

PRUEBAS DE EVALUACIÓN DE CALIDAD SUBJETIVA HDTV

Aritz Legarretaetxebarria, Josu Pérez, Maider Laka, Julen García

alegarreta@vicomtech.org, jperez@vicomtech.org, mlaka@vicomtech.org, jgarcia@vicomtech.org

Dpto. de TV digital y Servicios Multimedia, Vicomtech, Mikeletegi Pasealekua 57, 2009 Donostia, Spain.

Abstract- In the context of the research Project FURIA and with the aim of clarifying the optimum broadcasting parameters for HDTV content in Spain, we have carried out some subjective quality tests. The test has considered different types and nature and configuration parameters (resolution, bitrate and so on) and has been performed following the ITU-R BT.500-11 standard. In this article we describe the evaluation environment, the evaluation procedure, the tests performed, the results of the evaluation and the main conclusions derived from them. Those conclusions have been customized for the Spanish broadcasting context, aiming to support the broadcasting engineers in the definition of the HDTV broadcasting parameters.

I. INTRODUCCIÓN

La televisión en alta definición (HDTV, High Definition Television) tiene cada vez mayor penetración en el entorno del hogar, principalmente debido a la disponibilidad de pantallas de alta definición y el incremento de los contenidos de alta definición.

En todo el mundo se ha comenzado a emitir contenidos en HD mediante la televisión digital terrestre, satélite y por cable. A nivel estatal, la televisión digital terrestre ya ha comenzado a emitir en HD en diferentes cadenas autonómicas. Este es el caso de Aragón Televisión (2006), TV3 (2007) y la televisión de la Región de Murcia (2009), incluso la televisión pública TVE (2009) ha realizado algunas pruebas. En el caso de la televisión por satélite, la plataforma Digital+ comenzó sus emisiones en HD en el 2008 con el canal Canal+ HD y ha aumentado la oferta en varios canales más: Fox HD, Canal+ Liga HD, Canal+ HD Acción y HD Music. Y puntualmente TVE HD emitió los Juegos Olímpicos de Pekín mediante las plataformas de Canal+ e Imagenio.

En cuanto a los parámetros de emisión, hay diversidad de opciones. Antes de entrar en materia, hacemos un resumen que pretende explicar la terminología técnica empleada en el artículo.

- 720p/50 es un formato de alta definición con 720 líneas horizontales y 1280 píxeles por línea con un escaneado progresivo a 50 fotogramas por segundo [1].

- 1080i/25 es un formato de alta definición con 1080 líneas horizontales y 1920 píxeles por línea, con un escaneado entrelazado a 25 o 50 fotogramas por segundo [2].

En las primeras pruebas, Aragón Televisión emitía tanto en 720/50p como en 1080/25i usando la compresión MPEG4/H.264 a 11Mbps, en cambio TV3HD emitía en formato MPEG4/H.264 a 10,2Mbps en su primera fase con una resolución de 1440×1080i y TVE HD emite en 1080/25i

codificado en MPEG-4/H.264 AVC (Advanced video coding) a unos 10 Mbps.

En este sentido, ha habido gran actividad a nivel científico-técnico a la hora de definir cuál sería el mejor modo para emitir señales en alta calidad valorando las tendencias de las nuevas tecnologías [3]-[6]. La Unión Europea de Radiodifusión (EBU) hizo una recomendación a favor de un formato de emisión de escaneo progresivo [7], con 1280x720 píxeles de resolución y una frecuencia de 50 Hz (720p/50), aunque los decodificadores y las pantallas tengan que estar también preparadas para señales de 1080i/25 [8]. A pesar de estos estudios y de la recomendación de la EBU, existen fundadas incertidumbres dentro del sector audiovisual con respecto a la resolución óptima, tasas de compresión, estándares de codificación, etc. Es por ello que se han llevado a cabo unas pruebas de calidad subjetiva enmarcadas dentro del proyecto singular Furia, cuyo objetivo principal es desarrollar y validar la integración de las tecnologías emergentes para la difusión de contenidos audiovisuales. Para ello se ha tomado como referencia la recomendación ITU-R BT.500-11 (International Telecommunication Union, Radiocommunication Sector)[9].

A continuación se describen estas pruebas, los resultados de las mismas y las conclusiones extraídas de esos resultados, contextualizados en el trabajo previo identificado.

II. CONDICIONES GENERALES DEL ENTORNO DE EVALUACIÓN

A continuación se describe cómo ha sido acondicionada la sala para el correcto visionado de las secuencias de las pruebas.

