

RECOMENDACIÓN UIT-R BT.1129-2

**EVALUACIÓN SUBJETIVA DE SISTEMAS DE TELEVISIÓN DIGITAL
CON DEFINICIÓN NORMALIZADA (SDTV)**

(Cuestión UIT-R 211/11)

(1994-1995-1998)

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que varias administraciones y organizaciones en el mundo están evaluando actualmente los sistemas digitales y que es probable que en muchas partes del mundo la radiodifusión digital se convierta en el medio principal de difusión del siglo próximo;
- b) que las evaluaciones subjetivas son un elemento vital en el diseño y la comparación de los sistemas digitales;
- c) que en la Recomendación UIT-R BT.500 se describen algunos métodos preferidos para las evaluaciones subjetivas, muchos de cuyos detalles son apropiados en el contexto de la televisión digital;
- d) que implícitamente, y en virtud de las técnicas de transporte de datos utilizadas, los sistemas digitales ofrecen la oportunidad de introducir en el canal de radiodifusión trenes de bits correspondientes a varios programas o métodos de codificación jerárquica o ajustable por escalón;
- e) que en los sistemas de televisión digitales el contenido del programa puede tener considerable influencia en la calidad de imagen,

recomienda

- 1** que los métodos generales para la evaluación subjetiva de los sistemas digitales de definición normalizada, incluidos los que ofrecen múltiples trenes de bits de programa o métodos de codificación jerárquica o ajustable por escalón, sean los descritos en la Recomendación UIT-R BT.500;
- 2** que los procedimientos específicos para la evaluación subjetiva de los sistemas digitales, incluidos los que ofrecen múltiples trenes de bits de programa o métodos de codificación jerárquica o ajustable por escalón, sean los que se indican en el Anexo 1.

ANEXO 1

1 Introducción

En este Anexo, que se utilizará conjuntamente con la Recomendación UIT-R BT.500, se proporcionan detalles sobre la aplicación de los métodos generales indicados en la Recomendación a las evaluaciones subjetivas de los sistemas digitales cuyo nivel de calidad es el de los sistemas de televisión convencional o aproximado; los detalles de procedimiento que se ofrece en esta Recomendación, junto con la información general, tratan de las pruebas de códecs (o sistemas) utilizados para cursar material elaborado según la Recomendación UIT-R BT.601 por aplicaciones de contribución y distribución, así como los utilizados en las aplicaciones de emisión.

Las especificaciones de calidad en el caso de aplicaciones de distribución pueden expresarse en términos de la apreciación subjetiva de los observadores. En teoría, por tanto, esos códecs pueden evaluarse subjetivamente, contrastándolos con estas especificaciones. Sin embargo, la calidad de un códec diseñado para aplicaciones de contribución no podría especificarse teóricamente en términos de parámetros subjetivos de calidad de funcionamiento, porque su salida no está destinada a una visualización inmediata sino a tratamiento posterior en estudio, almacenamiento y/o codificación para transmisión ulterior. Dada la dificultad de definir esa calidad de funcionamiento para una diversidad de operaciones de tratamiento posterior, el enfoque preferido ha sido especificar la calidad de funcionamiento de una cadena de equipo, incluyendo una función de tratamiento posterior, a la que se considera representativa de una

aplicación práctica de contribución. Esta cadena podría constar típicamente de un códec, seguido por una función de tratamiento posterior de estudio (o de otro códec en el caso de evaluación de calidad de contribución básica) seguido todavía por otro códec antes de que la señal alcance al observador. La adopción de esta estrategia para las especificaciones de códecs destinados a aplicaciones de contribución significa que los procedimientos de medición que se dan en la presente Recomendación pueden también utilizarse para su evaluación.

Aunque se esté progresando al respecto, en la actualidad no se dispone de suficiente experiencia para dar detalles sobre métodos de evaluación objetiva de la calidad de imagen de los códecs. En materia de evaluación subjetiva, de la que existe mucha experiencia, se pueden hacer recomendaciones sobre condiciones de prueba y metodologías. Debe recordarse no obstante, al especificar objetivos de calidad o degradación, que los métodos existentes no pueden dar valoraciones subjetivas absolutas sino más bien resultados que están influidos en cierta medida por la elección de las condiciones de referencia y/o fijación. Pueden adoptarse las mismas metodologías para códecs de longitud de palabra fija y variable y para códecs de intratrama e intercuadro, aunque la elección de las secuencias de imágenes de prueba puede verse influenciada.

El método de evaluación más fiable para establecer un orden de jerarquía para los códecs de gran calidad consiste, en la actualidad, en evaluar todos los sistemas presentados al mismo tiempo y en condiciones idénticas. Las pruebas hechas independientemente, en las que se evalúan diferencias de calidad muy pequeñas, deben servir de guía más bien que de evidencia incuestionable de superioridad.

Una medida subjetiva útil puede ser la degradación determinada como una función de la velocidad a la que se producen los bits erróneos de transmisión en el enlace entre el codificador y el decodificador. En la actualidad no se tiene conocimiento experimental suficiente de estadísticas ciertas de errores de transmisión, que permitan recomendar parámetros para un modelo que tenga en cuenta las agrupaciones o ráfagas de errores. En tanto no se disponga de esta información, pueden utilizarse los errores con la distribución de Poisson.

