

# Modelo de Estimación de Calidad de Video: Video Quality Experts Groups

*Video Quality Estimation Models:*

*Video Quality Experts Groups*

---

*José Joskowicz<sup>1</sup>, Rafael Sotelo<sup>2</sup>*

Recibido: Agosto 2011

Aprobado: Junio 2012

---

**Resumen.-** El video digital, así como los contenidos multimedia, distribuidos a través de redes de telecomunicaciones, experimentan varios tipos de distorsiones o degradaciones durante el proceso de adquisición, compresión, procesamiento, transmisión y reproducción. Uno de los factores claves en el éxito de un sistema o servicio multimedia es poder lograr que la calidad de la experiencia percibida por el usuario resulte aceptable. La calidad de video es uno de los principales factores en la calidad global experimentada por los usuarios. Por tanto, para los operadores es necesario contar con herramientas que le permitan saber la clase de servicio que brindan a sus clientes. A los efectos de poder medir o estimar la calidad del video percibida, en la última década se han propuesto diversos modelos y se han realizado varias estandarizaciones. En este proceso, se destaca el trabajo del "Video Quality Experts Group" (VQEG), organismo que fue creado con el objetivo de avanzar en el campo de la evaluación de calidad de video, mediante la investigación de nuevas y avanzadas técnicas y métodos de evaluación, tanto subjetivas como objetivas. En el presente trabajo se realiza una revisión del trabajo del VQEG y de los avances y estandarizaciones recientemente realizadas en modelos de estimación de calidad de video.

**Palabras clave:** Calidad de Video; VQEG.

***Summary.-** The digital video and multimedia content distributed via telecommunication networks suffers from several types of distortion or degradation during the process of acquisition, compression, processing, transmission and reproduction. One of the key factors in the success of a multimedia system or service is to achieve an acceptable quality of experience for the user, depending on the application. Video quality is one of the key factors in the overall quality of the experience perceived by users. For the purposes of measuring or estimating the video quality perceived by users, in the last decade various models have been proposed and there have been several standardizations. The "Video Quality Experts Group" (VQEG) is an organization that was created with the goal of advance the field of video quality assessment by investigating new and advanced subjective and objective quality metrics and measurement techniques. In this paper we review the work performed by the VQEG, the progress in this area and the recently completed models for estimating video quality.*

**Keywords:** Video Quality; VQEG.

**1. Introducción.-** La utilización de aplicaciones de video y multimedia está creciendo rápidamente en la vida cotidiana. La televisión tiene ya décadas de funcionamiento, pero están emergiendo nuevos estándares, relacionados a la televisión digital. En el mercado de consumo masivo, diferentes proveedores están ofreciendo servicios y aplicaciones de video y contenido

---

<sup>1</sup> Dr. (c) Ing. José Joskowicz. Universidad de la República, Montevideo, Uruguay, josej@fing.edu.uy

<sup>2</sup> Dr. Ing. Rafael Sotelo. Universidad de Montevideo, Montevideo, Uruguay, rsotelo@um.edu.uy

multimedia a usuarios finales, incluyendo los distribuidores de televisión por cable, proveedores de Internet, y operadores de telefonía. En el mercado corporativo las aplicaciones de telefonía están bien establecidas y diversas aplicaciones con contenido de video están emergiendo. Video teléfonos, video conferencias y entrenamiento a distancia son algunos ejemplos. En todos estos escenarios, es fundamental garantizar una “Calidad de la Experiencia” (o QoE por sus siglas en inglés) apropiada al usuario final, de acuerdo al tipo de aplicación o servicio de que se trate. La QoE puede ser definida como el desempeño general de un sistema, visto desde la perspectiva del usuario. Diversos factores pueden afectar a la Calidad de la Experiencia, dependiendo de cuál sea la aplicación, el servicio y las expectativas que de ellos tengan los usuarios. Sin embargo, la calidad que perciben los usuarios del audio y el video son de los aspectos más importantes a considerar en la calidad final de la experiencia del usuario.

En una transmisión de video digital, cualquiera sea el medio, existe una diferencia de calidad entre el video original, previo a codificarse y transmitirse y el que llega al espectador o usuario, debido a degradaciones sufridas durante el proceso de adquisición, compresión, procesamiento, transmisión y reproducción. Por ejemplo, las técnicas utilizadas habitualmente en la codificación digital de video, para reducir el ancho de banda necesario para su transmisión, producen la pérdida de parte de la información que describe las imágenes originales. Ello genera distorsiones en las imágenes resultantes. Por otro lado, el sistema de transmisión sobre las que se transporta el video (por ejemplo redes de transmisión inalámbricas, redes de paquetes, Internet, redes de área local LAN o redes extendidas WAN), pueden introducir distorsiones adicionales, debido a las demoras, los errores y las pérdidas de paquetes, entre otros factores.

