

Bilboko Ingeniaritza Goi Eskola Teknikoa
Escuela Técnica Superior de Ingeniería de Bilbao

TESIS DE MÁSTER

TICRM

Máster y Doctorado en Tecnologías de la
Información y Comunicaciones en Redes Móviles

Estudio de la Calidad Experimentada (QoE)
en el Servicio de Navegación Web:
Contribución a la Validación de las
Recomendaciones ITU-T P.1501 e ITU-T G.1031

Autor **Eduardo SAIZ MACÍAS**
Directora **Dra. Ana Eva IBARROLA ARMENDÁRIZ**

Curso Académico
2013/2014

ÍNDICE

ÍNDICE DE FIGURAS	5
ÍNDICE DE TABLAS	7
1. INTRODUCCIÓN	8
2. CONTEXTO	10
2.1. QoXFERA	10
2.2. QoXCLOUD	13
3. OBJETIVOS	16
4. ESTADO DEL ARTE	18
4.1. QoS en el ámbito de las Telecomunicaciones	19
ITU - Unión Internacional de Telecomunicaciones	19
ISO – Organización Internacional de Normalización	20
ETSI – Instituto Europeo de Normas de Telecomunicaciones	20
4.2. QoE: Definición y modelos	21
Definición de QoE	21
Modelado de QoE sobre la base del Marco ITU-T G.1000	24
Modelos relacionales QoE-QoBiz	25
Modelos de QoE en el ecosistema de comunicaciones	27
QoE en el modelo global: De CSAT a QoXFERA	31
4.3. Métricas y metodologías de evaluación de la QoE	36
Naturaleza de la QoE	36
Recomendación ITU-T P.800 (1996)	40
Recomendación ITU-T G.1030 (2005)	42
Enfoque del ETSI STF 354: Introducción a la evaluación objetiva de QoE	46
Ejemplo de metodología de evaluación objetiva de QoE	48
4.4. QoE en el Servicio de Navegación Web	51
Estudio 1: Primer acercamiento a la medición de la calidad de navegación web	51
Estudio 2: Relación entre QoS y QoE en las sesiones de navegación web	54
Estudio 3: QoE en la navegación web en función del contenido	59
Estudio 4: Tiempos de espera en la QoE para servicios web	63
4.5. Conclusiones sobre el estado del arte	68
5. ARQUITECTURA PROPUESTA	70
5.1. Conceptos previos	70
Factores de influencia	72
5.2. Arquitectura propuesta en P.STMWeb	73
5.3. Arquitectura adoptada en el estudio	74
Equipos de usuario (UD)	75
Servidores EMU	76
Servidor Local de Contenidos (LCS)	78

5.4. Contenidos del LCS.....	80
Contenidos definidos en el borrador P.STMWeb.....	80
Contenidos adoptados en el estudio	81
Actualización de contenidos	82
6. METODOLOGÍA.....	83
6.1. Plan de pruebas.....	86
7. RESULTADOS	89
7.1. Análisis de los resultados	89
Resultados generales	89
Comparación de resultados por categoría de usuarios.....	90
Comparación de resultados por contenidos.....	91
Resultados sobre uso doméstico	92
7.2. Propuestas de mejora	93
Actualización de condiciones de test.....	93
Definición más precisa sobre el uso de cachés.....	94
Temporización.....	94
Definición de las tareas del participante.....	95
8. CONCLUSIONES.....	96
8.1. Valoración general del estudio	96
Aportaciones del estado del arte a la realización del estudio	96
Consideraciones sobre la evaluación de QoE objetiva	97
Validación de la recomendación ITU-T P.1501.....	97
Validación de la recomendación ITU-T G.1031	98
8.2. Líneas futuras	99
8.3. Publicaciones destacadas	100
9. GLOSARIO.....	101
9.1. Terminología.....	101
Calidad de Servicio y métricas.....	101
Equipos y módulos de la arquitectura de pruebas	101
Otros términos.....	101
9.2. Organizaciones y grupos	102
10. REFERENCIAS	103
ANEXOS	108

ÍNDICE DE FIGURAS

Figura 1.1 – Modelo de servicios tradicional vs. Modelo de servicios en la nube	8
Figura 2.1 –Capas del modelo QoXfera.....	10
Figura 2.2 – Especificación de las capas de QoXfera	11
Figura 2.3 – Metodología de QoXfera: Rotación y Traslación	11
Figura 2.4 – Relación de proyectos de NQaS con la arquitectura de capas de QoXfera	12
Figura 2.5 – Propuesta de plataforma en la nube de la ITU	13
Figura 2.6 – Arquitectura de QoXcloud e integración de QoXfera.....	14
Figura 4.1 – Enfoque de la QoS orientada al usuario de la ETSI TS 102 250-1.....	20
Figura 4.2 – Aspectos de la QoE (ETSI TS 102 250-1).....	22
Figura 4.3 – Proceso de formación de la Calidad (Qualinet).....	23
Figura 4.4 – Las cuatro vertientes de la QoS (ITU-T G.1000).....	24
Figura 4.5 – Relación de estándares de la ITU-T en el modelo de QoX (Stankiewicz et al)	25
Figura 4.6 – Relación entre QoS, QoE y QoBiz (Moorsel)	26
Figura 4.7 – Métricas de los conceptos QoS, QoE y QoBiz (Moorsel).....	26
Figura 4.8 – Marco para la gestión de la QoS y detalle de relaciones (Kilkki)	26
Figura 4.9 – Ecosistema de comunicaciones (Laghari).....	28
Figura 4.10 – Modelo de interacción de QoE en el ecosistema de comunicaciones (Laghari).....	28
Figura 4.11 – QoE en el ecosistema para un proveedor de aplicaciones/servicios (Qualinet).....	29
Figura 4.12 – Modelo de Satisfacción CSAT (Xiao).....	32
Figura 4.13 – Modelo de gestión global de QoS en un ISP (Ibarrola).....	32
Figura 4.14 – Metodología del modelo de gestión global de QoS en un ISP (Ibarrola)	32
Figura 4.15 – Satisfacción/Desatisfacción según la experiencia del cliente (Schindlholzer)	33
Figura 4.16 – Arquitectura de capas de QoXfera.....	36
Figura 4.17 – Tendencia general de la curva de relación QoS-QoE	40
Figura 4.18 – Conexión de referencia con ejemplos de dispositivos intermedios (ITU-T G.1030).....	43
Figura 4.19 – Descripción gráfica de la sesión de la prueba (ITU-T G.1030)	45
Figura 4.20 – Resultados de los participantes en una escala de 6 segundos (ITU-T G.1030).....	46
Figura 4.21 – Interacción entre dominios hacia la QoE total (Laghari)	47
Figura 4.22 – Comparación objetiva de diferentes resoluciones (Zinner)	50
Figura 4.23 – Comparación objetiva de diferentes tasas de frames para el vídeo ‘crowd run’ (Zinner).....	50
Figura 4.24 – Esquema de captura de experiencia de navegación web (Fallon).....	52
Figura 4.25 – Tiempos de carga de portales web en función del tipo de acceso (Fallon).....	53
Figura 4.26 – Métricas relacionadas con la descarga de un elemento de una página web (Cui).....	55
Figura 4.27 – Clustering de todas las sesiones del experimento controlado para los 3 terminales (Cui).....	57
Figura 4.28 – Clustering de las sesiones navegación libre (Cui)	58
Figura 4.29 – Temporización del experimento de Nguyen.....	59
Figura 4.30 – MOS vs Retardo y peticiones por segundo (Nguyen).....	60
Figura 4.31 – Los 4 tipos de contenidos del experimento de Nguyen	61
Figura 4.32 – MOS por tipo de contenido (Nguyen)	62

Figura 4.33 – Duración percibida vs Duración real (Seow).....	64
Figura 4.34 – Prueba ITU-T G.1030 [a] dentro de una sesión con distintos tiempos de espera [b] (Egger)	65
Figura 4.35 – Relación Estímulo-MOS [izq] vs. Relación KPI-MOS [dch] (Egger).....	66
Figura 5.1 – Relación QoS/QoE en la definición de un sistema de validación (S.H. Jeong)	71
Figura 5.2 – Arquitectura de red propuesta para la evaluación de QoE (P.STMWeb).....	73
Figura 5.3 – Arquitectura de red desplegada para la evaluación de QoE.....	74
Figura 5.4 – Arquitectura funcional del Servidor Emulador (EMU).....	77
Figura 5.5 – Arquitectura funcional del Servidor Local de Contenidos (LCS)	78
Figura 5.6 – Capturas de los contenidos alojados en el LCS: noticias y álbum fotográfico	82
Figura 6.1 – Ciclo de vida del estudio	83
Figura 6.2 – Fases del plan de pruebas y detalle de la fase de Registro	87
Figura 6.3 – Captura de la encuesta de condición.....	88
Figura 7.1 – Resultados MOS-BW para el total de participantes.....	89
Figura 7.2 – Gráficas comparativas de resultados MOS-BW por categoría de usuarios	90
Figura 7.3 – Gráficas comparativas de resultados MOS-BW por contenidos.....	91
Figura 7.4 – Aceptación de la calidad experimentada para su uso doméstico	92

ÍNDICE DE TABLAS

Tabla 4.1 – Evaluación de QoE: Subjetiva vs Objetiva (Brooks & Hestnes)	38
Tabla 4.2 – Evaluación de QoE: Cualitativa vs Cuantitativa (Brooks & Hestnes)	38
Tabla 4.3 – Tabla resumen para la clasificación de metodologías de evaluación de QoE.....	39
Tabla 4.4 – Marco para el proceso de evaluación de la calidad IP extremo a extremo (ITU-T G.1030).....	42
Tabla 4.5 – Mapeo de métricas objetivas a métricas subjetivas de QoE (Zinner)	48
Tabla 4.6 – Propiedades de las secuencias de referencia (Zinner).....	49
Tabla 4.7 – Características de las redes de acceso del estudio de Fallon	51
Tabla 4.8 – Características de los portales web del estudio de Fallon	52
Tabla 4.9 – Relación QoS-QoE del estudio de Cui.....	58
Tabla 4.10 – Resultados de valoración MOS por contenidos (Nguyen)	61
Tabla 5.1 – Características de los equipos de usuario de la arquitectura adoptada	76
Tabla 5.2 – Características de los servidores EMU de la arquitectura adoptada	76
Tabla 5.3 – Características del servidor local de contenidos (LCS) de la arquitectura adoptada.....	78
Tabla 5.4 – Características del portal de noticias del LCS.....	81
Tabla 5.5 – Características de la galería fotográfica del LCS	81
Tabla 6.1 – Características de los participantes del estudio	85
Tabla 6.2 – Condiciones de test de P.STMWeb.....	86
Tabla 6.3 – Posible secuencia aleatoria de velocidades	86
Tabla 6.4 – Temporización de las fases del experimento	88
Tabla 7.1 – Escala de condiciones del draft P.STMWeb frente a valores de referencia nacionales	93
Tabla 7.2 – Fases del experimento (versión estándar vs. versión corta).....	95

1.INTRODUCCIÓN

En las últimas décadas, las Tecnologías de la Información y Comunicaciones (TIC) han evolucionado enormemente gracias al auge de Internet y a la expansión del mercado de los dispositivos móviles. Sin embargo, tras años de crecimiento, el cambio en las condiciones del mercado junto a la situación económica actual, han traído consigo grandes recortes que han frenado drásticamente la inversión en este sector.

Esta situación de crisis ha supuesto también una oportunidad para el cambio. Muchos de los servicios que hasta hace relativamente poco tiempo eran ofertados únicamente por los operadores de telecomunicaciones, han pasado en la actualidad a ser prestados por terceros gracias al modelo de servicios en la nube, impulsando el auge de este modelo frente al tradicional (figura 1.1).

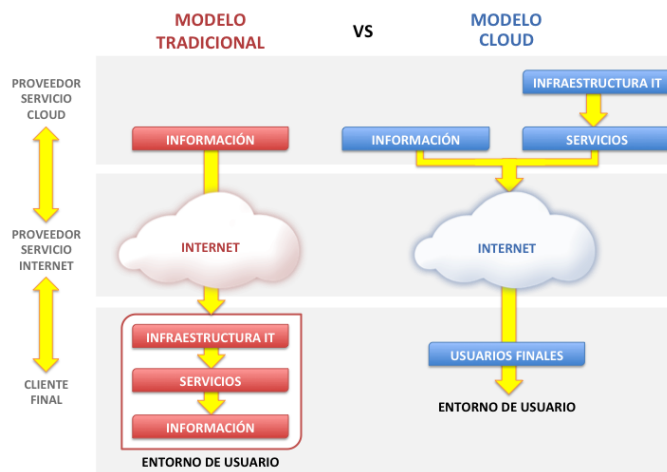


Figura 1.1 – Modelo de servicios tradicional vs. Modelo de servicios en la nube

En paralelo a los cambios en el ámbito económico-financiero, se han producido también una serie de cambios socioculturales de gran repercusión en los usuarios del sector. El perfil de los usuarios de los servicios de telecomunicación está evolucionando debido a este proceso de globalización: Cada vez existe una mayor demanda de conectividad y ubicuidad por parte de los usuarios, que, a su vez, exigen mejores condiciones de capacidad y calidad de servicio (QoS) de los servicios contratados. Estos cambios socioculturales han tenido especial repercusión en el ámbito de Internet, siendo en gran parte responsables de la proliferación de nuevos servicios en los que la percepción de la calidad por parte del usuario se convierte en un factor determinante a la hora de su contratación. Esto, a su vez, trae consigo un incremento en la complejidad en la implantación por parte del proveedor de políticas para la gestión de la QoS demandada por el cliente.

En este nuevo escenario, el concepto de calidad de servicio ha venido evolucionando desde su definición clásica, entendida como el análisis objetivo de los indicadores intrínsecos de calidad que definen el rendimiento de un servicio, hacia una visión subjetiva centrada en el punto de vista del cliente respecto a dicho servicio.

No es de extrañar, por tanto, la aparición de nuevos términos relacionados con la QoS que han pretendido abarcar diferentes enfoques orientados a la satisfacción del usuario y la fidelización del cliente, como son la calidad percibida por el usuario (QoP – Quality of Perception) [1], la calidad experimentada por el cliente (QoE – Quality of Experience) [2] o la calidad de negocio (QoBiz – Quality of Business) [3].

En vista de esta tendencia, tanto en el ámbito científico como en el entorno de los organismos de regulación y estandarización, se ha consolidado la idea de que el desarrollo de una gestión efectiva de la QoS no es posible desde el estudio aislado de cada enfoque. Por tanto, es necesario un análisis más complejo que abarque y relacione las diferentes visiones y vele por los intereses de los diferentes agentes que intervienen en la provisión de servicios: los proveedores, los usuarios y los reguladores. Esta interrelación de enfoques de calidad de servicio se engloba en lo que muy recientemente se ha dado en llamar QoX [4] con el objetivo de transmitir, de forma más certera, la nueva dimensión de lo que una gestión óptima de la QoS debería abordar.

A pesar de ello, existe una enorme complejidad a la hora de establecer dicha relación de una manera adecuada más allá de fundamentos teóricos, máxime si se pretende asentar dicho análisis en base a normativas y estándares existentes en la actualidad. Existen varios trabajos recientes [5][6][7] que han abordado de manera más o menos satisfactoria dicha complejidad, uno de los cuales, QoXfera [7], sirve de marco de desarrollo para el estudio del presente documento.

En este contexto, organismos de estandarización como la ITU, a través de sus grupos de trabajo, están fomentando el interesante planteamiento de abordar los servicios de telecomunicaciones ampliamente asentados y más que estudiados desde la vertiente clásica de la QoS, desde este nuevo prisma enfocado al cliente, generando nuevos borradores [8][9][10] para futuras recomendaciones en las que poder establecer una metodología de estudio de la QoE en dichos servicios.

En particular, el trabajo presentado en este documento se asienta en la validación del *draft* de la recientemente aprobada recomendación ITU-T P.1501 (Metodología de test subjetiva para el servicio de navegación web) [11], así como en la contribución a la recomendación G.1031 (Factores de QoE en el servicio de navegación web) [12].

Este trabajo, a su vez, se engloba en un proyecto más ambicioso, llamado QoXcloud, desarrollado en el seno del grupo de investigación NQaS (Networking Quality and Security) [13] de la Universidad del País Vasco (UPV/EHU), en el que se pretende dar un primer paso hacia evaluación de la QoX de los servicios en la nube, así como colaborar con la ITU en el proceso de validación de nuevas recomendaciones y estándares.

2. CONTEXTO

El presente trabajo está englobado dentro de un proyecto de mayor envergadura denominado QoXcloud [14]. El objetivo del proyecto QoXcloud es el de desarrollar una plataforma para la medida de la calidad experimentada por los usuarios de los servicios de telecomunicación a través de la nube. Con ello, se busca ofrecer una gestión eficaz de la calidad de servicio, así como proporcionar datos de QoX a todos los agentes que intervienen en la prestación de los servicios de telecomunicación en la nube.

2.1. QoXFERA

La plataforma QoXcloud se asienta sobre la arquitectura de referencia para la gestión global de la calidad de servicio (QoX) en los servicios de telecomunicaciones, denominada QoXfera [7]. Esta arquitectura permite la aplicación de un modelo de gestión de la QoX sobre escenarios reales de redes avanzadas (heterogéneas y *multi-play*) para los servicios de telecomunicación, con el objetivo resolver una demanda actual del mercado como es el control de la provisión de QoS en las redes de datos y la publicación de datos de mercado en este entorno.

Su arquitectura busca englobar los distintos enfoques de la QoS a través de un modelo esférico dividido en cuatro capas, cada una de las cuales se centra en la gestión y evaluación de diferentes conceptos de la calidad de servicio: La QoS intrínseca (indicadores objetivos de QoS), la QoS percibida (QoP), la QoS experimentada (QoE) y la calidad de negocio (QoBiz) (figura 2.1). La definición de estas capas y las relaciones con las capas adyacentes se basa fundamentalmente en la normativa y regulación existente, tomando como referencia los organismos de estandarización internacionales (ITU-T) y europeos (ETSI), así como las métricas definidas por asociaciones industriales como el Telemangement Forum (TMForum) [15] cuya incorporación al modelo QoXfera ha sido recientemente considerada a raíz de la tesis de máster presentada el año pasado al respecto de la '*definición de un modelo para la gestión de la calidad de negocio (QoBiz) en los servicios de telecomunicación*' [16].



Figura 2.1 – Capas del modelo QoXfera

En cuanto al enfoque metodológico de la arquitectura QoXfera, éste se centra en el análisis de la satisfacción del usuario basado en la estimación de la calidad experimentada, considerando sus expectativas y las políticas de optimización de la infraestructura para garantizar la rentabilidad del proveedor mediante la fidelización del cliente. Para ello, la especificación de capas presentada en la [figura 2.2](#) define, a su vez, un concepto novedoso de relaciones tanto internas a la capa, como relaciones intercapa, permitiendo al modelo esférico describir, mediante un proceso iterativo, movimientos similares a los de un planeta: rotación y traslación ([figura 2.3](#)).

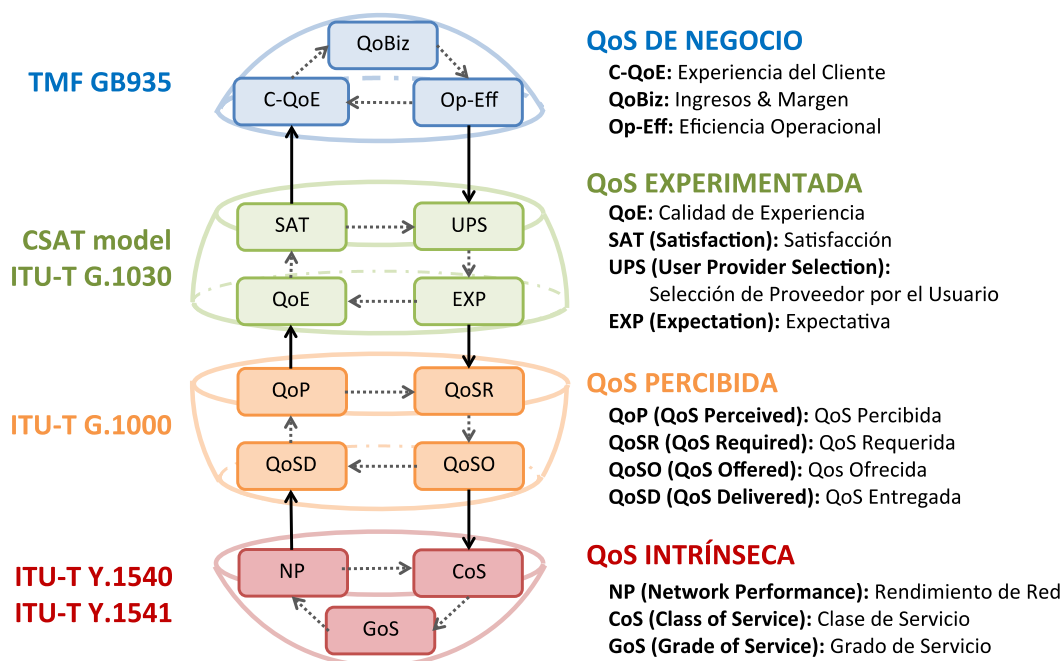


Figura 2.2 – Especificación de las capas de QoXfera

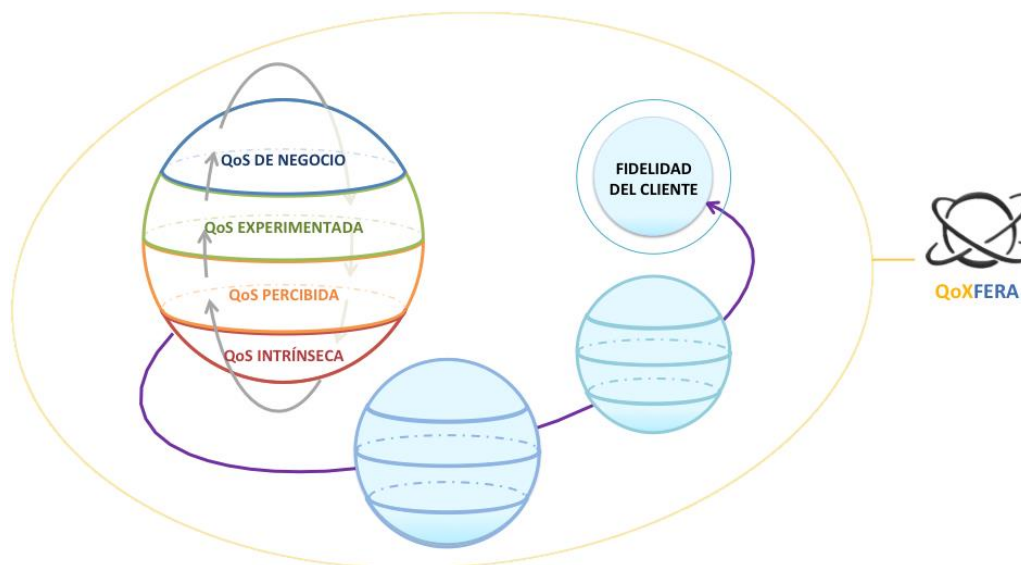


Figura 2.3 – Metodología de QoXfera: Rotación y Traslación

El movimiento rotacional sobre el eje vertical permite alimentar las capas adyacentes con la información obtenida en el análisis interno de cada capa. Asimismo, en el proceso iterativo de comunicación intercapa, se produce el movimiento de traslación. El objetivo de dicha iteración es la de favorecer la convergencia de las cuatro capas de QoS de cara a garantizar tanto la satisfacción del cliente como la rentabilidad del proveedor, como ya se ha citado anteriormente.

Por ejemplo, ciertos parámetros de rendimiento de red (QoS intrínseca) pueden influenciar la percepción del usuario (QoP). La evolución en el tiempo de dicha percepción, junto a las expectativas del usuario, afianzan la experiencia (QoE) de éste y, eventualmente, pueden afectar a su intención de cambio de compañía. Ante esta situación, el proveedor, buscando garantizar su rentabilidad, reacciona haciendo mejoras en la red, desplegando nuevos servicios y realizando campañas publicitarias que atraigan de nuevo al cliente y le ofrezcan un alto grado de satisfacción.



Figura 2.4 – Relación de proyectos de NQaS con la arquitectura de capas de QoXfera

El desarrollo de esta arquitectura, al igual que la plataforma QoXcloud, se ha llevado a cabo en el seno del grupo de investigación NQaS de la UPV/EHU [13]. Se puede ver en la figura anterior, cómo los diversos proyectos de investigación que se han venido realizando en dicho grupo han sido integrados en las diferentes capas de QoXfera. Así, podemos observar la integración del sistema de medición de QoS intrínseca en escenarios reales, denominado QoS METER [17], y el laboratorio para el establecimiento de escenarios de prueba LabQoS [18] en la capa más baja de la QoXfera.

La capa de QoS Percibida la integra el observatorio para el análisis de la calidad de servicio percibida OBAVAQoS, que junto con el sistema de encuestas (ENQoS), focalizado en el análisis de la QoS Experimentada por los usuarios, conforman el núcleo principal de la QoXfera, siguiendo la *'propuesta de una metodología para la aplicación de un modelo para la gestión global de la QoS en un ISP en el marco de la UIT-T G.1000'* descrita en la tesis doctoral de Eva Ibarrola [19].

Por último, la capa de QoS de Negocio implementa el *'modelo para la gestión de la calidad de negocio (QoBiz) en los servicios de telecomunicación'* de la tesis de máster defendida por Leire Cristobo [16].

2.2. QoXCLOUD

Si a la complejidad existente, en cuanto a los diferentes enfoques que deben ser tenidos en cuenta para la definición de una gestión óptima de la calidad de servicio, le sumamos la irrupción en los últimos años de un nuevo modelo de servicios de telecomunicación en auge, como son los servicios en la nube, con requerimientos de QoS distintos a los del modelo tradicional, el análisis resultante puede ser considerablemente más complicado de abordar. Esta crisis, sin embargo, puede suponer una oportunidad en la gestión y evaluación de la QoS, ya que la propia nube puede ser utilizada como herramienta para la globalización de las medidas y los resultados de QoX, lo que supone importantes ventajas a los diferentes actores del sector.

En vista de esto, el máximo órgano regulador en el ámbito de las telecomunicaciones, la Unión Internacional de Telecomunicaciones (ITU), ha mostrado interés en el desarrollo de una plataforma en la nube (figura 2.5) para la validación de sus recomendaciones, pruebas de interoperabilidad, desarrollo de medidas relacionadas con diferentes tecnologías y otros servicios que pudieran ser de interés tanto para ellos como para los miembros que participan en sus grupos de trabajo (reguladores/NRAs, operadores/proveedores y usuarios/clientes).

La propuesta de la ITU contempla validar como “agentes reconocidos por la ITU” a los participantes de esta iniciativa, así como a los sistemas o plataformas de medida que estos agentes desarrollen, y dejar accesibles los resultados de las pruebas realizadas por estos agentes al resto de la comunidad en la nube mediante la publicación de los resultados de las pruebas en las bases de datos de la ITU, internacionalmente reconocidas.

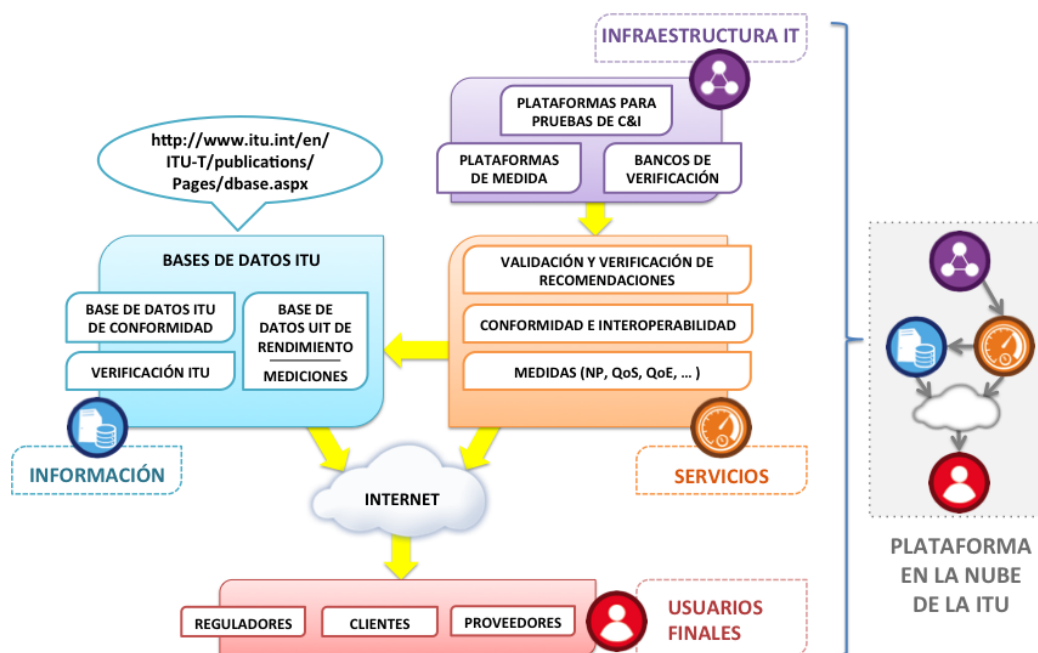


Figura 2.5 – Propuesta de plataforma en la nube de la ITU

En este interesante contexto se enmarca el proyecto QoXcloud [14], que propone el diseño de una plataforma para la medida de la calidad global experimentada por los usuarios de los servicios de telecomunicación a través de la nube que, en base a la arquitectura de referencia QoXfera, permita una gestión y evaluación eficaz de la calidad de servicio tal y como la experimentan los propios usuarios (QoX). Además, se busca alcanzar el ambicioso objetivo de ayudar a la ITU tanto en la certificación de agentes reconocidos, como en la validación de nuevas recomendaciones y servicios de interés.

La arquitectura propia de QoXcloud se ha definido teniendo en cuenta el diseño de QoXfera, con las adaptaciones necesarias para su funcionamiento en la nube. Cabe destacar el interés de que esta nueva plataforma en la nube sea compatible con la propuesta de la ITU de la figura anterior, por lo que se ha hecho un esfuerzo por mantener las entidades fundamentales de la arquitectura propuesta por la ITU de cara a una posible integración de QoXcloud en dicha plataforma (figura 2.6):

- **Entidad de Servicios:** Se basa en el diseño de QoXfera, haciendo especial hincapié en el cumplimiento de las recomendaciones y estándares en los que se asienta.
- **Infraestructura IT:** Dado que QoXfera hace uso de QoS METER [17] para su capa de QoS Intrínseca, se ha considerado oportuno aprovechar la infraestructura distribuida de la plataforma de medición de QoS existente para dar forma a esta entidad, con las debidas consideraciones para su adaptación a la nube.
- **Entidad de Información:** Se ha definido buscando su compatibilidad con las bases de datos de la ITU de cara a la validación de la plataforma para su integración.

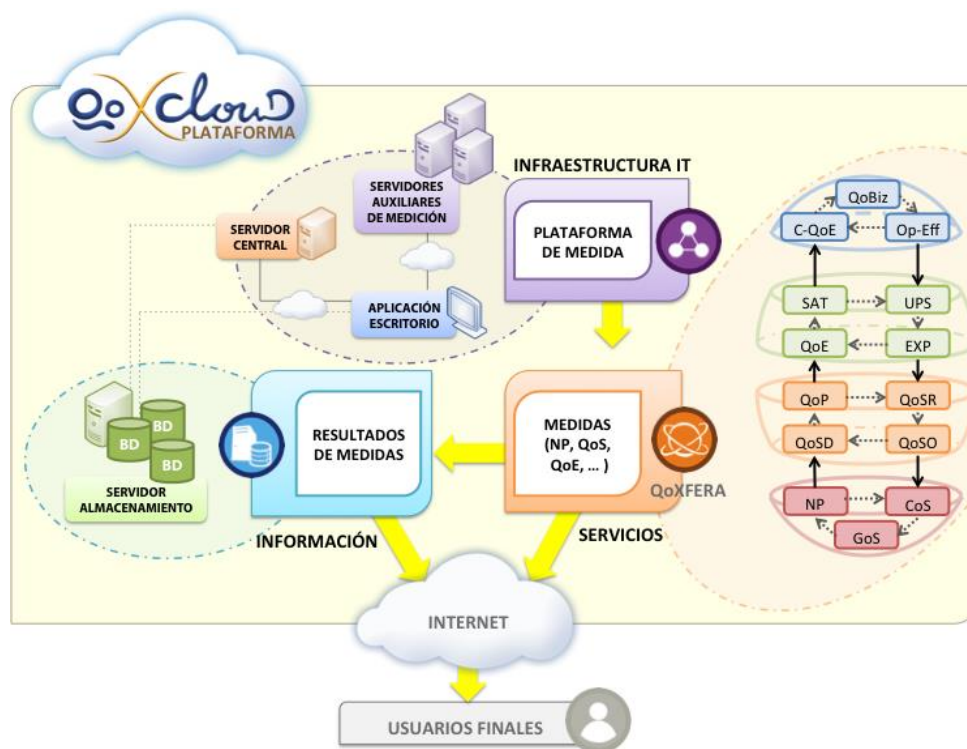


Figura 2.6 – Arquitectura de QoXcloud e integración de QoXfera

En dicho proceso de validación, se ha estimado oportuna la definición de dos prototipos que permitan demostrar la viabilidad de la plataforma QoXcloud en los objetivos para los que fue definida. En concreto, los prototipos consisten en la evaluación de la QoE en dos de los servicios de telecomunicaciones más extendidos:

- El Servicio de Acceso a Internet, tal y como se define en la Guía ETSI EG 202 057-4 [20].
- El Servicio Web (o servicio de navegación web), partiendo del *draft* P.STMWeb presentado en el grupo de trabajo SG12 de la ITU [8][9], y teniendo en consideración los factores y casos de uso para Web QoE del *draft* G.QoEWeb [10].

