

# Calidad de Experiencia en servicios multimedia sobre IP

Pablo Pérez<sup>1</sup>, Jaime J. Ruiz<sup>1</sup> y Narciso García<sup>2</sup>

<sup>1</sup>{pablo.perez | jaime.ruiz}@alcatel-lucent.com  
Alcatel-Lucent  
C/ María Tubau, 9  
28050, Madrid

<sup>2</sup>narciso@gti.ssr.upm.es  
Grupo de Tratamiento de Imágenes  
Universidad Politécnica de Madrid  
Ciudad Universitaria s/n  
28040 Madrid

**Resumen** — En este trabajo abordamos un esquema de medida de calidad de experiencia para servicios multimedia sobre IP. Esta arquitectura, denominada QuEM (*Qualitative Experience Measure*), es más adecuada que los esquemas tipo MOS para la monitorización de un gran volumen de tráfico multimedia en tiempo real: facilita la agregación de resultados y la interpretación de las medidas por los operadores. La arquitectura se basa en la detección y caracterización de eventos que degraden la calidad de experiencia durante un tiempo determinado y que puedan ser descritos de forma cualitativa. Cada tipo de evento es monitorizado por un detector específico denominado QuID (*Qualitative Impairment Detector*). En el artículo desarrollamos la arquitectura QuEM y proponemos un conjunto de QuIDs adecuado para la monitorización de servicios como IPTV o videoconferencia.

## I. INTRODUCCIÓN

En la operación de los actuales servicios de comunicación multimedia por IP, bien sean de videoconferencia (bidireccionales), bien de difusión (como IPTV), se plantea la necesidad de medir la Calidad de Experiencia (QoE). Por Calidad de Experiencia entendemos la calidad del proceso de comunicación según es percibido por el usuario. Se trata de *aquello que hay que maximizar en la prestación de un servicio*, pero no existe un criterio definido para medirlo.

Buena parte de los intentos de medida de calidad de experiencia que aparecen en la literatura se basan en el paradigma MOS (*Mean Opinion Score*). El MOS de una secuencia audiovisual no es más que la media de las valoraciones hechas por usuarios reales sobre la calidad de esa secuencia, medida en una escala de 1 (calidad mínima) a 5 (calidad máxima). Siguiendo este paradigma, el objetivo de una métrica de calidad debe ser tener una correlación lo mejor posible con el MOS obtenido de una prueba de calidad subjetiva con usuarios reales.

En esta línea van, por ejemplo, todos los trabajos del Video Quality Experts Group (e.g. [1]) que han dado lugar a estándares como el J.144 [2] para la medida de calidad de un contenido de vídeo degradado, teniendo acceso también al original (es lo que se denomina métrica con referencia completa, o *full reference –FR–*). Otra métrica muy usada en investigación es SSIM [3], por su combinación de sencillez y efectividad. Del mismo modo es posible encontrar métricas que estiman la calidad del audio [4], así como la calidad resultante de combinar audio y vídeo de distintas calidades [5].

Las métricas basadas en MOS ofrecen datos muy detallados para instantes puntuales del tiempo. Este tipo de análisis *intensivo* es adecuado para calibrar equipos de procesamiento de vídeo, por ejemplo. Pero para monitorizar la calidad de experiencia en una red compleja a lo largo del tiempo, resulta difícil procesar esos datos y obtener de ellos algo significativo; máxime cuando, en muchas pruebas de calidad subjetiva hechas con usuarios reales, la dispersión de los resultados tiende a ser bastante grande.

