



Secretaría de Estado de Telecomunicaciones
y para la Sociedad de la Información



CALIDAD DE VÍDEO EN ALTA DEFINICIÓN

Versión 1.0

Elaborado por

**Grupo Técnico del Foro de la Televisión
de Alta Definición en España**

Coordinado por
UNIVERSIDAD POLITÉCNICA DE MADRID

Marzo de 2008

Índice

1	OBJETIVO.....	3
2	INTRODUCCIÓN	4
3	METODOLOGÍA DEL ANÁLISIS DE CALIDAD	5
3.1	MÉTODOS DE CALIDAD SUBJETIVA	5
3.2	MÉTODOS DE CALIDAD OBJETIVA.....	6
4	ANÁLISIS DE CALIDAD SUBJETIVA	7
4.1	CARACTERÍSTICAS COMUNES DE LOS SISTEMAS DE EVALUACIÓN	7
4.2	DESCRIPCIÓN DE LOS DISTINTOS MÉTODOS DE EVALUACIÓN	8
4.3	COMPARACIÓN DE LOS MÉTODOS DE EVALUACIÓN SUBJETIVA.....	17
5	ANÁLISIS DE CALIDAD OBJETIVA	19
5.1	MÉTRICAS CON BASE MATEMÁTICA.....	19
5.2	VOEG	21
5.2.1	PSNR	22
5.2.2	CPqD.....	23
5.2.3	Tektronix/Sarnof	23
5.2.4	NHK/Mitsubishi Electric Corp.	23
5.2.5	KDD.....	23
5.2.6	École Polytechnique fédérale de Lausanne.....	24
5.2.7	NASA	24
5.2.8	KPN/Swisscom CT.....	25
5.2.9	NTIA	26
6	VALORACIÓN DE LA CONVIVENCIA ENTRE SD Y HD	27
6.1	CONCLUSIONES DE LA COMPARATIVA SD vs. SD	27
6.2	CONCLUSIONES DE LA COMPARATIVA SD vs. HD	29
6.3	CONCLUSIONES DE LA COMPARATIVA DE LA CRITICIDAD DE LAS IMÁGENES	30
6.4	CONCLUSIONES DE LA COMPARATIVA DEL TAMAÑO DE LAS PANTALLAS	32
6.5	CONCLUSIONES DE LA COMPARATIVA DE LA TASA BINARIA DE CODIFICACIÓN	33
6.6	CONCLUSIONES DE LA ESTIMACIÓN DE LAS DISTANCIAS PREFERIDAS DE VISUALIZACIÓN.....	35

1 Objetivo

El presente documento del Grupo Horizontal sobre calidad de vídeo del Foro de Alta Definición aborda la revisión de sistemas de medida de calidad de vídeo subjetivos y objetivos que permiten la valoración global de secuencias de vídeo de alta resolución, y presenta los resultados de estudios realizados en proyectos financiados por el Ministerio de Industria, Turismo y Comercio sobre el impacto de la convivencia entre las definiciones estándar y alta.

2 Introducción

Los grandes avances en la evaluación de la calidad de imagen se deben a la televisión y a la aparición de los sistemas digitales multimedia. El desarrollo de la Televisión Digital y la aparición de diferentes soluciones tecnológicas que permiten su explotación en diferentes ámbitos, asociados generalmente a servicios y prestaciones concretos, introducen nuevas consideraciones de calidad de servicio (QoS), con relaciones complejas entre los parámetros de calidad objetivos y la calidad de imagen subjetiva.

En base a estos criterios se evalúa la capacidad de un sistema visual o de una técnica de procesamiento de imagen.

Recientemente, se han introducido nuevos métodos de compresión para vídeo digital, y la evaluación de su calidad y prestaciones mediante algoritmos de medida de calidad objetiva es, sin duda, un elemento primordial.

El sistema de medida subjetivo MOS (*Mean Opinion Score*) constituye la medida de mayor difusión para la determinación de la calidad de imagen, pero como todo método subjetivo su ejecución acarrea una serie de desventajas obvias: el tiempo y los recursos necesarios, el coste y la imposibilidad de automatizar el proceso.

Por el contrario, las métricas de calidad objetiva pueden proporcionar de forma automática, o en un periodo relativamente corto de tiempo, una medida de calidad para una imagen o una secuencia de vídeo.

El propósito es alcanzar medidas de calidad objetivas con un alto grado de correlación con la calidad subjetiva, y así alcanzar una calidad de servicio óptima en los sistemas de televisión digital.

Este informe trata de proporcionar una visión sobre el estado del arte en la medición de calidad de vídeo digital, revisando la metodología de medida y sintetizando los diferentes métodos de medida que se emplean en la actualidad.

Existen diversas medidas de calidad de vídeo objetiva. Desde las que se limitan a calcular las diferencias entre las secuencias originales y comprimidas, hasta las que aplican cálculos complejos basados en el sistema visual humano (HVS - *Human Vision System*) [JAVU1].

Los principales problemas asociados al empleo de métricas objetivas¹ son el de, en numerosas ocasiones, necesitar tanto de la imagen original como de la codificada para realizar los cálculos, lo que impide su utilización en tiempo real o en sistemas como el control de tasa binaria; ser computacionalmente costosas, especialmente las basadas en el sistema visual humano; y el de no presentar siempre una buena correlación con las métricas subjetivas.

¹ Algunas métricas han sido diseñadas específicamente para un funcionamiento óptimo ante algunos tipos concretos de distorsión o en condiciones restringidas y fuera de los mismos presentan resultados poco útiles.

3 Metodología del Análisis de Calidad

Cuando se habla de criterios para la calidad de imagen, se pueden distinguir dos tipos básicos:

- ✎ Medidas Objetivas: analizan las diferencias físicas entre la imagen original y la modificada con alguna métrica de distorsión² (ya sea degradada, restaurada, etc.) e interpretan los resultados en el dominio espacial o frecuencial.
- ✎ Medidas Subjetivas: basadas en la evaluación de la calidad de imagen llevada a cabo por un observador humano.

Existen tres métodos básicos que permiten realizar tales medidas [VQEG]:

- ✎ Referencia completa: método aplicable cuando está presente al completo la señal de vídeo de referencia.
- ✎ Referencia reducida: método aplicable cuando sólo está disponible cierta información de la señal de vídeo de referencia.
- ✎ Sin referencia: método aplicable cuando no se dispone de ninguna información acerca de la señal de vídeo de referencia.

Los tres métodos tienen posibles aplicaciones, y provocan diferentes grados de precisión en las medidas, así como en términos de correlación entre los resultados obtenidos entre las medidas objetivas y subjetivas. [WINKLER]

3.1 Métodos de Calidad Subjetiva

Los métodos de medida de calidad subjetiva han sido estandarizados por la Unión Internacional de Telecomunicaciones (UIT) en su recomendación BT.500 [ITU1].

Los métodos pueden dividirse en dos grupos:

- ✎ Métodos de Estímulo Doble (Double Stimulus): Emplean tanto la señal de referencia como la de prueba.
- ✎ Métodos de Estímulo Único (Single Stimulus): Sólo utilizan la señal de prueba.

El más común del primer grupo es el denominado método de Doble Estímulo con Escala de Calidad Continua (DSCQS - Double Stimulus Continuous Quality Scale). En él, cada prueba consta de un par de estímulos: la referencia y la secuencia de prueba. Los estímulos se muestran dos veces cada uno con un orden elegido al azar. Los observadores evalúan cada estímulo con un valor dentro de una escala continua.

² Normalmente no pueden contemplarse todos los parámetros que inciden en la calidad. Además, variaciones en factores como la frecuencia de cuadro o en la resolución, que impiden la concordancia entre la señal original de referencia y de prueba, no pueden ser analizadas.

Otros métodos son el de Escala de Degradación con Doble Estímulo (DSIS - Double Stimulus Impairment Scale) o el de Voto Binario con Doble Estímulo (DSBV - *Double Stimulus Binary Vote*).

Los métodos de estímulo único sólo presentan una secuencia por prueba. Los observadores votan cada estímulo ajustándose a una escala de calidad de cinco valores. Estos métodos son apropiados cuando se trabaja en ausencia de referencia.

3.2 Métodos de Calidad Objetiva

Todos los análisis de calidad objetiva de video digital están basados en distintas medidas de las diferencias entre la secuencia original y la comprimida/degradada. [WANG]

Las medidas más sencillas se basan en el cálculo de estadísticas de error numérico entre la imagen distorsionada y la de referencia [LAMB]. Las más utilizadas son el Error Cuadrático Medio (MSE - *Mean Squared Error*) y la Relación Señal a Ruido de Pico (PSNR - *Peak Signal to Noise Ratio*). El mayor inconveniente que plantean es la relativa falta de correlación con las medidas subjetivas. [VORAN]

Existen también sistemas que buscan una mayor correlación con la distorsión percibida, o tratan de caracterizar el sistema visual humano. A ese modelo se corresponden los métodos de Tektronix/Sarnoff, NASA – Modelo de Calidad de Vídeo Digital (DVQ - *Digital Video Quality Model*) y la Métrica de Distorsión Perceptual de la Escuela Politécnica Federal de Lausana –EPFL- (PDM - *Perceptual Distortion Metric*).

Este informe se centra primordialmente en los métodos de medidas objetivas con referencia completa, aquel en el que se posee toda la información de la señal de vídeo de referencia. Este tipo de métodos evalúan el funcionamiento de un sistema mediante la comparación de una señal de vídeo de entrada sin ningún tipo de distorsión, la de referencia, y la señal de vídeo degradada que se obtiene a la salida del sistema.

4 Análisis de Calidad Subjetiva

La metodología empleada para el análisis de calidad subjetiva ha sido normalizada por la Unión Internacional de Telecomunicaciones en la recomendación BT.500 [ITU1].

Se utilizan métodos de evaluación subjetiva para determinar la calidad de funcionamiento de sistemas de televisión a través de mediciones que anticipan las reacciones de los posibles espectadores. En este aspecto, se comprende que no sería posible caracterizar totalmente la calidad de funcionamiento del sistema por medios objetivos. En consecuencia, es necesario complementar las mediciones objetivas con mediciones subjetivas.

En general, existen dos clases de evaluaciones subjetivas:

- ✎ Evaluaciones de calidad: determinan la calidad de funcionamiento del sistema en evaluación bajo condiciones óptimas.
- ✎ Evaluaciones de degradación: evalúan la capacidad de los sistemas de mantener la calidad en condiciones no óptimas relacionadas con la calidad de la transmisión o emisión.

4.1 Características comunes de los sistemas de evaluación

✎ Condiciones generales de observación

En las pruebas de evaluación de calidad hay que marcar una referencia común en cuanto a los parámetros de observación, como son la luminancia, el brillo y contraste de la pantalla, la distancia de observación, etc.

✎ Señales fuente

Las imágenes y secuencias almacenadas digitalmente son las señales fuente más reproducibles, y son, por consiguiente, las preferidas.

✎ Selección del material de prueba

En la práctica se deben emplear determinadas clases de imágenes de prueba para abordar problemas de evaluación específicos.

