Unión Internacional de Telecomunicaciones

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

P.910

(04/2008)

SERIE P: CALIDAD DE LA TRANSMISIÓN TELEFÓNICA, INSTALACIONES TELEFÓNICAS Y REDES DE LÍNEAS LOCALES

Calidad audiovisual en servicios multimedia

Métodos de evaluación subjetiva de la calidad vídeo para aplicaciones multimedios

Recomendación UIT-T P.910

Unión Internacional de Telecomunicaciones

RECOMENDACIONES UIT-T DE LA SERIE P

CALIDAD DE LA TRANSMISIÓN TELEFÓNICA, INSTALACIONES TELEFÓNICAS Y REDES DE LÍNEAS LOCALES

Vocabulario y efectos de los parámetros de transmisión sobre la opinión de los clientes	Serie	P.10
Características de los terminales vocales	Serie	P.30
		P.300
Sistemas de referencia	Serie	P.40
Aparatos para mediciones objetivas	Serie	P.50
		P.500
Medidas electroacústicas objetivas	Serie	P.60
Medidas relativas a la sonoridad vocal	Serie	P.70
Métodos de evaluación objetiva y subjetiva de la calidad vocal	Serie	P.80
Métodos de evaluación objetiva y subjetiva de la calidad vocal y de video	Serie	P.800
Calidad audiovisual en servicios multimedia	**Serie**	**P.900**
Aspectos de calidad de transmisión y de calidad de servicio en los puntos extremos de redes de protocolo Internet	Serie	P.1000
Comunicaciones implicando vehículos	Serie	P.1100
Modelos y herramientas para la evaluación de la calidad de los medios secuenciados	Serie	P.1200
Evaluación de las telerreuniones	Serie	P.1300
Directrices para el análisis, la evaluación y la información estadísticos de las mediciones de la calidad	Serie	P.1400
Métodos de evaluación objetiva y subjetiva de la calidad de servicios distintos a los servicios de voz o de vídeo	Serie	P.1500

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

Recomendación UIT-T P.910

Métodos de evaluación subjetiva de la calidad vídeo para aplicaciones multimedios

Resumen

La Recomendación UIT-T P.910 describe métodos de evaluación subjetiva no interactivos para determinar la calidad general de vídeo unidireccional para aplicaciones multimedios como la videoconferencia, el almacenamiento con recuperación, la telemedicina, etc. Los métodos se pueden utilizar con diversos fines, como selección de algoritmos, clasificación de la calidad de funcionamiento de sistemas audiovisuales y evaluación del nivel de calidad durante una conexión vídeo, sin que esta relación de objetivos sea exhaustiva. Esta Recomendación también describe las características de la secuencia fuente que se utilizarán, como duración, clase de contenido, número de secuencias, etc.

Historia

Edición	Recomendación	Aprobación	Comisión de Estudio	ID único*
1.0	ITU-T P.910	1996-08-30	12	11.1002/1000/3641
2.0	ITU-T P.910	1999-09-30	12	11.1002/1000/4751
3.0	ITU-T P.910	2008-04-06	9	11.1002/1000/9317

* Para acceder a la Recomendación, sírvase digitar el URL http://handle.itu.int/ en el campo de dirección del navegador, seguido por el identificador único de la Recomendación. Por ejemplo, http://handle.itu.int/11.1002/1000/11830-en.

PREFACIO

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

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

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución 1 de la AMNT.

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

NOTA

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

La observancia de esta Recomendación es voluntaria. Ahora bien, la Recomendación puede contener ciertas disposiciones obligatorias (para asegurar, por ejemplo, la aplicabilidad o la interoperabilidad), por lo que la observancia se consigue con el cumplimiento exacto y puntual de todas las disposiciones obligatorias. La obligatoriedad de un elemento preceptivo o requisito se expresa mediante las frases "tener que, haber de, hay que + infinitivo" o el verbo principal en tiempo futuro simple de mandato, en modo afirmativo o negativo. El hecho de que se utilice esta formulación no entraña que la observancia se imponga a ninguna de las partes.

PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente Recomendación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT no ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB en la dirección http://www.itu.int/ITU-T/ipr/.

ÍNDICE

Página

Recomendación UIT-T P.910

Métodos de evaluación subjetiva de la calidad vídeo para aplicaciones multimedios

1 Alcance

Esta Recomendación tiene por objeto definir métodos de evaluación subjetiva no interactivos para determinar la calidad de las imágenes de vídeo digital codificadas a las velocidades binarias especificadas en las clases para TV3, MM4, MM5 y MM6 especificadas en el Cuadro D.2 para aplicaciones tales como la videotelefonía, la videoconferencia y el almacenamiento con recuperación. Los métodos se pueden utilizar, por ello, con diversos fines, por ejemplo, la selección de algoritmos, la clasificación de la calidad de funcionamiento de sistemas de vídeo, y la evaluación del nivel de calidad durante una conexión vídeo, sin que esta relación de objetivos sea exhaustiva.

2 Referencias

Las siguientes Recomendaciones del UIT-T y otras referencias contienen disposiciones que, mediante su referencia en este texto, constituyen disposiciones de la presente Recomendación. Al efectuar esta publicación, estaban en vigor las ediciones indicadas. Todas las Recomendaciones y otras referencias son objeto de revisiones por lo que se preconiza que los usuarios de esta Recomendación investiguen la posibilidad de aplicar las ediciones más recientes de las Recomendaciones y otras referencias citadas a continuación. Se publica periódicamente una lista de las Recomendaciones UIT-T actualmente vigentes. En esta Recomendación, la referencia a un documento, en tanto que autónomo, no le otorga el rango de una Recomendación.

[UIT-T J.61] Recomendación UIT-T J.61 (1988), *Calidad de transmisión de los circuitos de televisión diseñados para ser utilizados en conexiones internacionales*.

[UIT-T P.800] Recomendación UIT-T P.800 (1996), *Métodos de determinación subjetiva de la calidad de transmisión*.

[UIT-T P.930] Recomendación UIT-T P.930 (1996), *Principios de un sistema de degradaciones de referencia para vídeo*.

[UIT-R BT.500-9] Recomendación UIT-R BT.500-9 (1998), *Método para la evaluación subjetiva de la calidad de las imágenes de televisión*.

[UIT-R BT.601-4] Recomendación UIT-R BT.601-4 (1994), *Parámetros de codificación de televisión digital para estudios*.

[UIT-R BT.814-1] Recomendación UIT-R BT.814-1 (1994), *Especificaciones y procedimientos de ajuste para establecer el brillo y el contraste en las pantallas*.

[IEC/TR 60268-13] Publicación 60268-13 de la CEI/TR (1998), *Sound system equipment – Part 13: Listening tests on loudspeakers* <http://webstore.iec.ch/webstore/webstore.nsf/artnum/022890>.

3 Términos y definiciones

En esta Recomendación se definen los términos siguientes:

3.1 gamma: Parámetro que describe la discriminación entre los pasos del nivel de gris de una presentación visual. La relación entre la luminancia de la pantalla y la tensión de la señal de entrada no es lineal, con la tensión elevada a un exponente gamma. Para compensar esta no linealidad,

se aplica generalmente a la cámara un factor de corrección que es una función inversa de gamma. El coeficiente gamma repercute también en la reproducción de los colores.

3.2 pruebas de optimización: Pruebas subjetivas que generalmente se llevan a cabo durante la elaboración o la normalización de un nuevo algoritmo o sistema. El objetivo de estas pruebas es evaluar la calidad de funcionamiento de nuevos elementos a fin de optimizar los algoritmos o los sistemas sometidos a estudio.

3.3 pruebas de calificación: Pruebas subjetivas que generalmente se efectúan para comparar la calidad de funcionamiento de sistemas o equipos comerciales. Estas pruebas se deben de llevar a cabo en condiciones de prueba que sean lo más representativas posible de las condiciones reales de utilización.

3.4 información de percepción espacial (SI, *spatial perceptual information*): Medida que generalmente indica el grado de detalle espacial de una imagen. Usualmente es mayor en escenas espacialmente más complejas. Esta información no constituye una medida de la entropía ni está asociada con la información definida en la teoría de la comunicación. Véase en la cláusula 5.3.1 la ecuación de SI.

3.5 información de percepción temporal (TI, *temporal perceptual information*): Medida que generalmente indica la cantidad de cambios temporales de una secuencia de vídeo. Usualmente es mayor en secuencias de alta velocidad. Esta información no constituye una medida de la entropía ni está asociada con la información definida en la teoría de la comunicación. Véase en la cláusula 5.3.2 la ecuación de TI.

3.6 transparencia (fidelidad): Concepto que describe la calidad de funcionamiento de un códec o un sistema en relación con un sistema de transmisión ideal sin ninguna degradación.

Se pueden definir dos tipos de transparencia:

El primer tipo describe el grado de ajuste de la señal procesada a la señal de entrada, o señal ideal, utilizando un criterio matemático. Si no existen diferencias, el sistema es totalmente transparente. El segundo tipo describe el grado de ajuste de la señal procesada a la señal de entrada, o señal ideal, para un observador humano. Si no se perciben diferencias bajo ninguna condición experimental el sistema se considera perceptivamente transparente. Se utiliza el término transparente sin referencia explícita a criterio alguno en el caso de sistemas que sean perceptivamente transparentes.

3.7 replicación; reiteración: Repetición de la misma condición de circuito (con el mismo material original) para el mismo sujeto.

3.8 fiabilidad de una prueba subjetiva:

a) Fiabilidad intraindividuo ("en el mismo sujeto"), se refiere a la concordancia entre calificaciones repetidas de un determinado sujeto con la misma condición de prueba.

b) Fiabilidad entre individuos ("entre sujetos"), se refiere a la concordancia entre calificaciones de diferentes sujetos con la misma condición de prueba.

3.9 validez de una prueba subjetiva: Concordancia entre el valor medio de las calificaciones obtenidas en una prueba y el valor verdadero que se pretende medir con la prueba.

3.10 condiciones de referencia: Condiciones simuladas añadidas a las condiciones de prueba para afianzar las evaluaciones procedentes de diferentes experimentos.

3.11 referencia explícita (referencia fuente): Condición utilizada por los evaluadores como referencia para expresar su opinión, cuando se emplea el método DCR. Esta referencia se visualiza primero dentro de cada par de secuencias. Por lo general, el formato de la referencia explícita es el utilizado a la entrada de los códecs sometidos a prueba (por ejemplo: [UIT-R BT.601-4], CIF, QCIF, SIF, etc.). En el cuerpo de esta Recomendación, se omitirán los términos "explícito" y "origen" siempre que en el contexto esté claro el significado de "referencia".

3.12 referencia implícita: Condición utilizada por los evaluadores como referencia para expresar su opinión sobre el material de prueba, cuando se emplea el método ACR. Si el experimentador sugiere la referencia implícita, debe ser perfectamente conocida por todos los evaluadores (por ejemplo, sistemas de televisión convencionales, realidad).

4 Abreviaturas

En esta Recomendación se utilizan las siguientes siglas:

ACR Índices por categorías absolutas (*absolute category rating*)

ACR-HR Índice por categoría absoluta con referencia escondida (*absolute category rating with hidden reference*)

CCD Dispositivo de acoplamiento de cargas (*charge coupled device*)

CI Intervalo de confianza (*confidence interval*)

CIF Formato intermedio común (*common intermediate format*)

NOTA – Un formato de imagen definido en [b-UIT-T H.261] para videotelefonía: 352 líneas × 288 píxels.

CRT Tubo de rayos catódicos (*cathode ray tube*)

DCR Índices por categorías de degradación (*degradation category rating*)

DV Usuario diferencial (*differential viewer*)

%GOB Porcentaje de bueno o mejor (*percent of good or better*) (proporción de votos bueno y excelente)

LCD Visualización de cristal líquido (*liquid crystal display*)

MOS Nota media de opinión (*mean opinion score*)

PC Comparación por pares (*pair comparison*)

%POW Porcentaje de mediocre o peor (*percent of poor or worse*) (proporción de votos mediocre y malo)

PVS Secuencia de vídeo procesada (*processed video sequence*)

QCIF Un cuarto de CIF (*quarter CIF*)

NOTA – Un formato de imagen definido en [b-UIT-T H.261] para videoteléfono: 176 líneas × 144 píxels.

S/N Relación señal/ruido (*signal-to-noise ratio*)

SI Información espacial (*spatial information*)

SIF Formato intermedio normalizado (*standard intermediate format*)

NOTA – Un formato de imagen definido en [b-ISO/CEI 11172] (MPEG-1): 352 líneas × 288 píxels × 25 tramas/s y 352 líneas x 240 píxels × 30 tramas/s.

SP Presentación simultánea (*simultaneous presentation*)

std Desviación típica (*standard deviation*)

TI Información temporal (*temporal information*)

VTR Magnetoscopio (*video tape recorder*)

5 Señal fuente

Para controlar las características de la señal fuente, las secuencias de prueba se deben definir de acuerdo con el objetivo de la prueba y registrar en un sistema de almacenamiento digital. Cuando el experimentador está interesado en comparar los resultados de diferentes laboratorios es necesario utilizar un conjunto común de secuencias fuente para eliminar otro origen de discrepancias.

5.1 Entorno de grabación

La fuente o las fuentes luminosas (lámparas o tubos fluorescentes) pueden situarse por encima de la cámara o a un lado de la misma. Cuando se coloquen las luces, se ha de tener en cuenta que la iluminación aérea o de techo es más propia de oficinas y se debe utilizar con escenas que representan el ambiente empresarial. Las luces de estudio y otras fuentes luminosas no típicas deben evitarse.