En la mayor parte de los escenarios de monitorización de calidad, por el contrario, se recurre a métricas de Calidad de Servicio (QoS), en particular de aquellos parámetros de la QoS que afectan al tráfico multimedia. Considerando un escenario habitual de transmisión usando RTP sobre IP, la métrica más extendida es el MDI (*Media Delivery Index*) [6], que proporciona de una forma compacta medidas de tasa de pérdidas y variación de la latencia (*jitter*), lo que se traduce en pérdidas provocadas por desbordamiento o vaciado de los *búferes* intermedios. Dado que las pérdidas son la principal causa de degradación de la QoE en servicios multimedia, se emplean diversas técnicas de protección contra errores, como pueden ser FEC o retransmisiones [7]. También existen métricas que permiten evaluar el resultado de estas medidas [8] y calcular así la tasa de pérdidas efectiva que afecta al usuario final.

Estas métricas, no obstante, consideran únicamente las causas de la degradación de la calidad multimedia (pérdidas y *jutter*), pero no sus efectos (artefactos en el vídeo). Si bien está demostrado que, para pérdidas uniformes y a largo plazo, la tasa de pérdidas es una buena medida del error agregado en el tiempo [9], el efecto de cada pérdida individual puede ser muy diferente en función de en qué parte del vídeo se haya producido el error [10].

En este contexto, resulta interesante diseñar una estrategia de medida de elementos de la calidad de experiencia que sea adecuada para la monitorización del tráfico multimedia IP en una red de comunicaciones real. El objetivo de este trabajo es precisamente plantear una solución a este problema: cómo evolucionar las medidas de calidad de servicio de red (tipo MOS) para, sin perder sus beneficios en entornos reales, introducir capacidades de medida de QoE.

El trabajo ha sido desarrollado en el marco del proyecto VISION, perteneciente al programa CENIT 2007 del Centro para el Desarrollo Tecnológico Industrial.

## II. DISEÑO DEL SISTEMA

### A. Requisitos

En este trabajo nos planteamos el diseño de un sistema para la monitorización de calidad de experiencia de un servicio multimedia. Debe cumplir, por tanto, los siguientes requisitos:

- Orientarse a un sistema de monitorización masiva de calidad de experiencia en una red de servicios multimedia. Tiene que ser posible obtener datos a lo largo del tiempo y agrupando un gran número de usuarios.
- Emplear únicamente medidas con referencia reducida o, preferiblemente, sin referencia. En un contexto de monitorización real es posible que no haya posibilidad de comunicar referencias, siquiera reducidas, a los equipos de medida.
- Ser significativo: debe trabajar con una escala de degradación que sea fácil de interpretar (y de reproducir) para el operador.

### B. Arquitectura

La mayor parte de los sistemas de monitorización basados en calidad de servicio de red, como MDI, cumplen los requisitos anteriores. Por tanto, en vez de intentar extender el paradigma MOS a un sistema de monitorización, nuestro punto de partida será el contrario: ampliar un esquema de medida de calidad de servicio para incluir conceptos de calidad de experiencia. Para ello partiremos de las siguientes hipótesis:

- Todo sistema de distribución multimedia extremo a extremo está diseñado de modo que, en condiciones normales de operación, ofrece una calidad *suficiente*. Con otras palabras: usaremos como calidad de referencia la ofrecida por el servicio multimedia en condiciones normales de funcionamiento. Esto incluye también la calidad de codificación de un contenido típico.
- Hay *eventos* puntuales que deterioran la calidad de experiencia. El más habitual es la pérdida de paquetes, que genera efecto de bloques en el vídeo y discontinuidades en el audio, entre otros efectos. Otros eventos podrían ser la caída puntual en la calidad de la codificación, la pérdida de sincronismo entre audio y vídeo, etc.

Por tanto, el objetivo de nuestro sistema será determinar cuándo suceden estos eventos y cuál es su grado de severidad. Para ello emplearemos una arquitectura como la que se muestra en la Figura 1, a la que denominaremos QuEM (*Quantitative Experience Measure*).

