



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

**UIT-T**

SECTOR DE NORMALIZACIÓN  
DE LAS TELECOMUNICACIONES  
DE LA UIT

**P.911**

(12/98)

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

Calidad audiovisual en servicios multimedios

---

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

Recomendación UIT-T P.911

(Anteriormente Recomendaciones del CCITT)

---

RECOMENDACIONES UIT-T DE LA SERIE P

**CALIDAD DE TRANSMISIÓN TELEFÓNICA, INSTALACIONES TELEFÓNICAS Y REDES LOCALES**

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 de evaluación objetiva y subjetiva de la calidad	Serie	P.80 P.800
<b>Calidad audiovisual en servicios multimedios</b>	<b>Serie</b>	<b>P.900</b>

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

### **MÉTODOS DE EVALUACIÓN SUBJETIVA DE LA CALIDAD AUDIOVISUAL PARA APLICACIONES MULTIMEDIOS**

#### **Resumen**

La presente Recomendación describe métodos de evaluación subjetiva no interactivos para evaluar la calidad audiovisual global en un sentido para aplicaciones multimedia como videoconferencia, aplicaciones de almacenamiento con recuperación, aplicaciones de telemedicina, etc. Estos métodos se pueden utilizar con diversos fines, por ejemplo, la selección de algoritmos, la clasificación de las prestaciones de sistemas audiovisuales y la evaluación del nivel de calidad durante una conexión audiovisual, sin que esta relación de objetivos sea exhaustiva. Cuando se evalúen aspectos interactivos se utilizarán los métodos de prueba conversacionales descritos en la Recomendación P.920. La presente Recomendación también destaca las características de las secuencias fuente a utilizar, como duración, tipo de contenido, número de secuencias, etc. Finalmente, aporta indicaciones sobre la relación entre la calidad de audio, de vídeo y audiovisual, derivadas de los resultados de las pruebas llevadas a cabo de forma independiente en diferentes laboratorios.

#### **Orígenes**

La Recomendación UIT-T P.911 ha sido preparada por la Comisión de Estudio 12 (1997-2000) del UIT-T y fue aprobada por el procedimiento de la Resolución N.º 1 de la CMNT el 3 de diciembre de 1998.

## 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.

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 *empresa de explotación reconocida (EER)* designa a toda persona, compañía, empresa u organización gubernamental que explote un servicio de correspondencia pública. Los términos *Administración*, *EER* y *correspondencia pública* están definidos en la *Constitución de la UIT (Ginebra, 1992)*.

## 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.

© UIT 1999

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.

## ÍNDICE

	<b>Página</b>
1 Alcance.....	1
2 Referencias.....	1
3 Términos y definiciones.....	2
4 Abreviaturas.....	3
5 Señal fuente.....	4
6 Métodos de prueba y diseño experimental.....	5
6.1 Índices por categorías absolutas (ACR, <i>absolute category rating</i> ).....	5
6.2 Índices por categorías de degradación (DCR, <i>degradation category rating</i> ).....	7
6.3 Método de comparación por pares (PC, <i>pair comparison</i> ).....	7
6.4 Evaluación continua de estímulo único de la calidad (SSCQE).....	8
6.5 Comparación de los métodos.....	9
6.6 Condiciones de referencia.....	9
6.7 Diseño experimental.....	9
7 Procedimientos de evaluación.....	10
7.1 Condiciones de observación y de escucha.....	10
7.2 Sistema de procesamiento y reproducción.....	11
7.3 Participantes.....	12
7.4 Instrucciones a los participantes y sesión de instrucción.....	12
8 Análisis estadístico y notificación de los resultados.....	12
Anexo A – Detalles relacionados con la caracterización de las secuencias de prueba.....	13
Anexo B – Clases de vídeo y audio y sus atributos.....	14
B.1 Clases de vídeo y sus atributos.....	14
B.2 Clases de audio y sus atributos.....	15
Anexo C – Consideraciones sobre la relación entre calidad de audio, de vídeo y audiovisual	17
Apéndice I – Secuencias audiovisuales de prueba.....	18
Apéndice II – Instrucciones para las pruebas audiovisuales.....	19
II.1 ACR.....	19
II.2 DCR.....	19
II.3 PC.....	20
Apéndice III – Bibliografía.....	20



## Recomendación P.911

# MÉTODOS DE EVALUACIÓN SUBJETIVA DE LA CALIDAD AUDIOVISUAL PARA APLICACIONES MULTIMEDIOS

(Ginebra, 1998)

## 1 Alcance

Esta Recomendación tiene por objeto definir métodos de evaluación subjetiva no interactivos para determinar la calidad audiovisual global en un sentido para las velocidades binarias especificadas en clases TV 3, MM 4, MM 5 y MM 6, como se indica en los cuadros B.2 y B.4, para aplicaciones tales como videoconferencia, almacenamiento con recuperación, telemedicina, etc. Los métodos se pueden utilizar con diversos fines, por ejemplo, la selección de algoritmos, la clasificación de la calidad de funcionamiento de sistemas audiovisuales y la evaluación del nivel de calidad durante una conexión audiovisual, sin que esta relación de objetivos sea exhaustiva. Cuando se evalúen aspectos interactivos se utilizará la Recomendación P.920.