Las condiciones de iluminación de la sala en el campo de visión pueden variar de 100 lux a unos 10 000 lux para uso en interiores. Se debe tener en cuenta la variación (frecuencia de la corriente alterna) de la luz (iluminación fluorescente) ya que puede provocar una fluctuación en la secuencia de vídeo grabada.

Las condiciones de iluminación, colores de los muros, reflectancia de la superficie, etc., se deben controlar y notificar detalladamente.

5.2 Sistema de grabación

5.2.1 Cámara

Las secuencias de imagen se deben registrar con una cámara CCD de alta calidad.

La relación señal/ruido de la señal vídeo de entrada puede afectar considerablemente la calidad de funcionamiento del códec.

Para definir la entrada vídeo se deben especificar los siguientes puntos:

- gama dinámica de las señales YUV;
- factor de corrección gamma (debe ser 0,45);
- anchura de banda/pendientes de los filtros;
- sensibilidad de la cámara en condiciones de muy baja iluminación y características del control automático de ganancia (AGC, *automatic gain control*), si se utiliza.

La relación S/N ponderada se debe medir de acuerdo con la Parte C, cláusula 3.2.1 de [UIT-T J.61] y debe ser superior a 45 dB (valor cuadrático medio).

La inestabilidad o las fluctuaciones de las señales de reloj podrían causar efectos de ruido. El dispositivo de temporización de la cámara requiere una estabilidad mínima de 0,5 ppm.

Se pueden utilizar sistemas de longitud focal fija o variable. Para terminales de pupitre se considera razonable una profundidad focal de 30 a 120 cm, mientras que para sistemas de múltiples usuarios sería más apropiada una profundidad focal de 50 cm a infinito. Para soportar la variación de iluminancia en la sala de grabación se deben utilizar filtros de iris ajustable o de densidad neutra. La cámara debe tener un sistema automático de equilibrado de los blancos para que pueda llevarse a cabo la adaptación a la temperatura de color de la fuente luminosa. La corrección de la temperatura de los blancos puede variar de 2 700° K (utilización en interiores con lámpara eléctrica) a 6 500° K (temperatura de luz diurna con cielo nublado).

5.2.2 Señal vídeo y formato de almacenamiento

Las señales fuente de vídeo suministradas por la cámara deben ser muestreadas de acuerdo con la Parte A de [UIT-R BT.601-4]. Para evitar la distorsión de la señal fuente, se debe almacenar en formato digital, por ejemplo, en computadora o formato de cinta D1 4:2:2.

5.3 Características de la escena

La elección de las escenas de prueba es un asunto importante. En particular, la información de percepción espacial y la información temporal de las escenas constituyen parámetros críticos. Estos parámetros desempeñan un papel crucial en la determinación del grado de compresión vídeo que es posible y, por consiguiente, del nivel de degradación que se produce cuando la escena se transmite por un canal de servicio de transmisión digital de velocidad fija. Se deben seleccionar escenas de vídeo de prueba adecuadas y pertinentes de modo que su información espacial y temporal sea coherente con los servicios vídeo que se supone que debe proporcionar el canal de servicio de transmisión digital. El conjunto de escenas de prueba debe abarcar la gama completa de información espacial y temporal de interés para los usuarios de los dispositivos sometidos a prueba.

En el Anexo A y en los Apéndices I y II figuran detalles sobre la caracterización de las secuencias de prueba y ejemplos de escenas de prueba adecuadas.

El número de secuencias se debe definir de acuerdo con el diseño experimental. Para evitar el cansancio de los observadores y lograr un mínimo de fiabilidad en los resultados, se deben elegir para las secuencias al menos cuatro tipos de escenas diferentes (es decir, distintos temas).

Las siguientes subcláusulas presentan métodos para cuantificar la información espacial y temporal de las escenas de prueba. Estos métodos de evaluación de la información espacial y temporal de las escenas de prueba son aplicables a las pruebas de calidad vídeo actuales y futuras. La ubicación de la escena de vídeo dentro de la matriz espacial-temporal es importante ya que la calidad de una escena de vídeo transmitida (especialmente después de pasar a través de un códec de baja velocidad binaria) depende a menudo en gran medida de dicha ubicación. Las medidas de información espacial y temporal que aquí se presentan pueden utilizarse para asegurar una cobertura apropiada del plano espacial-temporal.

Las medidas de información espacial y temporal que figuran a continuación son de valor único para cada trama en una secuencia de prueba completa. Esto da lugar a una serie temporal de valores que por lo general tendrán un cierto grado de variación. Las medidas de información de percepción que figuran más adelante eliminan esta variabilidad con una función de máximo (valor máximo para la secuencia). La propia variabilidad se puede estudiar convenientemente, por ejemplo con muestras de información espacial-temporal trama por trama. La utilización de distribuciones de información a lo largo de una secuencia de prueba permite también una mejor evaluación de las escenas con cortes de escena.

5.3.1 Medición de la información de percepción espacial

La información de percepción espacial (SI) se basa en el filtro Sobel. Primero se filtra cada trama vídeo (plano de luminancia) en un momento n (F_n) con el filtro Sobel [$Sobel(F_n)$]. A continuación se calcula la desviación típica de los píxels (std_{space}) de cada trama filtrada con el filtro Sobel. Esta operación se repite para cada trama de la secuencia de vídeo y da por resultado una serie temporal de información espacial de la escena. Se elige el valor máximo de la serie temporal ($\max_{time}$) como representación del contenido de información espacial de la escena. Este proceso se puede representar en forma de ecuación como sigue:

$$SI = \max_{time}\left\{std_{space}\left[Sobel\left(F_n\right)\right]\right\}$$

5.3.2 Medición de la información de percepción temporal

La información de percepción temporal (TI) se basa en la característica de diferencia de movimiento, $M_n(i, j)$, que es la diferencia entre los valores de píxels (del plano de luminancia) en la misma ubicación en el espacio pero en momentos o tramas sucesivos. $M_n(i, j)$ se define, como una función del tiempo (n) de la siguiente manera:

$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j)$$

donde $F_n(i, j)$ es el píxel en la i-ésima fila y j-ésima columna de la n-ésima trama en el tiempo.

La medida de la información temporal (TI) se calcula como el valor máximo en el tiempo ($\max_{time}$) de la desviación típica en el espacio (std_{space}) de $M_n(i, j)$ en todas las i y j.

$$TI = \max_{time}\{std_{space}[M_n(i, j)]\}$$

Un mayor movimiento en las tramas adyacentes dará lugar a valores de TI más elevados.

NOTA – Para escenas que contengan cortes, se pueden dar dos valores: uno en el que el corte de la escena se incluye en la medición de la información temporal, y otro en el que se excluye de la medición.

6 Métodos de prueba y diseño experimental

La medición de la calidad de imágenes percibida requiere la utilización de métodos de escala subjetiva. La condición para que esas mediciones sean significativas es que exista una relación entre las características físicas del "estímulo", en este caso la secuencia de vídeo presentada a los sujetos en una prueba, y la magnitud y naturaleza de la sensación causada por el estímulo.

Se han validado diversos métodos experimentales con distintos objetivos. Aquí se recomiendan tres métodos para aplicaciones que utilizan conexiones a las velocidades binarias especificadas en las clases para TV3, MM4, MM5 y MM6, especificadas en el Cuadro D.2. En los Apéndices III y IV se describen otros métodos de prueba.

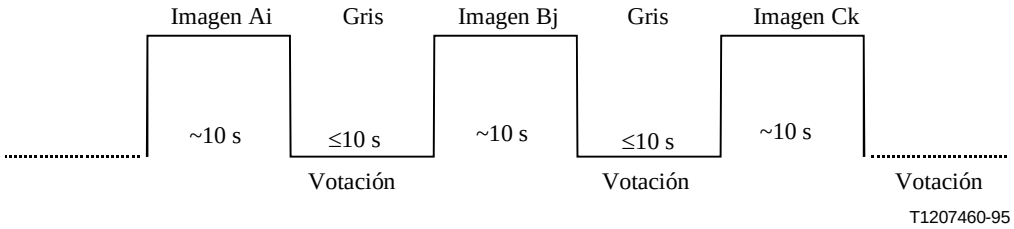
La elección final de uno de estos métodos para una aplicación determinada depende de varios factores, tales como el contexto, la finalidad y dónde se debe llevar a cabo la prueba en el proceso de desarrollo.

6.1 Índices por categorías absolutas (ACR, *absolute category rating*)

El método de los índices por categorías absolutas es un juicio de categorías en el que las secuencias de prueba se presentan una por vez y se califican independientemente en una escala de categorías. (Este método se denomina también método de evaluación con un solo estímulo.)

El método especifica que después de cada presentación se invite a los sujetos a evaluar la calidad de la secuencia mostrada.

En la Figura 1 se ilustra el diagrama de tiempos de la presentación del estímulo. Si se utiliza un tiempo de votación constante (por ejemplo, diversos usuarios simultáneamente desde una cinta), entonces el tiempo de votación debe ser igual o inferior a 10 s. El tiempo de presentación se puede reducir o aumentar de acuerdo con el contenido del material de prueba.



Ai Secuencia A en condición de prueba i
Bj Secuencia B en condición de prueba j
Ck Secuencia C en condiciones de prueba k

Figura 1 – Presentación del estímulo en el método ACR

Para evaluar la calidad global se debe utilizar la siguiente escala de cinco niveles:

5 Excelente

4 Bueno

3 Aceptable

2 Mediocre

1 Mala

Si se requiere una evaluación más discriminativa se puede utilizar una escala de nueve niveles. En el Anexo B figuran ejemplos de escalas apropiadas numéricas o continuas. En el Anexo B se dan también ejemplos de dimensiones de evaluación distintas de la calidad global. Dichas dimensiones pueden ser de utilidad para obtener más información sobre diferentes factores de calidad de percepción cuando el índice de calidad global es casi igual para determinados sistemas sometidos a prueba, aunque los sistemas se perciban claramente como diferentes.

Para el método ACR, se obtiene el número necesario de reiteraciones repitiendo las mismas condiciones de prueba en diferentes momentos de la prueba.

6.2 Índice por categoría absoluta con referencia escondida (ACR-HR)

El método de los índices por categorías absolutas es un juicio de categorías en el que las secuencias de prueba se presentan una por vez y se califican independientemente en una escala de categorías. El presente procedimiento de prueba debe incorporar una versión de referencia de cada secuencia de prueba mostrada como cualquier otro estímulo de prueba. Eso se llama "condición de referencia oculta". Durante el análisis de datos se computará una puntuación de calidad diferente (DMOS) entre cada secuencia de prueba y su correspondiente referencia (oculta). Este procedimiento se conoce como "referencia oculta".

El método especifica que después de cada presentación se invite a los sujetos a evaluar la calidad de la secuencia mostrada.

En la Figura 1 se ilustra el diagrama de tiempos de la presentación del estímulo. Si se utiliza un tiempo de votación constante (por ejemplo, diversos usuarios simultáneamente desde una cinta), el tiempo de votación deberá ser igual o inferior a 10 s. El tiempo de presentación puede reducirse o aumentarse en función del contenido del material de prueba.

Para evaluar la calidad global se debe utilizar la siguiente escala de cinco niveles:

5 Excelente

4 Bueno

3 Aceptable

2 Mediocre

1 Mala

Las puntuaciones del usuario diferencial (DV) se calculan por persona y por secuencia de vídeo procesada (PVS). La referencia oculta (REF) adecuada se utiliza para calcular el DV utilizando la fórmula siguiente:

$$DV(PVS) = V(PVS) – V(REF) + 5$$

donde V es la puntuación ACR del usuario. Con esta fórmula, un DV de 5 indica una calidad "excelente" y un DV de 1 indica una calidad "mala". Cualquier valor DV superior a 5 (es decir, cuando se considera que la secuencia de proceso tiene más calidad que su secuencia de referencia oculta asociada) se considerará válido generalmente. En cambio, puede aplicarse una función de aplastamiento de 2 puntos para evitar que esas puntuaciones de usuarios ACR-HR influyan indebidamente en la puntuación de opinión general:

$$DV\_aplastado = (7*DV)/(2+DV) \text{ cuando } DV > 5.$$

Si se requiere una evaluación más discriminativa se puede utilizar una escala de nueve niveles. En el Anexo B figuran ejemplos de escalas apropiadas numéricas o continuas. En el Anexo B se dan también ejemplos de dimensiones de evaluación distintas de la calidad global. Dichas dimensiones pueden ser de utilidad para obtener más información sobre diferentes factores de calidad de percepción cuando el índice de calidad global es casi igual para determinados sistemas sometidos a prueba, aunque los sistemas se perciban claramente como diferentes.

Para el método ACR, se obtiene el número necesario de reiteraciones repitiendo las mismas condiciones de prueba en diferentes momentos de la prueba.

El método ACR-HR solo debería utilizarse con un vídeo de referencia que un experto en la materia considere de calidad "buena" o "excelente" en la escala anterior de cinco niveles.

