

## RECOMENDACIÓN UIT-R BT.1663

**Métodos de observación especializada para evaluar la calidad de los sistemas destinados a la proyección digital de imágenes digitales en pantalla grande<sup>1</sup> en cines**

(Cuestión UIT-R 15/6)

(2003)

La Asamblea de Radiocomunicaciones de la UIT,

*considerando*

- a) que la UIT, al igual que otros organismos internacionales de normalización, está estudiando un nuevo servicio denominado imágenes digitales en pantalla grande (LSDI, *large screen digital imagery*);
- b) que probablemente varias aplicaciones se identificarán para el servicio de LSDI;
- c) que será necesario especificar requisitos de calidad de funcionamiento y verificar la conveniencia de soluciones técnicas consideradas para cada aplicación, teniendo en cuenta los requisitos de calidad de funcionamiento de cada una de ellas;
- d) que estas comprobaciones involucrarán necesariamente también pruebas de evaluación subjetiva bajo condiciones científicas rigurosas;
- e) que diversos métodos de prueba subjetiva pueden ofrecer diversos grados de sensibilidad para la detección de determinados tipos de degradación;
- f) que los métodos de evaluación subjetiva especificados en la Recomendación UIT-R BT.500, basados en una observación no especializada, requieren tiempo y son de costos elevados, debido al gran número de observaciones que deben efectuarse para poder diferenciar los sistemas desde el punto de vista de la calidad de funcionamiento;
- g) que se propone ahora un nuevo método de pruebas subjetivas, basado en el empleo de un pequeño número de observadores especializados, lo que proporcionará la capacidad comparable de diferenciar la calidad de funcionamiento de los diferentes sistemas mediante procedimientos más rápidos y menos costosos,

*recomienda*

**1** la utilización del método de prueba descrito en el Anexo 1, basado en el empleo de observadores especializados, para la evaluación subjetiva de las soluciones de LSDI, toda vez que limitaciones de tiempo o de presupuesto no permitan aplicar los métodos de observación no especializada que se especifican en la Recomendación UIT-R BT.500 y cuando la sensibilidad del método descrito en el Anexo 1 sea suficiente para diferenciar los sistemas evaluados,

---

<sup>1</sup> Las imágenes digitales en pantalla grande (LSDI) son una familia de sistemas de imágenes digitales aplicables a programas tales como representaciones dramáticas y comedias, acontecimientos deportivos y culturales, conciertos, etc., para una presentación en pantalla grande con calidad de alta resolución en salas cinematográficas, teatros, salas y otros locales adecuadamente equipados.

*recomienda además*

- 1 efectuar estudios como parte del proceso de prueba, para comprobar la sensibilidad del método de prueba descrito en el Anexo 1;
- 2 realizar otros estudios, a los efectos de mejorar la presente Recomendación, sobre el tratamiento de los resultados;
- 3 alentar a los administradores y a las organizaciones a poner a disposición, para las Administraciones Miembros de la UIT y los Miembros de Sector, todo material e instrumentos de prueba (por ejemplo, programas informáticos para generar presentaciones de imágenes antiguas o simétricas) que se estén elaborando, a fin de facilitar las futuras pruebas por otras organizaciones.

## **Anexo 1**

### **Observación especializada para evaluar la calidad de los sistemas destinados a la proyección digital de películas en LSDI en salas de cine**

#### **1 Introducción**

En los últimos años, la observación especializada se ha empleado frecuentemente para efectuar rápidas comprobaciones de la calidad de funcionamiento de un proceso de vídeo genérico.

El presente Anexo describe un método de prueba por observación especializada que permitirá una coherencia de resultados obtenidos en diferentes laboratorios al recurrir a los servicios de un número limitado de observadores especializados.

#### **2 Motivos del nuevo método basado en la «observación especializada»**

Conviene poner de relieve las ventajas que resultan de la aplicación de la metodología propuesta.

En primer lugar, una prueba de evaluación subjetiva formal suele requerir el empleo de por lo menos 15 observadores elegidos entre los «no especializados», la realización de pruebas prolongadas y la búsqueda continua de nuevos observadores. Tal número de observadores es necesario para lograr la sensibilidad necesaria de modo que los sistemas sometidos a prueba puedan diferenciarse y clasificarse, o juzgarse equivalentes de manera fiable.