## 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.

- [1] Recomendación UIT-T P.930 (1996), *Principios de un sistema de degradaciones de referencia para vídeo.*
- [2] Recomendación UIT-T P.920 (1996), *Métodos de prueba interactivos para comunicaciones audiovisuales.*
- [3] Recomendación UIT-R BT.601-5 (1995), *Parámetros de codificación de televisión digital para estudios con formatos de imagen normal 4:3 y de pantalla ancha 16:9.*
- [4] Recomendación UIT-R BT.500-8 (1998), *Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión.*
- [5] Publicación 60268-13 de la CEI/TR 3 (1998), *Sound system equipment – Part 13: Listening tests on loudspeakers.*
- [6] Recomendación UIT-T P.800 (1996), *Métodos de determinación subjetiva de la calidad de transmisión.*
- [7] Recomendación UIT-R BT.814-1 (1993), *Especificaciones y procedimientos de ajuste para establecer el brillo y el contraste en las pantallas.*
- [8] Recomendación UIT-R BT.1128-2 (1997), *Evaluación subjetiva de los sistemas de televisión convencional.*
- [9] Recomendación CCITT J.61 (1990), *Calidad de transmisión de los circuitos de televisión diseñados para ser utilizados en conexiones internacionales.*

- [10] Recomendación UIT-T P.810 (1996), *Aparato de referencia para ruido modulado*.
- [11] Recomendación UIT-T P.910 (1996), *Métodos de evaluación subjetiva de la calidad vídeo para aplicaciones multimedia*.
- [12] Recomendación CCITT G.722 (1988), *Codificación de audio de 7 kHz dentro de 64 kbit/s*.
- [13] Recomendación CCITT G.711 (1988), *Modulación por impulsos codificados (MIC) de frecuencias vocales*.
- [14] Recomendación CCITT G.728 (1992), *Codificación de señales vocales a 16 kbit/s utilizando predicción lineal con excitación por código de bajo retardo*.
- [15] Recomendación UIT-T G.114 (1996), *Tiempo de transmisión en un sentido*.
- [16] Recomendación UIT-R BS.775-1 (1993), *Sistema de sonido estereofónico multicanal con y sin acompañamiento de imagen*.
- [17] Publicación 60651 de la CEI (1979), *Sound level meters*.
- [18] Recomendación UIT-T P.931 (1998), *Retardo de las comunicaciones multimedia, sincronización y medición de la velocidad de tramas*.

### 3 Términos y definiciones

En esta Recomendación se definen los términos siguientes.

**3.1 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, CIF, QCIF, SIF, etc.). En el cuerpo de esta Recomendación, se omitirán los términos "explícito" y "fuente" siempre que en el contexto esté claro el significado de "referencia".

**3.2 gamma:** 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.

**3.3 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.

**3.4 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.

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

**3.6 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.

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

**3.8 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 comunicación. 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 ( $máx_{time}$ ) como representación del contenido de información espacial de la escena. Este proceso se puede representar en forma de ecuación como sigue:

$$SI = máx_{time} \{std_{space} [Sobel (F_n)]\}$$

**3.9 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 comunicación.

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

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

donde  $M_n(i,j)$  es la diferencia entre los valores de píxels en la misma ubicación en la trama, pero pertenecientes a dos tramas sucesivas, es decir:

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

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

**3.10 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 de distancia. Si no existen diferencias (es decir, distancia = 0), 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.

**3.11 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.

## 4 Abreviaturas

En esta Recomendación se utilizan las siguientes siglas.

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 × 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 × 144 píxels)
SI	Información espacial ( <i>spatial information</i> )
SIF	Formato intermedio normalizado ( <i>standard intermediate format</i> ) [formatos de imagen definidos en ISO/CEI 11172 (MPEG-1): 352 líneas × 288 píxels × 25 tramas/s y 352 líneas × 240 píxels × 30 tramas/s]
S/N	Relación señal/ruido ( <i>signal-to-noise ratio</i> )
SP	Presentación simultánea ( <i>simultaneous presentation</i> )
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, la secuencia de prueba se debe definir de acuerdo con el objetivo de la prueba y registrar en un sistema de almacenamiento digital. Se deben elegir escenas de prueba audiovisuales razonables y pertinentes para que ambas partes de vídeo y de audio sean consecuentes con los servicios que se pretende que proporcione el canal de servicio de transmisión digital. El conjunto de escenas de prueba debe contemplar la gama completa de información espacial y temporal e incluir cualquier tipo de señales de audio de interés para los usuarios de los dispositivos sometidos a prueba.

La duración de la secuencia fuente debe ser de unos 10 s, pero no inferior a 8 s. Ésta debe ser la duración real de la secuencia, lo que significa que las secuencias no se pueden obtener mediante repetición de secuencias más cortas. La finalización de una escena no debe interrumpir una frase oral o musical. Un periodo de silencio al principio y al final, que no dure más de 500 ms, puede hacer que la secuencia suene de forma más natural.

La calidad de las secuencias fuente debe ser tan alta como sea posible: la señal de vídeo debe registrarse en el formato UIT-R BT.601 4:2:2 y la de audio debe registrarse con una velocidad de muestreo de 48 kHz y por lo menos con 16 bits por muestra. Cuando el contenido lo requiera, el audio y el vídeo se deben sincronizar.

El anexo A enumera las categorías de audio y de vídeo que se pueden utilizar para caracterizar una secuencia audiovisual.

Cuando el experimentador esté interesado en comparar los resultados obtenidos por diferentes laboratorios, es necesario utilizar un conjunto común de secuencias fuente para eliminar otras fuentes de variación.