Conocer, poder medir y eventualmente predecir la calidad de video y contenido multimedia percibida por los usuarios es necesario para diversos tipos de aplicaciones, entre las que se pueden mencionar la monitorización, el cumplimiento de acuerdos de servicio basados en la calidad, el diseño de nuevas redes y la evaluación de redes existentes para el soporte de contenido multimedia. Todos estos factores hacen necesario disponer de herramientas que permitan estimar y cuantificar la calidad percibida por los usuarios en contenidos multimedia, de la manera más confiable posible. Las evaluaciones subjetivas son siempre los sistemas de medida más confiables, ya que en dichas evaluaciones se toman directamente la opinión de los usuarios, y se promedia entre un número apropiado de observaciones, obteniendo típicamente como métrica el promedio de opiniones (MOS, Mean Opinion Score). Existen estándares que indican cómo realizar estas evaluaciones. Sin embargo, su realización es costosa y compleja ya que es necesario implementarlas en ambientes controlados, disponer de un número importante de usuarios y de contenidos multimedia especialmente preparados. Adicionalmente no responde a la necesidad de evaluar en tiempo real en muchos escenarios posibles. Es por ello que en los últimos años se han comenzado a desarrollar sistemas, algoritmos y modelos matemáticos que permitan estimar la calidad percibida por los usuarios. Un sistema ideal de estimación de calidad percibida debería dar como resultado una calificación idéntica a la que se obtendría en pruebas subjetivas promediando los resultados de un gran número de individuos. Diversos modelos de este tipo han sido propuestos recientemente, y algunos de ellos han sido estandarizados por organismos internacionales como la ITU (International Telecommunications Union). El trabajo del VQEG (Video Quality Experts Group) se ha destacado especialmente, en el proceso de análisis y comparación de modelos de estimación de calidad de video, y en las sugerencias acerca de las estandarizaciones a realizar. En el presente trabajo se describe la evolución y estado actual de los avances y las estandarizaciones de modelos de estimación de calidad de video, realizando una revisión de los proyectos realizados y en curso del VQEG.

## **2. Métodos para medir la calidad percibida de video**

**2.1. Métodos Subjetivos.-** *La manera más confiable de medir la calidad de una imagen o un video es la evaluación subjetiva, realizada por un conjunto de personas que opinan acerca de su percepción. La opinión media, obtenida en forma directa mediante el “MOS” (Mean Opinion*

Score) o en forma comparativa mediante el “DMOS” (Degradation MOS) son las métricas generalmente aceptadas como medida de la calidad. Diversos métodos subjetivos de evaluación de video son reconocidos, y están estandarizados en las recomendaciones ITU-R BT.500-11 [1], especialmente desarrollada para aplicaciones de televisión y ITU-T P.910 [2], para aplicaciones multimedia. En todos los métodos propuestos, los evaluadores son individuos que juzgan la calidad en base a su propia percepción y experiencia previa.

La recomendación BT.500-11 detalla los métodos DSIS (Double Stimulus Impairment Scale), DSCQS, (Double Stimulus Continuous Quality Scale) SSCQE (Single Stimulus Continuous Quality Evaluation) y SDSCE (Simultaneous Double Stimulus for Continuous Evaluation). Por su parte, la recomendación P.910 describe los métodos ACR (Absolute Category Rating), DCR (Degradation Category Rating) y PR (Pair Comparison).

Las escalas de calidad utilizadas por los diferentes métodos subjetivos pueden ser continuas o discretas (típicamente entre 5 y 11 valores), según el caso. Para el MOS, la métrica más difundida, generalmente es aceptada una escala de 5 puntos, del 1 al 5, siendo 1 la peor puntuación (indicando calidad “Mala”) y 5 la mejor puntuación (indicando calidad “Excelente”).

**2.2. Métodos Objetivos.-** Los métodos subjetivos mencionados en la sección anterior son costosos, difíciles de realizar, e impracticables en aplicaciones de tiempo real. Por esto se hace necesario el uso de métodos objetivos y automáticos, que puedan predecir con fiabilidad la calidad percibida (i.e., el MOS), en base a medidas objetivas tomadas en algún punto del sistema.