2 Condiciones de observación

Las condiciones de observación generales para las evaluaciones subjetivas son las que se indican en el § 2.1 de la Recomendación UIT-R BT.500. Las condiciones de observación específicas para la evaluación subjetiva de los sistemas digitales se indican en los apartados que figuran a continuación.

2.1 Entorno del laboratorio

El entorno del laboratorio crea las condiciones críticas para la prueba de sistemas. Las condiciones de observación específicas para la evaluación subjetiva en laboratorio se indican en el Cuadro 1.

CUADRO 1

Condiciones de observación específicas para la evaluación subjetiva de sistemas digitales en laboratorio

Condición	Descripción	Valores
a	Relación entre la distancia de observación y la altura de la imagen	4 H y 6 H ¹⁾
b	Luminancia de cresta	70 cd/m ²
c	Ángulo de observación subtendido por la parte del fondo que satisface las especificaciones	≥ 43° de altura × 57° de anchura
d	Imagen	Pantalla de alta calidad. Tamaño ≥ 20 pulgadas (50 cm) ²⁾

¹⁾ 6 H es la distancia de observación nominal (DVD, design viewing distance) para evaluar sistemas digitales de definición normalizada, pero también es aceptable utilizar evaluadores con 4 H , siempre y cuando los resultados se indiquen por separado.

²⁾ Como hay ciertas indicaciones de que el tamaño de la imagen puede influir en los resultados de la evaluación subjetiva, se pide a quienes realizan los experimentos que comuniquen el tamaño de la pantalla y la marca y el modelo del aparato utilizado en los mismos.

2.2 Entorno doméstico

Este entorno ofrece los medios necesarios para evaluar la calidad de la cadena de televisión digital desde el punto de vista del consumidor. Las condiciones de observación específicas para la evaluación subjetiva de la televisión digital de definición convencional (SDTV) e el entorno doméstico se indican en el Cuadro 2.

CUADRO 2

Condiciones de observación específicas para la evaluación subjetiva de sistemas digitales en el entorno doméstico

Condición	Descripción	Valores
a	Relación entre la distancia de observación y la altura de la imagen	6 <i>H</i>
b	Tamaño de la pantalla para una relación de formato de 4/3	De 25 a 29 pulgadas ⁽¹⁾
c	Tamaño de pantalla para una relación de formato de 16/9	De 32 a 36 pulgadas ⁽¹⁾
d	Norma del monitor	SDTV
e	Luminancia de cresta	200 cd/m ²
f	Iluminación ambiental sobre la pantalla (La luz ambiental que incide sobre la pantalla debe medirse perpendicularmente a ésta)	200 Lux

⁽¹⁾ Este tamaño de pantalla cumple las normas en materia de distancia de observación preferida (PVD, preferred viewing distance) para una PVD = 6 *H*.

3 Métodos de evaluación

3.1 Evaluación de la calidad de imagen básica

Cuando se evalúa un códec para aplicaciones de distribución, esta calidad se refiere a las imágenes decodificadas después de un paso único a través de un par de códecs. En el caso de códecs de contribución, puede evaluarse la calidad básica después de varios códecs en serie, con el fin de simular así una aplicación típica de contribución.

Cuando la gama de calidades por evaluar es pequeña, lo que ocurrirá normalmente en el caso de códecs de televisión, la metodología de prueba a utilizar es la variante II del método de doble estímulo con escala de calidad continua que se describe en la Recomendación UIT-R BT.500. La secuencia fuente original se utilizará como condición de referencia. Se sigue debatiendo a propósito de la duración de la secuencia de presentación. En pruebas recientes efectuadas en códecs para vídeo en componentes con relación 4:2:2, se consideró ventajoso modificar la presentación con respecto a la que se da en la Recomendación UIT-R BT.500. Se utilizaron imágenes compuestas como referencia adicional para proporcionar un nivel de calidad inferior contra el cual juzgar el comportamiento del códec.

Se recomienda que en la evaluación se utilicen secuencias de al menos seis imágenes, más una adicional destinada a la capacitación antes del comienzo de la prueba. Las secuencias deben variar entre moderadamente críticas y críticas en el contexto de la aplicación de reducción de velocidad binaria que esté en consideración.

Para las evaluaciones subjetivas se puede seleccionar las imágenes y las secuencias de movimiento entre las enumeradas en la Recomendación UIT-R BT.1210, Anexo 1. A este respecto, cabe señalar que las imágenes y secuencias almacenadas digitalmente son las fuentes preferidas para las evaluaciones, puesto que son las señales de origen más reproducibles. Los materiales descritos en la Recomendación UIT-R BT.1210 pueden intercambiarse entre laboratorios para que las comparaciones de sistemas sean más significativas. El formato de cinta D-1 4:2:2 descrito en la Recomendación UIT-R BT.657 proporcionará una base para ese intercambio cuando se disponga de esas máquinas ampliamente y a costo reducido. También es posible efectuar el intercambio mediante formatos de cinta de computador.

