

Unión Internacional de Telecomunicaciones

UIT-R

Sector de Radiocomunicaciones de la UIT

Recomendación UIT-R BT.2095-1
(06/2017)

**Evaluación subjetiva de la calidad
de vídeo aplicando el Protocolo
de observación para expertos**

Serie BT
Servicio de radiodifusión
(televisión)



Unión
Internacional de
Telecomunicaciones

Prólogo

El Sector de Radiocomunicaciones tiene como cometido garantizar la utilización racional, equitativa, eficaz y económica del espectro de frecuencias radioeléctricas por todos los servicios de radiocomunicaciones, incluidos los servicios por satélite, y realizar, sin limitación de gamas de frecuencias, estudios que sirvan de base para la adopción de las Recomendaciones UIT-R.

Las Conferencias Mundiales y Regionales de Radiocomunicaciones y las Asambleas de Radiocomunicaciones, con la colaboración de las Comisiones de Estudio, cumplen las funciones reglamentarias y políticas del Sector de Radiocomunicaciones.

Política sobre Derechos de Propiedad Intelectual (IPR)

La política del UIT-R sobre Derechos de Propiedad Intelectual se describe en la Política Común de Patentes UIT-T/UIT-R/ISO/CEI a la que se hace referencia en el Anexo 1 a la Resolución UIT-R 1. Los formularios que deben utilizarse en la declaración sobre patentes y utilización de patentes por los titulares de las mismas figuran en la dirección web <http://www.itu.int/ITU-R/go/patents/es>, donde también aparecen las Directrices para la implementación de la Política Común de Patentes UIT-T/UIT-R/ISO/CEI y la base de datos sobre información de patentes del UIT-R sobre este asunto.

Series de las Recomendaciones UIT-R

(También disponible en línea en <http://www.itu.int/publ/R-REC/es>)

Series	Título
BO	Distribución por satélite
BR	Registro para producción, archivo y reproducción; películas en televisión
BS	Servicio de radiodifusión (sonora)
BT	Servicio de radiodifusión (televisión)
F	Servicio fijo
M	Servicios móviles, de radiodeterminación, de aficionados y otros servicios por satélite conexos
P	Propagación de las ondas radioeléctricas
RA	Radioastronomía
RS	Sistemas de detección a distancia
S	Servicio fijo por satélite
SA	Aplicaciones espaciales y meteorología
SF	Compartición de frecuencias y coordinación entre los sistemas del servicio fijo por satélite y del servicio fijo
SM	Gestión del espectro
SNG	Periodismo electrónico por satélite
TF	Emisiones de frecuencias patrón y señales horarias
V	Vocabulario y cuestiones afines

Nota: Esta Recomendación UIT-R fue aprobada en inglés conforme al procedimiento detallado en la Resolución UIT-R 1.

Publicación electrónica
Ginebra, 2018

© UIT 2018

Reservados todos los derechos. Ninguna parte de esta publicación puede reproducirse por ningún procedimiento sin previa autorización escrita por parte de la UIT.

RECOMENDACIÓN UIT-R BT.2095-1

Evaluación subjetiva de la calidad de vídeo aplicando el Protocolo de observación para expertos

(2016-2017)

Cometido

En este proyecto de nueva Recomendación se describe el método para la evaluación subjetiva de la calidad de vídeo de imágenes en movimiento aplicando el Protocolo de observación para expertos (EVP), con la participación de un reducido número de espectadores, todos ellos seleccionados entre expertos en el área de procesamiento de vídeo de que se trate.

