

Reemplazada por una versión más reciente



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

P.910

(08/96)

SERIE P: CALIDAD DE TRANSMISIÓN TELEFÓNICA
Calidad audiovisual en servicios multimedios

**Métodos de evaluación subjetiva de la calidad
vídeo para aplicaciones multimedios**

Recomendación UIT-T P.910
Reemplazada por una versión más reciente

(Anteriormente Recomendación del CCITT)

Reemplazada por una versión más reciente

RECOMENDACIONES DE LA SERIE P DEL UIT-T

CALIDAD DE TRANSMISIÓN TELEFÓNICA

Vocabulario y efectos de los parámetros de transmisión sobre la opinión de los clientes	Serie	P.10
Líneas y aparatos de abonado	Serie	P.30 P.300
Patrones de transmisión	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 para evaluaciones objetivas y subjetivas de la calidad	Serie	P.80 P.800
Calidad audiovisual en servicios multimedios	Serie	P.900

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

Reemplazada por una versión más reciente

RECOMENDACIÓN UIT-T P.910

MÉTODOS DE EVALUACIÓN SUBJETIVA DE LA CALIDAD VÍDEO PARA APLICACIONES MULTIMEDIOS

Resumen

Esta Recomendación tiene por objeto definir métodos de evaluación subjetiva no interactivos para determinar la calidad de las imágenes de vídeo digital codificadas a velocidades binarias bajas y medias (hasta 2 Mbit/s) para aplicaciones tales como la videotelefonía, la videoconferencia y el almacenamiento con recuperación.

Se tienen en cuenta los siguientes aspectos:

- Montaje de laboratorio para producir secuencias de prueba.
- Montaje de laboratorio para efectuar evaluaciones subjetivas.
- Características de las secuencias de prueba.
- Método de prueba y diseños experimentales.
- Análisis de los datos.

Esta Recomendación no trata los aspectos que ya están incluidos en otras Recomendaciones, como los siguientes:

- Condiciones de referencia del vídeo, definidas en la Recomendación P.930.
- Procedimientos para el ajuste del monitor, descritos en el Informe 1221 del CCIR.
- Métodos de prueba interactivos, definidos en la Recomendación P.920.

Orígenes

La Recomendación UIT-T P.910 ha sido preparada por la Comisión de Estudio 12 (1993-1996) del UIT-T y fue aprobada por el procedimiento de la Resolución N.º 1 de la CMNT el 30 de agosto de 1996.

Reemplazada por una versión más reciente

PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) 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 Conferencia Mundial de Normalización de las Telecomunicaciones (CMNT), 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 N.º 1 de la CMNT (Helsinki, 1 al 12 de marzo de 1993).

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.

© UIT 1996

Es propiedad. Ninguna parte de esta publicación puede reproducirse o utilizarse, de ninguna forma o por ningún medio, sea éste electrónico o mecánico, de fotocopia o de microfilm, sin previa autorización escrita por parte de la UIT.

Reemplazada por una versión más reciente

ÍNDICE

	Página
1 Alcance	1
2 Referencias.....	1
3 Términos y definiciones.....	2
4 Abreviaturas.....	3
5 Señal fuente.....	4
5.1 Entorno de grabación	4
5.2 Sistema de grabación	4
5.2.1 Cámara.....	4
5.2.2 Señal vídeo y formato de almacenamiento.....	5
5.3 Características de la escena	5
5.3.1 Medición de la información de percepción espacial	5
5.3.2 Medición de la información de percepción temporal	6
6 Métodos de prueba y diseño experimental	6
6.1 Índices por categorías absolutas (ACR)	6
6.2 Índices por categorías de degradación (DCR)	7
6.3 Método de comparación por pares (PC)	8
6.4 Comparación de los métodos.....	9
6.5 Condiciones de referencia.....	10
6.6 Diseño experimental	10
7 Procedimientos de evaluación	11
7.1 Condiciones de observación	11
7.2 Sistema de procesamiento y reproducción.....	11
7.3 Observadores	11
7.4 Instrucciones a los observadores y sesión de instrucción.....	12
8 Análisis estadístico y notificación de los resultados	12
Bibliografía	13
Anexo A - Detalles relacionados con la caracterización de las secuencias de prueba	14
A.1 Filtro Sobel	14
A.2 Cómo utilizar SI y TI para la selección de secuencias de prueba.....	15
A.3 Ejemplos	15
Anexo B - Escalas de evaluación adicionales.....	16
B.1 Escalas de índices	16
B.2 Dimensiones de calificación adicionales	18