A lo largo de este Anexo se hace énfasis en la importancia de comprobar los códecs digitales con secuencias de imágenes que sean críticas en el contexto de la reducción de la velocidad binaria en televisión. Parece por ello razonable preguntarse en qué medida es crítica una secuencia de imágenes particular para un objetivo determinado de reducción de la velocidad binaria, o si una secuencia es más crítica que otra. Una respuesta sencilla, aunque no especialmente útil,

sería decir que “criticidad” significa cosas muy distintas para diferentes códecs. Por ejemplo, podría ocurrir que una imagen fija que contuviera muchos detalles resultase crítica para un códec intratrama mientras que para un códec intercuadro, que es capaz de aprovechar similitudes de cuadro a cuadro, esa misma escena no representaría ninguna dificultad. Algunas secuencias que emplean textura móvil y movimiento complejo resultan críticas para toda clase de códecs, por lo que estos tipos de secuencias son los que más interesa generar o identificar. El movimiento complejo puede tomar la forma de movimientos que son predecibles para un observador pero no para los algoritmos de codificación, como por ejemplo un movimiento periódico tortuoso.

Un examen de posibles medidas estadísticas de criticidad de imagen, por ejemplo mediante métodos correlativos, métodos espectrales, métodos de entropía condicional, etc., ha puesto de manifiesto una medida sencilla pero útil basada en una medición de entropía autoadaptable intratrama/intercuadro. Este método se empleó en la “calibración” de secuencias de imágenes propuestas para utilización en las pruebas de códecs para 34, 45 y 140 Mbit/s en el UIT-R y demostró su utilidad para la selección de secuencias empleadas. La manera más sencilla de efectuar tales mediciones en secuencias de imágenes consiste en transferirlas a computadores de procesamiento de imágenes y someterlas a análisis por soporte lógico.

A continuación se dan algunas directrices de carácter general sobre cómo elegir material crítico, para el caso en que no se pueda acceder a las técnicas anteriores.

a) *Códecs intracampo de longitud de palabra fija*

Aunque es posible y válido evaluar estos códecs con imágenes fijas, se recomienda el empleo de secuencias móviles puesto que con ellas resulta más fácil observar los tratamientos del ruido de codificación y son más representativas de las aplicaciones de televisión. Si se emplean imágenes fijas en simulaciones de códecs por computador, se debe efectuar el tratamiento en toda la secuencia de evaluación, para preservar aspectos temporales de cualquier ruido de origen, por ejemplo. Las escenas elegidas deben contener el mayor número posible de los siguientes detalles: zonas estáticas con ciertas texturas y en movimiento (algunas con textura coloreada), objetos estáticos y en movimiento con bordes nítidos de alto contraste de diversas orientaciones (algunos de color); zonas estáticas uniformes semigrises. Del conjunto de secuencias, al menos una debe presentar ruido de origen justamente perceptible y por lo menos una debe ser sintética (es decir, generada por computador) de modo que esté libre de imperfecciones de cámara, tales como la abertura de exploración y persistencia de imagen.

b) *Códecs intercuadro de longitud de palabra fija*

Todas las escenas de prueba elegidas deben contener movimiento y el mayor número posible de los siguientes detalles: zonas con ciertas texturas y en movimiento (algunas coloreadas), objetos con bordes nítidos de alto contraste moviéndose en dirección perpendicular a esos bordes y con diversas orientaciones (algunos coloreados). Del conjunto de secuencias, al menos una debe tener ruido de origen justamente perceptible y por lo menos una debe ser sintética.

c) *Códecs intracampo de longitud de palabra variable*

Se recomienda que estos códecs se prueben con material de secuencias de imágenes en movimiento, por las mismas razones que los códecs de longitud de palabra fija. Hay que tener en cuenta que debido a su codificación de longitud de palabra variable y su memoria intermedia asociada, estos códecs pueden distribuir dinámicamente la capacidad de bits de codificación a través de la imagen. Así por ejemplo, si en la mitad de una imagen se presenta un cielo sin rasgos especiales que no necesita muchos bits para su codificación, se ahorra capacidad para otras partes de la imagen que pueden así reproducirse con calidad elevada, incluso si son críticas. La conclusión importante de todo esto es que si una secuencia de imágenes resulta crítica para un códec de este tipo, habrá que detallar el contenido de cada parte de la pantalla. Debe llenarse con textura en movimiento y estática, con tanta variación de color como se pueda y objetos con bordes nítidos de alto contraste. Al menos una secuencia del conjunto de prueba debe presentar ruido de origen justamente perceptible y por lo menos una debe ser sintética.

d) *Códecs intercuadro de longitud de palabra variable*

Este es el tipo de códec más complejo y el que necesita el material más exigente para forzarlo. No sólo hay que llenar cada parte de la escena con detalles, como en el caso del códec intracampo de longitud de palabra variable, sino que esos detalles deben, además, estar en movimiento. Por otra parte, puesto que muchos códecs emplean métodos de compensación de movimiento, el movimiento a través de la secuencia debe ser complejo. Ejemplos de movimiento complejo son: escenas que emplean simultáneamente el “zoom” y la panorámica, una escena que tenga como fondo una cortina agitada por el viento y en la que se aprecien sus detalles o su textura; una escena que contenga objetos que giran en un entorno tridimensional; escenas con objetos detallados que se aceleren a través de la pantalla. En todas las escenas debe abundar el movimiento de objetos con diferentes velocidades, texturas y bordes de alto contraste, así como un contenido de color variado. Por lo menos una secuencia del conjunto de prueba debe tener ruido de origen justamente perceptible, al menos una debe tener movimiento complejo de cámara generado por computador a partir de una imagen fija natural (de modo que esté libre de ruido y persistencia de imagen de la cámara), y una secuencia cuando menos debe ser generada completamente por computador.