Palabras clave

Calidad de vídeo, evaluación subjetiva, observación por expertos, televisión

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que las tecnologías de codificación de fuente para aplicaciones de televisión digital están mejorando continuamente tanto en eficiencia como en rendimiento visual;
- b) que la permanente evolución de las tecnologías de codificación de vídeo implica una demanda creciente de métodos de evaluación para determinar el rendimiento técnico y visual;
- c) que la eficacia de la compresión y el rendimiento visual de las nuevas tecnologías de codificación de fuente de vídeo requieren nuevos métodos de evaluación y calificación visuales más eficaces;
- d) que los métodos de evaluación especificados en las Recomendaciones UIT-R actuales son muy exigentes en lo que respecta a la inversión en tiempo y en recursos humanos y que a menudo no tienen en cuenta la evolución técnica de las pantallas ni la satisfacción del usuario final;
- e) que recientemente nuevos planteamientos en los protocolos de observación para expertos han mostrado mayor eficacia y mejores resultados en términos de tiempo y de coste global en comparación con los proporcionados por los métodos basados en observadores no expertos;
- f) que, aunque los resultados del Protocolo de observación para expertos no pueden sustituir a los resultados facilitados por un protocolo de evaluación subjetiva formal, se pueden considerar una indicación preliminar estimable del desempeño del sistema sometido a prueba;
- g) que la creciente evolución de la tecnología en materia de pantallas planas ha modificado drásticamente las condiciones de observación utilizadas habitualmente por los expertos;
- h) que ISO/CEI ya ha utilizado con éxito nuevos protocolos basados en la observación por expertos para la evaluación de nuevas tecnologías de codificación de fuente de vídeo,

recomienda

- 1** que, al evaluar las nuevas tecnologías de codificación de vídeo digital, se tenga en cuenta el uso del Protocolo de observación para expertos descrito en el Anexo 1;
- 2** que el Protocolo de observación para expertos se implante utilizando pantallas planas profesionales y el montaje de laboratorio que se describe en el Anexo 1.

NOTA 1 – En el Anexo 2 (informativo) se muestran los resultados de un experimento subjetivo realizado con el protocolo basado en la observación por expertos mencionado en el *considerando h*).

Anexo 1

Protocolo de observación para expertos para la evaluación de la calidad de vídeo

1 Montaje de laboratorio

1.1 Selección y montaje de la pantalla

La pantalla debe de ser una pantalla plana con características para aplicaciones profesionales (por ejemplo, estudios o unidades móviles de radiodifusión); la dimensión diagonal de la pantalla puede variar entre 22 pulgadas (mínimo) y 40 pulgadas (recomendado), aunque puede llegar a 50 pulgadas o más cuando se evalúan sistemas de imagen con una resolución de televisión de alta definición o superior.

Se permite utilizar una porción reducida de la zona activa de visualización en la pantalla, en cuyo caso la zona en torno a la parte activa de la pantalla se debe ajustar a gris medio. En estas condiciones de uso no se debe permitir que el monitor tenga una resolución diferente de la del original.

La pantalla debe permitir el ajuste y la calibración adecuada de la luminancia y el color mediante un fotómetro profesional. Para llevar a cabo la prueba la calibración, la pantalla debe cumplir los parámetros especificados en la Recomendación pertinente.

1.2 Distancia de observación

La distancia de observación a la que se situarán los expertos se debe elegir de conformidad con la resolución de la pantalla, a la altura de la parte activa de la pantalla, en función de la distancia de observación nominal que se describe en la Recomendación UIT-R BT.2022, aunque puede ser menor con arreglo a los requisitos relativos a las condiciones críticas de observación.

1.3 Condiciones de observación

No es imprescindible llevar a cabo los ensayos de protocolos de observación para expertos (EVP) en un laboratorio de pruebas, pero si es importante proteger el emplazamiento de las pruebas de perturbaciones acústicas o visuales (por ejemplo, se puede utilizar también un despacho o una sala de reuniones tranquilos).

Se debe eliminar cualquier fuente de luz directa o indirecta que incida en la pantalla; la luz ambiente debe ser tenue y se mantendrá con la menor intensidad posible que permita rellenar los formularios (si procede).