A. Sala de Pruebas

En primer lugar, se identificaron las condiciones necesarias para acondicionar la sala donde se iban a realizar las pruebas, conforme a la norma ITU-R BT.500-11. En esta norma se definen dos entornos diferentes, laboratorio y hogar, con diferentes condiciones de visión. El entorno orientado al hogar, que es el que se ha seleccionado para hacer estas pruebas, tiene como objetivo evaluar sistemas bajo unas condiciones que se pueden dar en un ambiente de hogar familiar medio.

B. Selección y caracterización de la pantalla

Las pantallas domésticas basadas en tubos de rayos catódicos (CRT Cathode Ray Tube) se descartaron debido a que estas no son capaces de reproducir digitalmente imágenes en HD, ni escaneos progresivos. Siguiendo los requisitos de los radiodifusores del proyecto y teniendo en cuenta que en

una pantalla PDP (plasma display panel) el ángulo de visión crítico es menor que en una pantalla LCD, las pruebas se realizaron con pantallas LCD.

Salmon y Drewery [4] llegaron a la conclusión de que para una imagen escaneada progresivamente a 1920x1028, hacía falta una pantalla de al menos 50 pulgadas para que el ojo humano sea capaz de apreciar toda la información mostrada en pantalla. Pero a su vez según la consultora GFK (*Gesellschaft für Konsumforschung*) [10], en la industria de las televisiones, los de 32 pulgadas son los más vendidos, seguidos por los de 40 y 42.

Teniendo esto en cuenta se han seleccionado dos televisores LCD, Panasonic TX-37LZD80F de 37 pulgadas y Full HD. La luminancia de pico máxima, se establece en 100cd/m² en una superficie blanca de la pantalla de unos 1800 cm² [11]. Para proporcionar un procesamiento equilibrado para ambos formatos, se desactivan las funcionalidades de reducción de ruido artificial y las de mejora de imagen.

C. Ubicación de los evaluadores

Antes de cada sesión de pruebas, se examinó la agudeza visual y el daltonismo de los evaluadores, basándose en las pruebas de Snellen (Figura 2), para examinar la agudeza visual, y la de Ishihara (Figura 1), para comprobar que la percepción de los colores es normal.

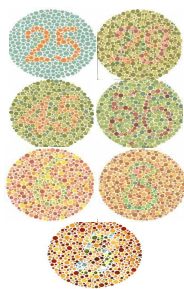


Figura 1. Grafica de Ishihara

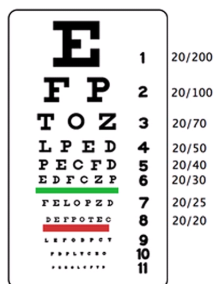


Figura 2. Grafica de Snellen

La prueba de Snellen valora la agudeza visual en función del grado de éxito en la identificación correcta de las letras en una gráfica conocida como "Gráfica de Snellen".

Con la finalidad de facilitar el posicionamiento de los observadores y para el tamaño de las pantallas elegidas, 37", se creó la Tabla 1 que se muestra a continuación, para las resoluciones de 720p y 1080i. Todo ello basándose en la herramienta de cálculo "Home Theater Plasma / LCD / Projector-&-Screen Room Design Calculator" [12].

La iluminación en la habitación no superaba el 10% del brillo máximo de la pantalla donde se reprodujeron las secuencias y se tuvo sumo cuidado en que ningún reflejo apareciera en la superficie frontal de dicha pantalla.

Agudeza/Resolución vertical	720p	1080i
20/50	95 cm	63 cm
20/40	118 cm	79 cm
20/30	159 cm	106 cm
20/20	238 cm	162 cm
20/15	317 cm	211 cm
20/10	475 cm	317 cm
20/7	679 cm	453 cm

Tabla 1. Distancias recomendadas para diferentes resoluciones, en pantallas de 37", según la agudeza visual

Con los datos obtenidos en cada prueba de Snellen se han colocado los observadores en frente de las pantallas, situando dos observadores por pantalla, para no superar el ángulo de visión máximo de 30° relativo al plano normal, que se describe en la recomendación de la ITU-T BT.500-11.

III. SELECCIÓN DE EQUIPAMIENTO Y CONTENIDOS

En esta sección se detalla la arquitectura de los equipos utilizados para mostrar las secuencias, así como la selección y el tratamiento de las mismas.