**2.2.1. Métricas Básicas.-** Las medidas objetivas más sencillas de estimación de la calidad del video están basadas en obtener las diferencias, píxel a píxel, entre las imágenes originales (previo a la compresión y transmisión) y las imágenes degradadas (luego de la recepción y reconstrucción). Las imágenes presentadas serán diferentes a las originales debido a que los sistemas de video utilizan técnicas de compresión con pérdida de información, y que los medios de transmisión a su vez pueden introducir factores distorsionantes (retardos, pérdida de paquetes, etc.).

Las medidas más simples son las de error cuadrático medio MSE (Mean Square Error) y su raíz cuadrada RMSE (Root Mean Square Error) y la relación señal a ruido de pico PSNR (Peak Signal to Noise Ratio), definidas en las ecuaciones (1) a (3). Estas métricas requieren de la referencia completa de la señal original para poder ser calculadas. Es decir, que para computarlas, además de la imagen recibida resultante, hace falta disponer de la imagen original.

$$MSE = \frac{1}{TMN} \sum_{n=1}^N \sum_{m=1}^M \sum_{t=1}^T [x(m,n,t) - y(m,n,t)]^2 \quad (1)$$

$$RMSE = \sqrt{MSE} \quad (2)$$

$$PSNR = 10 \log_{10} \left( \frac{L^2}{MSE} \right) \quad (3)$$

En las ecuaciones (1) a (3) la imagen tiene N x M píxeles y T cuadros, x, y son los píxeles de la imagen original y la distorsionada respectivamente. L es el rango dinámico que pueden tomar los valores de x o y, y toma el valor 255 para 8 bits por píxel.

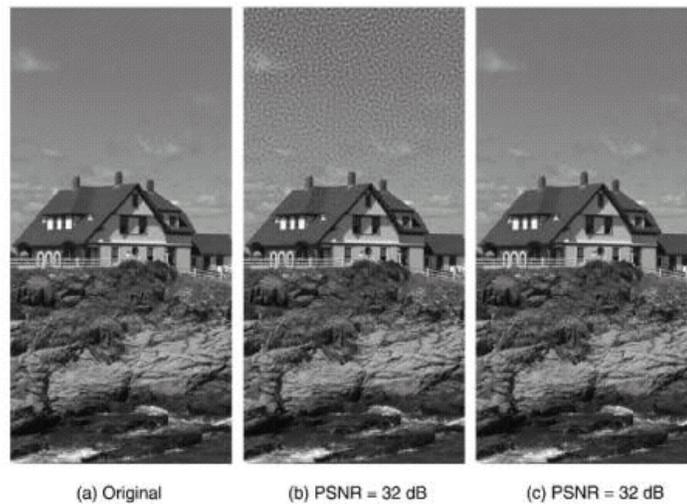
Estas métricas son fáciles de calcular, y tienen un claro significado. Por estas razones, han sido ampliamente usadas como métricas en la estimación de la calidad de video. Hay que poner

especial énfasis en la alineación espacial y temporal de las imágenes a comparar, ya que la referencia y la imagen degradada pueden estar desfasadas en el tiempo o en el espacio.

Sin embargo, también han sido ampliamente criticadas por no tener correlación directa con la calidad percibida por las personas. Por ejemplo, en la Figura I, tomada de [3], se muestran tres ejemplos de imágenes comprimidas, donde se puede ver claramente que con similares valores de MSE, la calidad percibida puede ser esencialmente diferente (comparar, por ejemplo, “Tiffany” con “Mandril”, sobre el lado derecho de la figura), lo que pone en duda la utilidad de este tipo de métrica como indicador de calidad. En la Figura II, presentada en [4], se puede ver como la misma imagen, con el mismo valor de PSNR, puede tener diferente calidad percibida, dependiendo del lugar en el que se presenten las degradaciones. En la Figura II(b), se nota claramente la degradación en el cielo (parte superior), mientras que en la Figura II(c), una degradación similar en la parte inferior prácticamente no es perceptible. Este fenómeno se conoce como “enmascaramiento”. En zonas texturadas o con gran “actividad espacial”, las degradaciones quedan “enmascaradas” y son menos percibidas por el sistema visual humano. El enmascaramiento también puede darse en video, donde cambios rápidos temporales pueden enmascarar cierta pérdida de calidad en cada cuadro.



