} \quad (\text{I.2-4})$$

donde:

x_n Disposición de imagen SIF de entrada

h_i Coeficiente del filtro (véase el Cuadro I.2)

$$S = \sum_{i=-6}^6 h_i$$

$0 \leq n < 352$ (horizontal)

$0 \leq n < 240$ (vertical)

En la ecuación (I.2-4), el valor de luminancia de la señal de entrada que identifica los límites de trama verticales y horizontales, x_n , con $n = 0, n = 352$ (horizontales) y $n = 0, n = 240$ (verticales), se amplía en 6 muestras más allá de los límites izquierdo, derecho, superior e inferior de la trama para permitir el cálculo de la señal de salida z_n .

CUADRO I.2/P.930
Valores de coeficientes para filtros multibanda de 13 derivaciones

Derivación del filtro	Coeficientes de derivación del filtro		
	Desplazamiento de eco, microsegundos		
	0,375	0,5	0,75
[-6],[6]	0	0	-1 a -30
[-5],[5]	0	0	0
[-4],[4]	0	-1 a -30	0
[-3],[3]	-1 a -30	0	0
[-2],[2]	0	0	0
[-1],[1]	0	0	0
[0]	175	175	175

El procedimiento seguido para llegar al método de simulación anterior fue de tipo experimental. Requiere un proceso de aproximaciones sucesivas al estudiar las diferentes maneras de agregar ecos y comparar a continuación los resultados con imágenes MPEG-1. Se introdujeron cambios en el método de simulación para intentar mejorarla, pero esos cambios no se pudieron hacer en tiempo real ya que era necesario volver a compilar el programa del VIRIS para producir los cambios. Se considera, no obstante, que con el método de simulación precedente se consiguió una simulación razonable de las degradaciones MPEG-1. Por supuesto, la perceptibilidad de las degradaciones MPEG-1 reales está correlacionada con el movimiento y la imagen, mientras que en el VIRIS la mayoría de las degradaciones simuladas no lo están. En consecuencia, el grado de semejanza entre las degradaciones simuladas y las degradaciones reales depende en buena medida de la secuencia de imagen que, en concreto, se utiliza. De cualquier modo, una de las preocupaciones principales al diseñar un sistema de degradación de referencia vídeo ajustable es que sea sencillo de realizar e independiente de cualquier esquema de codificación. Esto impide, actualmente, la correlación entre el movimiento en la imagen y la degradación simulada.

I.2.4 Ruido

Hay dos tipos de ruido producidos por VIRIS, ruido de cuantificación y ruido correlacionado con la señal.

I.2.4.1 Ruido de cuantificación

El ruido de cuantificación es un ruido creado a veces en la etapa de cuantificación del proceso de compresión. Se parece a un proceso de ruido aleatorio pero no es uniforme en toda la imagen. Se simula en el VIRIS sustituyendo el valor de pels de luminancia en ubicaciones aleatorias por un valor

aleatorio entre 16 (negro) y 255 (blanco de cresta). La ubicación aleatoria se determina extrayendo dos números al azar, el primero con un valor entre 1 y 240 para la ubicación fila del pel y el segundo con un valor entre 1 y 352 para su ubicación columna. Una vez determinada la ubicación del pel, se sustituye el valor de pels de luminancia por un valor aleatorio entre 16 y 255. A continuación se elimina esa ubicación de pel en las nuevas determinaciones de ubicaciones de pels de ruido aleatorio para esa trama. El número de valores de pels de ruido que se pueden cambiar en una determinada trama varía de 1 a 84 480 (352×240 pels) y se determina a partir de la entrada del usuario en el VIRIS. Con cada trama sucesiva, se cambia la semilla generadora de números aleatorios como parte de los datos de entrada en el VIRIS, de tal modo que las ubicaciones de pels de ruido cambian de trama a trama.

I.2.4.2 Ruido correlacionado con la señal

Ruido correlacionado con la señal es un término utilizado para describir el aspecto resultante de la combinación de degradaciones por vibración de contorno y ruido mosquito, tal como se observa en vídeo comprimido. El ruido correlacionado con la señal es una distorsión observada normalmente alrededor de los contornos de objetos en movimiento y se caracteriza por un diagrama de ruido tipo "halo" de contornos de objetos en movimiento y se caracteriza por un diagrama de ruido tipo "halo" de grano fino superpuesto cerca de, o sobre los, contornos de los objetos en movimiento. Una simulación de la combinación de estas degradaciones requiere extraer los contornos de una escena, determinar si están en movimiento y, finalmente, inyectar una degradación de ruido [1].

Los contornos se ubican en una escena utilizando la técnica de filtrado Sobel. Si el valor devuelto al aplicar el filtro Sobel en los nueve pixels adyacentes es superior a 50, la ubicación del pixel se clasifica como borde para los fines de la degradación de ruido correlacionado con la señal. Si la ubicación del pixel tiene además un grado de movimiento asociado (determinado al diferenciar el valor de luminancia de la ubicación del pixel en dos tramas sucesivas y verificar que la TI es superior a 2) se señala como ubicación que ha de ser degradada.

Para degradar una ubicación de pixel se agrega un número aleatorio entre $-\beta$ y β al valor de luminancia del pixel no degradado.

$$Y(x, y, t) = Y(x, y, t) + \text{random}(-\beta, +\beta) \quad (\text{I.2-5})$$

donde $Y(x, y, t)$ es el valor de luminancia de la ubicación del pixel (x, y) en el momento τ identificada como un contorno en movimiento.

Para cada ubicación de pixel que cumple la condición de ser un contorno en movimiento, el valor de luminancia de la imagen no degradada se modifica en una cantidad comprendida entre $-\beta$ y β . El nivel de degradación, especificado por β , controla la intensidad de la degradación. Al aumentar β aumenta la visibilidad de la degradación. Para que la correlación entre la degradación y la variación temporal sea más efectiva, los pixels que se han de degradar son, de hecho, la unión de pixels identificados como bordes en la trama previa y la trama en curso.

I.2.5 Trepidación

La trepidación se define como un movimiento, en principio uniforme y continuo, percibido como una serie de "instantáneas" distintas. Se observa normalmente en videoteléfonos y en algunos sistemas de teleconferencia vídeo de baja velocidad binaria. Es el resultado de la repetición de tramas de vídeo para reducir la cantidad de información vídeo que se requiere que un sistema transmita o procese. El sistema VIRIS utiliza un factor de repetición de tramas (FRF, *frame repetition factor*) para controlar el nivel de trepidación presente. Un FRF de 3, por ejemplo, significa que una de cada tres tramas sustituye a las dos tramas siguientes en la secuencia de vídeo. La velocidad efectiva de las tramas (EFR, *effective frame rate*) se calcula como $30/\text{FRF}$. La EFR del ejemplo con un FRF de 3 sería entonces 10 tramas por segundo.

I.3 Cálculo de la relación señal/ruido de cresta (PSNR)

En la actualidad, cada una de las degradaciones simuladas en el VIRIS se caracteriza objetivamente calculando la relación señal/ruido de cresta (PSNR, *peak signal-to-noise ratio*) en cada trama procesada y el valor medio de la PSNR en todas las tramas de una secuencia de imágenes procesada. La PSNR no ponderada sirve como medida para evaluar la distorsión de la secuencia procesada. Para cada trama procesada, k , se calcula el ruido RMS (cuadrático medio), N_{rms_k} con la fórmula siguiente:

$$N_{rms_k} = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^M [U_{ij} - I_{ij}]^2}{N \cdot M}} \quad (I.3-1)$$

donde:

U_{ij} = Valor de luminancia de pel no degradado en la ubicación fila i y columna j

I_{ij} = Valor de luminancia de pel degradado en la ubicación fila i y columna j

N = 240 para imagen SIF

M = 352 para imagen SIF

k = k -ésima trama.

