

Unión Internacional de Telecomunicaciones

**UIT-R**

Sector de Radiocomunicaciones de la UIT

**Recomendación UIT-R BT.2021-1**  
(02/2015)

**Métodos de evaluación subjetiva de los  
sistemas de televisión 3D estereoscópica**

**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   |
|------------|--|
| <b>BO</b>  | Distribución por satélite  |
| <b>BR</b>  | Registro para producción, archivo y reproducción; películas en televisión  |
| <b>BS</b>  | Servicio de radiodifusión (sonora)   |
| <b>BT</b>  | <b>Servicio de radiodifusión (televisión)</b>  |
| <b>F</b>   | Servicio fijo  |
| <b>M</b>   | Servicios móviles, de radiodeterminación, de aficionados y otros servicios por satélite conexos                  |
| <b>P</b>   | Propagación de las ondas radioeléctricas   |
| <b>RA</b>  | Radioastronomía  |
| <b>RS</b>  | Sistemas de detección a distancia  |
| <b>S</b>   | Servicio fijo por satélite   |
| <b>SA</b>  | Aplicaciones espaciales y meteorología   |
| <b>SF</b>  | Compartición de frecuencias y coordinación entre los sistemas del servicio fijo por satélite y del servicio fijo |
| <b>SM</b>  | Gestión del espectro   |
| <b>SNG</b> | Periodismo electrónico por satélite  |
| <b>TF</b>  | Emisiones de frecuencias patrón y señales horarias   |
| <b>V</b>   | 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, 2016

© UIT 2016

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.2021-1

**Métodos de evaluación subjetiva de los sistemas de televisión 3D estereoscópica**

(2012-2015)

**Cometido**

Esta Recomendación proporciona métodos de evaluación de sistemas de televisión 3D estereoscópica, incluyendo métodos generales de prueba, escalas de apreciación y condiciones de visualización.

La Asamblea de Radiocomunicaciones de la UIT,

*considerando*

- a) que se ha recopilado gran cantidad de información sobre métodos utilizados en diversos laboratorios para la evaluación de características de calidad de funcionamiento críticas de sistemas de televisión 3D;
- b) que el análisis de dichos métodos pone de manifiesto la existencia de un considerable nivel de acuerdo entre los laboratorios sobre un conjunto de aspectos a tener en cuenta en la realización de dichas pruebas;
- c) que es importante la adopción de métodos normalizados para el intercambio de información entre laboratorios;
- d) que la introducción de un servicio de televisión 3D puede exigir el desarrollo de nuevos formatos de imagen así como de técnicas de procesamiento y transmisión de imágenes, cuya calidad de funcionamiento debe evaluarse mediante metodologías subjetivas,

*recomienda*

**1** que se utilicen los métodos generales de prueba, escalas de apreciación y condiciones de visualización para la evaluación de la calidad de la imagen de televisión 3D estereoscópica en experimentos de laboratorio descritos en el Anexo 1 y, siempre que sea posible, en evaluaciones realizadas en funcionamiento real.

**Anexo 1****1 Dimensiones de la evaluación (perceptual)**

La televisión en 3D estereoscópica explota las características del sistema de visión binocular humana, recreando las condiciones que producen la percepción de la profundidad relativa de los objetos en el campo visual. El principal requisito de la tecnología actualmente disponible para la composición de una imagen estereoscópica es la captura de al menos dos vistas de la escena tomadas por dos cámaras alineadas horizontalmente. Las imágenes de los objetos de la escena tendrán posiciones relativas diferentes en función de la visión de que se trate, a saber, visión izquierda y visión derecha. La diferencia entre las posiciones relativas de las dos visiones es lo que comúnmente se denomina disparidad de imágenes (o paralaje) y normalmente se expresa en píxeles, distancia físicas (retiniana) o con una medición relativa (por ejemplo, porcentaje de la anchura de la pantalla). La disparidad de

imágenes debe distinguirse de la disparidad angular (retiniana). De hecho, la misma información de disparidad de imagen produce distintas disparidades angulares (retinianas) con distintas distancias de visualización. La magnitud y dirección de la percepción de profundidad se basa en la magnitud y dirección de las disparidades retinianas resultado de la imagen estereoscópica.

Los factores de evaluación generalmente aplicados a las imágenes de televisión monoscópica, tales como resolución, reproducción del color, representación del movimiento, calidad global, nitidez de la imagen, etc. también pueden aplicarse a los sistemas de televisión estereoscópica. Además, existen numerosos factores específicos de los sistemas de televisión estereoscópica. Entre ellos se incluyen factores como la resolución en profundidad, que es la resolución espacial en el sentido de la profundidad, el movimiento en profundidad, que determina si el movimiento en la dirección de la profundidad se reproduce sin discontinuidades ni distorsiones espaciales. Dos ejemplos bien conocidos de esto último son el *efecto teatro de marionetas*, que se produce cuando los objetos se perciben anormalmente grandes o pequeños, y el *efecto papel de cartón* (o aplanamiento de la profundidad de las imágenes), cuando los objetos se perciben de forma estereoscópica pero con una menor profundidad que hace que parezcan anormalmente delgados.

Se identifican tres dimensiones básicas de la percepción, que en su conjunto afectan a la calidad de experiencia de un sistema estereoscópico: *calidad de la imagen*, *calidad de profundidad* y *confort visual*. Algunos investigadores consideran que el impacto psicológico de las tecnologías de representación estereoscópica también podría medirse mediante conceptos más generales, como la *naturalidad* o el *sentido de presencia*.

### **Dimensiones perceptuales primarias**

*Calidad de la imagen*, que hace referencia a la calidad de imagen que proporciona el sistema. Es el principal parámetro que determina la calidad de funcionamiento de un sistema de vídeo. La calidad de la imagen está principalmente afectada por parámetros técnicos y otros errores introducidos, por ejemplo, por procesos de codificación y/o transmisión

*Calidad de la profundidad*, que hace referencia a la capacidad del sistema para ofrecer una sensación mejorada de profundidad. La presencia de factores monoculares, como perspectiva lineal, imagen borrosa, gradientes, etc. ofrece sensaciones de profundidad incluso en imágenes 2D. No obstante, las imágenes 3D estereoscópicas contienen además información de disparidad que proporciona información de profundidad que mejora la sensación de profundidad en comparación con los sistemas 2D.

*Confort visual*, que se refiere a la sensación subjetiva de comodidad/incomodidad que puede asociarse a la visión de imágenes estereoscópicas. Las imágenes estereoscópicas que hayan sido capturadas de forma incorrecta o sea representadas de forma inadecuada pueden generar una notable incomodidad.

### **Dimensiones adicionales de la percepción**

*Naturalidad*, que se refiere a la percepción de la imagen estereoscópica como una representación fiel de la realidad (es decir, el realismo perceptual). La imagen estereoscópica puede presentar distintos tipos de distorsiones que la hagan menos natural. Por ejemplo, los objetos estereoscópicos se perciben en ocasiones como anormalmente grandes o pequeños (efecto teatro de marionetas) o parecen anormalmente delgados (efecto papel de cartón).

*Sentido de presencia*, que se refiere a la experiencia subjetiva de estar en un lugar o entorno dado aunque uno se encuentra situado en otro distinto.

En esta Recomendación se presenta información relativa a los métodos y procedimientos para la evaluación de las tres dimensiones primarias señaladas: calidad de la imagen, calidad de la profundidad y confort visual. La presente Recomendación no incluye metodologías para la evaluación de la naturalidad ni del sentido de presencia, aunque su inclusión está prevista en una versión posterior.

## 2 Metodologías subjetivas

La Recomendación UIT-R BT.500 incluye varias metodologías para la evaluación de la calidad de la imagen. En todos los métodos se realizan un conjunto de pruebas de evaluación con un grupo de observadores utilizando una serie de secuencias de vídeo, que han sido procesadas por los sistemas en estudio (por ejemplo, un algoritmo con distintos parámetros, una tecnología de codificación que admite distintas velocidades binarias, diferentes casos de transmisión etc.). En cada prueba se pide a los observadores que evalúen una característica relevante (por ejemplo, la calidad de la imagen) de la secuencia o secuencias de vídeo utilizando una escala preestablecida. Los métodos difieren entre sí principalmente en el modo de presentación, es decir, la forma en la que se presentan al observador las secuencia de vídeo y en la escala utilizada por los observadores para evaluar dichas secuencias.

Las imágenes de prueba son imágenes estéreo binoculares seleccionadas en base a los elementos descritos en § 4. Los observadores evalúan las tres características siguientes:

- calidad de la imagen: efecto sobre la resolución de las imágenes 3D estereoscópicas de un sistema que incluye un trayecto entre las imágenes en prueba y la pantalla donde se muestran;
- calidad de la profundidad: efecto sobre la profundidad de percepción de las imágenes 3D estereoscópicas de un sistema que incluye un trayecto entre las imágenes en prueba y la pantalla donde éstas se muestran;
- confort visual: efecto sobre la facilidad de visualización de las imágenes 3D estereoscópicas de un sistema que incluye un trayecto entre las imágenes en prueba y la pantalla donde éstas se muestran;

Esta Recomendación incluye seis métodos de la Recomendación UIT-R BT.500. Dichos métodos se han utilizado satisfactoriamente en las últimas dos décadas en trabajos de investigación sobre calidad de imagen, calidad de profundidad y confort visual de tecnologías de imágenes estereoscópicas. Son los siguientes:

- método de estímulo único (SS, *single-stimulus*);
- método de escala de degradación con doble estímulo (DSIS, *double-stimulus impairment scale*);
- método de escala de calidad continua de doble estímulo (DSCQS, *double stimulus continuous quality scale*);
- método de comparación de estímulos (SC, *stimulus-comparison*);
- método de evaluación de calidad continua de estímulo único (SSCQE, *single stimulus continuous quality evaluation*);
- método de doble estímulo simultáneo para evaluación continua (SDSCE, *simultaneous double stimulus for continuous evaluation*).

Cuando ha sido adecuado, los métodos se han utilizado de forma ligeramente distinta, por ejemplo, con diferentes escalas de confort visual. En los Cuadros 1, 2 y 3 se resumen el modo de presentación y las escalas asociadas con los métodos de evaluación de la calidad de imagen, calidad de profundidad y confort visual, respectivamente.

En el punto siguiente se hace una breve descripción de cada metodología. Los elementos metodológicos comunes a todos los métodos se presentan en puntos posteriores.

### 2.1 Método de estímulo único (SS)

El procedimiento consta de una serie de pruebas de evaluación que pueden dividirse, cuando sea adecuado, en varias sesiones de pruebas separadas por interrupciones. En cada prueba, sólo se presenta la secuencia de un vídeo en «prueba» (es decir, una secuencia procesada con el sistema investigado) que se evalúa de forma independiente según la escala preestablecida.



