

Factores que Influyen en la Evaluación de QoE del Servicio de Video Streaming

Important Factors and Aspects about QoE Evaluation Relating to Video Streaming Service

Yeimar Andrés Parrado Leal^{1*} y Luis Guillermo Martínez Ballesteros^{2*}

¹Universidad Sergio Arboleda, Bogotá, Colombia *yeimar.parrado@correo.usa.edu.co

²Magister en Ingeniería de Telecomunicaciones, Universidad Nacional de Colombia. *lgmartinezb@gmail.com

Fecha de recepción del artículo: 28/08/2012 Fecha de aceptación del artículo: 28/12/2012

Resumen

El presente artículo muestra un análisis de cómo el dispositivo final y las interrupciones afectan la evaluación de la Calidad De La Experiencia (QoE) cuando un usuario hace uso del servicio de video streaming. Se logró realizando encuestas a 60 estudiantes los cuales observaron un video individualmente durante 1.53 minutos en uno de los dos dispositivos iPhone4 o iPad1 dos veces. La primera dejando precargar el buffer de datos 15 segundos y la segunda dejándolo precargar totalmente. Se observó que el usuario califica los dispositivos con base a la expectativa y a las interrupciones. Dichos factores tienen diferentes efectos sobre la evaluación de la QoE dependiendo del dispositivo. Para continuar con este trabajo finalmente se recomienda revisar variables como el tamaño de la pantalla y la resolución para así determinar con el usuario percibe el servicio dependiendo del cambio de dichas variables.

Palabras clave

QoC, QoE, QoS, Video Streaming, MD5.

Abstract

This paper presents an analysis of how the final device and interruptions affect the assessment of

Quality of Experience (QoE) when a user makes use of streaming video service which was achieved by surveying 60 students who watched a video individually for 1.53 minutes at one of the two devices iPad1 iPhone4 twice, leaving the first one preload the data buffer 15 seconds and the second leaving him totally preload, and based on the results was observed that the user qualifies the devices based on the expectation, the interruptions have different effects depending of QoE evaluation device. To continue this work finally recommends checking variables such as screen size and resolution in order to determine how the user perceives the service depending on the change of these variables.

Keywords

QoC, QoE, QoS, Video Streaming, MD5.

Introducción

Desde la aparición de los servicios sobre infraestructura de redes IP, los parámetros de Calidad De Servicio (QoS) como ancho de banda, retardos, pérdida de paquetes entre otros, eran los únicos tenidos en cuenta al momento de evaluar la calidad del servicio. En ese momento fueron suficientes elementos, puesto que los servicios no integraban audio, video y texto, sino que la mayoría de estos eran servicios

de un solo medio (por ejemplo la aplicación basada en File Transfer Protocol (FTP)) como el correo electrónico. Esto se debió a la limitación de ancho de banda con que se contaba [1, 2]. Sin embargo, desde hace un tiempo el incremento de servicios que integran medios multimediales ha forzado a buscar otras métricas como QoE y Calidad De Contenido (QoC), ya que los parámetros de QoS no garantizan que con su evaluación se logre saber la satisfacción de los usuarios que hacen uso de estos servicios [3]. Por esta razón, los operadores y las empresas que brindan servicios multimedia se han enfocado no solo en los parámetros de QoS, sino también en los parámetros de QoE y QoC, y así cada vez brindar la mejor experiencia al usuario final. Con base en lo anterior, se ha notado un creciente interés por crear diferentes técnicas de medición de la QoE (*Unión Internacional de Telecomunicaciones*, UIT, *European Telecommunications Standards Institute* (ETSI), Internet Protocol Television IPTV Group).