El número de expertos sentados frente al monitor puede ser variable en función del tamaño de la pantalla con el fin de garantizar una idéntica reproducción de la imagen y la misma presentación de estímulos para todos los observadores.

2 Observadores

Los observadores que participen en un experimento EVP deben ser expertos en la materia de estudio.

Los observadores no se seleccionarán necesariamente por su agudeza visual o daltonismo puesto que debe tratarse de personas cualificadas.

El número mínimo de observadores diferentes será de nueve.

Para lograr el número mínimo de observadores se puede repetir la prueba realizando el mismo experimento en la misma ubicación o en más de una ubicación. Las calificaciones obtenidas en diferentes ubicaciones que participen en una sesión de observación para expertos se pueden procesar juntas de forma estadística.

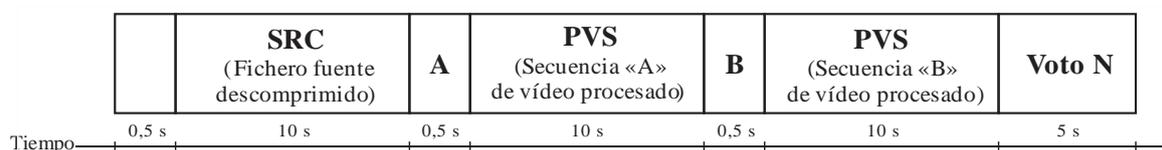
3 Célula básica de prueba

El material que se presente a los expertos se debe organizar a partir de una célula básica de prueba (BTC) para cada pareja de condiciones de codificación sometida a evaluación (véase la Fig. 1).

Los videoclips de secuencias fuente de referencia (SRC) y de secuencias de video procesadas (PVS) que se consideren en una BTC siempre se referirán a la misma secuencia de vídeo con el fin de que los expertos puedan identificar cualesquiera mejoras en la calidad visual aportadas por los algoritmos de compresión sometidos a prueba.

FIGURA 1

Temporización de una célula básica de prueba para el Protocolo de observación para expertos



BT.2095-01

La BTC se organizará de la forma siguiente:

- 0,5 segundos con la pantalla ajustada a gris medio (valor medio de la escala de luminancia);
- 10 segundos de presentación del videoclip de referencia descomprimido;
- 0,5 segundos para mostrar el mensaje «A» (primer video sometido a evaluación) sobre un fondo gris medio;
- 10 segundos de presentación de una versión deficiente del videoclip;
- 0,5 segundos para mostrar el mensaje «B» (segundo video sometido a evaluación) sobre un fondo gris medio;
- 10 segundos de presentación de una versión deficiente del videoclip;
- 5 segundos para mostrar un mensaje que solicite la opinión de los observadores.

Al mensaje «Voto» debe seguir un número que facilite la sincronización del formulario de puntuación.

3.1 Formulario de puntuación y escala de calificación

Como se muestra en la Fig. 1, la presentación de los videoclips se debe realizar de forma que se muestre en primer lugar la referencia sin deficiencias (SRC), seguida de dos secuencias de vídeo con deficiencias (PVS). El orden de presentación de la PVS se modificará de forma aleatoria para cada BTC y los observadores no deben conocer el orden de la presentación.

FIGURA 2

Ejemplo de formulario de puntuación para una sesión de observación para expertos de 24 BTC

Sesión 1

Voto 1	Voto 2	Voto 3	Voto 4	Voto 5
<input type="checkbox"/> A <input type="checkbox"/> B				
Voto 6	Voto 7	Voto 8	Voto 9	Voto 10
<input type="checkbox"/> A <input type="checkbox"/> B				
Voto 11	Voto 12	Voto 13	Voto 14	Voto 15
<input type="checkbox"/> A <input type="checkbox"/> B				
Voto 16	Voto 17	Voto 18	Voto 19	Voto 20
<input type="checkbox"/> A <input type="checkbox"/> B				
Voto 21	Voto 22	Voto 23	Voto 24	
<input type="checkbox"/> A <input type="checkbox"/> B				

Asiento

1 2 3

Tema

BT.2095-02

Se utiliza una escala numérica de 11 notas de 10 (degradación imperceptible) a 0 (degradación muy molesta).