### 2.1.1 Estructura de las pruebas del método de estímulo único

En cada prueba, la secuencia de vídeo en «Prueba» está precedida y seguida por la presentación de un campo gris-medio. El campo gris-medio precedente puede contener un objeto para la fijación, por ejemplo, el número de la prueba, con disparidad cero y que debe permanecer tres segundos o menos ( $\leq 3$  s). El campo gris-medio posterior puede contener un recordatorio de que debe realizarse la valoración, por ejemplo, incluyendo la frase «vote ahora», y debe durar el tiempo suficiente como para permitir que el observador vote (por ejemplo, hasta 10 s). La duración de la secuencia de vídeo en «Prueba» debe ser, por lo general, de unos 10 s<sup>1</sup>. La estructura de una típica prueba de estímulo único se muestra en la Fig. 1.

### 2.1.2 Escalas de apreciación del método de estímulo único

Para la evaluación de la calidad de imagen pueden utilizarse dos escalas etiquetadas: la escala discreta de cinco niveles y la escala de calidad continua estándar de la UIT (véase el Cuadro 1. Las etiquetas de calidad son «Excelente (*Excellent*)», «Buena (*Good*)», «Aceptable (*Fair*)», «Mediocre (*Poor*)» y «Mala (*Bad*)».

Las mismas escalas pueden utilizarse para evaluar la calidad de la profundidad (véase Cuadro 2). En este caso, se pide a los observadores que evalúen la calidad de la profundidad de la representación más que la calidad de la propia imagen.

Para la evaluación del confort visual pueden utilizarse dos escalas etiquetadas: una escala discreta de cinco niveles y la escala de confort continua (véase el Cuadro 3). Las etiquetas de confort son «Muy cómodo (*Very Comfortable*)», «Cómodo (*Comfortable*)», «Ligeramente incómodo (*Mildly Uncomfortable*)», «Incómodo (*Uncomfortable*)» y «Muy incómodo (*Extremely uncomfortable*)».

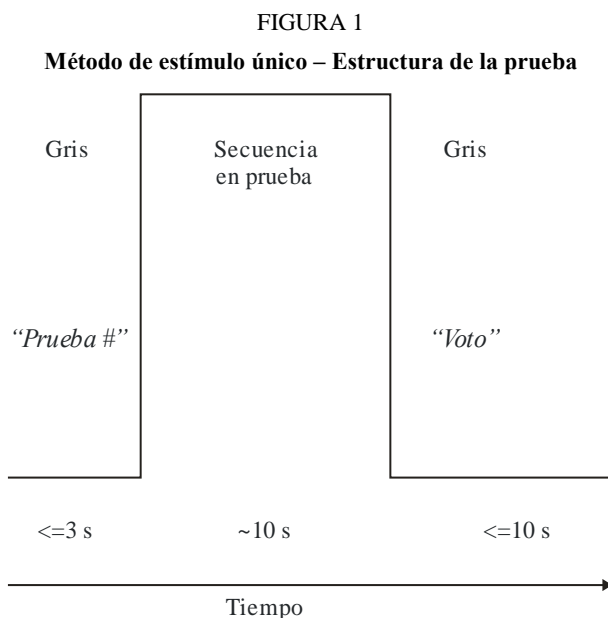
### 2.1.3 Datos de las notas de opinión del método de estímulo único

La valoración de cada secuencia examinada se denomina «*nota de opinión*». La media de dichas notas obtenidas para cada sistema investigado, se denomina nota media de opinión (MOS).

En el conjunto de las secuencias utilizadas pueden incluirse secuencias de vídeos de «Referencia», que son versiones de las secuencias en prueba que no han sido objeto de ningún procesamiento (véase § 8). Ello permite calcular la «*diferencia de la nota de opinión*», que es la diferencia aritmética entre la valoración de la secuencia en «Prueba» y de la versión de «Referencia» de la secuencia analizada. La media de las diferencias de las notas de opinión obtenidas para cada sistema analizado se denomina diferencia de nota media de opinión (DMOS).

---

<sup>1</sup> Algunos investigadores han abogado por la utilización de secuencias más largas en base al supuesto de que la apreciación plena de un contenido estereoscópico necesita más tiempo que la apreciación de un contenido monoscópico normal (2D). Hasta la fecha, existe escasa evidencia empírica a favor o en contra de dicho argumento.



BT.2021-01

## 2.2 Método de escala de degradación con doble estímulo (DSIS) (método de la EUR)

El método de doble estímulo (método UER) es cíclico en la medida en que se muestra al evaluador una imagen de referencia no degradada, y después la misma imagen degradada. A continuación, se le pide que opine sobre la segunda, con la primera en mente. En sesiones, que duran hasta media hora, se muestra al evaluador una serie de imágenes o secuencias en orden aleatorio y con degradaciones aleatorias que abarcan todas las combinaciones requeridas. La imagen no degradada se incluye en las imágenes o secuencias que deben evaluarse. Al final de la serie de sesiones, se calcula la nota media para cada condición de prueba y para cada imagen de prueba.

Este método utiliza la escala de degradación, cuyos resultados se suelen considerar más estables para degradaciones pequeñas que para degradaciones considerables. Si bien algunas veces se ha utilizado el método con una escala de degradaciones limitada, es más conveniente utilizarlo con una gama completa de degradaciones. La disposición general del sistema de prueba debería ser la que se indica en la Fig. 2.

### 2.2.1 Presentación del material de prueba

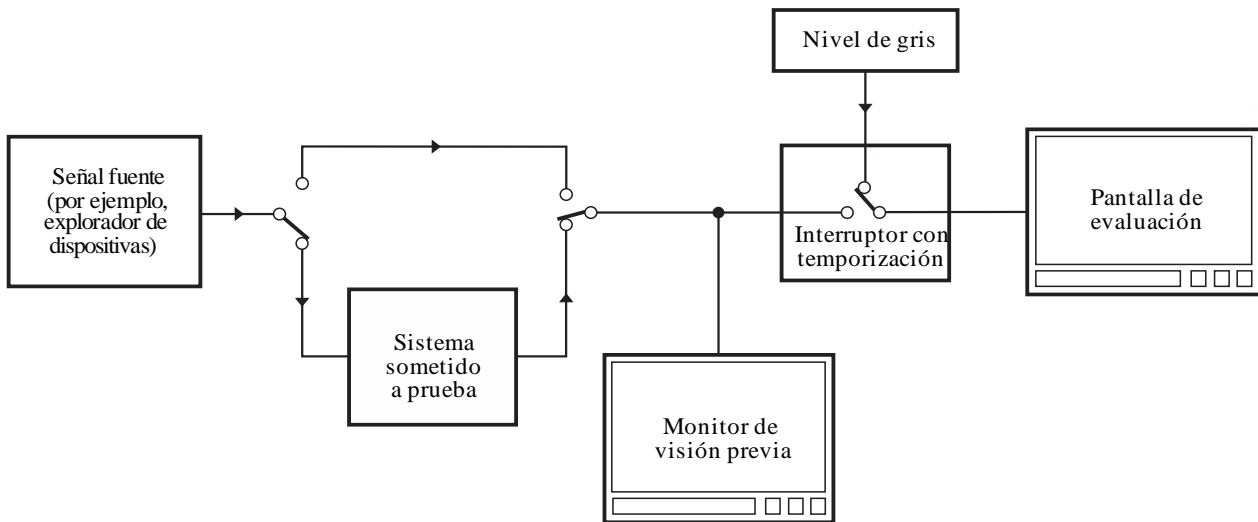
Una sesión de prueba consta de varias presentaciones. Hay dos variantes de la estructura de las presentaciones, la I y la II que se indican a continuación:

Variante I: La imagen o secuencia de referencia y la imagen o secuencia de prueba se presentan sólo una vez, como muestra la Fig. 3a).

Variante II: La imagen o secuencia de referencia y la imagen o secuencia de prueba se presentan dos veces, como muestra la Fig. 3b).

La variante II, que tiene una mayor duración que la variante I, puede aplicarse si es necesario discriminar entre degradaciones muy pequeñas o se están sometiendo a prueba secuencias en movimiento.

FIGURA 2  
Disposición general de los sistemas de prueba para el método de DSIS



BT.2021-02

### 2.2.2 Escalas de apreciación

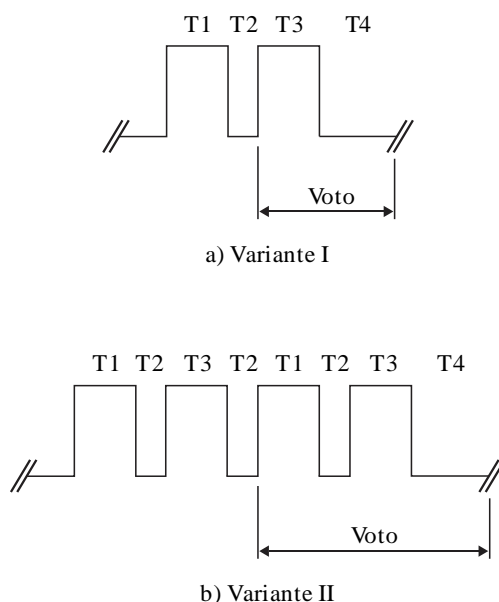
Debe utilizarse la escala de apreciación de cinco notas (véanse los Cuadros 1, 2 y 3). Los evaluadores deben utilizar un formulario que indique muy claramente la escala, y que cuente con cuadros numerados u otro medio para registrar las notas.

### 2.2.3 Datos de nota de opinión del método DSIS

El método de doble estímulo (método UER) es cíclico en la medida en que se muestra al evaluador una imagen de referencia no degradada, y después la misma imagen degradada. A continuación, se le pide que opine sobre la segunda, con la primera en mente. En sesiones, que duran hasta media hora, se muestra al evaluador una serie de imágenes o secuencias en orden aleatorio y con degradaciones aleatorias que abarcan todas las combinaciones requeridas. La imagen no degradada se incluye en las imágenes o secuencias que deben evaluarse. Al final de la serie de sesiones, se calcula la nota media para cada condición de prueba y para cada imagen de prueba.



FIGURA 3  
Estructura de presentación del material de prueba



*Fases de presentación:*

|             |  |
|-------------|--|
| T1 = 10 s   | Imagen de referencia                                     |
| T2 = 3 s    | Gris mediano producido por un nivel vídeo de unos 200 mV |
| T3 = 10 s   | Condición a evaluar                                      |
| T4 = 5-11 s | Gris mediano   |

La experiencia sugiere que prolongar los periodos T1 y T3 más allá de 10s no mejora la capacidad del evaluador para juzgar las imágenes o las secuencias.

BT.2021-03

## 23 Método de escala de calidad continua de doble estímulo (DSCQS)

El procedimiento consta de una serie de pruebas de valoración que pueden dividirse, cuando sea adecuado, en varias sesiones de prueba separadas por interrupciones. En cada prueba, se presentan dos veces dos versiones de la misma secuencia de vídeo, es decir, un total de cuatro presentaciones. La configuración generalizada del sistema de prueba es la que se representa en la Fig. 4.