Para calcular la PSNR en una secuencia de K tramas, se determina primero el valor medio del ruido por trama, N_{rms_k} , aplicando la ecuación (I.3-1). A continuación, se calcula el valor medio del ruido, N_{rms} , a lo largo de la secuencia de K tramas aplicando la ecuación (I.3-2) siguiente:

$$N_{rms} = \frac{1}{K} \sum_{k=1}^K N_{rms_k} \quad (I.3-2)$$

Por último, se calcula la PSNR con la ecuación (I.3-3) que se indica a continuación:

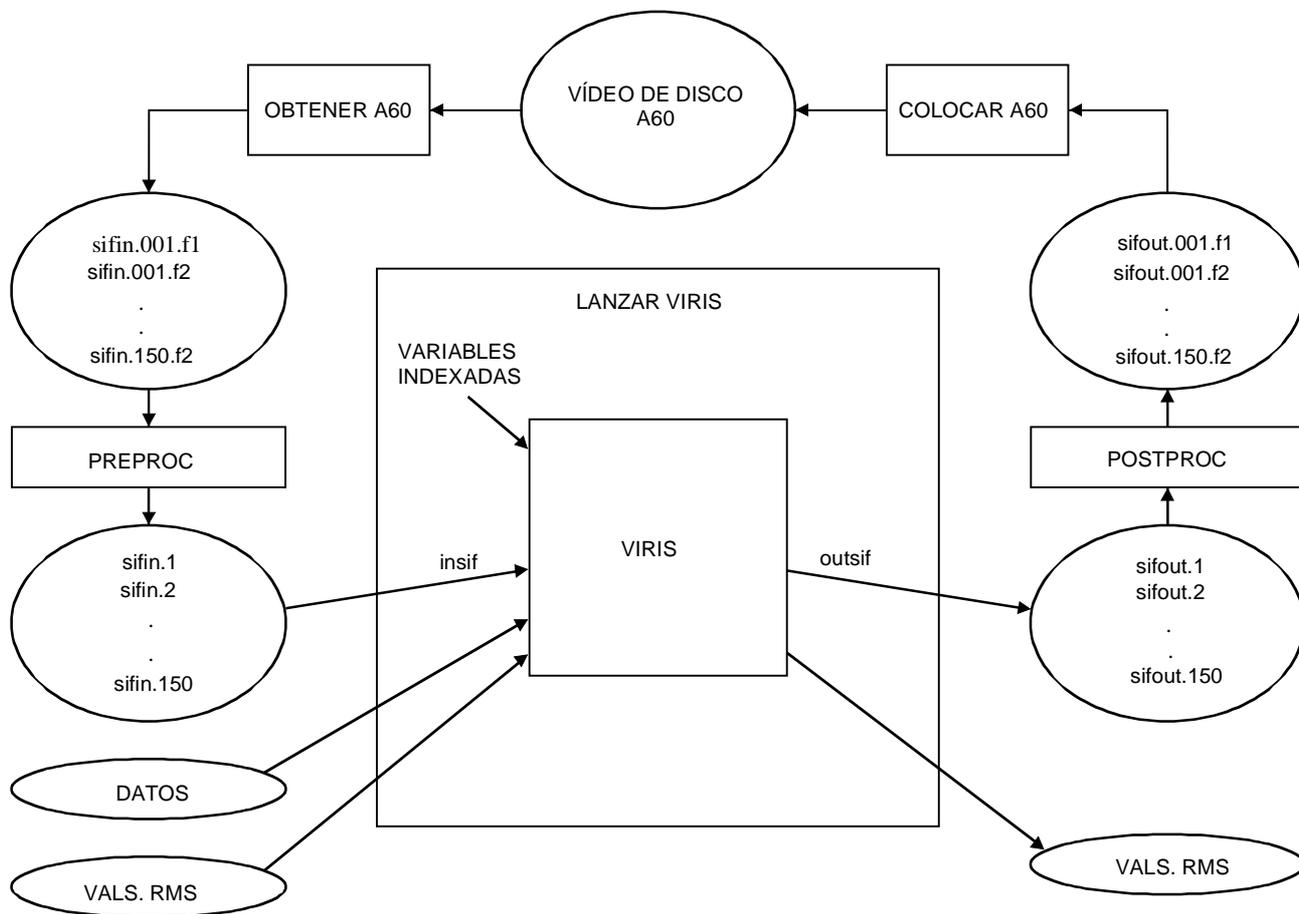
$$PSNR = 20 * \log_{10} \frac{S_p}{N_{rms}} \quad (I.3-3)$$

donde S_p es el número de niveles a los que se cuantifica la intensidad de luminancia. El sistema en el que funciona el VIRIS cuantifica los pels de luminancia con 8 bits, o sea $S_p = 255$.

Dado que la PSNR no ponderada descrita anteriormente está débilmente correlacionada con el sistema visual humano, quizás se requiera una medida objetiva más precisa. Este tema queda en estudio.

I.4 Programa del VIRIS

En la Figura I.1 se muestra el proceso por el que el sistema VIRIS degrada una trama de vídeo SIF por vez. El VIRIS inyecta degradaciones en una imagen SIF convertida a partir de una imagen CCIR 601 (utilizando el programa "preproc") almacenada en un disco vídeo. Con los niveles de VIRIS almacenados en ficheros de datos, se procesan las tramas y la secuencia degradada se vuelve a convertir en formato CCIR 601 (utilizando el programa "postproc") y se transfiere a un vídeo de disco digital para visualización.



T1207540-95

FIGURA I.1/P.930
Funcionamiento del VIRIS

El fichero de datos controla el nivel de las degradaciones añadidas a una secuencia de vídeo. Existen básicamente 6 parámetros de control del nivel de la degradación aplicables al sistema VIRIS:

- Nivel de distorsión de bloque, introducido como un número entero que representa el 0,1% de los 1320 bloques totales que se han de cambiar (el resultado se redondea al número entero más próximo). Por ejemplo, un nivel de distorsión de bloque de 10 da como resultado 13 bloques que están degradados ($10 \times 0,001 \times 1320 = 13,2$ que redondeado equivale a 13).
- Nivel de ruido de cuantificación, introducido como un número entero que representa el 0,001% de los valores de pels de luminancia total de 84480 (redondeado al número entero más próximo) que se han de cambiar. Por ejemplo, un nivel de ruido de 10 da como resultado un cambio en el valor de luminancia de 8 pels ($10 \times 0,00001 \times 84480 = 8,4$ que redondeado es 8).
- Nivel de ruido correlacionado con señal, introducido como un número entero que representa la gama de valores de luminancia por los que puede alterarse un pel. Por ejemplo, un nivel de 10 indica que el valor de pels de luminancia clasificados como bordes se puede alterar aleatoriamente en una gama de niveles de luminancia de -10 a 10 niveles con respecto a su valor original.

- d) Nivel de borrosidad, introducido como números enteros entre 0 y 6, en donde 0 indica sin borrosidad. Los números 1 a 6 seleccionan filtros de paso bajo con frecuencias de corte de 1,5; 1,0; 0,75; 0,5; 0,375 y 0,25 MHz.
- e) Desplazamiento de eco de vibración de contorno, que consta de un número entero entre 0 y 3, en donde 0 selecciona sin simulación su vibración de contorno y 1, 2 ó 3 seleccionan desplazamientos de 0,5; 0,75 y 0,375 milisegundos, respectivamente.
- f) Nivel de amplitud de eco de vibración de contorno. El elemento de datos introducido es un número entero entre -30 y -1 que representa el valor del coeficiente de derivación del filtro para el desplazamiento de eco particular seleccionado.

Con la realización VIRIS, el nivel de trepidación viene especificado por el factor de repetición de tramas (FRF, *frame repetition factor*), introducido como una línea de instrucción variable. Utilizando el fichero de datos, VIRIS inyecta degradaciones en el fichero de imagen SIF en base a los niveles especificados. Después de haber degradado cada trama, se calcula la PSNR para la trama en curso y la PSNR promedio a lo largo de todas las tramas procesadas.

I.5 Plan de pruebas subjetivas del VIRIS

Se efectuaron pruebas subjetivas para determinar la relación entre el nivel de degradación y la opinión subjetiva respecto a degradaciones simuladas por el sistema VIRIS. Por lo general, se utilizaron los procedimientos de prueba y las recomendaciones que figuran en la Recomendación UIT-R BT.500-6 como orientaciones para el diseño del plan de pruebas. Esta Recomendación da directrices sobre los métodos de prueba, las condiciones de observación y las escalas de clasificación que se emplean para evaluar la calidad de las imágenes de televisión. El procedimiento básico seguido en el plan de pruebas consistió en hacer que observadores no expertos evaluaran la calidad de secuencias de imágenes de televisión de 10 segundos que contenían diversos niveles de degradaciones del VIRIS. Los sujetos utilizaron una escala discreta de índices de calidad de cinco puntos que va desde EXCELENTE a INACEPTABLE. El plan de pruebas y los resultados indicados en las dos subcláusulas siguientes describen las relaciones entre nivel de degradación y opinión subjetiva para un subconjunto de degradaciones tratadas en I.2. De manera específica, no se han efectuado pruebas subjetivas en este momento para formular relaciones entre nivel de degradación y la opinión en el caso de distorsión de bloque, ruido correlacionado con la señal y degradación por trepidación basadas en los algoritmos descritos en este apéndice. La información presentada a continuación se extrajo de las cláusulas pertinentes de [2].

I.5.1 Secuencias de imágenes del VIRIS

Se seleccionaron tres secuencias de imágenes para las pruebas con el objeto de caracterizar las degradaciones de ruido y borrosidad. El número seleccionado en concreto representó un compromiso entre el deseo de disponer de una muestra amplia con la que representar distintos tipos de imágenes (con la ventaja adicional de proporcionar una variedad de temas a los sujetos de la prueba) y el tiempo de procesamiento requerido para preparar el material de prueba. Cada una de las secuencias tuvo una duración de 5 segundos y, después del procesamiento, se repitió durante el proceso de edición para crear una secuencia de 10 segundos.