Reemplazada por una versión más reciente

Página

Anexo C - Presentación simultánea de pares de secuencias	20
C.1 Introducción	20
C.2 Sincronización	20
C.3 Condiciones de observación	20
C.4 Presentaciones.....	21
Apéndice I - Secuencias de prueba	21
Apéndice II - Instrucciones para las pruebas de observación	22
II.1 ACR	22
II.2 DCR	22
II.3 PC.....	23

Reemplazada por una versión más reciente

Recomendación P.910

MÉTODOS DE EVALUACIÓN SUBJETIVA DE LA CALIDAD VÍDEO PARA APLICACIONES MULTIMEDIOS

(Ginebra, 1996)

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 velocidades binarias bajas y medias (hasta 2 Mbit/s) para aplicaciones tales como la videotelefonía, la videoconferencia y el almacenamiento con recuperación. Los métodos que se describen en las cláusulas siguientes sirven también para evaluar la repercusión de los errores de transmisión en esos sistemas de vídeo. 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 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.

- [1] Recomendación UIT-T P.930 (1996), *Sistema de degradación de referencia para vídeo*.
- [2] Recomendación UIT-T P.920 (1996), *Métodos de prueba interactivos para comunicaciones audiovisuales*.
- [3] Recomendación UIT-R BT.601-4 (1994), *Parámetros de codificación de televisión digital para estudios*.
- [4] Recomendación UIT-R BT.500-6 (1994), *Método para la evaluación subjetiva de la calidad de las imágenes de televisión*.
- [5] Publicación 268-13 de la CEI, *Sound system equipment: listening tests on loudspeakers*.
- [6] CCITT: *Manual sobre telefonometría*, Ginebra 1993.
- [7] Recomendación UIT-T P.800 (1996), *Métodos de determinación subjetiva de la calidad de transmisión*.
- [8] Recomendación UIT-R BT.814-1 (1994), *Especificaciones y procedimientos de ajuste para establecer el brillo y el contraste en las pantallas*.
- [9] Recomendación UIT-R BT.815-1 (1994), *Especificación de una señal para medir la relación de contraste de las pantallas*.
- [10] Informe 1213 del CCIR: *Imágenes y secuencias de prueba para la evaluación subjetiva de códecs digitales - Anexo al tomo XI - Parte 1*, 1990.

Reemplazada por una versión más reciente

- [11] Recomendación 567-3 del CCIR: *Calidad de transmisión de los circuitos de televisión diseñados para ser utilizados en conexiones - Tomo XII*, 1990.
- [12] Recomendación UIT-R BT.812 (1994), *Evaluación subjetiva de la calidad de las imágenes alfanuméricas y gráficas en servicios de teletexto y similares*.

3 Términos y definiciones

A los efectos de esta Recomendación, se aplican las siguientes definiciones:

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): 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 5.3.1 la ecuación de SI.

3.5 información de percepción temporal (TI): 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 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 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.

Reemplazada por una versión más reciente

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: Recomendación 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

A los efectos de esta Recomendación, se utilizan las siguientes abreviaturas:

ACR	Índices por categorías absolutas (<i>absolute category rating</i>)
CCD	Dispositivo de acoplamiento de cargas (<i>charge coupled device</i>)
CI	Intervalo de confianza (<i>confidence interval</i>)
CIF	Formato intermedio común (<i>common intermediate format</i>) (formato de imagen definido en la Recomendación H.261 para videotelefonía: 352 líneas x 288 píxels)
CRT	Tubo de rayos catódicos (<i>cathode ray tube</i>)
DCR	Índices por categorías de degradación (<i>degradation category rating</i>)
%GOB	Porcentaje de bueno o mejor (<i>percent of good or better</i>) (proporción de votos bueno y excelente)
LCD	Visualización de cristal líquido (<i>liquid crystal display</i>)
MOS	Nota media de opinión (<i>mean opinion score</i>)
PC	Comparación por pares (<i>pair comparison</i>)
%POW	Porcentaje de mediocre o peor (<i>percent of poor or worse</i>) (proporción de votos mediocre y malo)
QCIF	Un cuarto de CIF (<i>quart CIF</i>) (formato de imagen definido en la Recomendación H.261 para videoteléfono: 176 líneas x 144 píxels)
S/N	Relación señal/ruido (<i>signal-to-noise ratio</i>)
SI	Información espacial (<i>spatial information</i>)
SIF	Formato intermedio normalizado (<i>standard immediate format</i>) [formatos de imagen definidos en la Norma ISO 11172 (MPEG-1): 352 líneas x 288 píxels x 25 tramas/s y 352 líneas x 240 píxels x 30 tramas/s]
SP	Presentación simultánea (<i>simultaneous presentation</i>)