A. Servidor y arquitectura

La arquitectura de los equipos para realizar las pruebas, es la que se puede observar en la Figura 3. Se conecta la salida HD-SDI de la tarjeta externa del playout, en nuestro caso un HP xw8400 Workstation, a la entrada HD-SDI del convertidor BlackMagic HDLink Pro, que se encarga de enviar por su salida del loop la misma señal a otro convertidor de formatos, en este caso el Ediol VC300HD de Roland. Estos dispositivos convertirán esta señal y la sacan por las pantallas, en formato HDMI.

Para estas pruebas se ha permitido un escalado de la señal, lo que significa que se ha subido la escala de la señal de 720p a la escala nativa de la pantalla (up-scaling). Se podría haber mantenido la relación de pixel de la señal con la de la pantalla, pero esto nos daría una imagen más pequeña que la pantalla, lo que exigiría una reubicación de los observadores aumentando la complejidad de la prueba y haría que los evaluadores diferenciara los dos formatos a evaluar. Otro argumento para nuestra decisión de permitir el escalado fue el hecho de que en los hogares es habitual que se lleve a cabo esta práctica.

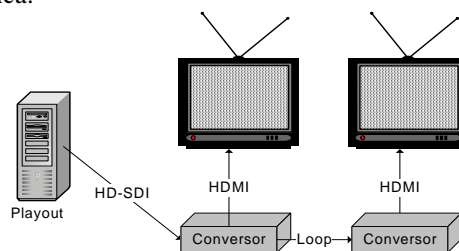


Figura 3. Arquitectura final para la segunda prueba

B. Estrategia y secuencias de visionado

Según la recomendación ITU-T BT.500-11 en las pruebas deben participar al menos 15 observadores, no expertos ni experimentados. Pero además de este grupo se ha creído oportuno contar con otro grupo de 15 observadores expertos, para contrastar los resultados obtenidos.

El grupo de evaluadores no expertos está compuesto fundamentalmente por personas de entre los 25 y 30 años con estudios superiores. Por otro lado el grupo de expertos lo conforman técnicos de video, expertos en tratamiento digital de señal, técnicos audiovisuales, productores de contenidos e ingenieros. Antes de comenzar se explicó la dinámica de la evaluación, mostrándoles unas secuencias de prueba.

Según la recomendación de la ITU-T BT.500 hay dos variantes para evaluar la degradación de una señal respecto a la señal de referencia. Se ha optado por la variante I de la primera de ellas "Double Stimulus Impairment Scale (DSIS), Variant I" mostrada en la Figura 4, que no tiene ninguna

repetición, a diferencia de la variante II que consta de una repetición.



Número de Test	2 segundos
 Secuencia A	10 segundos
Gris	3 segundos
 Secuencia A'	10 segundos
Tiempo de voto (Gris)	10 segundos
Tiempo total: 35 segundos	

Figura 4. Ciclo de las secuencias

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, Tabla 2.

Escala de cinco notas
Calidad
5. Imperceptible
4. Perceptible pero no molesta
3. Ligeramente molesta
2. Molesta
1. Muy molesta

Tabla 2. Escala de apreciación de cinco notas

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 frente de cada pantalla se encuentran sentados dos evaluadores y cada uno de estos deberá hacer la evaluación fijándose únicamente en la pantalla que tiene delante.

Los evaluadores no habían sido informados sobre cuál iba a ser el orden de los formatos HDTV que tenían que evaluar (ni su formato ni su tasa de compresión).

C. Codificación MPEG-4 AVC para las Secuencias

Se han codificado señales de referencia de 1080i y 720p sin compresión, con una profundidad de imagen de 24 bits por pixel, con unas tasas de compresión de 15 Mbit/s, 12 Mbit/s, 9 Mbit/s y 6 Mbit/s.

La codificación de los contenidos se ha llevado a cabo con el codificador profesional Tandberg EN8090, tanto para el formato 1080i como para el 720p. El tipo de codificación ha sido MPEG-4 AVC HD a 25 fotogramas por segundo para el formato 1080i y 50 fotogramas por segundo para el formato 720p.

D. Contenidos de las secuencias de evaluación

La selección del contenido adecuado para las pruebas era una cuestión fundamental y se decidió utilizar contenido crítico (aun sin llegar a ser excesivo), con detalles complejos y con secuencias en las que el movimiento era fundamental, como pueden ser imágenes deportivas. Además, el contenido que debe estar disponible en los dos formatos de HDTV expuesto a prueba debe ser idéntico. Por lo tanto, era necesario utilizar secuencias generadas con una sola cámara y una resolución lo suficientemente alta para permitir la conversión hacia abajo "downconversion" de los dos formatos

de HDTV. La Televisión Sueca (SVT) ha generado una película de 65 mm a 50 frames por segundo, evitando así los movimientos típicos de cine. El material es conocido con el nombre de "SVT High Definition Multi Format Test Set" [13]. Los contenidos utilizados son los siguientes: Crowd Run, Park Joy, Ducks Take Off, In To Tree y Old Town Cross.

