

Estimación automática de calidad de video en televisión digital abierta ISDB-T basada en indicadores objetivos y subjetivos

Automatic Estimation of Video Quality for ISDB-T free-to-air Digital Television Based on Subjective and Objectives Indicators

Diego Durán¹, Juan Pablo Garella², Marcos Juayek³, Jose Joskowicz⁴, Rafael Sotelo⁵

Recibido: Mayo 2014

Aceptado: Agosto 2014

Resumen.- El despliegue de la Televisión Digital Terrestre (TDT) ha comenzado en la mayoría de los países de América Latina que han adoptado el estándar Japonés-Brasileño ISDB-T. Un requisito fundamental para el exitoso despliegue de la tecnología es lograr una buena calidad de experiencia (QoE) percibida por los usuarios. Con el objetivo de brindar herramientas a los radiodifusores para que velen por la calidad de servicio que brindan a sus televidentes, se han desarrollado indicadores subjetivos así como objetivos de estimación de la calidad percibida de video en la señal de televisión. En este artículo se presenta el trabajo realizado así como los resultados obtenidos por el grupo de investigación.

Palabras clave: TDT; VQI; QoE; Interactividad.

Summary.- The deployment of Digital Terrestrial Television (DTT) has begun in most Latin American countries that have adopted the ISDB-T standard. A key requirement for the successful deployment of the technology is to achieve a good perceived quality of experience (QoE) by the users. With the aim of providing tools for broadcasters to ensure the quality of service they provide to their users, we have implemented subjective and objectives indicators for estimating the perceived video quality of digital television signals. This article describes the work and the results obtained by the research group.

Keywords: DTT; VQI; QoE; Interactivity.

1. Introducción.- El despliegue de la Televisión Digital Abierta (TVD) en Uruguay ya ha comenzado. Al igual que la mayoría de los países de la región, Uruguay ha adoptado la norma ISDB-T (con las modificaciones realizadas por Brasil al estándar japonés original) definida por la Asociación Brasileña de Normas Técnicas (ABNT) [1]. En el correr de este documento la sigla ISDB-T se utiliza para hacer referencia a la versión brasileña y no a la original de la norma. Acompañando la implementación de la TVD en el territorio uruguayo el gobierno ha incentivado a través de la Agencia Nacional de Investigación e Innovación (ANII) [2] la realización de proyectos vinculados a la temática. De esta manera surge el proyecto VQI – Video Quality Indicators [3], primer proyecto ejecutado en conjunto por la Facultad de Ingeniería de la Universidad de Montevideo (UM) y la Facultad de Ingeniería de la Universidad de la República

1 Facultad de Ingeniería, Universidad de Montevideo, Montevideo, Uruguay, dduran@correo.um.edu.uy

2 Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, jpgarella@fing.edu.uy

3 Facultad de Ingeniería, Universidad de Montevideo, Montevideo, Uruguay, mjuayek@correo.um.edu.uy

4 Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, josej@fing.edu.uy

5 Facultad de Ingeniería, Universidad de Montevideo, Montevideo, Uruguay, rsotelo@um.edu.uy

(UdelaR) y gestionado por la Fundación Julio Ricaldoni (FJR) [4]. Consistió en la generación de indicadores de calidad de video y en el desarrollo y puesta en marcha de un sistema automático de estimación de calidad percibida por los televidentes de TVD.

En gran medida, la aceptación de una nueva tecnología depende de la calidad de experiencia percibida (Quality of Experience - QoE) por parte del público. Poder contar con herramientas que permitan obtener esta información brinda a los operadores, en este caso radiodifusores, la posibilidad de velar por la calidad de servicio que ofrecen a sus usuarios. De esta manera y siguiendo esta línea de pensamiento se observó la necesidad de generar indicadores que permitan medir la calidad percibida por los usuarios.

Estimar la QoE no es una tarea sencilla, involucra un sinnúmero de factores que afectan la percepción global de la experiencia por parte del usuario. Por ello, se debió acotar el problema haciendo énfasis sólo en la calidad percibida de video, siendo uno de los principales factores que afectan la percepción global de la calidad.