A efectos del presente documento, se desgranará a lo largo de las próximas páginas el estudio de la calidad experimentada (QoE) en el Servicio Web, poniendo especial atención en la metodología de validación de dicho prototipo propuesto en QoXcloud, así como en la contribución de los resultados obtenidos a los grupos de trabajo de la ITU, que han aprobado recientemente las recomendaciones ITU-T P.1501 [11] e ITU-T G.1031 [12] gracias, entre otras, a las aportaciones realizadas desde el presente trabajo.

3. OBJETIVOS

El objetivo principal de este trabajo es el de realizar un **estudio metodológico para la evaluación de la Calidad Experimentada (QoE) en el Servicio de Navegación Web, basado en los marcos de referencia establecidos por los organismos internacionales, de forma que se garantice su adecuación a la normativa.**

Puesto que el análisis de la QoE en base a la estandarización existente en la actualidad se encuentra aún en una fase temprana de definición, es también objetivo de este trabajo **contribuir a la validación de las recomendaciones ITU-T P.1501 y G.1031, que abordan los factores y metodologías en el estudio de la QoE en la navegación web [11][12], con los resultados obtenidos en el estudio metodológico previamente realizado.**

En vista de lo anteriormente señalado, se pueden desglosar los objetivos principales en una serie de objetivos parciales que especifican, de forma más detallada, la finalidad del trabajo presentado:

- **Identificación de los criterios generales de QoS a considerar en el estudio de la calidad experimentada por el usuario.** Para ello, se desarrollará un estudio de las normas y estándares existentes, así como de diferentes investigaciones del ámbito científico, con el objetivo de determinar la influencia de los diferentes indicadores y criterios en el proceso de evaluación de la calidad experimentada en el Servicio de Navegación Web.
- **Desarrollo de una plataforma para la evaluación de la QoE en el Servicio de Navegación Web.** Una vez llevado a cabo el estudio de la estandarización existente y de los trabajos relacionados en el ámbito científico, se tomará como referencia de partida el *draft* P.STMWeb propuesto en el seno del grupo de trabajo SG12 de la ITU-T, en el marco de la propuesta de una nueva recomendación para la evaluación de la QoE en el Servicio Web. Esta plataforma servirá además como prototipo para la validación del proyecto QoXcloud, en el que se engloba el presente trabajo.
- Ante la diversidad de escenarios contemplados por el *draft*, se realizará una **selección de aquellos escenarios que se consideren más adecuados de cara a la aplicación de la metodología definida en P.STMWeb y la posterior fase de validación.**
- **Definición de los perfiles de usuario y selección de la muestra de población** que se someterá al estudio para la evaluación de la QoE sobre los escenarios definidos en la plataforma desarrollada. Dicho estudio se realizará a partir de la **definición de un plan de pruebas basado en la metodología definida en P.STMWeb**, a partir del cual se obtendrán los resultados de evaluación subjetiva en función de las valoraciones vertidas por los usuarios de la muestra en el sistema de encuestas integrado en la plataforma.

- **Análisis de los resultados obtenidos** tras la finalización de las pruebas de la plataforma en función de los distintos perfiles de usuario definidos, para los escenarios y condiciones de evaluación considerados. Para ello se realizará una minuciosa lectura e interpretación de los datos obtenidos, tanto de los recogidos por la plataforma a través del sistema de encuestas integrado, como de las opiniones y valoraciones vertidas por los usuarios de la muestra al respecto de la metodología aplicada.
- **Detección de las virtudes y carencias de la metodología definida en el *draft* P.STMWeb y propuesta de mejora por medio de una contribución a través del grupo de trabajo SG12 de la ITU.** Tan importante como el estudio de la QoE en el servicio de Navegación Web, es la puesta en común de los resultados obtenidos con el resto de entidades científicas que participan en la validación del borrador. A partir de dichas conclusiones se propondrá identificar y corregir los puntos débiles de la metodología propuesta, de cara a su incorporación en la redacción definitiva de las recomendaciones ITU-T P.1501 [11] e ITU-T G.1031 [12].

Como conclusión, cabe reiterar la importancia de este último punto, ya que la contribución en entornos científicos permite no sólo avanzar en la definición de las recomendaciones, sino también realimentar el diseño de la plataforma desarrollada con las modificaciones sugeridas. Prueba de ello, es la reciente publicación de la versión final de la recomendación ITU-T P.1501, en la que se han considerado algunas de las propuestas de mejora recogidas en las conclusiones del presente trabajo.

4. ESTADO DEL ARTE

A priori, el concepto de 'Calidad de Servicio' puede parecer sencillo de definir si se considera únicamente la visión objetiva del rendimiento de la red o de un determinado servicio. Sin embargo, las dificultades económicas en el sector de las telecomunicaciones en los últimos años, junto con las necesidades de comunicación global, que cada vez más demandan los usuarios, han propiciado un nuevo escenario en el que la satisfacción y la experiencia del cliente son determinantes en cuanto a lo que una gestión óptima de QoS debería abordar.

Muchos de los servicios que hasta hace poco eran ofrecidos por operadores, están siendo ahora ofrecidos por otros proveedores, abriendo nuevos modelos de negocio y aumentando la competitividad del sector. Esto hace que operadores y proveedores intenten reducir el coste de provisión de un determinado servicio sin comprometer la calidad del producto final, de cara a atraer la mayor cantidad de clientes con el máximo beneficio. Por su parte, el cliente, consciente de la diversidad de oferta existente, se muestra mucho más cauto a la hora de contratar un nuevo servicio y analiza con detenimiento si el gasto que va a realizar es acorde a sus expectativas, en función de su propia experiencia.

Por tanto, si se pretende abordar el concepto de QoS en la actualidad, no podemos limitarnos únicamente a los indicadores de rendimiento objetivo de la red (NP, Network Performance), sino que debemos considerar el enfoque subjetivo de la percepción y experiencia del cliente, así como de la calidad de negocio de operadores y proveedores.

Dado que el objetivo del presente trabajo es el de realizar una evaluación de la calidad experimentada (QoE) de los usuarios del servicio de navegación web, es fundamental realizar un estudio pormenorizado de la QoS en todas sus vertientes, tanto en el ámbito científico como en el de los organismos de estandarización, con la voluntad de identificar aquellos indicadores que permitan definir las condiciones de evaluación de QoE más adecuadas para los escenarios que se proponen en el presente trabajo.

Para ello, se ha estructurado el capítulo del estado del arte en cuatro apartados: El primero, sirve de introducción al concepto de QoS en el ámbito de las telecomunicaciones. Posteriormente, en el segundo apartado, se analiza en el propio concepto de QoE en base a la estandarización vigente y al establecimiento de relaciones de este concepto con el resto de vertientes de la QoS. El tercer apartado, por su parte, aborda modelos y metodologías propuestos tanto por los organismos de estandarización como en otras publicaciones del ámbito científico. Y el cuarto, finalmente, se centra en investigaciones directamente relacionadas con el Servicio de Navegación Web.

Por último, se incluye un apartado adicional en el que se establecen las conclusiones extraídas del estudio desarrollado en este capítulo.

4.1. QoS en el ámbito de las Telecomunicaciones

El término 'Calidad de Servicio' en el ámbito de las telecomunicaciones ha ido evolucionando a lo largo de los años desde su definición inicial. Ante la aparición de nuevos modelos de servicio en un mercado de enorme competencia, se ha pasado de la visión objetiva de QoS como sinónimo de la calidad de funcionamiento de la red (NP), a una visión más global, orientada a la satisfacción del cliente.

ITU - Unión Internacional de Telecomunicaciones

La ITU recoge en la recomendación ITU-T E.800 una de las primeras definiciones de QoS que ya en 1994 [21] apuntaba hacia esa globalidad, aunque centrándose principalmente en el funcionamiento del servicio:

Calidad de Servicio:

El efecto global de la calidad de funcionamiento de un servicio que determina el grado de satisfacción de un usuario de un servicio.

ITU-T E.800 (1994)

Esta versión de la recomendación se asienta sobre otras recomendaciones como la ITU-T I.350 [22] o el informe ETSI ETR 003 [23], que sólo contemplan aspectos del funcionamiento como factores a tener en cuenta en la evaluación de la QoS. Sin embargo, numerosos estudios han avalado la necesidad de ir más allá de las medidas de la calidad de funcionamiento de red y contemplar un marco más general en el que considerar otros aspectos que influyen la calidad de servicio [24][25][26].

Estos estudios analizan principalmente la importancia de la percepción de usuario y la calidad de servicio que éste experimenta y entienden que el análisis de la QoS no puede reducirse al estudio independiente de estos conceptos. Además, también alertan de los riesgos derivados de centrar la QoS únicamente en la satisfacción del usuario [27].

A raíz de estos y otros estudios, la ITU-T actualizó la recomendación E.800 en 2008 [28], recogiendo una nueva definición del concepto de calidad de servicio más orientado a la búsqueda de la satisfacción y cumplimiento de los requerimientos de los clientes. Además, en ese mismo texto se incluyó la diferenciación entre calidad de funcionamiento de red y calidad de servicio, quedando establecida la nueva definición de la siguiente manera:

Calidad de Servicio:

La totalidad de las características de un servicio de telecomunicaciones que determinan su capacidad de satisfacer las necesidades explícitas e implícitas del usuario del servicio.

ITU-T E.800 (2008)

ISO – Organización Internacional de Normalización

Otros organismos de normalización han realizado también diversas definiciones del concepto de calidad. En concreto, la familia de Normas ISO 9000, relacionada con sistemas de gestión y aseguramiento de la calidad, recoge en su versión de 2005 [29], la siguiente definición de calidad:

“El grado en el que un conjunto de características inherentes cumple con los requisitos”, entendiéndose por requisito “necesidad o expectativa establecida, generalmente implícita u obligatoria”, pudiendo éstos “ser generados por las diferentes partes interesadas”.

Norma ISO 9000 (2005)

Se pueden observar en la redacción de esta definición ciertos paralelismos con la revisión del concepto de QoS abordado en la recomendación ITU-T E.800, así como la inclusión de términos globales, no dependientes únicamente de características objetivas sino también abiertos a la subjetividad de las expectativas de las partes interesadas.

ETSI – Instituto Europeo de Normas de Telecomunicaciones

Por su parte, la ETSI, en su especificación técnica ETSI TS 102 250-1 de abril de 2011 [30], parte de la recomendación ITU-T E.800 para destacar el carácter extremo a extremo de la QoS, remarcando la idea de que el término ‘Calidad de Servicio’ no sólo debe abordar los aspectos técnicos sino también aquellos no técnicos que afecten al servicio. Esta puntualización implica la inclusión de una visión más global que aporta la evaluación de la calidad experimentada por los usuarios y clientes (figura 4.1).

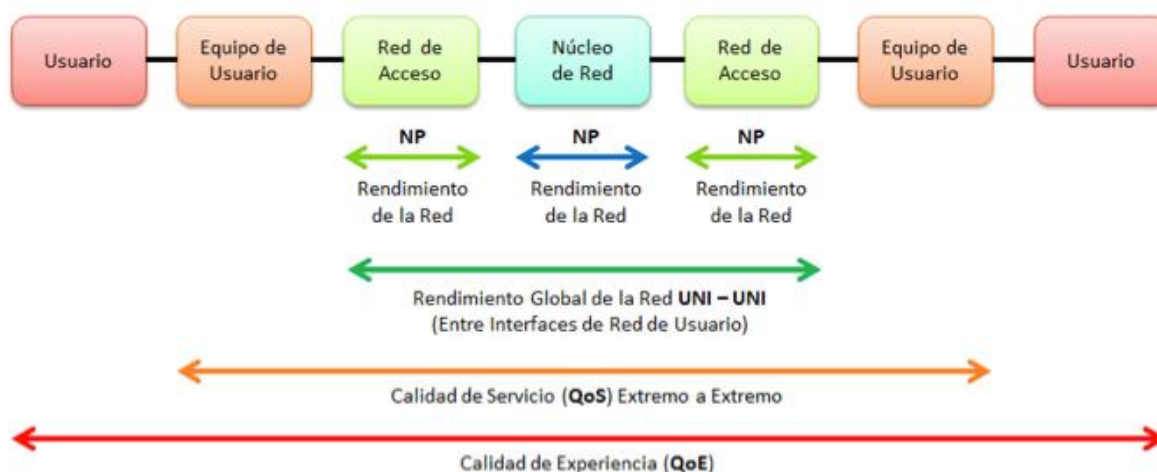


Figura 4.1 – Enfoque de la QoS orientada al usuario de la ETSI TS 102 250-1

4.2. QoE: Definición y modelos

Definición de QoE

La recomendación ITU-T P.10/G.100, en su enmienda 2 de 2008 [2] recoge la definición más comúnmente aceptada del concepto de 'Calidad de Experiencia':

Calidad de Experiencia:

La aceptabilidad general de una aplicación o servicio, tal y como es percibida subjetivamente por el usuario final.

ITU-T P.10/G.100 Amendment 2 (07/08)

En dicha definición puede apreciarse claramente una relación entre la percepción del usuario y la aceptación de un determinado servicio. Aun siendo relativamente clara esa relación, esta misma recomendación realiza un par de puntualizaciones adicionales al respecto del alcance de aplicación del concepto de QoE e insiste en la variabilidad en la aceptabilidad de un servicio por factores subjetivos ajenos al propio rendimiento objetivo de dicho servicio:

NOTA 1:

La calidad de experiencia incluye los efectos completos del sistema extremo a extremo (cliente, terminales, redes, infraestructura del servicio, etc.).

NOTA 2:

La aceptabilidad general puede verse influida por las expectativas del usuario y el contexto.

ITU-T P.10/G.100 Amendment 2 (07/08)

Estas consideraciones ya hemos podido anticiparlas en la representación gráfica del alcance de los conceptos de QoS y QoE (figura 4.1) que se describen en la especificación técnica ETSI TS 102 250-1 [30], citada en apartado anterior. Sin embargo, la diferenciación de ambos conceptos no es sencilla de explicar. La propia especificación de la ETSI da una explicación jerárquica de los conceptos, refiriéndose a la QoE de la siguiente manera:

La inclusión del propio usuario en la calidad global de las telecomunicaciones, extiende el concepto objetivo de Calidad de Servicio al altamente subjetivo de Calidad de Experiencia. La QoE varía de usuario a usuario ya que está influenciada por las experiencias personales y expectativas de cada individuo.

ETSI TS 102 250-1, V2.2.1 (2011)

Asimismo, dedica un apartado completo a detallar la relación entre QoS y QoE, del que se destaca la imagen de la figura 4.2, en la que se pueden observar la tipología de aspectos que contribuyen a definir tanto la QoS (objetivos), como la QoE (subjetivos).

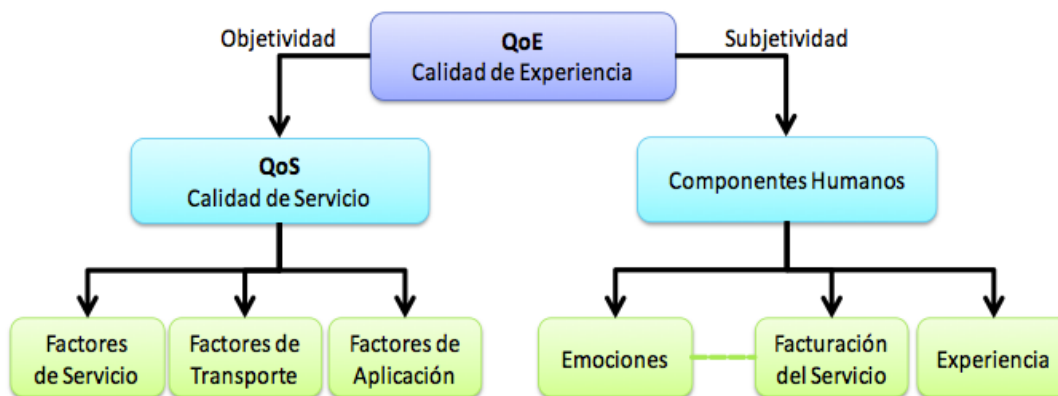


Figura 4.2 – Aspectos de la QoE (ETSI TS 102 250-1)

En ese mismo capítulo, la especificación de la ETSI introduce también la utilización de las escalas de opinión MOS (Mean Opinion Score) como mecanismo de evaluación de QoE, según se definen en la recomendación ITU-T E.800 [28]. En general, se puede definir una correlación entre la QoE subjetiva evaluada a través del MOS y ciertos parámetros objetivos de QoS. Habiéndose establecido dicha relación QoE/QoS para los parámetros de un determinado servicio, se pueden utilizar las escalas MOS de dos maneras:

- Dada una medición de QoS, se puede predecir la QoE esperada por un usuario
- Dado el objetivo de QoE de un usuario, se puede deducir el rendimiento de red requerido

Por último, además de las definiciones de los organismos de estandarización, existe otro acercamiento a la QoE promovido por Qualinet [31], un consorcio de investigación financiado por el Programa Internacional de Cooperación Europea en el Campo de la Investigación Científica y Técnica (COST) [32], cuyo objetivo es el de crear una red de investigación multidisciplinar sobre la QoE.

En su libro blanco de marzo de 2013 [33], se hace referencia a que el alcance de la QoE no está sólo limitado al uso de un sistema o servicio, sino que también depende de su contenido. Se parte del concepto de '*Experiencia*' como un flujo individual de percepción e interpretación de uno o múltiples eventos, sin ser necesariamente el resultado del juicio o evaluación de la calidad. Para Qualinet, el proceso de formación de la calidad en el individuo incluye dos caminos (figura 4.3):

- **El camino de la referencia**, que refleja la naturaleza contextual y temporal de la calidad, al partir de recuerdos y conocimientos previos que influyen en el proceso.
- **El camino de la percepción**, que parte de un evento físico originado en los órganos sensoriales del individuo, hasta que es interpretado por sus procesos cognitivos.

En vista de estos caminos, se puede determinar que la *calidad experimentada* está delimitada en el tiempo, en el espacio y el carácter. El sujeto puede formarse un concepto interno de calidad que, en este punto, se puede obtener y potencialmente cuantificar a partir de la limitación descriptiva del propio usuario.

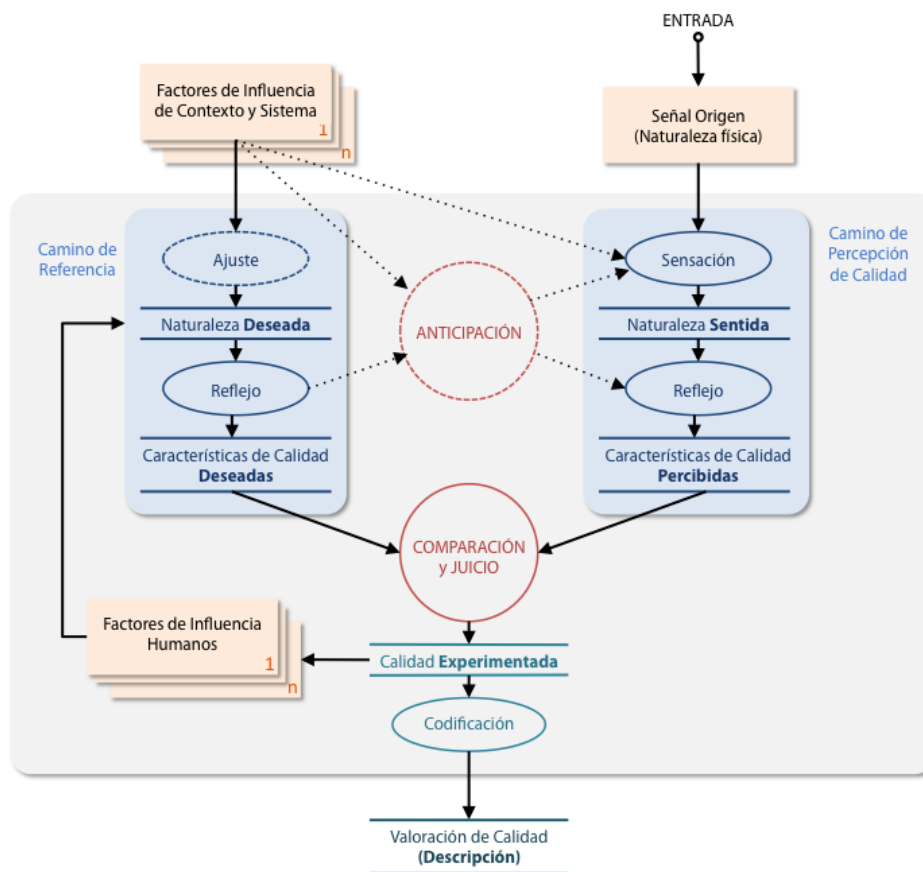


Figura 4.3 – Proceso de formación de la Calidad (Qualinet)

En vista de lo anterior, desde Qualinet [31] se propone la siguiente definición de QoE [33]...

Calidad de Experiencia:

Grado de placer o disgusto de un usuario de una aplicación o servicio. Es el resultado del cumplimiento de sus expectativas con respecto a la utilidad y/o el disfrute de la aplicación o servicio en función de la personalidad del usuario y su estado actual.

Qualinet White Paper, Marzo 2013

... donde el concepto de 'personalidad' se refiere a las 'características de la persona que dan cuenta de patrones consistentes de sentimientos, pensamientos y comportamientos' [34].

Comparando esta definición con la de la recomendación ITU-T P.10/G.100 Enmienda 2 [2], que definía la QoE en función de la aceptabilidad general de una aplicación o servicio, se observa una diferenciación respecto a este concepto.

Desde la visión de la ITU, se puede entender la aceptabilidad como la característica de un servicio que describe el grado de facilidad con la que la persona está dispuesta a utilizar dicho servicio. Sin embargo, Qualinet define la aceptabilidad como el resultado de una decisión basada parcialmente en la QoE. En otras palabras, el grado de placer o disgusto con un servicio, junto con otra serie de factores, llevan a la aceptabilidad del mismo; al contrario de lo definido por la ITU, en el que la aceptabilidad del servicio es la que define la QoE.

En cualquier caso, lo que sí es evidente en ambas definiciones es que la QoE no es un factor que dependa únicamente de la percepción del usuario respecto a la QoS intrínseca, sino que se deja influenciar por el contenido del servicio, las expectativas y objetivos del usuario, el contexto de uso, así como multitud de factores humanos, tanto físicos como cognitivos.

Modelado de QoE sobre la base del Marco ITU-T G.1000

Para abordar la relación de la QoE con otros aspectos de la calidad de servicio, es necesario detenernos previamente en la recomendación ITU-T G.1000 [1]. Esta recomendación parte de la definición de QoS de la recomendación ITU-T E.800, con la intención de definir un marco de referencia en el que se relacionen las diferentes vertientes de la QoS.

Gracias a la definición de este marco, se permite abordar de manera global la gestión de la QoS al cuantificar el problema desde diversos puntos de vista: el del cliente (mediante encuestas y pruebas subjetivas) y el del proveedor de servicio (por medio de mediciones de red). Solucionar un aspecto del problema en un lado (por ejemplo, en el proveedor) implica también la resolución en el otro (el lado del cliente).

En concreto, las cuatro vertientes de la QoS que se relacionan en la [figura 4.4](#) son:

- **Calidad de servicio ofrecida por el proveedor (QoSO – QoS Offered)**
Es la declaración del nivel de calidad que se espera que el proveedor de un servicio ofrezca al cliente.
- **Calidad de servicio entregada por el proveedor (QoSD – QoS Delivered)**
Es la declaración del nivel de calidad real conseguido y entregado al cliente.
- **Calidad de servicio percibida por el usuario (QoP – QoS Perceived)**
Es la declaración del nivel de calidad que el cliente cree haber experimentado.
- **Requisitos de calidad de servicio del cliente (QoSR – QoS Required)**
Es la declaración, en lenguaje corriente, del nivel de calidad requerido por las aplicaciones del cliente de un servicio.

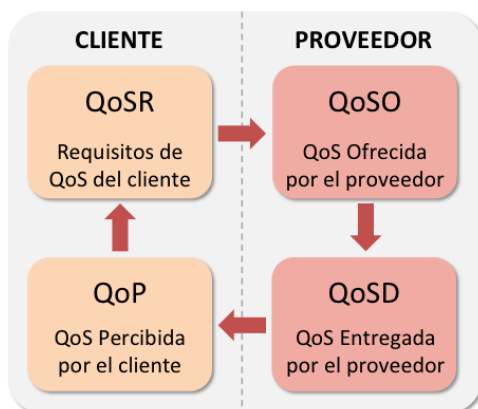


Figura 4.4 – Las cuatro vertientes de la QoS (ITU-T G.1000)

A raíz de la publicación de la recomendación ITU-T G.1000 en 2001, han sido numerosos los trabajos que han propuesto ir un paso más allá con la intención de extender el modelo de vertientes de la QoS, para englobar otros conceptos como la propia QoE o la Calidad de Negocio (QoBiz). Esta visión de calidad global, QoX, se describe perfectamente en el trabajo de Stankiewicz, "QoX: What is it really?" [4], del que se destaca el modelo general mostrado en la figura 4.5, en el que se relacionan las definiciones de calidad recogidas en los estándares de la ITU-T a los que se han venido haciendo referencia:

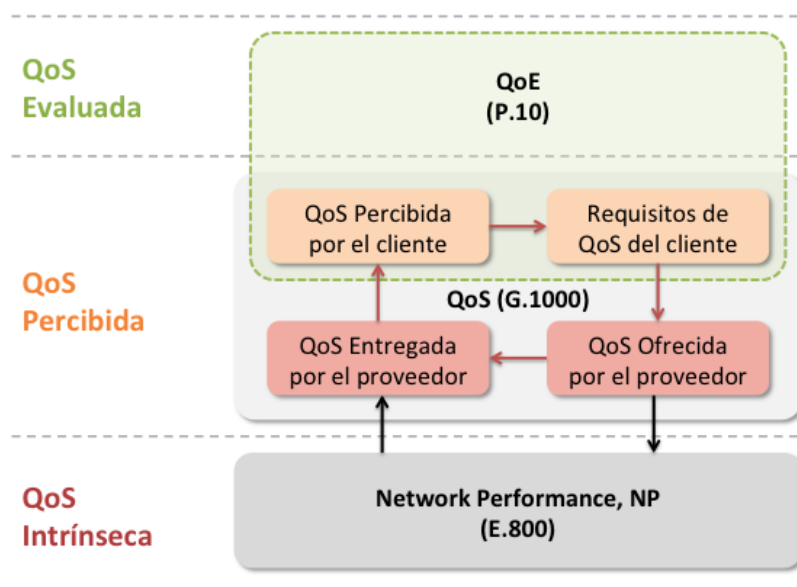


Figura 4.5 – Relación de estándares de la ITU-T en el modelo de QoX (Stankiewicz et al)

Al respecto de la QoE, se puede observar que no tiene una ubicación definida en los niveles del modelo de QoX que presenta Stankiewicz. En primera instancia, el término QoE parece mejor ubicado en la capa de QoS Percibida, superponiéndose con los conceptos de QoS subjetiva que define la ITU-T. Sin embargo, el autor lo ubica a caballo entre la capa de QoS Evaluada (*Assessed QoS*) y la capa de QoS Percibida. El motivo es que la QoE no sólo cubre la parte subjetiva del modelo definido por la ITU-T G.1000 sino que va más allá, ya que engloba la satisfacción del usuario y su atracción al servicio, así como la influencia de diversos factores, como la opinión de otros usuarios o, incluso, la política de precios del proveedor.

Modelos relacionales QoE-QoBiz

Entre los trabajos en los que plantea un modelo global, destaca Moorsel [5], uno de los primeros autores en establecer relaciones entre QoS, QoE y QoBiz. El trabajo de Moorsel coincide en el tiempo con la publicación de la recomendación ITU-T G.1000 (ambos son de 2001). Sin embargo, el concepto QoS difiere de las vertientes del marco propuesto en la recomendación, al asignarle únicamente el significado de calidad de funcionamiento intrínseco del servicio. A pesar de ello, sí asigna un significado similar al concepto de QoE de la ITU, relacionándolo con la calidad percibida y otros factores propios del usuario como consumidor. En las siguientes figuras podemos observar las métricas que asigna Moorsel a los conceptos QoS, QoE y QoBiz, así como la relación que define entre ellos.

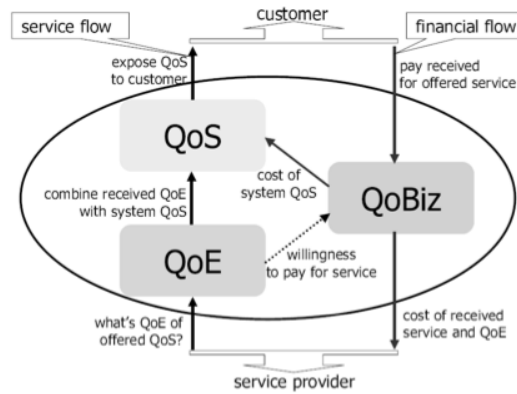


Figura 4.6 – Relación entre QoS, QoE y QoBiz (Moorsel)

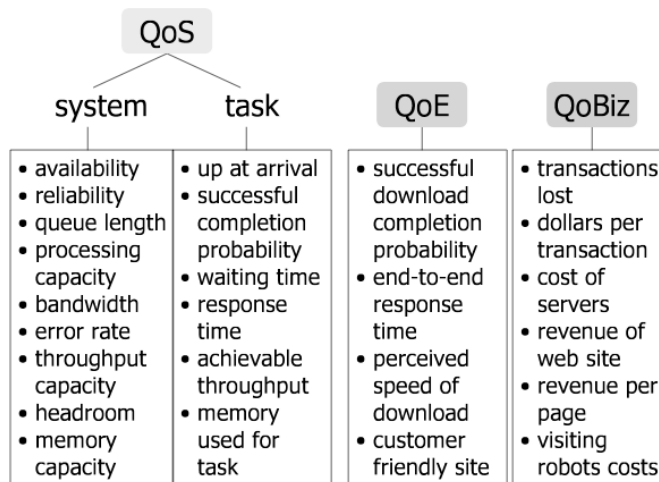


Figura 4.7 – Métricas de los conceptos QoS, QoE y QoBiz (Moorsel)

Por su parte, Kalevi Kilkki, en su artículo *"Quality of Experience in Communications Ecosystem"* [6], presenta un marco para el análisis de la QoS similar al de Moorsel (figura 4.8), aunque más cercano al de la recomendación ITU-T G.1000.

Kilkki divide el concepto de persona en usuario y cliente, tratando estos conceptos por separado. Por un lado, establece relación entre el cliente y el proveedor de servicios, siendo el primero el que contrata un producto al segundo.

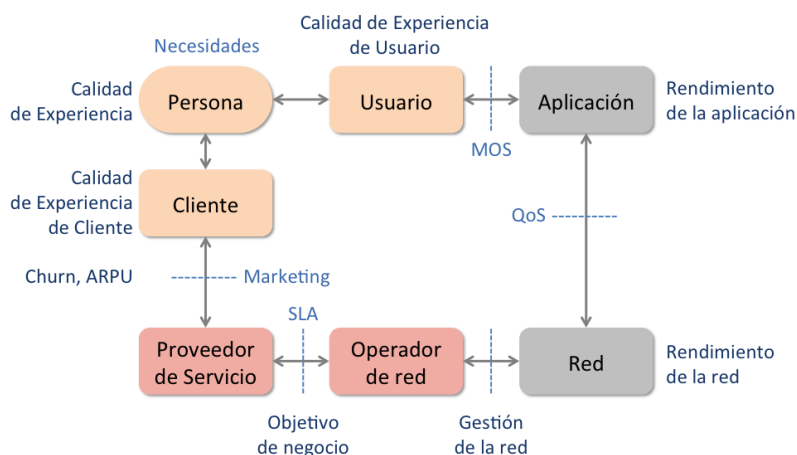


Figura 4.8 – Marco para la gestión de la QoS y detalle de relaciones (Kilkki)

El *usuario* puede ser o no la misma persona que contrató el servicio (el *cliente*) y, por tanto, compartir o no los mismos requerimientos y expectativas. Este motivo es el que origina la distinción de roles que se plasma en la [figura 4.8](#) y que se asemeja al marco de la recomendación ITU-T G.1000 en que ambos intentan vincular y hacer converger los puntos de vista de usuario/cliente y proveedor.

Esta distinción implica también la definición de la QoE a partir de tres conceptos:

- *QoE – Quality of Experience:* Calidad Experimentada
- *QoUE – Quality of User Experience:* QoE del usuario
- *QoCE – Quality of Client Experience:* QoE del cliente

La relación entre la aplicación y el usuario se establece a través del MOS, como ya vimos anteriormente en las directrices de la especificación técnica ETSI TS 102 250-1 [\[30\]](#). Por su parte, operadores y proveedores establecen su relación por medio de Acuerdos de Nivel de Servicio (SLA – Service Level Agreement). Este tipo de acuerdos es una herramienta útil para la formalización de las responsabilidades que se obtiene mediante negociación entre ambas partes, con el objetivo de alcanzar entendimiento común sobre el servicio prestado, su calidad, responsabilidades, prioridades, etc.

La recomendación ITU-T E.860 [\[35\]](#) establece un marco general para el establecimiento de acuerdos de nivel de servicio independiente del tipo de servicio y tecnología utilizados. Aunque no es objetivo directo de este trabajo analizar los SLA y las metodologías existentes para su especificación, se incluye a continuación la definición del concepto, dado su interés:

Acuerdo de Nivel de Servicio (SLA):

Es un acuerdo formal entre dos o más entidades que se alcanza después de un período de negociación con el fin de establecer las características del servicio, las responsabilidades y las prioridades de todas las partes.