✎ Observadores

En las pruebas deben participar al menos 15 observadores. No deben ser expertos, es decir, no deben estar directamente familiarizados con el estudio de la calidad de imagen ni con los métodos de compresión y codificación. El número de asesores necesarios depende de la sensibilidad y la fiabilidad del procedimiento de prueba adoptado.

✎ Instrucciones para la evaluación

Debe familiarizarse detenidamente a los observadores con el método de evaluación, los factores que definen la calidad, los tipos de degradaciones que

probablemente se produzcan, la escala de apreciaciones, la secuencia y la temporización.

✍ Sesión de evaluación

Una sesión debe durar entre media hora y una hora y media como máximo.

✍ Presentación de los resultados

Una vez realizados las pruebas subjetivas deben obtenerse las medidas estadísticas para cada evaluación solicitada. Habitualmente, se debe dar la media y el intervalo de confianza de la distribución estadística obtenida para los grados de evaluación definidos. Si lo que se evalúa es el cambio de degradación con un valor de parámetro variable, deben utilizarse técnicas de ajuste de curvas.

Los resultados deben darse junto con la información siguiente:

- ✍ Detalles de la configuración del experimento
- ✍ Detalles de los materiales de evaluación
- ✍ Tipo de la imagen fuente y de los monitores
- ✍ Número y tipo de evaluadores
- ✍ Sistemas de referencia utilizados
- ✍ Nota media global del experimento (la media estadística obtenida para cada parámetro en evaluación para el conjunto de los evaluadores).

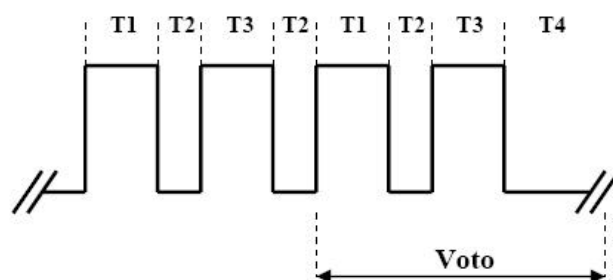
4.2 Descripción de los distintos métodos de evaluación

Este apartado resume los principales métodos aportados por la UIT-R para realizar medidas subjetivas de calidad en imágenes de televisión. Dichos métodos se detallan en su mayoría en la recomendación UIT-R BT.500-11 y se aplican según las características del experimento que se pretenda realizar. [WOLF]

✍ Método de escala de degradación con doble estímulo (DSIS)

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

En su variante II, que se utiliza para discriminar entre degradaciones muy pequeñas o cuando se están sometiendo a prueba secuencias en movimiento, la imagen o secuencia de referencia y la imagen o secuencia de prueba se presentan dos veces. Esto lo podemos apreciar en la siguiente figura:



Fases de presentación:

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

Figura 1: Estructura de presentación del material de prueba (DSIS)

A la hora de evaluar las diferencias observadas entre la imagen de referencia y la imagen a evaluar, se debe utilizar la escala discreta de apreciación de cinco notas:

Escala de cinco notas Calidad

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

Tabla 1: Escala de apreciación de cinco notas

Para este método, las imágenes y degradaciones deberían presentarse en una secuencia pseudoaleatoria y, preferentemente, en secuencias distintas para cada sesión, eligiéndose de tal manera que la mayoría de los observadores utilicen todas las notas de la escala de apreciación.

✍ Método de escala de calidad continua de doble estímulo (DSCQS)

El método es cíclico puesto que se pide al evaluador que observe un par de imágenes, ambas de la misma fuente, pero habiéndose transmitido una por el sistema que se evalúa, y la otra directamente desde la fuente. En lugar de evaluar la diferencia, ahora se pide que evalúe la calidad de ambas.

En sesiones que duran hasta media hora, se presenta al evaluador una serie de pares (aleatorios) de imágenes en orden aleatorio, y con degradaciones aleatorias que abarcan todas las combinaciones requeridas. Al final de las sesiones, se calculan las notas medias para cada condición de prueba y para cada imagen de prueba.

El gráfico de presentación de los datos sería similar al observado en el método anterior, con la diferencia de que ahora $T1 = \text{Secuencia de prueba A}$ y $T2 = \text{Secuencia de prueba B}$. La secuencia de prueba A y la secuencia de prueba B corresponden a la imagen directa de referencia o a la imagen transmitida por el sistema sometido a prueba, pero ésta vez la disposición varía aleatoriamente entre una condición de prueba y la siguiente, siendo el experimentador conocedor de ese dato.

Por tanto, una de las imágenes de prueba de cada par está degradada mientras que la otra puede o no contener una degradación constituyendo la imagen de referencia, cambiándose la posición de esta última de manera pseudoaleatoria.

A los observadores se les pide simplemente que evalúen la calidad global de imagen de cada presentación haciendo una marca en una escala vertical. Las escalas verticales se imprimen por pares para respetar la presentación doble de cada imagen de prueba. Las escalas ofrecen un sistema de evaluación continuo para evitar errores de cuantificación, pero están divididas en cinco segmentos de igual longitud que corresponden a la escala de calidad normal de cinco notas del UIT-R, actuando únicamente como referencias.

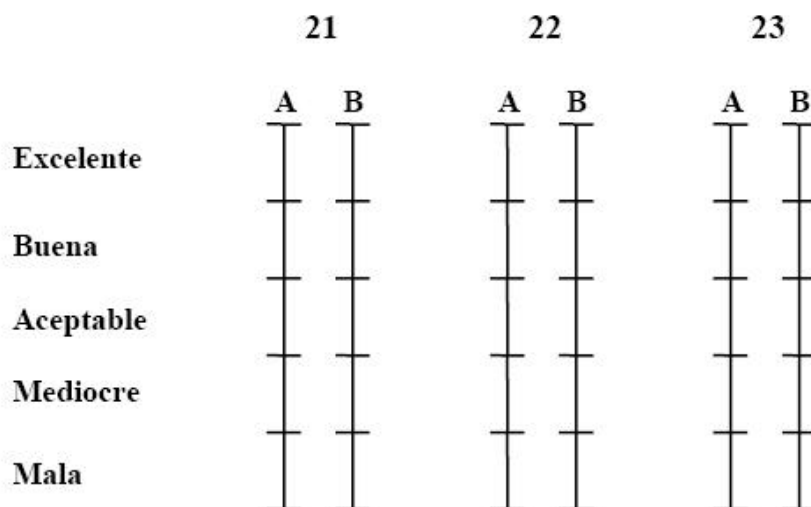


Figura 2: Ejemplo de evaluación de calidad con escalas continuas

Una vez realizados los tests, los pares de evaluaciones (de referencia y prueba) correspondientes a cada condición de prueba se convierten de mediciones de longitud en la hoja de resultados a resultados normalizados en la escala de 0 a 100. A continuación se calculan las diferencias entre la evaluación de la condición de referencia y la de prueba.

Es importante tener en mente que los resultados obtenidos por el método DSCQS no deberán tratarse como resultados absolutos sino como diferencias de resultados entre una condición de referencia y una condición de prueba.

✍ Métodos de estímulo único

En los métodos de estímulo único, se presenta una sola imagen o secuencia de imágenes y el evaluador proporciona un índice de toda la presentación.

La sesión de prueba consiste en una serie de pruebas de evaluación, que deberían presentarse según un orden aleatorio y, preferiblemente, en una secuencia aleatoria distinta para cada observador. Cuando se utiliza un orden aleatorio único de secuencias, hay dos variantes de la estructura de las presentaciones: estímulo único y estímulo único con repetición múltiple.

En el estímulo único, las imágenes o secuencias de prueba se presentan solamente una vez en la sesión de prueba, consistiendo una prueba típica en tres presentaciones: un campo de adaptación en gris medio, un estímulo y un campo de post-exposición en gris medio. El índice o los índices del observador pueden recogerse durante la presentación del estímulo o del campo de post-exposición.

En el estímulo único con repetición múltiple, las imágenes o secuencias de prueba se presentan tres veces organizando la sesión de prueba tres presentaciones, cada una de las cuales incluye todas las imágenes de secuencias que se han de

probar solamente una vez. Una prueba de evaluación típica consiste en dos presentaciones: un estímulo y un campo de post-exposición en gris medio, pudiéndose recoger el índice o los índices del observador durante la presentación del campo de post-exposición únicamente.

En general se han utilizado tres tipos de métodos de estímulo único en las evaluaciones de televisión:

- Métodos de apreciación por categorías de adjetivos

En las apreciaciones por categorías de adjetivos, los observadores asignan una imagen o secuencia de imágenes a una categoría elegida entre un conjunto de categorías que, por lo general, se definen en términos semánticos.

Las escalas de categorías que evalúan la calidad de imagen y la degradación de imagen, son las que se han utilizado más a menudo. En el siguiente cuadro podemos ver algunas de las escalas de la UIT-R.

Escala de cinco notas	
Calidad	Degradación
5. Excelente	5. Imperceptible
4. Buena	4. Perceptible, pero no molesta
3. Aceptable	3. Ligeramente molesta
2. Mediocre	2. Molesta
1. Mala	1. Muy molesta

Tabla 2: Escalas de apreciación de calidad y degradación de cinco notas

- Métodos de apreciación por categorías numéricas

El SSNCS es el método más aplicado en tales apreciaciones y se basa en una escala de categoría numérica de once notas. Cuando no se dispone de referencia, éste método nos permite mayor precisión en términos de sensibilidad y estabilidad.

- Métodos que no emplean escala de evaluación por categorías

Cuando las apreciaciones no se hacen por categorías, los observadores asignan un valor a cada imagen o secuencia de imagen mostrada.

Una posibilidad es realizar una apreciación por escala continua, donde el evaluador asigna cada imagen o secuencia de imagen a un punto de una línea trazada entre dos niveles semánticos. La escala puede incluir rangos adicionales

en puntos intermedios para fines de referencia. La distancia con respecto a un extremo de la escala se toma como índice para cada condición.

Así mismo se puede hacer una distribución por escala numérica, donde el evaluador asigna a cada imagen o secuencia de imágenes un número que refleja su nivel estimado en una dimensión especificada. En este caso, las escalas de números utilizados pueden ser restringidas o ilimitadas.

Con ambas formas se obtiene una distribución de números para cada condición. El método de análisis utilizado depende de la naturaleza de la apreciación y de información requerida.

- Métodos de comparación de estímulos

En los métodos de apreciación por categorías de adjetivos, los observadores asignan la relación entre miembros de un par a una categoría elegida entre un conjunto de categorías que, normalmente, se definen en términos semánticos. La escala de comparación de la UIT-R se muestra en la siguiente figura.

-3	Mucho peor
-2	Peor
-1	Ligeramente peor
0	Igual
+1	Ligeramente mejor
+2	Mejor
+3	Mucho mejor

Tabla 3: Escala de comparación de la UIT – R

Para éstos métodos, la manera en que se analizan las respuestas depende tanto de la apreciación como de la información requerida.