Se han hecho pruebas de cómo medir QoE en IPTV en busca de parámetros relevantes para hacer una correcta medición de QoE [4], y se ha demostrado mediante pruebas y comparaciones las ventajas de hacer uso de métodos híbridos para evaluar QoE [5]. También se realizó una encuesta de QoE donde hacen una relación entre la QoE y la intensidad de la señal [6]. Tominaga et al. recomendaron el uso de Absolute Category Rating (ACR5) para evaluar QoE en IPTV [7]. Basados en estas evaluaciones se observó que son pocas las medidas subjetivas de QoE que se realizan por el costo y la dificultad logística que se requieren al hacerlas [8–10]. También cuando las realizan, utilizan en las pruebas un número reducido de personas que puede resultar en una poca confiabilidad de los resultados. Cabe destacar el hecho de que en la mayoría de los casos no tienen en cuenta las características del dispositivo final. Con base en lo anterior, se realizó una evaluación subjetiva de la QoE en lo que se refiere a la reproducción de una secuencia de video streaming usando la aplicación Vimeo [11]. En una red wireless se realizaron pruebas con dos dispositivos diferentes (iPhone4, iPad1), basados en el método ACR [12] utilizando la escala Mean Opinion Score (MOS) [13]. Se realizó una en-

cuesta donde se analizó cómo la calidad de video influye en una buena o mala medida de QoE, así como también el tamaño de la pantalla del dispositivo y la velocidad de carga del buffer de datos durante la reproducción del video.

Existen coincidencias en las diferentes definiciones de la calidad de la experiencia (QoE está basada principalmente en la idea de que se trata de la aceptación de un servicio mientras es utilizado desde la perspectiva subjetiva de los usuarios) [14-16]. Las formas de medición de la QoE son objetivas, subjetivas o híbridas. En la medición objetiva se realiza una relación directa ente métricas de QoS y QoE pero en algunos casos, se encontró que a pesar de garantizar la QoS los resultados de QoE no fueron satisfactorios [3]. En la subjetiva se realiza una evaluación directa al usuario mediante encuestas. A pesar del costo y del tiempo que se requiere, da un resultado más fiable para evaluar QoE [16]. En la híbrida, se realiza una mezcla de la evaluación directa del usuario y se consideran también métricas de QoS debido a costo y tiempo [7, 10].

La medición de QoE de videos requiere el uso de métodos de escala subjetiva. La condición de tales medidas para que tengan sentido es que existe una relación entre características físicas del “estímulo”, para este caso el estímulo es causado por la secuencia de video presentado al usuario en la prueba, y la magnitud y naturaleza de la sensación causada [12]. La calidad de experiencia es evaluada normalmente haciendo tests controlados en los que se reproducen muestras de video a usuarios, a quienes se les pide que las califiquen mediante una escala determinada previamente. Las calificaciones establecidas por cada usuario son promediadas para así obtener la puntuación de opinión media (MOS), estandarizado en la recomendación UIT-T P.800.1 [13]. El hecho de que la calidad de experiencia sea una opinión dinámica hace que la evaluación sea diferente durante todo el tiempo que se hace uso del servicio por ejemplo, si un usuario hace uso de un servicio que durante la gran mayoría del tiempo ha sido de alta calidad pero baja durante la última parte de uso del servicio, la QoE global por parte del usuario

será muy baja. Por este motivo hay experiencias en las que se mide durante toda la prueba para realizar un promedio ejemplo de este caso [17].

Existen diferentes escalas de medición, pero la ACR brinda los mejores resultados para evaluar QoE. El método de categoría de calificación absoluta es un juicio de calificación en el que las secuencias de video de prueba se presentan una a la vez y se valoran de forma independiente en una escala de categorías (Este método también es llamado como solo de estímulo) [12]. El método especifica que después de cada presentación, a los usuarios se les pide que evalúen la calidad de la secuencia presentada. Según la recomendación de la ITU-T P.910 [12] los siguientes cinco niveles se usan en la escala de calificación de calidad: Excelente (5), Bueno (4), Aceptable (3), Regular (2), Malo (1).

Considerando lo anterior, el desarrollo de este artículo se orienta a crear un escenario de pruebas en el que se evaluará, de una manera subjetiva la QoE mediante una encuesta de la calidad de reproducción en una aplicación de video streaming como Vimeo usando dos dispositivos diferentes para realizar las pruebas, conectados a una red inalámbrica en un entorno en el cual muchos usuarios emplean la misma red wireless durante el día (red inalámbrica de la Universidad Sergio Arboleda). Con ello se obtuvo un mayor conocimiento de cómo repercute en la medida de la calidad percibida por un usuario, el tamaño del dispositivo que usa, la calidad de reproducción del video que observa, la cantidad de interrupciones que se puedan presentar durante la reproducción del video por la velocidad de carga del buffer de datos, entre otros aspectos.