ITU-T E.860 (2002)

Modelos de QoE en el ecosistema de comunicaciones

Partiendo del marco definido por Kilkki [\[6\]](#), otros autores han hecho un esfuerzo por definir un modelo que integre los principales aspectos relacionados con la QoE. En concreto, la propuesta de Laghari [\[36\]](#) define un ecosistema de comunicaciones que va más allá del modelo básico de QoE, al incluir conceptos relativos al factor humano y a las relaciones entre los dominios principales: Persona, Negocio y Tecnología.

Conectado con estos tres dominios, Laghari [\[36\]](#) incluye un cuarto, ausente en el modelo de Kilkki [\[6\]](#): el **contexto** ([figura 4.9](#)). Manteniendo el esquema de colores para las figuras de ambos modelos, podemos observar cómo la inclusión del concepto de contexto permite introducir un mecanismo de evaluación iterativa de la calidad derivado de la interacción entre los cuatro dominios.

Según Laghari [36], la QoE engloba todas las experiencias y necesidades de calidad objetivas y subjetivas de la persona en su interacción con la tecnología y las entidades de negocio en un contexto concreto. Entendiendo el comportamiento humano a partir de factores externos e internos a la persona, Laghari asigna dichos factores a los distintos dominios del ecosistema.

Así, los factores internos (factores biológicos, psicológicos y cognitivos) quedan enmarcados dentro del dominio relativo a la persona, mientras que los factores externos que afectan a su comportamiento, se dividen entre el resto de dominios. En concreto, el dominio tecnológico da cuenta de los factores de calidad intrínseca del servicio, el dominio de negocio recoge las características relativas a la imagen de la compañía y al precio de sus productos, y el dominio contextual engloba factores sociales, así como otra serie de variables dependientes del tiempo, del espacio e incluso de las condiciones climáticas.



Figura 4.9 – Ecosistema de comunicaciones (Laghari)

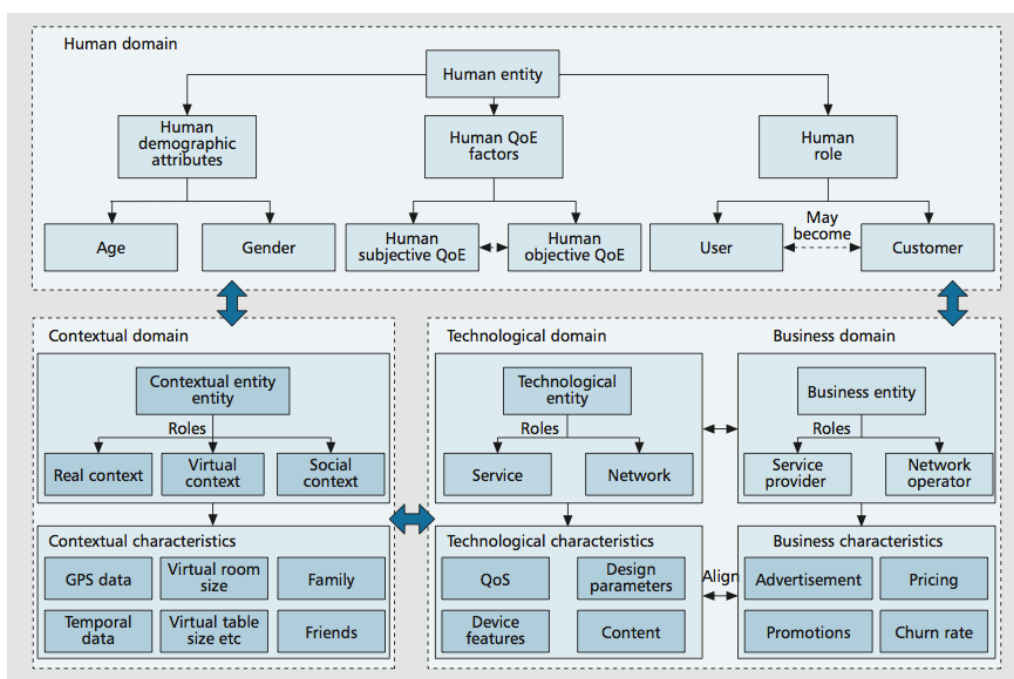


Figura 4.10 – Modelo de interacción de QoE en el ecosistema de comunicaciones (Laghari)

La variabilidad de factores que afecta directamente a la persona produce el citado proceso iterativo de evaluación de la calidad, permitiendo identificar los requisitos de QoE cuyo carácter se ha ido consolidando.

Este modelo, junto con el de Killki [6], es a su vez la base del ecosistema definido por Qualinet para su modelo de QoE [33], representado en la figura a continuación:

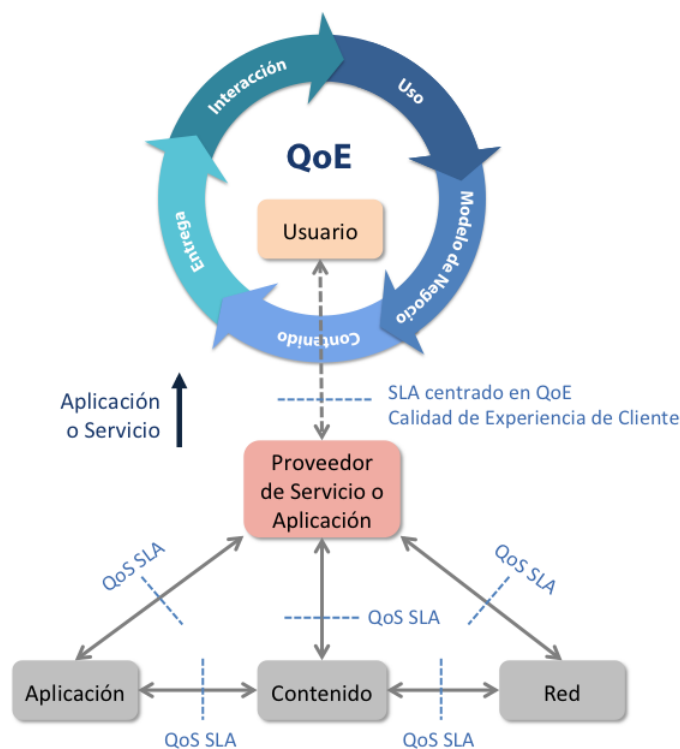


Figura 4.11 – QoE en el ecosistema para un proveedor de aplicaciones/servicios (Qualinet)

En esta nueva representación, los dominios definidos por Laghari [36] se corresponden casi directamente con los roles del ecosistema de Qualinet, representados en el diagrama circular de la parte superior de la figura 4.11: Tecnología (Entrega e Interacción), Mercado y Finanzas (Modelos de Negocio), Usuario (Uso) y Creativo (Contenido).

El contexto, en esta ocasión, no aparece directamente representado, aunque es considerado como factor de influencia del entorno del usuario en términos de características físicas, temporales, sociales y económicas, así como relativas a la tecnología. Podría ubicarse, por tanto, en un plano paralelo interrelacionado con todas las entidades y roles del modelo, pudiendo incidir sobre la calidad experimentada con distintos niveles de magnitud (micro o macro) y dinamismo (estático o dinámico), y en distintos patrones de ocurrencia, ya sea de manera aislada o combinada.

Una de las principales contribuciones al valor de QoE viene desde el contexto de la aplicación, por lo que es recomendable definir claramente el dominio de la aplicación objetivo en el estudio de la QoE. Generalmente, los dominios de aplicación pueden ser tanto unidireccionales como bidireccionales, y adoptar distintas modalidades de contenidos.

En concreto, en el libro blanco de Qualinet [33], se distinguen estos tres casos:

- Aplicaciones de transmisión (*streaming, broadcast,...*) de audio, vídeo, documentos...
- Aplicaciones conversacionales y colaborativas incluyendo aspectos artísticos y sociales
- Aplicaciones médicas, educacionales,...

A partir del contexto de aplicación se puede realizar una estimación en tiempo real del valor de QoE de cara a la toma de decisiones, como por ejemplo en la determinación del uso más adecuado de los recursos (tanto de infraestructura como de contenidos) [37]. Así, la modelización de la experiencia del usuario, junto a las métricas de rendimiento intrínseco, permiten simplificar el proceso de planificación de una aplicación.

Esto se representa en la figura 4.11, a través de las distintas interfaces en las que se definen acuerdos de nivel de servicio (SLA): Un SLA se puede definir en términos de QoE en las relaciones existentes entre proveedor y cliente, mientras que los acuerdos del proveedor del servicio con los proveedores de componentes (aplicación, contenido, red) se definen en términos de calidad intrínseca.

En cuanto a las áreas de aplicación de acuerdos definidos en términos de QoE, se identifican las cuatro siguientes:

- Aprendizaje multimedia
- Comunicación háptica
- Experiencia sensorial
- Servicios web y en la nube

Todas ellas tienen un componente centrado en el usuario de gran importancia, entre las que destaca la última, dado que es la más cercana al estudio del presente trabajo. En el texto de Qualinet [33], en clara referencia al trabajo de Hoßfeld [38] sobre la cuantificación de la QoE de *YouTube* a través de *crowdsourcing*¹, se analiza la tendencia creciente hacia las aplicaciones basadas en Internet en general y en la nube en particular.

Asimismo, detectan una falta de adaptación de los modelos de calidad existentes a este nuevo escenario en el que una cooperación flexible entre las partes involucradas puede permitir la compartición de contenidos, o la interacción y colaboración entre usuarios de manera transparente al mismo tiempo que se maximiza la QoE.

En una línea similar se mueve el proyecto QoXcloud [14] sobre el que se asienta el presente trabajo, en su intención de desarrollar una plataforma para la medida de la calidad global experimentada por los usuarios de los servicios de telecomunicación a través de la nube que, en base a la arquitectura de referencia QoXfera, permita una gestión y evaluación eficaz de la calidad de servicio tal y como la experimentan los propios usuarios (QoX).

¹ **Crowdsourcing:** Mecanismo consistente en la colaboración abierta distribuida. Los participantes en el proceso de cuantificación de QoE de YouTube se encontraban geográficamente dispersos durante los experimentos.

QoE en el modelo global: De CSAT a QoXFERA

A diferencia de los ecosistemas definidos en páginas anteriores, la arquitectura de QoXfera [7] no se centra fundamentalmente en la definición de la QoE y a sus relaciones, sino que pretende enmarcarla dentro de un modelo global, más amplio, considerando todos los matices subjetivos de la percepción, experiencia y satisfacción del usuario/cliente sobre el marco de referencia de la ITU-T G.1000 (figura 4.4), lo que nos devuelve a la definición de Calidad de Experiencia original de la ITU-T P.10/G.100 [2].

Para entender esta arquitectura, nos remontamos al modelo global de QoX sobre el marco de referencia ITU-T G.1000 presentado por Stankiewicz [4]. Pese a no definir una ubicación para el propio concepto de QoE dentro del modelo, este trabajo, junto al modelo de satisfacción CSAT de Xiao [39] (figura 4.12), supone el punto de partida de la propuesta de modelo de gestión global de la QoS de la doctora Ibarrola [19]. Este modelo de gestión constituye el núcleo principal del modelo QoXfera.

La figura 4.13 muestra el modelo de gestión global de QoS en un ISP (Internet Service Provider) desarrollado en la tesis doctoral de Ibarrola, el cual se divide fundamentalmente en tres grandes bloques:

- Un modelo de QoS que, a partir de los requerimientos del cliente, permite identificar los indicadores de calidad y desempeño con los que evaluar cuantitativamente la QoS del servicio o función sobre la que el cliente tiene puestas unas expectativas. Estos indicadores son de gran interés para el proveedor, como *input* de información.
- El marco de referencia de la ITU-T G.1000, como núcleo del modelo de gestión global.
- Y un modelo de satisfacción de usuario que permite actualizar las expectativas del cliente, así como su voluntad de continuar o no con el servicio contratado, lo que supone la realimentación del sistema de vuelta al modelo de QoS, con la actualización de los requerimientos del cliente.

En concreto, como modelo de QoS, se hace uso de una adaptación del modelo propuesto por Liberal [40] en el que, a partir de una representación matricial (figura 4.14, derecha), se identifican y relacionan los criterios de QoS que pueden ser de interés para los usuarios, con las expectativas de los clientes de un determinado servicio.

Para cada uno de los aspectos funcionales que caracterizan dichos criterios (QoS de la red, disponibilidad, asistencia al cliente y contenido), se identifican los indicadores de calidad (KQI – Key Quality Indicator) que pueden influir en la percepción del usuario. A su vez, estos KQI se relacionan con los indicadores de desempeño (KPI – Key Performance Indicator) que consisten en uno o más parámetros concretos, cuya medida puede o no ser inmediata. En este segundo caso, es necesario definir la relación de dependencia de dichos parámetros, así como los métodos de medida que permitan obtener una cuantificación de su valor.



Figura 4.12 – Modelo de Satisfacción CSAT (Xiao)

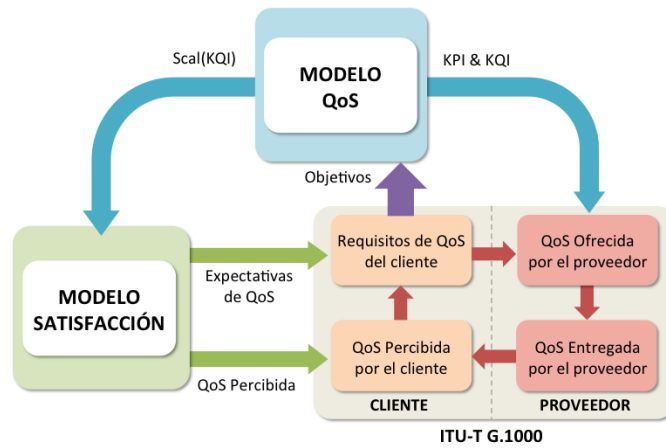


Figura 4.13 – Modelo de gestión global de QoS en un ISP (Ibarrola)

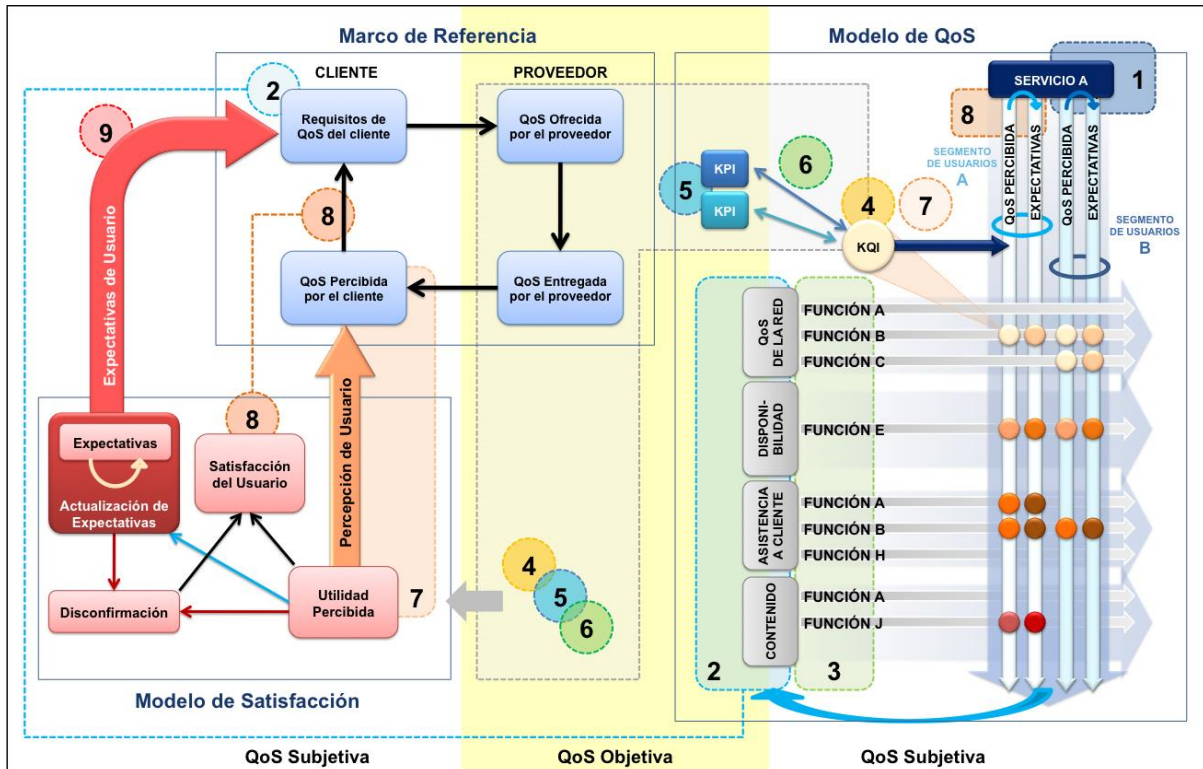


Figura 4.14 – Metodología del modelo de gestión global de QoS en un ISP (Ibarrola)

En resumen, los indicadores de calidad KQI del modelo de QoS constituyen los ‘puntos de cruce’ de la matriz y sirven al proveedor para establecer la base del acuerdo en cuanto a niveles de ‘QoS Ofrecida’ en el marco de referencia G.1000. De igual manera, el cliente puede utilizar dichos indicadores como mecanismo de evaluación de su satisfacción.

En cuanto al modelo de satisfacción, Ibarrola basa su contenido en el CSAT de Xiao [39] (figura 4.12), que toma como referencia el paradigma de la disconfirmación introducido por Anderson y Sullivan [41]. De acuerdo con este paradigma, el mecanismo de evaluación de la satisfacción de un cliente es el resultado de la diferencia o separación existente entre sus expectativas y la percepción del servicio recibido.

Tal y como expone Schindlholzer en su artículo *“The confirmation/disconfirmation paradigm: Why satisfied customers are not always satisfied”* [42], es necesario comprender la satisfacción del usuario de cara a establecer mecanismos que permitan evaluarla. Puede parecer una afirmación trivial y sin embargo no lo es tanto, si consideramos que un usuario que está percibiendo la calidad de un servicio ‘justo como esperaba’ puede mostrar satisfacción... o simplemente un sentimiento neutral puesto que todo ‘funciona como debe’ (figura 4.15).

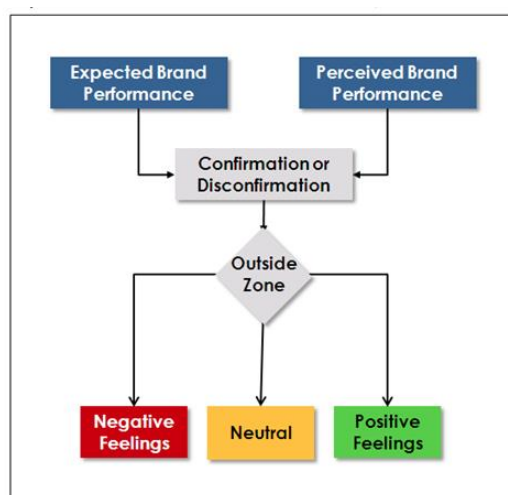


Figura 4.15 – Satisfacción/Desatisfacción según la experiencia del cliente (Schindlholzer)

Si la diferencia entre expectativas y percepción es desfavorable o muy desfavorable al cliente, lo más probable es que éste cambie de proveedor de servicio y se lleve consigo un aprendizaje que podría influir en la siguiente contratación. En cambio, si la percepción de la QoS supera con creces las expectativas del cliente, lo normal es que éste mantenga el contrato del servicio, aunque puede influenciar una revisión de las expectativas al alza del cliente, lo que en futuras iteraciones pase de sentimientos positivos a neutrales.

La exposición en el tiempo a la percepción de la calidad de un servicio y la revisión de expectativas que presenta este paradigma, es la esencia del concepto de experiencia (QoE), y es utilizada en el modelo CSAT para modelar la evolución de la satisfacción del usuario con el servicio, pudiendo observar desalineamientos entre la QoS percibida y la QoS requerida, que supondrán la información de entrada a una nueva iteración en la mejora de la QoS.

En esencia, la metodología del modelo de gestión global de la QoS propuesta por Ibarrola [19] puede resumirse en los siguientes pasos, que ya se han ido anticipando en párrafos anteriores, y que se corresponden con la numeración mostrada en la figura 4.14:

- 1 – *Identificación del escenario bajo análisis.* En esta fase se identifica el servicio (o servicios) a considerar en el proceso de gestión de la QoS, la topología del servicio, los segmentos de usuarios (ya que cada segmento puede tener consideraciones distintas de QoS) y las transacciones de usuario que se requieren para la adecuada provisión del servicio.
- 2 – *Identificación de los criterios de QoS de interés para los usuarios* para cada servicio y segmento. En concreto, se clasifican en torno a cuatro categorías: QoS de la red, disponibilidad, atención al cliente y calidad de los contenidos.
- 3 – *Identificación de los aspectos funcionales que caracterizan cada criterio de QoS.*
- 4 – *Identificación de indicadores de calidad (KQI).* A partir de los criterios de QoS identificados de interés para los usuarios, se definen los indicadores de calidad ('puntos de cruce' de la matriz) que pueden afectar o tener relevancia sobre el aspecto funcional de las transacciones de usuario y que, por tanto, puede influir en la QoS percibida.
- 5 – *Identificación de indicadores de desempeño (KPI).* Estos indicadores permiten controlar el nivel de calidad con el que se desempeña cada uno de los aspectos funcionales identificados de relevancia para los usuarios.
- 6 – *Identificación de relaciones KQI/KPI.* El mapeo entre indicadores es clave para poder obtener el valor final de la QoS conseguida para cada uno de ellos, así como los posibles desalineamientos entre la 'QoS Ofrecida' y la 'QoS Entregada' (o conseguida).
- 7 – *Modelado de la QoS percibida por el usuario.* Partiendo de los KQI se modela la 'QoS Percibida' que sirve de entrada al modelo de satisfacción. Se deben tener en cuenta la ponderación relativa a los criterios de QoS y a las preferencias/interés de los usuarios respecto a los distintos indicadores de calidad (KQI).
- 8 – *Análisis de la satisfacción del usuario.* Estudiando la QoS percibida y la disconfirmación, se modela la satisfacción global del usuario, que permite observar desalineamientos entre la QoS Percibida y la QoS Requerida.
- 9 – *Actualización de requerimientos de QoS y comienzo de una nueva iteración.* Partiendo de la última valuación de QoS percibida y las expectativas actuales, se actualiza la entrada del modelo de QoS con las expectativas a futuro y se fijan los nuevos requerimientos de QoS, con lo que se comienza de nuevo todo el proceso global de gestión de QoS.

Como conclusión a la descripción de este modelo, sólo queda decir que constituye el núcleo principal de la arquitectura QoXfera, desarrollada por la misma autora [7], en el seno del grupo de investigación NQaS [13].

Ya se ha tratado con anterioridad en este documento ([capítulo 2.1](#)) la metodología de funcionamiento del marco definido por QoXfera, en el que gracias a sus movimientos de rotación y traslación, se realiza un procedimiento iterativo en busca de la fidelización del cliente a través de la mejora en la gestión de la QoS en los cuatro niveles que se divide su arquitectura: QoS Intrínseca, QoS Percibida, QoS Experimentada y QoS de Negocio.

Esta realimentación entre capas es la que introduce Ibarrola en su modelo de gestión global de la QoS [19], a los mismos niveles que los expuestos por Stankiewicz [4] en su modelo de QoX ([figura 4.5](#)). Sobre dichas tres capas, QoXfera incluye el nivel de QoS de Negocio, definido en la tesis de máster de Cristobo [16], convirtiendo a QoXfera en el primer modelo global de QoX que engloba la totalidad de los aspectos objetivos y subjetivos de la QoS, desde los valores intrínsecos hasta la QoBiz, sin olvidarse del punto de vista del cliente.

Por ese motivo, y teniendo en cuenta el esfuerzo de asentar todo su desarrollo sobre la base de la normativa y regulación existente (ITU-T, ETSI, TMF) y en particular, sobre el marco de la ITU-T G.1000, esta arquitectura ha sido recientemente propuesta como modelo de referencia para una futura revisión de la recomendación ITU-T G.1000 [43].

En la representación gráfica de la [figura 4.16](#) puede observarse la integración del modelo de gestión global propuesto por Ibarrola [19], especialmente en las capas centrales del modelo, así como una definición más detallada de la capa de QoS Experimentada en la que podemos ver por primera vez el concepto de QoE firmemente asentado, como se puede apreciar en el extracto siguiente [44][7]:

Un determinado parámetro objetivo de rendimiento de la red puede tener influencia en la percepción del usuario final (QoP). Una exposición constante a dicha percepción, así como a otros factores, como la satisfacción del usuario con el servicio y sus expectativas, lleva a definir la propia calidad de experiencia (QoE) del usuario, que, eventualmente, puede afectar a su lealtad con el proveedor.

A new global Quality of Service model: QoXPHERE
IEEE Communications Magazine, Vol 52, Issue 1. January 2014

El modelo recoge también que la influencia de la satisfacción y las expectativas sobre la QoE no es unidireccional, sino que, debido a las iteraciones continuas de la esfera, la propia QoE, a instancias de la calidad percibida a lo largo del tiempo por el usuario, también puede modificar los valores de satisfacción y expectativas.

Por ello, se define en esta capa un sistema de encuestas basado en la recomendación ITU-T G.1030 [45] con el fin de determinar la evolución de estos factores, mediante la presentación al usuario de una serie de encuestas en distintos momentos del proceso iterativo.

Dichas encuestas se diseñan específicamente para cada servicio o aplicación a analizar, haciendo uso de las escalas de opinión MOS a las que ya se ha hecho referencia con anterioridad a lo largo de este mismo capítulo (Recomendación ITU-T E.800 [28]).

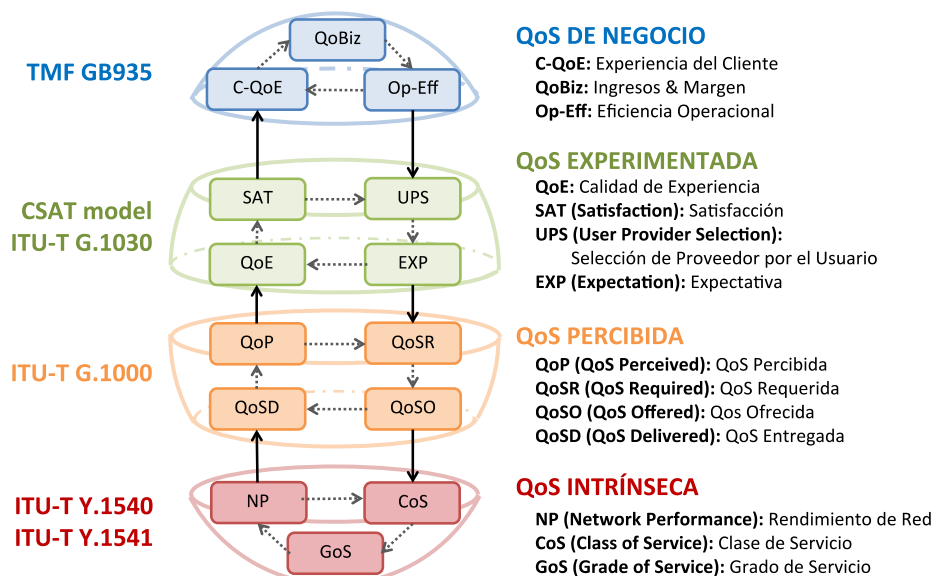


Figura 4.16 – Arquitectura de capas de QoXfera

4.3. Métricas y metodologías de evaluación de la QoE

Una vez realizado el estudio del concepto de Calidad de Experiencia a través de los diversos modelos de gestión de QoS presentados en el apartado anterior, analizamos ahora diversas metodologías para su evaluación. Algunas de ellas ya ha sido avanzada con anterioridad (figura 4.14), aunque en este apartado nos referiremos principalmente a las metodologías definidas en el ámbito de los organismos de estandarización, tanto para la evaluación de la QoE, como en su aplicación al Servicio de Navegación Web sobre el que versa este trabajo.

Naturaleza de la QoE

El autor quiere hacer constar que, dada la naturaleza 'difusa' del término QoE, se procurará explicar para cada recomendación y estudio a continuación, el enfoque que se da del concepto, siguiendo estas pautas:

- **Mediciones:** ¿Cualitativas o cuantitativas? ¿En base a estudios objetivos o subjetivos?
- **QoE:** ¿Hasta qué punto es medible, dependiente del contexto, dependiente del contenido o dependiente de las expectativas?

Para ello, como introducción a los siguientes apartados de este subcapítulo, se presenta una comparativa de estos enfoques, en base a los trabajos de Brooks [46] y Fiedler [47] a este respecto. Ambos autores realizan un exhaustivo estudio de los pros y contras de las metodologías de evaluación existentes, así como de la idoneidad de su utilización según su aplicación. Dada la importancia de dicho estudio, se analiza a continuación su aportación.

En su artículo [46], Brooks distingue dos puntos de vista principales para abordar las mediciones de QoE. Estos enfoques se corresponden, respectivamente, con las metodologías propuestas en diversas recomendaciones de la ITU, y las nuevas propuestas de medición objetiva de QoE de la ETSI, como podremos estudiar más adelante:

- **QoE subjetiva:** Enfoque basado en mediciones objetivas de parámetros técnicos, combinado con el estudio subjetivo de la percepción del usuario o la inferencia de su experiencia a partir de su opinión.
- **QoE objetiva:** Enfoque basado en mediciones objetivas de QoE, no dependiente de la opinión del usuario, pero sí combinable con información subjetiva obtenida a partir de la observación del participante.

Entre las metodologías del primer tipo, destacan las dos siguientes:

- **Pruebas de QoS percibida**

Se trata de un tipo de evaluación subjetiva, ya que se solicita al participante que evalúe la calidad de un determinado servicio. Estas metodologías, centradas en la percepción del usuario, suelen hacer uso de escalas MOS [49] como mecanismo de medición.

A favor de esta metodología está el hecho de que la obtención de datos de percepción de usuario se realiza en escenarios controlados sobre la base de estándares consolidados, lo que facilita la reproducción, intercambio y comparación de datos. Además, la utilización de escalas MOS permite aplicar técnicas estadísticas para desarrollar modelos predictivos.

En contrapartida, este tipo de mediciones se centra demasiado en el dominio de la percepción, lo que implica confiar en la capacidad humana para detectar o reconocer las variaciones en el servicio que se está evaluando. En los casos en los que éste sea el objetivo a analizar, este tipo de tests puede ser de gran utilidad. Sin embargo, cuando se trata de evaluar características de un servicio a partir de la percepción del usuario, y no el propio comportamiento del usuario, existe el riesgo de que este último influya en el resultado de la evaluación. El rendimiento del usuario y la experiencia no tienen por qué coincidir con la percepción del usuario.

- **Encuestas de evaluación subjetiva de QoE**

El concepto de la Calidad de Experiencia va más allá de la percepción del usuario, ya que considera conceptos adicionales como la usabilidad, la satisfacción y otros conceptos no necesariamente subjetivos. Sin embargo, según Brooks [46], las metodologías de evaluación de QoE existentes consideran este concepto desde una óptica completamente subjetiva. Para ello, se hace uso de encuestas de opinión y escalas de valoración, quedando la recogida de datos objetiva únicamente asociada a parámetros tecnológicos. Los valores de QoE se obtienen de manera inferida a partir de las mediciones técnicas y los valores subjetivos de usuario se utilizan para validar la relación entre los parámetros técnicos y el comportamiento del usuario.

En cuanto al segundo tipo, Brooks postula que la evaluación del comportamiento de un usuario no debe depender únicamente de su opinión. Un ejemplo de utilización de este enfoque es el de evaluar la propia experiencia del usuario sobre el uso real de un servicio.

Por ejemplo, si se pretende evaluar la usabilidad de un servicio más allá de parámetros técnicos de rendimiento, la metodología a aplicar debería considerar mediciones análogas a las referidas como ejemplo en la columna derecha de la [tabla 4.1](#). Estas mediciones de QoE se consideran objetivas, puesto que en ningún momento existe dependencia con la opinión del participante, sino con el grado de exactitud en la cumplimentación de una tarea o con factores numéricos que cuantifican objetivamente su experiencia.

La idoneidad de cada uno de estos enfoques depende en gran medida del objetivo a analizar, en función de si se busca validar el rendimiento objetivo de un servicio a partir de la opinión del usuario (QoE subjetiva) o si se pretende parametrizar la propia experiencia del usuario sobre la utilización de un determinado servicio (QoE objetiva).

En ambos casos, es necesario destacar la importancia de una adecuada selección de la muestra de población que se someterá al plan de pruebas, ya que la correcta definición del segmento de usuarios es vital para poder extraer las conclusiones oportunas del estudio.

	EVALUACIÓN SUBJETIVA DE QoE	EVALUACIÓN OBJETIVA DE QoE
Acción	Preguntar / Observar al usuario - Cuestionarios - Escalas de evaluación - Anotaciones - Métodos verbales / no verbales	Medir la actividad del usuario
Mediciones	- Efectividad - Eficiencia - Satisfacción - Disfrute - Presencia social - Impresión del interlocutor	- Grado de exactitud en la realización de las tareas de usuario - Pérdida o ganancia para el usuario - Tiempo total requerido para alcanzar el objetivo - Número de entradas (input) de usuario - Tiempo entre entradas (input) de usuario - Número de turnos requeridos para la comunicación - Número de interrupciones - Número de acciones simultáneas

Tabla 4.1 – Evaluación de QoE: Subjetiva vs Objetiva (Brooks & Hestnes)

	EVALUACIÓN CUALITATIVA DE QoE	EVALUACIÓN CUANTITATIVA DE QoE
Acercamiento	Descriptivo, verbal	Numérico
Mediciones	Escala MOS (escala cualitativa ordinal discreta) - Excelente - Bueno - Regular - Malo - Muy malo NOTA: Asignar números a una escala cualitativa puede condicionar el resultado estadístico. Los conceptos verbales no necesariamente son equidistantes.	Escalas de intervalo o relación (por ejemplo, con valores en los extremos) - 5 - 4 - 3 - 2 - 1 Escalas binarias - 1 / sí - 0 / no A priori más sencillo de procesar (comparativas, estadísticas...)