IV. PROCEDIMIENTO Y RESULTADOS DE LA EVALUACIÓN

A continuación se describen los procedimientos que se han seguido para evaluar los tests y sus correspondientes resultados.

A. Procedimiento de evaluación

La votación se llevo a cabo sobre papel, en el cual cada observador incluya los datos individualmente, según la planificación realizada por el supervisor. Todos los datos fueron transferidos a un archivo Excel, pero los primeros cuatro votos no computaron en las medias, para asegurarse de que los evaluadores tenían un margen de tiempo para familiarizarse con el contenido y el método de evaluación.

Para el cálculo de las medias se ha empleado:

$$\bar{u}_{jkr} = \frac{1}{N} \sum_{i=1}^N u_{ijkr} \quad (1)$$

Siendo u_{ijkr} la puntuación del evaluador i , para el test j , la secuencia k y la repetición r .

Siendo N el número de evaluadores.

La ITU-R BT.500-11 recomienda un valor de confianza del 95%, siendo el nivel de confianza empleada:

$$\delta_{jkr} = 1.96 \frac{S_{jkr}}{\sqrt{N}} \quad (2)$$

Con lo que la desviación estándar empleada es:

$$S_{jkr} = \sqrt{\frac{\sum_{i=1}^N (\bar{u}_{jkr} - u_{ijkr})^2}{(N-1)}} \quad (3)$$

B. Resultados de los test

En los siguientes gráficos se muestran las comparaciones entre los dos formatos analizados, 720p y 1080i, con diferentes tasas de compresión. En el Gráfico 1 se muestran los resultados globales. Se puede observar la puntuación media obtenida por cada formato y la desviación típica después de aplicarle el coeficiente de intervalo de 95%, para cada tasa de compresión.

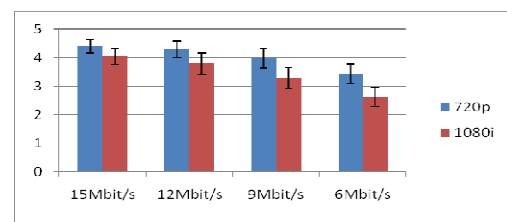


Gráfico 1. Resultados de los tests

En cuanto a las tasas de compresión, los resultados obtenidos son muy similares, pero en general, los vídeos en modo 720p han obtenido mejores puntuaciones.

Se muestran también los resultados desglosados entre los observadores expertos (Gráfico 2) y los no expertos (Gráfico 3).

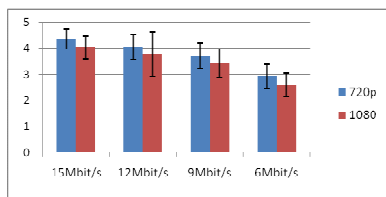


Gráfico 2. Resultados de los tests de observadores expertos

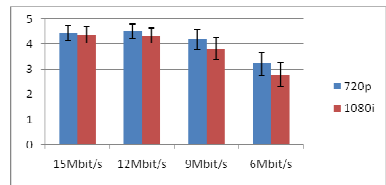


Gráfico 3. Resultados de los tests de observadores no expertos

Se mantienen las tendencias respecto a los resultados globales, pero el grupo de los expertos es más crítico respecto al grupo de los no expertos. Lo que más llama la atención, es que la degradación sufrida por la señal a 12 y a 15Mbps es prácticamente la misma en ambos modos para los evaluadores no expertos, conllevando a pensar que a partir de estas tasas de compresión un usuario común no aprecia la diferencia. Por lo tanto una tasa de compresión mínima de unos 11 o 12Mbps sería más que suficiente para ser emitida en alta definición, para una buena calidad de imagen en contenidos críticos.

Las pruebas han reforzado las conclusiones de una investigación anterior de Haglund [6], en las que se codificaron diversas secuencias de los formatos 720p/50 y 1080i/25 con una compresión MPEG-2. En estas pruebas se constató que las características por defecto del formato 720p/50 eran mejores que las del 1080i/25 y se pudo apreciar la preferencia de los observadores para este primer formato.

Nuestras pruebas han ampliado esta investigación inicial mediante el uso de compresión MPEG-4 AVC, unas características de visualización específicas y un método de evaluación específico. Las pruebas presentadas no deben interpretarse como pruebas de comparación directa entre los modos de 720p y 1080i, ya que hemos investigado cada formato de manera individual con el método de doble estímulo.