Las secuencias se seleccionaron con miras a obtener diversos grados de movimiento y detalle. Sus títulos, con una breve descripción, son como sigue:

- a) "Bond" – Esta secuencia se extrajo de un disco láser. Es una secuencia de dos personas, Bond y una mujer, cabalgando en el mismo caballo en un escenario desierto. Esta secuencia tiene pocos detalles y un movimiento de caballo "al paso" lento.

- b) "Persecución" – Esta secuencia se extrajo de una película de disco láser. Es una secuencia en la que un adolescente corre en monopatín perseguido por otros adolescentes. Esta secuencia tiene bastantes detalles y movimiento con tres cortes de escena.
- c) "Fútbol" – Esta secuencia se extrajo de una secuencia de 20 segundos que se había utilizado en pruebas anteriores para evaluar algoritmos MPEG-1. Se trata de una secuencia de un partido de fútbol que tiene bastantes detalles y movimiento moderado.

I.5.2 Condiciones de prueba del VIRIS

Para las degradaciones por ruido y borrosidad, el sistema VIRIS procesó 12 veces (2 degradaciones x 6 niveles de degradación) cada secuencia de imagen. Los niveles de degradación se seleccionaron en base a los resultados de pruebas de observación informales y en una prueba piloto limitada con observadores no expertos. En el Cuadro I.3 figuran las PSNR calculadas por el VIRIS para cada una de las tres imágenes y la PSNR media de las tres proyecciones para todos los niveles de degradación. El cuadro muestra también la entrada de datos en el VIRIS (leídos en un fichero de datos) para cada uno de los seis niveles de las dos degradaciones.

CUADRO I.3/P.930

PSNR para niveles de degradación del VIRIS

Nivel de degradación	Entrada en el VIRIS	PSNR media, dB			
		Bond	Persecución	Fútbol	Valor medio de las tres proyecciones
QN1	1	60,8	60,4	59,6	60,3
QN2	3	55,3	54,9	54,2	54,8
QN3	7	52,2	51,9	51,1	51,7
QN4	15	48,7	48,4	47,8	48,3
QN5	62	42,6	42,3	41,6	42,2
QN6	125	39,4	39,1	38,5	39,0
BLR1	1	47,2	41,8	42,7	43,9
BLR2	2	43,4	38,1	38,1	39,9
BLR3	3	40,0	35,4	34,2	36,5
BLR4	4	38,6	33,7	32,6	35,0
BLR5	5	36,3	31,8	30,1	32,7
BLR6	6	34,2	30,2	28,0	30,8

BLR Borrosidad
QN Ruido de cuantificación.

La PSNR varía considerablemente en las tres secuencias de imágenes con ambas degradaciones, por borrosidad y ruido. La variación con la degradación por borrosidad es del orden de 1 a 7 dB, dependiendo del nivel. La variación de la PSNR a lo largo de la secuencia con la degradación de ruido es considerablemente menor, del orden de 1 dB.

En las Figuras I.2 y I.3 se muestran gráficos de la PSNR media de las tres proyecciones, en relación con los dos niveles de degradación de entrada en el VIRIS. Se muestra también una función exponencial ajustada a los datos en cada uno de los gráficos para facilitar el paso de PSNR a entrada en el VIRIS para cada una de las degradaciones.

Como se vio previamente, las 150 tramas de cada secuencia de imágenes de cinco segundos procesadas por el VIRIS fueron convertidas del formato de imagen SIF al CCIR 601 y luego

transferidas al disco. A continuación, se transfirió cada secuencia dos veces seguidas a una cinta digital para convertirse en una secuencia de 10 segundos. Puesto que esta prueba subjetiva para caracterizar los efectos parásitos de ruido y borrosidad, caracterizaba también las degradaciones anteriores por distorsión de bloque y vibración de contorno, el VIRIS procesó realmente 72 secuencias de imágenes (3 secuencias \times 4 degradaciones \times 6 niveles de degradación) junto con las tres imágenes de referencia no procesadas de calidad SIF. Las secuencias se almacenaron en una cinta. Una vez almacenado todo el material de prueba en cinta digital, se registró en un grabador de vídeo de disco láser. Este sistema permite efectuar grabaciones de alta calidad del sistema NTSC y reproducción por acceso aleatorio del material grabado. Fue controlado por programa de computador para reproducir el material de prueba en diferentes órdenes aleatorios, según lo previsto en el diseño de las pruebas.

I.5.3 Método de prueba del VIRIS

En las pruebas se siguió un procedimiento de un solo estímulo en el que se utiliza una escala discreta de calidad de cinco puntos. Los observadores evaluaron cada una de las condiciones de prueba haciendo una marca junto a la categoría de valoración apropiada de la escala de calificación en la que figuraban los términos EXCELENTE, BUENA, ACEPTABLE, MEDIOCRE e INACEPTABLE. Para cada condición de prueba había una hoja de votación.

Participaron un total de 23 observadores no expertos, seleccionados de un grupo de individuos ajenos a la organización de las pruebas. Los sujetos fueron 3 hombres y 20 mujeres con edades comprendidas entre 25 y 67 años y una media de edad de 52,9. Todos ellos tenían una agudeza visual normal (igual o superior a 20/25) y visión normal de los colores.

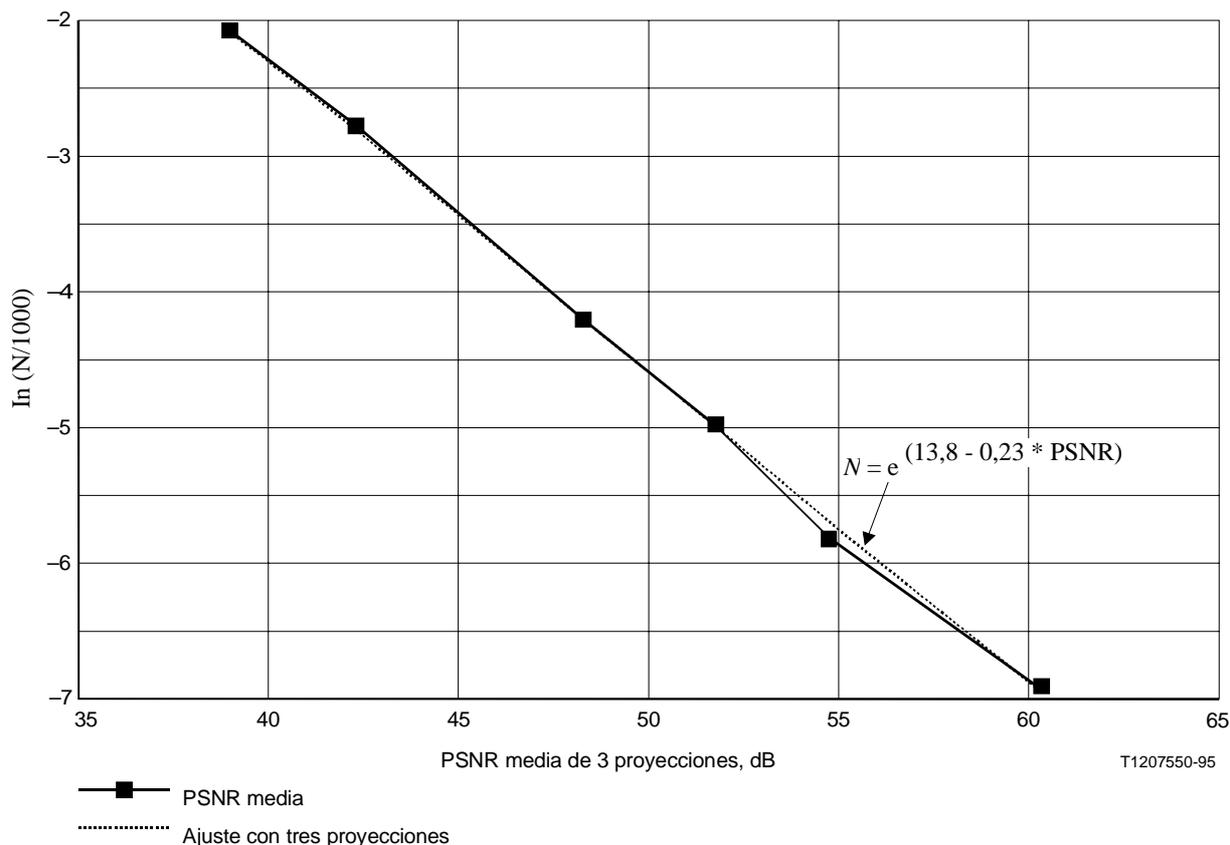


FIGURA I.2/P.930

PSNR de ruido de cuantificación con relación a entrada de datos de QN en el VIRIS