El Cuadro 1 indica el significado de las 11 notas de la escala numérica.

CUADRO 1

Significado de las 11 notas de la escala numérica

Nota	Elemento de degradación	
10	Imperceptible	
9	Ligeramente perceptible	En parte
8		En todo
7	Perceptible	En parte
6		En todo
5	Claramente perceptible	En parte
4		En todo
3	Molesto	En parte
2		En todo
1	Muy molesto	En parte
0		En todo

Se solicita a los observadores que rellenen un cuestionario con dos casillas (denominadas «A» y «B») para cada BTC para que anoten en cada una de ellas la calificación seleccionada de la escala numérica de 11 notas.

La Fig. 2 muestra un ejemplo de formulario de puntuación para una sesión constituida por 24 BTC.

Para cada BTC, los observadores rellenan tanto la casilla identificada por la letra **A** (para evaluar el videoclip presentado en primer lugar) como la identificada con la letra **B** (para evaluar el videoclip presentado en segundo lugar).

La presentación del videoclip original sin degradaciones permite a los expertos evaluar cualquier degradación con mayor facilidad.

Durante las «sesiones de capacitación» se debe explicar detalladamente el significado de la escala numérica de 11 notas como se describe a continuación.

3.2 Diseño de las pruebas y constitución de la sesión

El diseñador de la prueba establecerá aleatoriamente el orden de presentación de las BTC, de forma que no se muestre el mismo videoclip con o sin degradación dos veces consecutivas.

Todas las sesiones de observación deberán comenzar con una «fase de estabilización» que incluya las BTC «mejor» y «peor» así como dos sesiones de observación de «calidad media» elegidas entre las que figuran en cada sesión de prueba. Esto permite a los observadores formarse una opinión inmediata sobre la gama de calidades desde el inicio de la sesión de pruebas.

Si la sesión de observación dura más de 20 minutos, el diseñador de la prueba deberá dividirla en dos (o más) sesiones de observación diferenciadas que no excedan de 20 minutos cada una. En este caso, se incluirá la «fase de estabilización» antes de cada sesión de observación.

3.3 Capacitación

Aunque este procedimiento esté previsto para participantes expertos, antes de cada experimento se recomienda organizar una sesión corta de capacitación (con 5 ó 6 BTC).

Los vídeos utilizados durante la sesión de capacitación pueden ser los mismos que los que se utilicen durante la sesión real, aunque el orden de presentación debe ser diferente.

Los observadores deben ejercitarse en el uso de la escala de 11 notas analizando cuidadosamente los videoclips que se muestren inmediatamente después de los mensajes «A» y «B» en la pantalla y deben comprobar si pueden advertir alguna diferencia con el videoclip mostrado en primer lugar (la SRC).

4 Recopilación y tratamiento de los datos

Las notas se deben recopilar al final de cada sesión e introducir en una hoja de cálculo electrónica para obtener los valores MEDIOS.

Se aconseja realizar una «selección posterior» de los observadores utilizando una correlación lineal de Pearson.

La función «correlación» se aplicará tomando en consideración todas las notas de cada sujeto en relación con las notas medias de opinión (MOS); se puede fijar un valor umbral para determinar cada observador como «aceptable» o «rechazado» (la Recomendación UIT-T P.913 sugiere la utilización de un valor umbral de «rechazo» igual a 0,75).

5 Condiciones de uso de los resultados del Protocolo de observación para expertos

El Protocolo de observación para expertos (EVP) se puede emplear cuando no se pueda llevar a cabo un experimento de evaluación subjetiva formal debido a falta de tiempo o de recursos.