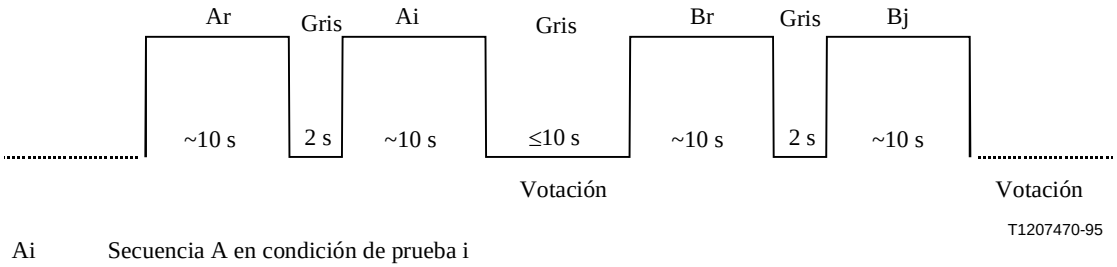
El método ACR-HR puede que no sea apto para analizar degradaciones inusuales que se producen en el primer y último segundo de la secuencia de vídeo. La falta de familiaridad del usuario con la secuencia de vídeo de referencia puede ocasionar que se pase por alto una degradación clara (por ejemplo, si una secuencia se detiene justo antes del final, el usuario puede que no sepa determinar si se trata de una pausa intencionada o de un error de red).

6.3 Índices por categorías de degradación (DCR, *degradation category rating*)

Los índices por categorías de degradación implican la presentación de las secuencias de prueba por pares: el primer estímulo presentado en cada par es siempre la referencia fuente, mientras que el segundo estímulo es la misma fuente presentada a través de uno de los sistemas sometido a prueba (este método se denomina también método de escala de degradación con doble estímulo).

Cuando se utilicen formatos de imagen reducidos (por ejemplo, CIF, QCIF, SIF), podría ser conveniente visualizar la referencia y la secuencia de prueba simultáneamente en el mismo monitor. En el Anexo C se dan directrices sobre este procedimiento de presentación.

En la Figura 2 se ilustra el diagrama de tiempos de la presentación del estímulo. Si se utiliza un tiempo de votación constante (por ejemplo, diversos usuarios simultáneamente desde una cinta), entonces el tiempo de votación debe ser igual o inferior a 10 s. El tiempo de presentación se puede reducir o aumentar de acuerdo con el contenido del material de prueba.



Ai Secuencia A en condición de prueba i
Ar, Br Secuencia A y secuencia B, respectivamente, en el formato fuente de referencia
Bj Secuencia B en condición de prueba j

Figura 2 – Presentación del estímulo en el método DCR

En este caso se invita a los sujetos a evaluar la degradación del segundo estímulo en relación con la referencia.

Para evaluar la degradación se debe utilizar la siguiente escala de cinco niveles:

5 Imperceptible

4 Perceptible, pero no molesta

3 Ligeramente molesta

2 Molesta

1 Muy molesta

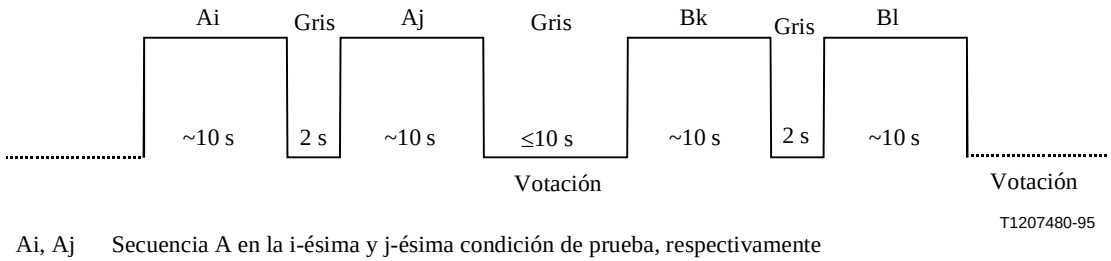
Para el método DCR, se obtiene el número necesario de reiteraciones repitiendo las mismas condiciones de prueba en diferentes momentos de la prueba.

6.4 Método de comparación por pares (PC, *pair comparison*)

El método de comparación por pares implica la presentación de las secuencias de prueba por pares, es decir, que en la misma secuencia se presenta primero a través de un sistema sometido a prueba y a continuación a través de otro sistema.

Con los sistemas sometidos a prueba (A, B, C, etc.) se forman generalmente todas las combinaciones n(n–1) posibles: AB, BA, CA, etc. De esta manera, todos los pares de secuencias se deben visualizar en los dos órdenes posibles (por ejemplo, AB, BA). Después de cada par se hace una apreciación sobre qué elemento del par se prefiere en el contexto del escenario de prueba.

En la Figura 3 se ilustra el diagrama de tiempos de la presentación del estímulo. Si se utiliza un tiempo de votación constante (por ejemplo, diversos usuarios simultáneamente desde una cinta), entonces el tiempo de votación debe ser igual o inferior a 10 s. El tiempo de presentación debe ser de unos 10 s y se puede reducir o aumentar de acuerdo con el contenido del material de prueba.



Ai, Aj Secuencia A en la i-ésima y j-ésima condición de prueba, respectivamente
Bk, Bl Secuencia B en la k-ésima y l-ésima condición de prueba, respectivamente

Figura 3 – Presentación del estímulo en el método PC

Cuando se utilicen resoluciones reducidas (por ejemplo, CIF, QCIF, SIF), podría ser conveniente visualizar cada par de secuencias simultáneamente en el mismo monitor. En el Anexo C se dan directrices sobre este procedimiento de presentación.

Por lo general, en el método PC no es necesario tener en cuenta el número de reiteraciones, ya que el propio método implica la presentación repetida de las mismas condiciones, aunque en pares diferentes.

Una variación del método PC utiliza una escala por categorías para apreciar en mayor grado las diferencias entre los pares de secuencias. Véanse [UIT-R BT.500-9] y [UIT-T P.800].

6.5 Comparación de los métodos

Una cuestión importante al elegir el método de prueba es la diferencia fundamental entre métodos que utilizan referencias explícitas (por ejemplo, DCR) y métodos que no utilizan ninguna referencia explícita (por ejemplo, ACR, ACR-HR y PC). Con esta segunda clase de métodos no se prueba la transparencia o la fidelidad.

Cuando se pruebe la fidelidad de transmisión con respecto a la señal fuente, se debe utilizar el método DCR. Con frecuencia es este un factor importante en la evaluación de sistemas de alta calidad. Durante mucho tiempo, el método DCR ha sido un método clave, especificado en [UIT-R BT.500-9], para la evaluación de imágenes de televisión cuya calidad típica representa los niveles sumamente altos de videotelefonía y videoconferencia. Se pueden utilizar también otros métodos para evaluar sistemas de alta calidad. Los comentarios específicos de la escala DCR (imperceptible/perceptible) son valiosos cuando la detección de la degradación por parte del observador es un factor importante.

Así pues, cuando sea importante comprobar la fidelidad con respecto a la señal fuente, se deberá utilizar el método DCR.

Este método se debe aplicar también para la evaluación de sistemas de alta calidad en el contexto de comunicaciones multimedios. La discriminación de la degradación imperceptible/perceptible en la escala DCR admite esto y también la comparación con la calidad de referencia.

El método ACR es sencillo y rápido de implementar y la presentación de los estímulos es similar a la del uso común de los sistemas. Este es el método adecuado, por ello, para las pruebas de calificación.

ACR-HR tiene todas las ventajas de ACR con respecto a presentación y velocidad. La ventaja principal de ACR-HR frente a ACR es que el impacto perceptual del vídeo de referencia puede eliminarse de la puntuación subjetiva. Eso reduce el impacto de la desviación de escena (por ejemplo, que a los usuarios les guste o no un vídeo de referencia), de la calidad del vídeo de referencia (por ejemplo, pequeñas diferencias en la calidad de la cámara) y del monitor (por ejemplo, calidad profesional frente a grado de consumidor) ante las puntuaciones finales. ACR-HR está bien adaptado para grandes experimentos si todos los vídeos de referencia son al menos de "buena" calidad. Ahora bien, ACR-HR puede que no sea sensible a algunas degradaciones que son fácilmente detectables con métodos diferenciales directos (por ejemplo, DCR). Por ejemplo, puede que ACR-HR no detecte una disminución sistemática en la ganancia de color (por ejemplo, colores sin brillo).

El mérito principal del método PC es su alto poder discriminatorio, que tiene un valor particular cuando algunos de los elementos en prueba tienen casi la misma calidad.

Cuando en una misma prueba se ha de evaluar un gran número de elementos, el procedimiento basado en el método PC suele ser demasiado largo. En este caso se puede efectuar primero una prueba con el método ACR o DCR, con un número limitado de observadores, seguida de una prueba con el método PC aplicado sólo a aquellos elementos que hayan recibido aproximadamente la misma evaluación.

6.6 Condiciones de referencia

Los resultados de las evaluaciones de calidad dependen a menudo no sólo de la calidad vídeo real, sino también de otros factores tales como la gama de calidades total de las condiciones de prueba, la experiencia y expectativas de los evaluadores, etc. Para controlar algunos de estos efectos, se pueden añadir condiciones de prueba ficticias y utilizarlas como referencias.

Una descripción de las condiciones de referencia y los procedimientos para generarlas figuran en [UIT-T P.930]. La introducción de la señal fuente como condición de referencia en una prueba PC se recomienda especialmente cuando las degradaciones introducidas por los elementos sometidos a prueba son pequeñas.

El nivel de calidad de las condiciones de referencia debe comprender como mínimo la gama de calidades de los elementos sometidos a prueba.

6.7 Diseño experimental

Se pueden utilizar diferentes diseños experimentales, tales como diseños aleatorizados completos, diseños de cuadrados latinos, grecolatinos y Youden, diseños de bloques repetidos, etc. [b-Kirk] cuya selección vendrá determinada por el objetivo del experimento.

Se deja a criterio del experimentador la selección de un método de diseño con el que satisfacer los objetivos de coste y precisión específicos. El diseño puede depender también de las condiciones que son de interés particular en una determinada prueba.

Se recomienda incluir al menos dos, y si fuera posible tres o cuatro, reiteraciones (es decir, repeticiones de condiciones idénticas) en el experimento. Existen varios motivos para la utilización de reiteraciones, siendo el más importante la posibilidad de medir la "variación en el mismo sujeto" empleando datos repetidos. Para comprobar la fiabilidad de un sujeto se puede utilizar el mismo orden de presentación en condiciones idénticas. Si se utiliza un orden de presentación diferente, la variación resultante en los datos del experimento está compuesta por el efecto de orden y por la variación en el mismo sujeto.

Las reiteraciones permiten calcular la fiabilidad individual de cada sujeto y, si fuera necesario, descartar resultados no fiables de algunos sujetos. Una estimación de la desviación típica en el mismo sujeto y entre sujetos es además un requisito previo para efectuar un análisis correcto de la varianza y para generalizar resultados a una población más amplia. Además, los efectos del aprendizaje en una prueba quedan compensados en cierta medida.

Se obtiene una mejora adicional en el tratamiento de los efectos del aprendizaje mediante una sesión de instrucción en la que se presentan al menos cinco condiciones al comienzo de cada sesión de prueba. Estas condiciones deben elegirse de modo que sean representativas de las presentaciones que se van a mostrar más tarde durante la sesión. Las presentaciones preliminares no se han de tener en cuenta en el análisis estadístico de los resultados de la prueba.

7 Procedimientos de evaluación

El Cuadro 1 enumera las condiciones típicas de observación utilizadas para la evaluación de la calidad de vídeo. Se deben especificar los conjuntos de parámetros reales utilizados en la evaluación. Para comparar los resultados de las pruebas, se deben fijar e igualar todas las condiciones de observación en todos los laboratorios para el mismo tipo de pruebas.

Tanto el tamaño como el tipo de monitor utilizado debe ser adecuado para la aplicación que se investiga. Cuando se presentan secuencias a través de un sistema basado en PC se deben especificar las características de visualización, por ejemplo densidad de puntos y de la pantalla, tipo de tarjeta de visualización de vídeo utilizada, etc.

En lo que respecta al formato de la visualización, es preferible utilizar toda la pantalla para la visualización de las secuencias. No obstante, cuando por alguna razón, las secuencias tengan que ser visualizadas en una ventana de la pantalla, el color de fondo de la pantalla debe ser gris en un 50% que se corresponde con Y = U = V = 128 (U y V sin signo).

7.1 Condiciones de observación

La prueba se debe efectuar con las siguientes condiciones de observación.

Cuadro 1 – Condiciones de observación

Parámetro	Valores
Distancia de observación (nota 1)	1-8 H (nota 2)
Valor de cresta de luminancia de la pantalla	100-200 cd/m (nota 2)
Relación entre la luminancia de la pantalla inactiva y la luminancia de cresta	≤ 0,05
Relación entre la luminancia de la pantalla, cuando se presenta únicamente el nivel de negro en una habitación totalmente oscura, y la correspondiente al blanco de cresta	≤ 0,1
Relación entre la luminancia de fondo detrás del monitor de imagen y la luminancia de cresta de la imagen (nota 3)	≤ 0,2
Cromaticidad del fondo (nota 4)	D_{65}
Iluminación de fondo de la habitación (nota 1)	≤ 20 lux
NOTA 1– Para una determinada altura de pantalla, es posible que la distancia de observación preferida por los participantes aumente cuando la calidad visual se degrade. A este respecto, para pruebas de calificación debe predeterminarse la distancia de observación preferida. La distancia de observación depende generalmente de las aplicaciones NOTA 2 – H indica la altura de la imagen. La distancia de observación se debe definir teniendo en cuenta no sólo el tamaño de la pantalla sino también el tipo de pantalla, el tipo de aplicación y el objetivo del experimento. NOTA 3 – Este valor indica un ajuste que permite la máxima detectabilidad de las distorsiones, para algunas aplicaciones se admiten valores menores o mayores o vienen determinados por la aplicación. NOTA 4 – Para monitores de PC la cromaticidad del fondo puede adaptarse a la cromaticidad del monitor.	