El bloque básico del sistema es el QuID (*Qualitative Impairment Detector*). Se trata de un elemento que monitoriza el flujo multimedia para detectar un único tipo de evento de degradación. Aunque en algunos casos el evento pueda detectarse por sus causas (p. ej. “pérdida de un paquete de vídeo”), la salida del QuID siempre caracteriza los efectos en la calidad de experiencia (p. ej. “efecto de macrobloques en la imagen durante dos segundos”). De esta manera, cada evento detectado tiene *valor semántico*: puede ser descrito por una persona.

El siguiente bloque se denomina STF (*Severity Transfer Function*), que asigna a cada evento una *severidad* basándose en un modelo de la degradación. Esta función se muestrea en el tiempo, obteniéndose un único valor cada cierto tiempo  $T_S$ .

La STF de cada evento puede ser configurada por el operador. Dado que los eventos tienen *significado*, es posible para cada operador definir sus propias severidades en función de los criterios que consideren oportunos. Así, por ejemplo, para un “efecto de macrobloques en la imagen durante dos segundos” un operador puede asignar una severidad alta y otro una severidad media (y, por ejemplo, poner el umbral de “severidad alta” en un artefacto de ese tipo de más de tres segundos de duración).

El valor de  $T_S$  también es configurable, aunque típicamente se usarán valores entre 10 y 60 segundos. Es aconsejable que el reloj de  $T_S$  esté sincronizado para todos los puntos de medida de la red. Esto se puede conseguir usando como reloj el del sistema y empleando algún protocolo tipo NTP para sincronizar todos los relojes.

Finalmente se define el valor de severidad global como el máximo de los valores de las severidades de todas las STF [11].

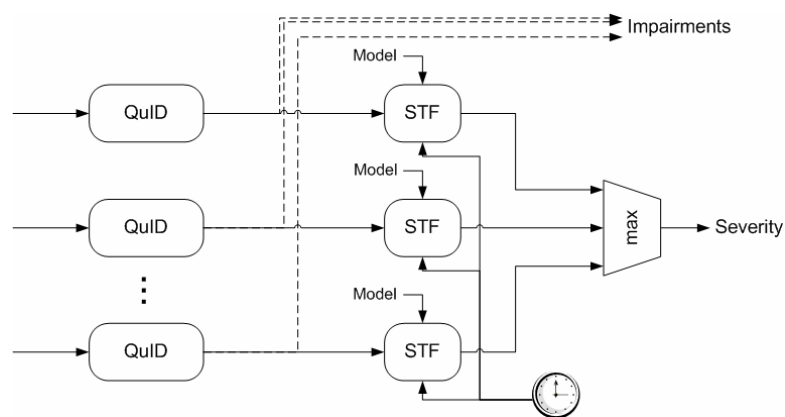


Figura 1. Arquitectura QuEM.

### C. Entorno de trabajo

Si bien la arquitectura QuEM es genérica, los QuIDs que se empleen dependerán del tipo de servicio o contenido multimedia que se esté monitorizando. Consideraremos la transmisión unidireccional de vídeo codificado en H.264 y audio en un formato tipo MPEG (MPEG-1, AAC o AC-3), todo ello multiplexado en una Trama de Transporte MPEG-2 y encapsulado en RTP/UDP/IP. Este escenario es típico de entornos de IPTV [7] y también es el utilizado en la arquitectura de referencia del proyecto VISION.

Como funcionalidad adicional, consideraremos que la trama de transporte está encapsulada en RTP de tal modo que el vídeo y el audio se transportan siempre en paquetes RTP diferente, y que cada paquete RTP lleva una extensión de la cabecera que indica el tipo de contenido (vídeo o audio) y el tipo de cuadro (en caso de que sea vídeo) [12]. Este tipo de encapsulado, incluido tanto en los productos de IPTV de Alcatel-Lucent como en el proyecto VISION, está específicamente diseñado para facilitar el trabajo de los sistemas que procesan la trama en la red y, si bien no es estrictamente necesario para el funcionamiento correcto de los QuIDs que proponemos, sí simplificará su diseño.