Reemplazada por una versión más reciente

std	Desviación típica (<i>standard deviation</i>)
TI	Información temporal (<i>temporal information</i>)
VTR	Magnetoscopio (<i>video tape recorder</i>)

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 Y U V;
- 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), si se utiliza.

La relación S/N ponderada se debe medir de acuerdo con la Recomendación 567-3, Parte C, subcláusula 3.2.1 [11] del CCIR 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

Reemplazada por una versión más reciente

de los blancos puede variar de 2700° K (utilización en interiores con lámpara eléctrica) a 6500° 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 [3]. Para evitar la distorsión de la señal fuente, se debe almacenar en formato digital, por ejemplo, en computadora o formato de cinta D-1 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 pixels (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:

Reemplazada por una versión más reciente

$$SI = \max_{\text{time}} \{ \text{std}_{\text{space}} [\text{Sobel}(F_n)] \}$$

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 pixels (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 pixel 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 ($\text{std}_{\text{space}}$) de $M_n(i,j)$ en todas las i y j.

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

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 de hasta 2 Mbit/s.

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.

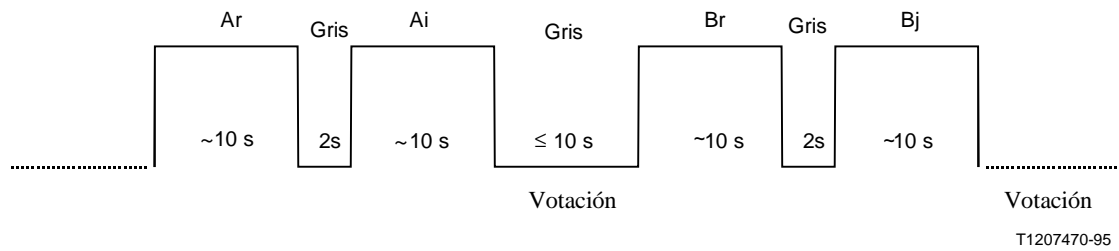
6.1 Índices por categorías absolutas (ACR)

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. El tiempo de votación debe ser igual o inferior a 10 segundos (s), dependiendo del mecanismo de votación utilizado. El tiempo de presentación se puede reducir o aumentar de acuerdo con el contenido del material de prueba.

Reemplazada por una versión más reciente



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/P.910

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.

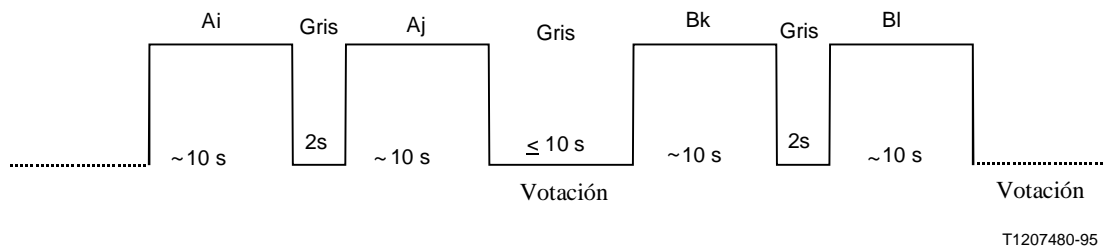
6.3 Método de comparación por pares (PC)

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. El tiempo de votación debe ser igual o inferior a 10 segundos, dependiendo del mecanismo de votación utilizado. El tiempo de presentación debe ser de unos 10 segundos y se puede reducir o aumentar de acuerdo con el contenido del material de prueba.