7.2 Sistema de procesamiento y reproducción

Existen dos métodos para la obtención de imágenes de prueba procedentes de grabaciones fuente:

a) mediante la transmisión o reproducción de grabaciones de vídeo en tiempo real a través de los sistemas sometidos a prueba, mientras los sujetos observan y responden;

b) procesando fuera de línea las grabaciones fuente a través del dispositivo sometido a prueba y grabando la salida para producir un nuevo conjunto de grabaciones.

En el segundo caso, se debe utilizar un VTR digital para reducir al mínimo las degradaciones que se pueden producir en el proceso de grabación. En cualquier caso, teniendo en cuenta que las degradaciones introducidas por los esquemas de codificación de baja velocidad binaria son generalmente más evidentes que las degradaciones introducidas por la modulación, se pueden utilizar VTR de calidad profesional tales como D2, MII y BetacamSP.

Se puede emplear un CRT, LCD, plasma, proyector, u otro tipo de monitor teniendo en cuenta el tipo de aplicación y el objeto del experimento. El tamaño y el tipo del monitor utilizado deben ser los apropiados para la aplicación que se investigue.

Los monitores se deben ajustar de acuerdo con los procedimientos definidos en [UIT-R BT.814-1].

7.3 Observadores

El número posible de sujetos en una prueba de observación (así como en pruebas de utilidad en terminales o servicios) varía de 4 a 40. Cuatro es el mínimo absoluto por razones estadísticas, mientras que difícilmente se obtengan mayores ventajas con más de 40 sujetos.

El número real para una determinada prueba debe establecerse, en la práctica, en función de la validez requerida y de la necesidad de efectuar una generalización de una muestra a una población mayor.

Por lo general en el experimento deben participar 15 observadores como mínimo. No deben intervenir directamente en evaluaciones de calidad de imagen como parte de su trabajo habitual y no han de ser evaluadores experimentados.

No obstante, en las primeras etapas del desarrollo de sistemas de comunicación vídeo y en experimentos piloto llevados a cabo antes de una prueba más amplia, los pequeños grupos de expertos (4-8) u otros sujetos críticos pueden proporcionar resultados indicativos.

Habitualmente, antes de una sesión, se deberá examinar a los observadores para determinar su agudeza visual normal o corregida a normal y su visión normal de los colores. Con respecto a la agudeza, no deben hacerse errores en la línea 20/30 de un diagrama de ojo normalizado [b-Snellen]. El diagrama se debe graduar para la distancia de observación de la prueba y la prueba de agudeza se debe efectuar en el mismo sitio en que se observarán las imágenes de vídeo (es decir apoyando el diagrama de ojo contra el monitor), y estando los sujetos sentados. Por lo que se refiere al color, no se deben perder más de 2 placas [b-Beck] de un total de 12.

7.4 Instrucciones a los observadores y sesión de instrucción

Antes de comenzar el experimento, se debe explicar a los sujetos el escenario de la aplicación prevista del sistema sometido a prueba. Además, se dará por escrito una descripción del tipo de evaluación, la escala de opinión y la presentación de los estímulos. La gama y tipo de degradaciones se debe presentar en pruebas preliminares, que pueden contener secuencias de vídeo distintas de las utilizadas en las pruebas reales.

No se debe tomar la deducción de que la peor calidad vista en la sesión de instrucción corresponda necesariamente al grado subjetivo más bajo de la escala.

Las preguntas referentes al procedimiento o al significado de las instrucciones se deben responder con cuidado para no influir en las apreciaciones y sólo antes de comenzar la sesión.

En el Apéndice III se propone un texto posible de las instrucciones que se han de dar a los evaluadores.

8 Análisis estadístico y notificación de los resultados

Los resultados se deben notificar junto con los detalles de la disposición experimental. Para cada combinación de variables de la prueba, se debe indicar el valor medio y la desviación típica de la distribución estadística de los grados de evaluación.

Se deberá calcular la fiabilidad de los sujetos a partir de los datos y se deberá informar sobre el método utilizado para evaluar dicha fiabilidad. En [UIT-R BT.500-9] y [CEI/TR 60268-13] se dan algunos criterios relativos a la fiabilidad subjetiva.

Resulta instructivo analizar la distribución acumulada de las notas de opinión. Puesto que las distribuciones acumuladas no dependen de la linealidad, podrían ser particularmente útiles para datos cuya linealidad sea dudosa, como los que se obtienen empleando los métodos ACR y DCR, junto con escalas por categorías sin gradaciones (por ejemplo, apreciación por categorías).

Los datos se pueden organizar, por ejemplo, como se muestra en el Cuadro 2 para el método ACR.

Cuadro 2 – Cuadro informativo con distribución acumulada de notas para el método ACR

Condición	Total de votos	Excelente	Buena	Aceptable	Mediocre	Mala	MOS	CI	Std	%GOB	%POW

Condición: Clasificación que indica una combinación de variables de prueba.
Total de votos: Cantidad de votos recogidos para esa condición.
Excelente, aceptable ... mala: Ocurrencia de cada voto.

Para evaluar la importancia de los parámetros de prueba se deben utilizar técnicas de análisis de varianza clásicas. Si la prueba está orientada a la evaluación de la calidad vídeo como función de un parámetro, quizás convenga utilizar técnicas de ajuste de curvas para la interpretación de los datos.

En el caso de comparaciones por pares, el método de cálculo de la posición de cada estímulo en una escala de intervalos, cuando la diferencia entre los estímulos se corresponde con la diferencia en preferencias, se describe en el *Manual sobre telefonometría*, sección 2.6.2C de [b-ITU-T Handbook].

Anexo A

Detalles relacionados con la caracterización de las secuencias de prueba

(Este anexo forma parte integrante de la presente Recomendación.)

A.1 Filtro Sobel

El filtro Sobel se realiza convolucionando dos núcleos 3 × 3 en la trama vídeo y extrayendo la raíz cuadrada de la suma de los cuadrados de los resultados de esas convoluciones.

Para y = Sobel(x), sea $x(i, j)$ el píxel de la imagen de entrada en la i-ésima fila y j-ésima columna. $Gv(i, j)$ será el resultado de la primera convolución y viene dado por:

$$\begin{aligned} Gv(i, j) \quad &= -1\times x(i-1, j-1) - 2\times x(i-1, j) - 1\times x(i-1, j+1) + \\ &+ 0\times x(i, j-1) + 0\times x(i, j) + 0\times x(i, j+1) + \\ &+ 1\times x(i+1, j-1) + 2\times x(i+1, j) + 1\times x(i+1, j+1) \end{aligned}$$

De manera similar, $Gh(i, j)$ será el resultado de la segunda convolución y viene dado por:

$$\begin{aligned} Gh(i, j) = \quad &-1\times x(i-1, j-1) + 0\times x(i-1, j) + 1\times x(i-1, j+1) + \\ &- 2\times x(i, j-1) + 0\times x(i, j) + 2\times x(i, j+1) + \\ &- 1\times x(i+1, j-1) + 0\times x(i+1, j) + 1\times x(i+1, j+1) \end{aligned}$$

La salida de la imagen filtrada con el filtro Sobel en la i-ésima fila y j-ésima columna viene dada por tanto, por la siguiente expresión:

$$y(i, j) = \sqrt{[Gv(i, j)]^2 + [Gh(i, j)]^2}$$

Los cálculos se efectúan para todas las $2 \le i \le N - 1$ y $2 \le j \le M - 1$, donde N es el número de filas y M es el número de columnas.

Se recomienda que los cálculos se efectúen sobre una subimagen de la trama vídeo para evitar efectos de borde no deseados y porque los bordes de extremo de una trama vídeo son invisibles normalmente para el observador de un CRT. Esto puede llevarse a cabo utilizando una subimagen adecuada, como se ilustra en el ejemplo de la Figura A.1 para formatos [UIT-R BT.601-4] de 625 y 525 líneas.

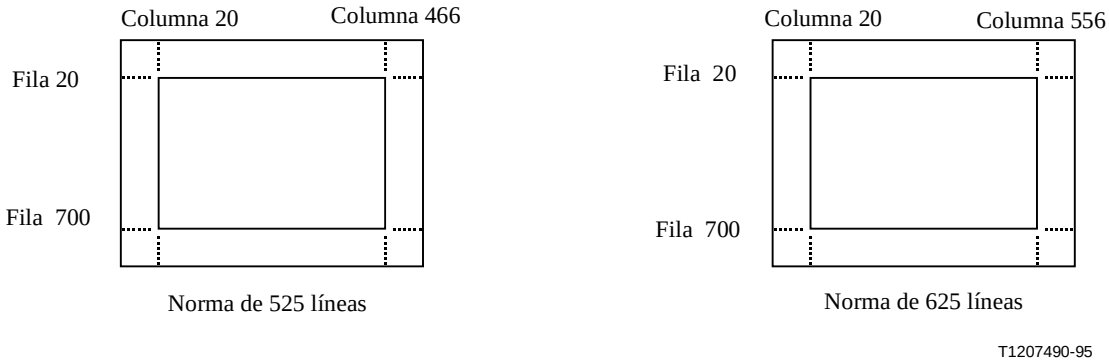


Figura A.1 – Subimágenes que deben utilizarse para calcular SI y TI para formatos [UIT-R BT.601-4] de 525 y 625 líneas

En [b-Gonzalez] figura más información sobre el filtro Sobel.

A.2 Cómo utilizar SI y TI para la selección de secuencias de prueba

Cuando se seleccionan secuencias de prueba, puede ser conveniente comparar la información espacial y la información temporal relativas halladas en las diversas secuencias disponibles. Por lo general, la dificultad de compresión está directamente relacionada con la información espacial y temporal de una secuencia.

Si en una determinada prueba se ha de utilizar un pequeño número de secuencias de prueba, quizá sea importante elegir secuencias que abarquen una gran porción del plano de información espacial-temporal (véase la Figura A.2). Cuando en una prueba se tengan que utilizar cuatro secuencias de prueba, podría ser conveniente elegir una secuencia de cada uno de los cuatro cuadrantes del plano de información espacial-temporal.

De manera alternativa, si se intenta escoger secuencias de prueba que sean equivalentes en dificultad de codificación, sería conveniente elegir secuencias que tuvieran valores de SI y TI similares.

A.3 Ejemplos

La Figura A.2 muestra las cantidades relativas de información espacial y temporal de algunas escenas de prueba representativas y cómo pueden ubicarse en un plano de información espacial-temporal.

A lo largo del eje TI = 0 (parte inferior del gráfico) se encuentran las escenas fijas y las de movimiento muy limitado (tales como l, f y a). Cerca de la parte superior del gráfico se encuentran escenas de mucho movimiento (tales como p, q e i). A lo largo del eje SI = 0 (borde izquierdo del gráfico) se encuentran escenas con mínimo detalle espacial (tales como l, k, x, u y f). Cerca del borde derecho del gráfico se encuentran escenas con máximo detalle espacial (tales como h y s). Los valores de SI y TI se obtuvieron utilizando las ecuaciones anteriores y vídeo muestreado espacialmente de acuerdo con las especificaciones [UIT-R BT.601-4]. En el Cuadro A.1 figura la lista de escenas de prueba del ejemplo por categoría del contenido de las mismas.

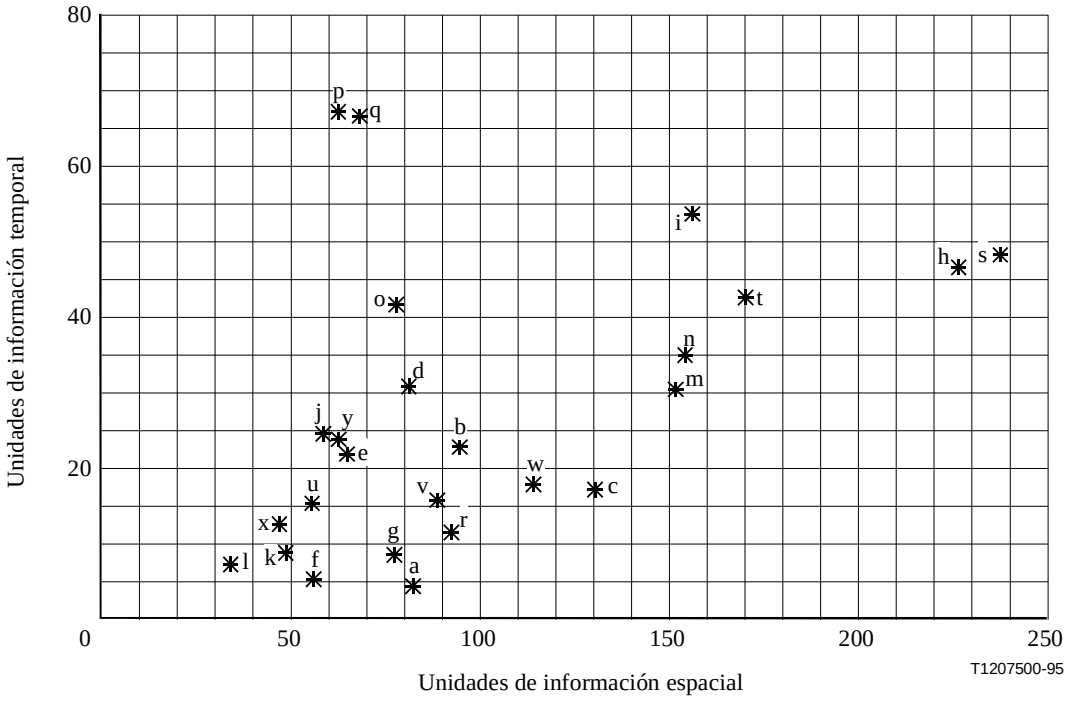


Figura A.2 – Gráfico espacial-temporal de un ejemplo de conjunto de escenas de prueba

Cuadro A.1 – Categorías del contenido de las escenas

Categoría	Descripción	Nombre y letra de la escena
A	Una persona, principalmente cabeza y hombros, detalle y movimiento limitados	vtc1nw(f), susie(j), disguy(k), disgal(1)
B	Una persona con gráficos y/o más detalle	vtc2mp(a), vtc2zm(b), boblec(e), smity1(m), smity2(n), vowels(w), inspec(x)
C	Más de una persona	3inrow(d), 5row1(g), intros(o), 3twos(p), 2wbord(q), split6(r)
D	Gráficos con indicación	washdc(c), cirkit(s), rodmap(t), filter(u), ysmite(v),
E	Gran movimiento del objeto y/o la cámara (ejemplos de televisión de radiodifusión)	flogar(h), ftball(i), fedas(y)

Anexo B

Escalas de evaluación adicionales

(Este anexo forma parte integrante de la presente Recomendación.)