### D. Agregación local de resultados

Una de las principales necesidades de los operadores a la hora de trabajar con sistemas de monitorización de calidad es disponer de un método de agregación de resultados, de modo que el volumen de datos con el que se trabaja sea manejable. La arquitectura QuEM, al proporcionar medidas discretas de severidad en intervalos discretos de tiempo, facilita después el procesado de la información.

El esquema empleado para agregar los resultados de la severidad QuEM es el mostrado en la Figura 2. En primer lugar los resultados de severidad son diezmados en un ratio 1:N, donde el valor de N es configurable (y puede ser N=1 si se desea evitar el diezmado). Este proceso produce un único valor de salida por cada N valores de entrada, siendo el valor de salida igual al valor de mayor severidad de las N entradas.

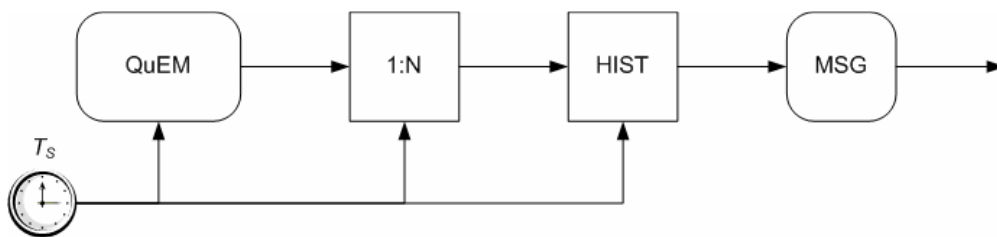


Figura 2. Agregación temporal.

Posteriormente se genera un histograma (HIST) con el número de muestras de cada nivel de severidad. Cada  $M=k*N$  muestras (es decir, cada  $M*T_s$  segundos) los valores del histograma se envían a un módulo de mensajería (MSG) que envía un mensaje a un sistema colector de estadísticas.

Este mismo esquema, que se aplica a la salida del módulo QuEM, puede aplicarse individualmente a cada una de las salidas de los STFs por separado.

### D. Agregación global de resultados

Una vez que los resultados se han generado en los puntos de medida, se envían a una colectora de estadísticas. La colectora agrega los resultados de distintos usuarios y de distintos flujos RTP si procede (por ejemplo, para el caso de canales de vídeo multicast, que se reciben de forma simultánea por muchos usuarios). De esta manera se pueden obtener resultados globales (agregados) de toda la distribución de una forma sencilla.

## III. MÉTRICAS DE DETECCIÓN DE EVENTOS

El siguiente paso en el diseño del sistema es determinar qué eventos de degradación se van a buscar y qué elementos de medida QuID se emplearán para detectarlos. La arquitectura QuEM únicamente permite emplear algoritmos que no requieran de información sobre la imagen sin degradar (es decir, métricas NR, *No Reference*). Propondremos dos alternativas, una empleando métricas basadas en medidas en la capa de transporte (en el flujo codificado, *bitstream-QuID*) y otra empleando medidas en la capa de medios (en el flujo decodificado, *baseband-QuID*).

Los eventos que detectaremos serán los siguientes:

- Audio entrecortado.
- Silencios en el audio (valor de audio por debajo de un umbral).
- Congelación en el vídeo o pantalla en negro.
- Efecto de *macrobloque* (típico efecto producido por la pérdida de paquetes de vídeo; a veces se denomina efecto de *pixelaciones*).

- Degradación en la calidad del vídeo codificado (que en H.264 se traduce típicamente en emborronamiento de la imagen).

Estos eventos constituyen la mayor parte de las causas de degradación que se observan en despliegues comerciales de servicios multimedia (como videoconferencia o IPTV). No consideraremos la degradación de la calidad del audio porque, dado que el audio requiere mucha menor tasa binaria que el vídeo para una misma calidad subjetiva, asumiremos que el audio siempre está codificado con calidad suficiente. Esto es así en la totalidad de los despliegues de servicios multimedia comerciales conocidos por los autores.