Las características del entorno de grabación y de los sistemas de grabación se deben definir de conformidad con 5.1/P.910 y 5.2/P.910 y B.1.1/P.800 y B.1.3/P.800, para las partes de vídeo y de audio respectivamente. En el anexo C se dan ejemplos de escenas de prueba adecuadas.

El número de secuencias debe estar definido en función del diseño experimental. Para evitar aburrir a los observadores y para lograr una fiabilidad mínima de los resultados, se deben elegir secuencias con cuatro tipos diferentes de escenas por lo menos (es decir, temas de asunto diferente).

## **6 Métodos de prueba y diseño experimental**

La medición de la calidad percibida de secuencias audiovisuales requiere la utilización de métodos de escala subjetiva. La condición para que estas mediciones sean significativas es que exista una relación entre las características físicas del "estímulo", en este caso la secuencia audiovisual presentada a los participantes en una prueba, y la magnitud y naturaleza de la sensación producida por el estímulo, en este caso la calidad audiovisual global.

Se han validado diversos métodos experimentales con distintos objetivos. Aquí se recomiendan tres métodos para la gama de velocidades binarias y las aplicaciones indicadas en el ámbito de la presente Recomendación.

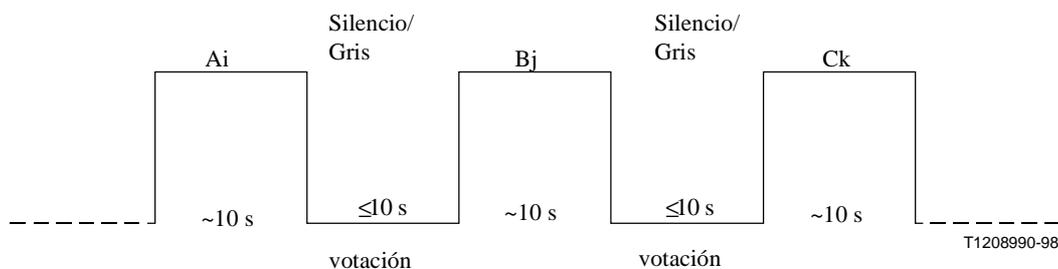
La elección final de uno de esos métodos para una aplicación determinada depende de varios factores, tales como el contexto, la finalidad y dónde se debe llevar a cabo la prueba en proceso de desarrollo.

### **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 tras otra y se califican de forma independiente en una escala de categorías. (Este protocolo es similar a los protocolos descritos en otras Recomendaciones UIT-T y UIT-R [11], [6] y [4]. En [4], este tipo de protocolo se indica como método de estímulo único.)

El método especifica que, después de cada presentación, se invite a los participantes a que evalúen la calidad de la secuencia presentada. A los participantes no se les da ninguna referencia explícita, pero ellos siempre utilizarán una referencia implícita.

El diagrama de tiempos para la presentación de estímulos se ilustra en la figura 1. El tiempo de votación debe ser inferior o igual a 10 s, dependiendo del mecanismo de votación utilizado. El tiempo de presentación se puede reducir o ampliar en función del 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 condición de prueba k

**Figura 1/P.911 – Presentación del estímulo en el método ACR**

Para evaluar la calidad global se debe utilizar la siguiente escala de cinco niveles (véase el cuadro 1).

**Cuadro 1/P.911 – Escala de calidad de cinco niveles**

5	Excelente
4	Buena
3	Aceptable
2	Mediocre
1	Mala

Si se requiere una evaluación más discriminativa, como en el caso de una evaluación de codificación con velocidad binaria baja, debe utilizarse la escala de nueve niveles siguiente (véase el cuadro 2).

**Cuadro 2/P.911 – Escala de calidad de nueve niveles**

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

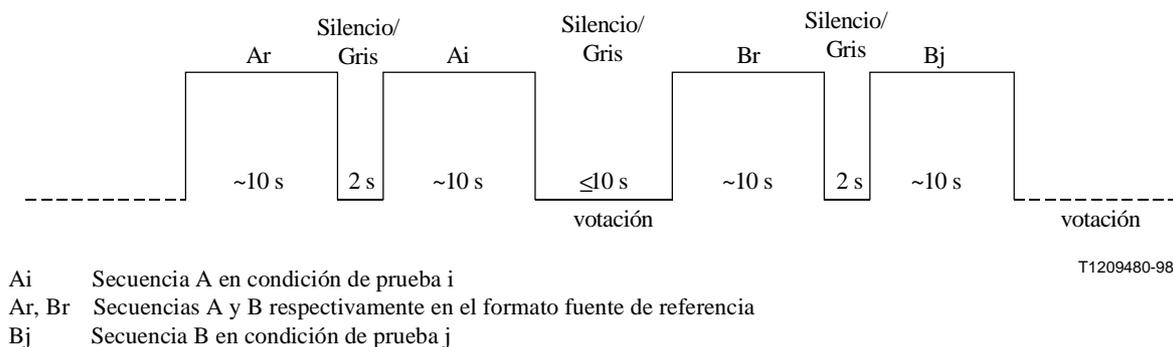
En el anexo B/P.910 figuran ejemplos de escalas apropiadas discretas o continuas. En ese anexo se dan también ejemplos de dimensiones de evaluación distintas de la calidad global. Dichas dimensiones pueden ser de utilidad para obtener más información sobre diferentes factores de calidad de percepción 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 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 sometidos a prueba. (Este protocolo es similar a los protocolos descritos en otras Recomendaciones UIT-T y UIT-R [11], [6] y [4]. En [4] se hace referencia a este tipo de protocolo como método de escala de degradación con doble estímulo.)