B.1 Escalas de índices

A menudo es necesario utilizar escalas de índices con más de 5 grados, en particular para evaluar códecs vídeo de baja velocidad binaria. Una escala adecuada a tal fin es la escala de 9 grados, en la que se utilizan las cinco categorías de calidad definidas verbalmente que se recomiendan en la cláusula 6.1 como etiquetas de calificación de los grados alternos de la escala, como se muestra en la Figura B.1.

9	Excelente
8	
7	Buena
6	
5	Aceptable
4	
3	Mediocre
2	
1	Mala

Figura B.1 – Escala de calidad numérica de 9 grados

En la Figura B.2 se muestra una ampliación de esta escala, en la que los puntos extremos han sido definidos verbalmente como "puntos de anclaje" que no se emplean para la clasificación. En esta definición verbal se utiliza alguna forma de referencia (por ejemplo, en la Figura B.2, se utiliza como referencia el original). La referencia puede ser explícita o implícita y ha de ser ilustrada claramente durante la fase de instrucción. Véanse también [CEI/TR 60268-13] y la sección 2.6, escala a), de [b-UIT-T Manual].

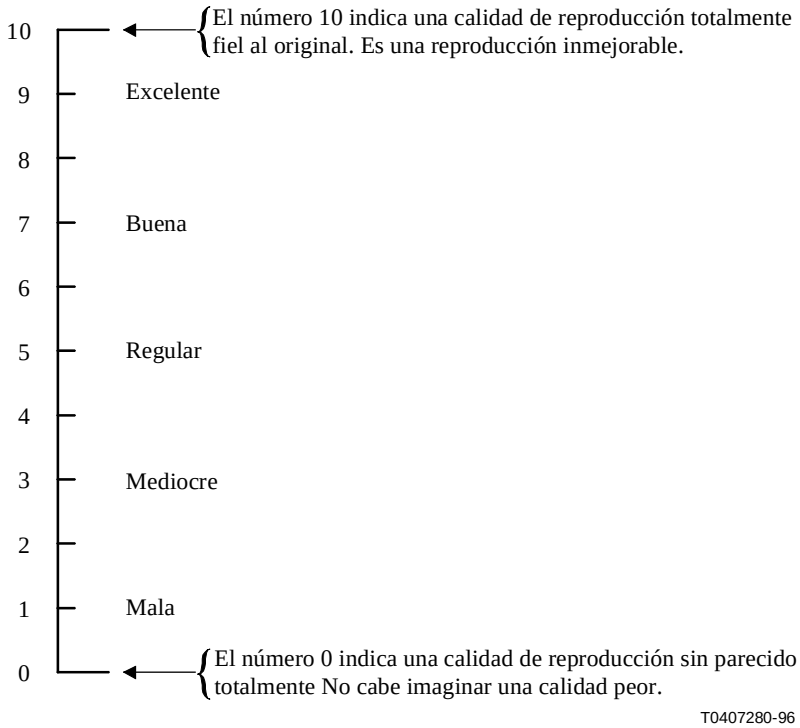


Figura B.2 – Escala de calidad numérica de 11 grados

Para ambos tipos de escalas, las respuestas de los sujetos se pueden registrar como números, que se escriben en una hoja de respuestas, o como marcas sobre la propia escala (en cuyo caso se ha de entregar en la hoja de respuestas una escala separada por cada condición de calificación). Cuando se requieran respuestas numéricas, se debería instar a los sujetos a que utilicen decimales (por ejemplo, 2,2 en lugar de 2), pero todavía tienen la posibilidad de utilizar enteros solamente.

Cabe señalar que quizá sea difícil traducir los nombres de la escala de categorías a los diferentes idiomas. Si se efectuara la traducción, la relación entre categorías resultante podría ser diferente de la del idioma original [b-Virtanen].

Una posibilidad adicional consiste en utilizar escalas continuas.

Puesto que los datos continuos se redondean por lo general con un grado de precisión razonable, se puede utilizar una escala de votación como la indicada en la Figura B.3 para simplificar la recogida de datos. Sólo se atribuyen calificaciones a los puntos extremos y se indica una marca en la mitad de la escala. Así se reduciría el sesgo debido a la interpretación de las calificaciones. Cada sector puede corresponder a un valor numérico específico y los datos pueden ser recogidos sin ambigüedades.

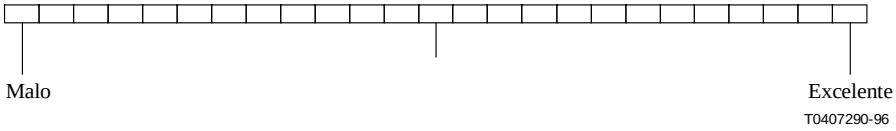


Figura B.3 – Escala casi continua para índices de calidad

B.2 Dimensiones de calificación adicionales

Si los sistemas evaluados en una prueba se consideran casi iguales en cuanto a calidad global y, por tanto, obtienen notas muy similares, quizás sea ventajoso calificar componentes de calidad adicionales en escalas separadas para cada condición. De esta manera es posible obtener información sobre características específicas cuando los objetos de prueba se perciben como significativamente diferentes, aun si la calidad global es en realidad casi la misma. Los resultados de tales pruebas adicionales pueden suministrar una valiosa información de diagnóstico sobre los sistemas sometidos a prueba.

A continuación se dan ejemplos de dimensiones de calificación a los que cabe considerar como definitorias de factores que contribuyen a la calidad de imagen global percibida, junto con una indicación que señala si el factor contribuye de forma positiva o negativa a dicha calidad:

– brillo (positivo);
– contraste (positivo);
– reproducción del color (positivo);
– definición del contorno (positivo);
– estabilidad del fondo (positivo);
– velocidad de recomposición de la imagen (positivo);
– inestabilidad (negativo);
– efectos "borrosidad" (negativo);
– efectos "mosquito" (negativo);
– imágenes/sombras dobles (negativo);
– halo (negativo).

Recientes investigaciones han demostrado que estos factores pueden combinarse en una calidad global predeterminada atribuyendo ponderaciones apropiadas a cada factor y sumándolos a continuación [b-RACE].

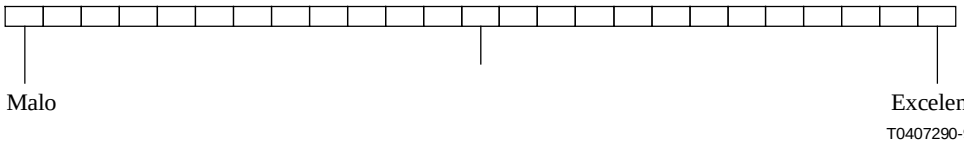
Para evaluar separadamente las dimensiones de la calidad vídeo global, se puede utilizar un cuestionario especial. En el cuestionario que sigue se dan ejemplos de las preguntas que se pueden formular tras la presentación de cada condición de prueba.

Cuestionario

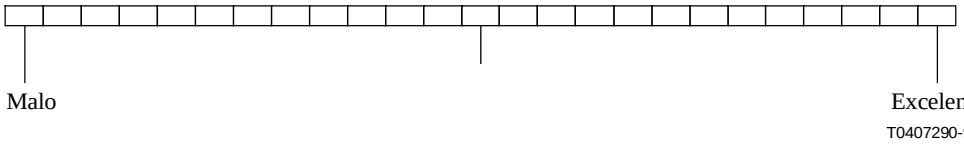
¿Quiere usted hacer el favor de responder a las siguientes preguntas relativas a la última secuencia mostrada?

Puede expresar su opinión poniendo una marca en las escalas que figuran a continuación.

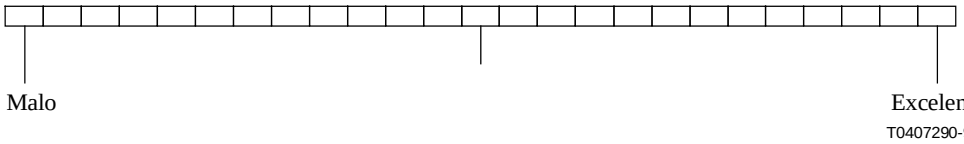
1) ¿Cómo calificaría el color de la imagen?



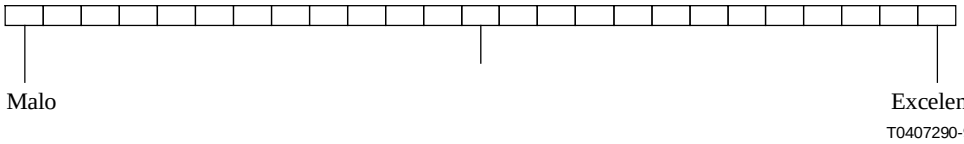
2) ¿Cómo calificaría el contraste de la imagen?



3) ¿Cómo calificaría los bordes de la imagen?

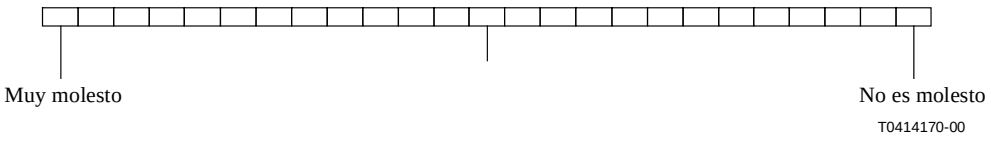


4) ¿Cómo calificaría la continuidad del movimiento?



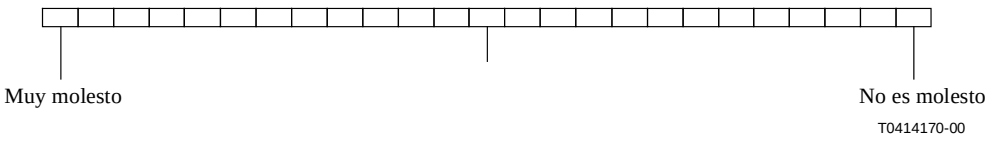
5) ¿Ha notado algún parpadeo en la secuencia? ☐ Sí ☐ No

Si ha notado un parpadeo, sírvase calificarlo en la escala siguiente



6) ¿Ha notado alguna borrosidad en la secuencia? ☐ Sí ☐ No

Si ha notado borrosidad, sírvase calificarla en la escala siguiente



NOTA – Cuando se utilizan estas escalas, se deben ilustrar cuidadosamente todas las categorías de calidad/degradación tenidas en cuenta (por ejemplo, continuidad del movimiento, parpadeo, borrosidad, etc.) durante las sesiones de instrucción.

Anexo C

Presentación simultánea de pares de secuencias

(Este anexo forma parte integrante de la presente Recomendación.)

C.1 Introducción

Cuando los sistemas que se evalúan en una prueba utilizan formato de imagen reducido, tal como CIF, QCIF, SIF, etc., y se emplean los métodos DCR o PC, puede que convenga visualizar simultáneamente las dos secuencias de cada par en un mismo monitor.

Las ventajas de utilizar presentación simultánea (SP) son como sigue:

1) La SP reduce considerablemente la duración de la prueba.

2) Si se utilizan dimensiones de imagen apropiadas, les resulta más sencillo a los sujetos evaluar las diferencias entre los estímulos.

3) Puesto que, con las mismas condiciones de prueba se reduce a la mitad el número de presentaciones, la atención de los sujetos es mayor normalmente cuando se utiliza SP.

La SP requiere precauciones particulares para que los sujetos puedan evitar el sesgo debido al tipo de presentación.

C.2 Sincronización

Las dos secuencias deben estar perfectamente sincronizadas, lo que significa que ambas deben comenzar y terminar en la misma trama y que la presentación visual debe estar sincronizada. Esto no impide que se puedan comparar secuencias codificadas con diferentes velocidades binarias, siempre que se aplique un muestreo temporal adecuado.

C.3 Condiciones de observación

Las secuencias se deben visualizar en dos ventanas presentadas una junto a la otra con un 50% de fondo gris (el tono gris se especifica en la cláusula 5.1), como se muestra en la Figura C.1. Para reducir el movimiento del ojo al pasar la atención de una a otra ventana, la distancia de observación debe ser de 8H, donde H indica la altura de imagen. La dimensión de la diagonal de los monitores ha de ser de 14 pulgadas como mínimo.