La Tabla I muestra el tipo de métricas que se emplearán en cada una de las dos estrategias. Describiremos primero las basadas en la capa de medios (*baseband-QuEM*), que se centran específicamente en los eventos de la degradación (y, por tanto, son directamente aplicables al esquema QuEM) y, posteriormente, las de capa de transporte (*bitstream-QuEM*), que tratan de estimar los efectos a partir de sus causas.

#### A. Baseband-QuEM

Los problemas en la transmisión que afectan al audio suelen producir bien cortes puntuales (que dan lugar a audio entrecortado o a discontinuidades o glitches) bien silencios prolongados (períodos donde el audio está por debajo de un cierto umbral). Ambos casos se pueden detectar midiendo la potencia de la señal de audio, bien con medidas de potencia instantánea (tipo picómetro), bien con medidas de potencia media (tipo vúmetro). Los silencios bruscos (caída brusca del nivel de potencia y recuperación posterior) son perceptibles para duraciones a partir de 30 milisegundos y comienzan a afectar significativamente a la calidad percibida a partir de los 200 ms [13]. Los silencios prolongados suelen ser significativos a partir de los dos segundos de duración [14]. En ambos casos, es posible fijar un modelo de severidad a partir únicamente de dos parámetros: la duración del evento y el número de eventos producidos durante el intervalo de muestreo  $T_S$ .

Tabla I. Métricas empleadas en los QuID

Evento	Baseband-QuEM	Bitstream-QuEM
Audio entrecortado	Discontinuidades en el audio	Pérdida de paquetes de audio
Silencios en el audio	Medida de nivel de audio	Estimación del nivel de audio
Congelación de vídeo	Detección de imagen congelada	Análisis de pérdidas y de movimiento
Efecto macrobloque	Detección de bloques (GBIM)	Visibilidad de la pérdida de paquetes
Degradación de codificación	Detección de emborronamiento	Análisis del parámetro QP

La imagen congelada supone una caída brusca en la información temporal de la imagen (el movimiento se detiene bruscamente). La imagen resultante puede ser una pantalla negra, pero no es necesario. Para detectarla, se monitoriza la información temporal de la imagen (energía de la diferencia entre una imagen y la anterior). Una caída brusca de la información temporal seguida de un intervalo largo de información temporal nula es un síntoma claro de imagen congelada. La duración del evento de congelación indicará su severidad.

Para la detección de la pérdida de paquetes, emplearemos un algoritmo de medida de efecto de bloques. Este tipo de algoritmos se diseñaron principalmente para la detección del efecto de bloques producido por la compresión de imágenes JPEG y vídeo tipo MPEG-2 y se centran en detectar estructuras de bloques (típicamente de 8x8 píxeles) en la imagen. Al generalizarse el uso de H.264, este efecto pasó a ser menos relevante en el vídeo codificado, ya que los bloques son ahora menores (desde 4x4 píxeles), la predicción espacial y temporal es más compleja y existe un filtrado específico en la imagen para eliminar este efecto. De hecho, este tipo de algoritmos no ofrece buenos resultados para determinar la calidad del vídeo codificado [15]. No obstante, en el caso de la pérdida de paquetes de vídeo, se pierde la información de macrobloques enteros: estructuras de 16x16 píxeles alineadas con los bordes de la imagen. En este caso los algoritmos de detección de bloques (buscando estructuras de 16x16, en vez de 8x8) sí son adecuados. En nuestro diseño proponemos el uso del algoritmo GBIM [16], que ofrece un buen compromiso entre efectividad y eficiencia.

El uso de las herramientas de protección frente a efecto de bloques hace que en H.264 la principal degradación introducida por la codificación de la imagen sea el emborronamiento. Existen diversas métricas de emborronamiento en la literatura. Nosotros hemos optado por incluir la de Marziliano [17], que se basa en la anchura de los bordes en el eje horizontal.