3.2 Evaluación de la calidad de la imagen después de la posproducción

Con esta evaluación se pretende facilitar la realización de apreciaciones sobre la idoneidad de un códec para aplicaciones de contribución con respecto a una determinada posproducción, por ejemplo la incrustación cromática, la cámara lenta o el “zoom” electrónico. La disposición de equipo mínima necesaria para tal evaluación consiste en un paso único a través del códec sometido a prueba, seguido del tratamiento posterior objeto de interés y a continuación, el observador. Sin embargo, puede ser más representativo de una aplicación de contribución el empleo de códecs adicionales después de la posproducción.

La metodología de la prueba que debe utilizarse es la variante II del método de doble estímulo con escala de calidad continua. Sin embargo, aquí la condición de referencia es la fuente sometida a la misma posproducción que las imágenes decodificadas. Si se considera ventajoso incluir una referencia de calidad inferior también ella deberá someterse a la misma posproducción.

Las secuencias de prueba necesarias para las evaluaciones de tratamiento posterior están sujetas exactamente a los mismos criterios de criticidad que las secuencias destinadas a otras aplicaciones digitales. Sin embargo, es posible que resulte difícil cumplir con esos criterios en el caso de secuencias de primeros planos de incrustación cromática porque normalmente tienen una proporción importante de fondo azul sin rasgos característicos.

Debido a las limitaciones de las posibilidades prácticas de tener que evaluar un códec con varios posproducciones, el número de secuencias de imágenes de prueba utilizadas puede ser como mínimo de tres, y una más disponible a efectos de demostración, por la imposición de tipo práctico de tener que evaluar probablemente un códec con varias posproducciones. La naturaleza de las secuencias dependerá de la tarea de posproducción que se estudie, pero debe variar entre moderadamente crítica y crítica en el contexto de reducción de la velocidad binaria de televisión y para el proceso que se considere. Para la evaluación de la cámara lenta puede servir una velocidad de visualización que sea la décima parte de la de origen.

3.3 Evaluación de la característica de fallo

En la evaluación subjetiva de las degradaciones de las imágenes de códec debidas a imperfecciones del canal de transmisión o emisión, debe elegirse al menos cinco, pero preferiblemente más, proporciones de bits erróneos o condiciones de transmisión/emisión seleccionadas, con separación aproximadamente logarítmica y que abarquen la gama que provoca las degradaciones de códec desde “imperceptible” a “muy molesta”.

Es posible que haga falta evaluar códecs con proporciones de bits erróneos de transmisión que provoquen transitorias visibles tan infrecuentes que no quepa esperar que se produzcan durante un periodo de secuencias de prueba de 10 s. El tiempo de presentación que aquí se sugiere es claramente inadecuado para tales pruebas.

Si es preciso grabar la salida de un códec en condiciones de proporción de bits erróneos bastante baja (lo que da lugar a un número pequeño de transitorios visibles en un periodo de 10 s) para montaje posterior en presentaciones de evaluación subjetiva, se debe tener la precaución de asegurarse de que la grabación utilizada es típica de la salida del códec observada en un intervalo de tiempo mayor.

Limitaciones de tipo práctico inducen a pensar que probablemente serán adecuadas tres secuencias de imágenes de prueba más una de demostración, puesto que hace falta explorar el comportamiento del códec con diversas proporciones de bits erróneos de transmisión. Las secuencias deben tener una duración del orden de 10 s, pero debe señalarse que los evaluadores pueden preferir una duración de 15 a 30 s. Estas deben variar entre moderadamente críticas y críticas en el contexto de reducción de la velocidad binaria de televisión.

Puesto que las pruebas abarcan la gama completa de degradaciones, el método de escala de degradación con doble estímulo es el apropiado y el que debe utilizarse.

3.4 Características de fallo de la imagen según su contenido

En el Apéndice 1 al Anexo 1 de la Recomendación UIT-R BT.500 se describe el concepto general de características de fallo de la imagen según su contenido. Para aplicar este concepto al sistema de televisión digital, se debe utilizar el siguiente procedimiento.