En segundo lugar, al recurrir a observadores no especializados, las formas tradicionales de prueba pueden no revelar diferencias que podrían resaltar en una exposición más prolongada, incluso para ojos no expertos.

En tercer lugar, las evaluaciones provisionales generalmente establecen medidas de calidad (o diferencias en la calidad), pero no identifican directamente los objetos u otras manifestaciones materiales a las que dichas medidas se refieren.

La metodología aquí propuesta trata de dar solución a estos tres problemas.

### **3 Definición de especialistas**

A los efectos de la presente Recomendación, un «observador especializado» es una persona que conoce el material empleado para la evaluación, sabe «qué mirar», eventualmente puede estar bien informado sobre los detalles del algoritmo utilizado para procesar el material vídeo que ha de evaluarse. En todos los casos, el «observador especializado» es una persona con larga experiencia en el terreno de la investigación de la calidad, alguien profesionalmente vinculado al tema específico de la prueba. Por ejemplo, cuando se organiza una serie de pruebas de «observación especializada» de un material dado de LSDI, deberían elegirse expertos en la producción o postproducción de películas o en la producción de contenido vídeo de alta calidad (por ejemplo, camarógrafos, retocadores de color, etc.); la selección ha de efectuarse teniendo en cuenta la posibilidad de establecer juicios subjetivos únicos sobre la calidad de imagen de LSDI y los artilugios de compresión.

### **4 Selección de evaluadores**

Toda prueba de observación especializada es una serie de evaluaciones basadas en las opiniones de evaluadores y en la que se emiten juicios sobre la calidad visual y/o la degradación aparente.

El grupo básico de expertos está formado por cinco a seis personas. Este número reducido facilita la tarea de elegir los evaluadores y permite llegar a decisiones más rápidas.

En función de las necesidades del experimento, se admite la utilización de más de un grupo básico de expertos, reunidos en un conjunto combinado más amplio (por ejemplo, provenientes de diferentes laboratorios).

Se entiende que los especialistas tienen tendencia a mejorar la clasificación cuando están probando su propia tecnología, por lo que debería evitarse la inclusión de personas directamente involucradas en el desarrollo del sistema sometido a prueba.

Todos los evaluadores deberían pasar por un examen de agudeza visual, normal o corregida (prueba de Snellen), así como de visión cromática normal (prueba de Ishihara).

### **5 Material de prueba**

Los materiales por probar deben seleccionarse de tal manera que sirvan de muestra de toda la gama de valores de producción y niveles de dificultad previstos en el contexto real en el cual los sistemas bajo prueba habrán de utilizarse. La selección debería favorecer los materiales más exigentes, aunque sin excesos. Por lo general, deberían emplearse de 5 a 7 secuencias de prueba.

El método de selección de materiales puede variar también en función de la aplicación para la cual el sistema bajo prueba se ha diseñado.

En tal sentido, no se dan más indicaciones aquí sobre las reglas de selección del material de prueba, dejando la decisión al autor de la concepción de la prueba, sobre la base de las consideraciones antes mencionadas.

### **6 Condiciones de observación**

Las condiciones de observación, que deberán describirse exhaustivamente en el informe sobre la prueba, han de ajustarse al Cuadro 1 y mantenerse de manera constante durante la prueba.

CUADRO 1

Condiciones de observación	Valores	
	Mínimo	Máximo
Dimensión de la pantalla (m)	6	16
Distancia de observación <sup>(1)</sup>	1,5 H	2 H
Luminancia del proyector (pantalla central, blanco máximo)	10 ftL	14 ftL
Luminancia de la pantalla (fuera del proyector)		<1/1 000 de la luminancia del proyector

<sup>(1)</sup> Ha de emplearse la presentación en «mariposa» cuando la distancia de observación sea menor que 1,5 H. En caso de emplearse una presentación «por yuxtaposición», la distancia de observación debería acercarse más al valor de 2 H.