El procedimiento más confiable para medir la calidad percibida de video es la evaluación subjetiva. Se debe realizar de acuerdo a recomendaciones de organismos internacionales, en ambientes controlados y con un número determinado de personas que opinan respecto de su percepción, garantizando que la muestra sea estadísticamente significativa. Su resultado es la opinión media, que es la métrica generalmente aceptada como medida de la calidad y se denomina comúnmente MOS (*Mean Opinion Score*). Estos procedimientos son costosos y difíciles de realizar, no es un método practicable en aplicaciones de tiempo real como es el caso de la TVD, ni replicable de manera permanente.

Por ello se vio la necesidad de utilizar métodos automáticos y objetivos, es decir, en los que no se precisa la participación de sujetos. Dentro de esta categoría se encuentran los métodos *Full reference (FR)*, *Reduce reference (RR)* y *No Reference (NR)*. Los métodos FR hacen uso de la señal original y la degradada para estimar la calidad percibida de video. Como se menciona en [5] una de las medidas más simples utilizadas es el error cuadrático medio (RMSE) y la relación señal a ruido de pico (PSNR) aunque cabe destacar que su fácil implementación se ve opacada por sus resultados, que están poco correlacionados con la percepción de calidad por parte del usuario. Este tipo de métodos es impracticable para una aplicación en tiempo real en TVD dado que difícilmente se cuente con la señal original. Por otro lado existen los modelos RR que hacen uso de un canal auxiliar por donde se envían datos asociados a la señal original. Por último, los modelos NR sólo utilizan la señal degradada. Estos son los modelos más deseables de utilizar debido a que no requieren un canal adicional para hacer llegar la información complementaria y solamente se basan en la imagen desplegada para estimar la calidad, al igual que lo hace un televidente. A diferencia de modelos FR y en algunos casos RR, lamentablemente aún no se han estandarizado modelos NR para TV por parte de ITU (*International Telecommunications Union*).

Se optó por partir de la base de un modelo presentado en [6], el cual fue propuesto por el grupo de investigación anteriormente. Este modelo fue adaptado a las características del sistema ISDB-T y sus degradaciones asociadas al proceso de codificación de fuente, codificación de canal y transmisión.

En el correr del proyecto se han desarrollado indicadores automáticos y objetivos de estimación de calidad de video, así como indicadores subjetivos, basados en una aplicación interactiva para que los propios televidentes puedan calificar la calidad de imagen de la señal de TVD desde su hogar. A su vez se diseñó e implementó un portal en Internet donde se publican los resultados de manera automática.

El documento está organizado de la siguiente manera. En la Sección 2 se describe al sistema de manera global mencionando cada una de sus partes. En la Sección 3 se presenta el indicador subjetivo de calidad percibida de video, basado en el concepto de *crowdsourcing*, que se definirá, y haciendo uso de una aplicación de televisión digital interactiva permite recolectar la opinión de los usuarios de TVD desde sus hogares. En la Sección 4 se describen las características del Modelo Objetivo propuesto, y las evaluaciones subjetivas realizadas con el fin de entrenar y validar los resultados obtenidos. En la Subsección 4.3 se detalla el funcionamiento de un autómata desarrollado para estimar la calidad. En la Sección 5 se hace referencia a la interfaz web donde se publican los resultados de los indicadores de calidad de video. Por último en las secciones 6 y 7 se encuentran las conclusiones y el trabajo a futuro respectivamente.

2. Descripción Global del sistema.- En la Figura I se ilustra el sistema global. Consta de una aplicación interactiva que corre sobre el middleware Ginga [7] instalado en los sintonizadores de TVD en los hogares de los usuarios. Esta aplicación permite consultar al televidente su opinión respecto a la calidad de la imagen que está percibiendo y, por medio de Internet como canal de retorno, enviar su opinión hacia la base de datos. Por otro lado se cuenta con un receptor dedicado a la estimación de calidad de video por medio de modelos objetivos. Este receptor graba las señales de los canales de TVD de acuerdo a una agenda preestablecida, las procesa y envía los resultados al servidor central para su almacenamiento. Estos indicadores se almacenan en una base de datos y se pueden consultar por medio de un portal web.