Reemplazada por una versión más reciente



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/P.910

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 las referencias [4] y [7].

6.4 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 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 [4], 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 multimedia. 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 aplicar 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.

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.

Reemplazada por una versión más reciente

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.

6.5 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 la Recomendación P.930 [1]. 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.

6.6 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. (Bibliografía, 5) 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.

Reemplazada por una versión más reciente

7 Procedimientos de evaluación

7.1 Condiciones de observación

La prueba se debe efectuar con las siguientes condiciones de observación, de acuerdo con [12]:

Distancia de observación	4H a 8H (véase la nota)
Valor de cresta de luminancia	70 cd/m ² a 200 cd/m ²
Relación de contraste de la pantalla sin iluminación de fondo	30 a 50
Relación luminancia de fondo/luminancia de pantalla máxima	~ 0,25
Iluminación	500 lux apróx.
Cromaticidad general	blanco

NOTA - 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.

Para una determinada altura de pantalla, es probable que la distancia de observación preferida por los sujetos aumente cuando se degrada la calidad visual. A este respecto, la distancia de observación preferida se debe determinar previamente para pruebas de calificación. En pruebas de optimización se deben utilizar distancias más críticas (por ejemplo, 4H-6H).

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 deberá ser 50% gris correspondiente a $Y=U=V=128$ (U y V sin signo).

Para la comparación de los resultados de la prueba, deben estar fijadas todas las condiciones de observación y ser iguales en todos los laboratorios para la misma clase de prueba.

7.2 Sistema de procesamiento y reproducción

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

- 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;
- 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 o un monitor LCD. 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 [8].

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.

Reemplazada por una versión más reciente

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 (Bibliografía, 3). 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 (Bibliografía, 4) 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 II 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 [4] y [5] 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 1 para el método ACR:

Reemplazada por una versión más reciente

CUADRO 1/P.910

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 [6].

Bibliografía

- [1] GONZÁLEZ (R.C.), WINTZ (P.): Digital Image Processing, 2nd Edition, *Addison-Wesley Publishing Co.*, Reading, Massachusetts 1987.
- [2] RACE Industrial Consortium Project 1018 HIVITS, WP B5, Picture Quality Measurement 1988.
- [3] Grahm-Field Catalogue Number 13-1240.
- [4] Pseudo Isochromatic Plates, engraved and printed by, *The Beck Engraving Co., Inc.*, Filadelfia y Nueva York, Estados Unidos.
- [5] KIRK (R.E.): Experimental Design – Procedures for the Behavioural Sciences, 2nd Editions, *Brooks/Cole Publishing Co.*, California 1982.
- [6] VIRTANEN (M.T.), GLEISS (N.), GOLDSTEIN (M.): On the use of Evaluative Category Scales in Telecommunications, HFT 1995, *Human Factors in Telecommunication Conference*, Melbourne 1995.

Reemplazada por una versión más reciente

Anexo A

Detalles relacionados con la caracterización de las secuencias de prueba

(Este anexo es parte integrante de la presente Recomendación)

A.1 Filtro Sobel

El filtro Sobel se realiza convolucionando dos núcleos 3x3 en la trama vídeo y extrayendo la raíz cuadrada de la suma de los cuadrados de los resultados de esas convoluciones.

Para $y = \text{Sobel}(x)$, sea $x(i,j)$ el pixel 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:

$$Gv(i,j) = -1*x(i-1,j-1) - 2*x(i-1,j) - 1*x(i-1,j+1) + \\ + 0*x(i,j-1) + 0*x(i,j) + 0*x(i,j+1) + \\ + 1*x(i+1,j-1) + 2*x(i+1,j) + 1*x(i+1,j+1)$$