3.4.1 Definición de la criticidad

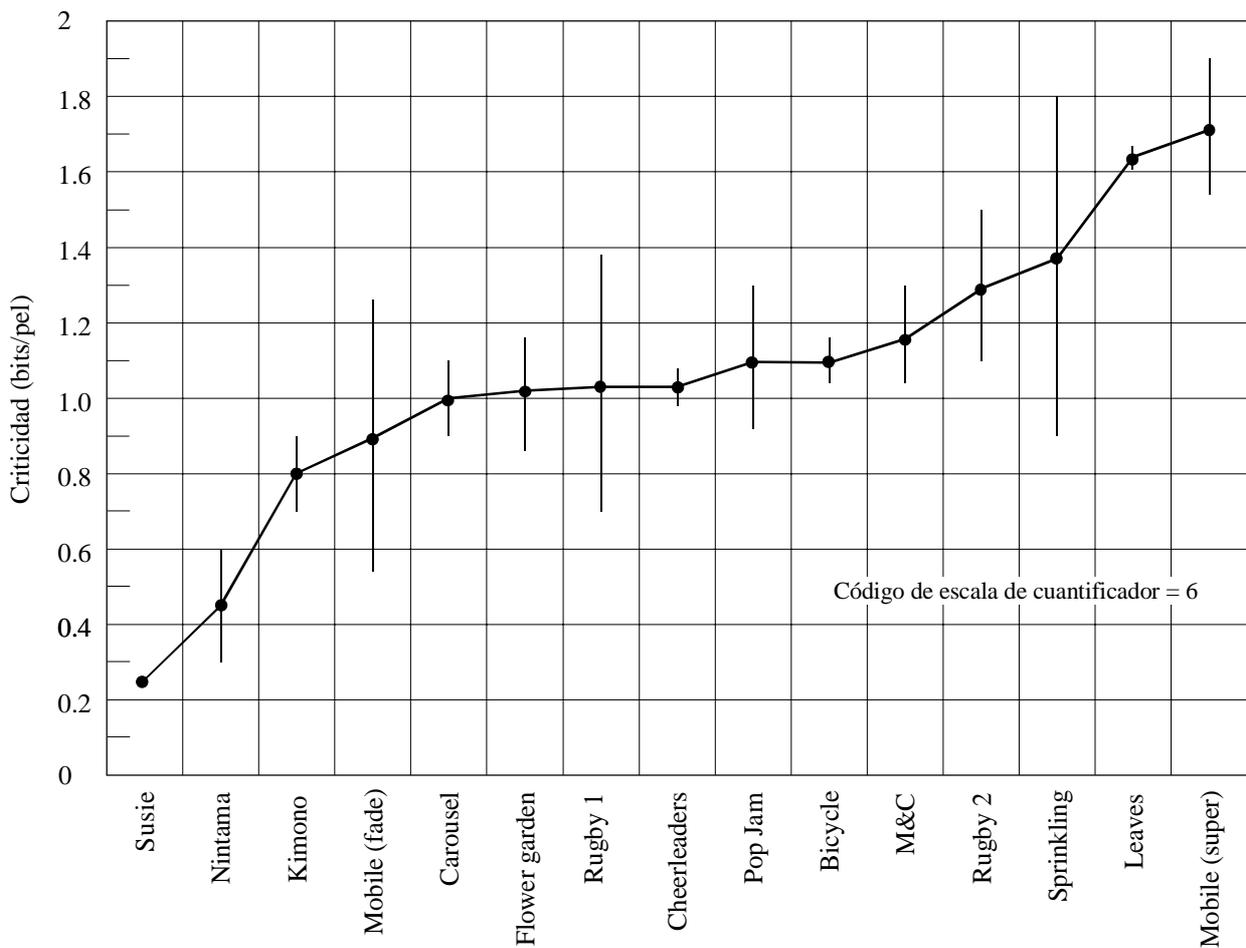
Conviene definir una medida determinada, llamada “criticidad”, que representa las características del sistema de televisión digital en prueba y que se mide por un método objetivo. Como ejemplo de sistema de televisión digital se utiliza MPEG-2 MP@ML y la criticidad aplicada se basa en el método de entropía de cuantificador fijo descrito en la Recomendación UIT-T BT.1210.

3.4.2 Procedimiento para obtener las características de fallo de la imagen según su contenido

- *Fase 1* : Medición de la criticidad de las secuencias de prueba utilizadas en la evaluación subjetiva.

Se mide la criticidad de las secuencias de prueba utilizadas para la evaluación subjetiva descrita en la fase 3 siguiente. En la Figura 1 se puede ver la desviación media y típica de cada secuencia en el sistema utilizado como ejemplo. La mayoría de las secuencias tienen una medida de criticidad de 0,8 a 1,4 bits/píxel. Algunas secuencias tienen una desviación típica grande porque el contenido de imagen varía considerablemente durante la secuencia.

FIGURA 1
Desviaciones media y típica de la criticidad de las secuencias de prueba