Tabla 4.2 – Evaluación de QoE: Cualitativa vs Cuantitativa (Brooks & Hestnes)

Además de la comparativa del acercamiento objetivo/subjetivo a la QoE, también se analiza la tipología de datos recogida según si es cualitativa o cuantitativa (tabla 4.2).

A este respecto, Brooks [46] sugiere la utilización de escalas cuantitativas ya que se expresan en términos más absolutos que las cualitativas. Las escalas cuantitativas permiten un estudio directo de media, desviación y correlación de coeficientes, ya que están basadas en respuestas numéricas. Las escalas cualitativas dependen de percepciones verbales, no siendo igual para todos los usuarios la 'distancia' entre los términos verbales. El autor del presente trabajo considera que, de utilizarse escalas cualitativas, se dé el mayor detalle posible para corregir distorsiones en la evaluación debido al lenguaje.

Por su parte, Fiedler, en su trabajo "A generic quantitative relationship between Quality of Experience and Quality of Service" [47], clasifica las métricas que permiten relacionar mediciones de QoE con valores intrínsecos de QoS en tres tipos principales:

- **Métricas de referencia completa (FR – Full Reference):** Tanto la evaluación subjetiva como los datos de rendimiento objetivo están disponibles para la extracción de comparaciones entre QoS y QoE. Un ejemplo de este tipo son los tests de laboratorio en los que se pide a los participantes que evalúen su percepción ante cambios controlados en el rendimiento del servicio.
- **Métricas sin referencia (NR – No Reference):** Los datos intrínsecos de QoS han de inferirse a partir de los datos subjetivos recogidos, al no contar con la información objetiva. Un ejemplo de este caso, es la evaluación subjetiva de la percepción del usuario de un servicio no controlado en un escenario real, al no poder distinguirse entre problemas de calidad referidos al servicio respecto a perturbaciones ajenas en la red.
- **Métricas de referencia reducida (RR – Reduced Reference):** En este caso también se realiza comparativa entre los datos de QoE y QoS, pero a diferencia de las métricas FR, la información objetiva de rendimiento no es dato fijo controlado. En este tipo de métricas se evalúa la QoS de un servicio o la red en función de las variaciones reales de *throughput* o pérdidas, a la vez que se solicita al usuario que evalúe su percepción, satisfacción,...

La siguiente tabla recoge las posibles combinaciones de evaluación de la QoE respecto a las métricas definidas por Fiedler. Nótese que en la tabla resumen no se especifica si las escalas son cuantitativas o cualitativas, ya que dependen en gran medida del diseño de la encuesta.

	MEDICIÓN QoE	MÉTRICA	DATOS QoS
TIPO 1	SUBJETIVA	FR	Controlados en laboratorio
TIPO 2		RR	Medidos en entorno real
TIPO 3		NR	Inferidos por percepción
TIPO 4	OBJETIVA*	FR	Controlados en laboratorio
TIPO 5		RR	Medidos en entorno real

* Los tipos 4 y 5, pese a considerarse mediciones objetivas, incluyen mediciones subjetivas de referencia

Tabla 4.3 – Tabla resumen para la clasificación de metodologías de evaluación de QoE

En el mismo artículo [47], Fiedler también plantea una relación cualitativa entre QoS y QoE independiente de las métricas definidas en las que se puede observar una estimación del valor que adopta la evaluación de la experiencia frente a mayor degradación de la QoS. La tendencia general que toma la curva de relación QoS-QoE se puede observar en la figura 4.17, que representa una relación exponencial decreciente del valor de QoE a mayor degradación de QoS.

Esta gráfica se ha validado a partir de los resultados de varios estudios de Calidad de Experiencia documentados en el artículo [47], por lo que puede ser una referencia útil en la fase de análisis de los resultados obtenidos a la finalización de este trabajo.

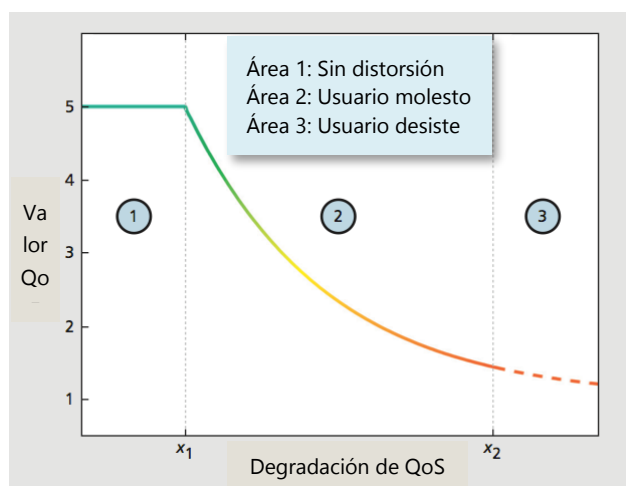


Figura 4.17 – Tendencia general de la curva de relación QoS-QoE

Recomendación ITU-T P.800 (1996)

Comenzamos el repaso a las métricas y metodologías definidas por los organismos de estandarización para la evaluación de la experiencia/percepción de un usuario sobre el funcionamiento de un servicio, con la recomendación ITU-T P.800 sobre los “*Métodos de determinación subjetiva de la calidad de transmisión*” [48]. Esta recomendación define una serie de métodos y pruebas de opinión al respecto de los servicios vocales que utilizan numerosos sistemas de transmisión. Pese a no ser la tipología de servicio que se analizará en el estudio de la QoE, sí es interesante observar algunas de las pautas que se detallan.

En concreto, se presentan metodologías para varios tipos de pruebas basadas en la opinión del usuario en entornos controlados de laboratorio. En referencia a la tabla 4.3, podrían categorizarse como Tipo 1 (pruebas de evaluación subjetiva con métricas FR). En todas ellas se sigue una secuencia muy similar de actuación:

Escenario y plan de pruebas

En todos los casos se parte de un escenario controlado en el que se recrean las condiciones normales del servicio a analizar. Para dicho escenario, se diseña el plan de pruebas del experimento teniendo en cuenta el número de condiciones de prueba, la precisión de los resultados y la capacidad para evaluar correctamente las conclusiones.

En este caso, puesto que se trata de servicios vocales, se introduce la condición de que las conversaciones establecidas tengan un comienzo y final naturales, salvo que se requiera como condición de prueba. Traducido a un servicio genérico, dicha condición equivaldría a que las transacciones de usuario necesarias en el testeo del servicio deben completarse con normalidad, salvo que se requiera el estudio subjetivo de condiciones de error.

Elección de los participantes

Para la elección de los participantes, se recomienda un muestreo de población en el que no existan sujetos que hayan participado recientemente en pruebas similares, tal que puedan verse influenciados los resultados por anteriores experiencias.

De igual manera, dejan fuera a usuarios expertos que pudieran conocer el funcionamiento del servicio o mecanismos de evaluación de calidad. La recomendación tampoco define la obligatoriedad de equilibrar el número de participantes por sexos, a menos que así se requiera en el experimento.

Mecanismo de evaluación

En cuanto a la evaluación de la opinión de los participantes, la ITU-T recomienda una serie de escalas de opinión [49]:

- **Escala de opinión MOS:** Asignación numérica a la siguiente escala cualitativa:
(1) Mala, (2) Mediocre, (3) Regular, (4) Buena y (5) Excelente.
- **Escala de dificultad:** Se formula una pregunta binaria (sí/no) al participante al respecto de alguna condición de dificultad en la utilización del servicio.
- **Otras escalas:** La recomendación distingue tres tipos:
 - Escala numerada (1, 2, 3, 4, 5) en la que sólo se indica la descripción de los extremos.
 - Escala numérica de mayor graduación (por ejemplo: de 1 a 100).
 - Representación gráfica de alguna propiedad sobre una línea recta graduada, de tal manera que la marca del participante indique su evaluación mediante proporcionalidad.

Instrucciones a los participantes

Los participantes reciben un documento explicativo antes de comenzar la prueba en el que se detalla la información no técnica del experimento y lo que se espera de él. Antes de comenzar el experimento se les pregunta si han comprendido todo lo expuesto, aclarándoles las dudas que hayan surgido. Se realiza una demostración del funcionamiento de la prueba y se procede a realización del test en sí mismo, bajo las condiciones previamente definidas.

Recopilación de datos

Los niveles de rendimiento objetivo de la transmisión se conocen previamente ya que se han definido en las condiciones de prueba. Sin embargo, también se realiza su medición mediante medidores controlados que permitan referir la evaluación del participante con cada instante del test. El experimentador recibe dos respuestas subjetivas del participante por cada condición: La nota de opinión sobre la transmisión y la decisión sobre la dificultad.

Tratamiento de resultados

Cada condición da lugar a dos opiniones (MOS y dificultad), mas los datos relativos al nivel de rendimiento objetivo de dicha condición. En casos particulares, también se puede disponer de datos adicionales como, por ejemplo, el comportamiento del usuario durante el servicio (recogida de datos objetiva de QoE sobre el usuario).

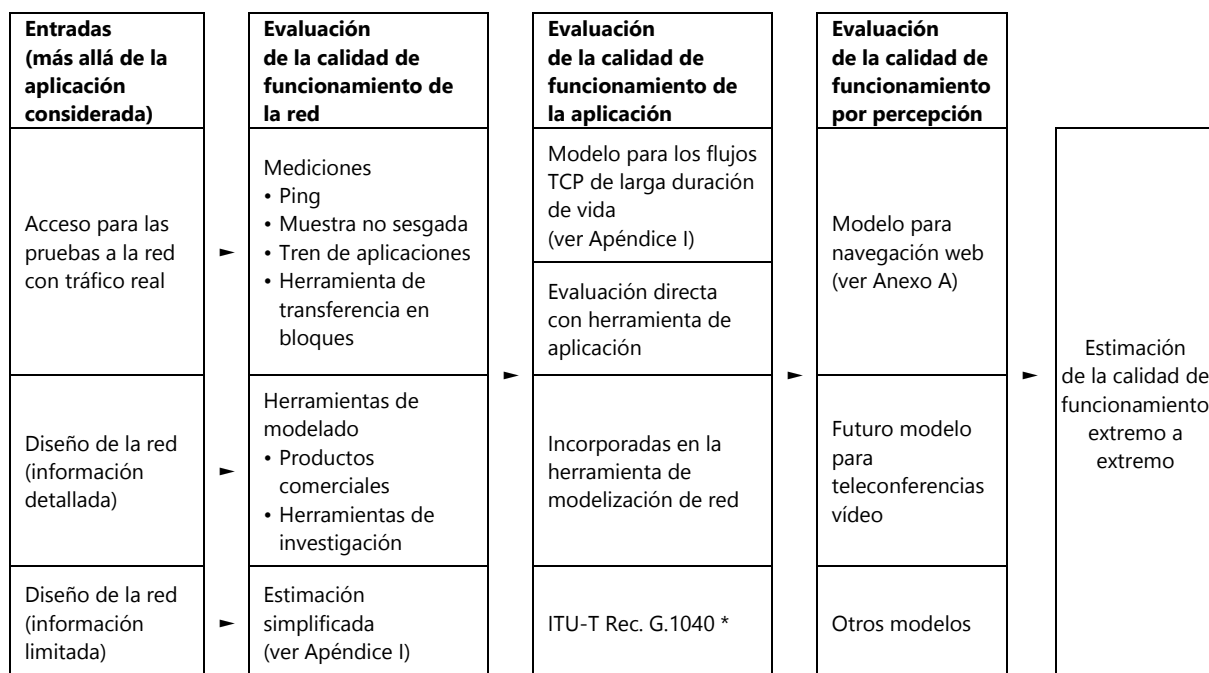
Para cada condición de prueba se debe calcular el promedio de las notas de opinión. Deben evaluarse los límites de confianza y llevarse a cabo pruebas de significado mediante técnicas convencionales de análisis de la varianza, verificando que no hay anomalías ni problemas inexplicables en el análisis de los resultados.

Si las pruebas se han llevado a cabo en varias etapas, se debe realizar el estudio para cada etapa. Por último, si es preciso, puede completarse el examen de los datos mediante gráficos que ilustren la nota media de opinión en función del parámetro sometido a prueba.

Como conclusión al estudio de las metodologías propuestas en esta recomendación, se puede decir que el modelo a seguir es muy extrapolable al estudio de otros servicios y condiciones, como podremos ir viendo a lo largo de las próximas páginas.

Recomendación ITU-T G.1030 (2005)

La recomendación ITU-T G.1030 [45] se centra en la "Estimación de la calidad de funcionamiento de extremo a extremo en redes IP para aplicaciones de datos". En concreto, se define un marco basado en la recomendación ITU-T G.1040 [50] en el que se definen las fases necesarias para la evaluación de la calidad de funcionamiento IP extremo a extremo:



* La Rec. ITU-T G.1040 proporciona la contribución de la red a la calidad de funcionamiento de las aplicaciones de transacciones. No proporciona la estimación completa de la calidad de funcionamiento de las aplicaciones, pero sí da una visión orientada a la aplicación de la calidad de funcionamiento de la red.

Tabla 4.4 – Marco para el proceso de evaluación de la calidad IP extremo a extremo (ITU-T G.1030)

La metodología que define el modelo recogido en el Anexo A de la recomendación², consiste fundamentalmente en experimentos en los que se manipulan los tiempos de respuesta y descarga de una sesión de navegación web, partiendo de condiciones de tráfico real (Metodología Tipo 2, [tabla 4.3](#)).

El alcance del modelo se limita a sesiones de navegación web que constan de dos pasos: un primero en el que se efectúa una petición de búsqueda, y un segundo en el que se muestra una página de resultados. También se proporciona una extensión simple para eventos de temporización, en la que se modela la influencia de la espera de una página simple. En la posterior revisión de 2014 [45] de la recomendación, se añade la puntualización de que este experimento no refleja la dinámica de la navegación web real, en la que se considera la experiencia del flujo, la interacción con el portal web y los modelos de tiempo de carga sucesivos a través de una sesión real de navegación³.

Escenario y caracterización de la red

En el Apéndice I de la recomendación se detalla la conexión de referencia para la realización de las pruebas de funcionamiento extremo a extremo en redes IP ([figura 4.18](#)). Dado que el experimento definido se realiza sobre un escenario real, es necesario considerar los dispositivos intermedios existentes (cortafuegos, traductores NAT,...). Aunque este tipo de dispositivos impiden la transparencia extremo a extremo, no pueden extraerse del escenario a considerar, puesto que son equipos existentes en muchos hogares y redes de empresas.

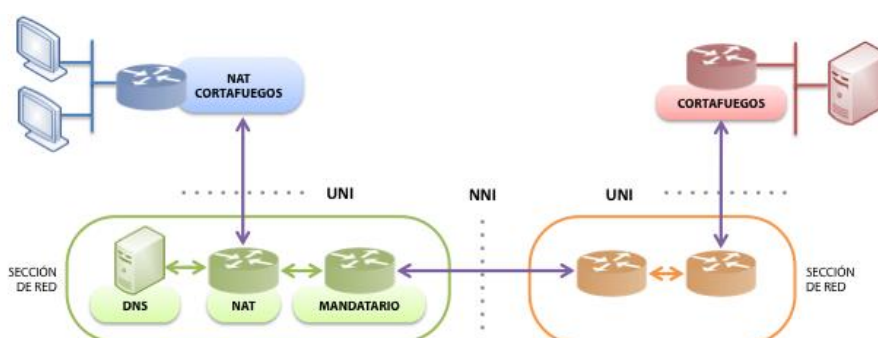


Figura 4.18 – Conexión de referencia con ejemplos de dispositivos intermedios (ITU-T G.1030)

Siguiendo el marco metodológico de la [tabla 4.4](#), se realiza una estimación de la red siguiendo las definiciones de retardos, pérdidas, ancho de banda, capacidad de TCP, transacción HTTP, tiempo de resolución DNS y tiempo de transferencia de datos contemplados en el Apéndice I de la recomendación. Con dicha información dispondremos de una estimación de la calidad de funcionamiento (en términos de QoS) extremo a extremo de la red IP en estudio.

² La Rec. ITU-T G.1030 ha sido actualizada (02/2014) a posteriori de la publicación de la nueva Rec. ITU-T G.1031, a la que se hace referencia en su texto. Fundamentalmente, no ha habido cambios considerables en la recomendación, salvo el renombrado del Anexo A: "Modelo de opinión para las aplicaciones de navegación por la web" como Apéndice II: "Modelo de opinión para las aplicaciones de búsqueda web simple", y la eliminación de la referencia a la inclusión futura de posibles nuevos modelos perceptuales.

³ Esta adición hace referencia a la Rec. ITU-T G.1031, donde se da información de influencia de dichos factores.

Percepción y expectativas

Una vez realizada la estimación de la red, se daría inicio al plan de pruebas por parte de los participantes en el experimento. Previamente a ello, desde la recomendación se plantea la consideración de ciertos factores al respecto de la percepción y las expectativas, que deben tenerse en cuenta tanto en el plan de pruebas como en la posterior recogida de impresiones por parte de los participantes.

En concreto, en la recomendación se definen tres regiones distintas de percepción:

- **Experiencia instantánea:** 0.1 segundos es aproximadamente el límite para tener la sensación de que el sistema responde instantáneamente, lo que es una referencia muy importante a considerar en servicios conversacionales.
- **Experiencia ininterrumpida:** 1 segundo es aproximadamente el límite para que un usuario mantenga ininterrumpido el curso de su pensamiento, aunque pierda la sensación de que el servicio está funcionando directamente. Este dato es necesario considerarlo en servicios interactivos (como los juegos).
- **Pérdida de atención:** 10 segundos es aproximadamente el límite para mantener la atención del usuario en el diálogo. Con mayor retardo, el usuario desea realizar otras tareas mientras espera a que el equipo termine. El valor de retardo es especialmente importante si existe variabilidad en el tiempo de respuesta.

Estas consideraciones pueden deducirse también en la recomendación ITU-T G.1010 [51] sobre “Categorías de calidad de servicio para los usuarios finales de servicios multimedia”, donde se hace referencia, entre otros, al servicio de la navegación web, y se especifica que:

Desde el punto de vista del usuario, el factor principal de calidad de funcionamiento es la rapidez con que se presenta la página una vez solicitada. Se aceptan retardos de varios segundos, pero no superiores a 10 segundos.

ITU-T G.1010 (2001)

Aplicando estos conceptos a los tiempos de descarga, los usuarios tienden a adoptar (y adaptar) su juicio de calidad y expectativas respecto al tiempo de descarga esperado. Los usuarios tienden a aceptar tiempos de descargas mayores si reciben información relativa respecto al tiempo previsto de descarga.

Definición del plan de pruebas

En base a las consideraciones anteriores, se define un plan de pruebas que imite al máximo posible una experiencia de navegación web en la vida real. En el caso del ejemplo recogido en el Anexo A (Apéndice II en la versión de 2014) de la recomendación ITU-T G.1030 [45], se utilizaron tres experimentos subjetivos de navegación web con escalas de tiempo de 6, 15 y 60 segundos respectivamente, representando contextos de red rápido, moderado y lento respectivamente.

En cada sesión, el sujeto realizó estas operaciones, tal y como describe la [figura 4.19](#):

- Petición, obtención y visualización de una página de búsqueda.
- Escritura y envío de un término de búsqueda en esta página.
- Obtención y visualización de la página de resultados.

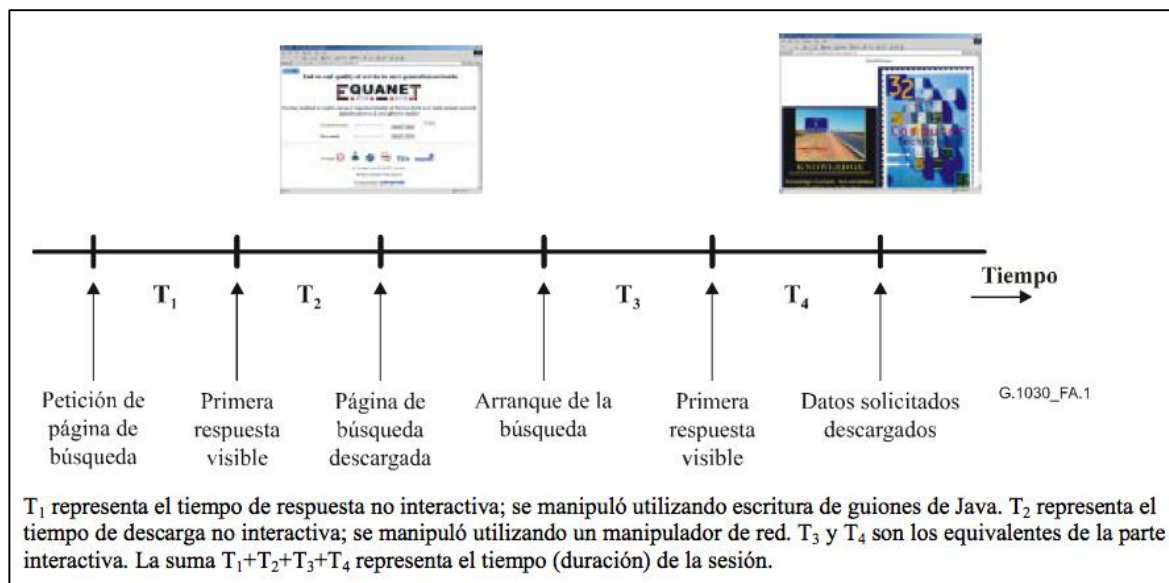


Figura 4.19 – Descripción gráfica de la sesión de la prueba (ITU-T G.1030)

La prueba completa se define para un total de 49 sesiones con diferentes combinaciones de valores T₁ a T₄, desde 0 al fondo de escala de tiempos especificado para cada sesión (6, 15 o 60 segundos). En el ejemplo mostrado, no se especifica si las duraciones totales de cada sesión se presentaron de manera ordenada o aleatoria al usuario ni si había tiempos de descanso para el participante definidos en algún momento de la duración total de la prueba.

Elección de los participantes

A diferencia de la metodología definida en la recomendación ITU-T P.800 [48], en este caso sí existen dos grupos de usuarios divididos en expertos y no expertos. La separación de ambos grupos permite desarrollar un modelo de predicción de calidad de navegación web en función de sus perfiles. No se especifican condiciones respecto al sexo ni grupo de edad.

Mecanismo de evaluación

Para los mecanismos de evaluación, en cambio, se siguen las directrices de la recomendación ITU-T P.800 [48], utilizando una escala de opinión MOS [49] con asignación numérica de cinco puntos sobre la escala cualitativa siguiente:

- (1) Mala, (2) Mediocre, (3) Regular, (4) Buena, (5) Excelente

Tratamiento de resultados

Para cada escala de tiempo, se recogen los resultados de opinión MOS del conjunto de participantes y se representan gráficamente las opiniones (ordenadas) frente a la duración de la sesión en segundos (abscisas), calculando la recta de regresión en cada escala.

La [figura 4.20](#) representa una de las varias gráficas comparativas documentadas en la recomendación, en la que se puede ver el caso de fondo de escala de 6 segundos para usuarios expertos frente a usuarios inexpertos. Según recoge este anexo, en los casos de 6 y 15 segundos, los expertos reaccionaron de manera más correlada (>0.9) que los no expertos, quienes tuvieron un comportamiento más dispar. En el caso de las sesiones de 60 segundos, ambos colectivos tuvieron reacción similar. El estudio recoge, entre sus conclusiones que en todos los casos se observa un comportamiento análogo: La calidad percibida disminuye linealmente con el logaritmo del tiempo de sesión.

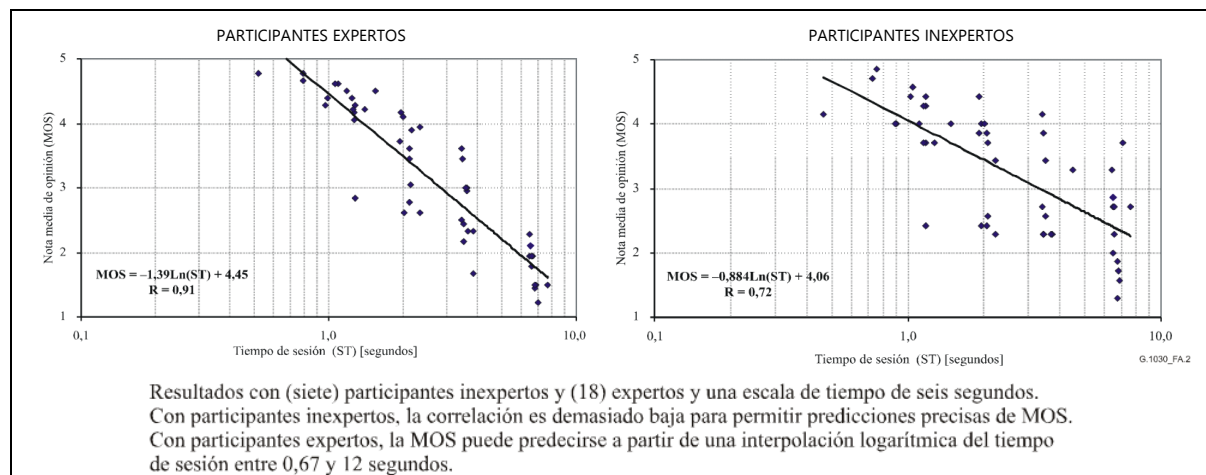


Figura 4.20 – Resultados de los participantes en una escala de 6 segundos (ITU-T G.1030)

Conclusiones

Como conclusión, las pruebas realizadas han permitido normalizar un modelo simple para los contextos de duración de 6, 15 y 60 segundos, que permite evaluar sesiones de navegación web con tiempos de sesión de 12, 38 y 155 segundos, respectivamente [45].

Gracias a este modelo, se ha podido inferir una formulación matemática que permite calcular la percepción del usuario relacionando el tiempo de sesión con el valor de opinión MOS⁴. Es decir, que se podrían definir nuevas metodologías de Tipo 3 ([tabla 4.3](#)) a partir de estas conclusiones. Ejemplo de ello es el trabajo de validación de la recomendación ITU-T G.1030 de Ibarrola [52], en el que se han extrapolado estos resultados a un entorno de aplicación virtualizado, lo que permite el estudio de la QoE en redes de nueva generación (NGN – Next Generation Network) en las que las características del enlace no quedan descritas completamente por el ancho de banda, los retardos o las pérdidas.

Enfoque del ETSI STF 354: Introducción a la evaluación objetiva de QoE

Además de las metodologías definidas por la ITU, el instituto europeo ETSI, a través del grupo de trabajo STF 354, viene definiendo una serie de informes técnicos sobre el factor humano (HF – Human Factor) y la QoE centrada en el usuario.

⁴ **Nota del autor:** Se omiten fórmulas y referencias gráficas adicionales en la explicación, dado que el modelo descrito puede consultarse en la recomendación ITU-T G.1030 [45]

En concreto, y a diferencia de la ITU, desde este grupo se promueve una evaluación de la QoE desde un punto de vista objetivo al margen de la opinión del usuario, en base a la parametrización del comportamiento del sujeto. De hecho, El informe ETSI TR 102 643 [53] parte de la definición de la ITU para QoE (ITU-T P.10/G.100 [2]) sobre la que se realiza una matización en cuanto a su medición⁵:

Calidad de Experiencia:

Medida del rendimiento de usuario basada tanto en mediciones psicológicas subjetivas como objetivas en la utilización de un servicio TIC o producto

ETSI TR 102 643 (2010)

Partiendo de esta matización, las mediciones psicológicas objetivas consideran funciones como el tiempo total en el que el usuario tarda en completar una tarea o la precisión de la tarea medida en el número de errores del sujeto en su realización. Dicho de otra manera, el término QoE incluye no sólo la percepción y satisfacción del usuario sino también la usabilidad del servicio, la efectividad del usuario y su eficiencia, permitiendo así poner en contexto las valoraciones subjetivas del participante.

Incluso sobre los citados conceptos se pueden realizar matizaciones adicionales derivadas del factor humano o de otras disposiciones como, por ejemplo, la ubicación geográfica o la situación financiera (propia del usuario o común al territorio), como representa la figura 4.21 extraída del trabajo de Laghari [36] a este respecto:

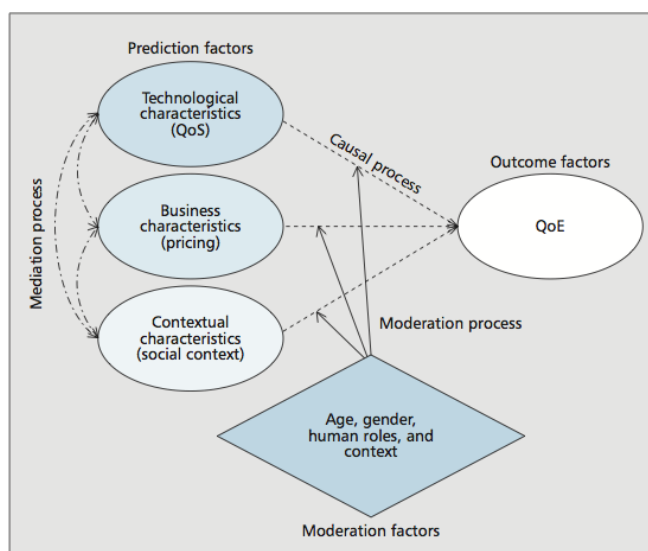


Figura 4.21 – Interacción entre dominios hacia la QoE total (Laghari)

Por tanto, definir una metodología para la evaluación de la QoE teniendo en consideración factores objetivos es algo complicado desde un punto de vista general, a menos que se especifique claramente su alcance en la aplicación o servicio específico bajo estudio.

⁵ **Nota del autor:** La definición de QoE de este grupo de la ETSI no se ha incluido en el subcapítulo anterior ya que hace referencia a su medición, lo que encaja más adecuadamente con el subcapítulo de metodologías.

Ejemplo de metodología de evaluación objetiva de QoE

Partiendo de la afirmación del último párrafo, se puede decir también que es complicado encontrar metodologías bien documentadas al respecto de métricas de QoE objetiva como la definida por Zinner [54].

En su trabajo, Zinner define una metodología para la determinación del valor de QoE óptimo, mediante métricas objetivas que permiten inferir la percepción del usuario respecto a un servicio de transmisión de vídeo en red que disponga de un mecanismo de control capaz de adaptar dinámicamente las características del vídeo (resolución, tasa de *frames*) en función del estado de la red.

La utilización de dichas métricas y su implementación directa en un hipotético mecanismo de control constatan la ventaja de la evaluación objetiva de la QoE en cuanto a su inmediatez, frente a las mediciones subjetivas, que requieren del muestreo de opiniones para correlar una relación entre QoS y QoE.

Métricas

En concreto, en esta metodología se hace uso de las métricas objetivas de QoE siguientes:

- **Métrica de Índice de Similitud Estructural (SSIM):**

Esta métrica, definida por Wang [55], se basa en la asunción de que la percepción visual humana está altamente adaptada a la extracción de información estructural. Dicho en términos más sencillos, el ojo humano es capaz de reconocer formas y figuras y asociarlas a objetos reconocidos, a pesar de que la imagen no sea nítida. Diversos estudios avalan la alta correlación de esta métrica con la calidad de imagen [55] y vídeo [56], en los que se han realizado pruebas comparativas con imágenes afectadas con distintas distorsiones.

- **Métrica de Calidad de Vídeo (VQM):**

Se trata de un método para la medición objetiva de la calidad de vídeo, estandarizado en las recomendaciones ITU-T J.144 [57] e ITU-BT.1683 [58] desde 2004. Las pruebas de validación llevadas a cabo por el Grupo Internacional de Expertos de Calidad de Vídeo (VQEG) mostraron una muy alta correlación entre las encuestas subjetivas de usuario y la medición objetiva de la calidad de vídeo.

Comparativamente, VQM presenta mayor correlación que SSIM a valores de QoE subjetivos. Sin embargo, es más rápido el tratamiento mediante SSIM, por lo que desde la metodología de Zinner [54] se plantea el estudio de ambas, previa definición de la tabla de mapeo de los valores QoE objetivos con valores de QoE subjetivos en términos de la escala MOS:

MOS	SSIM	VQM
5	$0.99 < x$	$0.20 > x$
4	$0.95 \leq x < 0.99$	$0.40 > x \geq 0.20$
3	$0.88 \leq x < 0.95$	$0.60 > x \geq 0.40$
2	$0.50 \leq x < 0.88$	$0.80 > x \geq 0.60$
1	$x < 0.50$	$x > 0.80$

Tabla 4.5 – Mapeo de métricas objetivas a métricas subjetivas de QoE (Zinner)

Escenario y caracterización de la red

Puesto que esta metodología busca poder determinar la QoE óptima para el diseño de un mecanismo de control adaptativo de un servicio de transmisión de vídeo a través de la red, será la propia red el escenario a contemplar.

En base a esto, podremos clasificar esta metodología en el Tipo 4 ó 5 de la [tabla 4.3](#) en función del conocimiento y control que se tenga de dicha red. En este caso, se parte de una situación inicial en la que se ha llevado a cabo la caracterización necesaria de los siguientes parámetros y condiciones:

- El entorno de red del usuario. Ejemplos: ADSL, FTTH, WiMAX...
- Las condiciones actuales de la red extremo a extremo: Pérdidas de paquetes, *throughput*...
- Parámetros relativos al usuario: Capacidades del terminal de usuario, SLAs...
- Información específica del servicio y la aplicación: Códec de video usado, tipo de contenido visualizado (deporte, música, un plató...). Esta información es importante ya que da cuenta de la profundidad del plano o de la variabilidad en el cambio de escenas.
- Otra serie de consideraciones sobre voz y tráfico web.