Metodología

Descripción del escenario de pruebas

Se realizó una encuesta a una muestra finita de una población de aproximadamente 4000 estudiantes que pertenecen a la sede en Bogotá de la Universidad Sergio Arboleda, con el fin de obtener resultados confiables sobre la evaluación de QoE [19].

Esta fue una muestra dio un 60% de confiabilidad (60 estudiantes) entre 17 y 25 años. La aplicación elegida para la reproducción del video fue Vimeo (aplicación que garantiza una alta calidad de reproducción de video streaming tanto en tablets como también en smartphones) y se reprodujo un video para que el usuario encuestado se centrara únicamente en la calidad de reproducción del mismo. Las características fueron: URL: <http://vimeo.com/20523652/>, pixelado: 1280x720 pixeles, tamaño: 101.55 megabytes (MB), formato: .mov y duración 1.53min.

Dispositivos usados

Se eligieron dos tipos de dispositivos diferentes. Por un lado un Smartphone (iPhone4) y TabletPC (iPad1). Se escogieron dispositivos de Apple puesto que estos tienen como característica principal la calidad en aplicaciones multimedia y serán un buen dispositivo para hacer pruebas de QoE enfocadas hacia la reproducción de video. A continuación se muestra las especificaciones de cada dispositivo.

Pruebas

Las pruebas se realizaron usando la red inalámbrica en las instalaciones de la Universidad Sergio Arboleda que cuenta con una red de 50 antenas wireless administradas por un dispositivo de control que está conectado al servicio de internet de 50MB. Se realizaron tres preguntas en el test ya que solo era necesario pedir una calificación de la calidad de reproducción del video (escala MOS) y unas apreciaciones sobre los factores que influían en la calificación (edad y uso habitual de smartphones o tablets). El procedimiento de evaluación se realizó teniendo en cuenta el entorno de pruebas, el test, el tipo de encuestados y el diseño de la encuesta [12].

Después de pedirle al estudiante que accediera a realizar la prueba de QoE se le pedía ubicarse en un sitio específico, elegido previamente en el entorno de la prueba. Se procedió a asignarle uno de los dos tipos de dispositivos y se le mostró la reproduc-

Tabla 1. Descripción de los dispositivos.

Características	iPhone4	iPad1
Medidas	115,2 mm de alto, 58,6 mm de ancho, 9,3 mm de espesor, 137 gramos de peso.	24,12 cm de alto, 18,57 cm de ancho, fondo: 0,88 cm, 601 gramos de peso.
Pantalla	Retina display, display multi-touch. Widescreen de 3,5 pulgadas (en diagonal).	Panorámica brillante multi-touch de 9,7 pulgadas (en diagonal) retro-luminada por led.
Resolución	De 960+640 a 326 píxeles por pulgada (PPP).	De 1.024*768 a 132 píxeles por pulgada (PPP).
Conectividad	802.11b/g/n Wi-fi (802.11n 24 GHz). Bluetooth 2.1 wireless technology.	Wi-fi (802, 11a, b, g, n) Tecnología bluetooth 2.1.
Procesador	Chip A4 de Apple un sólo núcleo a 1GHz.	Chip a5 de Apple de doble núcleo a 1GHz.

ción del video en los siguientes dos escenarios: sin/ con precarga de buffer. Durante la reproducción se contabilizó el número de interrupciones que se presentaron en la operación. Posteriormente se le pidió responder el test de QoE. Del total de 60 personas que se encuestaron, se realizó una división en dos grupos (30 personas con el iPhone 4, y las otras 30 personas con el iPad1).

Resultados y Análisis

En la Tabla 2 se observa que el promedio del número de interrupciones que ocurrieron en las pruebas en los dos dispositivos fueron en promedio mayores con el iPhone4 a las que se tuvieron con el iPad1 en los dos casos (carga o no precarga del buffer). Existen diferencias en las capacidades de procesamiento de un dispositivo o el otro como por ejemplo que el iPad1 maneja un chip a5 de doble núcleo mientras que el del iPhone4 maneja un chip a4 de un solo núcleo.