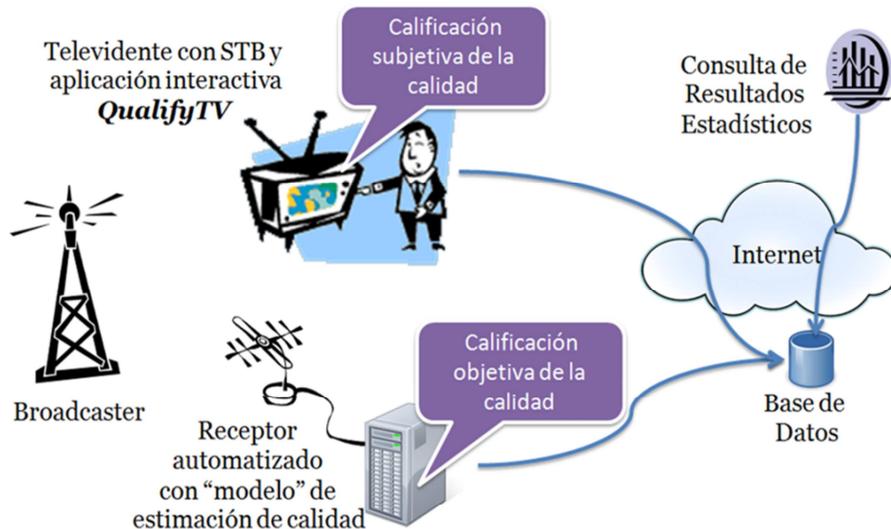


Figura I.- Ilustración global del sistema implementado.

3. Indicador Subjetivo de calidad de video.- Como se mencionó en secciones anteriores el procedimiento más confiable para medir la calidad percibida de video es la evaluación subjetiva. Esto requiere de estrictas pautas y metodologías para validar los resultados y es impracticable en aplicaciones de tiempo real. Por lo tanto se optó por otra alternativa, basada en el concepto de *crowdsourcing* y haciendo uso de la interactividad permitida en el estándar de TVD. El *crowdsourcing* es una nueva metodología que se vale de que los usuarios están conectados de manera masiva, dejando la posibilidad de realizar una amplia recolección de datos, algo que antes estaban disponibles sólo en condiciones de laboratorio y de manera reducida. Se ideó una aplicación llamada QualifyTV que corre sobre el middleware Ginga [7], plataforma de

interactividad de ISDB-T, desarrollada en los lenguajes NCL (Nested Context Language) y Lua. Ginga, nombre del middleware abierto del sistema brasileño de TV Digital (SBTVD), se subdivide en dos subsistemas principales interrelacionados, que permiten el desarrollo de aplicaciones siguiendo dos paradigmas de programación diferentes. Estos dos subsistemas se llaman Ginga-J (para aplicaciones procedurales Java) y Ginga-NCL (para aplicaciones declarativas NCL) [7].

Ginga-NCL proporciona una infraestructura de presentación de aplicaciones declarativas escritas en lenguaje NCL (Nested Context Language). NCL es un lenguaje de aplicación XML que permite definir presentaciones multimedia interactivas. Con NCL, se puede describir el comportamiento temporal de una presentación multimedia, asociar hiperenlaces (interacción con el usuario) entre los objetos multimedia que se presentan, definir comportamientos alternativos para una presentación (adaptación) y describir el layout de presentación entre varios dispositivos.

NCL es utilizado para desarrollar las presentaciones gráficas y las interacciones de la aplicación. Adicionalmente se utiliza el lenguaje Lua para extender algunas funcionalidades de NCL. Lua es un lenguaje de programación imperativo, estructurado y ligero que fue diseñado como un lenguaje interpretado con una semántica extensible, es utilizado por el subsistema Ginga-NCL como lenguaje de scripting.