Elección de los contenidos

Al igual que en las pruebas de medición subjetivas, se requiere definir las condiciones que van a testarse. En el caso de las pruebas subjetivas, es el usuario el que evalúa mediante la escala MOS. En el caso de esta prueba objetiva, se busca medir la respuesta del mecanismo de control adaptativo. En concreto, se hace uso de tres vídeos diferentes con las siguientes condiciones:

Nombre	BLUE SKY	CROWD RUN	PARK JOY
Número de <i>frames</i>	216	499	499
Tasa de <i>frames</i>	30	30	30
Ancho de banda medio (MB/s)	0.82	1.54	1.85
Duración (segundos)	7.2	16.63	16.63
Tipo de movimiento	bajo-medio	medio	medio

Tabla 4.6 – Propiedades de las secuencias de referencia (Zinner)

Se parte, en todos los casos, de vídeos en resolución 1080p codificados en H.264/SVC en diferentes tasas (1.875, 3.75, 7.5, 15 y 30 fps –*frames* por segundo) así como a diferentes resoluciones (480p, 640p, 960p, 1216p). El códec H.264/SVC permite incluir en un mismo fichero de vídeo, todas las capas superpuestas de las tasas y resoluciones definidas.

El caso de mayor resolución y tasa de *frames* (1216p y 30 fps) se considera como vídeo de referencia para la comparación posterior, suponiendo que el usuario realiza dicha petición, y asumiendo que éste dispone de equipamiento con capacidad suficiente para visualizarlo.

Plan de pruebas

El plan de pruebas consiste básicamente en evaluar todas las combinaciones de vídeo definidas para que sean evaluadas a través de software que implementa las métricas SSIM y VQM, de cara a determinar su idoneidad para utilizar dichas métricas como elemento decisor en el mecanismo de control citado.

Para el escalado del vídeo en la representación en destino, se realizó interpolación en base a dos técnicas distintas (interpolación bicúbica e interpolación del vecino más cercano). Se omite entrar en detalle en la explicación de dichas técnicas ya que no son objeto de estudio en este trabajo, aunque sí se comenta que en los resultados se analiza su idoneidad en función de la característica de vídeo que se vea afectada en la transmisión, debido a la supuesta actuación del mecanismo de control adaptativo del servicio.

Resultados obtenidos

Se han analizado los resultados en base a las dos características diferenciales de los vídeos: las resoluciones y la tasa de *frames*. La primera gráfica a continuación muestra la comparación objetiva entre los tres vídeos para ambas métricas, siendo las abscisas el valor relativo de la resolución normalizado al máximo tamaño. Es decir, la resolución 1216p toma el valor 1 en la escala representada en las abscisas, siendo el resto de resoluciones, una representación proporcional.

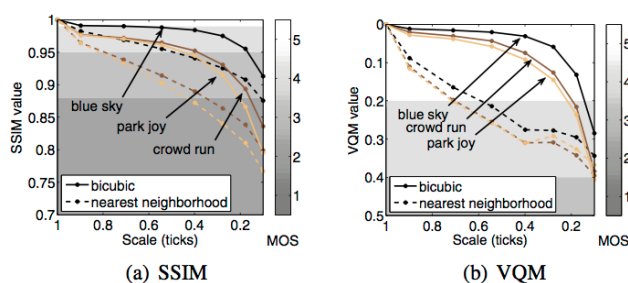


Figura 4.22 – Comparación objetiva de diferentes resoluciones (Zinner)

Se puede observar en la gráfica anterior cómo el usuario percibiría con mejor calidad la aplicación de la interpolación bicúbica cuando la condición de la red obligue a reducir la resolución del vídeo en su transmisión. En cambio, la [figura 4.23](#), referida a la variación en la tasa de *frames*, plantea resultados más cercanos en QoE en ambas interpolaciones, siendo la métrica SSIM más benevolente en la consideración de satisfacción por parte del usuario.

Como vemos, la utilización de métricas objetivas, validadas respecto a la percepción subjetiva de los usuarios, permite extraer conclusiones ante la idoneidad de una u otra característica de estudio, sin necesidad expresa de hacer depender toda la investigación de una buena selección de participantes. A cambio, puede darse que, para el análisis de un determinado servicio, no existan métricas definidas que permitan este tipo de metodologías, por lo que sería necesario seguir realizando encuestas subjetivas de percepción.

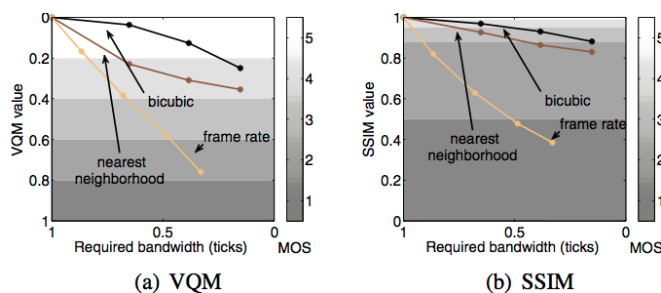


Figura 4.23 – Comparación objetiva de diferentes tasas de frames para el vídeo 'crowd run' (Zinner)

4.4. QoE en el Servicio de Navegación Web

En apartados anteriores se han abordado varios modelos y metodologías para la evaluación de la QoE en diversos servicios y aplicaciones. En este apartado, se pretende realizar un acercamiento a la QoE específicamente para el servicio en el que se basa el estudio del presente documento: el Servicio de Navegación Web.

Estudio 1: Primer acercamiento a la medición de la calidad de navegación web

Entre los trabajos que en los últimos años han venido publicándose al respecto de la calidad experimentada en la navegación web, destaca como un interesante punto de partida, la publicación de 2013 titulada *"Towards determining web browsing Quality of Experience"* [59].

En este documento, Fallon propone una metodología que permite realizar un acercamiento a la caracterización del servicio de navegación web mediante el uso de métricas objetivas centradas en el navegador. Teniendo en cuenta que más del 50% del tráfico de Internet actual se debe a la navegación web en general [60], y al uso del protocolo HTTP en particular, parece adecuado intentar comprender el funcionamiento de un navegador en el proceso de descarga de contenidos, con el objeto de determinar su influencia en el usuario.

En cuanto a la idoneidad de las métricas objetivas, Fallon destaca que el modelo subjetivo de opinión que la ITU describe en la recomendación ITU-T G.1030 [45] puede ser muy útil al comparar diferentes sesiones de navegación sobre una misma página web, pero puede no resultar adecuado para evaluar la experiencia de los usuarios sobre el servicio de navegación en general, ya que el rendimiento bruto de una página web no es el único criterio en el que los usuarios se basan al emitir su valoración.

Escenario: Red de acceso y navegador

Para las pruebas, Fallon [59] selecciona las 6 redes de acceso más comúnmente utilizadas por usuarios de Internet, cuya caracterización en términos de ancho de banda y retardo se puede observar en la tabla siguiente. No se detalla, sin embargo, la variabilidad de estos valores en el tiempo, ni la proximidad en el tiempo de la caracterización con las pruebas.

Tipo de Acceso	RETARDO (ms)	DESCARGA (Mbps)	SUBIDA (Mbps)
Red LAN corporativa	10	50,54	56,38
Red Wireless fija (alta ocupación)	72	2,94	0,05
Red Wireless fija (baja ocupación)	75	3,46	0,37
Banda ancha móvil (señal fuerte)	76	2,41	0,57
Banda ancha móvil (señal débil)	83	3,63	2,41
Acceso DSL rural	85	0,13	0,21

Mediciones de www.speedtest.net

Tabla 4.7 – Características de las redes de acceso del estudio de Fallon

En cuanto a la evaluación del comportamiento de las sesiones de navegación, se requiere de un mecanismo que permita medir cuantitativamente el rendimiento de dichas sesiones. Para ello, Fallon hace uso del navegador *Firefox* [59], dada la expansión global de este navegador y la disponibilidad de diferentes *add-ons* que facilitan la medición.

En concreto, se utilizan dos complementos: *iMacros* [61], que permite una ejecución controlada del navegador mediante la definición de las páginas a visitar en una macro prediseñada, y una versión modificada de *HttpFox* [62], que permite registrar las métricas de la sesión web, almacenando los tiempos de peticiones y descargas en un fichero.

En el documento [59] no se especifican las características al respecto de los terminales de pruebas en los que está instalado el navegador, ni si éste se encuentra disponible en la misma versión en todos los equipos. Sí se identifica, en cambio, la necesidad de que el cacheado de información haya sido deshabilitado, para forzar la carga de todas las peticiones web. Asimismo, se especifica que el máximo número de conexiones simultáneas que admite el navegador esté fijado en el valor por defecto: 32.

Contenidos

Este último dato es muy importante, si consideramos que el tamaño medio y el número medio de elementos contenidos en las páginas web se han duplicado en los últimos años [63]. Por ello, para la realización de las pruebas se han seleccionado 6 portales web diferentes, de distinta tipología y cantidad de elementos (tabla 4.8).

URL	TIPO	Nº VECES	KB BAJADOS		KB SUBIDOS		Nº PETICIONES	
			MEDIA	STDEV	MEDIA	STDEV	MEDIA	STDEV
www.ait.ie	Educación	49	801,6	646,6	26,1	1,2	46,4	2,5
www.ericsson.com	Corporativo	56	970,5	1054,8	26,8	3,3	45,6	5,8
www.google.com	Buscador	59	353,1	111,0	9,7	1,8	16,5	3,1
www.irishtimes.com	Noticias	32	1909,0	116,7	137,2	3,8	214,2	3,4
www.linkedin.com	Red social	41	709,9	267,6	71,9	14,7	120,3	20,8
www.youtube.com	Streaming Video (31s)	32	3078,9	302,3	66,4	4,5	97,8	4,7

Tabla 4.8 – Características de los portales web del estudio de Fallon

Metodología

Partiendo de las configuraciones especificadas, se define la lista de portales en el complemento *iMacros* [61] y se ejecuta la prueba. El motor del navegador realiza la carga sucesiva de dichos portales y va arrojando resultados al *plugin* modificado *HttpFox* [62], que se encarga de generar el documento con las métricas recogidas, tal y como se muestra en la figura a continuación. En vista de las métricas objetivas y el caso de escenario real, se puede clasificar esta metodología como Tipo 5 (Medición objetiva con métricas RR, tabla 4.3).

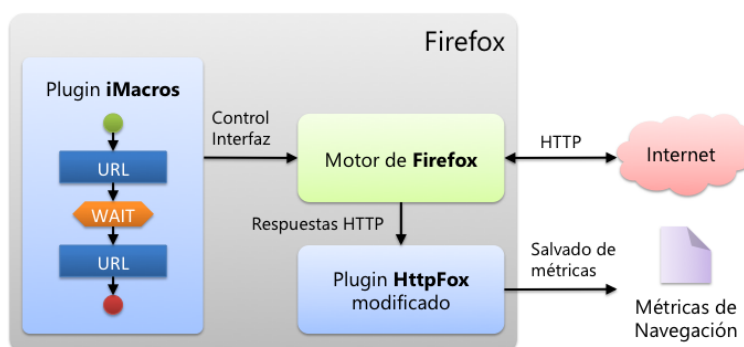


Figura 4.24 – Esquema de captura de experiencia de navegación web (Fallon)

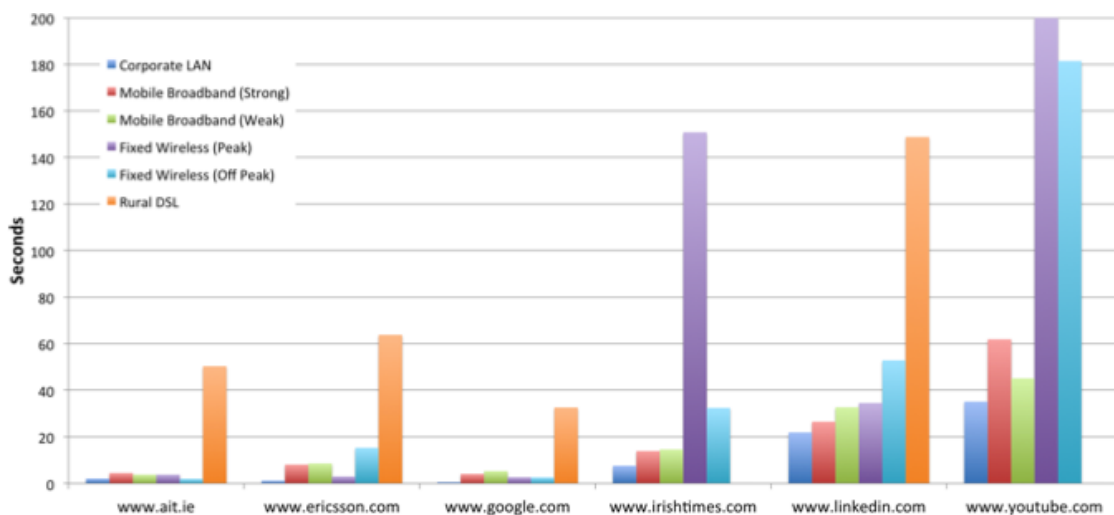


Figura 4.25 – Tiempos de carga de portales web en función del tipo de acceso (Fallon)

Resultados

Los resultados del estudio de Fallon [59] permiten realizar interesantes observaciones. En vista de los datos de caracterización de las distintas redes de acceso (tabla 4.7), se puede esperar que la red LAN corporativa se comporte unas diez veces mejor que el siguiente tipo de acceso con mejor ancho de banda. Sin embargo, los tiempos de carga de las páginas no muestran esa proporción. Esto se debe principalmente a que factores adicionales como el propio estado de carga del servidor, el uso de *proxies* y otros retardos propios de Internet que contribuyen a ese tiempo de carga.

En general, todos los tipos de accesos respondieron adecuadamente a la descarga de contenidos de sitios de contenidos mayoritariamente textuales (buscador, educación y web corporativa), notándose tiempos mayores en servidores más alejados geográficamente.

Sí destaca la importancia del número de conexiones que requiere cada página. Se observaron tiempos mucho mayores en páginas que superaban ampliamente el número máximo de 32 conexiones simultáneas. La red social (*LinkedIn*) requiere 120 conexiones, frente a las 40 de la web corporativa (*Ericsson*) pese a que esta última es más pesada. Esto tiene un claro efecto negativo en el tiempo de carga, como se aprecia en la figura 4.25.

Conclusiones

A raíz de estos resultados, desde el estudio [59] se destacan una serie de conclusiones. En primer lugar, se puede deducir que el tipo de contenidos afecta a la percepción de usuario. Los tiempos de carga son importantes, pero también debe tenerse en cuenta el orden en el que los elementos son presentados y la propia estructura de la web. De la misma manera, el tipo de dispositivo utilizado influye también en la experiencia navegacional del usuario.

En opinión del autor del presente documento, este estudio puede servir como referencia para una definición adecuada de las condiciones de partida en la realización de un estudio de evaluación de QoE en el servicio web, aunque las métricas objetivas limitan la posibilidad de evaluar las expectativas reales del usuario más allá de las estimaciones del observador.

La especial atención a la elección de los dispositivos a utilizar en el experimento, la caracterización precisa de la red utilizada y otra serie de observaciones adicionales, como deshabilitar la caché del navegador, pueden resultar determinantes para garantizar la validez de los datos obtenidos. La influencia de estos factores sobre las condiciones que se quieren evaluar ha de estar bien cuantificada y controlada, y el conocimiento de la opinión del usuario puede ser vital para dar sentido a los resultados objetivos:

Se puede estimar que un tiempo de carga mayor influye negativamente en el usuario, especialmente si comparamos dos páginas de tipología de contenidos similar. Sin embargo, no podemos extrapolar esa misma conclusión a todos los casos. Puede que el usuario tenga una mayor tolerancia a tiempos de carga mayores si es consciente de que el contenido que está descargando tiene unos requerimientos de ancho de banda mayores, o si puede hacer uso de dicho contenido mientras se está descargando (como en el *streaming* de vídeo).

Establecer una relación adecuada entre los indicadores objetivos de calidad y la experiencia del usuario no es una tarea sencilla, como se puede observar en el siguiente trabajo analizado en este capítulo.

Estudio 2: Relación entre QoS y QoE en las sesiones de navegación web

Precisamente a este respecto, Cui [64] presenta un estudio en el que se analizan distintas métricas y se propone una metodología para evaluar la QoE mientras se hace uso del servicio de navegación web, de cara a establecer una relación QoS-QoE mediante técnicas de *clustering*⁶ de datos de QoS.

Escenario

Al igual que en el caso anterior, se hace uso del navegador *Firefox* en este experimento, dado su potencial para integrar complementos externos. En concreto, se instala el navegador junto con un *plugin* que permite al usuario expresar su insatisfacción con el tiempo de carga experimentado en el renderizado de una página web, haciendo clic en el icono gráfico a modo de botón que se embebe en la barra de herramientas del navegador.

Este *plugin* permite distinguir la sesión a la que está asociada dicha valoración, al relacionar la acción del usuario sobre el citado botón con la pestaña que en ese momento se está visualizando en el navegador. Para la captura de las trazas de paquetes intercambiados, se hace uso de un *software* basado en *libpcap* [65] que permite almacenar las métricas obtenidas en una base de datos para su posterior procesado.

En cuanto al escenario propiamente dicho, se corresponde directamente con Internet. Se cuenta con tres terminales de usuario conectados en tres redes diferentes: Una red ADSL doméstica, una red corporativa de 100Mbps y la red de una residencia de estudiantes.

⁶ **Clustering (agrupamiento):** Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia se mide con respecto a todas las variables disponibles [66].

El equipo de la red ADSL se conecta mediante WiFi a una distancia muy cercana al *router*, mientras que los otros dos lo hacen por cable. Los tres equipos se encuentran en el mismo país, Francia, y los navegadores no tienen deshabilitada la caché, aunque su contenido se elimina al final de cada sesión web. Existe un cuarto equipo conectado a una red ADSL desde Turín, Italia, en similares condiciones al equipo francés.

Métricas

Para describir la QoS de una sesión web, Cui [64] hace referencia a las métricas específicas de retardo y pérdida de paquetes en base al intercambio de mensajes entre cliente y servidor a nivel TCP (figura 4.26).

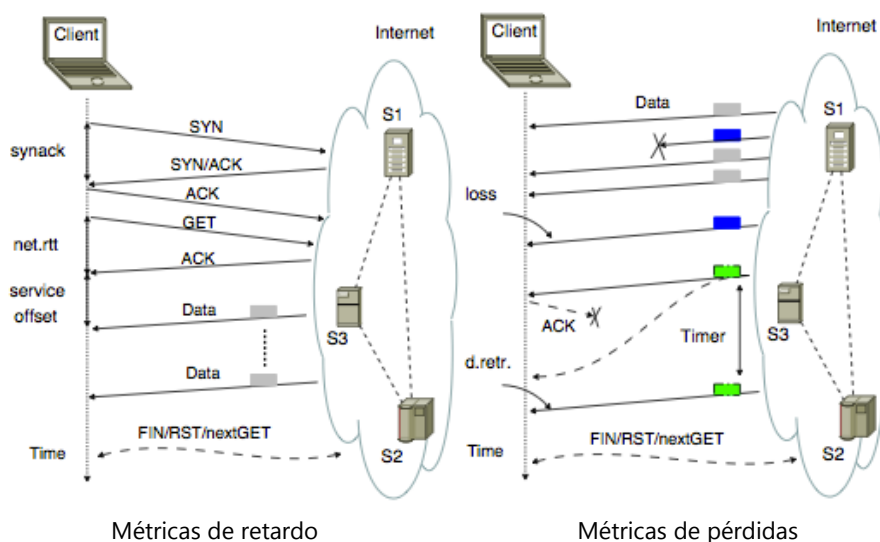


Figura 4.26 – Métricas relacionadas con la descarga de un elemento de una página web (Cui)

Cada página web contiene múltiples elementos que deben ser descargados para conformar la estructura que visualmente se ofrece al usuario. Para cada descarga individual (para cada elemento a descargar), se computan las métricas de retardo y pérdida de paquetes. Posteriormente, se calcula el valor medio global para cada métrica. De esta manera, se obtiene el indicador de desempeño (KPI) de la sesión web completa, en base a las métricas representadas en la figura 4.26: *synack*, *net.rtt*, *serv.off*, *rtt*, *loss*, *d.retr* [64] y *nr*, el número de total de peticiones GET durante toda la sesión.

Contenidos y plan de pruebas

Se definen dos experimentos de distinta naturaleza:

- **Experimento controlado**

Las pruebas se realizan desde los tres equipos franceses, teniendo como destino una lista predefinida de portales web geográficamente distribuidos: 3 en Francia, 4 en China, 2 en Estados Unidos y 1 en Canadá [64]. No se especifica el procedimiento exacto del plan de pruebas, aunque sí se detalla que el experimento se realiza simultáneamente desde los tres terminales de usuario con una duración total cercana a las 8 horas.

- **Navegación libre**

Con la intención de analizar un comportamiento más realista, los equipos pueden navegar libremente por cualquier página web tradicional (se excluyen redes sociales y similares). En este caso, participan los dos terminales conectados mediante ADSL (el francés y el italiano) y el de la red de estudiantes. Se evalúa la sesión web como 'pobre' cuando el usuario hace clic en el botón de insatisfacción del navegador. El valor de 'bueno' se obtiene cuando dicho botón no se utiliza para la sesión en curso.

Metodología

Dado que el escenario planteado es Internet y que el conocimiento de la red se limita a las mediciones simultáneas de las métricas recién descritas durante el tiempo en el que el participante está evaluando su experiencia navegacional, podemos clasificar la metodología de este trabajo como Tipo 2 (Evaluación subjetiva con métricas RR, [tabla 4.3](#)).

Puesto que las redes asociadas a todos los equipos participantes manejan distintas condiciones de acceso a Internet, los valores de los respectivos indicadores KPI se encuentran en diferentes unidades, por lo que todos ellos son normalizados a la unidad.

Para cada experimento, el procesamiento y comparación de los resultados entre los equipos se hace mediante *clustering* [66] en base al algoritmo de cuantización vectorial *k-means*⁷ por el que se dividen las muestras obtenidas en diferentes grupos mutuamente excluyentes. Cada muestra asignada a un clúster está separada una determinada distancia de error respecto al centroide de dicho clúster⁸. Queda únicamente determinar el número óptimo de clústeres a utilizar para reducir al máximo dicha distancia de error. Las pruebas empíricas realizadas determinan que el valor adecuado está entre 3 y 5. Se opta por utilizar 4 clústeres.

Análisis de resultados

Para ambos experimentos contemplados, se procesan los resultados obtenidos y se representan gráficamente para cada clúster mediante *boxplots* [67], que proveen el valor mediano de cada métrica, así como el cuartil inferior y superior. Dado el concienzudo análisis de los resultados bajo distintas consideraciones de agrupación en clústeres, se presenta en la [figura 4.27](#), únicamente como referencia, el ejemplo de análisis global de los resultados del experimento controlado. Otros análisis incluidos en la extensa documentación del estudio [64] incluyen análisis por portales visitados y por clientes, tanto en el experimento controlado como en el de navegación libre.

Al respecto de la [figura 4.27](#), se puede observar que los portales web asociados a los clústeres 1 y 4 tienen tiempos de carga en el orden de unos pocos segundos, mientras que los portales de los clústeres 2 y 3 están por encima de los 10 segundos.

⁷ **Algoritmo k-means:** Algoritmo consistente en la división de un conjunto de n observaciones en k grupos (clústeres) en el que cada observación se asigna al grupo más cercano a la media [66].

⁸ **Distancia de error:** Para el caso en estudio, se define como la distancia euclídea al cuadrado entre el valor de la observación en un clúster y el centroide del clúster. El centroide es el punto equidistante de todas las observaciones asignadas a dicho clúster [66].

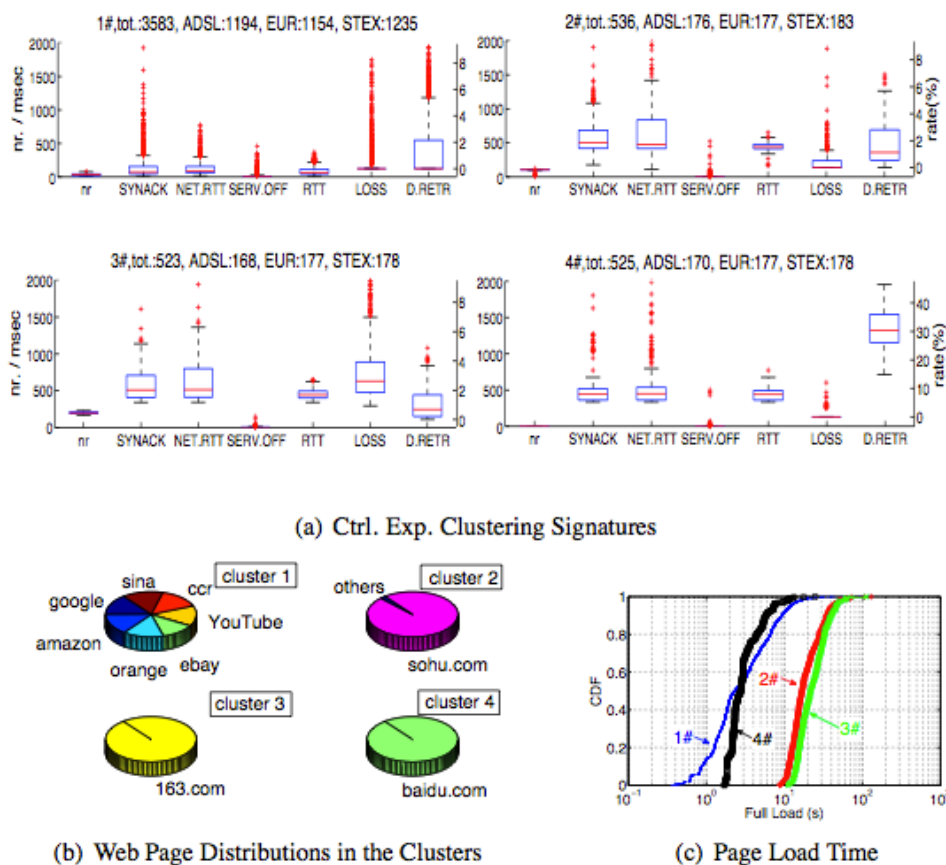


Figura 4.27 – Clustering de todas las sesiones del experimento controlado para los 3 terminales (Cui)

Se aprecia además, que los clústeres 2 y 3 contienen casi exclusivamente peticiones a una única página web respectivamente. Ambos portales (*163.com* y *sohu.com*) están alojados en China. Recordemos que las pruebas se están realizando desde Francia y que la conectividad entre ambos países tiene normalmente, unos tiempos de retardo de ida y vuelta (RTT) de unos 500ms [64]. Se puede apreciar, de hecho, en la métrica de establecimiento de conexión TCP (*synack*), cómo el tiempo de *handshake* inicial es mucho mayor en estos dos clústeres que en el caso del número 1. El clúster 4 también está asociado a una página china (*baidu.com*), en la que se observa el mismo efecto para la conexión TCP. Sin embargo, tiene un volumen de pérdidas (*loss*) mucho menor al de los clústeres 2 y 3, por lo que los tiempos de carga son mucho menores. También realiza menos peticiones de objetos (*nr*) para conformar la estructura web a mostrar al usuario, lo que redundará en su rapidez de carga.

El resto de análisis incluidos en el documento (por páginas, por clientes,...) generan gráficas similares de las que se extraen conclusiones comparativas de manera análoga a la presentada como referencia. Cabe destacar, sin embargo, una tabla en la que se recogen las valoraciones de los usuarios (bueno/pobre) respecto a las métricas obtenidas para el caso de la navegación libre cuya distribución en clústeres se presenta en la [figura 4.28](#).

Dicha tabla, presentada a continuación, da cuenta de la relación de parámetros de QoS considerados en el estudio, respecto a las valoraciones de QoE de los usuarios participantes.

CLUSTER	VALORES MEDIANOS [Las métricas con * se miden en milisegundos]							EVALUACIONES	
	NR	SYNACK *	NET.RTT *	SERV.OFF	RTT *	LOSS	D.RETR	TOTAL	% POBRE
#1	34	474	485	3	431	0.0 %	0.0 %	209	21 %
#2	10	82	124	12	70	0.0 %	0.0 %	399	6 %
#3	113	285	202	8	76	0.6 %	2.1 %	81	42 %
#4	44	1310	1122	96	505	11.7 %	1.6 %	44	64 %

Tabla 4.9 – Relación QoS-QoE del estudio de Cui

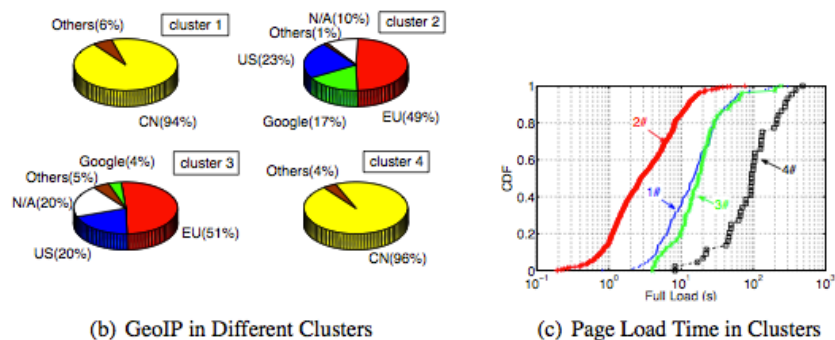


Figura 4.28 – Clustering de las sesiones navegación libre (Cui)

De la observación de la [tabla 4.9](#) se pueden extraer interesantes consideraciones: Se puede observar una clara relación entre el tiempo de carga y la disconformidad de los usuarios. Los clústeres 3 y 4 dan cuenta de ello. Probablemente, el usuario no aprecia si el tiempo de carga depende de retransmisiones por pérdidas o por la lejanía física al servidor que accede. Lo que sí es notable, en cambio, es que el número de peticiones de objetos que conforman la página no tiene una relación clara con la valoración negativa. Sí es cierto que dicha valoración es peor en el clúster 3 que en los clústeres 1 y 2 a este respecto. Sin embargo, comparando los clústeres 3 y 4, influye más el retardo en comenzar a cargar la página que el número de objetos a cargar. La carga progresiva de dichos objetos permite al usuario acceder y disfrutar de ciertos contenidos mientras otros están aún cargándose, lo que explica también el menor número de evaluaciones negativas del clúster 3 frente al 4.

Conclusiones

Al respecto del análisis exhaustivo llevado a cabo, Cui [64] realiza un brillante primer acercamiento a la relación QoS-QoE en un entorno real, que podría facilitar la inferencia de formulaciones matemáticas que relacionen las métricas clásicas de QoS con las valoraciones obtenidas de QoE.

Los resultados presentan una alta correlación entre retardos elevados y el nivel de insatisfacción de los usuarios. Sin embargo, *la correlación no implica causalidad* [64]. Pueden existir múltiples factores no contemplados que estén influyendo en la evaluación, como servidores DNS mal configurados, o *proxies* sobrecargados en algún punto de la red.

Como valoración personal, el autor de la presente tesis considera que, si bien los resultados obtenidos en este concienzudo trabajo pueden dar una idea de la QoE de los usuarios en el servicio de navegación web, también muestran lo extremadamente complejo que es establecer relaciones adecuadas entre la experiencia navegacional y las métricas intrínsecas.

Un análisis métrica a métrica, procurando mantener fijos el resto de valores intrínsecos en un escenario controlado, puede facilitar mucho la inferencia de una formulación relacional. Sin embargo, al considerar la aportación de múltiples factores, dicha formulación se complica al tener que definir un procedimiento que cuantifique el grado de aportación de cada una de esas métricas a la satisfacción del usuario.

Es por ello que, generalmente, los estudios que hasta la fecha han procurado abordar las relaciones entre QoS y QoE en el Servicio de Navegación Web, suelen centrarse en evaluar aspectos específicos del servicio, procurando mantener controlados el mayor número de factores adicionales que influyen en la experiencia navegacional del usuario.

Estudio 3: QoE en la navegación web en función del contenido

Un ejemplo de estudio de evaluación de QoE en entorno controlado es el que presenta Nguyen en su publicación *“Assessment of Quality of Experience for Web Browsing – As function of Quality of Service and content factors”* [68].

En este estudio, se pretende establecer una relación QoS-QoE a partir de la experiencia de los participantes en la navegación web a través de diversos contenidos en diferentes condiciones de retardo.

Escenario y plan de pruebas

A diferencia de los casos anteriores, este estudio se asienta sobre una red controlada que cuenta con 7 routers y 6 switches, en la que los parámetros de rendimiento de la red son modificados mediante la utilización del software de emulación de red **tc-netem** [69].

En concreto, las métricas de rendimiento intrínseco (KPI) que se modifican, son las relativas al retardo y la pérdida de paquetes. La elección de dichas métricas responde principalmente a la voluntad de relacionar los valores de QoE obtenidos en el experimento, con el indicador de calidad (KQI) objetivo que parece tener más influencia en la satisfacción del usuario: el tiempo de carga de la página web [70]. Este indicador depende, además de los KPI citados, de factores como el ancho de banda en la red de acceso a Internet o el *jitter*, entre otros.

Asimismo, también se limita el número de peticiones por segundo al servidor de contenidos mediante la utilización de **mod_qos** [71], un módulo de calidad de servicio del servidor web.

En cuanto al experimento, se define su duración en base a 36 sesiones consecutivas en las que se evalúan diferentes condiciones en la red. Cada sesión, de apenas 3 minutos de duración, es seguida por un cuestionario de evaluación. La duración total del experimento es de 108 minutos, como muestra la figura siguiente:

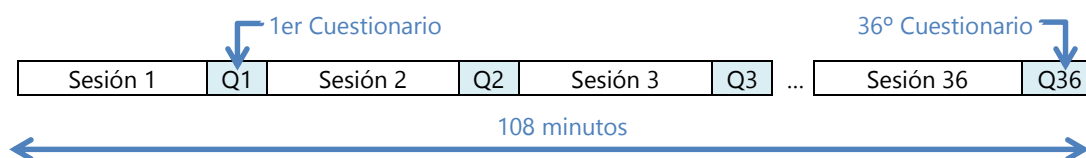


Figura 4.29 – Temporización del experimento de Nguyen

Participantes y mecanismo de evaluación

Para la realización de las pruebas se seleccionan 21 participantes, 12 de los cuales participan en una versión del experimento en la que se integra un factor de contenido de cuatro niveles, que se detalla más adelante. Todos los participantes son estudiantes universitarios.