Tabla 2. Promedios de interrupciones en la reproducción del video en cada dispositivo.

Dispositivo	iPhone4	iPad1
Promedio de Interrupciones sin precarga del buffer	14,3	9,3
Promedio de Interrupciones con precarga del buffer	9,4	5,6

En la Tabla 3 se muestran los resultados de la evaluación de QoE. Aunque existió gran diferencia en el número de interrupciones entre los dos dispositivos

(en iPhone4 fueron mayores las pausas), este factor no fue el único tenido en cuenta para dar la calificación de la QoE puesto que son medidas muy similares.

Tabla 3. Promedios de las calificaciones de QoE durante la reproducción del video en cada dispositivo.

Dispositivo	iPhone4		iPad1	
Tiempo de precarga del buffer	0 s	15 s	0 s	15 s
Calificación promedio	2,1	3,7	2,5	3,9

Este resultado se debe a que otro factor como la expectativa del usuario, jugó un papel importante ya que ante un dispositivo que se supone de mayor capacidad y calidad un usuario tiene una gran expectativa en todo aspecto y califica mas drásticamente si no cumple sus expectativas. Con referencia al análisis anterior en cuanto a la expectativa del usuario se esperaría que no fuera considerable la diferencia entre la evaluación de QoE cuando se deja cargar el buffer es decir con menos interrupciones, en este caso el mejor calificado es el iPad1 ya que el usuario, cuando no se enfoca en las interrupciones, cambia su atención de evaluación a otras características, sonido, imagen, tamaño de la pantalla, resolución. Aspectos en los que una tablet como el iPad1 supera a un Smartphone como el iPhone4 ya que, por ejemplo, es más cómodo visualmente ver un video en una pantalla más grande y con un sonido más fuerte.

Otro aspecto que se puede observar en la Tabla 3 es que se esperaría una calificación QoE mayor que 4 en los dos dispositivos cuando existieron menor cantidad de interrupciones, ya que si el servicio presenta menos defectos es de esperar una calificación sobresaliente. Esto no ocurrió debido a que como usuarios siempre se espera un servicio con calidad, tamaño de pantalla más grande, entre otros, y que el contenido del video no fue del gusto de algunos encuestados.

Tabla 4. Importancia de cada factor durante la evaluación de la QoE.

Dispositivo		iPhone4		iPad1	
Tiempo de precarga del buffer		0 s	15 s	0 s	15 s
Imagen	Muy importante	20	24	14	25
	Poco importante	6	4	9	5
	Nada importante	4	2	7	0
Audio	Muy importante	18	22	14	23
	Poco importante	8	6	9	4
	Nada importante	4	2	7	3
Interrupciones	Muy importante	30	5	21	10
	Poco importante	0	9	9	13
	Nada importante	0	16	0	7

de interrupciones y de la calificación de QoE. Por ejemplo, donde se muestra cómo en las pruebas con el iPhone4 fueron muy importantes las interrupciones cuando no se dejaba precargar el buffer, estas interrupciones llevaban a que los encuestados se enfocaran también de manera notable en la imagen y el audio (las interrupciones afectan cuando se deja precargar el buffer). En ninguno de los dos casos, con y sin precarga del buffer, la imagen y el audio fueron descartados por los encuestados como aspecto importante a la hora de dar una calificación sobre la percepción de la reproducción del video. Existe una relación directa entre interrupciones e imagen durante la reproducción de un video en el iPhone4 y el iPad1 que lleva a que estos aspectos sean de gran importancia al momento de dar una calificación cuantitativa de QoE.