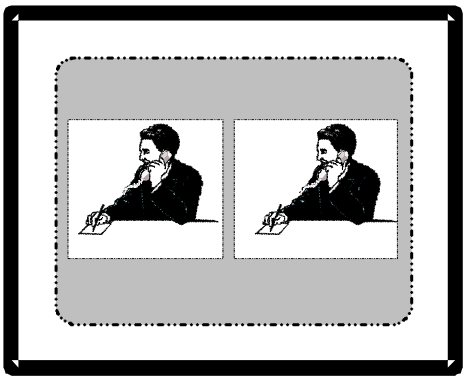


Figura C.1 – Posición relativa de las dos secuencias en SP

C.4 Presentaciones

En el método DCR, la referencia se debe situar siempre en el mismo lado (por ejemplo, a la izquierda) y los sujetos deben conocer las posiciones relativas de la referencia y las condiciones de prueba.

En el método PC, los pares de secuencias se deben visualizar en los dos órdenes posibles (por ejemplo, AB, BA). Esto significa que las secuencias que se presentaron en el lado izquierdo se presentan ahora en el derecho y viceversa.

Anexo D

Clases de vídeo y audio y sus atributos

(Este anexo forma parte integrante de la presente Recomendación.)

En la presente Recomendación, la calidad de vídeo más alta considerada es la de [UIT-R BT.601], vídeo codificado MIC lineal de 8 bit/muestra en formato 4:2:2, Y, C_R, C_B.

Cuadro D.1 – Definiciones de las clases de vídeo

TV 0	Sin pérdidas: vídeo de [UIT-R BT.601], 8 bits por muestra, utilizado para aplicaciones sin compresión
TV 1	Utilizada para postproducción completa, muchos montajes y capas de procesamiento, transmisión en plantas. Se utiliza también para transmisión entre lugares distantes y la planta. Perceptualmente transparente cuando se compara con TV 0
TV 2	Utilizada para modificaciones simples, pocos montajes, superposiciones de carácter/logo, inserción de programa y transmisión entre facilidades. Un ejemplo de difusión sería transmisión de red a afiliado. Otros ejemplos son un enlace regional de sistema de cable a un extremo local y un sistema de videoconferencias de alta calidad. Casi perceptualmente transparente cuando se compara con TV 0
TV 3	Utilizado para transmisión a domicilio/consumidor (sin cambios). Otros ejemplos son sistema de cable desde el extremo local al domicilio y videoconferencia de calidad media a alta. Aparecen perturbaciones cuando se compara con TV 2
MM 4	Todas las tramas están codificadas. Pocas perturbaciones en relación con TV 3. Videoconferencia de calidad media. Normalmente ≥ 30 trama/s
MM 5	Se pueden perder tramas en el codificador. Posibles perturbaciones perceptibles, pero con nivel de calidad útil para tareas designadas, por ejemplo videoconferencia de baja calidad
MM 6	Serie de imágenes fijas. No pretende proporcionar movimiento completo (ejemplos: vigilancia, gráficos)

Cuadro D.2 – Atributos de clases de vídeo

Clases de vídeo	**Formato espacial**	**Velocidad de trama entregada (Nota 1)**	**Latencia típica Variación de retardo (Nota 2)**	**Velocidad binaria de vídeo nominal (Mbit/s)**
TV 0	[UIT-R BT.601]	FR Máx	(Nota 2)	270
TV 1	[UIT-R BT.601]	FR Máx	(Nota 2)	18 a 50
TV 2	[UIT-R BT.601]	FR Máx	(Nota 2)	10 a 25
TV 3	[UIT-R BT.601]	FR Máx ocasional Repetición de tramas	(Nota 2)	1,5 a 8
MM 4a	[UIT-R BT.601]	~30 ó ~25 trama/s	Retardo <≈ 150 ms Variación <≈ 50 ms	~1,5
MM 4b	CIF	~30 ó ~25 trama/s	Retardo <≈ 150 ms Variación <≈ 50 ms	~0,7

Cuadro D.2 – Atributos de clases de vídeo

Clases de vídeo	Formato espacial	Velocidad de trama entregada (Nota 1)	Latencia típica Variación de retardo (Nota 2)	Velocidad binaria de vídeo nominal (Mbit/s)
MM 5a	CIF	10-30 trama/s	Retardo <≈ 1000 ms Variación <≈ 500 ms	~0,2
MM 5b	≤ CIF	1-15 trama/s	Retardo <≈ 1000 ms Variación <≈ 500 ms	~0,05
MM 6	CIF-16CIF	Límite → 0 trama/s	Sin restricciones	<0,05, Límite → 0 trama/s

NOTA 1 – Normalmente 30 tramas por segundo para sistemas 525 y 25 tramas por segundo para sistemas 625.

NOTA 2 – Todos los sistemas de difusión tienen latencia unidireccional constante, pero no necesariamente baja, y variación de retardo constante. Para la mayoría de las aplicaciones de difusión la latencia será baja, por ejemplo entre 50 y 500 ms. Para videoconferencias de alta calidad, y en general para tipos conversacionales de aplicaciones, la latencia debe ser preferentemente inferior a 150 ms (véase [b-UIT-T G.114]). Se permiten variaciones de retardo dentro de una gama dada pero no se deben llegar a percibir efectos perturbadores dinámicos.

Apéndice I

Secuencias de prueba

(Este apéndice no forma parte integrante de la presente Recomendación.)

La selección de las secuencias de prueba apropiadas es un punto clave en la planificación de la evaluación subjetiva. Cuando los resultados de las pruebas efectuadas con diferentes grupos de observadores o en diferentes laboratorios hayan de ser correlacionadas es importante disponer de un conjunto de secuencias de pruebas comunes.

En el Cuadro I.1 se describe un primer conjunto de dichas secuencias. En este cuadro se da, para cada secuencia, la siguiente información:

– la categoría (definida en el Cuadro A.1);

– una breve descripción de la escena;

– el formato fuente (625 ó 525 líneas, formato [UIT-R BT.601-4] o Betacam SP);

– los valores de la información espacial y la información temporal (definidas en las cláusulas 5.3.1 y 5.3.2, respectivamente).

Todas las secuencias indicadas en el Cuadro I.1 son de dominio público y se pueden utilizar libremente para evaluaciones y demostraciones. Algunas de las secuencias propuestas pertenecen a la biblioteca del CCIR descrita en [b-CCIR Informe 1213].

Se podrían utilizar convenientemente otras secuencias de la biblioteca del CCIR para aplicaciones particulares, tales como las basadas en el almacenamiento vídeo con recuperación.

El conjunto de secuencias de prueba está todavía en estudio. El conjunto de secuencias que figuran en el Cuadro I.1 se puede mejorar o ampliar de dos maneras por lo menos:

1) incluyendo secuencias representativas de una gama de aplicaciones más amplia (por ejemplo, videoteléfono móvil, aula distante, etc.);

2) haciendo que el formato fuente de todas las secuencias sea el formato [UIT-R BT.601-4] en las versiones de 525 y 625 líneas.

Cuadro I.1 – Secuencias de prueba para evaluación de calidad vídeo en aplicaciones multimedios

Secuencia	Categoría	Descripción	Formato fuente	SI	TI
washdc	D	Mapa de Washington DC con movimiento de mano y lápiz	Betacam SP (525 líneas)	130,5	17,0
3inrow	C	Hombres a la mesa, cámara panorámica	Betacam SP (525 líneas)	81,7	30,8
vtc1nw	A	Mujer sentada leyendo unas noticias	Betacam SP (525 líneas)	56,2	5,3
susie	A	Mujer joven al teléfono	ITU-R BT.601-4 525/625 líneas	58,7	24,6
flower garden	E	Paisaje, cámara panorámica	ITU-R BT.601-4 525/625 líneas	227,0	46,4
smity2	B	Vendedor en mostrador con revista	Betacam SP (525 líneas)	154,5	35,1

Apéndice II

Instrucciones para las pruebas de observación

(Este apéndice no forma parte integrante de la presente Recomendación.)

El material que sigue se puede utilizar como base para la instrucción de los evaluadores que participan en experimentos en los que se adoptan los métodos ACR, ACR-HR, DCR o PC.

Además, las instrucciones deben dar información acerca de la duración aproximada de la prueba, pausas, ensayos preliminares y otros detalles de utilidad para los evaluadores. Esta información no se incluye ya que depende de la implementación específica.

II.1 ACR y ACR-HR

Buenos días y gracias por estar aquí.

En este experimento verá usted secuencias de vídeo cortas en la pantalla situada frente a usted. Cada vez que se muestre una secuencia debe usted juzgar su calidad utilizando uno de los cinco niveles de la siguiente escala.

5 Excelente

4 Buena

3 Aceptable

2 Mediocre

1 Mala

Antes de efectuar su valoración observe cuidadosamente la secuencia de vídeo completa.

II.2 DCR

Buenos días y gracias por estar aquí.

En este experimento verá usted secuencias de vídeo cortas en la pantalla situada frente a usted. Cada secuencia se presentará dos veces en una sucesión rápida: en cada par sólo se procesa la segunda secuencia. Al final de cada presentación por pares debe usted evaluar la degradación de la segunda secuencia con respecto a la primera. Expresará su apreciación utilizando la siguiente escala:

5 Imperceptible

4 Perceptible pero no molesta

3 Ligeramente molesta

2 Molesta

1 Muy molesta

Antes de efectuar su valoración observe cuidadosamente el par completo de secuencias de vídeo.

II.3 PC

Buenos días y gracias por estar aquí.

En este experimento verá usted secuencias de vídeo cortas en la pantalla situada frente a usted. Cada secuencia se presentará dos veces en sucesión rápida: cada vez a través de un códec diferente. El orden de las secuencias y la combinación de los códecs en los pares varía de forma aleatoria. Al final de cada presentación por pares debe usted expresar su preferencia marcando una de las casillas mostradas a continuación. Marque la casilla 1 si prefiere la primera secuencia o la casilla 2 si prefiere la segunda secuencia del par.

1		2

Antes de efectuar su valoración observe cuidadosamente el par completo de secuencias de vídeo.

Apéndice III

El doble estímulo simultáneo para una evaluación contínua

(Este apéndice no forma parte integrante de la presente Recomendación.)

El doble estímulo simultáneo para una evaluación continua (SDSCE, *simultaneous double stimulus for a continuous evaluation*) es adecuado para evaluar el efecto de degradaciones dispersas, tales como errores de transmisión, en la fidelidad de la información visual. Este método se deriva del método de evaluación de calidad continua de estímulo único (SSCQE) descrito en [UIT-R BT.500-9].

III.1 Procedimiento de prueba

El panel de sujetos está observando dos secuencias simultáneamente: una es la referencia, y la otra es la condición de prueba. Si el formato de la secuencias es SIF o menor, pueden visualizarse las dos frecuencias lado a lado en el mismo monitor, o en otro caso deben utilizarse dos monitores alineados.

Se pide a los sujetos que comprueben las diferencias entre las dos secuencias y juzguen la fidelidad de la información de vídeo desplazando el cursor de un dispositivo de votación por microteléfono. Cuando la fidelidad es perfecta, el cursor debe hallarse en el máximo de la escala (codificado 100), y cuando la fidelidad es nula, el cursor debe estar en el mínimo de la escala (codificado 0).

Los sujetos saben cuál es la referencia y se les pide que expresen su opinión, mientras están observando las secuencias, en toda su duración.

III.2 La fase de formación

La fase de formación es una parte crucial de este método de prueba, ya que los sujetos podrían entender mal su tarea. Deben proporcionarse instrucciones por escrito para asegurarse de que todos los sujetos reciban exactamente la misma información, que incluyan una explicación de lo que van a ver, lo que tienen que evaluar (es decir, diferencia de calidad) y cómo deben expresar su opinión. Debe responderse a cualquier pregunta de los participantes para evitar en lo más posible cualquier influencia en la opinión por parte del administrador de la prueba.

Tras las instrucciones debe haber una sesión de demostración. De este modo los sujetos se familiarizan con los procedimientos de votación y el tipo de degradaciones.

Por último, debe hacerse una prueba simulada, en la que se muestren diversas condiciones representantivas. Las secuencias deben ser diferentes de las utilizadas en la prueba y reproducirse una tras otra sin interrupción alguna.

Una vez finalizada la prueba simulada, el experimentador debe comprobar que en el caso de condiciones de prueba iguales a las referencias, las evaluaciones se acerquen a 100, o en caso contrario repetir la explicación y repetir la prueba simulada.

III.3 Características del protocolo de prueba

Se aplican las siguientes definiciones a la descripción del protocolo de prueba:

- *Segmento de vídeo (VS, video segment)*: Un VS corresponde a una secuencia de vídeo.
- *Condición de prueba (TC, test condition)*: Una TC puede ser un proceso de vídeo específico, una condición de transmisión o ambas cosas. Cada VS debe procesarse al menos en una TC. Además, deben añadirse referencias a la lista de TC, a fin de hacer que se evalúen los pares "referencia/referencia".

- *Sesión (S)*: Una sesión es una serie de pares VS/TC diferentes sin separación y en orden seudoaleatorio. Cada sesión contiene al menos una vez todos los VS y TC, pero no necesariamente todas las combinaciones VS/TC. Todas las combinaciones de VS/TC deben ser votadas por el mismo número de observadores (pero no necesariamente los mismos).
- *Presentación de prueba (TP, test presentation)*: una presentación de prueba es una serie de sesiones que abarquen todas las combinaciones VS/TC.
- *Periodo de votación*: Se pide a cada observador que vote continuamente durante una sesión.