Para la recogida de sus valoraciones, se hace uso de una escala cualitativa de opinión MOS [49] de 5 niveles con la siguiente leyenda acerca de su experiencia con el servicio evaluado: (1) Inaceptable, (2) Insatisfecho, (3) Regular, (4) Satisfecho, (5) Muy satisfecho.

Antecedentes y definición de contenidos

Partiendo de los conocimientos adquiridos en estudios propios sobre QoE en los que se analizan métricas únicamente relativas a la red [72][73], Nguyen identifica un patrón en la evaluación de los usuarios participantes al que bautiza como '*regla de la naturaleza*'⁹ [68]. Los resultados de dichos estudios anteriores presentan una tendencia hacia la insatisfacción del usuario con el aumento del retardo, pero sin una clara distinción en el umbral que define el servicio como inaceptable (siguiendo la escala MOS anteriormente presentada).

En un estudio posterior del mismo autor [74], se introduce una variable adicional, el rendimiento de la aplicación. Al considerar la evaluación de la experiencia del usuario sobre un modelo más complejo en el que, además de las métricas de retardo y pérdida de paquetes, incluye el número de peticiones de descarga para una página web, se observa una relación bidireccional entre las perspectivas de rendimiento de la red y la aplicación. En otras palabras, a mayor número de peticiones de descarga por parte de la aplicación, más retardo y probabilidad de pérdidas. Igualmente, a mayores retardos y pérdidas en la red, más necesidad de solicitar retransmisión de descargas, lo que incide negativamente en la valoración del usuario, tal y como representa la gráfica bidimensional de la figura 4.30.

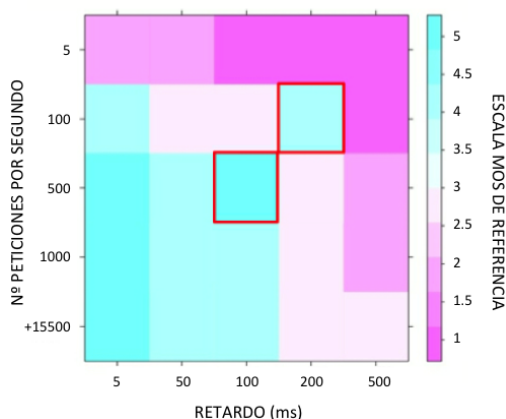


Figura 4.30 – MOS vs Retardo y peticiones por segundo (Nguyen)

Sobre esta gráfica, destaca la existencia de dos anomalías recuadradas en rojo, que a priori deberían haber tomado colores más rosados, lo que sugiere la posibilidad de que el propio contenido de las webs visitadas haya tenido alguna influencia en dicha valoración.

⁹ Cita literal del artículo [68]: 'Rule of nature'

Por ello, para el experimento abordado en el estudio que se recoge en estas páginas [68], se introduce una métrica adicional al respecto de los contenidos. En concreto, se definen cuatro tipologías diferentes en función de su complejidad, numerados de 1 a 4 en términos objetivos de mayor volumen de descarga (figura 4.31).

Como breve apunte, mayor volumen no implica mayor número de peticiones de descarga. Pueden ser el mismo número de objetos pero de mayor tamaño cada uno. Asimismo, no se consideran en el estudio las implicaciones psicológicas de los contenidos, sino que se asume en los participantes la capacidad de diferenciar incrementos o decrementos en los tiempos de descarga.

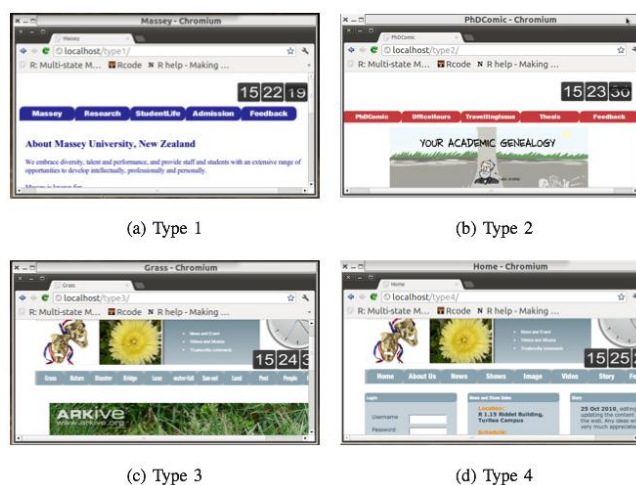


Figura 4.31 – Los 4 tipos de contenidos del experimento de Nguyen

Análisis de los resultados

La tabla 4.10 muestra las opiniones de los 12 participantes en el variante del experimento con factor de contenido, para un mismo caso de retardo y número de peticiones por segundo, para cada uno de los 4 tipos de contenido evaluados.

MOS	CONTENIDO			
	TIPO 1	TIPO 2	TIPO 3	TIPO 4
1	4	10	19	21
2	25	26	29	32
3	51	48	36	26
4	18	17	15	23
5	10	7	9	6
TOT	108	108	108	108

Tabla 4.10 – Resultados de valoración MOS por contenidos (Nguyen)

Pese a que no se consiguen valoraciones de MOS altas en ninguno de los contenidos, sí se observa cómo, con el aumento del tamaño de los objetos a descargar, existe una tendencia de evaluación negativa, especialmente en la parte más baja de la escala de opinión. Esta tendencia se puede observar claramente en la figura 4.32, donde se representan los valores de la tabla junto con la estimación del tiempo de descarga en segundos para una condición de ancho de banda de 512 Kbps (figura 4.32, REF).

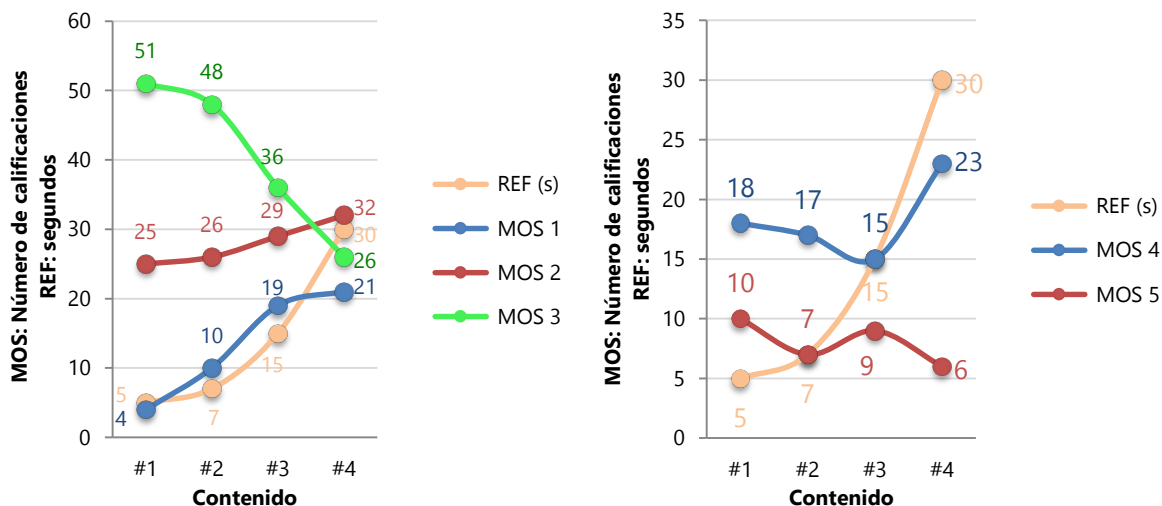


Figura 4.32 – MOS por tipo de contenido (Nguyen)

A mayor volumen de descarga, el valor MOS=3 pierde 'votantes', mientras los ganan las condiciones de MOS<3. Curiosamente, la tendencia no se mantiene con valores MOS>3, lo que evidencia una relación entre el contenido y el rendimiento de la aplicación, más allá de su tamaño.

Conclusiones

Con este experimento, Nguyen [68] certifica la existencia de relación entre el volumen de descarga y la evaluación de los usuarios. En cuanto a las anomalías visualizadas, la publicación recomienda la consideración de efectos psicológicos en el análisis de resultados relacionados con contenidos, y concluye afirmando que los usuarios prestan más atención al contenido que a la conectividad.

Si bien todo lo anteriormente expuesto puede ser cierto en muchos casos, se puede hacer una lectura adicional no considerada en la publicación y que tiene tintes objetivos, más allá de los factores psicológicos a los que se hacen referencia: La publicación [68] no considera la utilidad, la aplicación específica o la inmediatez de uso del contenido descargado.

Es cierto que el grado de utilidad de un contenido puede ser una medida subjetiva en función del interés específico de la persona que realiza el experimento, pero también es una característica objetiva, si la analizamos desde la perspectiva del rendimiento que aporta dicho objeto a la estructura de la web visualizada, tanto en el uso para el que se concibe, como en el orden en el que se carga respecto al resto de componentes de la web.

Por ejemplo, si el uso que se da a una simple fotografía (supongamos incluso que de gran tamaño) es el de ilustrar el fondo de una página web de un portal de noticias, el usuario puede valorar positivamente la experiencia de navegación por dicha web, ya que la utilidad principal de los contenidos de dicha web está en el texto y no en la fotografía. Por el contrario, si esa misma fotografía es el foco intencionado de una web, como el caso de una galería fotográfica, la valoración sí se verá, esta vez, muy influenciada por el tiempo de carga.

En otras palabras, se puede afirmar que, al respecto del contenido, entre los factores que determinan la experiencia del usuario no está únicamente lo que tardan en descargarse los objetos, sino el uso que se le da mientras se descargan, además de los factores psicológicos en los que valorar el interés o el gusto del usuario por el propio contenido visualizado. El ejemplo evidente de esto lo tenemos en los portales de *streaming* como *YouTube*, en el que se puede interactuar con los comentarios y disfrutar de los vídeos mientras aún se están descargando. Esta experiencia de navegación puede ser valorada positivamente por los usuarios, incluso a pesar de que el contenido visualizado no sea de su agrado.

Estudio 4: Tiempos de espera en la QoE para servicios web

De los estudios analizados hasta el momento se desprende el concepto de que el indicador que aparentemente influencia más la experiencia del usuario en la navegación web es el tiempo, en el sentido más amplio de la palabra. Concretamente, se ha hecho referencia a tiempos de retardo, tiempos debidos a la retransmisión de paquetes, tiempo de descarga natural de un determinado contenido o conjunto de objetos... En definitiva, tiempos de espera que el usuario percibe de manera diferente para contextos diferentes y en estados psicológicos diferentes.

El estudio que cierra el apartado dedicado a las investigaciones existentes en torno a la QoE en el Servicio de Navegación Web es, en sí mismo, un estado del arte de experimentos y resultados en los que se pone especial atención al concepto de tiempo de espera [70].

Como breve inciso, cabe destacar que los autores de *“Waiting times in Quality of Experience for web based services”* son miembros participantes en el consorcio Qualinet [31], y han sido recurrentemente referenciados en varios capítulos del presente documento [33][38][47], influenciando incluso el estudio de QoE que se recoge en esta tesis de máster [8][9][10], por lo que el autor de este trabajo estima de gran interés sus aportaciones y recomienda la lectura detallada de esta publicación [70], además de sus otros trabajos relacionados.

Aspectos psicológicos del tiempo de espera

A lo largo de este capítulo se ha hecho referencia a factores psicológicos como modificadores de la percepción del usuario en su experiencia de navegación, aunque no se ha realizado un análisis en profundidad en este aspecto. La realidad es que el factor humano (HF – Human Factor) [53][75] sigue siendo un terreno relativamente inexplorado y las investigaciones interdisciplinarias al respecto de los aspectos psicológicos que influyen en los participantes no se han abordado más allá de relaciones superficiales.

En cuanto a la percepción del tiempo de espera, Egger y compañía¹⁰ introducen la importancia psicológica del concepto de manera similar a la que Schindlholzer hacía referencia al paradigma de la disconfirmación [42]:

¹⁰ Al igual que en otras publicaciones citadas anteriormente, se hará referencia, a partir de este momento, al autor principal del documento en representación de todos los autores.

Una vez que el usuario experimenta la rapidez, se acostumbra a ella y amolda sus expectativas a mantener esa rapidez todo el tiempo. Toda percepción posterior de tiempos de espera similares, serán interpretados por el usuario como algo normal, mientras que un aumento del tiempo de espera se percibe como algo particularmente molesto.

Tan importante es este concepto, que muchos operadores de telecomunicaciones han cambiado su paradigma de calidad de analizar *“qué distorsiones son tolerables para asegurar cierto grado de satisfacción de usuario”* a analizar *“qué tiempos de espera son suficientes para asegurar cierto grado de satisfacción de usuario”* [70].

Esto que puede parecer un factor trivial, no lo es en absoluto. Existe un desajuste entre la duración objetiva de un suceso y el tiempo realmente percibido por el sujeto. El contexto en el que se produce el suceso, el estado anímico de la persona, las emociones o sentimientos que producen los contenidos que está visualizando en su navegador, entre miles de otras condiciones cognitivas, sensoriales y físicas, hacen que un mismo período de tiempo parezca más largo o más corto de lo que realmente es (figura 4.33) [76].

Mientras que la duración real refleja el tiempo objetivo, la duración percibida refleja tiempo psicológico subjetivo, susceptible de ser distorsionado por todos múltiples factores. Todos esos detalles y características de estímulos temporales tienen que ser considerados en la relación existente entre los tiempos de espera y la QoE. El problema es que desde el punto de vista tecnológico, apenas hay incursiones en el mundo de la psicología. Por ello, desde el documento de Egger [70], se miran con especial atención los avances a este respecto [77].

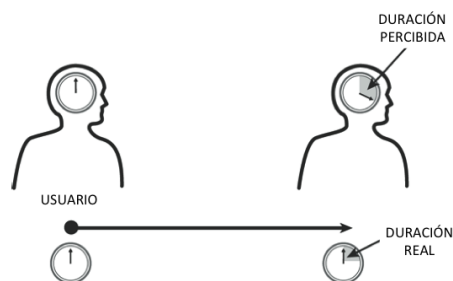


Figura 4.33 – Duración percibida vs Duración real (Seow)

Web QoE: Esperar para navegar

El concepto Web QoE hace referencia a la Calidad de Experiencia de servicios interactivos, accesibles a través de un navegador mediante el uso del protocolo HTTP, ya sea en navegación web, descarga de ficheros o gestión de emails.

Al respecto del servicio que se trata principalmente en esta tesis, la navegación web, Egger [70] refiere varios estudios en los que se identifica el tiempo de espera del usuario final como la clave determinante del valor de la QoE. A este respecto, analiza también la recomendación ITU-T G.1030 [45] y estudios de validación como el de Ibarrola [52], y concluye que en los últimos años, el concepto de navegación ha evolucionado desde la consulta puntual de una página, al concepto de sesión como concatenación de eventos de visualización de páginas a través de múltiples portales con diversas tipologías de contenidos.

El usuario ya no percibe la navegación web como una búsqueda aislada, si no como la secuencia total a modo de flujo inmersivo, lo que desemboca en la necesidad de revisión de la citada recomendación para dar cabida a esta nueva realidad (figura 4.34). El propio Egger es, de hecho, el promotor de los dos *drafts* de la ITU [8][9][10] que se toman como punto de partida en el desarrollo del presente estudio y que se detalla a partir del capítulo 5.

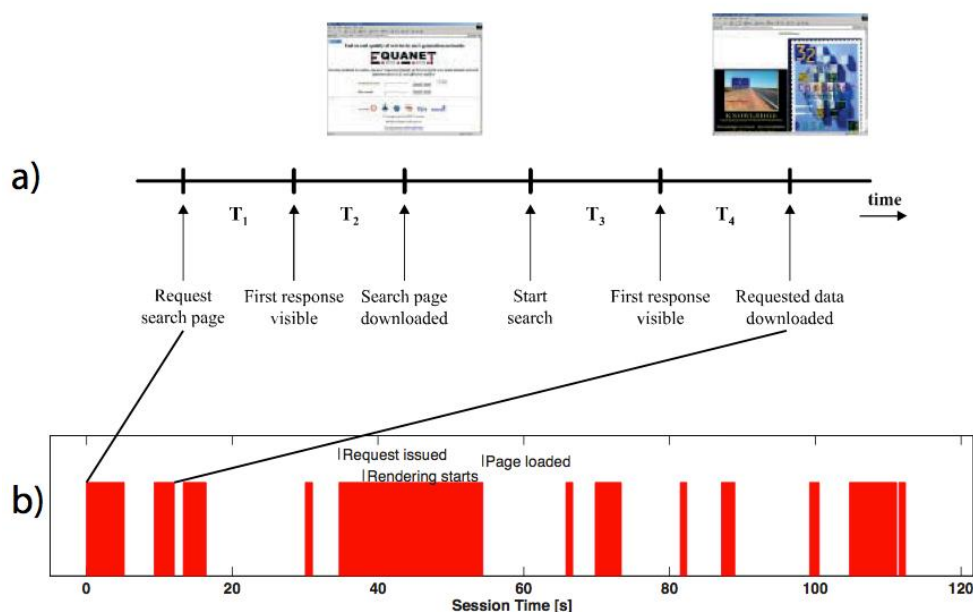


Figura 4.34 – Prueba ITU-T G.1030 [a] dentro de una sesión con distintos tiempos de espera [b] (Egger)

El documento [70] refiere dos posibles acercamientos a la evaluación de la QoE en el servicio de navegación web clásico:

- **Modo A:** Solicitar al participante la navegación por una o varias páginas web sin limitar el tiempo de prueba, cuantificando la suma de tiempos asociados a la realización de dicha tarea. Este caso es el que se corresponde con el modelo definido en la ITU-T G.1030, representado en la figura 4.34 con la letra 'a'.
- **Modo B:** Definir tiempos de sesión finitos en los cuales se solicita al usuario la realización de una determinada tarea bajo unas condiciones específicas de calidad intrínseca, incidiendo de manera directa en los parámetros de la red. Este caso es el que se representa con la letra 'b' en la figura 4.34.

Mientras que el modo A considera el tiempo de sesión total como variable independiente para la cual se evalúan los niveles de MOS obtenidos, el modo B hace lo propio con los parámetros intrínsecos de la red.

Es evidente, a la vista de esta comparativa, que el caso más realista de navegación lo ofrece el segundo modo, que, además, permite establecer relaciones QoS-QoE a partir de los parámetros de red que se definan como condición de evaluación. Sin embargo, dependiendo de la intención del estudio, nos puede interesar realizar un experimento del tipo A, en el que se tiene control sobre los tiempos de espera.

Independientemente de los ejemplos presentados en el documento para ambos casos [70], y que se dejan como consulta al interés del lector, sí se hace un breve resumen de las consideraciones y conclusiones al respecto de la evaluación de la QoE en el servicio web:

- **Estímulo frente a indicador intrínseco de calidad**

Del estudio 'modo A' en el que se busca establecer una relación entre un estímulo (el tiempo de espera) y la QoE, se observa la existencia de una relación lineal en base a una escala logarítmica del par estímulo-percepción (figura 4.35, izquierda). Sin embargo, esa misma relación no se da al utilizar parámetros intrínsecos (modo B) como condición. El ancho de banda no es un estímulo desde el desde el punto de vista psicológico, por lo que requiere de un ajuste que permita definir una interdependencia exponencial entre la QoE y parámetros de QoS como el citado ancho de banda (figura 4.35, derecha).

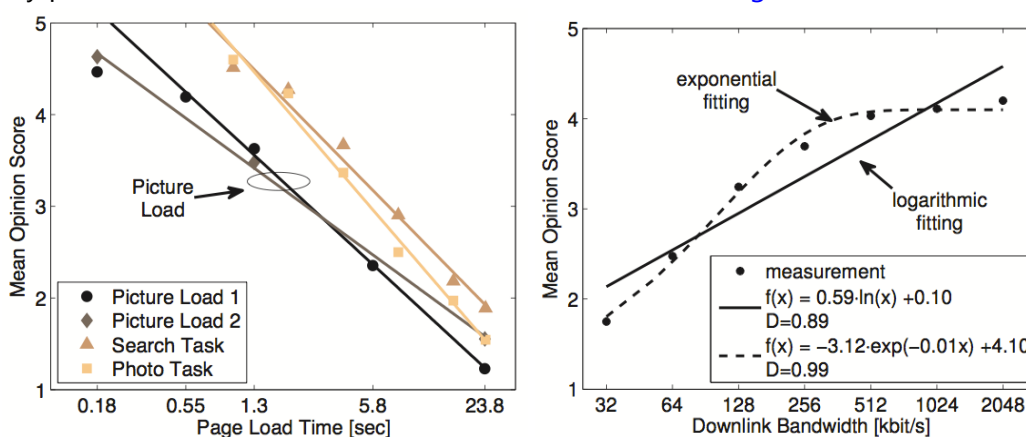


Figura 4.35 – Relación Estímulo-MOS [izq] vs. Relación KPI-MOS [dch] (Egger)

- **Tiempo frente a ancho de banda**

El mero hecho de utilizar el ancho de banda en lugar del tiempo como condición del experimento muestra las complicaciones existentes en la relación QoS-QoE. Ya no se trata de modificar el tiempo y obtener un valor de QoE de manera directa, sino que requiere de la consideración de múltiples factores. No existe una relación lineal entre el tiempo y el ancho de banda, debido, entre otras cosas, al funcionamiento interno de los protocolos utilizados en la navegación web (la fase de *slow start* TCP, la influencia de la congestión y el control de flujo en la red...) y a retardos propios a nivel de aplicación. Los distintos motores de los navegadores existentes introducen retardos propios diferentes.

- **Tiempo de carga de la aplicación frente al percibido**

Los puntos anteriores ya muestran la existencia de una no-linealidad entre los parámetros intrínsecos y los tiempos de carga (tanto a nivel de red como a nivel de aplicación). Incluso en el supuesto de que pudieran modelizarse o mapearse ambos conceptos, no se estarían considerando las variaciones en la percepción del tiempo real de carga por parte del participante (figura 4.33). Esto supone la inclusión de consideraciones psicológicas difíciles de parametrizar, complicando aún más el reto de medir o estimar el tiempo de espera tal y como lo percibe el usuario.

Video streaming: Esperar mientras se consume

Con respecto a los modelos de navegación web 2.0, como el de la transmisión de vídeo a través de portales como *YouTube*, el documento [70] recoge una extensa comparativa de los diversos modos de funcionamiento, y analiza los factores clave que influyen en la cuantificación de la QoE para este tipo de servicios.

De nuevo, sin entrar en detalle en los ejemplos contemplados [70], puesto que además esta variante no es objeto de estudio de la presente tesis, sí cabe mencionar la enorme complejidad que supone considerar, sobre la ya compleja base del flujo navegacional clásico de las páginas anteriores, nuevas implicaciones tanto psicológicas como de rendimiento en conceptos como la carga parcial del vídeo antes de su reproducción, la congelación de la imagen transferida ante la congestión de la red, la disponibilidad de un *buffer* en recepción de tamaño insuficiente o la aplicación de modelos adaptativos de resolución de vídeo.

De este último caso ya vimos en el estudio de Zinner [54] del apartado 4.3, que si bien una menor resolución en condiciones de congestión de la red permite una transferencia más fluida de los contenidos, también tiene una serie de implicaciones perceptuales en el consumidor que pueden influir en una mala valoración del servicio, a pesar de ofrecerse un rendimiento adecuado a nivel de red.

Conclusiones

Como conclusión a este extenso análisis de la importancia de los tiempos de espera en la evaluación de la QoE [70], destaca precisamente la necesidad de establecer relaciones interdisciplinares con expertos en psicología de cara a abordar el concepto de QoE en el sentido más completo del mismo. Al respecto de la relación lineal presentada entre los tiempos de espera y la QoE, puede ser una herramienta muy interesante en el diseño de mecanismos adaptativos que procuren mejorar la experiencia de usuario a través de la monitorización precisa de los tiempos que se desean mejorar.

Sin embargo, si lo que se quiere es evaluar la QoE respecto a condiciones específicas de red, en términos de parámetros de rendimiento clásico como el retardo, la pérdida de paquetes o el ancho de banda (todos ellos contribuyentes en diferente proporción según el contexto, al concepto de tiempo de espera), es necesario abordar una experimentación que contemple una experiencia de navegación lo más natural posible al participante. Como ya hemos visto, no sólo en este estudio [70], sino también en los anteriores de este mismo apartado [59][64][68], dicha experiencia de navegación es tremendamente difícil de monitorizar, habida cuenta de la enorme cantidad de factores que influyen en el sistema y que convierten la complicada relación QoS-QoE en un problema prácticamente inabordable.

En opinión personal del autor de este documento, lo más sensato parece ser abordar de manera independiente pero progresiva, los distintos factores que influyen en la QoE, como los que se definen en los *drafts* de referencia [8][9][10]. A fin de cuentas, es la línea que parece estar siguiéndose en organismos de estandarización tan importantes como la ITU.

4.5. Conclusiones sobre el estado del arte

A lo largo de este capítulo hemos podido desgranar las diferentes definiciones del término QoE a lo largo de los últimos años. Si bien es cierto que hay cierta unanimidad en utilizar la definición de referencia de la recomendación ITU-T P.10/G.100 [2], vemos que hay múltiples formas de intentar abordarla.

En los modelos y metodologías expuestos, la QoE pasa de asimilarse prácticamente a la percepción del usuario, en los casos más sencillos, a incluir relaciones con la Calidad de Negocio (QoBiz) o distinciones entre los conceptos de usuario y cliente de un servicio, en los modelos más completos. Incluso, se aborda la idoneidad de utilización de métricas objetivas en su evaluación. A pesar de ello, en todos los casos se aprecia la tendencia creciente actual de estudiar la QoS desde un punto de vista centrado en el usuario.

Habida cuenta del título del presente trabajo, merece mención especial la metodología definida en la recomendación ITU-T G.1030 [45] al respecto del estudio simple de la búsqueda de contenidos en una única página web. Teniendo en cuenta la evolución del servicio de navegación web en los últimos años, y, en particular, de los contenidos ofrecidos (imágenes, vídeos, servicios de mapas...), se requiere una actualización en la evaluación de la percepción y satisfacción del usuario respecto a la navegación web, mediante la definición de una metodología que permita evaluar la QoE del usuario en un flujo de navegación real.

Por ello, en el seno de los grupos de trabajo de la ITU se ha planteado recientemente un borrador que propone abordar las carencias de la ITU-T G.1030, y que constituye la base sobre la que se realiza el presente trabajo. Como es evidente, la descripción base de la metodología de evaluación de la QoE descrita en el citado *draft* P.STMWeb [8][9] no ha sido desgranada en estas páginas, ya que se le reserva una mayor atención en siguientes capítulos. Sin embargo, si se ha pretendido aportar, como se verá más adelante, alguna de las consideraciones abordadas en el estado del arte. De hecho, la metodología definida en la ITU-T P.800 [48] es probablemente el ejemplo de aplicación más reproducido y, como podrá verse más adelante, influencia en gran medida el trabajo presentado.

Quedan fuera de estudio, por razones obvias, las metodologías de carácter objetivo, ya que el *draft* a validar requiere de una consulta directa a participantes. Sin embargo, desde este trabajo se quiere hacer constar el alto grado de interés que puede suscitar este acercamiento a la parametrización de factores humanos para la validación mediante métricas objetivas de servicios, aplicaciones o incluso recomendaciones.

El trabajo del grupo ETSI STF 354 ha sido bastante prolífico en los últimos años, incluyendo varios informes técnicos y una guía de referencia (ETSI EG 202 670 [75]) a este respecto. Sin embargo, a título personal, el autor considera que aún hace falta mucho recorrido para poder disponer de métricas adecuadas para analizar *únicamente* la QoE desde un punto de vista objetivo. Incluso, es probable que muchos servicios sigan aportando mayor información a partir de la percepción directa de los usuarios.

Por último, en cuanto a los trabajos referidos al respecto del estudio de la QoE en el Servicio de Navegación Web, cabe destacar la meticulosidad con la que se tratan las métricas intrínsecas de QoS y la búsqueda de relaciones subjetivas a través de la opinión de los usuarios participantes.

El Servicio de Navegación Web tiene muchas áreas de influencia, desde la propia red de acceso, hasta la disponibilidad del servidor que ofrece los contenidos, pasando por todo el equipamiento intermedio de la red. No es posible realizar, a día de hoy, un análisis que contemple la totalidad de los factores que influyen en la experiencia del usuario. Internet es un elemento vivo y cambiante. Cuando un usuario navega por uno o varios portales web, puede parecer que únicamente hace uso de los protocolos habituales (TCP, HTTP...). Sin embargo, existen multitud de comunicaciones adicionales, como peticiones de resolución de dirección (DNS), o simples balizas de señalización entre otros nodos de la red que, sin tener relación directa con nuestra navegación, influyen incrementando retardos, congestionando la red, provocando retransmisiones...

Desde el estudio que se describe en el presente documento, se ha buscado dar un paso hacia la evaluación de la QoE, procurando controlar al máximo todos esos factores externos. Por ello, en el presente trabajo, se ha tomado la determinación de llevar a cabo el experimento diseñado en un entorno de laboratorio controlado en el que poder extraer conclusiones respecto a unas condiciones de QoS muy concretas, como se puede empezar a ver a partir del próximo capítulo.

5. ARQUITECTURA PROPUESTA

Una vez analizados diversos modelos y metodologías para la evaluación de la QoE, nos centramos principalmente en aquellos relativos al servicio que se aborda en el presente trabajo. En concreto, tal y como se adelanta en las conclusiones del capítulo anterior, se toman como referencia las recomendaciones ITU-T G.1030 [45] e ITU-T P.800 [48].

El objetivo del estudio es, en definitiva, el de poder evaluar o determinar la QoE del servicio web mediante la presentación al usuario de una experiencia de navegación completa a través de varias páginas web con distintos tipos de contenidos, con las que puede interactuar con total libertad, bajo el control de diferentes condiciones de test que se van sucediendo a lo largo de la duración de la prueba.

Este estudio toma como punto de partida el *draft* P.STMWeb [8][9] surgido en el grupo de trabajo SG12 de la ITU, cuya pretensión es la de abordar los aspectos no contemplados en la recomendación ITU-T G.1030, como ya se ha comentado en el capítulo anterior.

5.1. Conceptos previos

A lo largo de las próximas páginas y capítulos se hará referencia a una serie de conceptos cuyo significado o alcance puede ser complicado de entender sin una definición clara. Algunos ya han aparecido en capítulos anteriores aunque sin detallar claramente el alcance de su significado. Es por ello que, en este punto, se presta atención a dichos conceptos:

- **Condición:**

Período de tiempo en el cual se aplica una determinada configuración de QoS para su posterior evaluación por parte del usuario. El punto de aplicación de dicha configuración de QoS puede darse en cualquier punto de la red entre el equipo de usuario y el servidor web que aloja los contenidos. Típicamente se modificarán los parámetros de la red en el enlace directo que une el equipo de usuario con el resto de la plataforma de evaluación, como veremos más adelante. El tiempo de condición termina cuando al usuario se le muestra un formulario en el que evaluar uno o varios aspectos de calidad al respecto de la condición experimentada.

- **Tarea:**

Dícese del conjunto de instrucciones que se dan al usuario sobre cómo acceder a determinada información u contenido durante el tiempo de una condición de test. Este concepto atañe más concretamente a la fase de metodología y plan de pruebas del capítulo siguiente, aunque se introduce en este punto para mejor comprensión.

- **Contenido:**

Distintas tipologías de objetos estáticos o dinámicos que se presentan al usuario durante una sesión de navegación web, y con los que éste puede interactuar.

• **Sesión (de navegación web):**

Una sesión de navegación web se inicia en el momento en el que, de manera voluntaria, el usuario realiza una petición de búsqueda y acceso a una página web, con el objetivo de realizar una determinada tarea. La finalización de la sesión es típicamente el instante en el que dicha tarea finaliza. Un ejemplo de sesión puede ser la participación del usuario en un flujo de navegación web que encadene la visita e interacción con varias páginas webs de manera consecutiva.

• **Factor de influencia:**

Durante las sucesivas condiciones que se van presentando al usuario a lo largo de la duración total del experimento, existen elementos de diferente índole que pueden modificar parcial o completamente la experiencia del usuario. Dichos elementos son conocidos como factores de influencia.

Dada la importancia de los factores de influencia en el estudio de la QoE, es necesario realizar una identificación de los mismos (figura 5.1) [78] de cara a evaluar el grado de acción que tienen sobre la percepción y experiencia del usuario, por encima de los parámetros que realmente se quieren evaluar. Para ello, se tiene muy en cuenta el borrador G.QoE-Web [10] en el que se definen los factores de influencia en el entorno de navegación web.