- Métodos que no utilizan una escala de apreciación por categorías

Cuando las apreciaciones no se hacen por categorías, los observadores asignan un valor a la relación entre los elementos de un par de evaluación. Éste método puede revestir dos formas:

- En la apreciación con escala continua, el evaluador asigna cada relación a un punto de una línea trazada entre dos notas.
- En la segunda forma, el evaluador asigna a cada relación un número que refleja el nivel estimado en una dimensión especificada, pudiendo estar la gama de números utilizada limitada o no. El número asignado puede

describir la relación en términos absolutos o en términos de la relación en un par estándar.

✎ Método de evaluación de calidad continua de estímulo único (SSCQE)

A medida que la televisión se va introduciendo en el mundo digital se van sumando nuevos factores que es preciso tener en cuenta. Éstos nuevos factores, tales como la comprensión, hacen que incluso dentro de breves muestras de vídeo codificado digitalmente, la calidad pueda variar mucho dependiendo del contenido de la escena y las degradaciones puedan ser de muy corta duración.

Para analizar estos casos, se ha creado el método SSCQE donde se considera conveniente que la calidad subjetiva del vídeo codificado digitalmente se mida de manera continua, observando los sujetos participantes el material una sola vez, sin una referencia fuente.

Para implementar adecuadamente este método, se debe utilizar un sistema de registro electrónico con un mecanismo de desplazamiento lineal deslizante, conectado a un computador, donde se registre la evaluación de la calidad por parte de los participantes.

La presentación de prueba definida en este método sustenta la realización completa de una prueba y puede comprender una o varias sesiones de prueba, es decir, una o varias series de una o más combinaciones diferentes de segmento de programa/parámetro de calidad sin separación y dispuestas en orden pseudoaleatorio. Cada sesión de prueba contiene por lo menos una vez todos los segmentos de programa y parámetros de calidad, pero no necesariamente todas las combinaciones segmento de programa/parámetro de calidad. Según lo definido, un segmento de programa corresponde a un tipo de programa procesado de acuerdo con uno de los parámetros de calidad objeto de evaluación.

A la hora de presentar los resultados, se pueden trazar gráficas para segmento de programa que nos muestren la media de las observaciones, en una escala de 0 a 100 en función del tiempo de observación. Para realizar esto último, hay que ser conscientes de los diferentes sesgos basados en la memoria y que por tanto dependen de la duración del segmento de programa. Actualmente se están estudiando diversas formas para mitigar el efecto introducido por dichos sesgos en el análisis de los resultados.

✎ Método de doble estímulo simultáneo para evaluación continua (SDSCE)

La idea de evaluación continua fue introducido por la UIT-R para intentar solventar las deficiencias que presentaban los métodos anteriores para la

medición de la calidad de vídeo de esquemas de compresión digital. Estas deficiencias se basan principalmente en los artefactos digitales que dependen en gran medida del contenido espacial y temporal de la imagen fuente, que pueden suceder en un servicio real y que con la temporización usada (generalmente de 10 s) no es suficiente para que el observador analice.

Es por tanto que se introdujo el método SSCQE, que es capaz de medir la calidad vídeo en secuencias más largas, representativas del contenido vídeo y de la estadística de errores. Pero para reproducir las condiciones de observación que estén lo más próximas posibles a las situaciones reales, en el SSCQE no se utilizan referencias, mientras que para evaluar la fidelidad se han de introducir condiciones de referencia.

Así pues, el SDSCE ha sido elaborado a partir del SSCQE, con ligeras diferencias en cuanto a la manera de presentar las imágenes a los sujetos y con respecto a la escala de apreciación.

En el SDSCE el grupo de sujetos observa dos secuencias al mismo tiempo, generalmente presentadas en dos monitores alineados, siendo una la referencia y la otra la condición de prueba. Los espectadores, conocedores de este dato, deben comprobar las diferencias entre las dos secuencias y juzgar la fidelidad de la información vídeo en una escala de 0 a 100.

Al igual que en los demás métodos, debe existir una fase de entrenamiento donde se entreguen todas las instrucciones y se respondan a todas las dudas de los espectadores. Tras esto seguirá una sesión de demostración para familiarizarse con el procedimiento del voto y la clase de degradaciones y posteriormente se realizará la prueba simulada.

La fiabilidad de los sujetos puede ser evaluada cualitativamente comprobando su comportamiento cuando se muestran los pares de referencia/referencia. En estos casos, se espera que los sujetos den evaluaciones muy próximas a 100. Esto prueba que por lo menos han comprendido su tarea y que sus votos no son aleatorios.

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

- Desviación sistemática: un observador puede ser demasiado optimista o pesimista (incluso puede haber entendido mal el proceso de votación). Esto puede conducir a una serie de votos con desviación sistemática a la serie media, si no completamente fuera de la gama.
- Inversiones locales: a veces los observadores votan sin preocuparse mucho de observar y seguir cuidadosamente la calidad de la secuencia visualizada.

A la hora de evitar dichos efectos, el entrenamiento de los participantes se posiciona como el mejor remedio, no descartándose la eliminación de los resultados obtenidos por los observadores incoherentes.

✍ *Metodología de valoración subjetiva para calidad de vídeo (SAMVIQ)*

El método SAMVIQ está pensado ante todo para realizar evaluaciones subjetivas de los contenidos de video en un dominio multimedia. Este método se hizo necesario porque hay algunas diferencias importantes entre la televisión y la información multimedia. Así encontramos diferentes temporizaciones en las variaciones del refresco de la imagen, una tasa de actualización del video variante, diferentes distancias de visualización, etc , diferencias que hacen más difícil evaluar la calidad de las imágenes multimedia que las de la televisión. La experiencia muestra que esta combinación influye fuertemente en la calidad subjetiva final percibida. [SUNNA]

Hay muchas diferencias entre los métodos descritos en la UIT-R BT.500-11 expuestos con anterioridad y el método SAMVIQ. La principal es la forma en que las secuencias son mostradas al evaluador. Con el SAMVIQ las secuencias de video se muestran en forma de multiestímulos, por tanto el usuario puede elegir el orden del test y corregir sus votos. Un evaluador puede empezar y terminar su proceso de evaluación cuando desee y le permite establecer su propio ritmo para realizar las evaluaciones, modificar marcas anteriores, repetir una escena cuando sea necesario, etc. Con esto se consigue que no se produzca una presentación secuencial continua como en el método DSCQS, donde el evaluador puede caer en errores al juzgar las secuencias debido a la carga de concentración que tiene que tener y evitar el "efecto grupo" al ser cada visualización individual.

Las sesiones test están organizadas de tal forma que todas las escenas se producen consecutivamente y para cada escena es posible reproducir y valorar cada secuencia en cualquier orden. Entre diferentes escenas, el orden de las secuencias se aleatoriza para prevenir que los espectadores voten siempre en un orden determinado. La escala de apreciación que se utiliza es la misma que para los métodos de estímulo único.

Este test, para el que es preciso un software a la hora de computar y mostrar las imágenes, no precisa de una distancia de visualización específica. El observador ajusta su propia distancia óptima acordando a su preferencia en la observación del vídeo.

4.3 Comparación de los métodos de evaluación subjetiva

La utilización de uno u otro método de los descritos en el apartado anterior depende del problema que se pretende evaluar. Como se muestra en la tabla 4, cada uno de los problemas más usuales lleva un test asociado para el estudio de la calidad subjetiva en la percepción de la televisión.

METODOLOGÍA					
Parámetros	DSIS	DSCQS	SSCQE	SDSCE	SAMVIQ
Referencia explícita	Sí	No ⁽¹⁾	No	Sí	Sí (sin comprimir)
Referencia oculta	No	Sí ⁽¹⁾	No	No	Sí
Alto fijo	No	Sí ⁽¹⁾	No	No	No (referencia oculta)
Bajo fijo	No	Sí ⁽³⁾	No	No	Sí
Escala	5 grados	Mala-Excelente (escala continua de calidad)	Mala-Excelente (escala continua de calidad)	Mala-Excelente (escala continua de calidad)	Mala-Excelente (escala continua de calidad)
Longitud de la secuencia	10 s	10 s	=5 min	10 s	10 s
Formato de imagen	Todos	Todos	Todos	Todos	Todos
Dos estímulos simultáneos	No	No	No	Sí	No
Presentación del material de test	Variante 1: sólo una vez Variante 2: dos veces sucesivas	Dos veces sucesivas (doble estímulo)	Una vez	Una vez	Tantas veces como el usuario desee (multiestímulo)

Videos por exposición	2	2	1	2	Un máximo de 10 ⁽⁴⁾
Votar	Sólo en la secuencia de test	Sólo en la secuencia de test y en la de referencia	Sólo en la secuencia de test	Se vota la diferencia entre la secuencia de test y la de referencia, que se muestran de modo simultáneo	Secuencia de test y de referencia
Posibilidad de cambiar el voto antes de proceder	No	No	No	No	Sí
Evaluación continua de calidad	No	No	Sí	Sí	No
Número mínimo de votos aceptados	15	15	15	15	15
Criterio de rechazo		Sí, pero cambiante			Sí
Observadores por dispositivo	Uno o muchos	Uno o muchos	Uno o muchos	Uno o muchos	Sólo uno
Dispositivos	Todos (principalmente TV)	Todos (principalmente TV)	Todos (principalmente TV)	Todos (principalmente TV)	Todos (sobre todo PC, PDA, etc)
Resultados de calidad	Relativos, dependientes de la referencia de calidad	Relativos, dependientes de la secuencia comparada	Relativos	Relativos, dependientes de la referencia de calidad	Absolutos
Estándar	ITU-R BT.500-11	ITU-R BT.500-11	ITU-R BT.500-11	ITU-R BT.500-11	ITU-R BT.700

Tabla 4: Características de las cinco recomendaciones

¹ No es obligatorio (Podía ser cualquier secuencia de test)

² Diferentes tasa binarias en un intento para evitar los efectos contextuales

5 Análisis de Calidad Objetiva

5.1 Métricas con base Matemática

En la mayoría de las aplicaciones de tratamiento de imagen se trabaja con unas métricas estándar [OLSS] para determinar la calidad de la imagen. A continuación se detallan las más comunes.

La notación usada hace referencia a:

N, M: dimensión vertical y horizontal de la imagen en píxeles

$x(i,j)$ e $r(i,j)$ imagen original y restaurada respectivamente en notación matricial.

\mathbf{x} y \mathbf{r} : imagen original y restaurada respectivamente en notación lexicográfica.

✎ Error Cuadrático Medio (MSE):

$$MSE = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (x(i,j) - r(i,j))^2}{NM} = \frac{(\mathbf{x} - \mathbf{r})(\mathbf{x} - \mathbf{r})^T}{NM}$$

Donde T indica transpuesta.