#### B. Bitstream-QuEM

El objetivo de la arquitectura *bitstream-QuEM* es conseguir hacer las mismas medidas que en *baseband-QuEM*, pero midiendo únicamente en la capa de transporte sin decodificar el vídeo ni el audio. Se trata, pues, de predecir los efectos a partir de una medida, siempre limitada, de sus causas. Se asume una cierta pérdida de precisión con respecto a las medidas en capa de medios a cambio de una reducción drástica de la carga computacional. Por ello nos restringiremos a medidas que puedan obtenerse en la capa NAL de H.264, llegando únicamente hasta la cabecera de slice (*slice\_header*) y sin decodificar, en ningún caso, la codificación entrópica CAVLC o CABAC.

Para detectar audio entrecortado, analizamos las pérdidas de audio. Los estándares de audio genérico bajo estudio no incluyen ningún tipo de predicción temporal entre distintas Unidades de Acceso, de modo que la duración del artefacto de

audio será similar al tamaño de la ráfaga de pérdida. Esto se ha comprobado experimentalmente a partir de una trama de audio MPEG-1 layer 2, a la que se han introducido pérdidas de longitud controlada. En la figura 3(a) se muestran los resultados: la duración del silencio ( $L$ ) es igual a la duración de la ráfaga ( $D$ ) más un cierto offset fijo ( $\alpha$ ), con un coeficiente de determinación  $R^2 = 0.98$ .

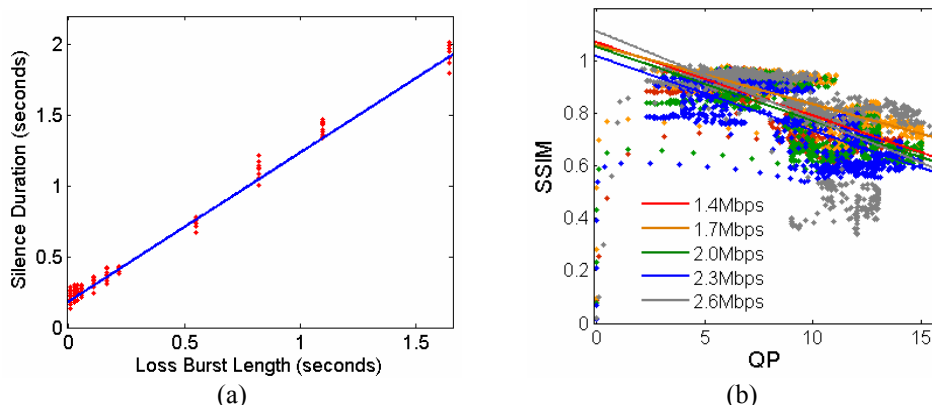


Figura 3. (a) Duración del silencio de audio en función de la longitud de la pérdida. (b) SSIM en función de QP estimado

Los silencios de larga duración pueden producirse por una ráfaga larga de pérdidas de paquetes de audio (en cuyo caso se detectarán del modo indicado anteriormente) o por un silencio anterior al proceso de codificación. En este caso se trata de medir la intensidad media del audio a partir de su codificación. Esto se puede calcular de forma directa a partir de los coeficientes de subbandas del audio, bien a partir las herramientas de control de rango dinámico del audio incluidas en AC-3, DAB (para MPEG-1) y AAC, que son una medida indirecta de la energía media de la trama.

El equivalente a los silencios de larga duración en el audio son las imágenes congeladas en el vídeo. Las congelaciones se pueden producir antes del bloque codificador o bien por problemas de transmisión (pérdidas). En el primer caso es posible detectar imágenes congeladas porque el tamaño de todos los cuadros B y P se reduce prácticamente a cero, ya que es posible predecir la imagen a partir de una simple copia de la imagen anterior. En el caso de las pérdidas de transmisión, se suelen producir congelaciones cuando se pierde la cabecera de un cuadro I, y el decodificador se queda un GOP completo sin poder decodificar nada. Esta pérdida es posible detectarla fácilmente a partir de las cabeceras extendidas RTP descritas en II.C.