Figura I.- Evaluaciones de imágenes. Izquierda: imagen original. Derecha: imagen degradada  
Arriba: “Tiffany”, MSE=165; Medio: “Lago”, MSE=167; Abajo: “Mandril”, MSE=163



*Figura II.- Evaluaciones de imágenes. Izquierda imagen original. Centro y Derecha: imágenes degradadas*

**2.2.2. Métodos Perceptuales.-** En los últimos tiempos se ha realizado un gran esfuerzo para desarrollar nuevos modelos que tengan en cuenta las características de percepción del sistema visual humano y que permitan calcular métricas objetivas que lo simulen, tratando de lograr mejores resultados que los que se obtienen con las métricas MSE, RMSE y PSNR. Existen varias propuestas de métricas de medida, con diversa complejidad y precisión de sus resultados. El sistema visual humano es extremadamente complejo, y puede detectar fácilmente ciertos tipos de distorsión, mientras que puede pasar por alto otras, dependiendo de diversos factores. Estos factores pueden incluir el tipo de aplicación que esté siendo utilizada (TV, video conferencia, video telefonía, etc.), el lugar de la imagen en donde se produce la degradación (generalmente las degradaciones son menos visibles en regiones con muchos detalles o “actividad espacial”, o con gran movimiento, y son más visibles en imágenes estacionarias, o en fondos poco texturados). Incluso la calidad percibida puede depender del tipo de dispositivo utilizado y del tamaño del monitor. En general, el sistema de visión humano juega un rol fundamental, y la ciencia no tiene aún una comprensión total del mismo.

En forma genérica, los métodos objetivos de medida de calidad pueden clasificarse según si lo que se analiza es el propio video (el medio o contenido) o si lo que se analiza son los parámetros de codificación y transmisión.

Los modelos basados en el análisis del contenido basan su funcionamiento en la inspección del propio video, y típicamente buscan degradaciones de diversa índole (efectos de bloques, borrosidad, etc.). Dentro de esta categoría, los modelos a su vez se clasifican según la disponibilidad total, parcial o nula de la señal original.

Los métodos con disponibilidad total de la señal original o FR (Full Reference) se basan en la disponibilidad de la señal original, la que puede ser contrastada con la señal degradada, cuadro a cuadro. Los métodos que utilizan métricas del tipo FR pueden ser utilizados para categorizar en forma objetiva un sistema de transmisión, un códec, el efecto de un reducido ancho de banda, o de diversos factores que degraden una señal, en ambientes controlados, básicamente dedicados a pruebas. Sin embargo, no son adecuados para aplicaciones de tiempo real (TV, video conferencias, etc.), ya que no es posible disponer de las señales originales junto con las señales degradadas.

Los métodos con disponibilidad parcial de la señal original o RR (Reduced Reference) se basan en enviar, en un “canal auxiliar”, junto con el video codificado, algunos parámetros que

caractericen a la señal, y que sirvan de referencia en el receptor para poder estimar la calidad percibida. Puede pensarse en la reserva de un pequeño ancho de banda (comparado con el del video) para el envío de este tipo de información adicional.

Los modelos del tipo NR (No Reference) intentan estimar la calidad percibida basándose únicamente en el análisis de la señal recibida. Son los métodos más complejos de implementar, pero no requieren de otra información que la propia señal de video. Los seres humanos no necesitan señales de referencia, ni información adicional para juzgar la calidad de una señal de video. Se basan en sus experiencias previas, y en las expectativas que tengan respecto al sistema. Los modelos del tipo NR buscan realizar el mismo trabajo que el sistema visual humano y son por tanto los más complejos de desarrollar.

En la Figura III se esquematiza la forma en que trabajan los modelos FR, RR y NR. Se observa que los modelos del tipo FR toman como entrada la señal original completa y la señal degradada, los modelos RR toman como entrada la señal degradada y un conjunto resumido de características de la señal original, a través de un canal auxiliar, y los modelos NR toman como entrada únicamente la señal degradada. Esta figura muestra claramente que un modelo del tipo NR sería el ideal, ya que no requiere que sea enviada ninguna información adicional a la señal transmitida. En segundo término un modelo del tipo RR serviría para monitorizar la calidad utilizando ciertos recursos adicionales como el “canal auxiliar”. Finalmente, un modelo FR es útil a efectos de pruebas de laboratorio, pero no en entornos reales, ya que requiere disponer en forma completa del video original.