✎ Raíz del Error Cuadrático Medio (RMSE):

$$RMSE = \sqrt{MSE}$$

✎ Error Cuadrático Medio Normalizado (NMSE):

$$NMSE = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (x(i,j) - r(i,j))^2}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (x(i,j))^2} = \frac{(\mathbf{x} - \mathbf{r})(\mathbf{x} - \mathbf{r})^T}{\mathbf{xx}^T}$$

✎ Relación Señal a Ruido (SNR):

$$SNR(\text{dB}) = 10 \log(NMSE)$$

✎ Relación Señal a Ruido de Pico (PSNR), para imágenes de 8 bits/píxel:

$$PSNR(\text{dB}) = 10 \log \frac{MSE}{x(i,j)_{\max}^2} = 10 \log \frac{MSE}{\|\mathbf{xx}^T\|_{\max}} = 10 \log \frac{MSE}{255^2} = 20 \log \frac{255}{RMSE}$$

✎ Relación Señal a Error (SER):

$$SER = 20 \log_{10} \frac{x(i,j)_{\max}}{RMSE}$$

≡ Desviación Estándar (STD):

$$STD = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (x(i,j) - r(i,j))^2 = \frac{1}{NM} ((\mathbf{x} - \mathbf{r})(\mathbf{x} - \mathbf{r})^T)$$

donde:

$$= \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} x(i,j) - r(i,j)}{NM}$$

Tales índices forman un conjunto de parámetros de calidad objetiva conocidos como *Medidas de Distorsión Comparativa*, y explican el hecho de que grandes valores de error corresponden a una pésima calidad de imagen. Su mayor inconveniente es que sólo son eficientes cuando los errores se comportan como ruido adicional que no está correlacionado con la señal. Es decir, no se puede distinguir entre pocos errores de gran amplitud (los cuales son molestos para los observadores) y muchos errores con baja amplitud (que desde un punto de vista subjetivo pueden ser imperceptibles). Estas medidas de distorsión comparativa pueden ser más eficientes desde el punto de vista del dominio frecuencial cuando se aplica a imágenes previamente filtradas con filtros de modelado del SVH (Sistema Visión Humano). Con ello se puede dar una mejor estimación de la importancia del error para el ojo humano.

El parámetro de calidad de imagen objetivo puede incluir en ciertos casos la aplicación de un modelo perceptual de visión humana.

La visión humana es un mecanismo muy complejo que ha sido estudiado durante años; el modo en que el ojo humano percibe el color y el movimiento juega un papel muy importante en la definición de calidad de vídeo. Por ejemplo, debido a la anatomía y características físicas del ojo humano, éste es más sensible a los colores rojizos y menos a los tonos amarillos y azules [POIR93]. De esta manera, la pérdida de detalles en las dimensiones del azul y el amarillo será menos evidente para el ojo humano, y por ello debieran tener menos peso en los parámetros de medida de distorsión de vídeo.

5.2 VQEG

Los estudios de medidas perceptuales de calidad de vídeo están dirigidos por un grupo de Expertos de Calidad de vídeo (Vídeo Quality Expert Group [VQEG]), que colaboran con los grupos de estudio 9 y 12 de la UIT-T [UIT-T9][UIT-T12] y el grupo 6 de la UIT-R [UIT-R6].

La primera tarea del VQEG fue la de valorar el funcionamiento de los algoritmos de medida de calidad de vídeo, disponiendo de la señal de referencia completa, existentes hasta el momento.

El VQEG distribuyó las primeras conclusiones a modo de borrador en Marzo del 2000. En la publicación se mostraban los resultados de las pruebas efectuadas sobre diez modelos sometidos al juicio del VQEG por diez asociaciones diferentes, que calculaban valores objetivos comparados con evaluaciones subjetivas sobre un amplio rango de sistemas de vídeo y de fuentes de secuencias. Las pruebas contrastaban el comportamiento de los modelos propuestos con los test de valoración subjetiva sobre las mismas imágenes, y también comparaban con el algoritmo de referencia PSNR. El propósito era verificar la predicción de los modelos propuestos en términos de:

- ✍ Precisión: capacidad del modelo para predecir la calidad subjetiva
- ✍ Monitorización: grado en que la predicción del modelo sigue las pautas marcadas por las tasas de calidad subjetiva
- ✍ Consistencia: grado en que el modelo mantiene la precisión en la predicción sobre un rango de secuencias y sistemas de vídeo, es decir, robustez frente a diversas entradas de vídeo.

Se generaron alrededor de 26000 muestras de opiniones subjetivas basadas en 20 fuentes de secuencias diferentes, procesadas por 16 sistemas de vídeo y evaluadas en 8 laboratorios independientes repartidos por todo el mundo.

Los sondeos subjetivos se organizaron en cuatro cuadrantes: calidad de contribución a 50 Hz, calidad de distribución a 50 Hz, calidad de contribución a 60 Hz y calidad de distribución a 60 Hz. La calidad de contribución incluía vídeo con tasas comprendidas entre 3 y 50 Mbps, mientras que distribución quedaba limitado a vídeo con tasas entre 768 Kbps y 4.5 Mbps.

La evaluación subjetiva siguió fielmente el método DSQCS desarrollado en la Recomendación UIT-R BT.500-9 [UIT500]. El plan de pruebas objetivas y subjetivas incluía procedimientos para validar el análisis de los resultados subjetivos y cuatro métricas para hacer la comparativa de los datos objetivos con los subjetivos.

Conclusiones al análisis de calidad objetiva de televisión digital

Las conclusiones esenciales que se obtuvieron tras el análisis de las experiencias formalizadas por el VQEG fueron, en ese momento, las siguientes:

- ✍ Un modelo perceptual no es capaz de representar de un modo totalmente fiel el conducta de un estudio subjetivo
- ✍ Estadísticamente, ningún modelo perceptual mejora al resto en todas las condiciones de referencia.
- ✍ Estadísticamente, ningún modelo perceptual mejora el comportamiento del modelo PSNR en todas las condiciones de r eferencia.
- ✍ Basándose en la evidencia actual, no existe un método único que pueda ser recomendado por la UIT en este momento.

Aún así, el trabajo del VQEG no fue en vano, supuso una mejora considerable en el entendimiento del problema de evaluación de calidad de vídeo perceptual, y en los requerimientos del usuario.

A pesar de no poder recomendar un método, en los apartados siguientes se describen, de modo general, los ocho modelos que fueron validados por el VQEG y documentados en la publicación final de Marzo de 2000 [VQEG4], además del modelo general de PSNR.

- ✍ PSNR
- ✍ CPqD
- ✍ Tektronix -Sarnoff
- ✍ KDD
- ✍ NHK/Mitsubishi Electric Corp
- ✍ EPFL
- ✍ NASA-DVQ (Digital Video Quality)
- ✍ KPN/Swisscom CT
- ✍ NTIA

5.2.1 PSNR

Destaca que el PSNR precisa un grado muy alto de normalización para ser usado con confianza. La normalización requiere alineamiento espacial y temporal así como correcciones en la ganancia y el offset. El método de normalización es objeto de otra Recomendación bajo estudio.

5.2.2 CPqD

Este método, también conocido como Evaluación de Imagen basado en la Segmentación, es un sistema diseñado para obtener predicciones de calidad sobre un grupo de escenas predefinidas. CPqD implementa la valoración de la calidad de vídeo usando parámetros objetivos basados en la segmentación de imágenes. Las escenas son segmentadas en planos y regiones de bordes y texturas, y para cada uno de estos contextos se definen parámetros objetivos. Se diseña también un modelo perceptual que predice las valoraciones subjetivas, para calcular las relaciones entre las medidas objetivas y los resultados de la valoración de pruebas subjetivas, aplicadas sobre un conjunto de escenas reales procesadas con diversos sistemas de vídeo.

En una segunda versión desarrollada de este mismo método, se agregaba un clasificador de escenas que permitía que el sistema de evaluación fuese independiente de la escena.

5.2.3 Tektronix/Sarnof

Este prototipo está basado en un modelo de discriminación visual que simula la respuesta espacio-temporal del mecanismo visual humano, y las magnitudes de diferencias perceptuales en el mecanismo de salida entre las secuencias de origen y la procesada. A partir de estas diferencias se calcula una métrica general de discriminación entre las dos secuencias.

Este modelo fue diseñado bajo condiciones de procesamiento hardware de imágenes estándar a gran velocidad, de modo que representa una solución relativamente sencilla para el procesamiento con ordenadores.

5.2.4 NHK/Mitsubishi Electric Corp.

El modelo emula las características de la visión humana usando filtros tridimensionales, que son aplicados a las diferencias entre la señal original y la procesada. Las características de los filtros son variadas basándose en los niveles de luminancia. La calidad de salida obtenida se calcula como la suma de las medidas obtenidas de cada filtro.

El hardware disponible actualmente permite medir calidad de imagen en tiempo real y se usa en varios entornos de difusión tales como la monitorización en tiempo real de señales distribuidas.

5.2.5 KDD

El MSE se calcula restando la señal de prueba de la señal de referencia y se hace pasar por una serie de filtros obtenidos a partir del modelo de visión humano.

5.2.6 École Polytechnique fédérale de Lausanne

La métrica de distorsión perceptual (PDM) suministrada por este grupo de estudio está basada en un modelo espacio-temporal del sistema de visión humano.

Se compone de cuatro fases por las que pasan tanto la señal de referencia como la procesada:

- ✍ La primera convierte la entrada a un espacio de color opuesto.
- ✍ La segunda implementa una descomposición perceptual espacio-temporal en canales visuales separados de diferente frecuencia temporal, frecuencia espacial y orientación.
- ✍ La tercera fase modela los efectos del patrón de enmascaramiento mediante la simulación de los mecanismos de excitación e inhibición, de acuerdo con un modelo de control de ganancia de contrastes.
- ✍ La cuarta y última etapa de la métrica actúa como fase de detección y sondeo y obtiene una medida de distorsión a partir de las diferencias entre los sensores de salida de la referencia y la secuencia procesada.

5.2.7 NASA

El modelo propuesto por la NASA se denomina DVQ (Digital Video Quality). Esta métrica supone un intento de incorporar algunos aspectos de la sensibilidad visual humana en un algoritmo de procesamiento de imagen simple. La simplicidad es un objetivo importante si lo que se busca es una métrica aplicable en tiempo real y con requisitos de procesamiento computacional escasos.

Uno de los elementos más complejos y que más tiempo consumen en otros modelos es el filtrado espacial, usado para simular los múltiples filtros espaciales paso banda, característicos de la visión humana. Este paso se acelera si se usa la transformada discreta del coseno [GONZ92] para la descomposición en canales espaciales. La aplicación de la DCT supone una potente ventaja ya que existe hardware y software disponible para realizar esta transformación, e incluso en muchas aplicaciones la transformación puede ya haberse realizado como parte del proceso de compresión.

La entrada al sistema de medición son un par de secuencias de imágenes de color, la de referencia y la de prueba, que son sometidas a una serie de transformaciones:

- ✍ El primer paso consiste en muestreos y transformaciones de color que sirven para limitar el procesamiento a una región de interés y además para expresar la secuencia en términos de un espacio de color perceptual. Este paso también se ocupa del desentrelazado y la corrección gamma de la señal de vídeo.

- ✍ Las secuencias son sometidas a un bloqueo y a una transformada discreta del coseno y los resultados son entonces transformados a un contraste local.
- ✍ El siguiente paso es el filtrado espacial y temporal y una operación de enmascaramiento de contraste.
- ✍ Finalmente las diferencias enmascaradas se muestran en dimensiones espacial, temporal y cromática para obtener la medida de calidad.