### 2.3.1 Presentación del material de prueba

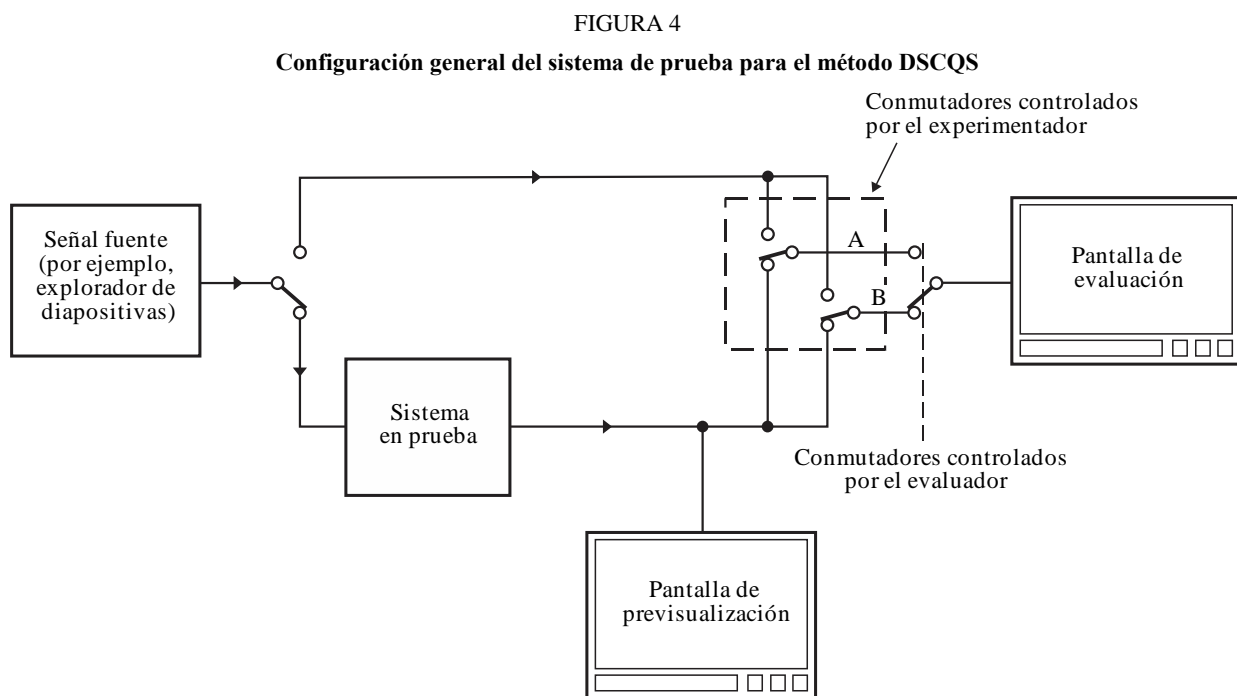
Una sesión de prueba consta de un cierto número de presentaciones. En el caso de la variante I, en cada presentación el evaluador puede pasar de la señal A a la B y viceversa hasta que haya cuantificado mentalmente la calidad asociada a cada señal. El observador puede repetir esa maniobra típicamente dos o tres veces durante periodos de hasta 10 s. En la variante II, en la que se utiliza simultáneamente un conjunto de observadores, la pareja de señales se visualizan durante periodos de tiempo iguales una o más veces antes de registrar resultados, de forma que los evaluadores cuantifiquen las calidades de las mismas; entonces se muestran de nuevo las dos imágenes una o más veces al tiempo que los observadores registran sus apreciaciones. El número de repeticiones depende de la duración de las secuencias de prueba. En el caso de imágenes fijas, puede ser adecuado repetir cinco veces una secuencia durante 3-4 s, realizándose la votación en las dos últimas repeticiones. En el caso de imágenes en movimiento con señales molestas que varían con el tiempo, puede ser adecuado repetir dos veces una secuencia de 10 s, realizándose la votación en la segunda. En la Fig. 5 se muestra la estructura de las presentaciones.

### 2.3.2 Escalas de apreciación del método DSCQS

En el método DSCQS, se pide a los observadores que valoren las secuencias de vídeo A y B. Para la evaluación de la calidad de la imagen y la calidad de la profundidad puede utilizarse escala de calidad continua estándar de la UIT (véanse los Cuadros 1 y 2). Para la evaluación del confort visual, debe utilizarse la escala continua de confort con las etiquetas «Muy cómodo (*Very Comfortable*)», «Cómodo (*Comfortable*)», «Ligeramente incómodo (*Mildly Uncomfortable*)», «Incómodo (*Uncomfortable*)» y «Muy incómodo (*Extremely Uncomfortable*)» como se muestra en el Cuadro 3.

### 2.3.3 Datos de nota de opinión del método DSCQS

Las valoraciones de las versiones de «Prueba» y de «Referencia» de cada secuencia obtenidas en cada prueba se utilizan para calcular la diferencia de las notas de opinión. Estas últimas permiten calcular la el parámetro DMOS de cada sistema investigado.



BT.2021-04

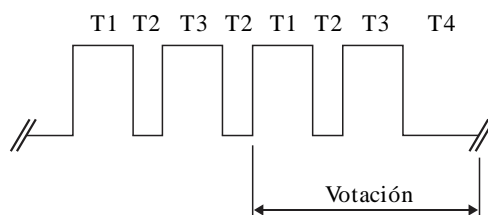
Existen dos variantes de este método, denominadas I y II, que se describen a continuación.

**Variante I:** El evaluador, que normalmente está solo, puede conmutar entre dos líneas, A y B, hasta que considere que tiene una opinión definitiva para cada una de ellas. Las líneas A y B se alimentan con la imagen de referencia o con la imagen que se transmite a través del sistema en prueba, aunque la señal que se inserta en una u otra línea varía aleatoriamente de una prueba a otra, de forma que el observador aprecia la diferencia pero no recibe al respecto ninguna información.

**Variante II:** Las imágenes de las líneas A y B se muestran sucesivamente y varias veces a un conjunto de evaluadores a fin de que cuantifiquen su opinión de cada una. En cada presentación, las líneas A y B se alimentan tal como en la variante I. La estabilidad de los resultados de esta variante con una gama limitada de valores de calidad se considera aún sujeta a estudio.

FIGURA 5

Método de escala de calidad continua de doble estímulo – Estructura de la prueba

*Fases de presentación:*

|             |                       |
|-------------|-----------------------|
| T1 = 10 s   | Secuencia de prueba A |
| T2 = 3 s    | Nivel gris-medio      |
| T3 = 10 s   | Secuencia de prueba B |
| T4 = 5-11 s | Nivel gris-medio      |

BT.2021-05

## 2.4 Método de comparación de parejas (PC)

En el método de comparación de parejas (PC, *pair comparison*), un conjunto de secuencias de «Prueba», es decir, secuencias procesadas por diferentes sistemas (por ejemplo, diferentes velocidades binarias, diferentes algoritmos, etc.), se comparan por parejas (es decir, se muestran de dos en dos). Los observadores deben valorar qué elemento de la pareja prefieren en el contexto de la prueba. El número de valoraciones necesarias es función del número de sistemas analizados. Los sistemas en prueba (X, Y, Z, etc.) se combinan formando todas las parejas posibles,  $n(n-1)$ , XY, ZY, YZ, etc. Además, las parejas deben visualizarse en los dos órdenes posibles (por ejemplo, XY, YX).

### 2.4.1 Estructura de prueba del método PC

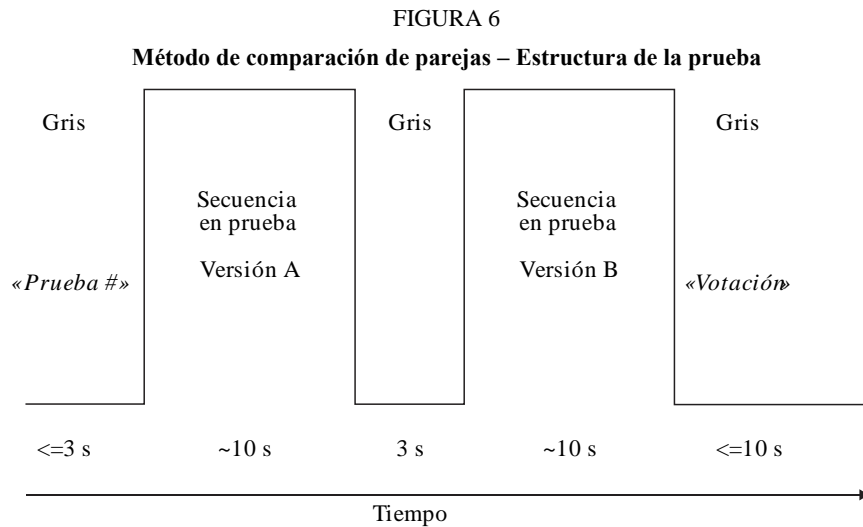
La prueba comienza con la presentación de un campo gris-medio que puede contener un objetivo de fijación, por ejemplo, el número de la prueba, con disparidad cero y durante un tiempo igual o inferior a 3 s ( $\leq 3$  s). A continuación, se presentan las secuencias a comparar. La duración de cada secuencia en prueba debe ser de unos 10 s. Las secuencias pueden presentarse simultáneamente en dos pantallas (o una junto a otra en la misma pantalla) o sucesivamente (por ejemplo, AB) en la misma pantalla. En este último caso, las secuencias están temporalmente separadas por la presentación de un gris-medio de 3 s de duración. La prueba termina con una pantalla gris-medio que puede recordar que debe votarse, por ejemplo, con la expresión «vote ahora», que permanece al menos el tiempo necesario para que el observador proporcione su apreciación (por ejemplo,  $\leq 10$  s). En la Fig. 6 se representa una prueba PC típica.

### 2.4.2 Escalas de apreciación del método PC

Puede solicitarse a los observadores que proporcionen una valoración de preferencia simple utilizando una escala binaria (por ejemplo, A es la preferida), o bien, que proporcionen una preferencia graduada (por ejemplo, A es mucho mejor que B). Pueden utilizarse las mismas escalas para la calidad de imagen, la profundidad de imagen y el confort visual (véanse los Cuadros 1, 2 y 3).

### 2.4.3 Datos de nota de opinión del método PC

La valoración del método de PC se realiza en términos de preferencias.



## 2.5 Método de evaluación de calidad continua de estímulo único (SSCQE)

Los niveles de calidad de imagen, calidad de profundidad y confort visual pueden fluctuar ampliamente a lo largo del tiempo cuando se utilizan fragmentos breves de vídeos estereoscópicos codificados digitalmente; dichas fluctuaciones pueden depender del contenido de la escena y de la duración de las señales molestas (breves o largas) que afecten a las tres dimensiones básicas. El método SSCQE se ha elaborado con el fin de abordar el impacto de dicha dinámica. En el método SSCQE la calidad de imagen, la calidad de profundidad y el confort visual de las secuencias de vídeo estereoscópico se evalúan de forma continua (es decir, según cambian a lo largo del tiempo). Esta metodología se considera, por lo general, más representativa de los patrones de observación reales en el hogar.