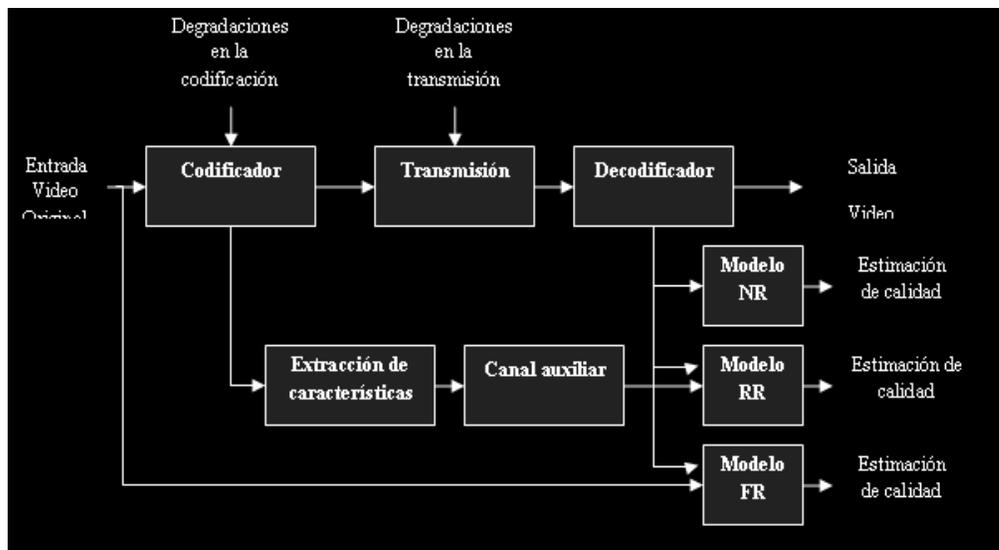


Figura III.- Esquema de modelos FR, RR y NR

**3. Los proyectos del VQEG.-** El VQEG (Video Quality Experts Group) [5] es un organismo formado en 1997 por profesionales “expertos” en la temática de calidad de video, provenientes de la industria, la academia y las organizaciones de estandarización. Tiene como objetivo proveer un foro apropiado para el intercambio de información y el trabajo en conjunto hacia fines relacionados con la evolución de modelos y métricas de estimación de calidad de video. VQEG ha desarrollado un enfoque sistemático para realizar la comparación del desempeño y la validación de las propuestas de modelos de estimación de calidad de video, sobre la base de contrastar los resultados de los modelos con pruebas subjetivas. El objetivo del VQEG es proporcionar información a los organismos internacionales de estandarización acerca del

desempeño de diversos modelos propuestos, a los efectos de definir una métrica estándar y objetiva de calidad percibida de video digital o VQM (Video Quality Metric).

El VQEG ha trabajado en la evaluación de modelos específicos para diferentes aplicaciones. Algunos de los proyectos del VQEG han terminado en Recomendaciones aprobadas por ITU. Otros se encuentran aún en proceso de evaluación o estudio. A continuación se describen los principales proyectos de VQEG.

**3.1. FR-TV (Full Reference TV).**- *El proyecto FR-TV fue el primero del VQEG, y evaluó modelos del tipo FR de estimación de calidad percibida de video para aplicaciones de TV en formato estándar (SDTV). En la fase I de FR-TV, llevada a cabo entre 1997 y 2000, se evaluaron 9 propuestas de modelos, presentadas por diversas empresas, universidades y organismos. Las evaluaciones fueron realizadas con diversos tipos de material de video, incluyendo 20 tipos diferentes de contenidos (entre los que hay deportes, animaciones, escenas de interiores y exteriores, etc.) y con velocidades desde 768 kb/s hasta 50 Mb/s. Los resultados de los modelos propuestos fueron contrastados contra los resultados de pruebas subjetivas, las que se realizaron sobre una base de más de 26.000 opiniones subjetivas, tomadas en 8 laboratorios independientes en diferentes partes del mundo, utilizando el método DSCQS.*

Como resultado de la fase I de FR-TV, dependiendo de la métrica de comparación utilizada, siete u ocho de los modelos propuestos resultaron estadísticamente equivalentes entre sí, y a su vez, equivalentes a los resultados obtenidos con el PSNR [6]. Este resultado fue realmente desalentador, ya que indica que no existen diferencias apreciables entre el sencillo cálculo del PSNR y los sofisticados métodos propuestos. En base a estos resultados, el VQEG ha realizado una segunda fase de pruebas, llamando nuevamente a interesados en contrastar sus modelos. La denominada “fase II” para FR-TV fue realizada entre los años 2001 y 2003 y los resultados finales fueron publicados en agosto de 2003 [7]. El objetivo de esta segunda fase era obtener resultados más discriminatorios que los obtenidos en la fase I. Las pruebas fueron realizadas en tres laboratorios independientes, en Canadá, Estados Unidos e Italia y se evaluaron los modelos propuestos por seis proponentes. Los resultados, presentados en [7], indicaron mejoras en los modelos propuestos de entre 17% y 21% respecto al PSNR. Sobre la base de estos resultados, ITU ha estandarizado en 2004 la Recomendación ITU-R BT.1683 [8] y la Recomendación ITU-T J.144 [9], con los cuatro mejores algoritmos: BTFR de British Telecom [10], Yonsei de la Universidad Yonsei de Corea [11], CPqD del Centro de Investigación y Desarrollo en Telecomunicaciones de Brasil [12] y NTIA de National Telecommunications and Information Administration) [13].