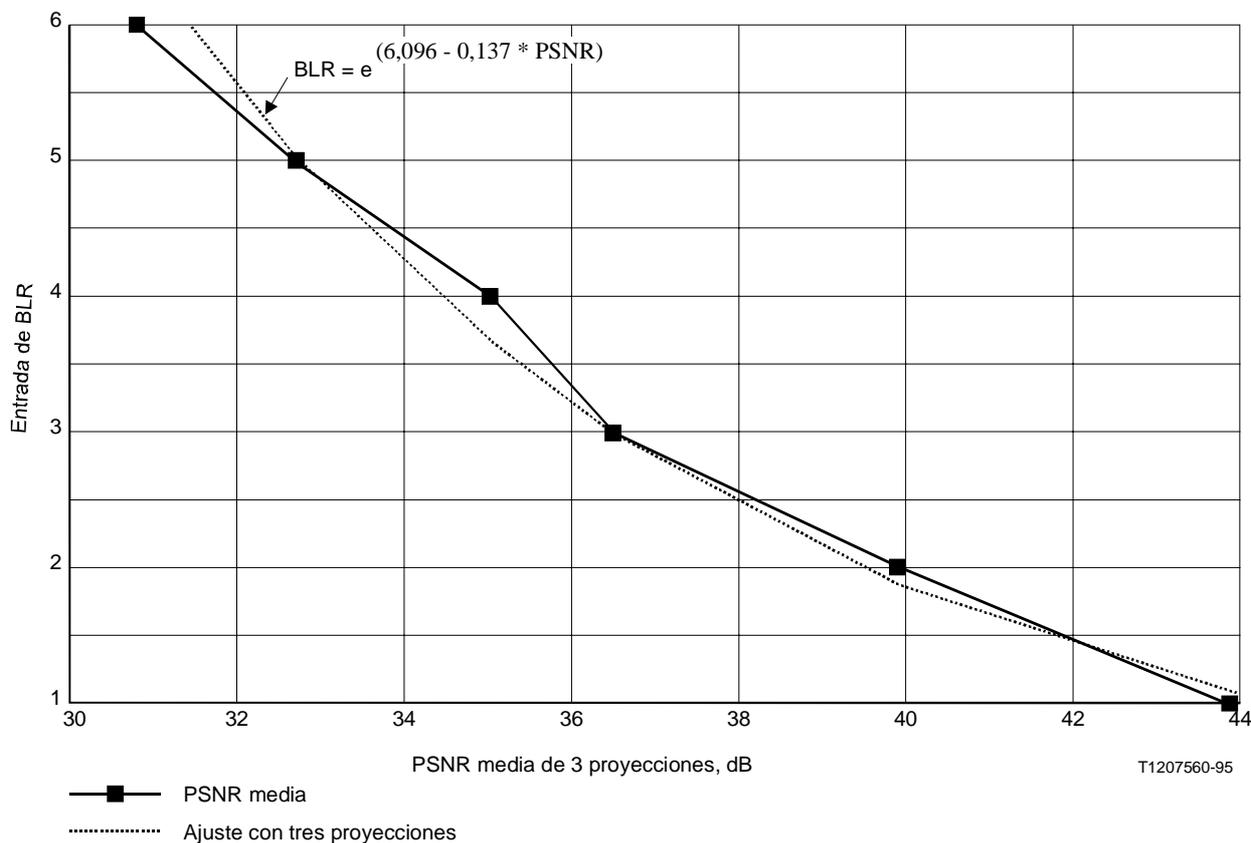


FIGURA I.3/P.930

PSNR de borrosidad con relación a entrada de datos de BLR en el VIRIS

I.5.4 Condiciones de observación del VIRIS

Los sujetos de la prueba observaron un receptor de televisión. Se sentaron a una distancia de observación de 7,5 veces la altura ($7,5H$) de la imagen (a unos 2,30 m frente a la pantalla), lo que viene a ser el valor medio de las dos distancias de observación utilizadas en pruebas anteriores de algoritmos de velocidad binaria comprimida. Una distancia de observación de $5H$ se encuentra en la mitad de la gama recomendada por el UIT-R y una distancia de $10H$ es aproximadamente la distancia de observación media en Estados Unidos. La distancia de $7,5H$ utilizada en este estudio se estableció como un compromiso entre una distancia ($5H$) que el autor consideró demasiado corta para una observación confortable en las pruebas efectuadas y el deseo de obtener valoraciones más críticas, que podrían conseguirse con una distancia de $10H$.

La sala de pruebas era de $3,65 \times 6,10$ m, con paredes aisladas acústicamente de color gris claro e iluminación incandescente de techo, controlada. La iluminación del ambiente y la luminancia de la pantalla se midieron con un fotómetro digital. La iluminación del ambiente de la sala de pruebas se redujo unos 6,5 lux en las posiciones de asiento de los sujetos, ajustando los reguladores de intensidad luminosa de las lámparas de techo. La iluminación del ambiente resultante en el receptor fue de 1,1 lux aproximadamente. Se ajustaron los controles del receptor para tener una luminancia de cresta de 24,8 mililambert aproximadamente y una luminancia mínima de 0,16 mililambert aproximadamente para una relación de contraste en torno a 150. Para efectuar estos ajustes se utilizaron señales de prueba de 100 IRE y ráfaga de negro procedentes de un generador digital.

I.5.5 Procedimientos de prueba del VIRIS

En el transcurso de la prueba subjetiva se utilizó un total de 81 condiciones de prueba básicas. Las condiciones consistieron en 72 secuencias SIF (3 secuencias \times 4 degradaciones \times 6 niveles de degradación) procesadas por el VIRIS (los algoritmos anteriores de distorsión de bloque y vibración

de contorno y los algoritmos actuales de borrosidad y ruido de cuantificación) y 3 repeticiones de cada una de las tres secuencias de vídeo SIF no procesadas. Cada condición duró 20 segundos e incluyó un mensaje de 2 segundos en la pantalla del receptor identificando el número de la condición, una secuencia de imagen de 10 segundos y un periodo de votación de 8 segundos con un mensaje en pantalla en el que se solicitaba al observador que emitiera su voto sobre la calidad de la imagen. Cada una de las 81 condiciones fue probada dos veces, lo que hizo un total de 162 condiciones.

Las pruebas se efectuaron con tres observadores a la vez, salvo en una prueba en la que solo participaron dos (un observador convocado no se presentó). Los sujetos se sentaron frente al receptor a una distancia de observación de 7,5H. Anotaron su nombre y posición de asiento en la primera página de un cuaderno que contenía una hoja de votación para cada práctica y condición de prueba. El conductor de la prueba leyó a continuación las instrucciones correspondientes. Tras contestar a algunas preguntas, se llevó a cabo una sesión práctica de 11 condiciones, para familiarizar a los sujetos con el material y los procedimientos de prueba.

La prueba se dividió en dos partes, con una pausa de 20 minutos para un refrigerio entre ellas y un breve intervalo de 2 a 3 minutos en cada parte después de mostrar 40 secuencias. Cada parte incluyó las 81 condiciones de prueba básicas y duró unos 30 minutos. Las condiciones se presentaron, en ambas partes, en un orden pseudoaleatorio con dos constricciones, a saber:

- a) el primero, segundo y tercer grupo de 27 condiciones contenían cada uno todos los niveles de degradación para los cuatro tipos de degradación incluidas las tres imágenes SIF de referencia, y
- b) cada uno de los tres grupos de 27 condiciones contenía un número igual de las tres secuencias de imágenes.

Estas constricciones tenían por objeto compensar la influencia en los resultados de cualquier efecto aprendizaje, tal como la súbita constatación de un aspecto particular de una degradación sólo después de que la prueba haya avanzado ya bastante, o quizás un cambio gradual en los criterios de valoración de la calidad del sujeto mientras la prueba sigue su curso. Los órdenes aleatorios fueron diferentes en cada mitad de la prueba y se cambiaron para cada grupo de sujetos.

I.5.6 Resultados de la prueba subjetiva del VIRIS

I.5.6.1 Análisis de los datos

El análisis de los datos consistió en calcular la nota media de opinión (MOS, *mean opinion score*) y los intervalos de confianza del 95% para cada nivel de degradación y SIF de referencia, para cada una de las tres secuencias de imágenes. En primer lugar se asignaron los números 5, 4, 3, 2 y 1 para las opiniones de votación de los sujetos de EXCELENTE, BUENA, ACEPTABLE, MEDIOCRE e INACEPTABLE, respectivamente. La MOS de cada condición se calculó aplicando la siguiente fórmula:

$$MOS = \frac{1}{N} \sum_{i=1}^N V_i \quad (\text{I.5-1})$$

donde:

V_i Voto

N Número de votos

El número total de votos de las condiciones de degradación fue de 46 (23 sujetos \times 2 repeticiones) y para las condiciones de SIF de referencia fue de 138 (23 sujetos \times 6 repeticiones). Se computaron los intervalos de confianza (CI, *confidence intervals*) del 95% de las MOS a partir de las medias de los

23 sujetos para cada nivel de degradación y SIF de referencia aplicando la fórmula de la ecuación (I.5-2) [4]:

$$95\% \text{ CI} = \mu \pm \frac{t \cdot \sigma}{\sqrt{N}} \quad (\text{I.5-2})$$

donde:

95% CI Intervalo de confianza del 95%

μ MOS

t Estadístico t de dos colas para un nivel de significación de 0,05 = 2,069

σ Desviación típica de las medias de los sujetos

N Número de medias de sujetos (23)

I.5.6.2 Resultados del VIRIS

En el Cuadro I.4 se muestran los resultados, incluida la PSNR, la MOS y el CI del 95% para todos los niveles de degradación y SIF de referencia de cada una de las tres secuencias de imágenes.