El radiodifusor es el encargado de emitir la aplicación QualifyTV al aire, y el televidente puede activarla e interactuar con ella. Es aquí donde por medio de un menú se le indica al espectador que califique la calidad de imagen de la señal del 1 (Mala) al 5 (Excelente); dicho valor se envía a una base de datos y se guarda para su procesamiento y posterior publicación. En la Figura II se puede ver la aplicación corriendo en un televisor.





Figura II.- Aplicación interactiva: *QualifyTV*. La aplicación le pide al televidente que opine sobre su percepción, posteriormente guarda y envía la información recolectada a la base de datos vía Internet.

Los requisitos que se exigen en las evaluaciones subjetivas difícilmente se cumplan en los hogares de los espectadores de TVD que opinan respecto de su percepción. Otro factor a tener en cuenta es que, en general la actividad de mirar televisión es colectiva no permitiendo de esta manera la recolección de la opinión de cada individuo de manera independiente. Igualmente la adquisición de dicha información en grandes cantidades puede permitir un análisis estadístico mediante el uso de minería de datos, de gran ayuda para el despliegue de la TVD en el país.

4. Indicador Objetivo de Calidad de Video.- En esta sección se describe el modelo objetivo propuesto por el grupo de investigación, el proceso de entrenamiento y validación y finalmente la implementación de un autómata encargado de la adquisición, procesamiento y envío de la estimación de calidad de video para su posterior publicación.

4.1. Modelo Objetivo.- Durante los últimos años, se han propuesto diferentes modelos paramétricos para la estimación de la calidad de vídeo. En [6] se presenta una revisión de modelos paramétricos publicados por diferentes autores, en donde se describen brevemente y se contrastan sus desempeños utilizando un conjunto de clips de vídeo comunes, en diferentes escenarios de codificación y transmisión. Basándose en los resultados obtenidos, José Joskowicz et al. presenta un modelo paramétrico nuevo y más general que tiene en cuenta las degradaciones obtenidas en el proceso de codificación, así como también características extraídas del contenido del video y su transmisión sobre redes de paquetes IP. Dado que las características de transmisión de la TVD son diferentes a las redes de paquetes IP, se utilizó este modelo como base y se adaptó al sistema ISDB-T. Se tuvieron en cuenta la influencia de las degradaciones asociadas a las pérdidas de paquetes en transmisión ISDB-T y también la resolución HD, no prevista en el modelo inicial.

A continuación se presenta la formulación general del modelo, sobre la base de [6]:

$$MOS_p = 1 + I_c I_p \quad (1)$$

donde MOS_p es la predicción de MOS que realiza nuestro modelo. I_c está asociada a la pérdida de calidad por el proceso de codificación mientras que I_p a la pérdida de calidad debida al

proceso de transmisión, en nuestro caso, será principalmente debido a los paquetes perdidos. I_c , varía entre 0 y 4 y se calcula de la siguiente manera [6]:

$$I_c = v_3 \left(1 - \frac{1}{1+(ab/v_4)^{v_5}} \right) \quad (2)$$

donde,

$$v_3 = 4 + 4 (f_{\max} - f) (k_1 s + k_2 e^{-k_3 (f_{\max} - f) ab}) \quad (3)$$