Varios de los métodos del tipo FR estandarizados utilizan una arquitectura interna donde las señales original y degradada son alineadas espacial y temporalmente, para luego aplicarles alguna transformada o cálculo sobre la luminancia o crominancia en cada señal, típicamente segmentadas en regiones espaciales y temporales pequeñas y delimitadas. Los datos resultantes son utilizados para calcular diferentes “características” de la imagen o el video. Dentro de los elementos utilizados por estos modelos se encuentran la detección de bordes, el contraste, el movimiento y la textura en cada región de la imagen. Para cada una de estas características se calcula un “error” (diferencia entre señal original y degradada) aplicando algún tipo de métrica. Estos errores son luego promediados, ponderados e integrados para obtener finalmente una única medida de calidad para toda la secuencia de video.

Las experiencias del proyecto FR-TV sirvieron como base para los siguientes proyectos. Adicionalmente, las bases de datos de videos de prueba y los resultados de las pruebas subjetivas han servido como referencia para gran parte de los trabajos de investigación posteriores.

**3.2. RRNR-TV (Reduced Reference/No Reference TV).**- *El proyecto RRNR-TV evaluó modelos del tipo RR y NR de estimación de calidad percibida de video para aplicaciones de TV en*

formato estándar (SDTV). En la fase I de RRNR-TV, llevada a cabo entre 2000 y 2009, se evaluaron 3 proponentes, cada uno con modelos del tipo RR y NR [14]. Las secuencias de prueba incluyeron formatos de 525 y 626 líneas, en 60 Hz y 50 Hz respectivamente, y se admitieron para los modelos RR canales de referencia de 15 kb/s, 80 kb/s y 256 kb/s. Los modelos NR, tal como ya ha sido señalado, no requieren canales de referencia, y basan sus medidas únicamente en la señal de video degradada. Se utilizaron codificaciones en MPEG-2 y H.264/AVC. Las degradaciones incluían tanto las producidas por el proceso de codificación, como por errores en la transmisión del canal.

Todos los modelos del tipo NR fueron retirados de las pruebas, y de los informes finales, debido a su bajo desempeño. De los modelos RR, algunos resultaron superiores a las métricas del tipo PSNR, y sobre la base de estos resultados, en enero de 2010 ITU-T ha estandarizado la recomendación ITU-T J.249 [15], contemplando los modelos propuestos por NTIA [16], la Universidad de Yonsei, y con ciertas restricciones el modelo propuesto por la empresa NEC de Japón.

**3.3. MM (MultiMedia).**- Si bien el término “Multimedia” se refiere a aplicaciones que pueden combinar texto, gráficos, video y sonido, en la primera fase de este proyecto del VQEG, se incluyeron modelos que evalúan únicamente la calidad de video. Los resultados y el foco de la evaluación de la fase I del proyecto MM de VQEG se centra en aplicaciones de video para dispositivos móviles o de tipo PDA (Personal Digital Assistant) y en servicios multimedia distribuidos a través de Internet, con tasas de bits menores a 4 Mb/s, vistos en pantallas de PC con una resolución máximo de  $640 \times 480$  pixels.

La fase I del proyecto MM [17] evaluó modelos de estimación de calidad percibida de video para aplicaciones multimedia en formatos VGA (Video Graphics Array,  $640 \times 480$  pixels), CIF (Common Intermediate Format,  $352 \times 288$  pixels) y QCIF (Quarter Common Intermediate Format,  $176 \times 144$  pixels) de 25 y 30 cuadros por segundo, y en modalidades FR, RR y NR. Las degradaciones introducidas incluían las propias del sistema de codificación y también errores en la transmisión. Se incluyeron secuencias codificadas en diferentes códecs, incluyendo Windows Media 9 (VC-1), H.264, Real Video, H.261, H.263, MPEG4, MPEG2, Cinepak, DivX, Sorenson3 y Theora.