### 2.5.1 Forma general del protocolo de prueba

Las sesiones de prueba se presentan a los sujetos en el formato siguiente:

- *Segmento de programa (PS; programme segment)*: un PS corresponde a un tipo de programa (por ejemplo, deportes, noticias, películas) procesado de conformidad con el parámetro de calidad (QP) objeto de evaluación (por ejemplo, velocidad binaria); cada PS debe tener una duración mínima de 5 min.
- *Sesión de prueba (TS, test session)*: una TS es una serie de una o más combinaciones diferentes PS/QP sin separación entre ellas y dispuestas en un orden pseudoaleatorio. Cada TS contiene, al menos una vez, todos los PS y las QP pero no necesariamente todas las combinaciones PS/QP; cada TS debe tener una duración de entre 30 y 60 min.
- *Presentación de prueba (TP, test presentation)*: una TP es la representación completa de una prueba. Una TP puede dividirse en varios TS para cumplir los requisitos de duración máxima y evaluar la calidad para todas las parejas PS/QP. Si el número de parejas PS/QP está limitado, una TP puede estar compuesta de la misma TS repetida para que la prueba se realice durante un periodo de tiempo suficientemente largo.

Para la evaluación de la calidad de servicio puede añadirse una señal de audio. En ese caso, la elección del audio de acompañamiento, realizado antes de la prueba de calidad de funcionamiento, tiene una importancia similar a la selección del vídeo.

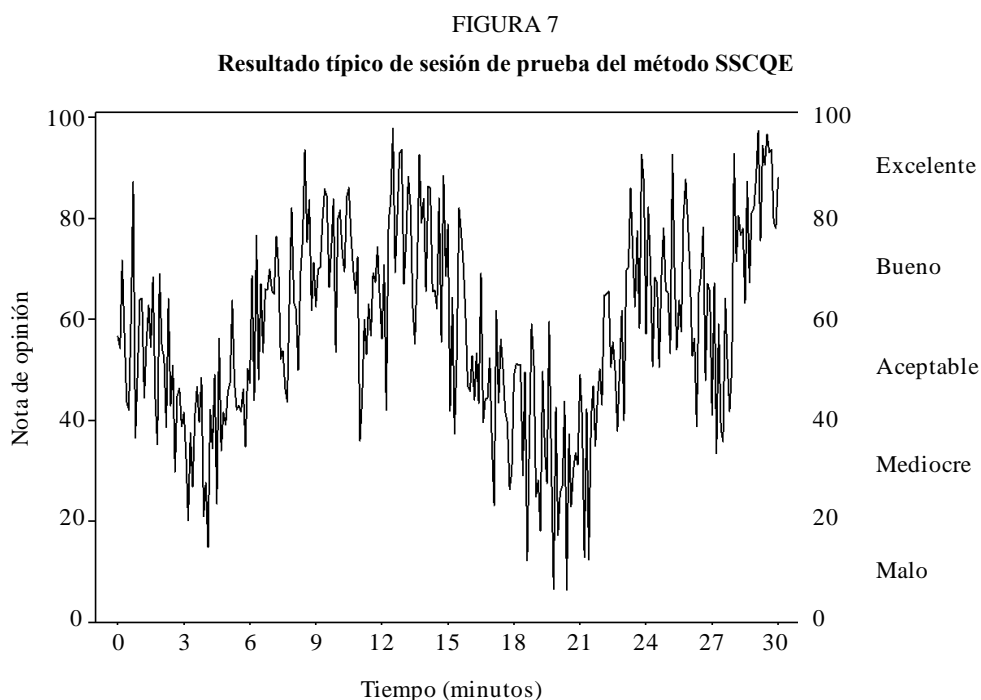
El formato de prueba más sencillo utiliza un único segmento de programa (PS) y un único parámetro de calidad (QP).

### 2.5.2 Escalas de apreciación del método SSCQE

Para la evaluación de la calidad de imagen y de la calidad de profundidad, debe utilizarse la escala de calidad continua estándar de la UIT (véanse los Cuadros 1 y 2). Para el confort visual, debe utilizarse la escala continua de confort que se muestra en el Cuadro 3. La Fig. 7 incluye el ejemplo de una *sesión de prueba* utilizando una escala de calidad.

### 2.5.3 Datos de nota de opinión del método SSCQE

Para calcular la valoración media de calidad en función del tiempo  $q(t)$  deben recopilarse los datos de todas las sesiones de prueba. Los resultados pueden presentarse como la media de las apreciaciones de calidad de todos los observadores por segmento de programa, contenido de vídeo o sesión de prueba.



## 2.6 Método de doble estímulo simultáneo para evaluación continua (SDSCE)

El método SSCQE es capaz de medir la calidad de vídeo estereoscópico en secuencias más largas, representativas del contenido de vídeo estereoscópico y de la estadística de errores. Para reproducir las condiciones de observación que estén lo más próximas posibles a las situaciones reales, en el SSCQE no se utilizan referencias.

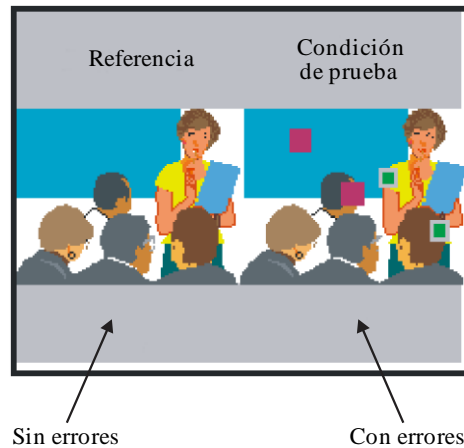
Cuando hay que evaluar la fidelidad, se han de introducir condiciones de referencia. El SDSCE ha sido elaborado a partir del SSCQE, con ligeras diferencias en cuanto a la manera de presentar las imágenes a los sujetos y con respecto a la escala de apreciación. El método fue propuesto a MPEG para evaluar la solidez contra los errores a velocidades binarias muy bajas, pero puede ser aplicado adecuadamente a todos los casos en los que hay que evaluar la fidelidad de la información visual afectada por la degradación que varía en función del tiempo.

### 2.6.1 Presentación del material de prueba

El grupo de sujetos observa dos secuencias al mismo tiempo: una es la referencia, la otra es la condición de prueba. Si el formato de las secuencias es de formato de imagen normalizado (SIF) o

más pequeño, las dos secuencias pueden ser visualizadas juntas en el mismo monitor; en los demás casos se debe utilizar dos monitores alineados (véase la Fig. 8).

FIGURA 8  
Ejemplo de formato de visualización



BT.2021-08

### 2.6.2 Escalas de apreciación del método SDSCE

Se pide a los sujetos que comprueben las diferencias entre las dos secuencias y juzguen la fidelidad de la información vídeo moviendo el cursor de un dispositivo de voto manual. Cuando la fidelidad es perfecta, el cursor debe estar en la parte superior de la escala (codificada 100), cuando la fidelidad es nula, el cursor debe estar en la parte inferior de la escala (codificada 0). Puede utilizarse la escala de calidad continua normalizada de la UIT (véanse los Cuadros 1, 2 y 3).

Los sujetos conocen cuál es la referencia y se les pide que expongan su opinión, durante todo el tiempo que están observando las secuencias.

### 2.6.3 Datos de la nota de opinión del método SDSCE

Las siguientes definiciones se aplican a la descripción del protocolo de prueba:

- *Segmento vídeo*: un segmento vídeo corresponde a una secuencia vídeo.
- *Condición de prueba*: una condición de prueba puede ser un proceso vídeo específico, una condición de transmisión, o ambos. Cada segmento vídeo debe ser procesado de acuerdo con una condición de prueba por lo menos. Además, se deben añadir referencias a la lista de condición de prueba, con el fin de hacer pares de referencia/referencia que se han de evaluar.
- *Sesión*: una sesión es una serie de diferentes segmentos vídeo/condiciones de prueba pares sin separación y arregladas en un orden pseudoaleatorio. Cada sesión contiene por lo menos una vez todos los segmentos vídeo y condiciones de prueba pero no necesariamente todas las combinaciones de segmento vídeo/condición de prueba.
- *Presentación de prueba*: una presentación de prueba es una serie de sesiones para abarcar todas las combinaciones de segmento vídeo/condición de prueba. Todas las combinaciones de segmento vídeo/condición de prueba deben ser votadas por el mismo número de observadores (pero no necesariamente los mismos observadores).
- *Periodo de votación*: se pide a cada observador que vote continuamente durante una sesión.
- *Segmento de votos*: un segmento de 10 s de votos; todos los segmentos de votos se obtienen utilizando grupos de 20 votos consecutivos (equivalentes a 10 s) sin ninguna superposición.



Una vez efectuada la prueba, uno (o más) ficheros de datos están disponibles con todos los votos de las diferentes sesiones (S) que representan todo el material de voto de la presentación de prueba (TP). Se puede efectuar una primera comprobación de la validez de los datos verificando que cada par de segmentos vídeo/condiciones de prueba ha sido presentado y que un número equivalente de votos ha sido asignado a cada uno de ellos.

Los datos recopilados durante la ejecución de las pruebas realizadas de acuerdo con este protocolo pueden ser procesados de tres maneras diferentes:

- Análisis estadístico de cada segmento vídeo separado.
- Análisis estadístico de cada condición de prueba separada.
- Análisis estadístico global de todos los segmentos vídeo/condiciones de prueba pares.