En la Figura 1 se puede observar cuál fue el aspecto más importante en cada caso de prueba. Por ejemplo, cuando se realizó la prueba a los encuestados con el iPhone4, el factor más trascendente cuando no se dejó precargar el video fueron las interrupciones para todos los encuestados, y en el caso en que se dejó precargar el buffer, el aspecto más resaltado fue la imagen por 24 de los encuestados. Las personas que realizaron las pruebas con el iPhone4 se enfocaron más en las interrupciones como se puede ver en la Tabla 4 y la Figura 1, mientras que en

Con lo descrito en la Tabla 4 soportamos los análisis anteriormente hechos sobre los promedios

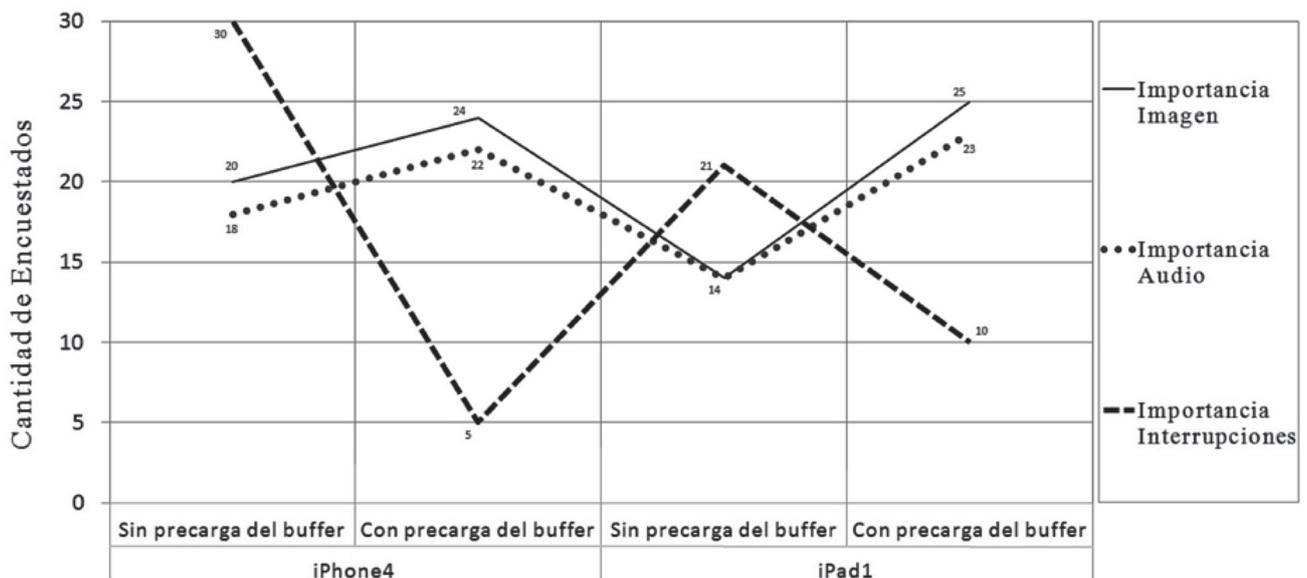


Figura 1. Aspectos más importantes para los encuestados en cada caso de prueba realizado.

el caso del iPad1 el aspecto más notable cuando no se dejaba precargar el video continuo. Cada persona tuvo una expectativa mayor sobre un dispositivo como el iPad1 y por tanto, no solo se fijó en las interrupciones, sino también en la imagen o el audio. Se puede decir entonces que las personas que realizaron las pruebas con el iPad1 se enfocaron más en la imagen ya que eran más altas las expectativas hacia la tablet de Apple.

Comparando la Figura 1 con la Tabla 2 se puede analizar que la mejor calificación de QoE en promedio que hicieron los encuestados fue en el caso en que se presentaron una menor cantidad de interrupciones es decir cuando se dejaba precargar el buffer del video. Cabe resaltar que las personas que realizaron las encuestas tenían una experiencia previa en el uso de smartphones y tablets, y ninguno conocía el portal de videos Vimeo elegido para realizar las pruebas.

Conclusiones

Resaltamos la dificultad de desarrollar un estudio de QoE subjetiva por las distintas maneras en que piensa cada persona de lo que es algo bueno o malo al momento de evaluar un servicio multimedia como el de video, también cada encuestado en ocasiones es difícil de persuadir en enfocarse solamente en ciertos aspectos ya que por ejemplo muchos de ellos no desligaban el concepto de QoC.