$$v_4 = c_1 s^{c_2} + c_3 \quad (4)$$

$$v_5 = c_4 s^{c_5} + c_6 \quad (5)$$

La variable b representa al bitrate; a es un parámetro cuyo valor depende de la resolución de la imagen; f es el frame rate, siendo $f_{\max} = 25$ fps, c_1 a c_6 y k_1 a k_3 coeficientes del modelo. Por último s es una característica llamada *Sum of Absolute Differences* (SAD) extraída del contenido del video por un algoritmo conocido, el cual se detalla a continuación. En el área de compresión de video las técnicas de estimación de movimiento son utilizadas con el objetivo de disminuir la redundancia temporal de una secuencia de imágenes. El método más difundido es el llamado BMA (Block Matching Algorithm). Consiste en dividir un cuadro cualquiera, llamémoslo k , en bloques rectangulares de tamaño $M \times N$, por ejemplo: 8×8 . Luego se define una ventana de búsqueda y se compara un bloque i del cuadro k con los bloques dentro de una determinada ventana de búsqueda en el cuadro anterior, el $k-1$. El vector de movimiento se genera cuando de acuerdo al cómputo de alguna métrica se encuentra el bloque del cuadro $k-1$ que mejor represente al bloque original del cuadro k dentro de la ventana de búsqueda. La métrica utilizada más frecuentemente es el SAD (Sum of Absolute Differences), y consiste en la suma de las diferencias absolutas entre la luminancia del bloque i del cuadro k respecto a los bloques ubicados dentro de la ventana de búsqueda en el cuadro anterior, el $k-1$. Siguiendo este razonamiento la característica s es el promedio del cómputo de la métrica SAD asociada a cada uno de los vectores de movimientos obtenidos para cada bloque de cada cuadro del clip de video.

I_p por su parte, varía entre 0 y 1. Sería 0 en el caso de que el proceso de transmisión redujera la calidad a un mínimo, resultando en un MOS_p de 1. En cambio si el proceso de transmisión no altera la calidad I_p sería 1. La transmisión de video en TVD se basa en distintas estructuras constituyentes. Una de ellas es el *slice* que representa un tirilla o sucesión de bloques en un cuadro de video. El *slice* es una unidad que permite la utilización de técnicas de corrección de errores; de esta forma, en el receptor se tendrán *slices* sin errores que se verán bien en la pantalla, y otros afectados por errores que se verán mal, de distinta forma a la imagen original. Para el cómputo del parámetro I_p se tiene en cuenta el número total de *slices* afectados y se lo divide por el número total de *slices* presentes en el video analizado de acuerdo a la siguiente fórmula:

$$I_p = \left(\frac{1}{1+Kp} \right) \quad (6)$$

En donde p es precisamente la relación mencionada anteriormente y K es una constante que se utiliza para ajustar el modelo.

4.2. Entrenamiento y Validación: Pruebas Subjetivas.- Con el objetivo de ajustar el modelo de estimación de calidad se realizó una etapa de entrenamiento y luego una de validación. Para ello se realizaron evaluaciones subjetivas, obteniéndose una base de entrenamiento y otra de validación con clips de video calificados, es decir con su MOS asociado. Para la ejecución de las

evaluaciones se siguieron los procedimientos establecidos en la recomendación ITU-R BT.500-13 [8]. En la etapa de entrenamiento se utilizaron cinco videos extraídos de bases de datos internacionales como por ejemplo de *Consumer Digital Video Library* [9]. La selección se hizo en base a escoger distintas categorías tales como espectáculos, deportes o animaciones, y de distinto rango asociado a la cantidad de movimiento del contenido, es decir, clips rápidos (con mucho movimiento) y lentos (en los que hay pocos cambios). La selección de los videos de pruebas es fundamental dado que se quiere cubrir todas las posibilidades en las que está previsto que funcione el modelo. Luego cada video fue codificado utilizando el codec H.264 [10] en distintos *bitrates*, *group of pictures* (GoP) y resoluciones de pantalla. También fue degradado utilizando diferentes técnicas de introducción de errores asociados al canal de transmisión. Fueron utilizados errores en ráfaga y aleatorios en distintas configuraciones. Obteniéndose así una base de cien videos en calidad estándar (SD) y cien en alta definición (HD). Para la etapa de validación se utilizaron videos grabados de transmisiones de TVD de aire de dos broadcasters distintos de la ciudad de Montevideo por medio de una placa sintonizadora. En total la base de validación constó de cien videos distintos en SD y cien en HD.

Para facilitar y evitar errores en el despliegue de los videos y la recolección de los datos se desarrolló especialmente para el proyecto una aplicación que gobierna las pruebas subjetivas. Esta aplicación está disponible para su utilización en la página web del proyecto [3].

El resultado de la etapa de entrenamiento del modelo puede verse en la Figura III, en donde se muestra su dispersión. Entre los valores predichos por el modelo y los obtenidos en las pruebas subjetivas se obtuvo una Correlación de Pearson de 0,90 y un Error Cuadrático Medio (RMSE) de 0,43.