5.2.8 KPN/Swisscom CT

El algoritmo desarrollado por KPN/Swisscom CT se conoce como Métrica de Calidad de Vídeo Perceptual, ó PVQM, y se divide en una serie de pasos que se resumen a continuación:

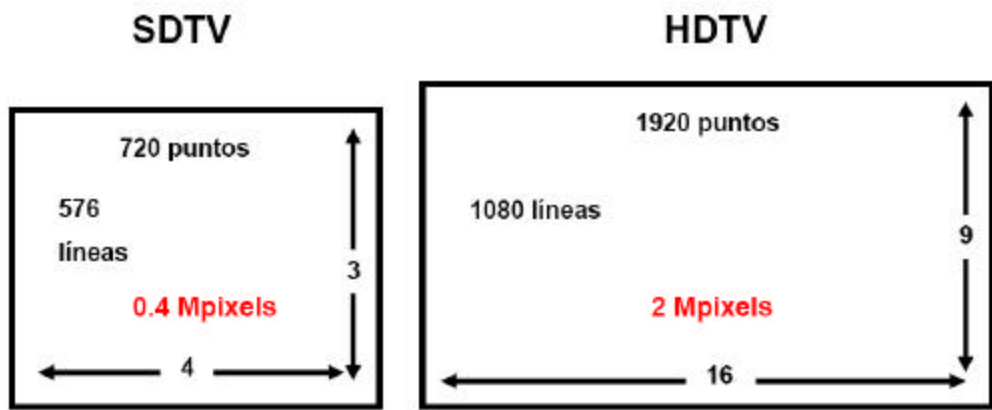
- ✍ Primera aproximación a la función de sensibilidad de contraste del sistema visual humano, en las componentes de luminancia y crominancia. PVQM usa un adaptador especial para la luminosidad y el contraste de la secuencia de vídeo distorsionada, aprovechando que los cambios globales en la luminosidad y el contraste sólo tienen un impacto limitado en la percepción subjetiva de calidad. El análisis espacial de la luminancia se basa en la detección de bordes de la componente Y. Se sabe que el sistema visual humano es mucho más sensible a la definición en la componente de luminancia que a la de las componentes de color. Además, el sistema visual humano tiene una sensibilidad al Contraste que disminuye con el aumento de la frecuencia espacial.
- ✍ Cálculo de la señal de bordes de luminancia, entendida como la señal que representa los aspectos más significativos de la imagen. La imagen de bordes se obtiene calculando el gradiente local de la componente de luminancia usando un filtrado espacial similar a Sobel [GONZ87] para cada cuadro de la secuencia, y promediados después en espacio y tiempo.
- ✍ Cálculo del error de crominancia como la media ponderada de los errores de color en ambas componentes, azul y roja, dominando la crominancia roja.
- ✍ Relación de los tres criterios anteriores para obtener un único indicador de calidad, que presente una correlación alta con la percepción subjetiva de la calidad de la secuencia de vídeo.

5.2.9 NTIA

Este modelo de calidad de vídeo digital aprovecha características de ancho de banda reducido, que se extraen de regiones espacio-temporales de escenas de vídeo procesadas de entrada y salida. Estas propiedades caracterizan los detalles espaciales, el movimiento y el color presentes en la secuencia de vídeo.

- ✍ Las características espaciales describen la actividad de los bordes de la imagen ó el gradiente espacial. Los sistemas de vídeo digital pueden añadir bordes ó difuminarlos.
- ✍ Las características temporales determinan la actividad de las diferencias temporales ó el gradiente temporal entre cuadros sucesivos. Los sistemas de vídeo digital pueden añadir movimiento a la secuencia ó reducirlo.
- ✍ Las características de color describen la actividad de la información de color. Del mismo modo que con el gradiente espacial y temporal, un sistema de vídeo podría añadir información de color o reducirla.
- ✍ Los parámetros de ganancia y pérdida se calculan comparando las características obtenidas para la señal de vídeo de entrada y la señal obtenida a la salida del sistema. Las funciones de comparación tratan de emular la sensibilidad del sistema visual humano a cada uno de los efectos provocados, mediante la definición de umbrales de sensibilidad, enmascaramientos, etc.

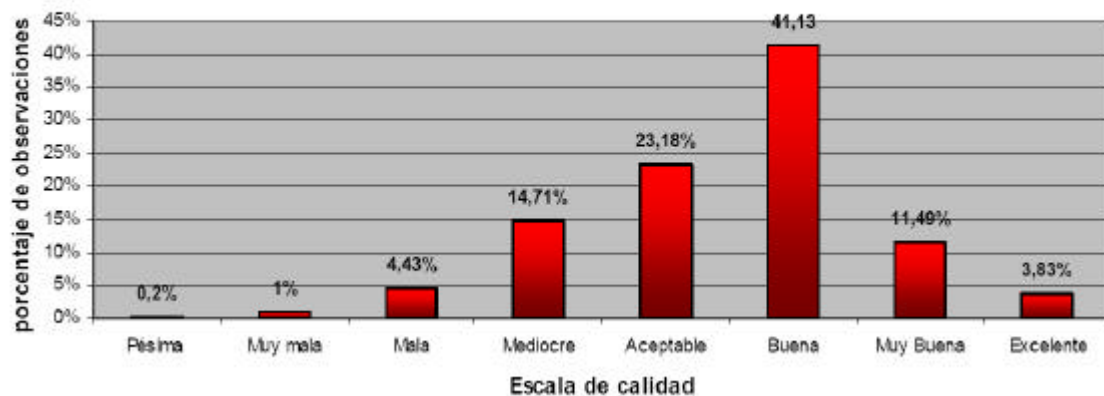
6 Valoración de la convivencia entre SD y HD



6.1 Conclusiones de la comparativa SD vs. SD

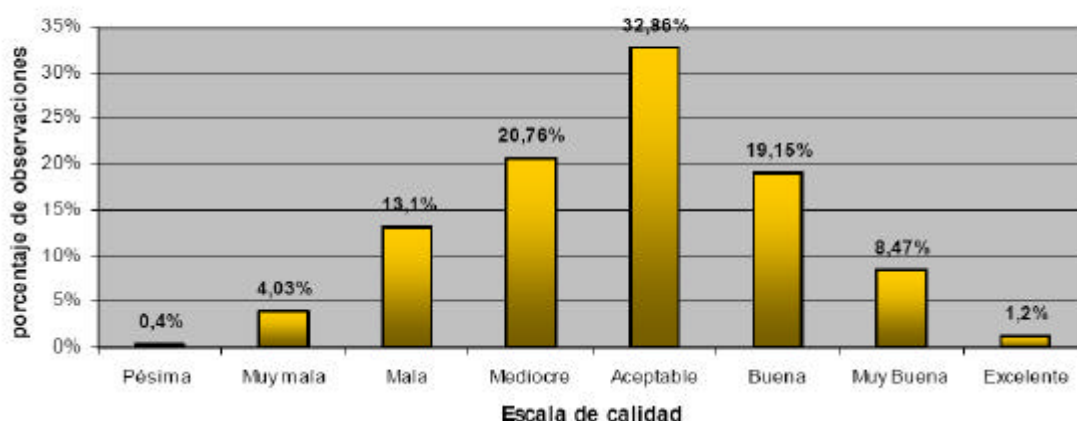
Tras una primera visualización de contenidos de definición estándar sin experiencia previa de la calidad ofrecida por la alta definición, codificados a una tasa binaria equivalente a la que actualmente se está emitiendo para la televisión digital terrestre, los espectadores valoran la calidad de éstas imágenes con un valor medio aproximado de 5.5 sobre una escala de 8 notas. En términos calificativos, dentro de nuestra percepción, dicha puntuación se encuentra entre los valores de "aceptable" y "bueno".

Evolución de la calidad subjetiva SD



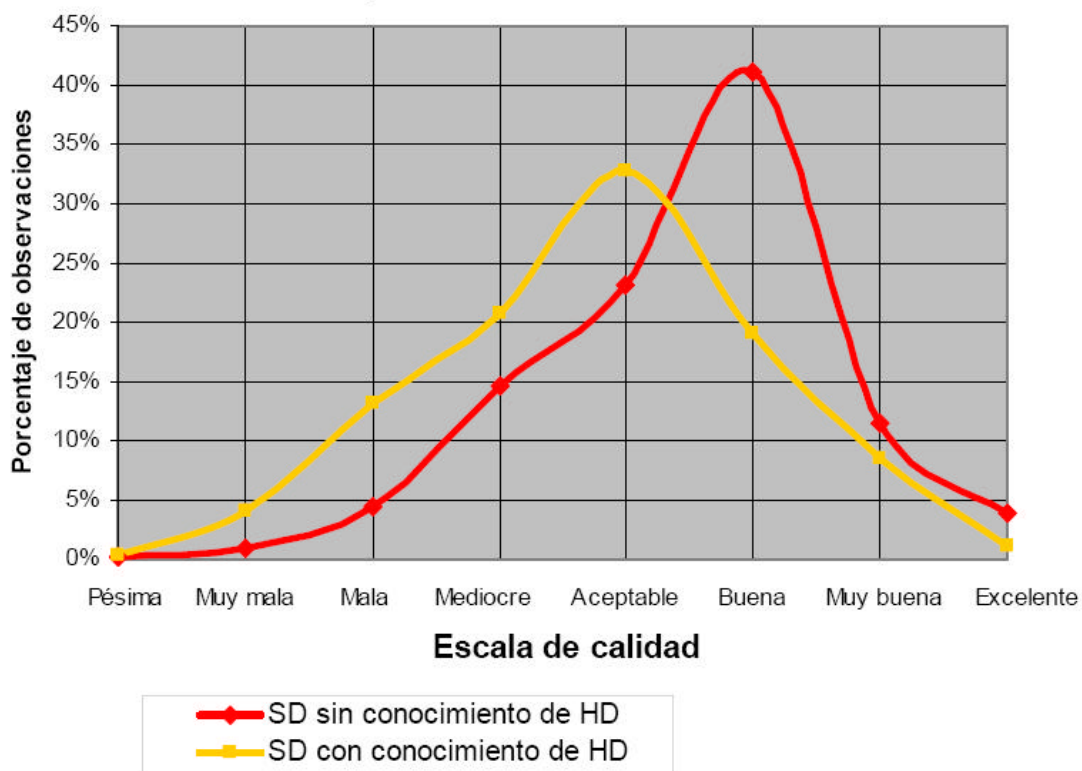
Este valor medio decrece considerablemente hasta la cantidad de 4,8 para los mismos contenidos y condiciones tras la visualización de alta definición, situándolo entre las etiquetas "mediocre" y "aceptable" más cerca de ésta última.

Evolución de la calidad subjetiva SD tras visualizar HD



Tras estos datos se observa como influye el conocimiento previo o la experiencia en la visualización de la alta definición sobre nuestra percepción de la televisión actual y se puede prever como, a medida que crezca el mercado de la televisión digital de alta definición y se haga más popular entre los espectadores, la valoración sobre la televisión digital actual empezará un descenso en términos de calidad subjetiva.