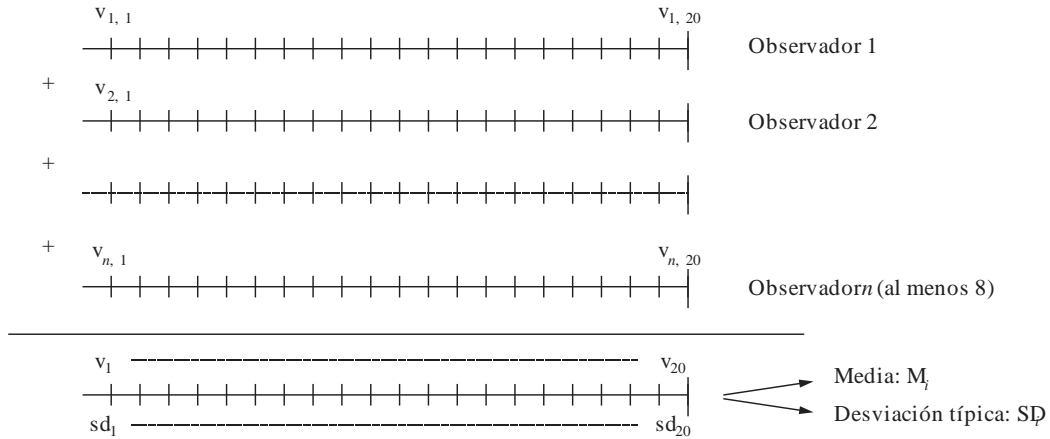
En cada caso se requiere un análisis de múltiples pasos:

- Se calculan los valores medios y las desviaciones típicas para cada voto por acumulación de los observadores.
- Se calcula el promedio y la desviación típica para cada segmento de votos, como se ilustra en la Fig. 9. Los resultados de este paso pueden ser representados en un diagrama temporal, como se muestra en la Fig. 10.
- Se analiza la distribución estadística de los valores medios calculados en el paso anterior (es decir, correspondiente a cada segmento de votos), y su frecuencia de aparición. Para evitar el efecto de novedad debido a las anteriores combinaciones de segmentos vídeo × condiciones de prueba, se rechazan los primeros 10 s de votos para cada muestra de segmento vídeo × condición de prueba.
- La característica global de molestia se calcula acumulando las frecuencias de ocurrencia. En este cálculo se deben tener en cuenta los intervalos de confianza, como se muestra en la Fig. 11. Una característica global de molestia corresponde a esta función de distribución estadística acumulada mostrando la relación entre los valores medios para cada segmento de votación y su frecuencia de aparición acumulada.

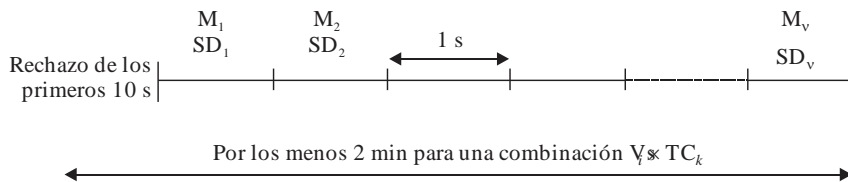
FIGURA 9

Procesamiento de datos

a) Cálculo de la nota media,  $V$ , y la desviación típica,  $SD$ , por instante de voto de los observadores para cada secuencia de votación de cada combinación segmento de vídeo (VS) × condición de prueba (TC)



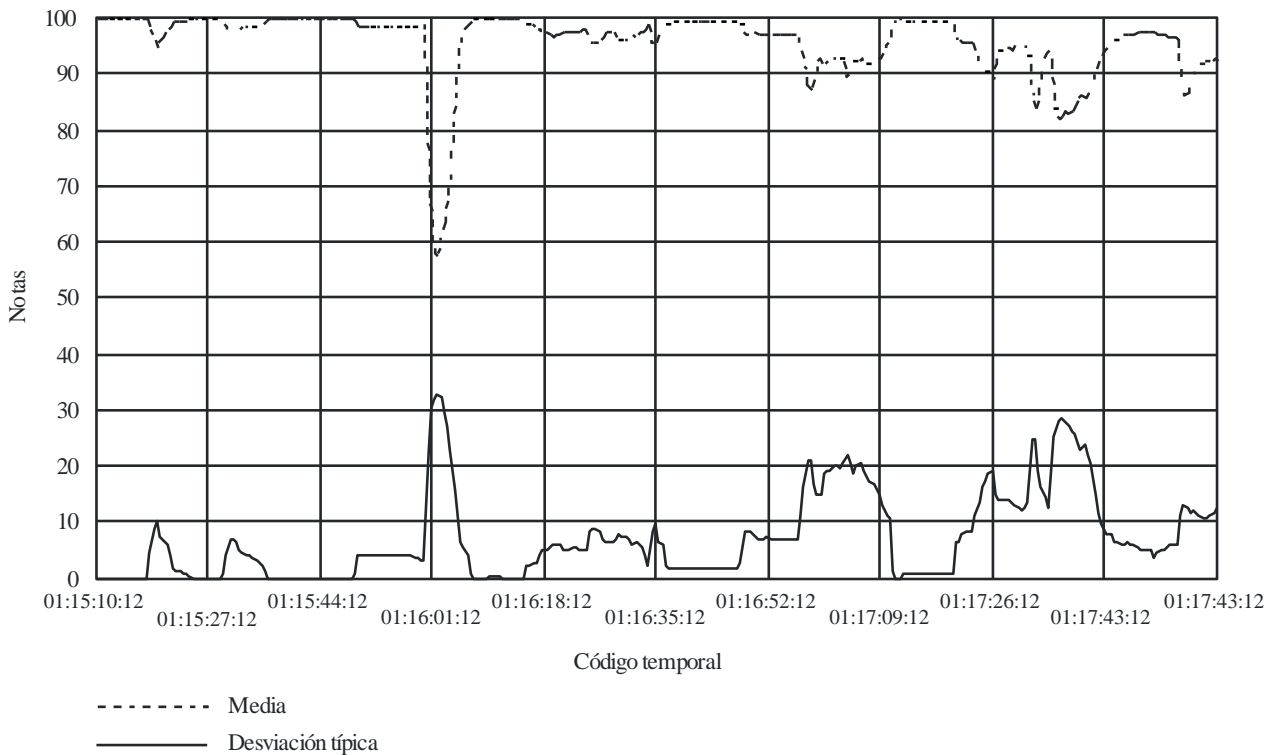
b) Cálculo de  $M$  y  $SD$  por secuencia de votación de 1 s para cada combinación VS/TC



BT.2021-09

FIGURA 10

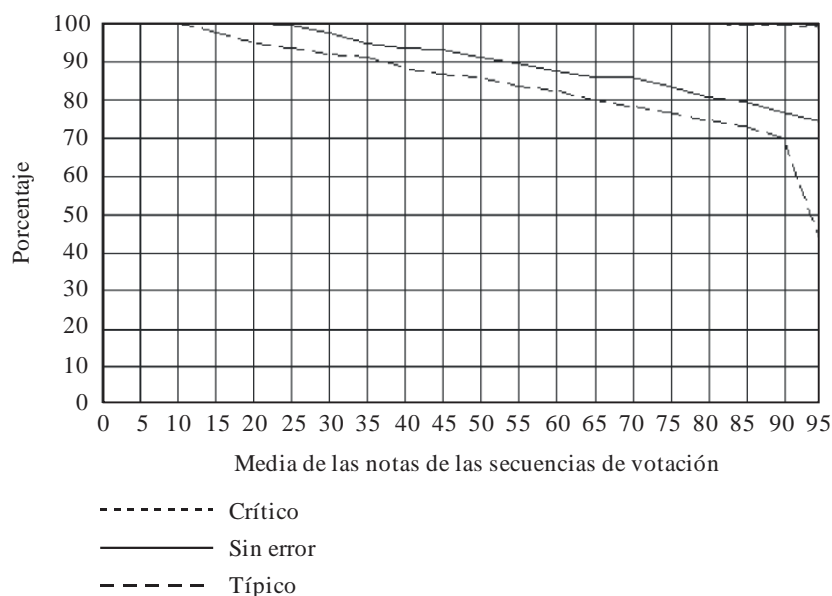
Diagrama temporal



BT.2021-10

FIGURA 11

Características globales de molestia calculadas a partir de las distribuciones estadísticas e incluido el intervalo de confianza




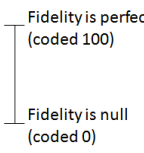
BT.2021-11

CUADRO 1

Métodos subjetivos de evaluación de la calidad de la imagen




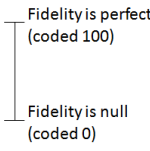
| Modo de presentación   | Duración de la secuencia | Escala binaria | Escala discreta  | Escala continua |
|--|--------------------------|----------------|--|-----------------|
| Métodos de estímulo único (SS) descritos en la Recomendación UIT-R BT.500, Anexo 1, § 6.1                              | ~10 s                    |                | 5 Excelente<br>4 Bueno<br>3 Aceptable<br>2 Mediocre<br>1 Malo  |                 |
| Método de escala de degradación con doble estímulo (DSIS) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 4      |                          |                | 5 Imperceptible<br>4 Perceptible, pero no molesta<br>3 Ligeramente molesta<br>2 Molesta<br>1 Muy molesta |                 |
| Método de escala de calidad continua de doble estímulo (DSCQS) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 5 | ~10 s                    |                |  |                 |

CUADRO 1 (*fin*)

| Modo de presentación   | Duración de la secuencia | Escala binaria | Escala discreta   | Escala continua  |
|--|--------------------------|----------------|---|--|
| Métodos de comparación de estímulos (SC) descrito en la Recomendación UIT R BT.500, Anexo 1, § 6.2                             | ~10 s                    | A vs. B        | -3 Mucho peor<br>-2 Peor<br>-1 Ligeramente peor<br>0 Igual<br>1 Ligeramente mejor<br>2 Mejor<br>3 Mucho mejor |  |
| Método de evaluación de calidad continua de estímulo único (SSCQE) descrito en la Recomendación UIT R BT.500, Anexo 1, § 6.3   | ~3-5 min                 |                |   |   |
| Método de doble estímulo simultáneo para evaluación continua (SDSCE) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 6.4 |                          |                |   |  |

CUADRO 2

## Métodos subjetivos de evaluación de la calidad de la profundidad

| Modo de presentación   | Duración de la secuencia | Escala binaria | Escala discreta   | Escala continua   |
|--|--------------------------|----------------|---|---|
| Métodos de estímulo único (SS) descritos en la Recomendación UIT-R BT.500, Anexo 1, § 6.1                                      | ~10 s                    |                | 5 Excelente<br>4 Bueno<br>3 Aceptable<br>2 Mediocre<br>1 Malo   |    |
| Método de escala de degradación con doble estímulo (DSIS) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 4              |                          |                | 5 Imperceptible<br>4 Perceptible, Pero no molesta<br>3 Ligeramente molesta<br>2 Molesta<br>1 Muy molesta      |   |
| Método de escala de calidad continua de doble estímulo (DSCQS) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 5         | ~10 s                    |                |   |   |
| Métodos de comparación de estímulos (SC) descrito en la Recomendación UIT R BT.500, Anexo 1, § 6.2                             | ~10 s                    | A vs. B        | -3 Mucho peor<br>-2 Peor<br>-1 Ligeramente peor<br>0 Igual<br>1 Ligeramente mejor<br>2 Mejor<br>3 Mucho mejor |   |
| Método de evaluación de calidad continua de estímulo único (SSCQE) descrito en la Recomendación UIT R BT.500, Anexo 1, § 6.3   | ~3-5 min                 |                |   |  |
| Método de doble estímulo simultáneo para evaluación continua (SDSCE) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 6.4 |                          |                |   |  |