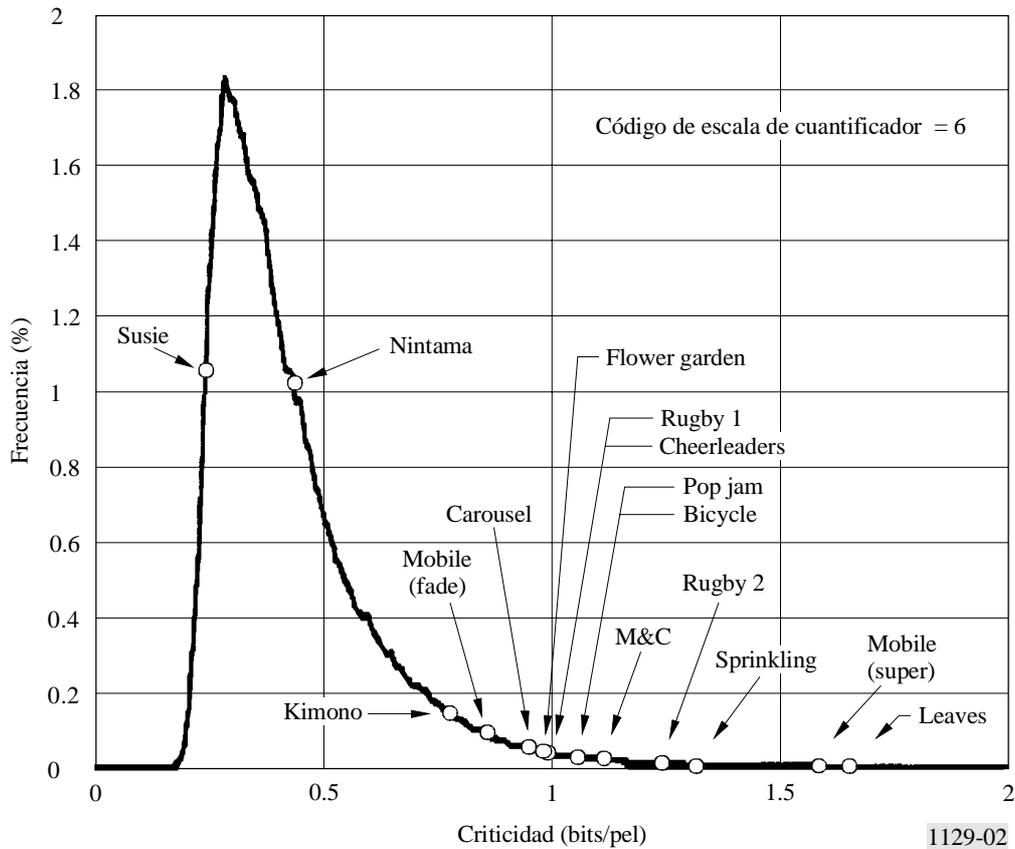
1129-01

- *Fase 2* : Medición de la distribución de la criticidad de los programas de radiodifusión durante un largo periodo de tiempo.

La distribución de la criticidad de los programas de televisión transmitidos se mide durante un periodo de tiempo suficientemente largo, por ejemplo, una semana. En la Figura 2 se presenta un ejemplo de la distribución medida durante una semana, 130 en total, de señales de radiodifusión NTSC, convertidas para la medición en las señales Y/C componentes. La frecuencia de aparición de criticidad en los programas de televisión se calculó cada 5×10^{-3} bits/píxel. En esta figura se presenta también la criticidad correspondiente a las secuencias de prueba utilizadas para la evaluación subjetiva.

FIGURA 2

Desviaciones de la criticidad de los programas transmitidos y criticidad de las secuencias de prueba



- Fase 3 : Evaluación subjetiva de la calidad de imagen del sistema probado y obtención de la relación entre la criticidad y la calidad de imagen subjetiva.

La calidad de imagen del sistema de televisión digital se evalúa por el método de escala de calidad continua de doble estímulo (double stimulus continuous quality scale, DSCQS). La relación entre la criticidad y las notas obtenidas en la evaluación se deriva combinando el resultado de la evaluación subjetiva y la criticidad obtenida en la fase 1. En la Figura 3 se presenta la calidad de imagen del sistema utilizado como ejemplo a las velocidades binarias de 4, 6, 9 y 15 Mbit/s. La diferencia de calidad (% de DSCQS) en la figura representa la degradación en relación con la secuencia de referencia original cuyos componentes tienen una relación 4:2:2. En la Figura 4 se puede ver la relación entre la criticidad y la diferencia de calidad. En este ejemplo, se dio por supuesto una relación lineal entre criticidad y calidad de imagen y se derivaron las rectas de regresión utilizando el método de los mínimos cuadrados. En la figura se ilustra la recta de regresión a cada velocidad binaria. En general es posible aplicar una relación no lineal según los resultados de la evaluación.

- Fase 4 : Establecer las características de fallo de la imagen según su contenido (calidad en función de la frecuencia de aparición) combinando los resultados de la fase 3 (criticidad en función de la calidad) y de la fase 2 (criticidad en función de la frecuencia de aparición)

Combinando los resultados obtenidos en las fases 2 y 3 se obtienen las características de fallo de la imagen según su contenido, es decir, la distribución de la calidad de imagen de programas de televisión con codificación digital. La degradación de imagen en los programas de televisión transmitidos se convierte en frecuencia acumulativa de aparición. En la Figura 5 se presentan las características de fallo de la imagen según su contenido del sistema utilizado como ejemplo.

FIGURA 3
 Resultado de la evaluación subjetiva (MP@ML en 6H)