En todas las tasas de compresión, los resultados de los evaluadores son favorables al modo de 720p, incluso a una tasa de compresión de 6 Mbit/s, el resultado quedó por encima de "ligeramente molesta". Una posible justificación de estos resultados, se debe a que un codificador MPEG-4 AVC comprime señales progresivas de manera más eficiente que señales entrelazadas.

V. CONCLUSIONES Y RECOMENDACIONES PARATRAJOS FUTUROS

Una calidad de imagen aceptable, basándose en la escala de apreciación que se muestra en la Tabla 2, se situaría en una puntuación de "4". Según las evaluaciones realizadas para el modo de 720p, las tasas de compresión a 9 Mbps se

sitúan muy cercanas a este valor. En cambio para 1080i, habría que irse hasta los 12Mbps para poder conseguir la misma calidad de imagen.

Según los resultados de las pruebas, en el contexto español, con una clara presencia de la televisión digital terrestre, la resolución 1280 x 720 con un escaneado progresivo es la mejor opción para la transmisión terrestre, donde hay menor ancho de banda para el vídeo en comparación con el satélite y el cable. Actualmente la configuración de cada canal de la TDT tiene un ancho de banda disponible cercano a los 20 Mbps. Con la codificación empleada en las pruebas y conforme a los resultados se podrían multiplexar dos programas HD 720p, siendo muy recomendable aplicar multiplexación estadística.

Se plantea como trabajo futuro la realización de estas pruebas visualizando los contenidos en unas pantallas con un tamaño mínimo de 50 pulgadas, con el fin de comparar los resultados con los obtenidos en este artículo con Salmon y Drewery [4].

También se deberían realizar estas pruebas con el método de evaluación de Hoffmann [11], con la finalidad de poder contrastar los resultados obtenidos mediante diferentes métodos de evaluación.

AGRADECIMIENTOS

Los autores de este artículo quieren agradecer a las entidades que forman parte del proyecto Furia por la cesión del equipamiento necesario para llevar a cabo estas pruebas de calidad subjetiva. El proyecto singular Furia, ha sido parcialmente financiado por el Ministerio de Industria, Turismo y Comercio (TSI-020100-2008-630).

REFERENCIAS

- [1] SMPTE 296M, 1280 _ 720 Scanning, Analog and Digital Representation and Analog Interface Society of Motion Pictures and Television Engineers. New York, Tech. Rep. SMPTE 296M,
- [2] SMPTE 274M, 1920_1080 Scanning and Analog and Parallel Digital Interfaces for Multiple Picture Rates Society of Motion Pictures and Television Engineers. New York, Tech. Rep. SMPTE 274M, 1998.
- [3] H. Hoffmann, T. Itagaki, D. Wood, and A. Bock, "Studies on the bit rate requirements for a HDTV format with 1920 1080 pixel resolution, progressive scanning at 50 Hz frame rate targeting large flat panel displays", IEEE Trans. Broadcasting, vol. 4, pp. 420-434, Dec. 2006
- [4] R. A. Salmon and J. D. Drewery, Test of Visual Acuity to Determine the Resolution Required of a Television Transmission System BBC, Tech. Rep. WHP 092, 2004
- [5] S. B. Gauntlett, Technologies for Efficient Emission of HDTV Across Europe Research and Development British Broadcasting Cooperation, UK, Tech. Rep. WHP104, 2005
- [6] L. Haglund, Overall Quality Assessment When Targeting Wide XGA Flat Panel Displays SVT Corporate development technology. Stockholm/ Munich, 2002
- [7] EBU-R112, EBU Statement on HDTV Standards European Broadcasting Union. Geneva, Tech. Rep. R-112, 2004
- [8] European Industry Association EICTA, Conditions for High Definition Labelling of Display Devices January 19, 2005, 2005
- [9] ITU-R BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures International Telecommunication Union. Geneva, Tech. Rep. ITU-R BT.500-11, 2003
- [10] Impulsa TDT [Online]: <http://www.impulsatdt.es/>
- [11] H. Hoffmann, T. Itagaki, D. Wood, Tobias Hinz, and Thomas Wiegand, "A Novel Method for Subjective Picture Quality Assessment and Further Studies of HDTV Formats", IEEE Trans. Broadcasting, Vol. 52, No. 4, Dec. 2006
- [12] Carlton Bale, 2008 Version 4.0, [Online].
- [13] L. Haglund, 2006, The SVT High Definition Multi Format Test Set. Stockholm, Swedish Television. Stockholm