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

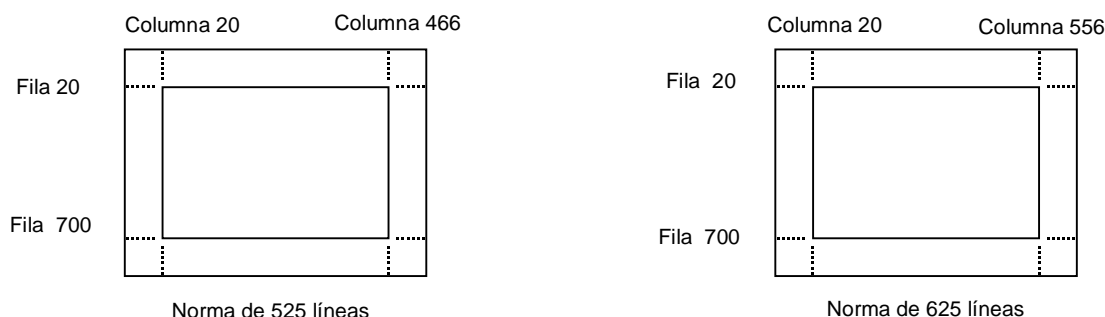
$$Gh(i,j) = -1*x(i-1,j-1) + 0*x(i-1,j) + 1*x(i-1,j+1) + \\ - 2*x(i,j-1) + 0*x(i,j) + 2*x(i,j+1) + \\ - 1*x(i+1,j-1) + 0*x(i+1,j) + 1*x(i+1,j+1)$$

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:

$$y(i,j) = \{[Gv(i,j)]^2 + [Gh(i,j)]^2\}^{0,5}$$

Los cálculos se efectúan para todas las $2 \leq i \leq N-1$ y $2 \leq j \leq M-1$, 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 [3].



T1207490-95

FIGURA A.1/P.910

Subimágenes que deben utilizarse para calcular SI y TI para formatos UIT-R BT.601-4 de 525 y 625 líneas [3]

Reemplazada por una versión más reciente

En (Bibliografía, 1) 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 [3]. En el Cuadro A.1 figura la lista de escenas de prueba del ejemplo por categoría del contenido de las mismas.

Reemplazada por una versión más reciente

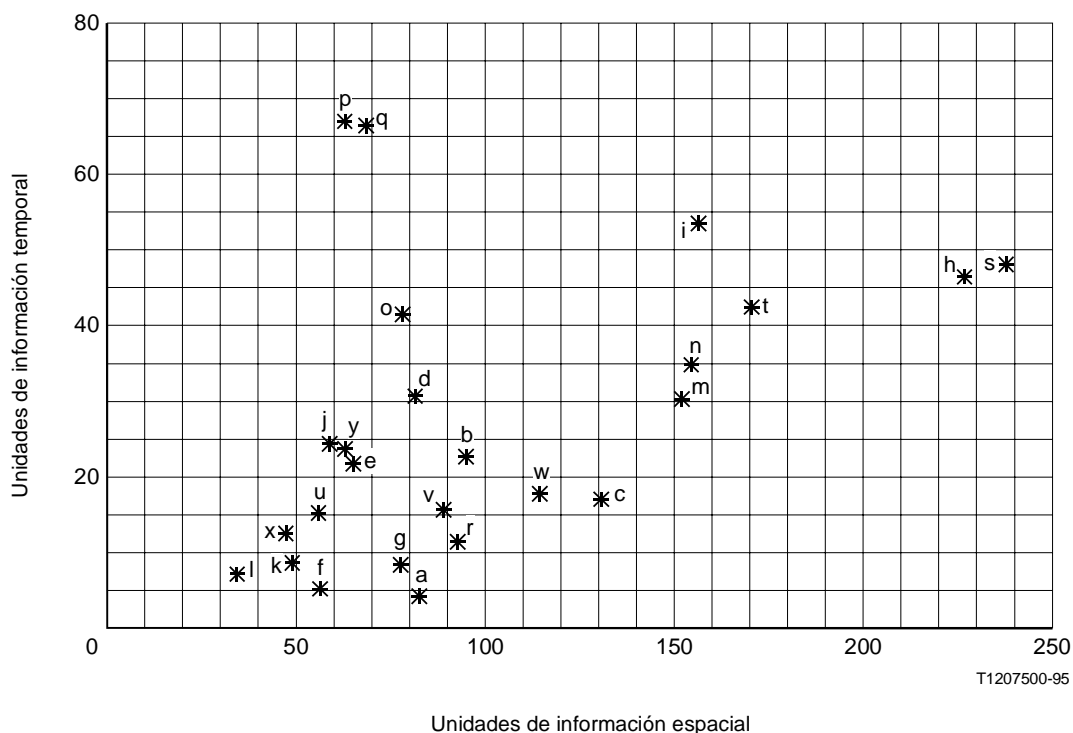


FIGURA A.2/P.910

Gráfico espacial-temporal de un ejemplo de conjunto de escenas de prueba

CUADRO A.1/P.910

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(l)
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), fball(i), fedas(y)

Anexo B

Escalas de evaluación adicionales

(Este anexo es 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 6.1 como etiquetas de calificación de los grados alternos de la escala, como se muestra en la Figura B.1.