El EVP requiere menos tiempo que la evaluación subjetiva formal y puede ejecutarse en un entorno «informal», en el supuesto de que el entorno en el que se lleva a cabo esté protegido de perturbaciones externas visuales y auditivas.

Las únicas condiciones obligatorias se refieren a las condiciones ambientales de iluminación y de observación (pantalla, ángulo de observación y distancia de observación) que se describen en los párrafos anteriores.

6 Limitaciones en el uso de los resultados del EVP

Aunque el EVP está proporcionando resultados aceptables con solo nueve observadores, las MOS obtenidas por un experimento EVP no pueden sustituir los resultados obtenidos mediante un experimento de evaluación subjetiva formal.

Los datos de las MOS obtenidos mediante el EVP se podrían utilizar para disponer de una indicación preliminar del grado de perturbación.

Los datos de las MOS obtenidos mediante el EVP se podrían utilizar para efectuar una clasificación preliminar de los esquemas de procesamiento de vídeo sometidos a evaluación.

Cuando se considere conveniente o necesario, se puede llevar a cabo un experimento EVP en paralelo en diferentes ubicaciones, siempre que las condiciones de observación, la distancia de observación y el diseño de la prueba sean idénticos.

Si el experimento se realiza en diferentes ubicaciones y el número de observadores expertos implicados en el mismo EVP es superior a 15, se podrían procesar los datos subjetivos brutos para obtener los datos de MOS, desviación típica e intervalo de confianza que pueden contribuir a establecer una clasificación más precisa de los casos sometidos a prueba. En este último caso se pueden realizar análisis de estadística inferencial más precisos, por ejemplo, pruebas T-Student.

Anexo 2 (informativo)

Aplicación del protocolo de observación por expertos y su comportamiento en presencia de un gran número de expertos evaluadores

En este Anexo informativo se presentan los resultados de dos evaluaciones subjetivas de vídeos con codificación HD y UAD utilizando EVP, realizadas durante la 117ª reunión del MPEG. Para realizar las evaluaciones se aplicaron las disposiciones de la Recomendación UIT-R BT.2095 a fin de establecer una clasificación rápida y fiable de dos métodos de codificación fuente distintos.

Dado que a la 117ª reunión del MPEG asistió un elevado número de expertos, el número de evaluadores que participaron en las dos sesiones EVP superó con mucho los 9 que se aconsejan en la Recomendación UIT-R BT.2095. A la prueba EVP HD asistieron 30 expertos y a la prueba EVP UAD asistieron 32 expertos.

La elevada participación de evaluadores expertos brindó la oportunidad de analizar los datos MOS a fin de verificar el nivel de fiabilidad inherente al uso de la Recomendación UIT-R BT.2095 para clasificar vídeos codificados.

Para esta evaluación se formaron cuatro grupos de observadores (a saber, de 9, 12, 15 y 18 observadores) y se compararon los valores MOS obtenidos por el grupo de 9 expertos con los obtenidos por los grupos de 12, 15 y 18 expertos observadores.

El objetivo era comparar la clasificación establecida por los 9 expertos (y, por tanto, conforme con el protocolo EVP) y las clasificaciones establecidas por los grupos de 12, 15 y 18 expertos (semejante a evaluación subjetiva formal).

Como puede verse en la Figura 3 (experimento con contenido UAD) y en la Figura 4 (experimento con contenido HD), los resultados de las clasificaciones son muy semejantes en los cuatro casos considerados.

Tomando los resultados obtenidos por el grupo de 18 observadores como una especie de «verdad fundamental», se pueden trazar los gráficos de las Figuras 3 y 4 clasificando los puntos de prueba según los valores MOS obtenidos por el grupo de 18 observadores (línea roja continua).

Las demás líneas del gráfico muestran los resultados obtenidos por los grupos de 9 (línea roja discontinua), 12 (línea azul discontinua) y 15 (línea verde continua) observadores.