En la figura 2 se ilustra el diagrama de tiempos de la presentación de estímulos. El tiempo de votación debe ser igual o inferior a 10 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.



**Figura 2/P.911 – Presentación del estímulo en el método DCR**

En este caso se invita a los participantes 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 (véase el cuadro 3).

**Cuadro 3/P.911 – Escala de evaluación 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.

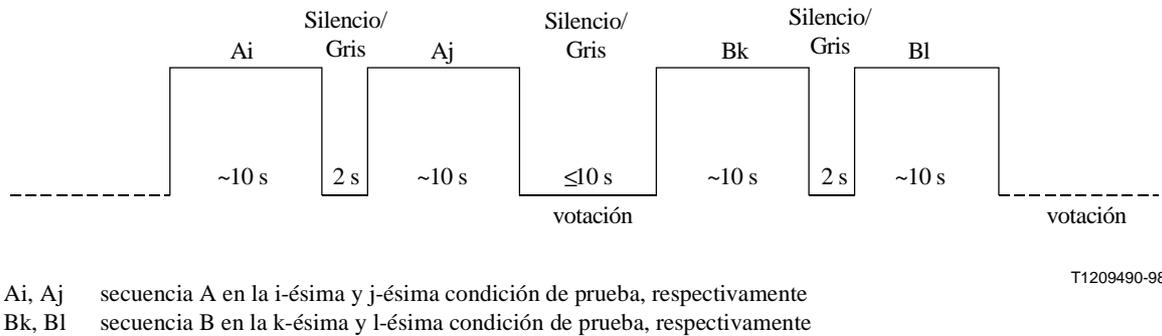
## 6.3 Método de comparación por pares (PC, *pair comparison*)

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. Se puede incluir la secuencia fuente que sería tratada como un sistema adicional sometido a prueba.

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 s y se puede reducir o aumentar en función del contenido del material de prueba.



**Figura 3/P.911 – Presentación del estímulo en el método PC**

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 variante 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 [6].

#### 6.4 Evaluación continua de estímulo único de la calidad (SSCQE)

El método de evaluación continua de estímulo único de la calidad (SSCQE, *single stimulus continous quality evaluation*) se ha incluido en [4] para tener en cuenta las variaciones temporales de la calidad en transmisiones digitales.

El método considera secuencias de larga duración (de 3 a 30 min). Cuando se presentan las secuencias, la evaluación de la calidad es continua. La evaluación continua se realiza mediante deslizadores que los participantes deben mover mientras observan y/o escuchan los programas o escenarios. Es importante destacar que no existe referencia para la determinación de la evaluación subjetiva.

Los resultados se presentan mediante curvas impresas que indican el porcentaje de tiempo durante el cual la calificación subjetiva ha sido superior a un valor determinado, en una escala de 0 a 100. En esta escala, 100 representa la calidad perfecta para el servicio o servicios considerados.

La ventaja de este método es que resulta adecuado para la consideración de variaciones temporales de la calidad y para la obtención de estimaciones de calidad globales. Permite la comparación entre servicios cuando se pueden parametrizar estos servicios. Por lo tanto, el método puede ser un elemento útil para la definición de un servicio. Además, este método se puede utilizar tanto en un laboratorio como en el domicilio del usuario. Una desventaja de este método es que carece de referencia. Esto hace que no resulte adecuado para pruebas que requieran un alto grado de discriminación, por ejemplo, para la comparación entre codificadores.

Se está estudiando la conveniencia del método de evaluación continua de estímulo único de la calidad (SSCQE) para la gama de velocidades binarias y para el tipo de aplicaciones consideradas en la presente Recomendación.

## **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, PC y SSCQE). Con esa segunda clase de métodos no se prueba la fidelidad.

Cuando se pruebe la fidelidad de transmisión con respecto a la señal fuente, se debe utilizar el método DCR. Éste es con mucha frecuencia un factor importante en la evaluación de sistemas de alta calidad. 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 debe 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. Por ello, ACR es el método adecuado para las pruebas de calificación.

El mérito principal del método PC es su gran capacidad de discriminación, 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 las Recomendaciones P.930 [1] y P.810 [10], para vídeo y audio respectivamente. 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. (véase la 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 (tres o cuatro en algunos casos) 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**

### **7.1 Condiciones de observación y de escucha**

El cuadro 4 enumera las condiciones típicas de observación y de escucha utilizadas para la estimación de la calidad audiovisual. 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 y de escucha 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 los transductores de visualización y de audio, por ejemplo densidad de puntos y de la pantalla, tipo de tarjeta de visualización de vídeo utilizada, características de los auriculares, cascos y altavoces, etc.

En particular, en el caso de presentación mediante altavoces, debe indicarse la cantidad de altavoces y su posición en relación con la imagen.

Los parámetros operacionales, tales como el nivel de señal, de las secuencias de pruebas deberán concordar con los de la señal de ajuste utilizada para verificar las condiciones de observación y escucha. Deberá comunicarse cualquier modificación operativa efectuada para que las secuencias fuente o las secuencias procesadas cumplan este requisito.

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).

La sincronización entre la señal de audio y la de vídeo deberá medirse de conformidad con la Recomendación P.931 [18], y habrá que informar al respecto.