## 7 Metodología

### 7.1 Series de evaluación

Cada serie de evaluación (definida como el conjunto de sesiones de prueba de un grupo determinado de observadores) debería comprender dos fases (por ejemplo, Fase I y Fase II).

#### 7.1.1 Fase I

La Fase I consiste en una prueba subjetiva formal realizada en un entorno controlado (véase el § 6), gracias a la cual se obtendrán resultados válidos, sensibles y repetibles. En esta fase, los especialistas clasifican cada uno el material presentado, empleando la escala de clasificación que se describe más adelante. Los miembros del equipo no deben debatir entre sí lo que estén viendo ni controlar las presentaciones. En el curso de esta fase, los especialistas NO deberían conocer el sistema de codificación sometido a prueba ni el orden de presentación del material que se prueba. El material bajo prueba debería presentarse de manera aleatoria, para evitar toda parcialidad en el juicio.

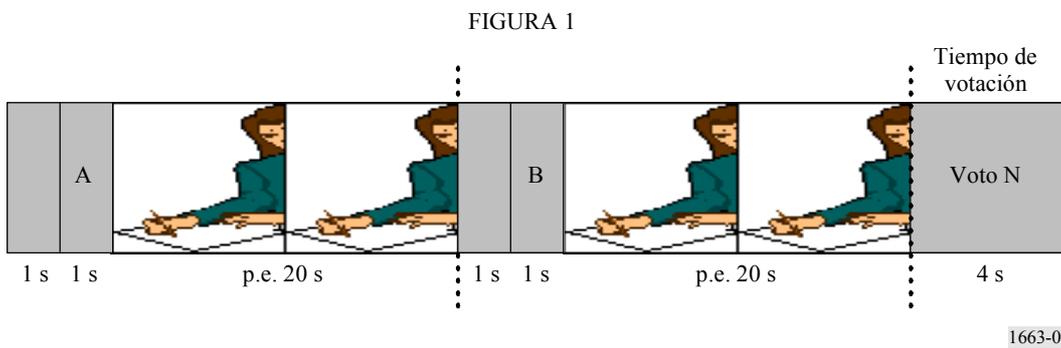
##### 7.1.1.1 Presentación del material

El método de presentación combina elementos de doble estímulo simultáneo para evaluación continua (SDSCE) (véase la Recomendación UIT-R BT.500, § 6.4) y el método de escala de calidad continua de doble estímulo (DSCQS) (véase la Recomendación UIT-R BT.500, § 5). Es posible referirse a él como método de doble estímulo simultáneo.

Con arreglo al método SDSCE, en cada prueba se presentará en una pantalla dividida el material correspondiente a dos imágenes. En la mayoría de los casos, una de las imágenes de origen será la referencia (es decir, la imagen fuente) y la otra la imagen de prueba; en otros casos, las dos imágenes se habrán extraído de la imagen de referencia. La referencia estará constituida por el material de origen presentado de manera transparente (es decir, sin someterlo a otra compresión que la correspondiente al medio de grabación de la fuente). El material de prueba será el material de fuente procesado a través de uno de los sistemas bajo prueba. La velocidad binaria y/o el nivel de calidad corresponderán a las especificaciones del guión de la prueba. A diferencia del método SDSCE, los observadores no conocerán las condiciones representadas por ambos miembros del par de imágenes.

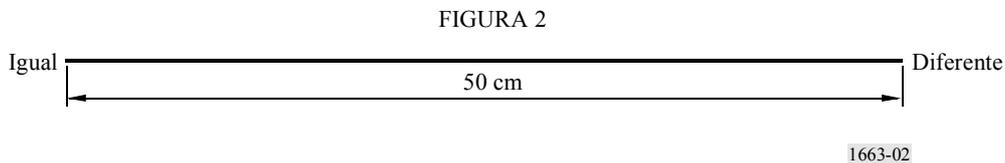
La presentación en pantalla dividida podrá efectuarse mediante el método tradicional de división de la pantalla sin efecto de simetría, o bien mediante la técnica «en mariposa», en la que la imagen de la derecha de la pantalla es su equivalente especular. Puesto que se presentarán imágenes completas a lo ancho, sólo la mitad de cada una podrá observarse a la vez. En cada presentación, en cada lado de la pantalla se presentará la misma mitad de la imagen.