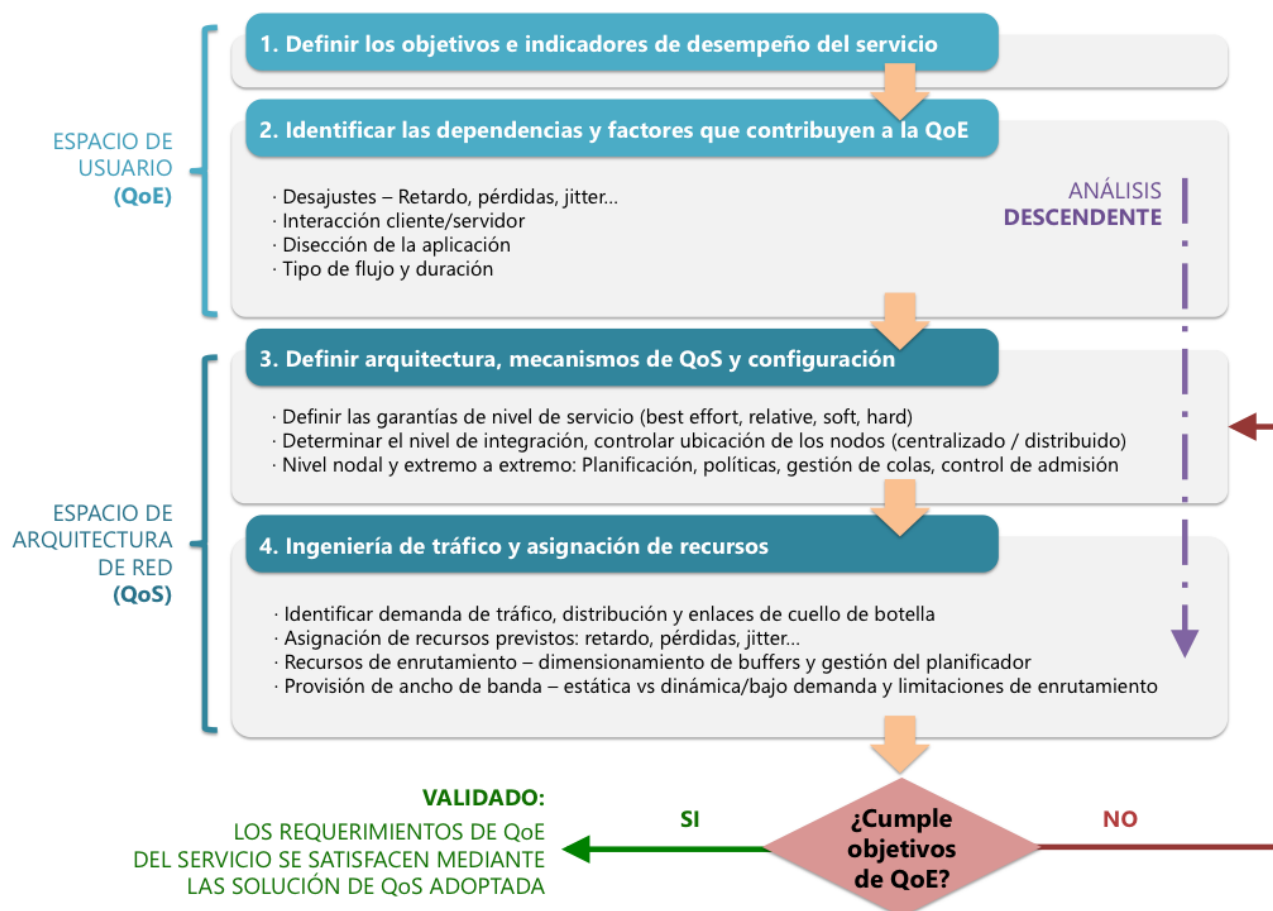


Figura 5.1 – Relación QoS/QoE en la definición de un sistema de validación (S.H. Jeong)

Factores de influencia

En general, se pueden agrupar los factores que influyen en el valor de la QoE en tres categorías principales [33]:

- **Factores de influencia del usuario**

El borrador no especifica este tipo de factores, aunque el autor del presente trabajo considera que en esta categoría podrían incluirse aspectos psicológicos del usuario, aspectos económicos respecto al precio del servicio, e incluso aspectos cognitivos y físicos de la persona como, por ejemplo, su capacidad visual (miopía, vista cansada...).

- **Factores de influencia del contexto**

En esta categoría de factores que pueden influir en el comportamiento de usuario y su valor de QoE se encuentran las siguientes dimensiones:

- *Ubicación desde la que se accede al servicio:* Domicilio, oficina, café...
- *Grado de interactividad:* Alto, medio, bajo
- *Tipo de tarea:* Entretenimiento, negocio,...
- *Urgencia de la tarea:* Urgente o casual

- **Factores de influencia del sistema**

Probablemente, los factores más relacionados con la calidad intrínseca del sistema, divididos a su vez en factores de contenido, de red y relativos cliente.

- Factores de red:

Contribución de la red al tiempo de transacción (ITU-T G.1040 [50]).

Capacidad disponible en el enlace (ITU-T Y.1541 [79])

Tiempos de respuesta del servidor...

- Factores de contenido:

Estos factores van más allá de la tipología de contenido (vídeo, imágenes, texto), que suele ser lo que comúnmente se relaciona con la QoE. También se han definido factores asociados como por ejemplo:

Número, tipo, tamaño y orden de objetos mostrados en pantalla.

Orden de aparición de los distintos elementos.

Estrategia de renderizado del navegador usado...

- Factores de cliente:

Típicamente factores relacionados con el equipamiento de usuario, tales como el sistema operativo, capacidad de procesamiento del equipo, la configuración de la pila TCP/IP en el equipo, el navegador utilizado...

A la vista de estos factores, es importante considerar aquellos que puedan tener influencia directa en los resultados, puesto que las pruebas se realizan en un entorno de laboratorio controlado con un equipamiento y condiciones determinadas. La experiencia previa del usuario en este servicio, teniendo en cuenta el lugar habitual en el que navega por internet, la velocidad contratada o el uso que hace del servicio pueden influenciar los resultados.

5.2. Arquitectura propuesta en P.STMWeb

Partiendo de estos factores, el *draft* P.STMWeb [8] define una arquitectura de red para la evaluación subjetiva de la navegación web en un entorno de laboratorio controlado (típicamente una red de área local), en el que figuran los siguientes dispositivos (figura 5.2):

- **Terminal del cliente (UD – User Device):** Equipo con conectividad de red que dispone de un navegador desde el que el participante accederá a los contenidos disponibles de manera libre, pudiendo alternar páginas visitadas, introducir comentarios en fotografías, etc. En este mismo equipo se presentará la encuesta de valoración subjetiva al usuario.
- **Emulador de red (EMU):** Puesto que estamos en un entorno de red controlado, el emulador de red será el encargado de introducir variaciones en los parámetros de la red que se consideren en el estudio. Dichos parámetros constituyen los indicadores de calidad (KQI) susceptibles de influir en la percepción del usuario.
- **Router:** Proporciona conectividad a Internet a los dispositivos conectados a la red LAN (Local Area Network) desplegada.

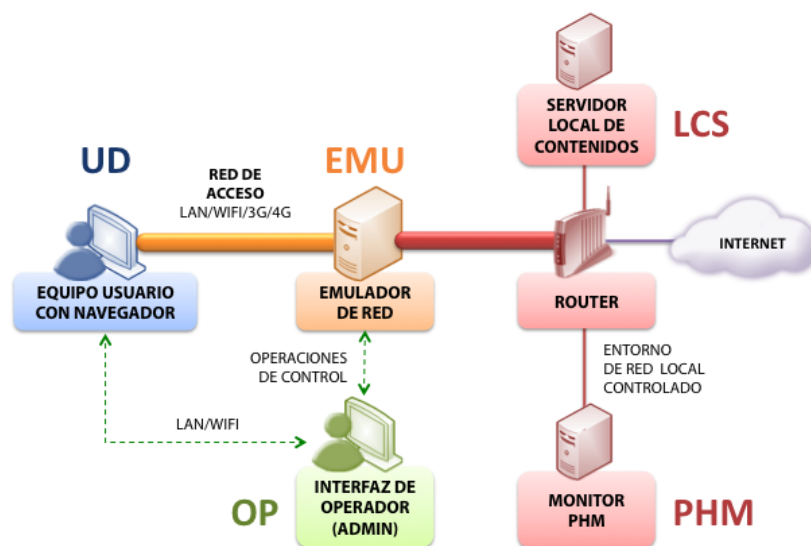


Figura 5.2 – Arquitectura de red propuesta para la evaluación de QoE (P.STMWeb)

A dicha lista se añaden también una serie de equipos opcionales que completan la arquitectura propuesta en el borrador, en función de la variante metodológica que se aplique. Veremos más adelante que la metodología propuesta contempla la utilización de contenidos tanto alojados en un servidor local a la red como en el dominio de Internet.

- **Monitor PHM (Page Health Monitor):** Si se estima la realización de pruebas con contenidos alojados fuera de la red LAN, este equipo es el encargado de identificar potenciales influencias de la red en su tramo no controlado. Es decir, debe verificar el correcto funcionamiento de los portales web que se utilizarán en las pruebas de navegación, así como posibles congestiones u otras complicaciones de la red, para evitar que afecten, en la medida de lo posible a la evaluación de los participantes.

- **Servidor Local de Contenidos (LCS):** Si se realizan pruebas de navegación sin acceso real a Internet, éste servidor será el encargado de mostrar los contenidos al participante, de la manera más real posible. Para ello deberá contar con uno o varios portales en los que el usuario pueda interactuar libremente durante el experimento.
- **Interfaz de Operador (OP):** Adicionalmente se puede contar con un equipo que permita monitorizar el correcto funcionamiento de los dispositivos principales del sistema.

Volviendo al equipamiento de usuario, se especifica además que el terminal del cliente (UD) puede ser cualquier equipo con capacidad para la navegación web y que, salvo que el objetivo sea el de analizar expresamente las variaciones en la experiencia entre múltiples dispositivos, los participantes deberán realizar las pruebas siempre con un mismo tipo de equipo, ya sea un ordenador de sobremesa, un portátil, una tableta o un teléfono móvil. Asimismo, se especifica la recomendación de utilizar el mismo software en todos los UD disponibles (Sistema operativo, versión de navegador, etc.).

De esta manera, se unifican todos los posibles factores de influencia que pudieran tener cualquier tipo de efecto en la percepción de los participantes que realizasen el experimento a través de distintos dispositivos. Ante distintas situaciones de partida, los resultados obtenidos podrían no ser comparables.

5.3. Arquitectura adoptada en el estudio

A partir de la propuesta de arquitectura definida en el *draft* P.STMWeb [8], se ha desarrollado una plataforma de validación cuya arquitectura comparte las características principales a excepción del monitor PHM, ya que se ha limitado la navegación web del experimento a los contenidos disponibles en el LCS. La siguiente figura representa la red desplegada para la realización del presente estudio:

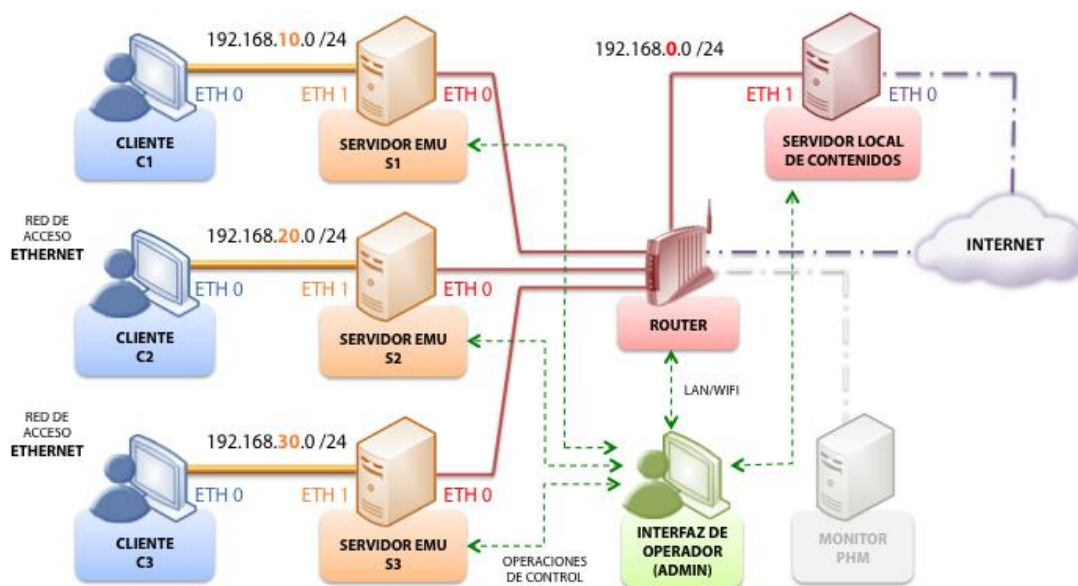


Figura 5.3 – Arquitectura de red desplegada para la evaluación de QoE

Como se puede ver, la red consta de tres terminales de usuario conectados directamente a tres servidores que actúan de emulador de red, ante la imposibilidad de disponer de equipamiento más específico. Dichos tres servidores disponen del mismo *hardware* y *software* instalado y cada uno gestiona únicamente el enlace directo con un terminal UD.

De esta manera, con las configuraciones internas de rutado de los servidores EMU, los clientes pueden acceder a la red LAN en la que se encuentra el servidor de contenidos LCS. El *router*, por su parte, no realiza ninguna labor de filtrado de peticiones desde los clientes hacia Internet a pesar de que las pruebas se realizan con contenidos del LCS, para evitar procesamiento innecesario de paquetes en el único equipo que rige la topología de la red.

Todas las comunicaciones mostradas en la [figura 5.3](#) se realizan a través de Ethernet, a excepción del equipo de administración (OP), que accede a la red a través del punto de acceso inalámbrico del *router* para la gestión de los cuatro servidores de la plataforma.

Nótese también que cada enlace UD-EMU constituye una subred diferente a la red LAN base del sistema, a modo de red de acceso de cada terminal UD a los contenidos del LCS y al resto de Internet. Este enlace es el punto de acción del EMU. Para cada condición del experimento a evaluar, el servidor emulador se encarga de variar los parámetros intrínsecos del enlace que han sido identificados como indicadores de desempeño (KPI) a partir de los indicadores de calidad (KQI) considerados de más relevancia en el estudio.

Asimismo, se han tenido en cuenta los factores de influencia del sistema, tanto en el dimensionamiento de los enlaces UD-EMU, como en el resto de la red LAN, asegurando que, en el caso más exigente de utilización de la red, el tráfico agregado de los tres enlaces cliente-emulador no supera ni la máxima capacidad de la red LAN, ni el umbral de funcionamiento óptimo del *router* en la gestión del tráfico, ni la capacidad del servidor LCS en la gestión de las transacciones para la obtención y visualización de los contenidos.

Equipos de usuario (UD)

En cuanto a los factores de influencia del sistema referidos al cliente, se ha procurado que todos los equipos tuvieran las mismas prestaciones computacionales.

Dada la limitación en el número de equipos disponibles en el laboratorio en el que se han llevado a cabo las pruebas, se ha procurado que, al menos, todos los equipos tuviesen capacidad suficiente (muy por encima de la necesaria), para llevar a cabo la navegación web en primer plano, sin otros procesos consumiendo recursos en *background*.

En todos los casos se ha contado con el mismo sistema operativo y navegador, para evitar implementaciones distintas de la pila de protocolos TCP/IP que pudieran variar la percepción de la navegación a los participantes.

El resto de características contempladas se presenta en la [tabla 5.1](#), a continuación.

	Cliente #C1	Cientes #C2 & #C3
Hardware (PC)		
Marca	Dell	Dell
Modelo	OptiPlex 790	OptiPlex 745
Procesador	Intel Core i3 – 3.10 GHz	Intel Core 2 Duo – 2.13 GHz
Memoria (RAM)	4 GB	2 GB
Interfaz de Red	Ethernet 10/100	Ethernet 10/100
Software		
Sistema Operativo	Windows 7 Professional SP1 – 32 Bit	
Navegador	Firefox 24.0 – Todas las cachés deshabilitadas	
Monitor		
Pantalla	17" (4:3)	
Resolución	Pantalla: 1280 x 1024 Navegador: 1200 x 650 – El usuario puede redimensionar	
Luminancia	250 cd/m2	
Calibración	Todas las pantallas calibradas a D65	

Tabla 5.1 – Características de los equipos de usuario de la arquitectura adoptada

Como último apunte, cabe destacar los factores de influencia de contenido considerados en los terminales de usuario (UD). Puesto que el plan de pruebas ([capítulo 6.2](#)) consiste en la evaluación sucesiva de diferentes condiciones de navegación web en el mismo conjunto de portales de contenidos alojados en el LCS, se ha tomado en consideración deshabilitar la caché de los navegadores. De esta manera, se evita que objetos precargados en anteriores sesiones de navegación puedan influir en el caso concreto de evaluación.

Servidores EMU

A diferencia de los equipos de usuario, se ha hecho el esfuerzo por que los tres servidores de emulación de red tuvieran los mismos componentes *hardware* y *software*, como establece la [tabla 5.2](#), para evitar diferencias en los cambios introducidos en cada enlace UD-EMU.

	Servidores EMU #S1, #S2 & #S3
Hardware (PC)	
Marca	Dell
Modelo	OptiPlex GX 620
Procesador	Intel Pentium 4 HT
Memoria (RAM)	2 GB
Interfaz de Red	Ethernet 10/100 – Enlace UD-EMU Ethernet 10/100 – Enlace EMU-ROUTER
Software	
Sistema Operativo	Ubuntu Server 12.04 LTS – Linux 3.2.0
Software	Apache2, SSH (gestión remota), tc, wondershaper

Tabla 5.2 – Características de los servidores EMU de la arquitectura adoptada

En cuanto a su funcionalidad, no sólo se encarga de aplicar los parámetros de las condiciones en testeo a lo largo de la duración de cada experimento, sino que también actúa de *gateway* dando salida al UD al resto de la red LAN (y por tanto a los contenidos del LCS).

Su arquitectura interna ([figura 5.4](#)) consta principalmente de un módulo de control, un módulo de operación y una interfaz web a través de la que el usuario participante accede al experimento.

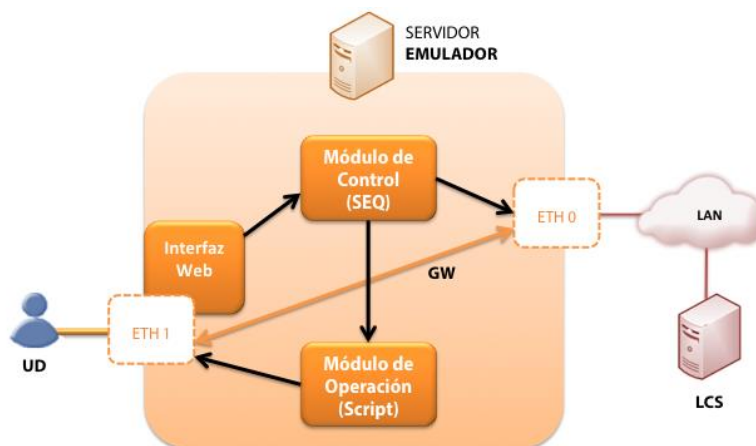


Figura 5.4 – Arquitectura funcional del Servidor Emulador (EMU)

- **Interfaz web:**

Este módulo presenta al usuario una página web a través de la cual el usuario tiene acceso a la realización del experimento. Toda la lógica de temporización de las condiciones de testeo se inicia en el momento en el que el usuario clicca por primera vez en la página de inicio que se le presenta.

- **Módulo de Control (SEQ):**

El módulo de control o secuenciador (SEQ) del EMU es el que gobierna el funcionamiento del equipo, y por tanto, del experimento en curso. A partir de la orden inicio por parte del usuario, el secuenciador intercala tiempos de navegación libre y tiempos de encuesta de opinión en los que el usuario debe valorar las condiciones recién experimentadas.

Durante dichos tiempos, el secuenciador da órdenes al módulo de operación, quien activará o desactivará las condiciones a emular en el enlace UD-EMU. Asimismo, el SEQ también se encarga de redirigir al usuario el portal web alojado en el LCS sobre el que el participante podrá interactuar durante el tiempo de navegación.

- **Módulo de Operación:**

Siguiendo la explicación del módulo secuenciador, el módulo de operación es el encargado de ejecutar los *scripts* que modifican los atributos de la interfaz de red que comparte el EMU con el UD. Para dichas modificaciones que permiten emular distintos funcionamientos en el enlace UD-EMU, se hace uso del juego de comandos **tc-netem** [69] de Linux, apoyados sobre el software **wondershaper** [80], que a su vez se basa en **tc**, y que ha tenido que ser modificado para contemplar las condiciones de testeo definidas.

- **Funcionalidad de gateway:**

Aunque no figura como módulo propiamente dicho, puesto que el EMU no es ni un *router* ni un *gateway* en sí mismo, esta funcionalidad es vital para el correcto funcionamiento de la arquitectura en su conjunto. Todas las transacciones que no están directamente relacionadas con el EMU, deben ser referidas al equipo correspondiente en la red LAN de manera transparente. Para ello, nos servimos del comando **iptables** [81], que permite establecer reglas de rutado entre los interfaces del emulador.

El lector puede encontrar más información sobre el diseño de alto y bajo nivel del EMU y cada uno de los módulos, así como la secuencia completa de funcionamiento y un extracto del código desarrollado, en el [Anexo III](#) de este documento.

Servidor Local de Contenidos (LCS)

La [figura 5.5](#) presenta la arquitectura funcional de este servidor y su interconexión con el resto de las instalaciones. Como se puede ver, consta principalmente de cuatro módulos: El servidor web, un módulo de sesión, un módulo de actualización y un módulo de persistencia, así como una interfaz web a través de la cual se accede a los contenidos.

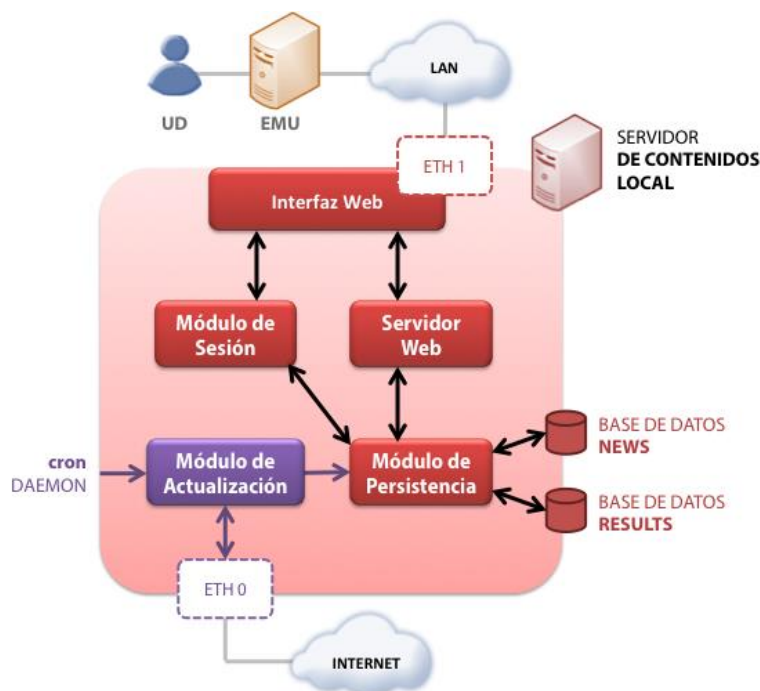


Figura 5.5 – Arquitectura funcional del Servidor Local de Contenidos (LCS)

La [tabla 5.3](#), por su parte, da cuenta de sus componentes internos. Se puede apreciar, a diferencia del LCS de la arquitectura propuesta en el *draft* P.STMWeb [8] ([figura 5.2](#)), que este servidor dispone de una segunda interfaz con conexión directa a Internet. La funcionalidad de dicha interfaz se explica en el apartado referido al módulo de actualización.

Servidor Local de Contenidos (LCS)	
Hardware (PC)	
Marca	Dell
Modelo	OptiPlex GX 620
Procesador	Intel Pentium 4 HT
Memoria (RAM)	2 GB
Interfaz de Red	Ethernet 10/100 – Enlace LCS-ROUTER Ethernet 10/100 – Enlace LCS-Internet
Software	
Sistema Operativo	Ubuntu Server 12.04 LTS – Linux 3.2.0
Servidor Web	Apache2
Registro / Sesión	Java Server Faces (JSF) - Servidor Glassfish 3.1.2
Bases de datos	MySQL

Tabla 5.3 – Características del servidor local de contenidos (LCS) de la arquitectura adoptada

- **Interfaz web:**

Se trata de un módulo básico del LCS. Cuando el usuario, de manera transparente a través del EMU, realiza una petición de cualquiera de los contenidos servidos por el LCS, la interfaz evalúa la petición y la redirige al módulo correspondiente. Todas las peticiones relacionadas con el registro del usuario en el sistema, la evaluación de las condiciones de test y el procesamiento de los resultados son redirigidas al módulo de sesión. El resto de peticiones (relativas a la navegación web en sí misma) se tratan en el servidor web.

- **Servidor Web:**

Este módulo es el núcleo del LCS. Consiste principalmente de dos partes: Los contenidos web disponibles para la navegación libre del usuario, y el propio servidor, que permite el acceso a dichos contenidos. En el caso concreto implementado, se hace uso de un servidor Apache2 con tecnología PHP5 habilitada. En cuanto a los contenidos propiamente dichos, serán analizados adecuadamente en el [apartado 5.4](#) de este mismo capítulo, por lo que se evita redundar en este punto.

- **Módulo de Sesión:**

Este módulo es el responsable de procesar el registro de los participantes en el sistema y monitorizar sus respuestas durante toda la duración del experimento. Cada vez que finaliza el período de navegación libre en el que se está aplicando una determinada condición sobre el enlace con el UD, el secuenciador del EMU realiza una petición a este módulo para que devuelva al usuario el correspondiente formulario de evaluación. Este módulo es, en realidad, otro servidor web dentro de la misma arquitectura del LCS, pero de funcionamiento completamente independiente al servidor de contenidos.

- **Módulo de Persistencia:**

Tanto el servidor web como el módulo de sesión requieren del acceso a la información almacenada en las bases de datos que dependen del LCS. Este módulo es el encargado de procesar todas las peticiones de ambos servidores.

La implementación particular de este módulo es un servidor MySQL estándar con dos bases de datos diferenciadas: una para sesión y otra para contenidos.

- **Módulo de Actualización:**

Pese a ser un módulo menor de la arquitectura funcional del LCS, es probablemente el más complejo. Su funcionalidad principal consiste en ejecutar periódicamente un *script* que es capaz de localizar en Internet nuevos contenidos e insertarlos en la base de datos correspondiente.

Puesto que el participante se va a enfrentar en repetidas ocasiones a los mismos portales web durante el proceso de evaluación de cada una de las condiciones definidas para el experimento, es muy importante que se mantenga la atención en las características a evaluar, evitando la pérdida de interés al visitar una y otra vez los mismos contenidos.

En otras palabras, se busca reducir los factores de influencia de contenido que pudieran desembocar en el aburrimiento del participante y, a la postre, influir en su evaluación por motivos no relacionados con la modificación de parámetros en el enlace UD-EMU.

Dada la complejidad e importancia de este módulo, se recomienda al lector que visite, de nuevo, el [Anexo III](#) de este documento, en el que encontrará información al respecto de los *scripts* utilizados, así como el procedimiento de búsqueda, descarga y almacenamiento de nuevos contenidos por medio del servicio **cron** [82] del sistema operativo Linux.

5.4. Contenidos del LCS

A lo largo de todo este capítulo se ha estado haciendo referencia de manera genérica a los contenidos del servidor web del LCS a los que el usuario tiene acceso durante los tiempos de condición (tiempos de navegación web). En este subcapítulo se describe la tipología de contenidos que define el *draft* P.STMWeb [8] a este respecto, así como los contenidos finalmente escogidos para la realización del este estudio.

Contenidos definidos en el borrador P.STMWeb

El *draft* P.STMWeb [8] define cuatro contenidos tipo, desde el más estático y ligero al más dinámico y pesado, en términos de interactividad del usuario y tamaño de descarga para su visualización.

- **Noticias:** Contenido en el que predomina fundamentalmente el texto, con alguna imagen ilustrativa. El usuario puede navegar por las distintas secciones del portal y acceder a noticias recientes. En general, se trata de contenido de bajo peso en los que el tamaño de cada página es de unos pocos *kilobytes* (KB). Salvo que se incluyan vídeos o contenidos de más tamaño, las páginas suelen tener tamaños muy similares.
- **Álbum fotográfico:** Ya sea una galería fotográfica de contenido estático o una red social en la que se pueden visualizar y comentar las fotografías, este tipo de contenidos suele requerir un mayor ancho de banda. Además, las páginas de previsualización de *thumbnails* de toda una galería y la propia visualización de cada imagen pueden tener tamaños muy dispares, según el número de elementos y la resolución de las imágenes.
- **Tienda online:** Generalmente, contenidos basados en la búsqueda por parte de los usuarios de un determinado producto. Son páginas dinámicas capaces de presentar múltiples resultados y gestionar el carro de la compra de los artículos seleccionados. Normalmente, son una mezcla equilibrada de imágenes y texto descriptivo.
- **Mapas:** Este tipo de contenidos son probablemente los que más recursos requieren. Partimos de la misma premisa de búsqueda de la tienda online, pero suponen más esfuerzo en la descarga de imágenes e instrucciones para alcanzar el destino. La descarga de imágenes es mayor cuanto más grado de detalle especifique el usuario. Por ejemplo, una vista satélite o a pie de calle requiere más recursos que un plano 2D dibujado.

Además de estas cuatro categorías, dentro de la navegación web existen otras tipologías de portales y contenidos. Sin embargo, los cuatro tipos definidos son suficientes para analizar el rendimiento y la experiencia del usuario en la evaluación de la QoE del servicio web.

Contenidos adoptados en el estudio

De dichas cuatro categorías tipo, en el presente estudio se ha hecho uso de las dos primeras. Los motivos que justifican dicha elección son principalmente dos: Por un lado, el tiempo limitado para presentar contribuciones a la mejora del *draft* P.STMWeb. Por otro, la voluntad de realizar el experimento con contenidos controlados alojados en un servidor local, para no depender de factores externos que pudieran influir en la valoración de los participantes por causas ajenas a las definidas en las condiciones de evaluación. Esta segunda causa complementa la primera, puesto que el desarrollo de los portales de noticias y fotografías ha tenido que realizarse dentro del tiempo acotado con el que se disponía antes de la finalización del plazo para presentar la contribución.

Para la implementación de ambos portales, se han tomado como referencia los trabajos previos del Centro de Investigación de las Telecomunicaciones de Viena (FTW) [83], que precisamente es impulsor del borrador. Partiendo del núcleo del código desarrollado por FTW, se ha adaptado a contenidos en castellano y se ha pretendido mimetizar el funcionamiento de portales existentes que pudieran ser fácilmente reconocibles por el grupo de participantes. En concreto, estas son las características de cada uno de ambos portales:

Portal de Noticias	
Características	
Servidor web	Apache2
Lenguajes	HTML, PHP5 & JavaScript
Imágenes	Todas las noticias contienen una imagen JPEG
Contenidos	Extraídos de feeds RSS en castellano de www.europapress.es
Apariencia	Mimetiza el aspecto de la web de www.europapress.es
Actualización	Se ejecuta un <i>script</i> que añade nuevas noticias varias veces al día
Tamaño de páginas	
Tamaño de la estructura de la página	3,5 KB
Tamaño medio del contenido (texto de la noticia + foto)	379 bytes
Desviación estándar del tamaño del contenido	73 bytes
Tamaño total medio	~3,9 KB

Tabla 5.4 – Características del portal de noticias del LCS

Galería fotográfica	
Características	
Servidor web	Apache2
Lenguajes	HTML, PHP5 & JavaScript
Imágenes	Todas las noticias contienen una imagen JPEG
Apariencia	Mimetiza el aspecto de la web de www.facebook.com
Contenidos	5 álbumes con una media de 40 fotos por álbum
Tamaño de páginas	
Tamaño medio de las fotos	114,50 KB
Desviación estándar del tamaño de las fotos	10,15 KB

Tabla 5.5 – Características de la galería fotográfica del LCS

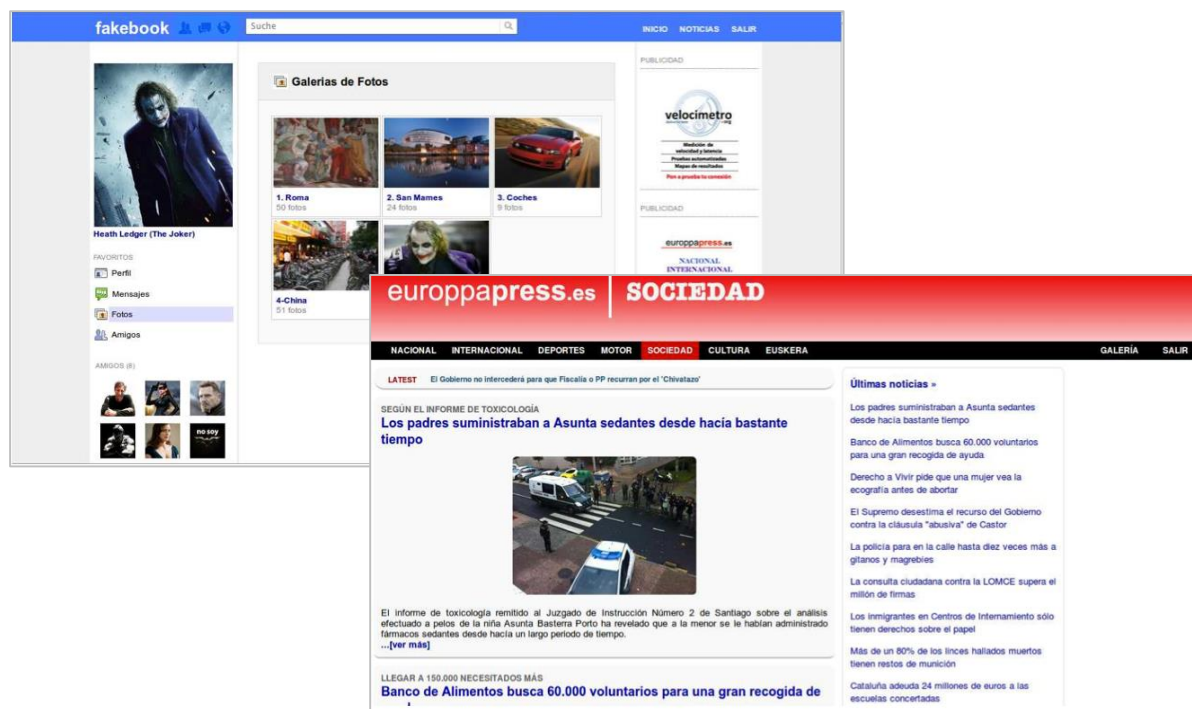


Figura 5.6 – Capturas de los contenidos alojados en el LCS: noticias y álbum fotográfico

Actualización de contenidos

Como ya se anticipara anteriormente al describir el módulo de actualización del servidor de contenidos, se ha procurado ofrecer información actual a los participantes para que durante la realización de la prueba pudieran mantener el interés en las noticias publicadas, lo cual permite acercarnos a una situación más real de navegación. El usuario que hace uso de estos portales, puede decidir cambiar de noticia o entrar a leerla en profundidad, a la vista de su interés por el tema. Si dicha información está anticuada, es muy probable que el participante realice la prueba a desgana o sin prestar suficiente atención a los aspectos que se evalúan. De esta manera, podemos limitar algunos factores de influencia del usuario.