CUADRO 3

## Métodos subjetivos de evaluación del confort visual

| Modo de presentación   | Duración de la secuencia | Escala binaria | Escala discreta   | Escala continua |
|--|--------------------------|----------------|---|-----------------|
| Métodos de estímulo único (SS) descritos en la Recomendación UIT-R BT.500, Anexo 1, § 6.1                                      | ~10 s                    |                | 5 Muy cómodo<br>4 Cómodo<br>3 Ligeramente incómodo<br>2 Incómodo<br>1 Muy incómodo                            |                 |
| Método de escala de degradación con doble estímulo (DSIS) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 4              |                          |                | 5 Imperceptible<br>4 Perceptible, pero no molesta<br>3 Ligeramente molesta<br>2 Molesta<br>1 Muy molesta      |                 |
| Método de escala de calidad continua de doble estímulo (DSCQS) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 5         | ~10 s                    |                |   |                 |
| Métodos de comparación de estímulos (SC) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 6.2                             | ~10 s                    | A vs. B        | -3 Mucho peor<br>-2 Peor<br>-1 Ligeramente peor<br>0 Igual<br>1 Ligeramente mejor<br>2 Mejor<br>3 Mucho mejor |                 |
| Método de evaluación de calidad continua de estímulo único (SSCQE) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 6.3   | ~3-5 min                 |                |   |                 |
| Método de doble estímulo simultáneo para evaluación continua (SDSCE) descrito en la Recomendación UIT-R BT.500, Anexo 1, § 6.4 |                          |                |   |                 |



### 3 Condiciones generales de la observación

Las condiciones de observación (incluida la luminancia de la pantalla, la iluminación de fondo, la distancia de observación etc.) deben ser consistentes con las utilizadas en 2D, tal como se describe en la Recomendación UIT-R BT.2022 (Documento 6/20), Condiciones generales de visualización para la evaluación subjetiva de la calidad de imágenes de televisión de definición normalizada y televisión de alta definición en monitores de pantalla plana. El motivo de dicha consistencia es doble. Por una parte, en la práctica los usuarios verán las imágenes de TV 3D en las mismas condiciones de representación y visualización que las imágenes de 2D. En segundo lugar, los avances en términos de calidad de funcionamiento de las tecnologías de vídeo TV 3D a menudo deberán medirse en relación con (es decir, «comparadas con») el avance de las normas de las tecnologías de vídeo de TVAD.

En la Recomendación UIT-R BT.2022 (Documento 6/20) se especifican dos posibles criterios para la selección de la distancia de visualización. Debe seleccionarse la distancia de visualización nominal (DVD, *design viewing distance*). La DVD de un sistema digital es la distancia a la que dos píxeles adyacentes presentan, para el ojo del observador, un ángulo subtendido de 1 arc-min.

Cuando se expresa el múltiplos de la altura de la imagen, la DVD para una resolución de imagen de  $1280 \times 720$  (Recomendaciones UIT-R BT.1543 y UIT-R BT.1847) es de 4,8H; y la de la familia de sistemas de resolución de imagen TVAD de  $1920 \times 1080$  (Recomendación UIT-R BT.709) es de 3,1H (para imágenes estáticas).

En el Cuadro 4 se presentan con fines ilustrativos de distancias de visualización nominales en metros para un conjunto representativo de tamaños en diagonal.

CUADRO 4

#### Distancia de visualización nominal en metros para distintos tamaños de TV en diagonal

| Tamaño de la diagonal<br>(pulgadas) | Sistema de imagen de $1920 \times 1080$        | Sistema de imagen de $1280 \times 720$         |
|-------------------------------------|--|--|
|                                     | Distancia de visualización nominal<br>(metros) | Distancia de visualización nominal<br>(metros) |
| 32                                  | 1,24   | 1,88   |
| 42                                  | 1,62   | 2,47   |
| 52                                  | 2,01   | 3,06   |
| 62                                  | 2,39   | 3,64   |
| 72                                  | 2,78   | 4,23   |
| 82                                  | 3,17   | 4,82   |
| 92                                  | 3,55   | 5,41   |
| 102                                 | 3,94   | 5,99   |

Cuando dos píxeles adyacentes subtienden un ángulo de 1 arc-min desde la perspectiva del ojo del observador, la menor disparidad angular (retinal) que puede ser representada por el sistema a la distancia de observación nominal (es decir, la resolución de profundidad del sistema) es igual a 1 arc-min (o su equivalente 60 arc-s). Las investigaciones han mostrado que casi el 97% de la población puede distinguir disparidades horizontales iguales o menores a 140 arc-s y que al menos el 80% puede detectar disparidades horizontales de 30 arc-s. Por tanto, la mayoría de los observadores no tendrán dificultades para resolver la disparidad más pequeña representable en los sistemas de vídeo 3D actuales a la distancia de visualización nominal.

## 4 Material de prueba

La selección del material de prueba debe estar asociada a la cuestión experimental que pretende abordarse en el correspondiente estudio. En general, el contenido de las secuencias de prueba (deporte, películas, etc.) y sus características espacio-temporales deben ser representativas de los programas distribuidos por el servicio analizado.

Además, el contenido de las secuencias estereoscópicas también debería ser, por lo general, de visualización cómoda. El confort de visualización de imágenes estereoscópicas depende de manera crítica de las disparidades de la imagen (paralaje) y de las condiciones de visualización. En consecuencia, deben tomarse las precauciones necesarias para garantizar que las disparidades no excedan los límites señalados en el punto siguiente, salvo que el estudio tenga por objetivo específicamente la medición del confort visual. Además, cuando sea posible, se medirán e informará de las estadísticas: media, desviación típica y distancia (mín/máx) de la distribución de la disparidad de la secuencia de prueba.

El paralaje, las inconsistencias entre las imágenes izquierda y derecha y la distribución y cambios del paralaje, pueden ser elementos a considerar en la selección de imágenes de prueba de fácil visualización estereoscópica. La relación entre las imágenes estereoscópicas 3D de fácil visualización y el paralaje, las inconsistencias entre las imágenes izquierda y derecha y la distribución y cambios del paralaje se describen en los puntos siguientes.

### 4.1 Límites del confort visual

Un paralaje o disparidad excesiva causa incomodidad visual, probablemente porque empeora el conflicto entre acomodación y vergencia (movimiento binocular en el cual ambos ojos se desplazan en direcciones opuestas). Por tanto, se ha sugerido que para minimizar el conflicto entre acomodación y vergencia, las disparidades en la imagen estereoscópica deben ser suficientemente pequeñas de forma que las profundidades percibidas de los objetos se encuentren en una «zona de confort». Se han propuesto varias formas de definir dichos límites. Un enfoque utiliza la medición del paralaje de la pantalla, expresado como un porcentaje del tamaño horizontal de la misma, para especificar los límites de la visualización cómoda. Se han sugerido valores del 1% para las disparidades cruzadas/negativas y del 2% para disparidades no cruzadas/positivas (para un valor total de aproximadamente un 3%). Según otro posible enfoque, la zona de confort está delimitada por la profundidad de campo del ojo. Para condiciones típicas de visualización de televisión (radiodifusión), los investigadores han supuesto una profundidad de campo de  $\pm 0,2D$  (dioptrías) y  $\pm 0,3D$  (dioptrías). Para un sistema de TVAD con una resolución de  $1920 \times 1080$  (Recomendación UIT-R BT.709) observado desde la distancia de visualización nominal de  $3,1H$ , dichos valores corresponden a aproximadamente  $\pm 2\%$  y  $\pm 3\%$  del paralaje de la pantalla. Finalmente, un tercer enfoque especifica los límites del confort en términos de disparidad retinal y fija dichos límites en  $\pm 1^\circ$  del ángulo visual de las disparidades positiva y negativa.

Es interesante señalar que los distintos enfoques convergen en los mismos límites de confort. Recuérdese que a la distancia de visualización nominal, dos píxeles adyacentes subtienden un ángulo de 1 arc-min desde el ojo de observador. Por tanto, 60 píxeles corresponden a un ángulo visual de  $1^\circ$ . Ello permite especificar fácilmente los límites de confort en términos de disparidad retinal (para un observador medio). Por ejemplo, para sistemas TVAV de resolución  $1920 \times 1080$  (Recomendación UIT-R BT.709), 1% ( $\sim 19.2$  píxeles) corresponde a aproximadamente 20 arc-min, 2% a  $\sim 40$  arc-min y 3% a  $\sim 60$  arc-min (equivalente a  $1^\circ$ ).

Debe observarse que aunque a la distancia de visualización nominal dos píxeles adyacentes siempre subtienden un ángulo de 1 arc-min, la separación física (por ejemplo, en mm) entre dichos píxeles aumenta para pantallas más grandes (el número de píxeles es el mismo, pero aumenta el tamaño de la pantalla). Por tanto, los límites más altos (por ejemplo,  $\pm 3\%$ ) pueden hacer que para las pantallas

de mayor tamaño la distancia física entre puntos correspondientes (es decir, el paralaje de dos vistas en mm) sea superior a la distancia entre pupilas de un observador medio (~63-65 mm). Ello puede dar lugar a una creciente incomodidad.

#### **4.2 Discrepancias entre las imágenes izquierda y derecha**

En sistemas 3D estereoscópicos, la imagen binocular se forma presentando la imagen izquierda y derecha a los respectivos ojos. Si existen discrepancias entre ambas imágenes, puede producirse estrés psicofísico y, en algunos casos, puede no ser posible la visualización 3D. Por ejemplo, cuando se filman y se muestran en pantalla programas estereoscópicos de TV 3D, pueden producirse entre las imágenes izquierda y derecha distorsiones geométricas tales como inconsistencia del tamaño, desplazamiento vertical o error de rotación. Es conveniente que las imágenes de prueba estén libres de dichas distorsiones geométricas. Para más información véase el § 3.2.1 del Anexo 4 al Informe UIT-R BT.2160-2.

Los elementos relativos a discrepancias entre las imágenes izquierda y derecha que deben considerarse cuando se seleccionen imágenes estereoscópicas TV 3D de prueba que sean de fácil visualización son los siguientes:

- discrepancia geométrica, incluyendo el tamaño, desplazamiento vertical y rotación;
- discrepancia en el brillo, incluyendo los niveles de blanco y de negro;
- diafonía.

#### **4.3 Distancia, distribución y cambio en el paralaje**

Las distribuciones del paralaje están correladas con el confort visual de las imágenes estereoscópicas.

La distribución de paralaje de las imágenes estereoscópicas es discontinuo durante los cuadros de cambio de escena. Cuando se produce un paralaje extremo o cambios bruscos en el paralaje se produce un efecto visual incómodo, por lo que es importante mantener un nivel adecuado de paralaje de las imágenes. Para más información véase el § 3.2.2 del Anexo 4 al Informe UIT-R BT.2160-2.

Por lo general, dado que los estudios que utilizan secuencias de prueba estereoscópicas pueden generar un cierto grado de incomodidad visual, es recomendable utilizar, siempre que sea posible, material de prueba cuyas disparidades no excedan los límites de confort, aunque ocasionalmente puedan permitirse situaciones puntuales en las que se superen dichos límites.