Se pudo observar que las interrupciones durante la reproducción de video en las pruebas realizadas tienen un impacto distinto, que está directamente relacionado con el dispositivo y con la poca o mucha experiencia que tenían en el uso de smartphones o tablets.

Referencias

1. Herrera Pérez, Enrique, Introducción a las telecomunicaciones modernas, 1 vol, México, D, F. 2004.
2. Figueiras, Aníbal R, Una panorámica de las telecomunicaciones, Madrid, 2002.
3. Jimenez Rodriguez, Erick, Nunome Toshiro, y Tasaka Shuji, "Assessment of User Behavior and QoE in Multi View Video And Audio Ip Transmission", 15th Asia-Pacific Conference on Communications, APCC 2009-188.
4. HuoI, Yusong, SuI Yujie, WangI Zhenhua, WuI Jun, y Yan Ma, "Measurement Based Modeling and Assessment of Qoe Index in P2p Iptv Services", IC-BN MET 2010.
5. Piamrat, Kandaraj, Viho Cesar, Ksentini Adlen, y Bonnin Jean Marie, "Quality of Experience Measurements For Video Streaming Over Wireless Networks", 2009 Sixth International Conference on Information Technology.
6. DeMoor, Katrien, Wout Joseph, Istv'an Ketyk'ó, Emmeric Tanghe, Tom Deryckere, Luc Martens, y Lieven De Marez, "Linking Users' Subjective QoE Evaluation to Signal Strength in an IEEE 802.11b/G Wireless LAN Environment", EURASIP Journal on Wireless Communications and Networking.
7. Tominaga, Toshiko, Takanori Hayashi, Okamoto Jun, y Takahashi Akira, "Performance Comparisons of Subjective Quality Assessment Methods for Mobile Video", NTT Service Integration Laboratories, NTT Corporation.
8. Mok, Ricky K. P., Cha, Edmond W. W., y Chang, Rocky K. C., "Measuring the Quality of Experience of HTTP Video Streaming", Department of Computing, The Hong Kong Polytechnic University, Hungghom, Kowloon, Hong Kong.
9. K.-T. Chen, C.-Y. Huang, P. Huang, y C.-L. Lei, "Quantifying Skype User Satisfaction," Proc. Of SIGCOMM'06, Pisa, Italy, September 11-15, 2006.
10. A.G. Davis, D. Bayart, y D.S. Hands, "Hybrid No-Reference Video Quality Prediction," Proc. of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Bilbao, Spain, May 13-15, 2009.
11. Vimeo, LLC, 555 West 18th Street, New York, Vimeo [sitio en internet], disponible en: <http://vimeo.com/>. Acceso continuo mes de enero, febrero y marzo de 2012.

12. "Subjective video quality assessment methods for multimedia applications" Recommendation ITU-T P.910 04/08.
13. "Mean Opinion Score (MOS) terminology", Recommendation ITU-T P.800.1, 07/06.
14. "Reference guide to quality of experience assessment methodologies", Recommendation ITU-T G.1011, 06/10.
15. E.L. Law, V. Roto, M. Hassenzahl, A.P. Vermeeren, y J.Kort, "Understanding, Scoping and Defining User Experience: A Survey Approach", Proceedings of the 27th international conference on Human factors in computing systems, Boston, MA, USA: ACM, 2009, pp. 719-728.
16. Brooks, Peter, "User Measures of Quality of Experience: Why Being Objective and Quantitative is Important", Journals & Magazines IEEE, 2010.
17. Kuan-Ta, Chen, y Chi-Jui Chang, "Quadrant of Euphoria: A Crowdsourcing Platform for QoE Assessment", National Taiwan University.
18. Kuan-Ta, Chen, Cheng Chun TU, y WeiCheng Xiao, "OneClick: A Framework for Measuring Network Quality of Experience", IEEE Communications Society.
19. Spagni de Barletta, Beatriz, Estadística Basica Probabilidad, 3ª ed, Universidad Nacional del Litoral, Santa Fe, Argentina, 2005.