Comparación SD - SD tras visualizar HD

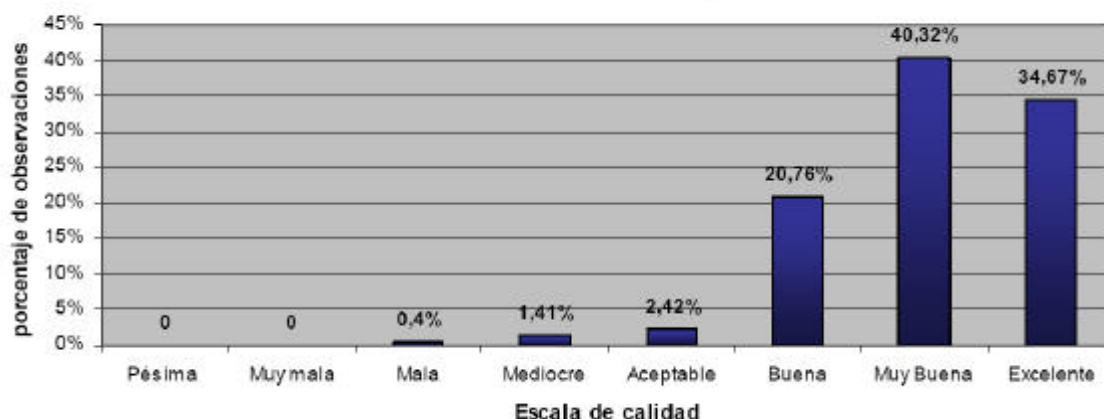


Además se observa como aumenta la incertidumbre en las valoraciones tras el conocimiento de la alta definición sobre datos a los que en un principio (antes de dicho conocimiento) se les encasillaba claramente bajo un nivel de calidad. Dicha incertidumbre ha llevado generalmente a la percepción de las mismas secuencias una escala de calidad inferior (en algunos casos iguales y dos por debajo) que la calidad observada inicialmente.

6.2 Conclusiones de la comparativa SD vs. HD

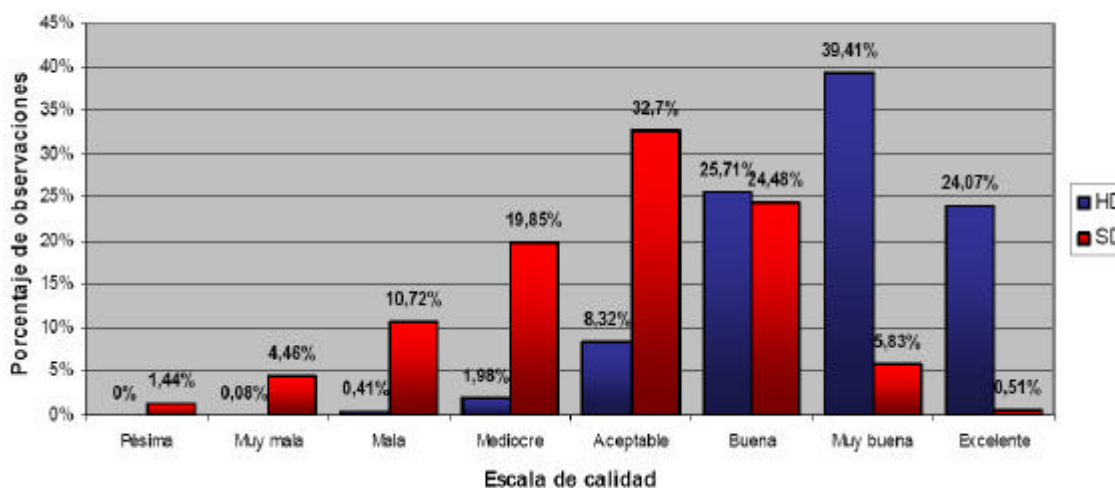
La primera vez que un observador visualiza alta definición, ésta le causa un cierto impacto obteniendo una valoración media de aproximadamente 7 dentro de nuestra escala de 8 notas. Valoración que como se puede observar es bastante alta y se centra en el término “muy buena”.

Evolución de la calidad subjetiva HD



A medida que se visualizan contenidos de alta definición con mayor regularidad, el impacto causado inicialmente se ve paliado, lo que causa un ligero descenso en la valoración media de la cantidad inicial de 7 hasta el valor estable de 6.75. Estos datos provienen cada uno de uno de los dos estudios realizados, siendo los del segundo estudio más estable como se apunta, por corresponder, además de a un mayor número de observaciones, a dichas observaciones con conocimiento de ambos estándares de definición de imágenes.

Comparación SD - HD



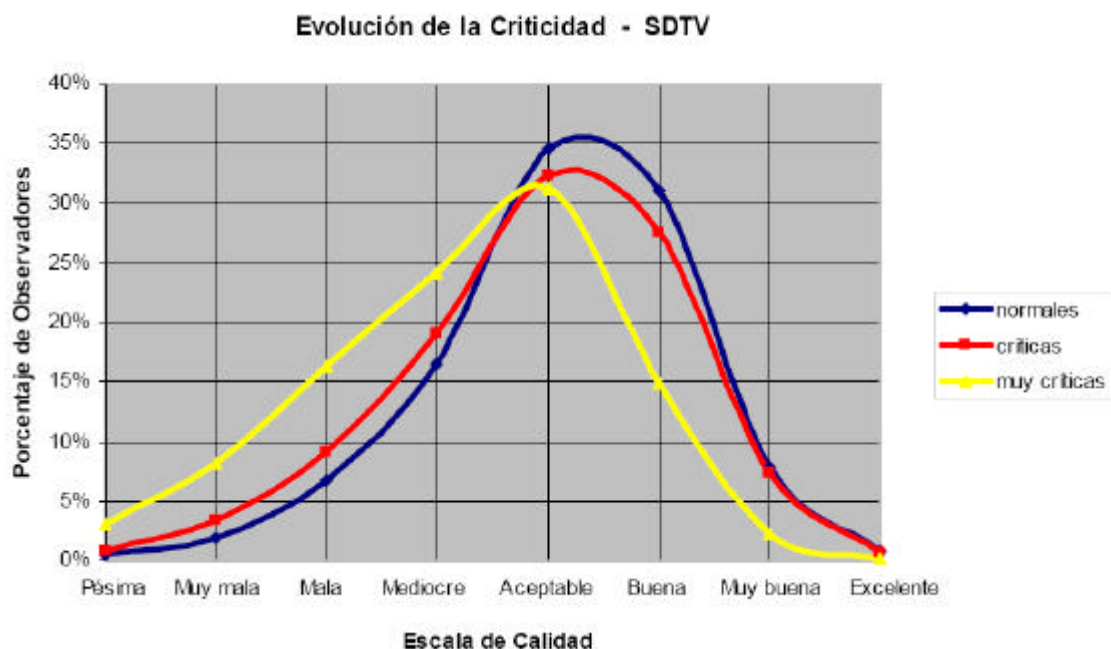
Es curioso observar así mismo como para ambos estudios, tras superar la experiencia del conocimiento de la alta definición, el valor medio de las percepciones de definición estándar se mantiene constante en el valor 4.8, magnitud cercana a la valoración “aceptable”. Por tanto, la diferencia observada entre las percepciones de alta definición y definición estándar, independiente mente de la criticidad de las secuencias, el tamaño de la pantalla y la tasa binaria de codificación, se mantiene constante en aproximadamente dos valoraciones en

nuestra escala de calidad, es decir, la calidad de la definición estándar se visualiza dos puntos de ocho posibles por debajo de la de alta definición.

6.3 Conclusiones de la comparativa de la criticidad de las imágenes

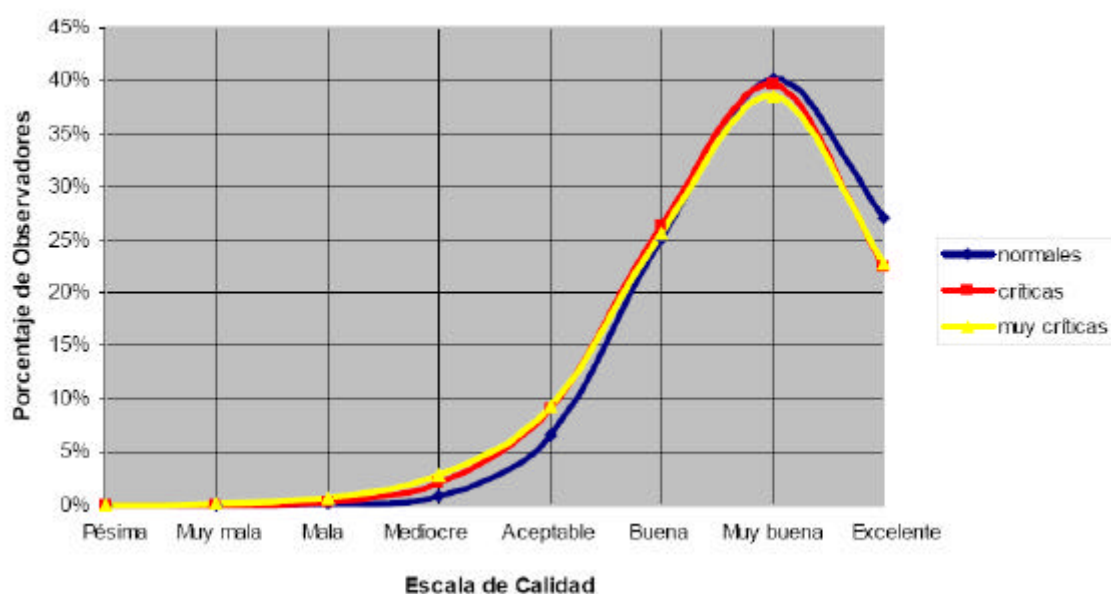
Sobre este estudio y los siguientes hay que destacar en primer lugar que las conclusiones se deben a la influencia que las demás variables tienen sobre la ya comentada comparación o diferencia entre SD y HD, y no afectan tan directamente a la diferencia entre las variables aisladas para cada definición por separado por la imposibilidad de separar psicológicamente ambos conceptos.

Sobre el nivel de criticidad se las secuencias y su influencia en la diferencia de percepción, en los resultados se ve como para contenidos de alta definición tal diferencia es inapreciable, es decir, se consiguen eliminar los artefactos causantes del empeoramiento de la calidad tales como el efecto de bloques y el efecto Moiré y las percepciones se equiparan. Estos mismos efectos son los causantes de que en definición estándar si que se produzca un empeoramiento de la calidad a medida que se aumenta la criticidad de las imágenes.