Se evaluaron cinco proponentes, cada uno con diversos modelos FR, RR y NR. Las empresas Psytechnics, Opticom y NTT (Nippon Telegraph and Telephone), así como la Universidad Yonsei de Corea presentaron modelos del tipo FR. Los resultados obtenidos en todos estos modelos, para VGA, CIF y QCIF resultaron estadísticamente mejores al PSNR. VQEG ha sugerido a ITU la estandarización de estos modelos, la que ha sido realizada en agosto de 2008 en la Recomendación ITU-T J.247 [18].

Únicamente la Universidad de Yonsei presentó modelos del tipo RR. Esta propuesta resultó estadísticamente mejor que el PSNR, aún siendo el PSNR una métrica del tipo FR. VQEG ha sugerido a ITU la estandarización de estos modelos, la que ha sido realizada en agosto de 2008 en la Recomendación ITU-T J.246 [19].

Psytechnics y Swissqual propusieron modelos del tipo NR. Los resultados en general no han sido lo suficientemente buenos como para que VQEG sugiriera su estandarización.

**3.4. HD-TV (High Definition TV).**- El proyecto HD-TV comenzó en 2004 y el objetivo fue evaluar modelos del tipo FR, RR y NR, para la predicción de la calidad de video percibida en aplicaciones de televisión digital de alta resolución (HDTV). Las pruebas se limitaron a códecs MPEG-2 y H.264, incluyendo eventuales errores de transmisión. La resolución de pantalla es 1080i @ 50 / 60 campos por segundo y 1080p @ 25 / 30 cuadros por segundo.

El reporte final de VQEG fue aprobado en junio de 2010 [20]. Se presentaron modelos propuestos por las empresas NTT, OPTICOM, SwissQual, Tektronix y la Universidad Yonsei. El modelo del tipo FR que tuvo mejor desempeño fue el propuesto por SwissQual, seguido por el de Tektronix. VQEG ha propuesto estandarizar por lo menos uno de estos modelos, y ITU-T lo realizó en la Recomendación ITU-T J.341 en enero de 2011 [21]. El único proponente que presentó modelos del tipo RR fue Yonsei con resultados aceptables. VQEG ha indicado que estos modelos podrían ser estandarizados. Finalmente no fueron presentados modelos del tipo NR.

**3.5. Hybrid Perceptual / Bitstream.-** El objetivo del proyecto HBS (Hybrid Perceptual Bit-Stream) es la evaluación de modelos FR, RR y NR de estimación de la calidad percibida en servicios de video y multimedia entregados a través de redes IP. Las aplicaciones incluyen IPTV y distribución de contenido multimedia a través de Internet, tanto a terminales fijos como a dispositivos móviles. A diferencia de los otros proyectos, los modelos del proyecto HBS tienen como datos de entrada el flujo de bits de la red IP, tal como llega al decodificador. De esta manera, los modelos pueden hacer uso tanto del video decodificado, como de los parámetros de red. Este tipo de modelos está alineado con las propuestas del grupo de estudio SG12 de ITU-T, donde se han diferenciado dos tipos de modelos, según el tipo de información del flujo de datos utilizada. Estos modelos se han llamado, provisionalmente, P.NAMS (Non-intrusive parametric model for Assessment of performance of Multimedia Streaming) y P.NBAMS (Non-intrusive Bit-stream model for Assessment of performance of Multimedia Streaming). Los modelos P.NAMS utilizan únicamente información del cabezal de los paquetes IP, sin considerar el contenido (son útiles, por ejemplo, cuando el contenido del paquete está cifrado y no es posible decodificarlo). Por el contrario, los modelos P.NBAMS pueden utilizar no solamente información del cabezal, sino también el contenido (el que es accesible cuando el contenido no está cifrado). En el caso de los modelos HBS, adicionalmente, pueden contar con la señal de video decodificada.

Al momento de escribir este artículo, el proyecto se encuentra avanzando, con el plan de trabajo aprobado en abril de 2011 [22].

**4. Conclusiones.-** Conocer la calidad de video percibida por los usuarios es de interés para diversos tipos de aplicaciones, entre las que se pueden mencionar la evaluación de códecs de video, la planificación de redes (anchos de banda, porcentaje admitido de pérdida de paquetes, etc.) o la evaluación del desempeño de redes existentes frente al tráfico impuesto por nuevas aplicaciones multimedia. Los métodos de medida subjetivos (i.e., obtener un promedio de opiniones de personas “reales”) son complejos de realizar e impracticables en aplicaciones de tiempo real. Esto ha llevado al desarrollo de diversos modelos objetivos de estimación de calidad de video. En la última década, el VQEG ha realizado un extenso trabajo en la evaluación objetiva y sistemática de diversos modelos de estimación, contrastando los resultados obtenidos por los modelos respecto a pruebas subjetivas, realizadas en laboratorios independientes. Estos trabajos han permitido la estandarización por parte de ITU de varias recomendaciones, aplicables en diferentes contextos. Existen ya diversas empresas y organismos que ofrecen sistemas de estimación de la calidad de video, tanto para aplicaciones comerciales como de investigación, basadas en las recomendaciones de ITU.