### **5 Equipamiento para la experimentación**

Los aparatos utilizados en los experimentos (servidor de vídeo, pantalla, etc.) deben permitir visualizar secuencias de prueba 3D con la máxima resolución, por ejemplo, utilizando un formato de empaquetamiento de trama HDMI. Ello ofrece una mayor flexibilidad respecto al conjunto de estudios que pueden realizarse.

Hasta la fecha, no se ha normalizado una pantalla TV 3D de referencia. En consecuencia, es previsible que la mayoría de los investigadores utilicen pantallas TV 3D de gran consumo. Dado que las características de dichas pantallas pueden variar de un fabricante a otro, se insta a los investigadores a que informen de los ajustes relevantes de la pantalla utilizados en cada estudio.

## **6 Observadores**

### **6.1 Tamaño de la muestra**

En general, es recomendable utilizar al menos 30 observadores. No obstante, el número real dependerá de los objetivos específicos de la investigación, y teniendo en cuenta que las consideraciones sobre el tamaño del grupo para estudios de TV 3D no difieren de los de 2D.

### **6.2 Examen de la visión de los observadores**

Los observadores deberán pasar un examen de agudeza visual, daltonismo y visión estereoscópica aplicando las pruebas clínicas vigentes, tales como las cartas de Snellen para la agudeza visual; los diagramas de Ishara, o equivalente, para el color, y la prueba de Randot, o equivalente, para la visión estereoscópica. Obsérvese que las pruebas de visión estereoscópica como la de Randot, la de la mosca o la de Frisby, miden, por lo general, disparidades retinales de entre 20 y 400 arc-s. Se insta a que los investigadores informen de las estadísticas relevantes de las capacidades estereoscópicas de los observadores participantes en un estudio. Si fuera necesario un análisis más detallado de dichas capacidades estereoscópicas, los investigadores pueden utilizar los materiales de prueba que se incluyen en el Apéndice 1.

## **7 Instrucciones a los observadores**

Las instrucciones a los observadores deben adaptarse a las dimensiones investigadas (por ejemplo, calidad de profundidad, confort, etc.). En particular, las directrices que es necesario dar a los observadores para la realización de estudios sobre imágenes en 3D deben ser más estrictas que las típicamente utilizadas en la evaluación de la calidad de imágenes en 2D, ya que los participantes pueden experimentar malestar visual. En general, los estudios sobre 3D exigen dar mayor grado de información a los participantes sobre las razones del estudio, así como los posibles efectos negativos que pudiera producir la exposición a los estímulos.

## **8 Duración de la sesión**

Si el material de visualización se considera confortable, la prueba podría tener una duración tan larga como la de los estudios de 2D (es decir, ~20-40 minutos con interrupciones entre ellas). Si el material de prueba tiene un paralaje excesivo, y por tanto es potencialmente molesto, debe limitarse la duración.

## **9 Utilización de material de vídeo de referencia**

Los investigadores pueden tener interés en incluir una secuencia de referencia en el conjunto de secuencias de prueba, en caso de estar disponible. La referencia es normalmente una versión de la secuencia de prueba sin procesamiento alguno (es decir, es la secuencia fuente original). Para los estudios estereoscópicos, la principal referencia es la secuencia estereoscópica original no procesada. Sin embargo, el plan del experimento puede incluir también como referencia la versión monoscópica (es decir, sólo una de las vistas de la secuencia fuente original); así por ejemplo, en los estudios del confort visual puede ser de utilidad como referencia el confort visual de la referencia monoscópica. La versión monoscópica de la referencia debe presentarse en modo 3D (por ejemplo, presentar la visión izquierda a ambos ojos con la misma configuración del hardware 3D que para la secuencia estereoscópica real). La inclusión de la referencia en el experimento proporciona dos ventajas importantes. En primer lugar, ofrece la oportunidad de medir la transparencia (también denominada

fidelidad) del algoritmo o tecnología investigada<sup>2</sup>. En segundo lugar, la inclusión de la referencia supone disponer de un anclaje de alta calidad que puede ayudar a estabilizar las apreciaciones<sup>3</sup>.

## 10 Variabilidad de las respuestas

Las apreciaciones de observadores en experimentos de evaluación subjetiva son, por lo general, bastante variables. Las diferencias entre observadores puede que simplemente reflejen las características de la población de referencia y, por tanto, pueden resolverse aumentando el tamaño de la muestra.

Sin embargo, parte de la variabilidad puede deberse a cambios en los patrones de respuesta de los observadores individuales durante el experimento. Dichos cambios implican una modificación de los criterios de evaluación debidos, por ejemplo, a la mayor práctica en la realización de la prueba, al aprendizaje de las características de los efectos molestos, etc.). Para minimizar los efectos negativos de dicha variabilidad, los investigadores deben establecer procedimientos de entrenamiento adecuados (tarea, nivel de degradación, etc.), utilizar la aleatorización múltiple (es decir, presentar las secuencias de prueba en órdenes aleatorios diferentes a los distintos observadores) y replicar las secuencias (lo que, además, permitiría medir los posibles cambios en los patrones de respuesta).

## 11 Criterios de rechazo de observadores

Los criterios para el rechazo de observadores para los métodos identificados en § 2 se describen en la Recomendación UIT-R BT.500.

## 12 Análisis estadístico

Los análisis estadísticos para la investigación de sistemas de imágenes en 3D son los mismos que para sistemas de imágenes en 2D.

# Apéndice 1

## Materiales de prueba para las pruebas de visión

### 1 Prueba de visión

En el Cuadro 5 se enumeran los diagramas utilizados en las pruebas de visión. Las 12 pruebas se han seleccionado de conformidad con la jerarquía del sistema de visión humano, desde el nivel más bajo al más alto. A continuación se describen ocho pruebas de visión (VT, *visión test*) principales, siendo las cuatro restantes utilizadas en pruebas clínicas. Los observadores deben tener un estereopsis

---

<sup>2</sup> La transparencia (fidelidad) es un concepto que describe la calidad de funcionamiento de un códec o de un sistema en relación con un sistema de transmisión ideal sin degradación. Es fácil ver que la transparencia puede medirse comparando las apreciaciones asignadas a la secuencia de referencia con las asignadas a la secuencia procesada con el algoritmo o la tecnología investigada.

<sup>3</sup> Se reconoce que la estabilidad de las apreciaciones en el espacio (es decir, realizadas en distintos laboratorios) y en el tiempo (es decir, realizadas en el mismo laboratorio pero en momentos diferentes) también puede mejorarse utilizando anclajes de baja calidad. No obstante, la UIT tiene planes para definir de forma inminente anclajes de baja calidad para la evaluación de tecnologías de imágenes estereoscópicas.

normal, es decir, deben pasar la prueba VT-04 de la estereopsis fina y la VT-07 para la estereopsis dinámica. Las restantes seis pruebas se utilizan para una caracterización más detallada. Los diagramas de prueba deben visualizarse a una distancia de tres veces la altura de la pantalla en las que se proyectan.

CUADRO 5

**Materiales de prueba estereoscópica para una prueba de visión**

| N.º | Elemento                   | Prueba de   | Contenido  |
|-----|----------------------------|---|--|
| 1   | Recepción simultánea       | Capacidad de percibir simultáneamente imágenes presentadas dicópticamente y en la posición correcta   | Se presenta la imagen de una jaula en un ojo y la de un león en el otro  |
| 2   | Fusión binocular           | Capacidad de percibir dos imágenes dicópticas en los ojos izquierdo y derecho como una sola imagen  | La imagen para un ojo tiene dos puntos y la del otro ojo tiene tres puntos, con un punto común   |
| 3   | Estereopsis gruesa         | Capacidad de percibir imágenes que se presentan de forma dicóptica con un cierto paralaje como una sola imagen con una profundidad considerable | Las imágenes para ambos ojos son una estereopareja de imágenes de una libélula con sus alas extendidas   |
| 4   | Estereopsis fina           | Capacidad de percibir imágenes que se presentan de forma dicóptica con un cierto paralaje como una sola imagen con una profundidad reducida     | Se presentan nueve romboides de prueba, cada uno de los cuales tiene cuatro círculos, y sólo uno de los círculos tiene un pequeño paralaje         |
| 5   | Límite de fusión cruzada   | Capacidad de percibir imágenes que se presentan de forma dicóptica con disparidades cruzadas como una sola imagen                               | Se presenta una estereopareja de barras cuyo paralaje varía a razón de 10'/s   |
| 6   | Límite de fusión sin cruce | Capacidad de percibir imágenes presentadas de forma dicóptica con disparidades no cruzadas como una sola imagen                                 | Se presenta una estereopareja de barras cuyo paralaje varía a razón de 11'/s   |
| 7   | Estereopsis dinámica       | Capacidad de percibir la profundidad en imágenes de un estereograma de puntos aleatorios en movimiento  | Estereograma de puntos aleatorios dinámicos  |
| 8   | Agudeza binocular          | Agudeza binocular, incluyendo cualquier asimetría de la agudeza monocular que pueda impedir una buena estereopsis                               | Caracteres E con diversas orientaciones y tamaños  |
| 9   | Estrabismo horizontal      | Desviación horizontal del ojo que el paciente no puede evitar   | Líneas verticales y horizontales   |
| 10  | Estrabismo vertical        | Desviación vertical del ojo que el paciente no puede evitar   | Líneas verticales y horizontales   |
| 11  | Aniseiconia                | Condición en la que la imagen ocular de un objeto visto por un ojo difiere en tamaño y forma respecto a como lo ve el otro ojo                  | La imagen izquierda consiste en caracteres «[o]» y la derecha consiste en caracteres «o]», donde el carácter «o]» tiene la misma posición en ambas |



## CUADRO 5 (fin)

**Materiales de prueba estereoscópica para una prueba de visión**

| N.º | Elemento   | Prueba de   | Contenido   |
|-----|------------|---|---|
| 12  | Cicloforia | Desviación de uno de los ojos alrededor del eje anteroposterior cuando se evita la fusión | La imagen izquierda consiste en la superficie de un reloj y la derecha en las manecillas del reloj marcando las seis en punto |

NOTA 1 – Estos materiales están en el formato 1125/60/I (véase la Recomendación UIT-R BT.709).

NOTA 2 – Estos materiales pueden obtenerse del Institute of Image Information and Television Engineers (ITE), 3-5-8 Shibakoen, Minato-ku, Tokio 105-0011, Japón. Tel.: 81-3-3432-4677, correo electrónico: [ite@ite.or.jp](mailto:ite@ite.or.jp).

Las imágenes en miniatura siguientes, a derecha e izquierda, se colocan una junto a otra con fines explicativos para una fusión sin cruces.

**1) VT-01: Percepción simultánea (prueba del león)**

Prueba la capacidad de percibir simultáneamente imágenes presentadas dicópticamente y en la posición correcta. Se presenta la imagen de una jaula en un ojo y la de un león en el otro, cuya posición se desplaza a razón de 12'/s. El tamaño de cada imagen se fija a 10°, de tal forma que los observadores pueden capturar las imágenes en sus paramáculas. Los observadores con una visión normal pueden ver al león dentro de la jaula durante un cierto tiempo del periodo de presentación.