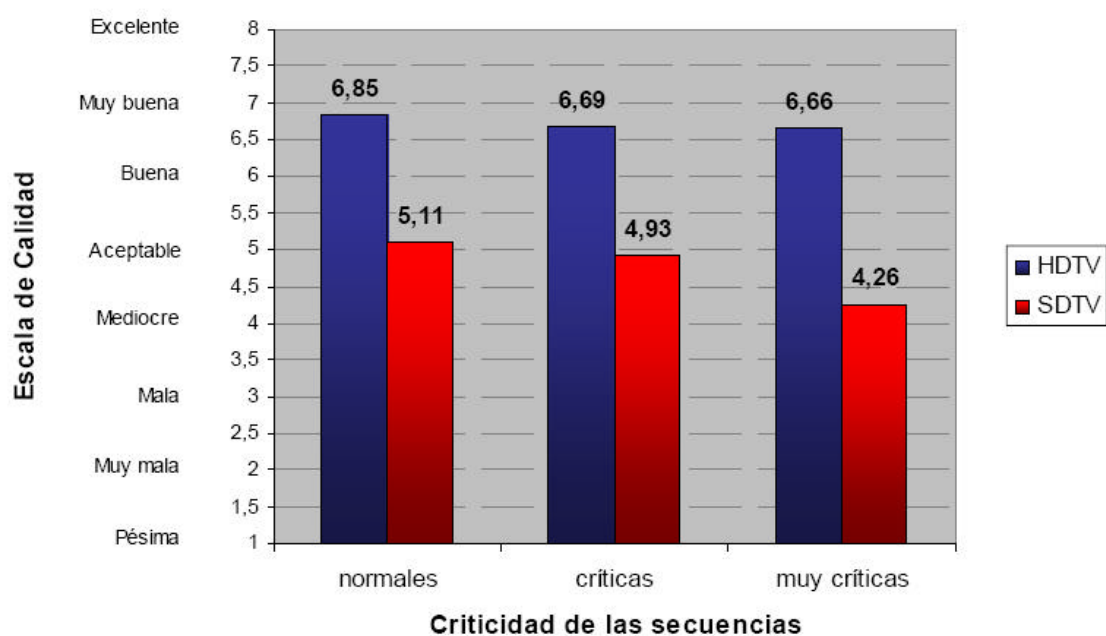
Dicho empeoramiento de calidad que se observa en la figura, es un dato bastante importante a considerar porque la calidad global de un video o secuencia de imágenes se aproxima a la calidad de la peor de dichas imágenes, o lo que es lo mismo, la calidad inferior de las secuencias muy críticas tendrá un peso mayor que una buena calidad de las secuencias normales.

Evolución de la Criticidad - HDTV



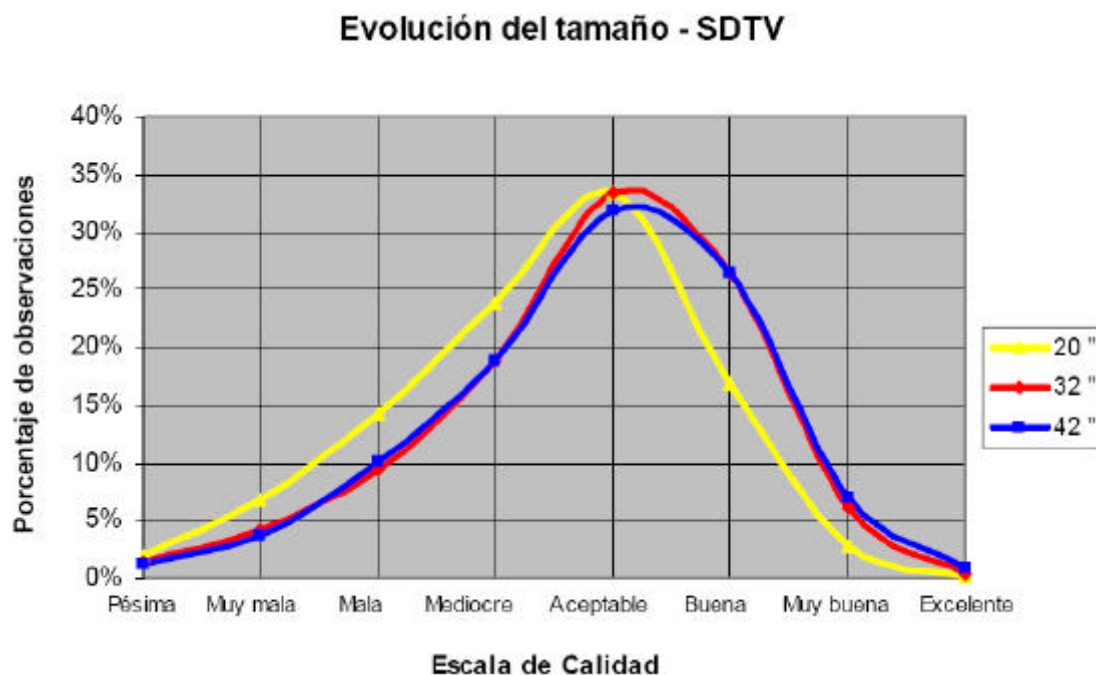
Para las secuencias de criticidad "normal" ya explicadas, la media de las valoraciones para televisión estándar se sitúa 0.85 por encima de la misma media para las secuencias "muy críticas", lo que pone de manifiesto la problemática de éstas últimas recientemente comentada.

Valores medios relativos a la criticidad



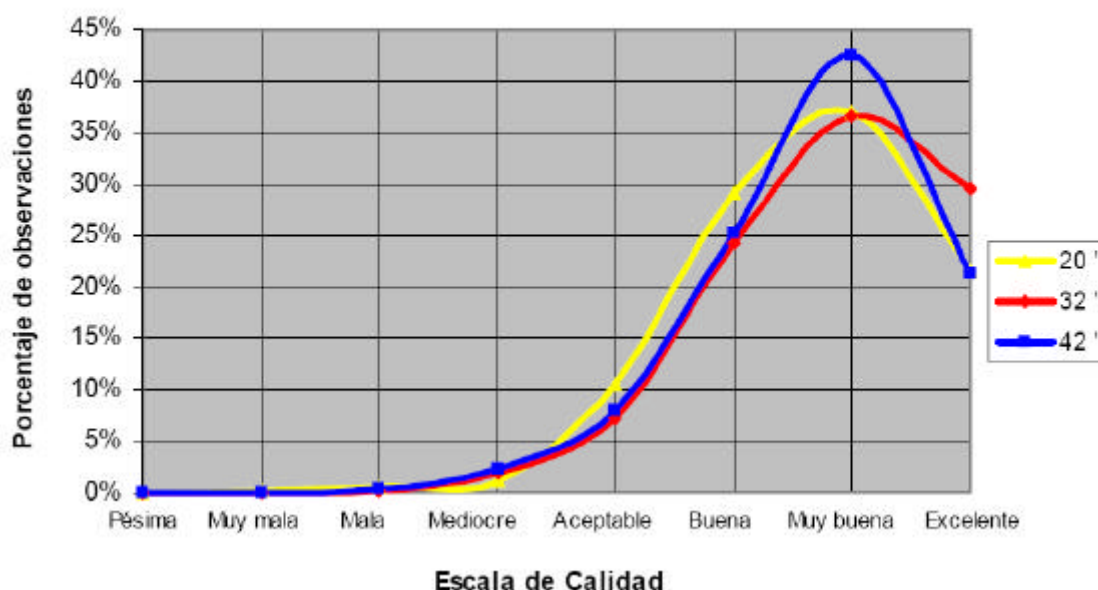
6.4 Conclusiones de la comparativa del tamaño de las pantallas

A la hora de obtener conclusiones en este apartado, hay que tener en cuenta que la distancia de visualización a cada pantalla es proporcional según la relación de la página 28 al tamaño de ésta, lo que en teoría induciría a una igualdad en la percepción solo modificada por el impacto que tiene el tamaño de la pantalla en sí.



Los resultados reflejan una débil variación de la diferencia entre calidades de alta definición y definición estándar debida al tamaño de la pantalla, es decir, no se observa un efecto negativo por el hecho de aumentar la superficie de presentación como teóricamente se produciría si la distancia de observación fuera homogénea para los diferentes tamaños.

Evolución del tamaño - HDTV

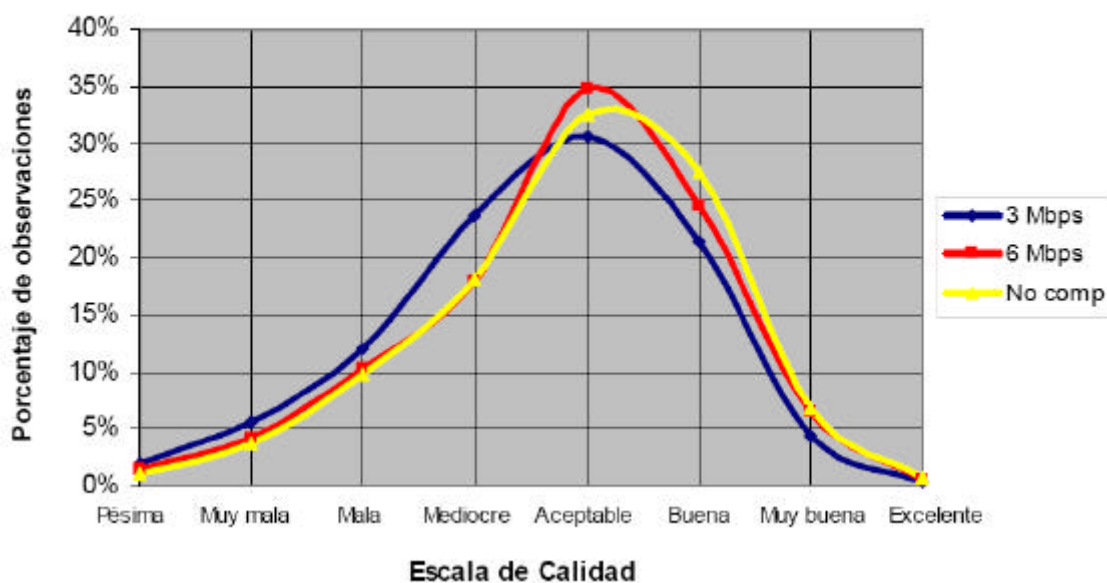


Las curvas obtenidas para esta variable, considerándola independiente de las demás son muy parejas para las dos definiciones y se mantiene en todo momento la diferencia de dos puntos en nuestra escala de calidad para los tres tamaños analizados. En alta definición además se percibe como para el caso de mayor superficie (42") se produce un pico da calidad subjetiva, máximo que se debe en parte al mayor impacto que producen estos terminales de grandes dimensiones.

6.5 Conclusiones de la comparativa de la tasa binaria de codificación

La elección de los bitrates o tasas binarias de codificación se ha realizado de tal manera que se pudiera comparar directamente alta definición con estándar independientemente del bitrate (secuencias sin comprimir) y posteriormente se ha elegido aquellos bitrates que actualmente ofrecen "garantías buenas" en la emisión de contenidos (SD en MPEG -2 y HD en MPEG -4) y aquellos que se encuentran en el límite inferior estimado para poder emitir.