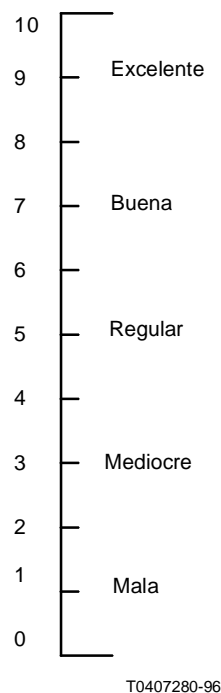
Reemplazada por una versión más reciente

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

FIGURA B.1/P.910

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 [5] y la sección 2.6, escala a, de [6].



El número 10 indica una calidad de reproducción totalmente fiel al original. Es una reproducción inmejorable.
El número 0 indica una calidad de reproducción sin parecido alguno con la original. No cabe imaginar una calidad peor.

FIGURA B.2/P.910

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

Reemplazada por una versión más reciente

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 (Bibliografía, 6).

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.

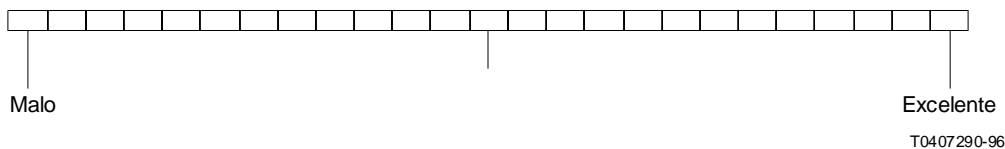


FIGURA B.3/P.910

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)

Reemplazada por una versión más reciente

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. Véase en (Bibliografía, 2) más información.

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.

Reemplazada por una versión más reciente

Anexo C

Presentación simultánea de pares de secuencias

(Este anexo es 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.

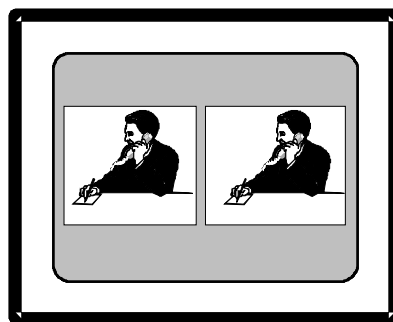
La SP requiere precauciones particulares para que los sujetos puedan evitar el sesgo debido al tipo de presentación.

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 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.



T1207510-95

FIGURA C.1/P.910

Posición relativa de las dos secuencias en SP

Reemplazada por una versión más reciente

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.

Apéndice I

Secuencias de prueba

(Este apéndice no es 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 BetacamSP);
- los valores de la información espacial y la información temporal (definidas en 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 el Informe 1213 del CCIR [10].

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/P.910

Secuencias de prueba para evaluación de calidad vídeo en aplicaciones multimedia

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

Reemplazada por una versión más reciente

Secuencia	Categoría	Descripción	Formato fuente	SI	TI
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 es 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, 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 realización específica.

II.1 ACR

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

Reemplazada por una versión más reciente

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.

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

Reemplazada por una versión más reciente

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie B	Medios de expresión
Serie C	Estadísticas generales de telecomunicaciones
Serie D	Principios generales de tarificación
Serie E	Red telefónica y RDSI
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión
Serie H	Transmisión de señales no telefónicas
Serie I	Red digital de servicios integrados (RDSI)
Serie J	Transmisiones de señales radiofónicas y de televisión
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	Mantenimiento: sistemas de transmisión, circuitos telefónicos, telegrafía, facsímil y circuitos arrendados internacionales
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Equipos terminales y protocolos para los servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos y comunicación entre sistemas abiertos
Serie Z	Lenguajes de programación