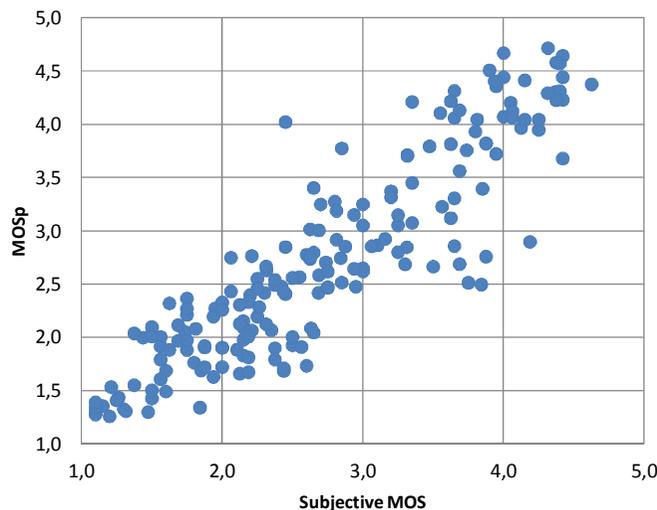


Figura III.- MOS estimado por el modelo objetivo vs. MOS obtenido en la etapa de entrenamiento.

Por otro lado la dispersión del modelo en la etapa de validación obtuvo una correlación de Pearson $PC = 0,92$ y un error cuadrático medio $RMSE = 0,46$. Se puede observar que los resultados obtenidos son muy buenos, y recordando que en esta etapa se utilizaron señales captadas de transmisiones de canales de aire en la ciudad de Montevideo, se comprueba la usabilidad del modelo con emisiones de TVD en ambientes reales.

El modelo objetivo de calidad de video fue diseñado y entrenado con un conjunto de videos de prueba especialmente preparado, y posteriormente validado con otro conjunto independiente de

videos grabados de señales reales de Televisión Digital Terrestre de distintos canales de Montevideo, Uruguay. En la siguiente sección se detalla el sistema automático responsable de aplicar este modelo a distintos segmentos de programación de TVD ofrecido por los broadcasters.

4.3. Autómata de calidad.- Se desarrolló un sistema que se encarga de la adquisición, ejecución del modelo y envío de los resultados al servidor central. Una ilustración del sistema implementado se puede ver en la Figura IV, consta de un servidor con una placa sintonizadora con antena UHF, una agenda de grabación que permite programar los horarios, canales y la duración de cada grabación. Luego, a los videos grabados se les aplica el Modelo Objetivo desarrollado y descrito en 5.1 y se devuelven los resultados, que son enviados hacia el servidor central para su publicación vía Web.



Figura IV.- Sistema encargado de la adquisición, procesamiento, aplicación de modelo objetivo, y envío de información asociada a la estimación de calidad percibida de video hacia el servidor central.

Video Quality Indicators

User: mjuayek, [Cerrar Sesión](#)

Administración **Consulta de evaluaciones subjetivas** Consulta de evaluaciones objetivas

Fecha Inicio: 01/04/2013 03:00 p.m.

Fecha Fin: 01/04/2013 05:00 p.m.

Canal: Todos

Fecha	Hora	Canal	Cant. Votos	MOS	Desv Estandar	BitRate Prom	
2013-04-01	15:00	10	2	5.0000	0.0000	1.5	Ver Detalle
2013-04-01	15:00	13	2	5.0000	0.0000	1.5	Ver Detalle
2013-04-01	15:00	2	2	5.0000	0.0000	1.5	Ver Detalle
2013-04-01	15:00	25	2	5.0000	0.0000	1.5	Ver Detalle
2013-04-01	15:00	5	6	3.3333	1.9664	1.5	Ver Detalle
2013-04-01	16:00	5	1	2.0000		1.5	Ver Detalle

Figura V.- Interfaz web de consulta de los resultados obtenidos de estimación de calidad.