**Cuadro 4/P.911 – Condiciones típicas de observación y de escucha utilizados en la evaluación de la calidad audiovisual**

Parámetro	Valores
Tamaño de la habitación (nota 7)	Especificar L × W × H
Distancia de observación (nota 5)	1-8 H
Valor de cresta de luminancia de la pantalla	100-200 cd/m <sup>2</sup>
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 1)	≤ 0,2
Cromaticidad del fondo (nota 6)	D <sub>65</sub>
Iluminación de fondo de la habitación (nota 1)	≤ 20 lux
Nivel de ruido de fondo (nota 2)	≤ 30 dBA
Nivel de escucha (nota 3)	~80 dBA
Tiempo de reverberación (nota 4)	< 500 ms, ∀f > 150 Hz
<p>NOTA 1 – Este valor indica un ajuste que permite la máxima detectabilidad de las distorsiones, se permiten valores superiores para algunas aplicaciones o vienen determinados por la aplicación.</p> <p>NOTA 2 – Si los niveles de ruido de fondo utilizados en la aplicación son significativamente más altos, debe utilizarse el ruido Hoth (véase la Bibliografía [9]) para entornos de tipo oficina. El ruido Hoth se debe generar en una habitación con niveles de ruido de fondo bajos (≤ 30 dBA) y se deben medir en el dominio acústico. Si para algún tipo especial de aplicación se utiliza otro tipo de ruido de fondo, se debe especificar la densidad espectral de potencia y el nivel en dBA.</p> <p>NOTA 3 – Este valor indica un ajuste que permite la máxima detectabilidad de las distorsiones, para algunas aplicaciones se admiten valores menores o mayores. El ajuste de nivel se mide utilizando el valor máximo en la secuencia de audio mediante el promediado rápido de la publicación CEI [17]. Cuando se utilizan altavoces, se podría ajustar el nivel de sonido en función de la distancia de observación (véase la Bibliografía [11]).</p> <p>NOTA 4 – Este valor sólo es importante para presentaciones con altavoces, tiempos de reverberación superiores conducen generalmente a una disminución en la detectabilidad de las distorsiones.</p> <p>NOTA 5 – 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.</p> <p>NOTA 6 – Para monitores de PC la cromaticidad del fondo puede adaptarse a la cromaticidad del monitor.</p> <p>NOTA 7 – El tamaño de la habitación es importante sólo para presentaciones con altavoces.</p>	

## 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 audiovisuales en tiempo real a través de los sistemas sometidos a prueba, mientras los participantes observan, escuchan 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 grabar la señal en formato digital para reducir al mínimo las degradaciones que se pueden producir en el proceso de grabación. Las condiciones de prueba se pueden grabar en un dispositivo de almacenamiento componente de la Recomendación UIT-R BT.601. En cualquier caso, teniendo en cuenta que las degradaciones introducidas por esquemas de codificación de baja velocidad binaria son generalmente más evidentes que las degradaciones introducidas por dispositivos de calidad profesional, se pueden utilizar VTR tales como D2, MII y BetacamSP.

### **7.3 Participantes**

El número posible de participantes en una prueba de observación y de escucha (así como en pruebas de utilidad en terminales o servicios) varía de 6 a 40. Cuatro es el mínimo absoluto por razones estadísticas, mientras que difícilmente se obtengan mayores ventajas con más de 40.

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 participantes como mínimo. No deben intervenir directamente en evaluaciones de calidad de imagen ni de audio como parte de su trabajo habitual y no han de ser evaluadores experimentados.

Habitualmente, antes de una sesión, se debe examinar a los observadores para determinar su agudeza visual normal o corregida a normal y su visión normal de los colores.

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

### **7.4 Instrucciones a los participantes y sesión de instrucción**

Antes de comenzar el experimento, se debe explicar a los participantes 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 audiovisuales distintas de las utilizadas en las pruebas reales.

No se deberá deducir que la peor calidad experimentada 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 5 para el método ACR.

**Cuadro 5/P.911 – 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
<b>Condición</b>	Clasificación que indica una combinación de variables de prueba										
<b>Total de votos</b>	Cantidad de votos recogidos para esa condición										
<b>Excelente, aceptable ... mala</b>	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 audiovisual 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 , y en [6].

## ANEXO A

### Detalles relacionados con la caracterización de las secuencias de prueba

La caracterización de secuencias audiovisuales se basará en características de contenido tanto de audio como de vídeo. Los cuadros A.1 y A.2 ilustran las categorías de las partes de vídeo y de audio respectivamente. Los cuadros no pretenden ser exhaustivos.

Según esta clasificación una escena de videoteléfono pertenecerá a la categoría A-I.

**Cuadro A.1/P.911 – Categorías del contenido de vídeo**

Categoría	Descripción
A	Una persona, principalmente cabeza y hombros, detalle y movimiento limitados
B	Una persona con gráficos y/o más detalles
C	Más de una persona
D	Gráficos con indicación
E	Gran movimiento del objeto y/o la cámara, más allá de la gama utilizada normalmente en videoteleconferencias

**Cuadro A.2/P.911 – Categorías del contenido de audio**

<b>Categoría</b>	<b>Descripción</b>
I	Conversación/un orador
II	Conversación/varios oradores
III	Conversación con música de fondo
IV	Música/un instrumento
V	Música/varios instrumentos

**ANEXO B****Clases de vídeo y audio y sus atributos****B.1 Clases de vídeo y sus atributos**

En la presente Recomendación, la calidad de vídeo más alta considerada es la de la Recomendación UIT-R BT.601, vídeo codificado MIC lineal de 8 bit/muestra en formato 4:2:2, Y, C<sub>R</sub>, C<sub>B</sub>.