Para detectar el efecto de las pérdidas de paquetes, se pueden usar métodos como [9][10]. Estos métodos están diseñados para trabajar fundamentalmente como métricas de referencia reducida, ya que suelen requerir de información sobre el proceso de codificación del vídeo, o bien decodificarlo parcial o totalmente. No obstante, son un buen punto de partida para predecir el efecto de la pérdida de paquetes. A partir de ello, es posible diseñar medidas *bitstream-NR* adaptadas a H.264 [18].

Finalmente es necesario considerar también la degradación debida a los defectos en la codificación en el vídeo. En este aspecto es preciso indicar que, en la mayor parte de los casos, esta degradación será menos severa que las mencionadas anteriormente (siempre es menos molesto un vídeo con menos calidad que la pérdida total de información en buena parte de la imagen o la pérdida de audio). Elegiremos como indicador de la calidad de codificación el valor del parámetro de cuantificación QP incluido en la cabecera del *slice* H.264, que es, en las implementaciones comerciales típicas, una aproximación al valor real de QP empleado en los macrobloques de la imagen. En la figura 3(b) se muestran las variaciones de SSIM [3] frente al valor de QP para una misma secuencia de vídeo codificada a cinco regímenes binarios diferentes. Cuantitativamente se observa cierta correlación entre ambos parámetros ( $R^2 = 0.5$ ). Cualitativamente podemos observar que, a partir de determinados valores de QP (10 en el ejemplo, aproximadamente) la calidad de la secuencia se ve comprometida. Por tanto, es posible asignar a determinados umbrales de QP eventos de degradación de diversas severidades.

#### IV. DESPLIEGUE

Los sistemas de medida basados en arquitecturas QuEM pueden desplegarse de forma extensiva en la red de comunicaciones multimedia. En particular, un sistema basado en *bitstream-QuEM* podría estar integrado incluso en los terminales de usuario, de modo que se recopilaren estadísticas de usuario final a lo largo de la red, sin impactar perceptiblemente al rendimiento del terminal.

La principal ventaja de un sistema *bitstream-QuEM* es que los eventos que se caracterizan son reproducibles. Empleando el mismo análisis que se utiliza para predecir las degradaciones es posible simularlas: con pérdidas controladas de paquetes de vídeo y/o audio, controlando la salida de vídeo y audio (para forzar congelaciones o silencios) o recodificando el vídeo con valores de QP fijos y elevados. De esta manera se proporciona al operador un método para fijar las severidades *a priori*, en un entorno de laboratorio, con una reproducción bastante fidedigna de la degradación en la calidad de experiencia producida por cada uno de los efectos bajo estudio.

El despliegue de un sistema QuEM se producirá en tres pasos. En primer lugar es preciso tener un adecuado diseño del servicio multimedia y determinar cuál es el funcionamiento esperado del mismo (la *calidad de referencia*). En segundo lugar se

fijarán las severidades de los distintos eventos a partir de las simulaciones mencionadas en el párrafo anterior. Finalmente se puede desplegar el sistema y comenzar a recopilar los datos de calidad de experiencia de los usuarios.

## V. CONCLUSIONES

En este trabajo hemos descrito QuEM, una arquitectura para la monitorización de calidad de experiencia orientada a proporcionar datos útiles y manejables al operador de red. En este sentido hemos partido de los sistemas de monitorización de calidad de servicio de red, cuya métrica más relevante respecto a la calidad multimedia es el MDI, y los hemos mejorado para introducir capacidades específicas que modelen los efectos en la calidad final introducidos por las degradaciones en la red.