Con arreglo al método DSCQS, el par de imágenes se presenta dos veces sucesivamente, una vez para familiarizarse con la imagen y evaluarla, y otra vez para confirmar la impresión y clasificar. Cada secuencia durará entre 15 y 30 s. Es posible rotular cada secuencia al comienzo de cada fragmento, para asistir a los evaluadores (véase la Fig. 1 como ejemplo de división de pantalla sin efecto especular).



**7.1.1.2 Escala de evaluación**

El criterio de aceptación en las aplicaciones de LSDI es que la imagen de prueba (es decir, comprimida) no pueda distinguirse de la referencia. Para evaluar los sistemas sometidos a prueba pueden emplearse varios métodos corrientes de clasificación. Uno de los métodos que se propone consiste en las escalas de comparación de estímulo, sugerido en la Recomendación UIT-R BT.500, § 6.2. Un ejemplo concreto de escala es la de IGUAL-DIFERENTE, de tipo no categórico (continuo) que se describe en el § 6.2.4.2 de la Recomendación UIT-R BT.500:



**7.1.1.3 Sesiones de evaluación**

El proceso de evaluación puede implicar más de una sesión, en función del número de condiciones de prueba, y deberá comprender dos tipos de pruebas: pruebas iniciales y pruebas de confirmación. En una prueba inicial, una mitad de la pantalla muestra la referencia, mientras que la otra mitad muestra la prueba. En una prueba de confirmación ambas mitades muestran la referencia. La prueba de confirmación tiene por objeto medir la posibilidad de juicio parcial.

Para cada sistema probado, se requieren en cada secuencia de prueba las siguientes pruebas iniciales:

CUADRO 2

<b>Flanco izquierdo de la pantalla</b>	<b>Flanco derecho de la pantalla</b>
Referencia de la mitad de la izquierda	Prueba de la mitad de la izquierda
Referencia de la mitad de la derecha	Prueba de la mitad de la derecha
Prueba de la mitad de la izquierda	Referencia de la mitad de la izquierda
Prueba de la mitad de la derecha	Referencia de la mitad de la derecha

De preferencia, cada uno de los casos descritos debería repetirse por lo menos dos veces. Para cada sistema se requiere, en cada secuencia de prueba, las siguientes pruebas de confirmación:

CUADRO 3

<b>Flanco izquierdo</b>	<b>Flanco derecho</b>
Referencia de la mitad de la izquierda	Referencia de la mitad de la izquierda
Referencia de la mitad de la derecha	Referencia de la mitad de la derecha

También aquí es preferible efectuar, para cada uno de los casos descritos, por los menos dos repeticiones.

Las pruebas deberían consistir en sesiones de más de una hora de duración, con pausas de descanso de 15 min. Las pruebas iniciales y de confirmación resultantes de la combinación de secuencias de códec y de prueba deberían dividirse en sesiones según un criterio de asignación pseudoaleatorio. Aunque sea más complicado, vale la pena imponer algunas restricciones a este proceso. Por ejemplo, en caso de efectuarse cuatro sesiones, podría asignarse aleatoriamente cada una de las pruebas iniciales a un códec dado y las secuencias de pruebas a una posición determinada aleatoriamente en una de estas sesiones. Este método tiene la ventaja de asegurar una distribución de las pruebas iniciales del sistema a lo largo de toda la serie de pruebas.

#### 7.1.1.4 Procesamiento de los resultados de las pruebas

Para una prueba inicial dada, el resultado de la prueba es la distancia entre el extremo «IGUAL» de la escala y la marca establecida por el observador, expresada en valores de 0 a 100 de la escala. Los resultados se analizarán como nota media de opinión (MOS), método que se empleará para establecer una clasificación de los sistemas probados. Según el número de observaciones efectuadas por sistema (observadores  $\times$  secuencias de prueba  $\times$  repeticiones), los datos pueden someterse a un análisis de varianza (ANOVA)<sup>2</sup>. El resultado de pruebas de confirmación puede utilizarse para derivar una diferencia de evaluación básica de «azar».