Véanse los cuadros B.1 y B.2.

**Cuadro B.1/P.911 – Definiciones de las clases de vídeo**

TV 0	Sin pérdidas: vídeo de la Recomendación 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 le 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 le compara con TV 2.
MM 4	Todas las tramas están codificadas. Pocas perturbaciones en relación con TV 3. Videoconferencia de calidad media. Normalmente $\geq 25$ trama/s para sistemas de 625 líneas y $\geq 30$ trama/s para sistemas de 525 líneas.
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 B.2/P.911 – Atributos de clases de vídeo**

<b>Clases de vídeo</b>	<b>Formato espacial</b>	<b>Velocidad de trama entregada (Nota 1)</b>	<b>Latencia típica Variación de retardo (nota 2)</b>	<b>Velocidad binaria de vídeo nominal, Mbit/s</b>
TV 0	Rec. UIT-R BT.601	FR Máx	(Nota 2)	270
TV 1	Rec. UIT-R BT.601	FR Máx	(Nota 2)	18 a 50
TV 2	Rec. UIT-R BT.601	FR Máx	(Nota 2)	10 a 25
TV 3	Rec. UIT-R BT.601	FR Máx ocasional Repetición de trama	(Nota 2)	1,5 a 8
MM 4a	Rec. UIT-R BT.601	~30 ó ~25 trama/s	Retardo $\approx$ 150 ms Variación $\leq$ 50 ms	~1,5
MM 4b	CIF	~30 ó ~ 25 trama/s	Retardo $\approx$ 150 ms Variación $\leq$ 50 ms	~0,7
MM 5a	CIF	10-30 trama/s	Retardo $\approx$ 1000 ms Variación $\leq$ 500 ms	~0,2
MM 5b	$\leq$ CIF	1-15 trama/s	Retardo $\approx$ 1000 ms Variación $\leq$ 500 ms	~0,05
MM 6a	CIF-16CIF	Límite $\rightarrow$ 0 trama/s	Sin restricciones (Nota 3)	$<0,05$ , Límite $\rightarrow$ 0 trama/s
MM 6b	CIF-16CIF	Límite $\rightarrow$ 0 trama/s	Sin restricciones (Nota 3)	$<0,05$ , Límite $\rightarrow$ 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 [15]. Se permiten variaciones de retardo dentro de una gama dada pero no se deben llegar a percibir efectos perturbadores dinámicos.

NOTA 3 – Difiere únicamente en la velocidad binaria de audio.

## **B.2 Clases de audio y sus atributos**

En la presente Recomendación, el nivel más alto de calidad de audio considerado es 5,1 de sonido ambiente, audio codificado con MIC lineal de 20 bit/muestra [16]. Al referirse a las clases de audio debe indicarse la disposición del canal, por ejemplo, mono, estéreo, 3 canales, canales 3/1, etc.

Véanse los cuadros B.3 y B.4.

**Cuadro B.3/P.911 – Definiciones de clases de audio**

TV 0	Calidad de estudio, 20 bits/muestra, 48 kHz sin pérdidas, MIC lineal.
TV 1	Utilizada para posproducción completa, muchos montajes y capas de procesamiento, transmisión entre plantas. Se utiliza también para transmisión entre lugares distantes y la planta. Perceptualmente transparente cuando se compara con TV 0.
TV 2	Distribución primaria: Utilizada para modificaciones sencillas, pocos montajes, mezcla de programas y transmisión entre facilidades. Un ejemplo de difusión sería la transmisión de red a afiliado. Otros ejemplos son un enlace regional de sistemas de cable a un extremo local y un sistema de videoconferencia de alta calidad. Casi perceptualmente transparente cuando se compara con TV 0.
TV 3	Utilizado para entregas a domicilio/consumidor (sin cambios). Otros ejemplos son sistema de cable entre un extremo local y un domicilio y videoconferencia de calidad media a alta. Aparecen pequeñas perturbaciones audibles cuando se compara con TV 2.
MM 4	Pequeñas perturbaciones audibles en comparación con TV 3, utilizando conversación y audio. Videoconferencia de calidad media. Normalmente utiliza toda la anchura de banda de audio (20-20 000 Hz), pero se pueden aceptar anchuras de banda hasta banda ancha (50-7000 Hz) (Nota).
MM 5	Pequeñas perturbaciones audibles en relación con una referencia de banda estrecha (banda de telefonía de 300-3400 Hz) utilizando conversación y música. Posibles perturbaciones perceptibles, pero el nivel de calidad puede ser útil para tareas designadas, por ejemplo, videoconferencia de baja calidad.
MM 6	Perturbaciones audibles severas en relación con una aplicación de telefonía de banda estrecha (300-3400 Hz). La conversación sigue sin embargo siendo inteligible.
NOTA – La anchura de banda se define como una anchura de banda perceptible (es decir, la anchura de banda del sistema sometido a prueba sin tener en cuenta componentes inaudibles) del sistema sometido a prueba.	