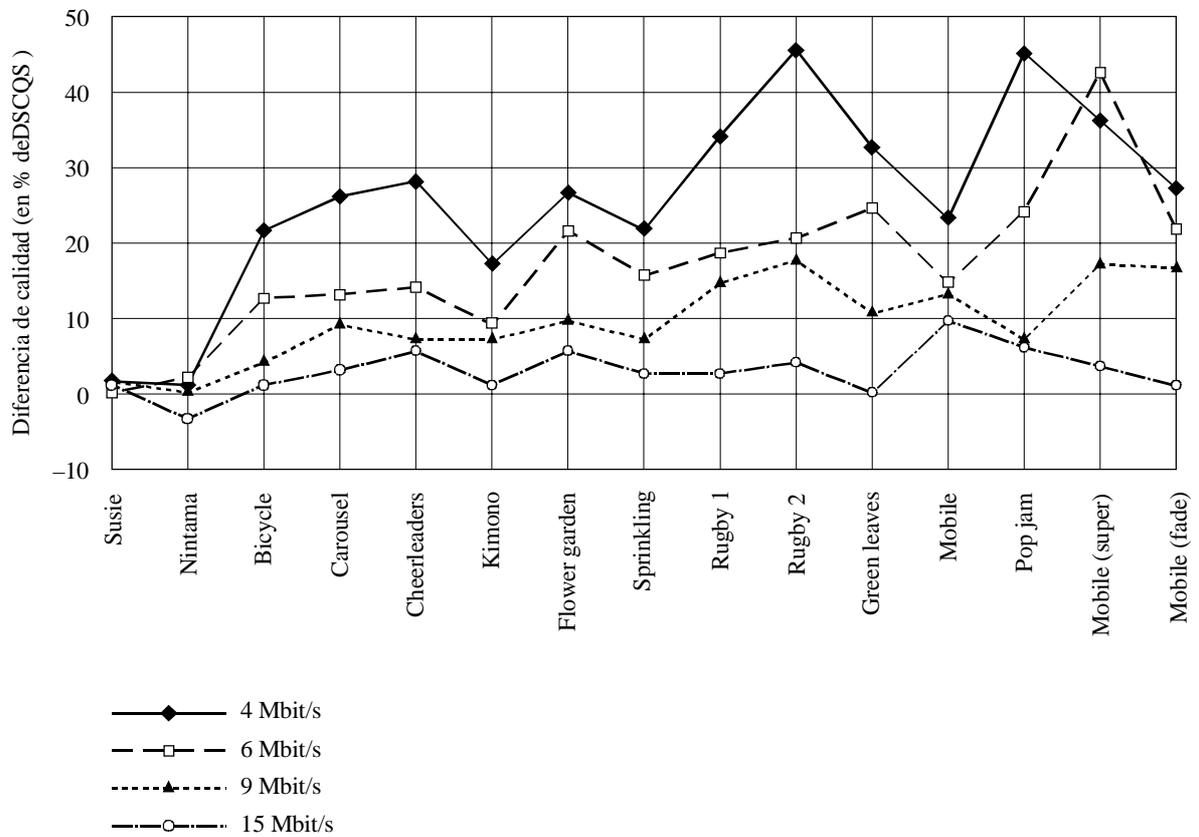
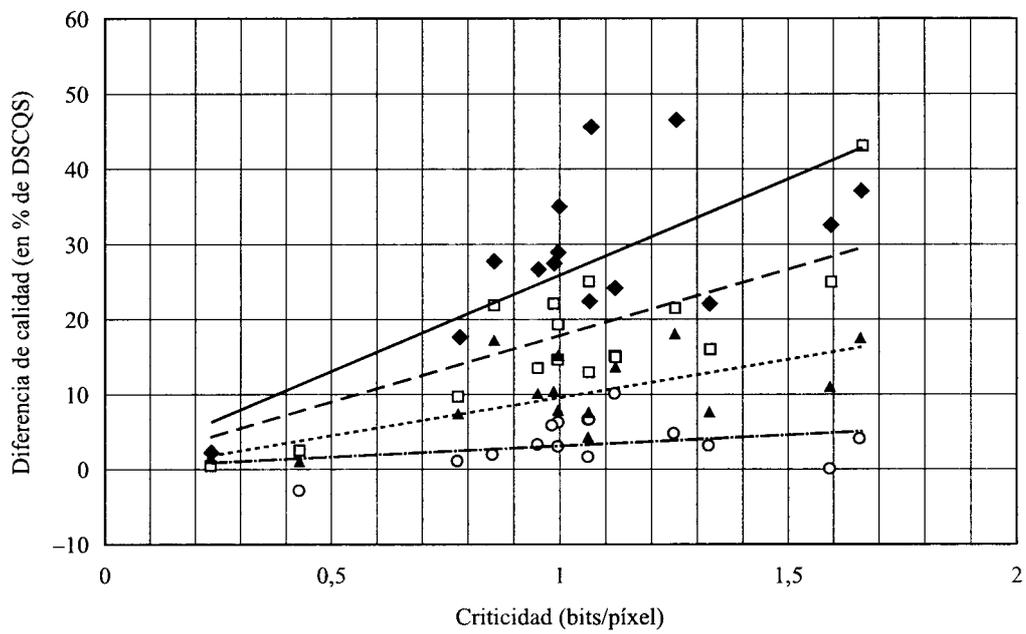