5. Publicación de Resultados.- El indicador objetivo obtenido, así como las opiniones subjetivas para cada una de las señales de televisión abierta están accesibles desde una aplicación web de consulta desarrollada especialmente para tal fin. Esta aplicación permite la comparación de calidad entre las diversas señales, y colaborando de esta manera a que los radiodifusores velen por entregar la mejor calidad de imagen posible.

La aplicación web permite consultar en un rango de fecha específico los resultados de las evaluaciones objetivas y subjetivas. En el primer caso se puede consultar el valor obtenido en un determinado momento por un determinado canal, mientras que en las evaluaciones subjetivas se muestra el promedio de las calificaciones de los usuarios a un canal en un determinado horario, pudiendo además obtener el detalle puntual de las calificaciones de un horario particular. La Figura V muestra una captura de pantalla de esta aplicación.

6. Conclusiones.- Acompañando la transición de la televisión abierta analógica hacia la digital, se ha incentivado tanto a la comunidad académica como a empresas vinculadas al rubro a desarrollar e innovar haciendo uso de la tecnología venidera. Siguiendo esta línea, el grupo de investigación ha implementado un sistema capaz de generar indicadores tanto subjetivos como objetivos de estimación de calidad percibida de video. Basándose en el concepto de *crowdsourcing*, se ha diseñado e implementado una aplicación interactiva capaz de interactuar con los espectadores de TVD recolectando la opinión de su percepción sobre la calidad de video. Esta aplicación corre sobre el middleware Ginga. Con respecto a la automatización y medición objetiva de la calidad se logró implementar un modelo existente propuesto por el grupo de investigación, adaptado para medir la influencia de la pérdida de paquetes del *bitstream* en la calidad de video percibida en los espectadores. Esta y otras líneas de investigación han sido generadas con el objetivo de acercarse lo mejor posible la calidad global de experiencia.

7. Trabajo a futuro.- El trabajo ha dejado abiertas puertas para seguir avanzando en líneas prometedoras, tanto en el ámbito de la TVD como de otros medios de transmisión de contenidos audiovisuales. Se prevé la continuación del proyecto, incluyendo la evaluación del audio y generar indicadores para la experiencia multimedia. También se buscará la introducción de modelos NR capaces de inferir estimaciones de percepción a partir del contenido como pueden ser defectos típicos de video como *blurring* o *blockiness* entre otros.

8. Referencias.

- [1] Asociación Brasileña de Normas Técnicas [online]. Disponible en: <http://www.abnt.org.br>. [Consulta: 3 de Mayo de 2014].
- [2] Agencia Nacional de Investigación e Innovación [online]. Disponible en: <http://www.anii.org.uy>. [Consulta: 3 de Mayo de 2014].
- [3] Video Quality Indicators [online]. Disponible en: <http://ingenieria.um.edu.uy/vqi>. [Consulta: 3 de Mayo de 2014].
- [4] Fundación Julio Ricaldoni [online]. Disponible en: <http://www.ricaldoni.org.uy>. [Consulta: 3 de Mayo de 2014].
- [5] J. Joskowicz y R. Sotelo, "Modelo de estimación de calidad de video: video quality experts groups," Memoria investig. ing. (Facultad Ing., Univ. Montev.), n°10, pp. 7-17, 2012.
- [6] J. Joskowicz, R. Sotelo and J. C. López Ardao, "Towards a general parametric model for perceptual video quality estimation," IEEE Transactions on Broadcasting, vol 59, n°4, pp. 569- 579, Dec., 2013.

- [7] Sitio Oficial del Middleware Ginga [online]. Disponible en: <http://www.ginga.org.br>. [Consulta: 3 de Mayo de 2014].
- [8] Recommendation ITU-R BT.500-13 Methodology for subjective assessment of the quality of television pictures, 01/2012.
- [9] The Consumer Digital Video Library [online]. Available: <http://www.cdvl.org>. [Consulta: 9 de Mayo de 2014].
- [10] Information technology. Coding of audio-visual objects. Part 10: Advanced Video Coding, ISO/IEC 14496-10:2012, 2012.