**Cuadro B.4/P.911 – Atributos de las clases de audio**

<b>Clases de audio</b>	<b>Frecuencia de muestreo</b>	<b>Latencia típica Variación de retardo (Nota 1)</b>	<b>Velocidad binaria nominal, kbit/s/canal (Nota 2)</b>
TV 0	48,0	(Nota 1)	960
TV 1	44,1 ó 48,0 ó 32,0	(Nota 1)	250-500
TV 2	44,1 ó 48,0 ó 32,0	(Nota 1)	120-300
TV 3	44,1 ó 48,0 ó 32,0	(Nota 1)	50-150
MM 4a	44,1 ó 48,0 ó 32,0	Retardo $\leq 150$ ms Variación $\leq \pm 20$ ms	20-100
MM 4b	12-32	Retardo $\leq 150$ ms Variación $\leq \pm 20$ ms	10-50
MM 5a	12-20	Retardo $\leq 400$ ms (Nota 3) Variación $\leq \pm 100$ ms	4-32
MM 5b	8-12	Retardo $\leq 400$ ms (Nota 3) Variación $\leq \pm 200$ ms	2-16

**Cuadro B.4/P.911 – Atributos de las clases de audio (*fin*)**

<b>Clases de audio</b>	<b>Frecuencia de muestreo</b>	<b>Latencia típica Variación de retardo (Nota 1)</b>	<b>Velocidad binaria nominal, kbit/s/canal (Nota 2)</b>
MM 6a	8	Retardo $\leq 400$ ms (Nota 4) Variación $\leq \pm 200$ ms	<8
MM 6b	8	Retardo $\leq 4000$ ms (Nota 4) Variación $\leq \pm 2000$ ms	<8

NOTA 1 – Todos los sistemas de difusión tienen latencia 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 videoconferencia de alta calidad, y en general para tipos de aplicaciones de conversación, se prefieren latencias unidireccionales inferiores a 150 ms. Se permiten variaciones de retardo en la gama dada, pero no se deben llegar a percibir efectos perturbadores dinámicos.

NOTA 2 – El canal de mejora de baja frecuencia (presentación 0,1 en 5,1) requiere sólo una velocidad binaria un poco más alta. Para una señal de audio de N canales se introducirán correlaciones mutuas, dando como resultado una velocidad binaria total que puede ser significativamente inferior a N veces la velocidad de cada canal.

NOTA 3 – Para esta aplicación, la sincronización entre el audio y el vídeo puede no implicar la calidad más alta de conversación en el enlace de comunicación.

NOTA 4 – Conferencias audiográficas.

## ANEXO C

### **Consideraciones sobre la relación entre calidad de audio, de vídeo y audiovisual**

La parte principal de la presente Recomendación trata de la calidad audiovisual global en un sentido de secuencias audiovisuales. Las relaciones entre las calidades de audio, de vídeo y audiovisual se encuentran en estudio, pero se dispone de algunos resultados estables. El más importante es la correspondencia entre las calidades de audio en un sentido y de vídeo en un sentido, obtenidas a partir de experimentos subjetivos sólo de audio y sólo de vídeo, y la calidad audiovisual global en un sentido. Cuatro laboratorios diferentes encontraron resultados de correspondencia similares a pesar de que las condiciones experimentales eran bastante diferentes ([13]-[17] de Bibliografía). Se realizó una serie de experimentos ([13] y [14] de Bibliografía) utilizando borrosidad para el vídeo y ruido multiplicativo para degradaciones de audio. En [15] de Bibliografía se describe un experimento similar con televisiones comerciales utilizando la calidad de difusión de televisión como la calidad más alta de referencia de audio y de vídeo. Se utilizaron distorsiones similares para las degradaciones de vídeo (VIRIS 601, ANSI [8] extensión V3, V6, V8) pero limitando en banda las distorsiones de audio (desde la anchura de banda completa, 20-20 000 Hz, hasta banda estrecha 300-3400 Hz). En [16] de Bibliografía se utilizó una evaluación de calidad de teleconferencia de vídeo en un sentido con varios tipos de distorsiones de codificación. Se utilizaron tres códecs de vídeo diferentes (dos propios y uno H.261) y cuatro códecs de audio diferentes. En [17] de Bibliografía se utilizó también una evaluación de calidad de teleconferencia en un sentido con varios tipos de distorsiones de codificación. En este caso se introdujeron además condiciones no codificadas (banda completa para audio y señal PAL para vídeo). Los resultados de los experimentos, que utilizaron las escalas de calidad ACR de nueve niveles o de cinco niveles indicadas en 6.1, muestran que:

- 1) El vídeo domina la percepción global. Cuando las varianzas de la calidad de audio y de la calidad de vídeo son similares, la varianza de la calidad audiovisual global está dominada por la varianza de la calidad de vídeo. La correlación entre la calidad de vídeo y la audiovisual global es mayor que la correlación entre la calidad de audio y la audiovisual global.
- 2) Se puede predecir la calidad audiovisual global en un sentido a partir de la calidad de audio en un sentido y de la de vídeo en un sentido obtenidas de los experimentos subjetivos sólo de audio y sólo de vídeo.
- 3) La correspondencia más estable en los laboratorios entre las calidades de audio y de vídeo tomadas por separado y la calidad audiovisual global demostró ser  $MOS_{av} = \alpha + \beta * MOS_a * MOS_v$ . Esta correspondencia se basa en cuatro conjuntos de experimentos subjetivos y la correlación entre la calidad audiovisual global predicha y la medida varió entre 0,93 a 0,99. El valor de  $\alpha$  varió entre 1,1 y 1,5 y los valores de  $\beta$  entre 0,107 y 0,121, cuando todas las escalas se hacen corresponder en una escala de nueve valores. De forma provisional, se recomiendan los valores 1,3 para  $\alpha$  y 1,1 para  $\beta$ .