Sólo se presentaron los resultados obtenidos para los algoritmos de la simulación de degradaciones que se describe en I.2. Si bien la prueba subjetiva real caracterizó un conjunto de cuatro degradaciones: distorsión de bloque, vibración de contorno, borrosidad y ruido de cuantificación, se han modificado los algoritmos para simular las dos primeras degradaciones, y, por tanto, los resultados obtenidos durante esta prueba subjetiva no son aplicables a los algoritmos de simulación actuales. En la subcláusula I.6 se describe una segunda prueba subjetiva para caracterizar la degradación por vibración de contorno basada en la técnica de simulación descrita en I.2.

Las MOS para las tres imágenes de referencia, "Bond", "Persecución" y Fútbol" fueron de 4,0, 3,8 y 4,2, respectivamente. Estos índices de opinión son algo bajos para las condiciones de referencia, pero no imprevistos ya que el formato de imagen SIF limita la calidad de la imagen debido a su limitada resolución horizontal y vertical. En una prueba anterior con una escala continua de calidad, el índice medio de seis secuencias de imágenes SIF fue de en torno a 60, estando situada la región de opinión de calidad BUENA entre los índices de 60 y 80. Los resultados actuales con secuencias de imágenes SIF no son incoherentes con los resultados anteriores.

La anchura de los intervalos de confianza (CI) del 95% varió de 0,3 a 0,7. En general, las anchuras de los CI se reducen a medida que disminuye la calidad de la imagen, lo que indica que los sujetos coincidieron más en sus valoraciones de calidad de imagen mediocre que en sus valoraciones de calidad de imagen buena. Normalmente, cuando la gama de degradaciones abarque la gama completa de la escala de calificación, las anchuras de los CI serán menores en ambos extremos de la escala de calificación y mayores para calidades comprendidas entre dichos extremos.

Resultados de las pruebas subjetivas para degradaciones del VIRIS

Degradación	Secuencia de imágenes								
	Bond			Persecución			Fútbol		
Tipo/nivel	PSNR (dB)	MOS	Anchura CI X	PSNR (dB)	MOS	Anchura CI X	PSNR (dB)	MOS	Anchura CI X
SIF de ref.	-	4,0	0,7	-	3,8	0,6	-	4,2	0,7
QN1	60,8	3,9	0,6	60,4	3,8	0,6	59,6	3,9	0,6
QN2	55,3	3,4	0,6	54,9	3,6	0,6	54,2	3,3	0,6
QN3	52,2	2,7	0,5	51,9	3,4	0,6	51,1	3,3	0,6
QN4	48,7	2,5	0,4	48,4	3,0	0,5	47,8	2,7	0,5
QN5	42,6	1,8	0,4	42,3	1,8	0,4	41,6	2,2	0,4
QN6	39,4	1,8	0,4	39,1	1,7	0,3	38,5	2,0	0,4
BLR1	47,2	3,8	0,7	41,8	3,3	0,6	42,7	3,7	0,6
BLR2	43,4	3,4	0,6	38,1	2,7	0,5	38,1	3,0	0,5
BLR3	40,0	3,0	0,5	35,4	2,0	0,4	34,2	2,6	0,5
BLR4	38,6	2,3	0,5	33,7	1,7	0,4	32,6	2,1	0,4
BLR5	36,3	1,9	0,4	31,8	1,5	0,3	30,1	1,3	0,3
BLR6	34,2	1,6	0,3	30,2	1,3	0,3	28,0	1,3	0,3
PSNR	Relación señal/ruido de cresta								
MOS	Nota media de opinión								
MOS X	Intervalo de confianza del 95%								
SIF de ref.	SIF de referencia								
QN	Ruido de cuantificación								
BLR	Borrosidad.								

En las Figuras I.4 y I.5 se muestran gráficos de la MOS en función de la PSNR de las degradaciones por ruido y borrosidad, respectivamente para las tres secuencias de imágenes. En las figuras se muestra también una función logística ajustada a los datos de las tres proyecciones para cada una de las degradaciones. La función logística es de la forma dada por la ecuación (I.5-3):

$$MOS = K_1 + \frac{K_2}{1 + e^{-K_3(PSNR - K_4)}} \tag{I.5-3}$$

donde:

- K_1 Asíntota inferior de la MOS = 1,0
- K_2 Asíntota superior de la MOS - $K_1 = 3,2$
- K_3 Parámetro de pendiente
- K_4 Valor de la PSNR cuando $MOS = K_1 + \frac{K_2}{2}$

Al ajustar la función logística, se fijó la asíntota inferior, K_1 , en 1,0 y la asíntota superior, $K_1 + K_2$, en 4,2 en todos los casos. El valor de la asíntota inferior de 1,0 es la MOS más baja que puede resultar con la escala de calificación de cinco puntos. El valor de la asíntota superior de 4,2 representa el valor promedio de MOS más alto obtenido en el presente estudio para una de las

condiciones de referencia SIF (secuencia "Fútbol"). Los dos parámetros de ajuste fueron así, K_3 , un parámetro de pendiente, y K_4 , el valor de la PSNR cuando la MOS es el valor intermedio entre los valores asintóticos inferior y superior.