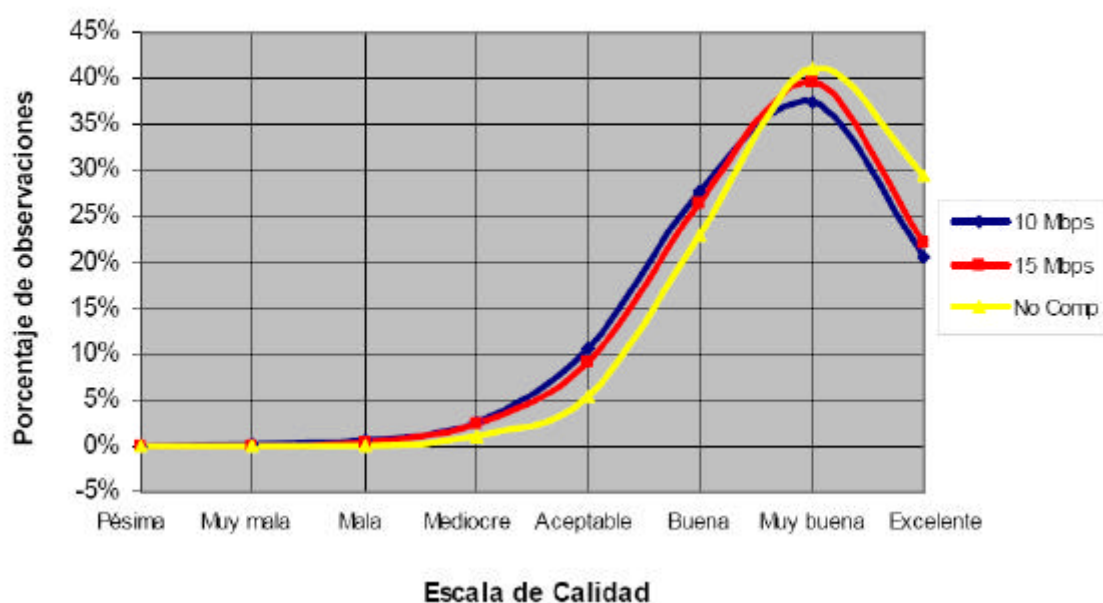
Evolución de la tasa binaria - SDTV



Los resultados obtenidos para alta definición son muy parejos. Puesto que la tasa binaria inferior usada fue de 10 Mbps, este estudio nos refleja que, a la hora de comparar con contenidos en definición estándar, con esa tasa binaria y los actuales codificadores ya se podrían obtener resultados bastante buenos y un incremento en la tasa binaria de codificación en alta definición no produciría un aumento de la distancia existente entre ambos formatos.

En definición estándar se observa una mayor influencia de la tasa binaria (de entre las tres elegidas) a medida que se llega a tasas inferiores, pero ese empeoramiento como ya se ha explicado se ve enmascarado por la comparación SD – HD por lo que no se puede extrapolar a una definición por separado.

Evolución de la tasa binaria - HDTV

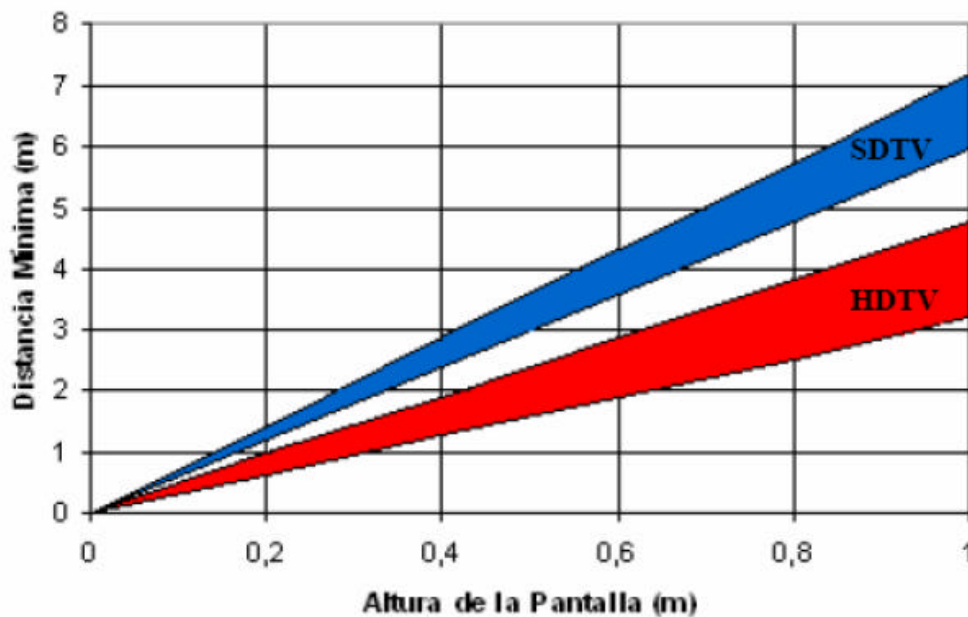


Se observa así mismo como a medida que las tasas binarias descienden, la distancia entre la SD y la HD aumenta ligeramente, puesto que en alta definición se mantiene más o menos constante pero en definición estándar varía ligeramente como se acaba de explicar.

6.6 Conclusiones de la estimación de las distancias preferidas de visualización

Para la realización de las pruebas los observadores se situaron a la distancia mínima de visualización para cada tamaño de pantalla. Dicha distancia no se suele corresponder con la distancia preferida de visualización.

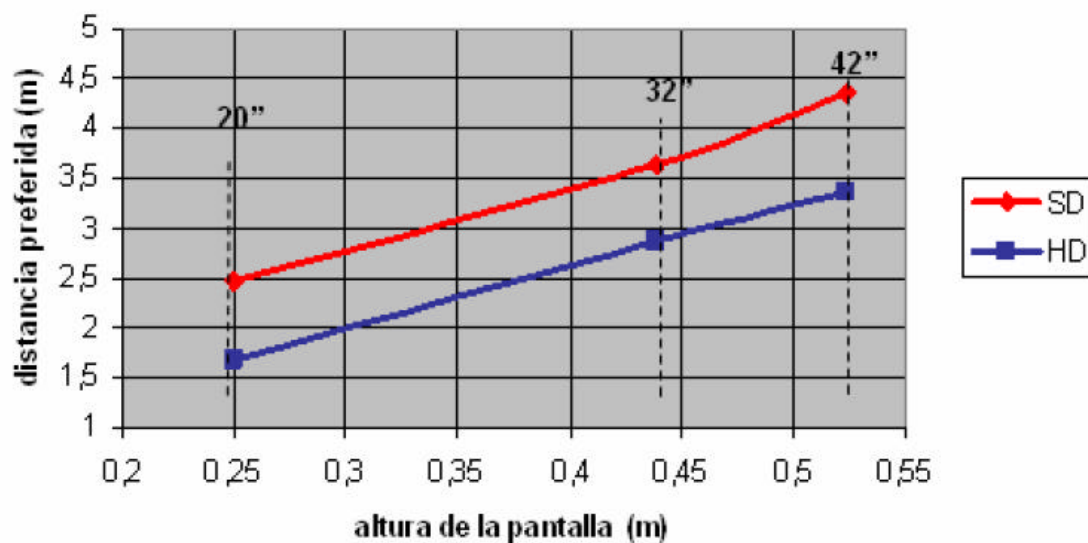
Rango Distancias de Visualización



Tras calcular la media de las distancias preferidas de visualización para las diferentes definiciones se obtiene que, al igual que se predijo de forma teórica, la distancia aumenta con el tamaño de la pantalla de forma lineal (según los tres tamaños analizados) y la diferencia entre ambas se mantiene más o menos constante en lugar de seguir un incremento proporcional al tamaño.

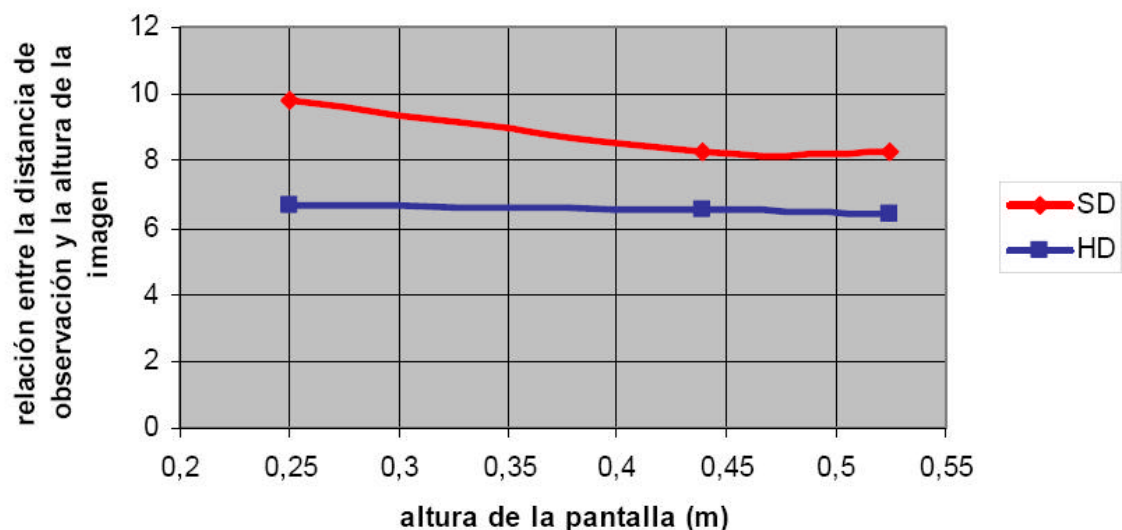
Si comparamos las distancias preferidas con las distancias mínimas a las que se situaron los observadores, se obtiene un incremento considerable de la distancia para las primeras, incremento ahora sí proporcional al tamaño de la pantalla. Es decir, a medida que aumenta el tamaño de la pantalla, la distancia preferida media de los observadores se aleja cada vez más de la distancia mínima.

Evolución de la distancia preferida según el tamaño



Si destacamos datos, para pantallas de 42" la distancia preferida media para visualizar contenidos de definición estándar es de 4.35 metros, distancia que se sale de las dimensiones de un salón de un hogar medio español, lo que indica que para este tamaño de terminal, los usuarios domésticos visualizarían los contenidos a una distancia menor a la que les gustaría .

Distancia preferida (PVD)



El descenso de la distancia preferida de visualización para los contenidos de alta definición lleva implícito un aumento de los ángulos de visualización (tanto horizontal como vertical) que deriva en una mejor percepción si cabe de la señal por la sensación de estar más metido en la imagen y la eliminación de posibles interferencias visuales externas.

Para terminar, se ha valorado también la distancia a la que los observadores visualizan con calidad semejante la alta definición y la definición estándar. Para los tres puntos obtenidos (correspondientes a los tres tamaños de monitor) la curva conseguida también sigue una progresión lineal, con una pendiente mayor que las anteriores. Esto significa que para mayores dimensiones la distancia a la que se observan las mismas calidades es proporcionalmente mayor que para dimensiones medianas y pequeñas y en ambos casos toman valores no aceptados en la mayoría de los hogares españoles. Estas distancias tan grandes acentúan, desde otro punto de vista, la diferencia actual existente entre definición estándar y alta definición.

Igualdad en la percepción SD/HD