El modelo QuEM se basa en métricas cuyos resultados se pueden *describir* y, por tanto, son significativos para las personas. Por ello son también repetibles, siendo posible simular eventos que produzcan efectos conocidos *a priori*. Además se emplea un modelo claro de integración de distintas medidas en paralelo, de medidas a lo largo del tiempo y de los resultados de diferentes usuarios. En todo ello este modelo supera al paradigma MOS como arquitectura para la monitorización de calidad en despliegues comerciales con miles de usuarios.

Finalmente, el modelo se basa en tecnologías existentes. Al ser un modelo modular basado en subsistemas independientes (los QuID), es posible mejorar cada una de las métricas empleadas por separado sin afectar al funcionamiento del resto del sistema. Además puede aplicarse tanto con medidas en capa de red como en capa de medios. Por todo ello es una aproximación muy adecuada para la monitorización de calidad de experiencia en servicios de comunicaciones multimedia.

## AGRADECIMIENTOS

Agradecemos a Silvia Varela, Jesús Macías y Alfredo Pérez sus valiosas contribuciones al trabajo presentado en este artículo. Este trabajo ha sido parcialmente financiado por el CDTI, en el marco del proyecto CENIT-VISION 2007-1007.

## REFERENCIAS

- [1] Video Quality Experts Group, “Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II”, August 2003.
- [2] ITU-T J.144 “Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference”, February 2004.
- [3] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity”, IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004.
- [4] ITU-R BS.1387-1 “Method for objective measurements of perceived audio quality”, January 2001.
- [5] D. S. Hands, “A Basic Multimedia Quality Model”, IEEE Transactions on Multimedia, vol. 6, No. 6, December 2004
- [6] J. Welch and J. Clark, “A proposed Media Delivery Index”, IETF RFC 4445, April 2006.
- [7] ETSI TS 102 034, “Digital Video Broadcasting (DVB); Transport of MPEG-2 Based DVB Services over IP Based Networks”.
- [8] T. Friedman, R. Caceres, A. Clark, “RTP Control Protocol Extended Reports (RTCP XR)”, November 2003.
- [9] A. R. Reibman, V. A. Vaishampayan and Y. Sermadevi, “Quality monitoring of video over a packet network”, IEEE Transactions on Multimedia, Vol. 6, No. 2, April 2004.
- [10] A. R. Reibman and D. Poole, “Characterizing packet-loss impairments in compressed video”, IEEE International Conference on Image Processing, Oct. 2007
- [11] S. Jumisko-Pyykkö, V. Kumar, and J. Korhonen, “Unacceptability of instantaneous errors in mobile television: from annoying audio to video”, in MobileHCI, NY, USA, 2006, pp. 1–8, ACM Press
- [12] D. C. Robinson and A. Villegas, “Intelligent wrapping of video content to lighten downstream processing of video streams”, European Patent EP2071850, June 2009.
- [13] R. Pastrana, J. Gicquel, C. Colomes and C. Hocine, “Sporadic Signal Loss Impact on Auditory Quality Perception”, <http://wireless.feld.cvut.cz/mesaqin2004/contributions.html>, 2004
- [14] A. R. Reibman and A. R. Wilkins, “Video Outage Detection: Algorithm and Evaluation”, Proc. 27<sup>th</sup> Picture Coding Symposium, Chicago 2009
- [15] A. Leontaris and A. R. Reibman, “Comparison of Blocking and Blurring Metrics for Video Compression”, Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp- 585-588, March 2005
- [16] H. R. Wu and M. Yuen, “A Generalized Block-Edge Impairment Metric for Video Coding”, IEEE Signal Processing Letters, Vol. 4, No. 11, Nov 1997.
- [17] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, “A no-reference perceptual blur metric”, in Proceedings of the International Conference on Image Processing, vol. 3, pp. 57–60, Rochester, NY, September 22--25, 2002
- [18] P. Pérez, J. Macías, J. J. Ruiz and N. García “Effect of packet loss in Video Quality of Experience”, Bell-Labs Technical Journal, submitted.