FIGURA 12

Diagrama de prueba para VT-01



Imagen derecha

Imagen izquierda

BT.2021-12

**2) VT-02: Fusión binocular (prueba de los 4 puntos de Worth)**

Prueba de la capacidad de percibir dos imágenes dicópticas en los ojos izquierdo y derecho como una sola imagen. La imagen para un ojo tiene dos puntos y la del otro ojo tiene tres puntos, con un punto común. Los observadores con una visión normal ven cuatro puntos.

FIGURA 13

Diagrama de prueba para VT-02



Imagen derecha

Imagen izquierda

BT.2021-13

### 3) VT-03: Estereopsis gruesa (prueba de la libélula)

Prueba de la capacidad de percibir imágenes que se presentan de forma dicóptica con un cierto paralaje como una sola imagen con una profundidad considerable. Las imágenes para ambos ojos son una estereopareja de imágenes de una libélula con sus alas extendidas. Los observadores con una visión normal perciben las alas delante de la pantalla de visualización.

FIGURA 14

Diagrama de prueba para VT-03



Imagen derecha

Imagen izquierda

BT.2021-14

### 4) VT-04: Estereopsis de detalle (prueba del círculo)

Prueba la capacidad de percibir imágenes que se presentan de forma dicóptica con un cierto paralaje como una sola imagen con una profundidad reducida. Se presentan nueve romboides de prueba, cada uno de los cuales tiene cuatro círculos, y sólo uno de los círculos tiene un pequeño paralaje. Los observadores con visión normal pueden percibir el círculo con el pequeño paralaje delante de la pantalla de visualización. El Cuadro 6 muestra el número de prueba, las respuestas correctas y el ángulo de estereopsis a 3 H.

CUADRO 6

Respuestas correctas y paralaje

| Número de prueba | Repuesta correcta | Ángulo de estereopsis a 3 H («) |
|------------------|-------------------|---------------------------------|
| 1                | Abajo             | 480                             |
| 2                | Izquierda         | 420                             |
| 3                | Abajo             | 360                             |
| 4                | Arriba            | 300                             |
| 5                | Arriba            | 240                             |
| 6                | Izquierda         | 180                             |
| 7                | Derecha           | 120                             |
| 8                | Izquierda         | 60                              |
| 9                | –                 | 0                               |

FIGURA 15

Diagrama de prueba para VT-04

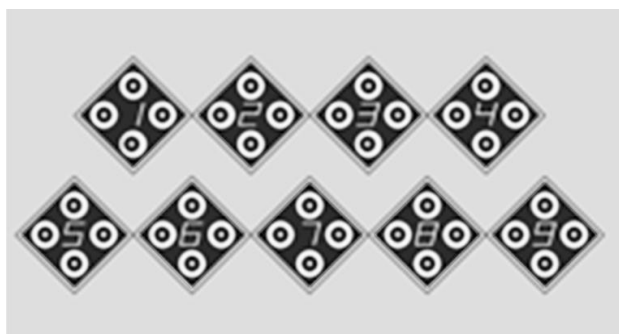


Imagen derecha

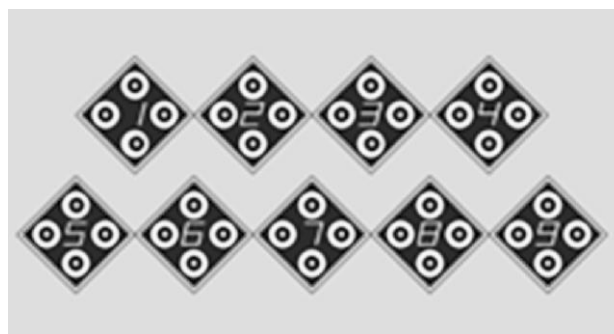


Imagen izquierda

BT.2021-15

5) **VT-05: Límite de fusión cruzada (prueba de la barra)**

Prueba la capacidad de percibir imágenes que se presentan de forma dicóptica con disparidades cruzadas como una sola imagen. Se presenta una estereopareja de barras cuyo paralaje varía a razón de 10'/s. Pueden medirse los límites de fusión de las series ascendentes y descendentes. Se pide a los observadores que informen del momento en que detectan la ruptura de fusión, es decir, tan pronto como perciben imágenes dobles en las series ascendentes, así como de la recuperación de la fusión, es decir, tan pronto como perciben las imágenes dicópticas como una imagen única en las series descendentes.

FIGURA 16

Diagrama de prueba para VT-05



Imagen derecha



Imagen izquierda

BT.202116

**6) VT-06: Límite de fusión sin cruce (prueba de la barra)**

Prueba la capacidad de percibir imágenes presentadas de forma dicóptica con disparidades no cruzadas como una sola imagen. Las imágenes que se presentan son las mismas que en el caso cruzado anterior, pero se invierten las imágenes derecha e izquierda.

FIGURA 17

Diagrama de prueba para VT-06



Imagen derecha



Imagen izquierda

BT.202117

**7) VT-07: Estereopsis dinámica (prueba del estereograma de puntos aleatorios dinámicos)**

Prueba la capacidad de percibir la profundidad en imágenes de un estereograma de puntos aleatorios en movimiento. Los observadores con visión normal pueden percibir una forma rectangular y un movimiento sinusoidal en profundidad en el estereograma de puntos aleatorios dinámicos.

FIGURA 18

Diagrama de prueba para VT-07

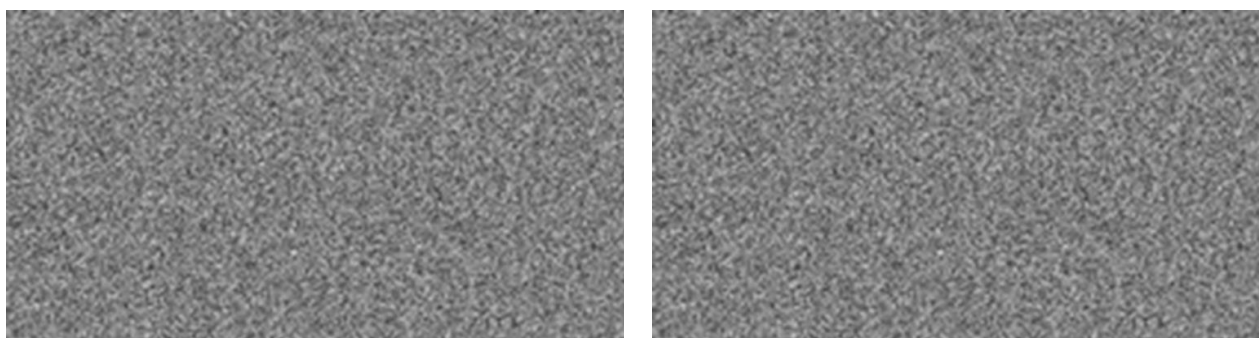


Imagen derecha

Imagen izquierda

BT.202148

### 8) VT-08: Agudeza binocular (prueba de agudeza)

Prueba la agudeza binocular con fusión binocular, incluyendo cualquier asimetría de la agudeza monocular que pueda impedir una estereopsis adecuada. Las imágenes tienen cuatro columnas y cinco líneas que consisten en caracteres E con diversas orientaciones y tamaños. Las dos columnas centrales pueden verse con ambos ojos; las dos columnas de la izquierda sólo pueden verse con el ojo izquierdo y las dos columnas de la derecha sólo pueden verse con el ojo derecho. Los observadores con una visión normal pueden decir correctamente la orientación de los caracteres E. Los tamaños de los caracteres se corresponden con agudezas de 1,0, 0,5, 0,33, 0,25 y 0,125 a 3 H.

FIGURA 19

Diagrama de prueba para VT-08



Imagen derecha

Imagen izquierda

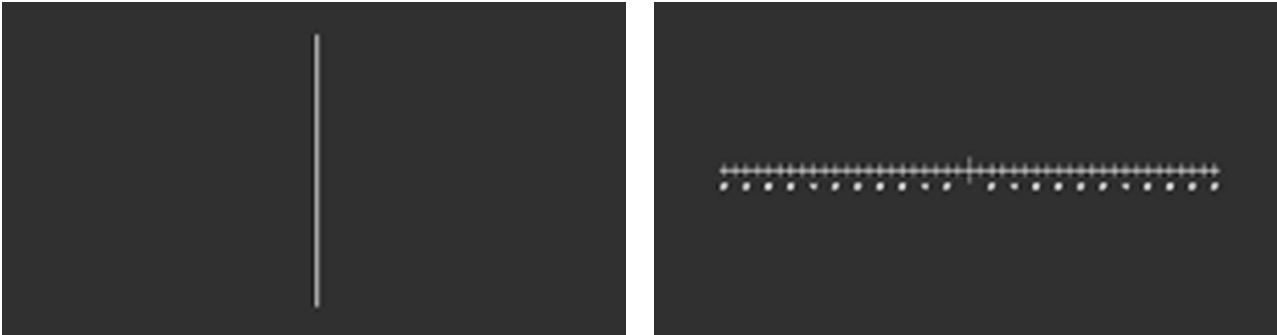
BT.202149

### 9 y 10) VT-09: Estrabismo horizontal (prueba horizontal de Maddox) y VT-10: Estrabismo vertical (prueba vertical de Maddox)

Estos diagramas miden la desviación horizontal y vertical del ojo. Los ejes visuales tienen una posición relativa entre ellos distinta, según las condiciones fisiológicas. Las imágenes constan de una línea vertical y otra horizontal. Los observadores con una visión normal perciben el punto de cruce de las líneas, aproximadamente, en el centro de las mismas. Los números junto a las marcas indican las dioptrías prismáticas para una distancia pupilar (PD) de 65 mm a 3,02 H.

FIGURA 20

Diagrama de prueba para VT-09



BT.202120

FIGURA 21

Diagrama de prueba para VT-10



BT.202121

### 11) VT-11: Aniseiconia («[ ]»prueba de caracteres)

Condición en la que la imagen ocular de un objeto visto por un ojo difiere en forma y tamaño de la vista por el otro ojo. La imagen izquierda consiste en caracteres «[o]» y la derecha consiste en caracteres «o]», donde el carácter «o» tiene la misma posición en ambas. Los observadores con una visión normal perciben los caracteres «[» y «]» con el mismo tamaño y la misma altura.

FIGURA 22

Diagrama de prueba para VT-11



BT.202122

**12) VT-12: Cyclophoria (prueba del reloj)**

Desviación de uno de los ojos alrededor del eje anteroposterior cuando se evita la fusión. La imagen izquierda consiste en la superficie de un reloj y la derecha en las manecillas del reloj marcando las seis en punto.

FIGURA 23

Diagrama de prueba para VT-12



BT.202123

---