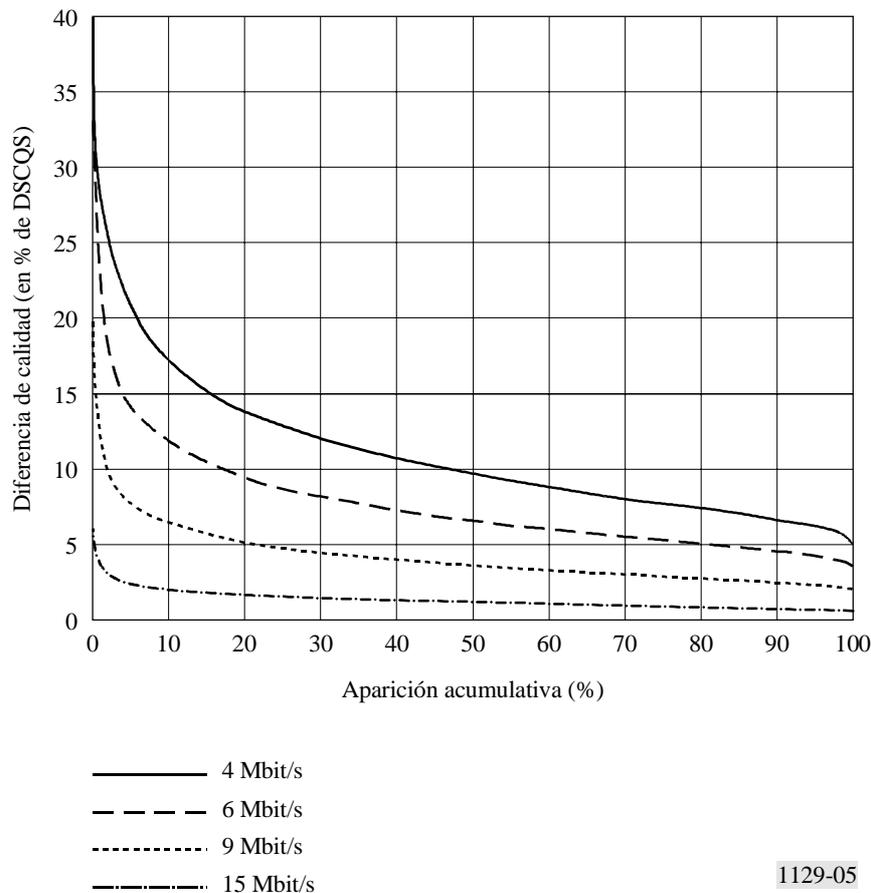
FIGURA 4
 Relación entre la criticidad y la nota de evaluación (MP@ML at 6H)



- ◆ 4 Mbit/s
- 6 Mbit/s
- ▲ 9 Mbit/s
- 15 Mbit/s

FIGURA 5

Frecuencia acumulativa de aparición de degradación de imagen
(MP@ML en 6 H)



1129-05

4 Observaciones relativas a la aplicación

Cuando no hace falta una apreciación de la calidad o la degradación absolutas de un códec sino sólo su orden de jerarquía, o cuando se desea la confirmación del orden de jerarquía obtenido a partir de los resultados del método de doble estímulo, se debe utilizar el método de las comparaciones de pares de estímulos.

Tal como se describe en la Recomendación UIT-R BT.500, el método proporciona una comparación sensible y una manera de medir la relación entre pares de sistemas. Es posible una extensión del método, para jerarquizar las calidades o las degradaciones de más de dos sistemas. En este enfoque, el orden de jerarquía global se deriva de la jerarquización de todos los pares posibles de secuencias de imágenes efectuada por los observadores.

El análisis se complica por el hecho de que un observador puede, por ejemplo, clasificar a la imagen A como mejor que la imagen B, y a la imagen B mejor que la C, pero también a la C mejor que la imagen A. Es lo que se denomina una "triada intransitiva".

El número de presentaciones necesarias aumenta con el cuadrado del número de secuencias de imágenes de prueba y de códecs, lo cual representa una desventaja de este método que puede llegar a hacerlo impracticable.

Si el canal de radiodifusión se utiliza para entregar trenes de bits de programas múltiples o métodos de codificación jerárquica o ajustable por escalón, quizás sea necesario adaptar la metodología de evaluación para tener en cuenta lo siguiente:

- puede que el criterio de servicio aceptable no sea la transparencia de la codificación en la fuente; en vez, puede ser la capacidad del sistema de suministrar, con una velocidad binaria determinada, una alternativa viable para el servicio convencional. Por consiguiente, como sucede con la referencia en las pruebas de calidad, quizás sea conveniente utilizar el material que puede entregar un sistema convencional en condiciones de recepción típicas, en vez de entregarlo en forma digital sin compresión. Además, quizás sea conveniente utilizar material de prueba

seleccionado para representar la gama de programas actuales y futuros (véase el Apéndice 1 del Anexo 1 a la Recomendación UIT-R BT.500). En las pruebas, las condiciones de observación serán las indicadas en la Recomendación UIT-R BT.500 y en el § 1 de este Anexo, y el método general de prueba será el de doble estímulo con escala de calidad continua (Recomendación UIT-R BT.500, § 5); y,

- es importante tener en cuenta la capacidad del sistema de mantener la integridad de cada tren de bits de programa en condiciones de carga total de canal y las degradaciones de la transmisión. Por consiguiente, en las pruebas de degradación, quizás resulte conveniente asegurar la carga total del canal y utilizar una gama de niveles de degradación seleccionada de manera que represente la gama de condiciones de recepción posibles (véase el Apéndice 2 del Anexo 1 a la Recomendación UIT-R BT.500). En las pruebas, las condiciones de observación serán las indicadas en la Recomendación UIT-R BT.500 y en el § 1 de este Anexo, y el método de prueba general será el de doble estímulo con escala de degradación (Recomendación UIT-R BT.500, § 4).

NOTA 1 – Cuando se evalúan sistemas analógicos y digitales en el mismo contexto, es importante elegir un conjunto de materiales de prueba que refleje una dificultad equilibrada para los sistemas analógicos y digitales. En este caso y para profundizar el análisis, quizás resulte conveniente aplicar el método con escala multidimensional al que se hace referencia en el Cuadro 2 del Anexo 1 a la Recomendación UIT-R BT.500.