---

<sup>2</sup> Un total de 10-20 observaciones en la condición de orden inferior de interés es suficiente para aplicar tratamientos estadísticos inferenciales del tipo ANOVA.

### 7.1.2 Fase II

Uno de los principales objetivos de la Fase II es ajustar el orden relativo de los resultados de la Fase I, cuya precisión y fiabilidad puede reducirse debido al número limitado de observadores y/o de evaluaciones efectuadas. Otro objetivo importante es hacer variar las observaciones en cuanto a las características de percepción de las imágenes en las que se hubieran basado las evaluaciones de la Fase I.

En esta parte, el grupo de expertos analiza el material presentado. Aquí están autorizados a comentar el material presentado, repetir su presentación, total o parcialmente, todas las veces que sea necesario para su análisis y/o demostración, y a elaborar por consenso una evaluación y descripción de lo que hayan visto. Si los observadores especializados lo desean, puede hacerse una «presentación trucada», empleando por ejemplo modalidades tales como movimiento lento, imágenes consecutivas y cuadros fijos. Estas técnicas requerirán cierta interacción y la intervención del director de la prueba.

#### 7.1.2.1 Agrupamiento del material sometido a prueba

Para llevar a cabo la prueba de la Fase II adecuadamente, es necesario agrupar el material sometido a prueba según su contenido, en lo que se denomina conjunto básico de observación para expertos (BES), o sea, todas las secuencias codificadas obtenidas de la misma secuencia fuente deben agruparse y ordenarse, de conformidad con los resultados de la Fase I.

Los materiales de prueba se ordenarán desde los valores inferiores de MOS hasta los valores más altos. El número de BES corresponderá al número de secuencias utilizadas en la prueba.

#### 7.1.2.2 Subsesión de prueba de observación especializada básica

La subsesión de prueba de expertos de observación básica (BEV) es una sesión de debate, en la que los expertos examinan todo el material incluido en un BES; una tarea consiste en confirmar o modificar el orden de clasificación resultante de la prueba formal de la Fase I. Por consiguiente, ha de confirmarse o modificarse la relativa visibilidad de las diferencias.

#### 7.1.2.3 Plan de la Fase II

En la Fase II, deben llevarse a cabo todas las BEV. Los expertos sabrán que el orden de presentación es el resultado de la clasificación de la Fase I. Los expertos no tendrán conocimiento de las relaciones entre los sistemas propuestos y su clasificación.

La Fase II se llevará a cabo como un esfuerzo colectivo que dé por resultado opiniones de consenso entre los evaluadores.

Antes de iniciarse la Fase II, se encargará a los evaluadores, eventualmente mediante un texto escrito, a ejecutar las tareas siguientes:

- observar el material en cada BEV;
- debatir la clasificación del material en cada BEV; en caso de desacuerdo del grupo, definir un nuevo orden de clasificación;
- comentar cada caso, con inclusión de observaciones de detalle sobre la característica de las diferencias observadas, en caso de existir;
- documentar la clasificación, así como sus comentarios y observaciones.

El director de la prueba tendrá a su cargo la responsabilidad de reunir todos los comentarios de los grupos y señalar las discrepancias. Mientras prosiguen las pruebas, los resultados de las Fases I y II de los distintos grupos se mantendrán en secreto, para no influir a los grupos siguientes.

En la medida de lo posible, el director de la prueba está autorizado a identificar las divergencias y apoyar la solución de las mismas mediante pruebas adicionales para los resultados controvertidos. El objetivo de esta última medida es lograr un consenso general.

## **8 Informe**

El Informe final de la prueba estará a cargo del director de la misma.

Dicho Informe comprenderá la siguiente información:

- resultados de la Fase I (con inclusión de cuadros de MOS, así como los resultados de los análisis estadísticos que corresponda);
  - comentarios de los expertos durante la Fase II;
  - comentarios sobre toda reclasificación;
  - cualquier información pertinente sobre condiciones de observación, características de la señal de entrada, procesamiento de la señal, características del proyector, ajuste del proyector, cromaticidad, selección de observadores y condiciones de la prueba;
  - una caracterización completa de la calidad de funcionamiento del dispositivo de visualización (tiempo medio entre fallos, etc.);
  - resumen y conclusiones.
-