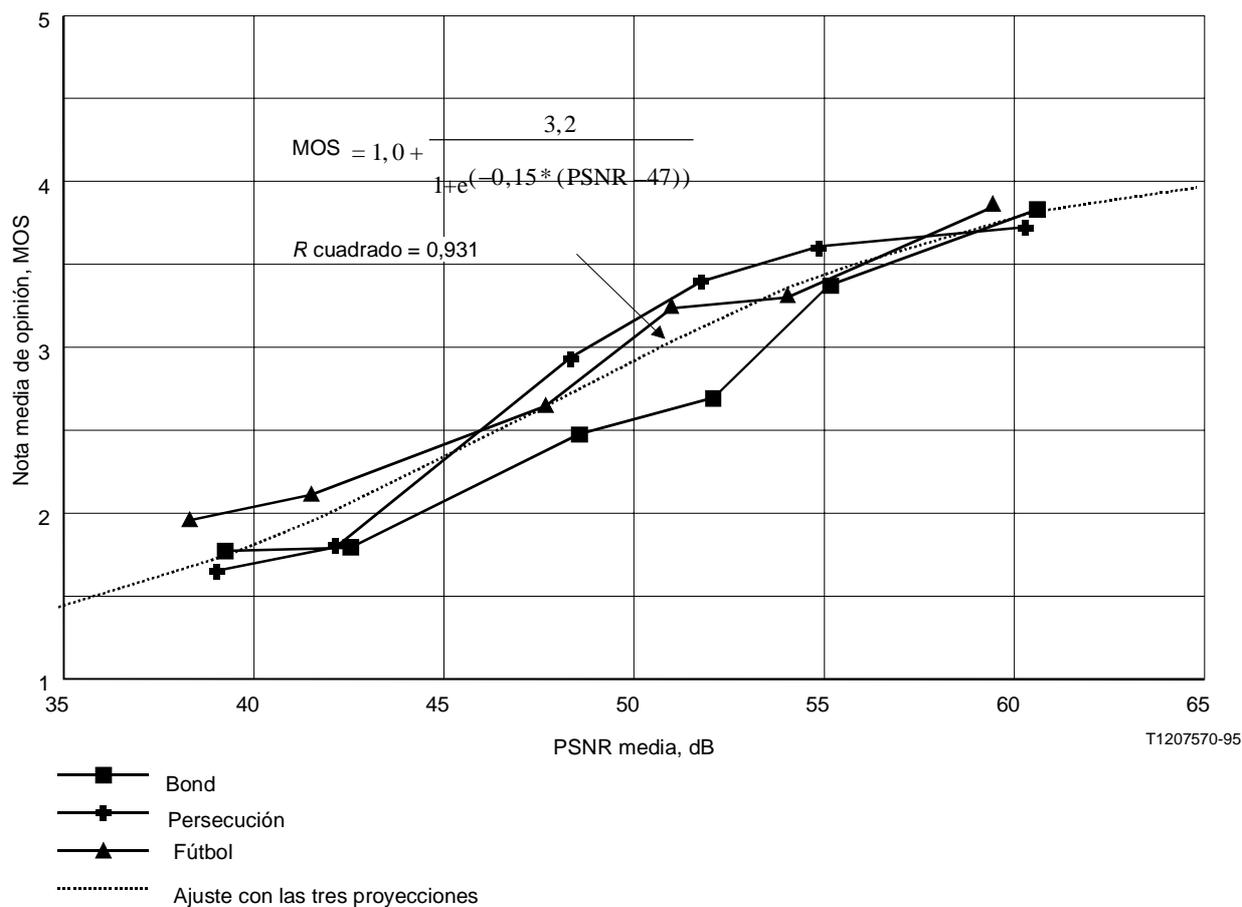


FIGURA I.4/P.930

PSNR de ruido de cuantificación en relación con la MOS

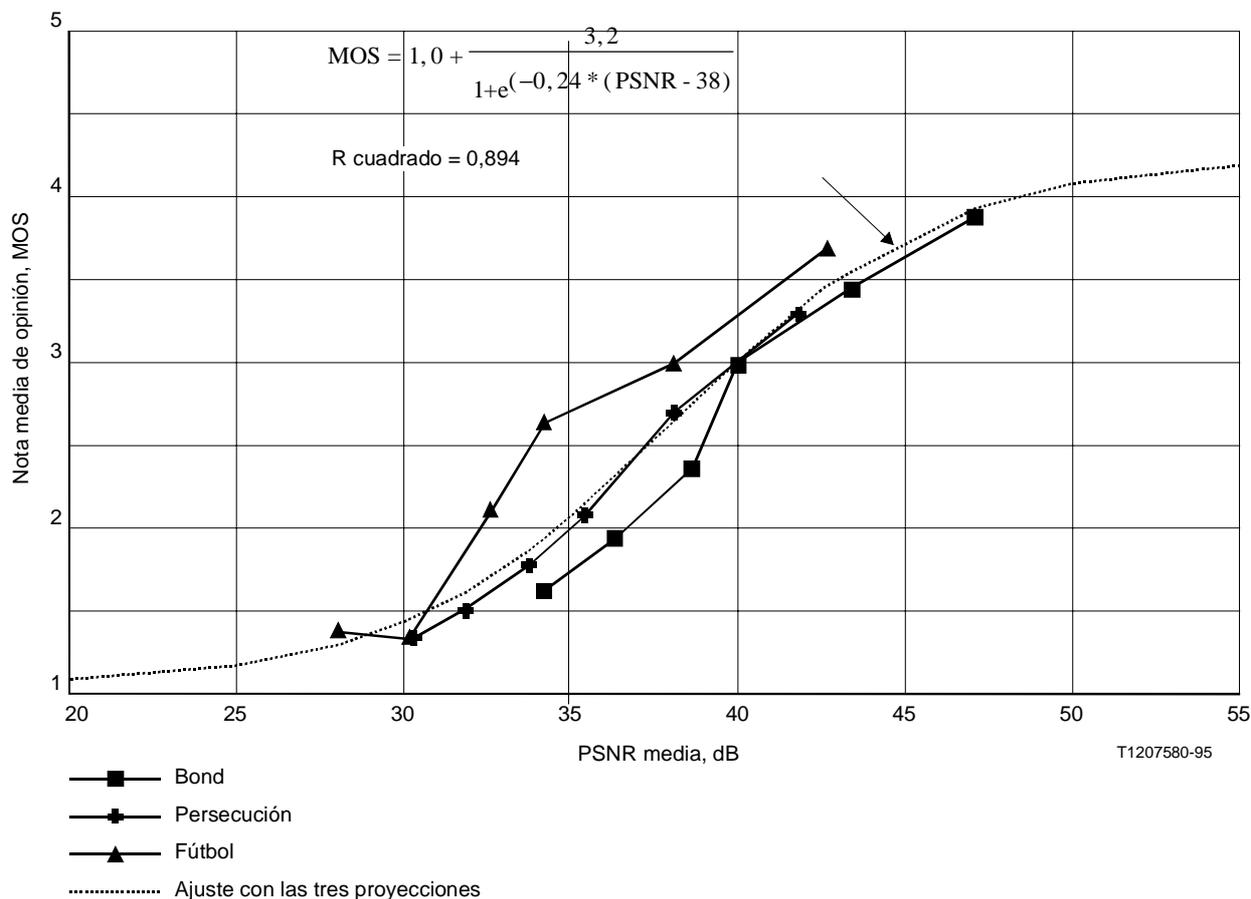


FIGURA I.5/P.930

PSNR de borrosidad en relación con la MOS

Respecto a los resultados del ruido de cuantificación de la Figura I.4, la reacción subjetiva a las PSNR del ruido fueron algo diferentes dependiendo de la imagen que, en concreto, se estaba observando. Sin embargo, las gamas de las MOS de las tres proyecciones son bastante concordantes y la función logística parece que se ajusta bien a los datos de las tres proyecciones con R cuadrado = 0,9309. (R cuadrado es un estadístico de la bondad del ajuste que varía de 0 a 1, en donde 1 representa el ajuste óptimo.)

La Figura I.5 indica que la reacción subjetiva a las PSNR de borrosidad dependió más de las imágenes que en el caso de las PSNR de ruido y distorsión de bloque, con una separación ligeramente mayor entre las MOS de las tres proyecciones. La función logística se ajusta bien todavía a los datos de las tres proyecciones con R cuadrado = 0,8941.

I.6 Plan de pruebas subjetivas de la vibración de contorno

Se efectuó una segunda prueba subjetiva para determinar la relación entre el nivel de degradación y la opinión subjetiva respecto a la degradación por vibración de contorno de acuerdo con la descripción de la sección I.2. Se probaron también otras condiciones, entre ellas, una condición de SIF de referencia, una condición de NTSC de referencia y dos condiciones MPEG-1, una a 1,3 Mbit/s y otra a 1,0 Mbit/s.

Por lo general, se utilizaron los procedimientos de prueba y las Recomendaciones que figuran en la Recomendación UIT-R BT.500-6 como orientaciones para el diseño del plan de pruebas. Esta Recomendación da directrices sobre los métodos de prueba, las condiciones de observación y las

escalas de calificación que se emplean para evaluar la calidad de las imágenes de televisión. El procedimiento básico seguido en el plan de pruebas consistió en hacer que observadores no expertos que evaluaran la calidad de secuencias de imágenes de televisión de 10 segundos que contenían diversos niveles de degradaciones del VIRIS, las condiciones de SIF y NTSC de referencia y las condiciones MPEG-1. Los sujetos utilizaron una escala discreta de índices de calidad de cinco puntos que va de EXCELENTE a INACEPTABLE. En las siguientes subcláusulas figuran detalles del plan de pruebas.

I.6.1 Secuencias de imágenes de vibración de contorno

Se seleccionaron tres secuencias de imágenes para estas pruebas. El número seleccionado en concreto representó un compromiso entre el deseo de disponer de una muestra amplia con la que representar distintos tipos de imágenes (con la ventaja adicional de proporcionar una variedad de temas a los sujetos de la prueba) y el tiempo de procesamiento requerido para preparar el material de prueba. Cada una de las secuencias tuvo una duración de 5 segundos y, después del procesamiento, se repitió durante el proceso de edición para crear una secuencia de 10 segundos.

Las secuencias se seleccionaron con miras a obtener diversos grados de movimiento y detalle. Sus títulos, con una breve descripción, son como sigue:

"Automóviles de carrera" – Esta secuencia se extrajo de una secuencia de 20 segundos que muestra automóviles de carrera en movimiento sobre un fondo de pista estática. La secuencia incluye cuatro cortes de escena para diferentes automóviles y contiene algún detalle (particularmente los números de los automóviles) y movimiento moderado. Procede de un vídeo corto de demostración de TVAD que fue telecargado en formato NTSC.

"Tenis de mesa" – Esta secuencia se extrajo de una secuencia de 20 segundos que muestra a dos personas jugando al tenis de mesa e incluye detalle y movimiento moderados. Hay tres cortes de escena con fondo estacionario en dos de ellas y de movimiento lento en la restante. La secuencia procede de una cinta digital D1 y fue una de las secuencias "típicas" utilizadas para evaluar la calidad de funcionamiento del algoritmo MPEG.

"Fútbol" – Esta secuencia se utilizó en la prueba del VIRIS anterior. Es una secuencia de un partido de fútbol que tiene bastantes detalles y movimiento moderado. No tiene cortes de escena y el fondo se mueve lentamente. Procede de un vídeo corto de demostración de TVAD que fue telecargado en formato NTSC.

I.6.2 Condiciones de prueba del VIRIS y preparación del material de prueba

Se seleccionaron ocho niveles de degradación por vibración de contorno como condiciones de prueba en base a pruebas de observación informales de laboratorio. En el Cuadro I.5 figuran las PSNR calculadas por el VIRIS para cada uno de los niveles de degradación por vibración de contorno, primero como valor medio de cada secuencia de imágenes y a continuación como valor medio de las tres proyecciones.