De la comparación de los resultados representados en las Figuras 3 y 4 se desprende lo siguiente:

- los gráficos de los grupos de 15 y 18 observadores muestran una pendiente homogénea de los valores MOS de mayor calidad a menor calidad;
- los gráficos de los grupos de 9 y 12 observadores muestran algunas «inversiones» de la clasificación con respecto al grupo de 18 observadores, aunque la variación de la puntuación es bastante limitada.

Como conclusión puede decirse que los experimentos EVP aquí descritos demuestran un muy buen rendimiento del protocolo EVP, lo que confirma lo indicado en la Recomendación UIT-R BT.2095, a saber, que la utilización del protocolo EVP, aunque no puede considerarse un sustituto completo del experimento subjetivo formal, puede considerarse un procedimiento de evaluación estable con el que se obtienen resultados muy cercanos a los que se consiguen cuando se dispone de muchos más observadores y se realiza una evaluación subjetiva formal.

FIGURA 3
Clasificación del experimento UAD en función del número de evaluadores

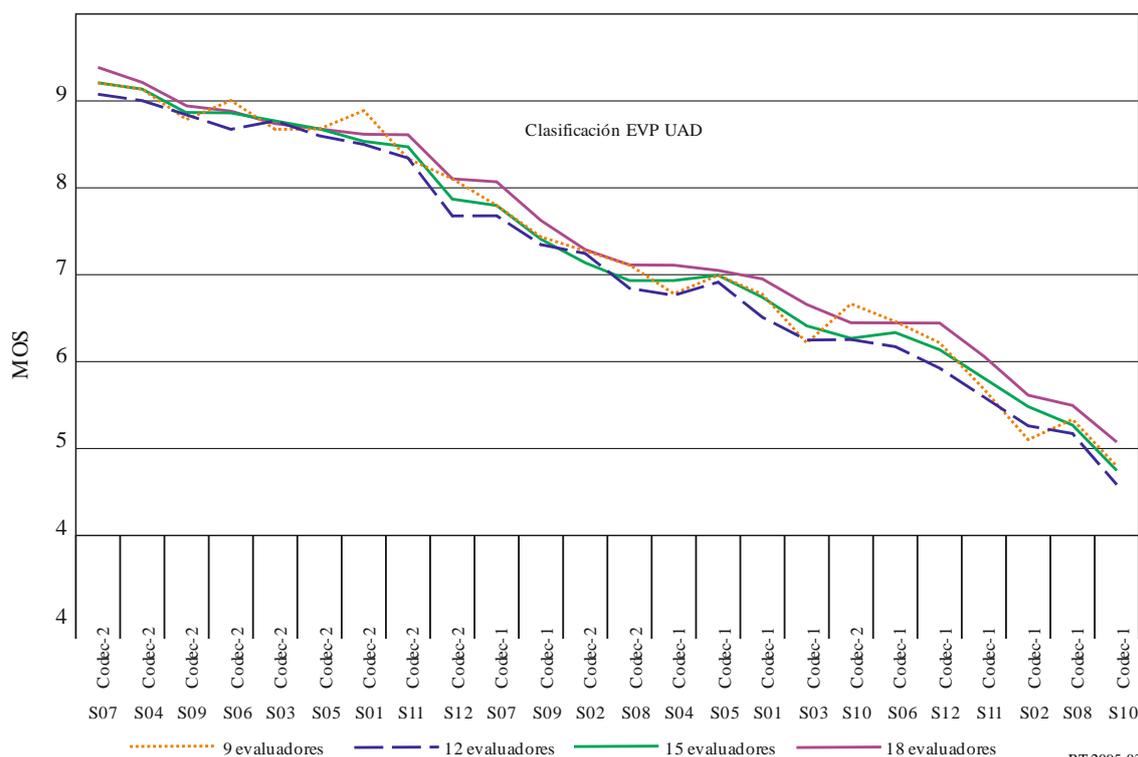


FIGURA 4
 Clasificación del experimento HD en función del número de evaluadores