El VQEG continúa su trabajo con nuevos proyectos, entre los que se encuentran la evaluación de calidad para video en 3D (3DTV) y la evaluación de calidad para funciones específicas de reconocimiento o Quality Recognition Tasks (QART). Asimismo, se están iniciando nuevas fases del proyecto de HDTV, tomando las experiencias de la primera fase ya terminada, y esperando poder estandarizar modelos del tipo NR y del proyecto Multimedia, teniendo en cuenta en este caso modelos que combinen audio y video.

Hasta el momento, todos los modelos estandarizados aplican a ámbitos específicos, y tienen diversas restricciones en su aplicación. Se destaca en particular que no hay modelos NR (sin referencia) estandarizados. En el mejor de los casos es necesario enviar sobre un canal auxiliar

información con ciertas características sobre el video original. Por otra parte, si bien los desempeños de los modelos estandarizados son aceptables, aún no se ha llegado al punto de ser estadísticamente equivalentes a las pruebas subjetivas. Aún está abierto el camino para obtener modelos más generales y precisos, que apliquen en diferentes ámbitos y entornos, y que tengan desempeños equivalentes a las pruebas subjetivas.

## 5. Referencias

- [1] Recommendation ITU-R BT.500-11 *Methodology for the subjective assessment of the quality of television pictures*, 06/2002
- [2] Recommendation ITU-T P.910 *Subjective video quality assessment methods for multimedia applications*, 09/1999
- [3] Furth, B. and Marqure, O.: *The handbook of Video Databases: Design and Applications*, September 2003
- [4] Winkler, S. *Digital Video Quality, Vision Models and Metrics*. John Wiley & Sons Ltd, 2005
- [5] VQEG – *Video Quality Experts Group* <http://www.its.bldrdoc.gov/vqeg/>
- [6] Final Report from the Video Quality Experts Group on the Validation of Objective models of Video Quality Assessment, VQEG June, 2000
- [7] Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II, VQEG, August 2003
- [8] Recommendation ITU-R BT.1683 *Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference*, Junio 2004
- [9] Recommendation ITU-T J.144, *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference*. February 2004
- [10] Bourret, A. J.; Hands, D. S.; Bayart, D.; Davies, A.-G.: *Method and System for Video Quality Assessment*. US Patent No. 2006/0152585 A1, July 13, 2006
- [11] Cho, S.; Choe, J.; Jeong, T.; Ahn, W. and Lee, E.: *Objective video quality assessment*. Optical Engineering Vol. 45 (1), January 2006
- [12] Lotufo, A.; Da Silva, R.; Falcao, W.D. F.; Pessoa, A.X.: *Morphological image segmentation applied to video quality assessment*. IEEE Proceedings in Computer Graphics, Image Processing and Vision, SIGGRAPI Proceedings, pp 468-475, October 1998
- [13] Pinson, M.H. and Wolf, S.: *A New Standardized Method for Objectively Measuring Video Quality*. IEEE Transactions on Broadcasting, Volume 50, Issue 3, September 2004, pp. 312-322
- [14] Validation of Reduced-Reference and No-Reference Objective Models for Standard Definition Television, Phase I, VQEG, 2009
- [15] Recommendation ITU-T J.249, *Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference*. January 2010
- [16] Wolf, S. and Pinson, M.H.: *Low Bandwidth Reduced Reference Video Quality Monitoring System*. First Int'l Workshop on Video Proc. and Quality Metrics, Jan 2005
- [17] Final Report of VQEG's Multimedia Phase I Validation Test, 19 September 2008
- [18] Recommendation ITU-T J.247 *Objective perceptual multimedia video quality measurement in the presence of a full reference*. August, 2008
- [19] Recommendation ITU-T J.246, *Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference*. August, 2008

[20] Report on the Validation of Video Quality Models for High Definition Video Content, VQEG Version 2.0, June 30, 2010

[21] Recommendation ITU-T J.341, *Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference*. January 2011

[22] Hybrid Perceptual/Bitstream Group Test Plan, Versión 2.9, April 2011