PSNR para niveles de degradación de vibración de contorno VIRIS

Nivel de vibración de contorno, EB	Valor medio de PSNR, dB			
	Automóviles de carrera	Fútbol	Tenis de mesa	Valor medio de las tres proyecciones
-1	42,9	43,5	41,1	42,5
-3	41,8	40,7	39,5	40,7
-5	40,3	37,8	37,3	38,5
-10	36,5	32,1	32,6	33,7
-15	33,2	28,2	29,1	30,2
-20	30,3	25,1	26,4	27,3
-25	27,7	22,7	24,1	24,8
-30	25,3	20,5	22,1	22,6

NOTA – El nivel de vibración de contorno (EB, *edge busyness level*) corresponde a valores de coeficiente de derivación del filtro de la gama variable de coeficientes dada para los tres filtros multibanda del Cuadro I.2.

Las PSNR de las secuencias "fútbol" y "tenis de mesa" fueron bastante próximas, difiriendo en 0,5 a 2,4 dB en toda la gama de niveles de vibración de contorno. La PSNR de la secuencia "automóviles de carrera" difirió de la PSNR de las otras dos secuencias en 0,6 a 5,2 dB en toda la gama de niveles de vibración de contorno. En la Figura I.6 se muestra un gráfico de la PSNR media de las tres proyecciones en relación con el nivel de vibración de contorno de entrada en el VIRIS. En el gráfico se muestra también una función lineal ajustada a los datos para facilitar el paso de PSNR a entrada de nivel de vibración de contorno en el VIRIS.

Además de las ocho condiciones de degradación por vibración de contorno, la prueba incluyó condiciones de calidad NTSC, calidad SIF y dos condiciones MPEG-1, una a 1,3 Mbit/s y otra a 1,0 Mbit/s, lo que hizo un total de 12 condiciones de prueba. Cada una de las tres proyecciones de secuencias de prueba se sometió a cada una de las 12 condiciones de prueba, con un total de 36 secuencias de prueba. Durante el procesamiento y edición, las secuencias se almacenaron en una cinta digital D1 y posteriormente se transfirieron a un grabador de vídeo de disco láser. El grabador de disco láser permitió registrar imágenes NTSC de alta calidad y la reproducción por acceso aleatorio del material grabado. Fue controlado por un programa de computador para reproducir el material de prueba en diferentes órdenes aleatorios, según lo previsto en el diseño de prueba.

I.6.3 Método de prueba subjetiva de la vibración de contorno

En las pruebas se siguió un procedimiento de un solo estímulo en el que se utiliza una escala discreta de calidad de cinco puntos. Participaron un total de 27 observadores no expertos, seleccionados de un grupo de individuos ajenos a la organización de las pruebas. Los sujetos fueron 3 hombres y 24 mujeres con edades comprendidas entre 19 y 71 años y una media de edad de 47,3. Los observadores evaluaron cada una de las condiciones de prueba haciendo una marca junto a la categoría de valoración apropiada de la escala de calificación en la que figuraban los términos EXCELENTE, BUENA, ACEPTABLE, MEDIOCRE e INACEPTABLE. Para cada condición de prueba había una hoja de votación.

I.6.4 Procedimientos de prueba subjetiva de la vibración de contorno

Las condiciones de observación fueron las mismas que las de la primera prueba del VIRIS (véase I.5.4). Las pruebas se efectuaron con tres observadores a la vez, salvo en una prueba en la que sólo participaron dos (un observador convocado no se presentó). Después que el conductor de la prueba leyera las instrucciones correspondientes y contestara a algunas preguntas, se llevó a cabo una sesión práctica de 12 condiciones, para familiarizar a los sujetos con el material y los procedimientos de prueba. Cada condición de prueba duró 20 segundos e incluyó un mensaje de 2 segundos en la pantalla del receptor identificando el número de la condición, una secuencia de imagen de 10 segundos y un periodo de votación de 8 segundos con un mensaje en pantalla en el que se solicitaba al observador que emitiera un voto sobre la calidad de la imagen. Cada una de las 36 condiciones de prueba se probó dos veces, lo que hizo un total de 72 condiciones.

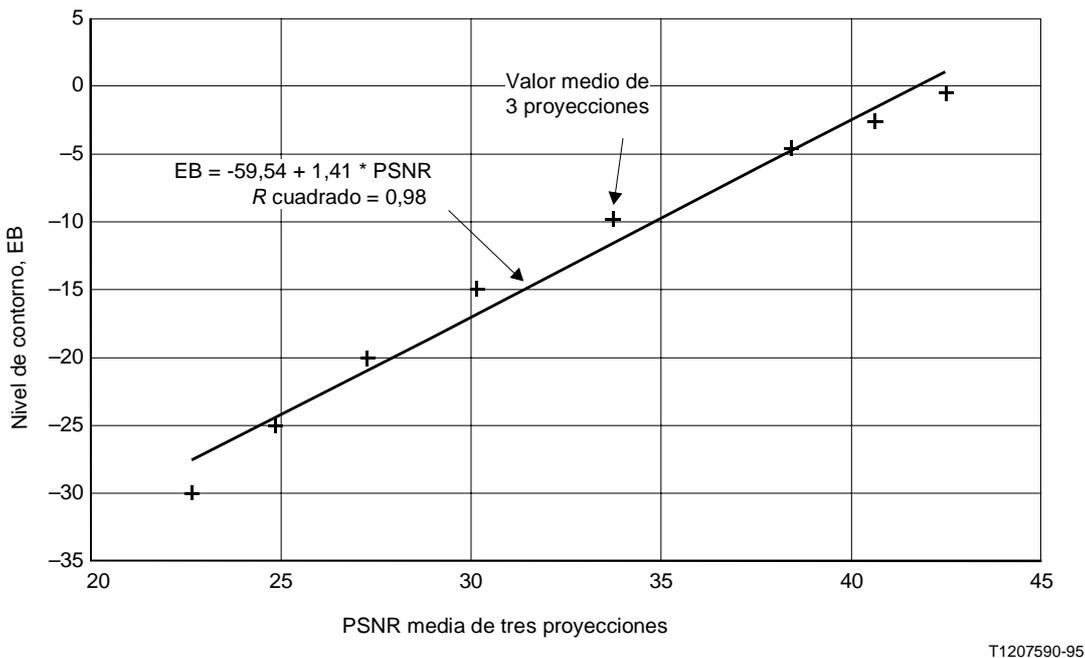


FIGURA I.6/P.930

PSNR de vibración de contorno con relación a la entrada de datos EB a VIRIS

La prueba se dividió en dos partes, con una pausa de 10 minutos para un refrigerio entre ellas. Cada parte incluyó las 36 secuencias de prueba diferentes y se efectuó en 12 minutos. Las condiciones se presentaron, en ambas partes, en un orden pseudoaleatorio con dos constricciones, a saber:

- el primero, segundo y tercer grupo de 12 secuencias de imágenes contenían cada uno las 12 condiciones de prueba básicas, y
- cada uno de los tres grupos de 12 secuencias de imágenes contenía un número igual de las tres secuencias de imágenes.

Éstas tenían por objeto compensar la influencia en los resultados de cualquier efecto de aprendizaje, tal como la súbita constatación de un aspecto particular de una degradación sólo después de que la prueba haya avanzado bastante, o quizás un cambio gradual en los criterios de valoración de la calidad del sujeto mientras la prueba sigue su curso. Los órdenes aleatorios fueron diferentes en cada mitad de la prueba y se cambiaron para cada grupo de sujetos.

I.6.5 Resultado de la prueba subjetiva de vibración de contorno

Se calculó la nota media de opinión (MOS) para cada una de las 12 condiciones de prueba para las 3 secuencias de imágenes. La MOS de las condiciones de referencia NTSC fue de 4,7, 4,6 y 4,6 para las secuencias "automóviles de carrera", "fútbol" y "tenis de mesa". Las MOS de las condiciones

de referencia SIF para las secuencias correspondientes fueron 3,2, 3,6 y 3,8, respectivamente, lo que indicaba una caída substancial en la calidad de NTSC a SIF. Esta pérdida de calidad se debe sobre todo a la disminución de la resolución horizontal y vertical que se produce en la conversión de formato de imagen CCIR 601 a SIF. Teniendo en cuenta que los algoritmos MPEG1 utilizan SIF como formato de imagen de entrada, la calidad asociada con SIF representa también el límite de la calidad que se puede obtener con algoritmos MPEG-1.