La validez del modelo se aprobó únicamente para señales sincronizadas de audio y de vídeo, y sólo dentro de los límites del experimento en que las distorsiones de audio y vídeo son aproximadamente igual de perturbadoras. Es posible diseñar un experimento en el que la relación (ítem 3) no fuese válida. Por ejemplo, si los tratamientos incluyen una gama de calidades de audio de las que se pretende determinar la inteligibilidad, y una pequeña gama de calidades de vídeo, cabe esperar que la evaluación de audio domine en la calidad combinada de audio y vídeo.

## APÉNDICE I

### Secuencias audiovisuales de prueba

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 prueba 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 los cuadros A.1 y A.2);
- una breve descripción de la escena;
- el formato fuente (625 ó 525 líneas, formato UIT-R BT.601 o BetacamSP);
- la frecuencia de muestreo de la señal de audio;
- los valores de la información espacial y temporal (definidas en 5.3.1/P.910 y 5.3.2/P.910 respectivamente).

Todas las secuencias indicadas en el cuadro I.1 son de dominio público y se pueden utilizar libremente para evaluaciones y demostraciones.

**Cuadro I.1/P.911 – Secuencias de prueba para evaluación de  
calidad audiovisual en aplicaciones multimedios  
(Por determinar)**

Secuencia	Categoría (A + V)	Descripción	Formato fuente	Lenguaje	Frecuencia de muestreo	SI	TI
	A-I						
	B-III						
	D-I						
	D-II						
	etc.						

## APÉNDICE II

### Instrucciones para las pruebas audiovisuales

El material que sigue se puede utilizar como base para las instrucciones a 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 implementación específica.

#### II.1 ACR

Buenos días y gracias por estar aquí.

En este experimento se le presentarán secuencias audiovisuales cortas. 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

Su evaluación debe reflejar su opinión sobre la calidad combinada global **de audio y de vídeo**.

Antes de efectuar su valoración observe y escuche cuidadosamente la secuencia completa.

#### II.2 DCR

Buenos días y gracias por estar aquí.

En este experimento se presentarán secuencias audiovisuales. Cada secuencia se presentará dos veces en una sucesión rápida: la primera secuencia es la secuencia de referencia, la segunda secuencia la secuencia procesada. Al final de cada presentación por pares debe usted evaluar la degradación de la segunda secuencia con respecto a la primera. Expresará su apreciación de la diferencia entre la segunda secuencia y la primera utilizando la siguiente escala:

- 5 Imperceptible
- 4 Perceptible pero no molesta
- 3 Ligeramente molesta
- 2 Molesta
- 1 Muy molesta

Su evaluación debe reflejar su opinión sobre la calidad combinada global **de audio y de vídeo**.

Antes de efectuar su valoración observe y escuche cuidadosamente la secuencia completa.

### II.3 PC

Buenos días y gracias por estar aquí.

En este experimento se presentarán secuencias audiovisuales cortas. Cada secuencia se presentará dos veces en sucesión rápida: cada vez a través de un sistema diferente. El orden de las secuencias y la combinación de 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
---	--	---

Su evaluación debe reflejar su opinión sobre la calidad combinada global **de audio y de vídeo**.

Antes de efectuar su valoración observe y escuche cuidadosamente la secuencia completa.

## APÉNDICE III

### Bibliografía

- [1] GONZALEZ (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] Gramm-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.
- [7] GUILFORD (P.): Psychometric methods, MCGRAW-HILL, Nueva York, 1954.
- [8] ANSI T1A1.5 contribution 96-109: *VIRIS for ITU-R 601 digital images*, mayo de 1996.
- [9] HOTH (D.F.): Room noise spectra at subscribers' telephone locations, *J.A.S.A.*, Volume 12, pp. 99-504, abril de 1941.
- [10] BECH (S.): Calibration of Relative Level Differences of a Domestic Multichannel Sound Reproduction System, *Journal of Audio Engineering Society*, abril de 1998.

- [11] ITU-T Delayed Contribution D.40 (1990), *Some speech quality aspects to be considered in multimedia services*, NTT.
- [12] *Handbook on Telephony*, ITU, Ginebra, 1992.
- [13] ITU-T COM 12-20 (diciembre de 1993), *Experimental combined audio/video subjective test method*, Bellcore.
- [14] ITU-T COM 12-37 (septiembre de 1994), *Extension of combined audio/video quality model*, Bellcore.
- [15] ITU-T COM 12-19 (febrero de 1998), *Relations Between Audio, Video and Audiovisual Quality*, KPN Research, Netherlands.
- [16] ITU-T COM 12-64 (noviembre de 1998), *Results of an audiovisual desktop video teleconferencing subjective experiment*, USA.
- [17] ITU-T COM 12-61 (noviembre de 1998), *Study of the influence of experimental context on the relationship between audio, video and audiovisual subjective quality*, France Telecom/CNET.



## SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie B	Medios de expresión: definiciones, símbolos, clasificación
Serie C	Estadísticas generales de telecomunicaciones
Serie D	Principios generales de tarificación
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 multimedios
Serie I	Red digital de servicios integrados
Serie J	Transmisiones de señales radiofónicas, de televisión y de otras señales multimedios
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	RGT y mantenimiento de redes: 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
<b>Serie P</b>	<b>Calidad de transmisión telefónica, instalaciones telefónicas y redes locales</b>
Serie Q	Conmutación y señalización
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 y comunicación entre sistemas abiertos
Serie Y	Infraestructura mundial de la información
Serie Z	Lenguajes de programación