III.4 Procesamiento de los datos

Una vez realizada una prueba, se dispone de uno o más ficheros de datos que contienen todos los votos de las diferentes sesiones (S), que representan el material de votación completo de la presentación de prueba (TP). Puede hacerse una primera comprobación de la validez de los datos verificando que se ha considerado cada par VS/TC y que se ha asignado a cada uno de ellos un número equivalente de votos.

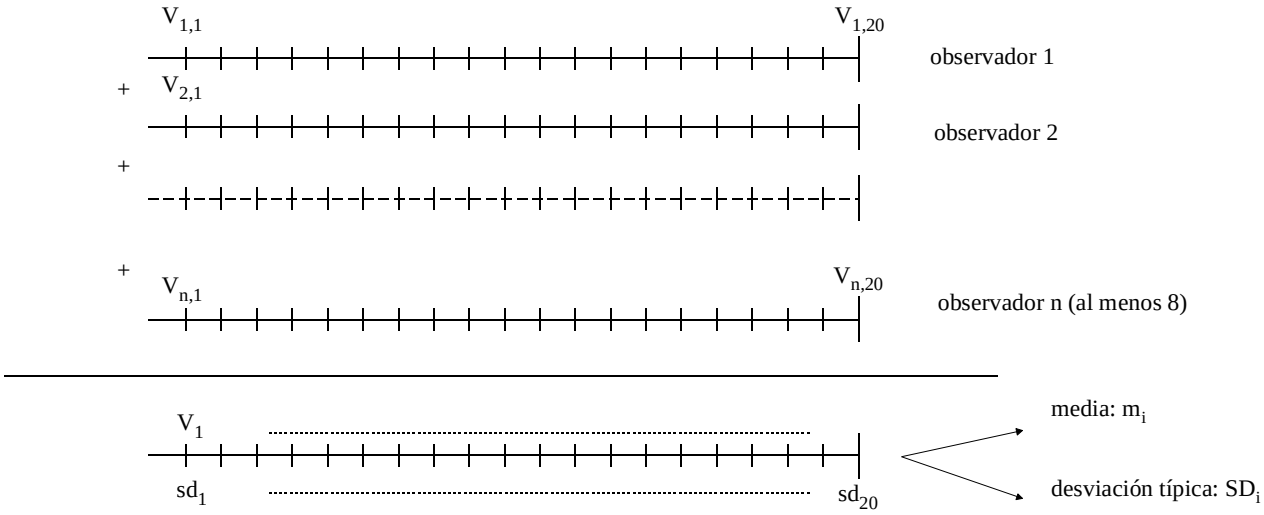
Los datos de las pruebas realizadas con arreglo a este protocolo pueden procesarse de tres maneras diferentes:

- Análisis estadístico de cada VS por separado.
- Análisis estadístico de cada TC por separado.
- Análisis estadístico global de todos los pares VS/TC.

Se necesita en cada caso un análisis en varias etapas:

- Se calculan las medias y las desviaciones típicas para cada punto de voto por acumulación de los observadores, como se ilustra en la Figura III.1.
- Cada VS se considera como un conjunto de segmentos de votación de duración máxima 10 segundos. Dado que ni la condición de reciente ni el efecto de olvido repercuten en la evaluación de secuencias que duran no más 10 s se calculan, la media y la desviación típica de las medias calculadas en la etapa anterior para cada segmento de votación, como se ilustra en la Figura III.1. Cuando se necesita información detallada sobre la variabilidad de la calidad, la duración del segmento de votación debe ser breve (en torno a un segundo). Los resultados de esta etapa pueden representarse en un diagrama de tiempos, como se muestra en la Figura III.2.
- Se analizan la distribución estadística de las medias calculadas en la etapa anterior (es decir, correspondientes a cada segmento de votación) y su frecuencia de aparición. A fin de evitar el efecto reciente debido al par anterior VS/TC, se rechazan los primeros 10 segundos de votos sobre cada muestra VS/TC. En la Figura III.3 se incluye un ejemplo.
- La característica de molestia global se calcula acumulando las frecuencias de ocurrencia. Deben tenerse en cuenta en este cálculo los intervalos de confianza, que se muestran en la Figura III.4. Una característica de molestia global corresponde a esta función de distribución estadística acumulativa mostrando la relación entre las medias de cada uno de los segmentos de votación y su frecuencia acumulativa de aparición.

1) Cálculo de la nota media (V) y de la desviación típica (sd) de los observadores para cada secuencia de votación de cada combinación VS/TC



2) Cálculo de la media (M) y de la desviación típica (SD) por secuencia de votación de 1 s para la combinación VS/TC

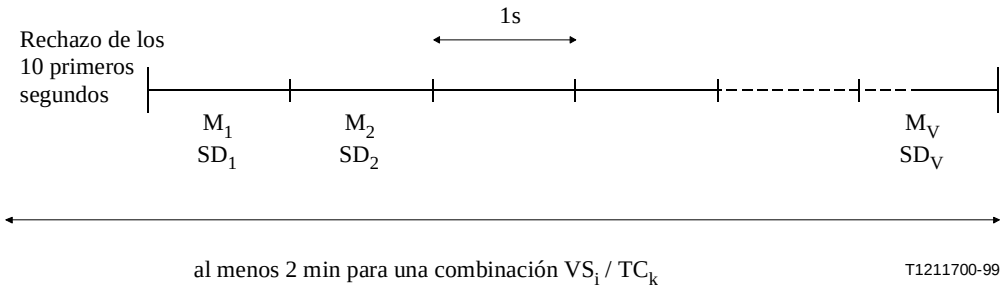


Figura III.1 – Procesamiento de los datos

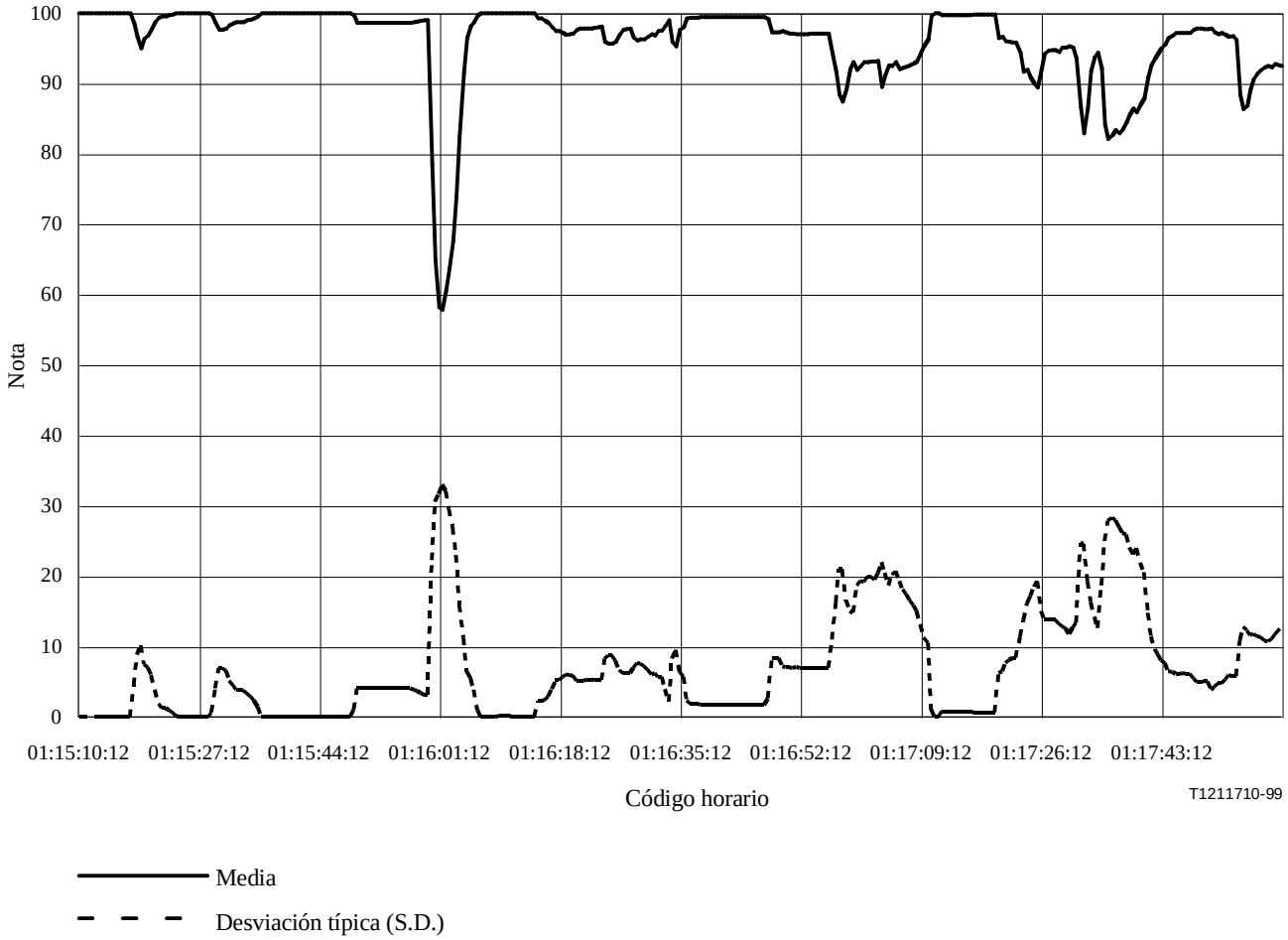


Figura III.2 – Diagrama de tiempo en bruto

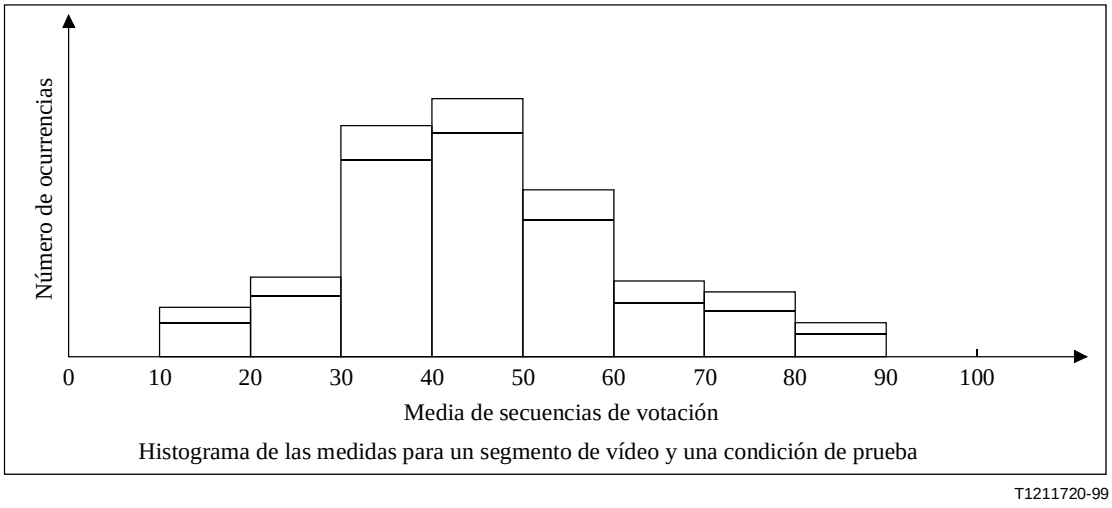


Figura III.3 – Relación entre las características de degradación y su número de ocurrencias

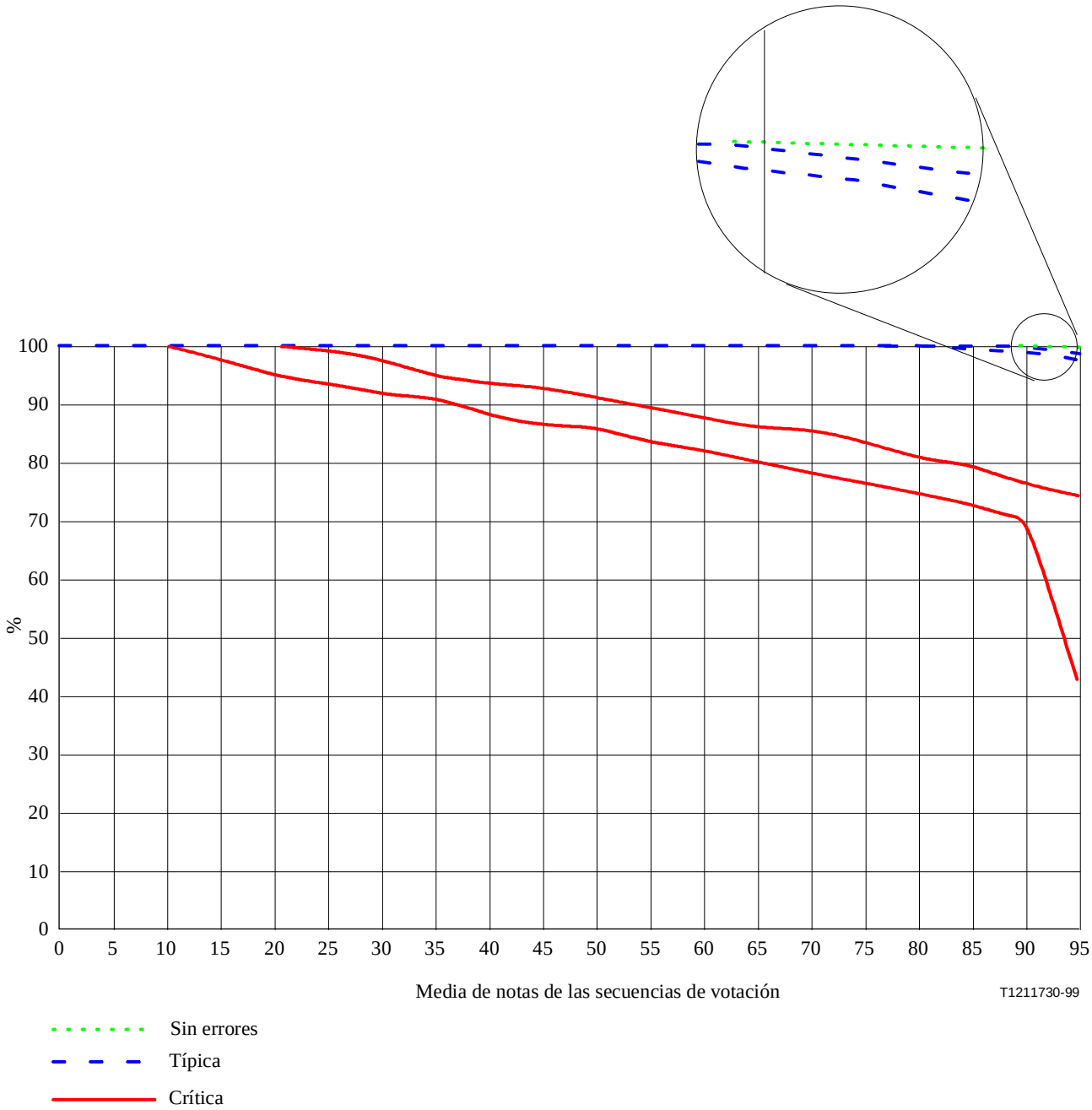


Figura III.4 – Característica de molestia global calculadas a partir de distribuciones estadísticas e incluido el intervalo de confianza

III.5 Fiabilidad de los sujetos

La fiabilidad de los sujetos puede evaluarse cuantitativamente comprobando su comportamiento cuando se muestran pares "referencia/referencia". En estos casos, se espera que los sujetos den evaluaciones muy próximas a 100, lo cual demuestra que al menos entendieron su tarea y que no están dando votos de manera aleatoria.