La condición de referencia SIF con la secuencia "fútbol" tuvo una MOS de 3,6 que es 0,6 inferior a la obtenida en la prueba del VIRIS. Sin embargo, en la prueba anterior no se utilizaron las condiciones de referencia NTSC como se hizo en esta prueba y, como resultado, el material de estímulo no abarcó la gama completa de la escala de calificación. Puesto que los sujetos probablemente intentan utilizar todas las calificaciones de la escala para valorar el material de estímulo, quizás hayan atribuido a las condiciones de referencia SIF de la prueba anterior una valoración mayor que la que le hubieran atribuido si se hubiese presentado una gama completa de material de estímulo. El propósito de incluir las condiciones NTSC en la segunda prueba subjetiva fue de obtener una cobertura total de la escala de calificación para reducir o eliminar esta especie de problema de anclaje. De este modo, la calidad SIF obtenida en la segunda prueba es probablemente más indicativa de la calidad SIF real que los resultados obtenidos en la primera prueba subjetiva.

En la Figura I.7 se muestran gráficos de la MOS en función de la PSNR de la degradación por vibración de contorno del VIRIS para las tres secuencias de imágenes. En la figura se muestra también una función logística ajustada a los datos de las tres proyecciones. La función logística tiene la misma forma que la utilizada y descrita en el estudio anterior. A propósito de la Figura I.7, la reacción subjetiva a la degradación por vibración de contorno fue algo diferente para las tres secuencias de imágenes de la prueba. Los resultados de las secuencias "fútbol" y "tenis de mesa" fueron similares, pero claramente diferentes de los resultados de la secuencia "automóviles de carrera", especialmente en las PSNR superiores. No obstante, a pesar de las diferencias en los resultados de las tres secuencias de imágenes, la función logística proporciona un buen ajuste a los datos de las tres proyecciones con R cuadrado = 0,90. R cuadrado es un "estadístico de la bondad del ajuste" que varía de 0 a 1, en donde 1 representa el ajuste óptimo. Indica también la proporción de la varianza de los datos que tiene en cuenta el modelo.

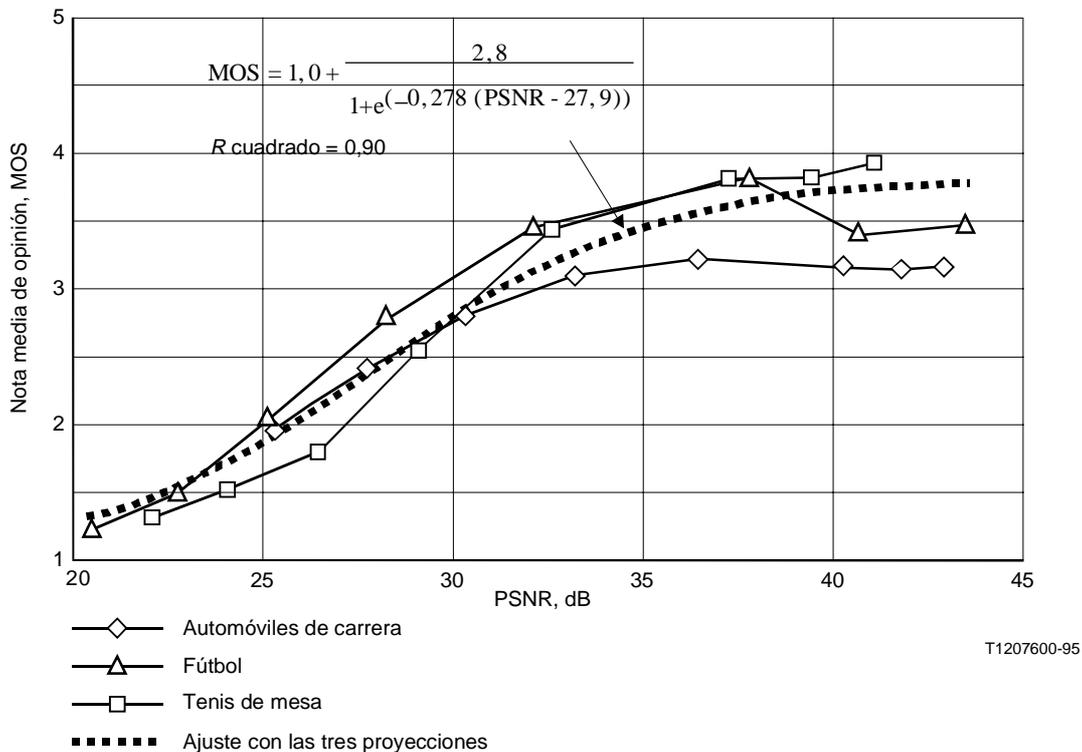


FIGURA I.7/P.930

PSNR de vibración de contorno en relación con la MOS

I.7 Conclusiones

Se necesita un sistema de degradación de referencia vídeo ajustable para la evaluación de la calidad de imagen de sistemas vídeo digitales comprimidos. Dicho sistema ha de producir una amplia gama de calidades de imagen introduciendo degradaciones de imágenes simuladas similares a las producidas por los codificadores digitales. El sistema de degradación de referencia vídeo (VIRIS) descrito en este apéndice es el comienzo hacia ese objetivo.

El VIRIS puede introducir simulaciones de degradaciones de distorsión de bloque, borrosidad, vibración de contorno, ruido correlacionado con la señal, ruido de cuantificación y trepidación (movimiento con sacudidas) en las secuencias de imágenes SIF. Estas degradaciones se pueden introducir solas o en combinación. Se pueden caracterizar objetivamente en términos del nivel de degradación y de la relación señal/ruido de cresta media (PSNR) en todas las tramas de la secuencia de imágenes. Los resultados de la prueba subjetiva con tres de estas degradaciones (la distorsión de bloque, el ruido correlacionado con la señal y la trepidación no han sido caracterizadas subjetivamente) indican que la reacción subjetiva a las degradaciones simuladas abarca una amplia gama de calidades de imagen. Estas relaciones se muestran en las Figuras I.4, I.5 y I.7 para el ruido de cuantificación, la borrosidad y la vibración de contorno, respectivamente, para cada una de las tres secuencias de imágenes probadas. Una función logística ajustada a los datos de las tres proyecciones relaciona la MOS con la PSNR para cada una de las tres degradaciones.

Las Figuras I.2, I.3 y I.6 muestran gráficos de las PSNR medias de las tres proyecciones y el nivel de entrada en el VIRIS para cada una de las tres degradaciones. En los gráficos se muestra también una función exponencial ajustada a los datos para relacionar la PSNR con la entrada de datos en el VIRIS para cada una de las 4 degradaciones; para el VIRIS se utilizó una función lineal. El empleo de estas funciones junto con las funciones logísticas mostradas en las Figuras I.4, I.5 y I.7 facilita el paso de

la MOS deseada a la entrada de datos en el VIRIS para cada una de las cuatro degradaciones de las escenas utilizadas en los resultados de la prueba. No se recomienda, por ahora, utilizar estas funciones para predecir los resultados con otras escenas; este tema ha de ser objeto de un estudio ulterior.

Ha de proseguir el trabajo en relación con el sistema VIRIS, trabajo que incluye la simulación y la prueba subjetiva de diversas combinaciones de degradaciones. El objetivo es relacionar la opinión subjetiva deseada, MOS, con la entrada de datos en el VIRIS para diferentes combinaciones de degradaciones. El resultado puede tener la forma de una relación simple para una combinación de todas las degradaciones o, quizás, de relaciones múltiples (que incluirían las 5 elaboradas en el presente estudio) para combinaciones específicas de degradaciones. El trabajo futuro exige un mayor conocimiento acerca de las degradaciones reales provocadas por los algoritmos de compresión de los codificadores vídeo digitales, así como la manera de simular estas degradaciones, y el desarrollo de nuevos métodos con los que correlacionar mediciones objetivas de calidad con una opinión subjetiva.

Bibliografía

- [1] UIT-T COM 12-53 (1995), *Addition of Motion Correlated Artifacts to VIRIS*, Bellcore.
- [2] UIT-T COM 12-21 (1993), *VIRIS, An Experimental Video Reference Impairment System*, Bellcore.
- [3] UIT-T COM 12-35 (1994), *Modification of Video Reference Impairment System (VIRIS) to improve Edge Busyness Impairment Simulation*, Bellcore.
- [4] HAYS (William L.): *Statistics or Psychologists*, Holt, Rinehart and Winston, 1963.
- [5] ANIL (Jain): *Fundamentals of Digital Image Processing*, Prentice Hall, 1989
- [6] *Digital Filter Design Package*, Version 2.12, Atlanta Signal Processors Incorporated, diciembre, 1986.

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie B	Medios de expresión
Serie C	Estadísticas generales de telecomunicaciones
Serie D	Principios generales de tarificación
Serie E	Red telefónica y RDSI
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión
Serie H	Transmisión de señales no telefónicas
Serie I	Red digital de servicios integrados (RDSI)
Serie J	Transmisiones de señales radiofónicas y de televisión
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	Mantenimiento: sistemas de transmisión, circuitos telefónicos, telegrafía, facsímil y circuitos arrendados internacionales
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Equipos terminales y protocolos para los servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos y comunicación entre sistemas abiertos
Serie Z	Lenguajes de programación