Además, la fiabilidad de los sujetos puede comprobarse utilizando procedimientos próximos a los descritos en [UIT-R BT.500-9] para el método SSCQE.

En el procedimiento SDSCE, la fiabilidad de los votos depende de los dos parámetros siguientes:

Desplazamientos sistemáticos – Durante una prueba, un observador puede ser demasiado optimista o demasiado pesimista, o puede incluso haber entendido mal los procedimientos de votación (por ejemplo, el significado de la escala de votación). Esto puede conducir a una serie de votos sistemáticamente más o menos desplazados con respecto a la serie media, si no completamente fuera de gama.

Inversiones locales – En los otros procedimientos de prueba conocidos, los observadores pueden a veces votar sin poner mucha atención en la observación y el seguimiento de la calidad de la secuencia visualizada. En este caso, la curva de voto total puede estar "relativamente" dentro de la gama media. No obstante, pueden observarse las inversiones locales.

Estos dos efectos indeseables (comportamiento atípico e inversiones) podrían evitarse. La formación de los participantes es por supuesto muy importante, no obstante debe ser posible utilizar una herramienta que permita detectar y, si es necesario, descartar a los observadores menos consecuentes.

Apéndice IV

La evaluación por objetos

(Este apéndice no forma parte integrante de la presente Recomendación.)

Las funcionalidades por objetos deben evaluarse en la escena completa y en los objetos aislados, lo cual se debe a que en general una escena compuesta de objetos codificados independientemente puede ser "utilizada" como ha sido producida por el autor, pero en algunos casos puede también ser manipulada y cada objeto aislado puede ser utilizado en un contexto completamente diferente. Por esta razón es importante tener un equilibrio entre la calidad global de toda la escena y la calidad de la textura y de los contornos de cada objeto aislado.

Por tanto, las funcionalidades por objetos (que son la escalabilidad de los objetos y la escalabilidad de la calidad por objetos) debe evaluarse en dos pasadas:

Evaluación de la imagen completa – Ésta es una prueba clásica con la secuencia completa, que está incluyendo todos los VO. Los métodos de evaluación pueden ser el ACR (véase la cláusula 6.1) o el DCR (véase la cláusula 6.3), según la gama de velocidades binarias y la criticidad de las secuencias fuente.

Evaluación por objetos (OBE) – En esta prueba se visualizará sólo uno de los VO en un fondo gris y se pedirá a los sujetos que evalúen la calidad/degradación (por el método de prueba utilizado en la evaluación de la imagen completa) del VO mostrado. Tiene que especificarse el porcentaje de velocidad binaria que se gasta en el VO. El VO evaluado se extraerá de la misma secuencia codificada exacta que se utilizó en la evaluación de la imagen completa.

La Figura IV.1 ilustra las dos pruebas que hay que efectuar para la evaluación de la escalabilidad de los objetos.

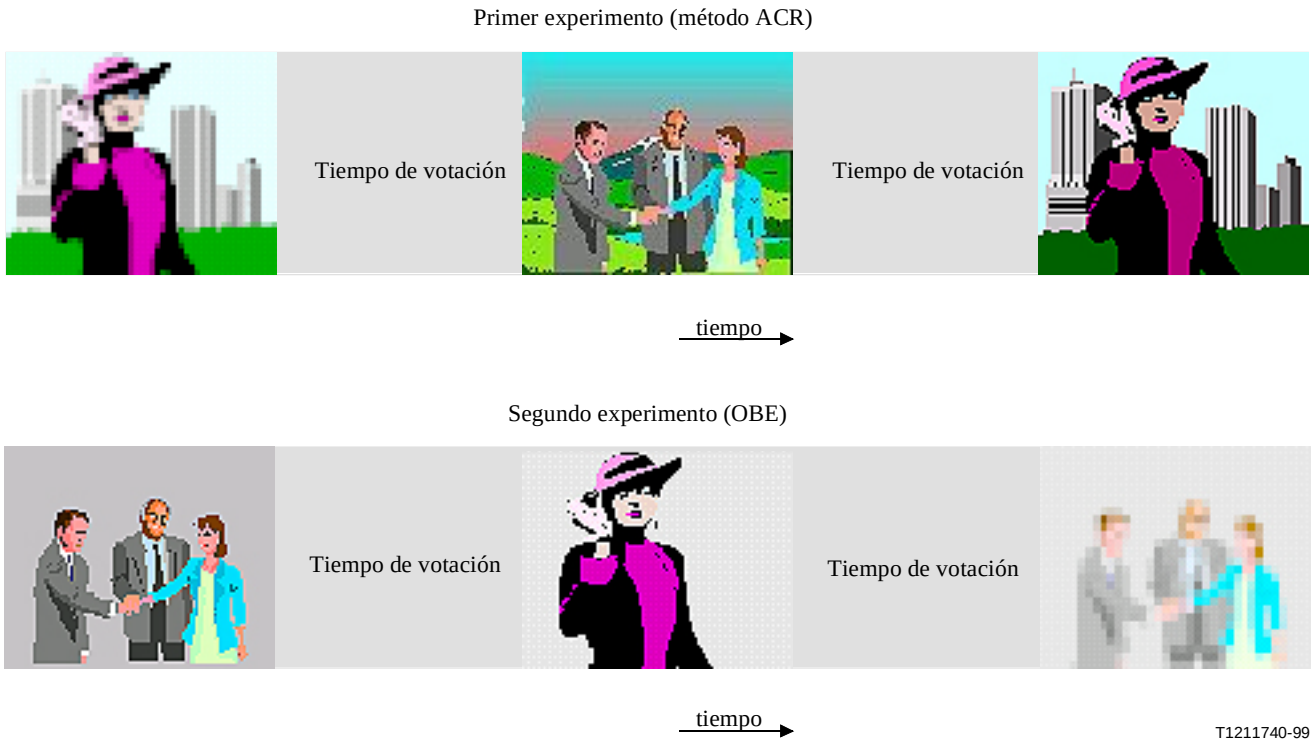


Figura IV.1 – Pruebas para evaluar la escalabilidad de los objetos

En el caso de una escalabilidad de calidad por objetos, deben efectuarse pruebas separadas para evaluar la escalabilidad en el espacio y la escalabilidad en el tiempo, y sólo deberá aplicarse una OBE.

Tanto para la escalabilidad en el espacio como en el tiempo, la OBE debe aplicarse para evaluar en una misma pasada ambos VO codificados a velocidades binarias "base" y los mismos VO codificados a velocidades binarias mejoradas especificadas.

En general la evaluación de las funcionalidades por objetos debe tener en cuenta la calidad de la trama completa y la calidad de los objetos aislados. La primera evaluación debe efectuarse por métodos normalizados y la segunda con la OBE.

Para establecer una comparación entre los diferentes sistemas basados en la codificación por objetos, el experimentador debe especificar por adelantado el peso relativo para asignar la calidad global y la calidad de objeto individual.

En casos particulares, merecerá también la pena utilizar criterios de evaluación por tareas en lugar de evaluaciones de calidad tradicionales. Por ejemplo, en la evaluación de un sistema de monitorización a distancia a utilizar en un garaje, la escalabilidad de la calidad debe evaluarse en términos de legibilidad de las placas de los vehículos. La tarea será decidida caso a caso por el experimentador, según la finalidad de la prueba y el tipo de aplicación que se investigue.

Por último, la evaluación de la calidad de los objetos puede aplicarse para investigar la repercusión de la calidad de los objetos aislados en la calidad global de la escena. Los resultados de dicho estudio podrían ser utilizados para optimizar los esquemas de codificación por objetos.

Apéndice V

Escala de evaluación adicional por DRC

(Este apéndice no forma parte integrante de la presente Recomendación.)

Una escala de degradación de 9 grados podría utilizarse, como la que muestra la Figura V.1. En esta escala, el grado 8 corresponde al umbral de perceptibilidad de la degradación, es decir el nivel de degradación en el que el observador no está completamente seguro de percibir degradación.

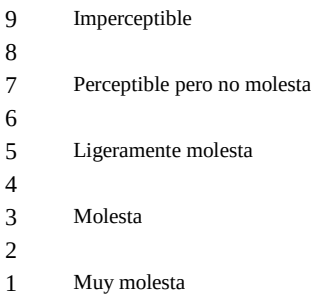


Figura V.1 – Escala de degradación numérica de 9 grados

Bibliografía

[b-UIT-T G.114] Recomendación UIT-T G.114 (2003), *Tiempo de transmisión en un sentido*.

[b-UIT-T H.261] Recomendación UIT-T H.261 (1993), *Códec vídeo para servicios audiovisuales a p × 64 kbit/s*.

[b-UIT-T P.920] Recomendación UIT-T P.920 (1996), *Métodos de prueba interactivos para comunicaciones audiovisuales*.

[b-UIT-T Manual] Manual UIT-T (1993), *Manual de telefonetría*, UIT, Ginebra.

[b-UIT-R BT.812] Recomendación UIT-R BT.812 (1992), *Evaluación subjetiva de la calidad de las imágenes alfanuméricas y gráficas en servicios de teletexto y similares*.

[b-UIT-R BT.815-1] Recomendación UIT-R BT.815-1 (1994), *Especificación de una señal para medir la relación de contraste de las pantallas*.

[b-CCIR Informe 1213] Informe 1213 del CCIR (1990), *Imágenes y secuencias de prueba para las evaluaciones subjetivas de códecs digitales*, Anexo al Volumen XI, Parte 1.

[b-Gonzalez] Gonzalez, R.C. and Wintz, P. (1987), *Digital Image Processing*, 2ª edición, Addison-Wesley Publishing Co., Reading, Massachusetts.

[b-RACE] RACE Industrial Consortium Project 1018 HIVITS, WP B5, *Picture Quality Measurement*, 1988.

[b-Snellen] Snellen Eye Chart.

[b-Beck] *Pseudo Isochromatic Plates* (1940), impreso por The Beck Engraving Co., Inc., Philadelphia y New York, Estados Unidos.

[b-Kirk] Kirk, R.E. (1982), *Experimental Design – Procedures for the Behavioural Sciences*, 2ª edición, Brooks/Cole Publishing Co., California.

[b-Virtanen] Virtanen, M.T., Gleiss, N. and Goldstein, M. (1995), *On the use of Evaluative Category Scales in Telecommunications*, Human Factors in Telecommunication Conference, Melbourne.

[b-Guilford] Guilford, P. (1954), *Psychometric methods*, McGraw-Hill, Nueva York.

[b-ISO/IEC 11172] ISO/IEC 11172:1993, *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s*.

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie D	Principios de tarificación y contabilidad y cuestiones económicas y políticas de las telecomunicaciones/TIC internacionales
Serie E	Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión, sistemas y redes digitales
Serie H	Sistemas audiovisuales y multimedia
Serie I	Red digital de servicios integrados
Serie J	Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedia
Serie K	Protección contra las interferencias
Serie L	Medio ambiente y TIC, cambio climático, ciberdesechos, eficiencia energética, construcción, instalación y protección de los cables y demás elementos de planta exterior
Serie M	Gestión de las telecomunicaciones, incluida la RGT y el mantenimiento de redes
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	**Calidad de la transmisión telefónica, instalaciones telefónicas y redes de líneas locales**
Serie Q	Conmutación y señalización, y mediciones y pruebas asociadas
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Terminales para servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos, comunicaciones de sistemas abiertos y seguridad
Serie Y	Infraestructura mundial de la información, aspectos del protocolo Internet, redes de próxima generación, Internet de las cosas y ciudades inteligentes
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación

Impreso en Suiza
Ginebra, 2017