Para el proceso de actualización, se hace uso de varios *feeds* de noticias del portal de la agencia Europa Press [84], los cuales tienen que pasar por un proceso de filtrado en varias fases para adecuarse al formato requerido para ser almacenados en la base de datos de noticias del LCS (figura 5.5).

Todo el proceso de creación del *pipe* de noticias a partir de las fuentes de Europa Press, la posterior adaptación de los contenidos y su inserción en la base de datos del LCS está documentado en el documento descriptor de la plataforma desarrollada incluido como Anexo III de este trabajo.

6. METODOLOGÍA

En este capítulo se desgrena la metodología adoptada en la realización del estudio de QoE que se describe en el presente documento, prestando especial atención al plan de pruebas diseñado para llevar a cabo la validación. En concreto, el ciclo de vida que rige el proceso seguido es el siguiente:



Figura 6.1 – Ciclo de vida del estudio

Paso 1: *Análisis de P.STMWeb y G.QoE-Web*

Partimos del análisis de los borradores P.STMWeb [8] y G.QoE-Web [10] con el objetivo de definir el alcance de aplicación del estudio de la calidad experimentada que se presenta en este documento. En otras palabras, se busca identificar los indicadores de calidad (KQI) relativos al Servicio de Navegación Web que puedan afectar en las transacciones del usuario, y, por tanto, influir en su experiencia de navegación.

Para ello, se analizan características específicas del nivel de red, como el ancho de banda, el retardo o la pérdida de paquetes, así como otra serie de factores a nivel de aplicación, como puede ser el tiempo de carga de las páginas. En este punto, se hace una selección de los parámetros que serán objeto de estudio y se identifican los factores de influencia del sistema que pudieran afectar en el proceso de evaluación de la QoE.

En concreto, y siguiendo las directrices del *draft* P.STMWeb [8], se plantea analizar la experiencia del usuario respecto a la variación en el ancho de banda de su acceso al sistema, con unas condiciones de retardo fijas en un entorno controlado. Dicho entorno se define en el paso 2 de la metodología del presente estudio.

Paso 2: Definición del escenario y contenidos

La descripción de este paso se corresponde con el capítulo anterior de este documento, tanto en el diseño y desarrollo de la plataforma de evaluación adoptada como escenario para la realización de las pruebas, como en la definición de los contenidos a utilizar durante las mismas.

Se recuerda que, pese a que la red desplegada cuenta con acceso a Internet, las pruebas se conducen dentro de la zona controlada, que incluye la red LAN en la que se encuentra el Servidor Local de Contenidos y los enlaces UD-EMU sobre los que se aplican las distintas condiciones de test (figura 5.2).

Se trata, por tanto, de un caso de evaluación subjetiva de la QoE con métricas de referencia completa (métricas FR), ya que no sólo se conoce el estado de la red, sino que se controla a necesidad para las sucesivas condiciones que se presentan al usuario. De acuerdo a la clasificación de metodologías expuesta en el capítulo 'Estado del arte', el presente estudio se corresponde con el Tipo 1 de la tabla 4.3.

Paso 3: Definición del plan de pruebas

Partiendo de la definición del escenario de pruebas y del desarrollo de los mecanismos necesarios para la aplicación de las condiciones de test a nivel de red por parte del EMU (recordemos la selección del ancho de banda como criterio de influencia), se define el plan de pruebas a seguir por los participantes, poniendo especial interés en la definición de los mecanismos de evaluación oportunos. Para ello, se hace referencia al borrador de plan de pruebas asociado al *draft* P.STMWeb [9], cuyo contenido puede consultarse en el Anexo I de este documento. Dada la importancia de este paso, se remite al lector al apartado 6.1 de este mismo capítulo, en el que se detalla paso por paso su contenido.

Paso 4: Elección de participantes

Definido el plan de pruebas, se realiza la selección de participantes que se van a someter al experimento. En concreto, se establecen dos categorías de usuarios: **expertos** y **no expertos**. Esta distinción no atiende tanto al uso del servicio de navegación, sino a los conocimientos técnicos del servicio y a los procedimientos de evaluación subjetiva.

De hecho, el grupo de los usuarios no expertos lo integran alumnos de la Escuela Técnica Superior de Ingeniería de Bilbao de la Universidad del País Vasco (UPV/EHU). Todos ellos son estudiantes de último curso del grado de ingeniería de telecomunicaciones, por lo que, como es obvio por sus estudios, no son totalmente ajenos al servicio. Por su parte, el grupo de expertos lo forman profesores e investigadores del Departamento de Ingeniería de Comunicaciones de la misma universidad.

Se ha procurado equilibrar el número de participantes, consiguiéndose la cifra de 23 sujetos en cada grupo. Desgraciadamente, a efectos comparativos de los resultados, esta paridad no se ha logrado también en la proporción de sujetos por género. Se puede encontrar más información destacada respecto a los participantes en la tabla 6.1 mostrada a continuación.

Asimismo, en el [Anexo IV](#) se presenta la hoja de instrucciones facilitada a los participantes, donde se detallan las tareas a seguir en cada fase del experimento.

	Expertos	No Expertos
Número	23	23
Prueba estándar	12	23
Prueba corta	11	-
Género	18 varones – (78,26 %) 5 mujeres – (21,74%)	15 varones – (65,22 %) 8 mujeres – (34,78 %)
Edad (media)	30,65	21,83
Edad (mediana)	28	21
Uso semanal de internet	Más de 20 horas/semana	15 a 20 horas/semana
Experiencia en navegación	15 años usando Internet	10 años usando Internet

Tabla 6.1 – Características de los participantes del estudio

Paso 5: Pruebas de validación

Una vez instruidos a los participantes en el procedimiento a seguir, se inician las pruebas de validación. Para ello, se dispone de tres equipos ([figura 5.3](#)) desde los que se puede realizar, en paralelo, el experimento de navegación definido en el plan de pruebas.

El motivo principal que origina esta decisión es el de poder ofrecer simultáneamente la prueba a múltiples participantes en el plazo de las dos semanas en las que se llevaron a cabo las pruebas de validación. Recordemos que los participantes son tanto alumnos como investigadores y profesores universitarios y que las pruebas se llevaron a cabo en horario laboral, por lo que fue necesario acomodar su disponibilidad a los tiempos disponibles en el laboratorio para realizar el experimento.

Paso 6: Análisis de los resultados

Una vez finalizada la fase de validación, se realiza el tratamiento de los resultados. Para ello, se hace uso del *software* y lenguaje matemático *MatLab*, con el que se obtiene información sobre la validez de los resultados, así como una representación gráfica de los mismos, lo que facilita el posterior análisis y extracción de conclusiones sobre el estudio realizado. Además de los datos obtenidos por las encuestas, se tiene muy en consideración la experiencia arrojada por los participantes a partir de la exposición verbal de su percepción durante la realización del experimento.

El [capítulo 7](#) de este documento desgrana los resultados obtenidos, tanto por grupo de usuario, como en función de diversas condiciones definidas en el plan de pruebas, por lo que se evita ahondar más en este punto en este momento.

Paso 7: Contribución de propuestas

Por último, se cierra la primera iteración del ciclo de vida con la presentación de los resultados y las propuestas de mejora a los borradores en estudio [\[8\]\[9\]\[10\]](#). Una vez revisadas y publicadas las modificaciones en las recomendaciones ITU-T P.1501 [\[11\]](#) y G.1031 [\[12\]](#), se daría inicio a una segunda vuelta, adaptando dichos cambios para una nueva fase de validación. Este punto se describe en el [capítulo 8](#) sobre conclusiones y líneas futuras.

6.1. Plan de pruebas

En anteriores capítulos ya han quedado definidos tanto el escenario sobre el que se realizan las pruebas de evaluación de QoE, como los parámetros intrínsecos de QoS sobre los que actúa el servidor emulador de red (EMU) para configurar las diferentes condiciones de test.

En este apartado, por tanto, se especifican las condiciones de test del experimento, así como la secuencia de aparición en la que se mostrarán dichas condiciones y su mecanismo de evaluación. Para ello, tomamos como referencia el borrador del plan de pruebas asociado al *draft* P.STMWeb [9], cuyo contenido puede consultarse en el [Anexo I](#) de este documento.

Condiciones de test

Se establece que los participantes evalúen 6 condiciones distintas de ancho de banda en el enlace descendente con valores múltiplos entre 64 y 2048 kbit/s, tanto para el portal de noticias como para el álbum fotográfico, lo que supone un total de 12 condiciones entre ambos tipos de contenido, tal y como presenta la [tabla 6.2](#) a continuación. En todos los casos, el retardo de ida y vuelta queda establecido en un valor fijo de 20ms.

Enlace descendente [kbit/s]	64	128	256	512	1024	2048	Condiciones
Portal de noticias	x	x	x	x	x	x	6
Álbum fotográfico	x	x	x	x	x	x	6
RTT (Retardo de ida y vuelta)	20 ms (tiempo fijo)						
TOTAL							12

Tabla 6.2 – Condiciones de test de P.STMWeb

El *draft* no especifica valores de ancho de banda para el enlace ascendente. Sin embargo, por coherencia con el tipo de conexiones domésticas existentes (generalmente asíncronas), en este trabajo se ha establecido una limitación en el ancho de banda ascendente igual a la mitad del valor del enlace descendente. Por ejemplo, para la condición de 512 kbit/s de bajada, se asigna un valor de 256 kbit/s de subida.

Fases y temporización

Asimismo, se establece que el participante debe someterse a dos fases completas de evaluación, separadas por un período de descanso de 10 minutos. Cada fase debe constar, por tanto, de las 12 condiciones de la [tabla 6.2](#), con una única aparición por sesión para cada condición. El orden de presentación de las condiciones en cada fase debe ser completamente aleatorio en cuanto al valor de ancho de banda.

Es decir, suponiendo una hipotética secuencia aleatoria de velocidades, como podría ser la mostrada en la [tabla 6.3](#), el participante deberá evaluar su experiencia de navegación de manera consecutiva para cada velocidad, primero para el portal de noticias y, seguidamente, para el álbum fotográfico, antes de pasar al siguiente valor.

Secuencia	Primera	Segunda	Tercera	..		
Velocidad [kbit/s]	256	128	2048	512	64	1024

Tabla 6.3 – Posible secuencia aleatoria de velocidades

A priori, el participante sabrá únicamente que debe evaluar las condiciones, alternando la navegación libre por el portal de noticias, con la navegación a través del álbum fotográfico, pero desconociendo que para cada par de contenidos (noticias/fotos) que se le presenten, estará evaluando la misma condición de velocidad.

El tiempo total de experimento se completa con tres bloques previos a la realización de las fases de evaluación. Dichos tres bloques son:

- **Fase de instrucciones:** El usuario tiene unos 5 minutos para leer el documento explicativo y consultar cualquier duda que tuviera al respecto de la realización de la prueba.
- **Fase de registro:** En los 4 minutos siguientes, el usuario deberá rellenar un formulario como el de la [figura 6.2](#). Es fundamental que responda con sinceridad y precisión a los datos de registro, ya que su respuesta permite identificar los distintos perfiles de usuario que realizan el test. Esta fase no está incluida en el *draft* como tal, pero se ha considerado de importancia de cara a la obtención de datos estadísticos adicionales.

Figura 6.2 – Fases del plan de pruebas y detalle de la fase de Registro

- **Fase de entrenamiento:** Esta fase es similar a las fases de evaluación posteriores, salvo que se trata de una versión reducida con sólo tres condiciones: El caso a mejor velocidad, el caso a peor velocidad y un caso intermedio. En esta fase de entrenamiento, los usuarios evalúan su experiencia de igual manera que en las fases posteriores, con la diferencia de que al completar las encuestas para cada condición, se les muestra su valoración en comparación con el valor que hipotéticamente deberían haber evaluado. Así, asimilan el funcionamiento del sistema y pueden ajustar su criterio al aprendizaje de esta fase.

Considerando tiempos de navegación libre (tiempos de condición) de 150 segundos, mas tiempos de evaluación (tiempo para rellenar la encuesta) de 30 segundos, la evaluación completa de cada condición se realiza en 180 segundos. La [tabla 6.4](#), mostrada a continuación, recoge la temporización de las fases y la duración completa del experimento.

Tarea	Duración [seg]	Duración [min]	Condiciones
Instrucciones	300	5	
Registro de usuario	240	4	
Entrenamiento	540	9	3
Fase 1	2160	36	12
Descanso	600	10	
Fase 2	2160	36	12
Total	6000 seg	100 min	27

Tabla 6.4 – Temporización de las fases del experimento

Evaluación

Por último, solamente queda determinar cómo se realiza la evaluación de cada condición presentada al participante. De nuevo, acudimos al borrador del plan de pruebas del P.STMWeb [9], donde se definen dos preguntas concretas y las escalas cualitativas asociadas:

- **Escala de opinión MOS:**

Por favor, califique la calidad global de la conexión que ha estado usando según la escala indicada: (1) Muy mala, (2) Mala, (3) Regular, (4) Buena, (5) Muy buena.

- **Pregunta de respuesta binaria:**

¿Aceptaría esta calidad de conexión para uso habitual en su casa?: Sí / No

Cuestionario

Por favor, califique la calidad global de la conexión que ha estado usando según la escala indicada

Muy mala

¿Aceptaría esta calidad de conexión para uso habitual en su casa?

No

Enviar

Figura 6.3 – Captura de la encuesta de condición

Estas dos preguntas por cada condición, junto con otra serie de opiniones recogidas de los participantes a la finalización del experimento, han permitido obtener los resultados que se muestran en el capítulo a continuación, así como extraer las propuestas de mejora al *draft* P.STMWeb [8][9], recogidas en la contribución realizada al SG12 de la ITU a este respecto.

7. RESULTADOS

7.1. Análisis de los resultados

Tras la conclusión de las pruebas de evaluación por parte de los 46 participantes en el experimento (recordemos: 23 expertos y 23 no expertos, [tabla 6.1](#)) se han obtenido los resultados presentados a continuación.

Todas las gráficas y valores a continuación se han evaluado a partir de los resultados de opinión sobre la escala MOS de 5 niveles con intervalos de confianza del 95%, lo que permite eliminar valoraciones espurias que puedan afectar a la correlación de los valores obtenidos.

Resultados generales

La primera gráfica que se presenta a continuación, muestra el valor medio del MOS obtenido respecto a los distintos anchos de banda evaluados por el conjunto total de participantes, para cada tipo de contenido (noticias y fotografías):

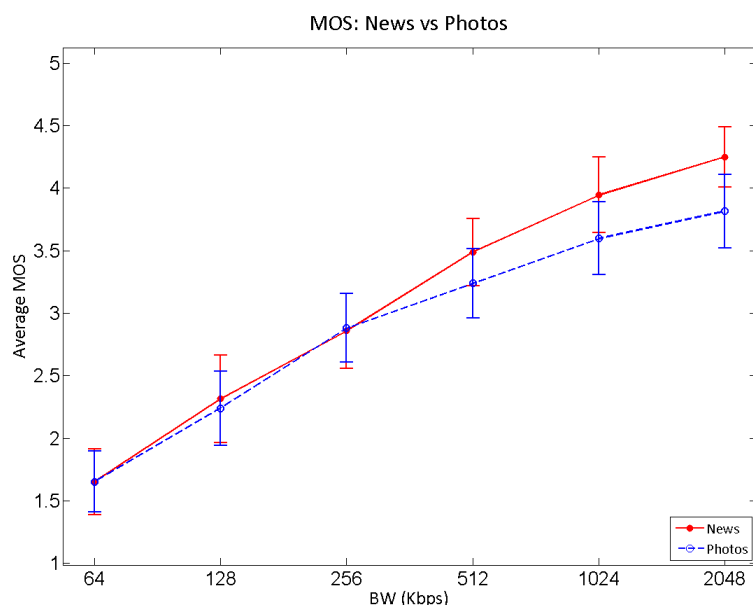


Figura 7.1 – Resultados MOS-BW para el total de participantes

Se puede observar en la gráfica, cómo ambos tipos de contenidos presentan cierta linealidad en los resultados, siguiendo un patrón similar al que marcaba la curva de relación QoS-QoE de la [figura 4.17](#). Se destaca la existencia de una menor tolerancia a contenidos de mayor peso, como es el caso del álbum fotográfico, tal y como se observa en azul en la figura.

Este efecto se aprecia mejor a mayores velocidades, ya que la diferencia en los tiempos de carga entre texto y noticias se hace más evidente. En condiciones de peor ancho de banda, puesto que los tiempos de carga de las páginas no son buenos en ningún caso, ambas gráficas tienen un aspecto muy similar.

Observando la actuación de los participantes durante la realización de las pruebas ratifica la diferenciación en la valoración por contenidos:

Cuando los participantes navegaban libremente por el portal de noticias, no prestaban especial atención a las imágenes, sino que hacían recorridos rápidos por el texto, que generalmente era lo primero en cargarse, antes de decidir si pasar a la siguiente noticia.

Sin embargo, en los contenidos del álbum fotográfico, y, en especial, en la carga de fotos de mayor peso, muchos usuarios mostraban cierta impaciencia al observar la carga progresiva de la imagen que habían seleccionado. La 'frustración' de tener que estar a la espera mientras se completaba la carga, incide directamente en la valoración recibida.

Comparación de resultados por categoría de usuarios

La tendencia general mostrada en la figura 7.1 se mantiene en la distinción por grupos de usuarios, con algunas matizaciones, como se puede observar a continuación:

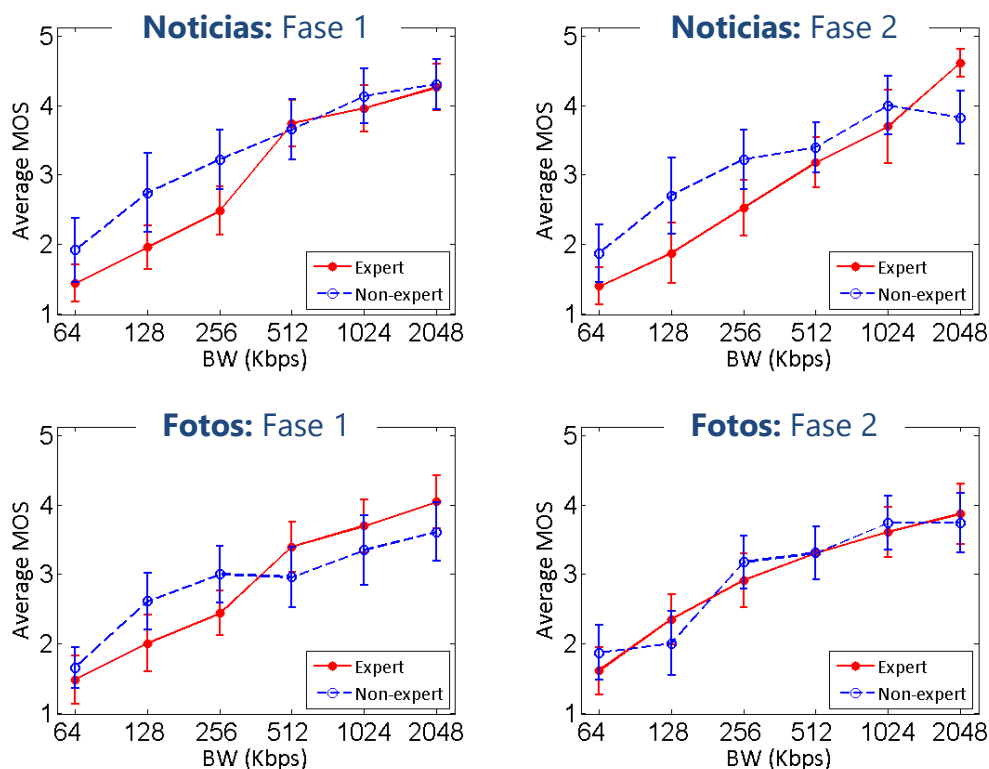


Figura 7.2 – Gráficas comparativas de resultados MOS-BW por categoría de usuarios

Se sigue apreciando la tendencia ascendente, tanto en el caso de los usuarios expertos, como en el de los no expertos. Sin embargo, se aprecian mayores irregularidades en la valoración en el caso de estos últimos, quizá por una mayor dificultad en su adaptación al experimento.

En cuanto a los usuarios expertos, prácticamente en todos los casos se muestran más exigentes en su experiencia de navegación, pudiendo incluso apreciarse un pequeño aumento en su satisfacción sólo en velocidades por encima de los 512 kbit/s.

Comparación de resultados por contenidos

Si reorganizamos las gráficas anteriores para comparar la experiencia recogida en función de los contenidos, nos encontramos la figura siguiente:

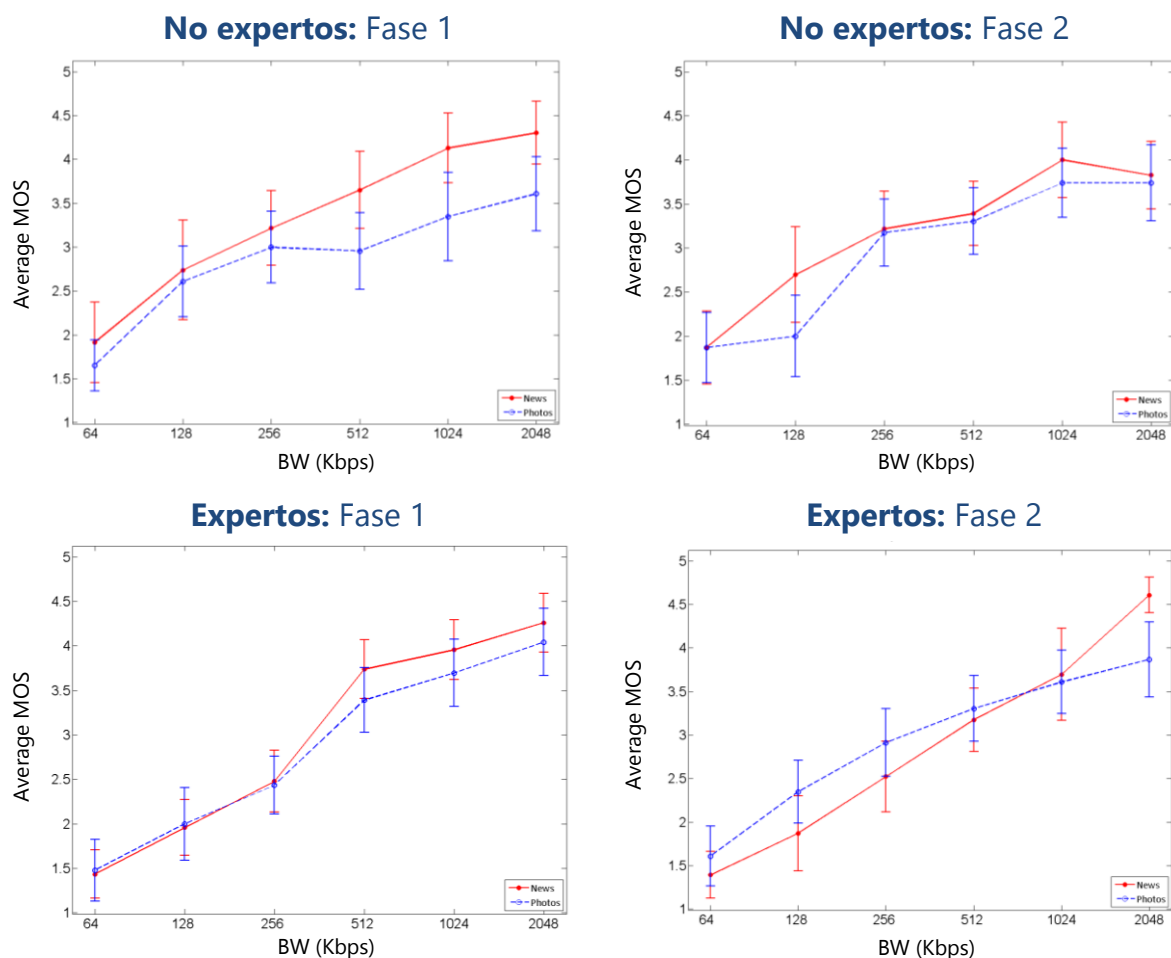


Figura 7.3 – Gráficas comparativas de resultados MOS-BW por contenidos

En el caso de los usuarios no expertos, se puede apreciar que los contenidos más pesados son valorados con menor puntuación. Esto es especialmente notable en la fase 1, donde la experiencia de carga de imágenes pesadas se deja ver en las valoraciones de las condiciones de más ancho de banda: El usuario está experimentando una carga medianamente aceptable de la página en su totalidad, pero la imagen principal parece tener ciertas dificultades, lo que se traduce en una penalización en la valoración, si lo comparamos con la misma condición para el contenido fundamentalmente textual del portal de noticias.

Los usuarios expertos, por su parte, mantienen una tendencia más estable en sus valoraciones, aunque destaca la valoración superior en el contenido fotográfico respecto a las noticias en la segunda fase del experimento. Esta aparente incongruencia puede tener sentido si tenemos en consideración que los usuarios expertos son más conscientes del procedimiento del test. Muchos de ellos observaron que, pese a la aleatoriedad de la aparición de condiciones, existía una relación entre la velocidad experimentada en la ronda de noticias y la subsiguiente ronda de navegación por el álbum fotográfico.

Resultados sobre uso doméstico

Otro análisis interesante es el de los resultados arrojados por la pregunta '¿Aceptaría esta calidad de conexión para uso habitual en su casa?'. La [figura 7.4](#) recoge las respuestas ofrecidas por los participantes tras la evaluación de las sucesivas condiciones experimentadas en la navegación por el portal de noticias¹¹. De la observación de dicha figura pueden extraerse, en primera instancia, dos conclusiones directas:

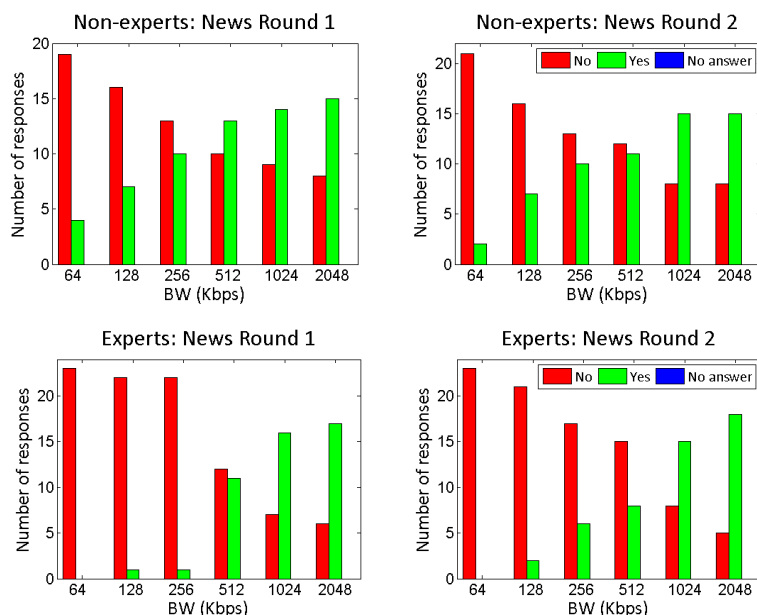


Figura 7.4 – Aceptación de la calidad experimentada para su uso doméstico

Por un lado, se puede apreciar claramente un aumento de la valoración positiva con mejores condiciones de navegación, tanto en el caso de los usuarios expertos como en el de los no expertos, lo cual mantiene la tendencia de que, a más valoración en la escala MOS, más probabilidad de aceptación de las condiciones en el entorno doméstico. La segunda es que los usuarios expertos mantienen también aquí una posición más crítica y exigente que los no expertos, especialmente en las experiencias de navegación a menor velocidad, lo que cuadra con las gráficas de la [figura 7.3](#) en la que se aprecia ese salto en la valoración MOS.

Gracias a la información recogida en las conversaciones mantenidas con los participantes (y en particular con los expertos) a la finalización del experimento, cabe destacar que, en muchos de los casos, los participantes decidieron clicar en 'Sí' a esta pregunta, condicionados por la fase de entrenamiento en la que se mostraba la comparativa entre lo que ellos inicialmente habían valorado y lo que se esperaba de ellos. De no haber recibido esa información, la gran mayoría habría votado 'No' a pesar de haber valorado la condición con valores altos en la escala MOS.

El motivo de esto no es otro que el de la existencia de un desfase evidente entre las velocidades analizadas en el estudio (marcadas por el plan de pruebas anexo al [draft P.STMWeb \[9\]](#)) y las que a día de hoy se contratan en el entorno doméstico nacional.

¹¹ Las gráficas del contenido fotográfico, así como el resto de resultados, pueden consultarse en el [Anexo II](#).

7.2. Propuestas de mejora

Actualización de condiciones de test

Como se puede suponer en vista del último párrafo, la primera propuesta de mejora al *draft* está directamente relacionada con los valores de ancho de banda que se toman como condición. En lugar de establecer unos valores fijos como los que indica el *draft* en su plan de pruebas [9], convendría definir la escala de condiciones en función de las expectativas y la propia experiencia previa de los participantes. Para ello, se pueden plantear dos opciones:

- **Adecuar la escala al ancho de banda contratado por el participante en su domicilio.**

A partir de la obtención de este valor en la fase de registro, se definiría una escala de condiciones cuyo valor central fuera el contratado por el participante. De esta manera, se podría observar el grado de aceptabilidad de su conexión doméstica respecto a otras condiciones mejores y peores.

Esta información podría ser de gran utilidad para el participante. Sin embargo, para realizar un estudio en el que los resultados puedan ser comparables de acuerdo a mismos valores de condición, se requeriría una muestra de población mucho mayor a la disponible en este estudio, tal que existieran subgrupos suficientemente grandes de participantes con la misma velocidad de acceso contratada.

- **Adecuar la escala a las velocidades más comunes en el territorio del estudio.**

Para ello, sería necesario realizar un estudio previo de las velocidades más comúnmente contratadas por los usuarios. En concreto para el territorio español, la mejor opción es consultar las publicaciones sobre niveles de QoS que trimestralmente deben realizar los operadores de telecomunicaciones a través de la web del Ministerio de Industria, Turismo y Comercio [86].

En función de las publicaciones de tecnologías más recurrentes, se podría elaborar una escala con seis condiciones que pudiera servir como base para el experimento. Tomando como referencia las publicaciones más recientes, la escala de debería ser similar a la siguiente, donde se puede observar el enorme grado de desfase con los valores propuestos en el borrador:

Escala de condiciones	1°	2°	3°	4°	5°	6°
Según Draft P.STMWeb	64 Kbps	128 Kbps	256 Kbps	512 Kbps	1024 Kbps	2048 Kbps
Según Publicación QoS Ministerio	3 Mbps	6 Mbps	10 Mbps	20 Mbps	30 Mbps	50 Mbps

Tabla 7.1 – Escala de condiciones del draft P.STMWeb frente a valores de referencia nacionales

Desde este estudio se considera más adecuada la segunda opción, ya que, además de obtener resultados comparables para las mismas velocidades, se abre la opción a la comparación con otros estudios realizados en el mismo ámbito nacional, para una muestra de población similar o, incluso, de otra comunidad o territorio.

Teniendo en cuenta además, que no todos los países tienen el mismo despliegue tecnológico, ni todos los dispositivos de usuario con acceso a Internet permiten el mismo rango de velocidades, definir una recomendación con una escala adaptable a la necesidad del estudio tiene mucho más sentido que una escala de valores fijada que, en muchos casos, puede estar anticuada antes incluso de realizarse el estudio.

Definición más precisa sobre el uso de cachés

En el diseño de la arquitectura de pruebas del [capítulo 5](#), se hace referencia en varias ocasiones a la necesidad de determinar y minimizar los factores de influencia que pudieran afectar negativamente a una adecuada valoración de las condiciones durante el tiempo de realización del experimento. De todos los factores detectados y tratados, existe uno en particular para el que definir una política de acción es especialmente determinante: la caché del navegador.

Mantener la caché deshabilitada, tal y como se ha optado en la realización de este trabajo, permite evitar que objetos precargados en anteriores sesiones de navegación puedan influir en el caso concreto de evaluación. Por el contrario, si la caché se mantiene activa, existe un riesgo elevado de que no exista independencia entre condiciones y el usuario tenga dificultad para distinguir lo que realmente está experimentando.

Por ejemplo, si en una sesión anterior se hubieran visitado fotografías de gran tamaño y éstas quedaran almacenadas en la caché, al evaluar una condición de bajo ancho de banda y querer visitar las mismas imágenes, éstas aparecerían rápidamente al no realizarse la petición de descarga de las mismas desde el servidor de contenidos. De esta manera, el usuario puede percibir una experiencia de navegación a alta velocidad cuando en realidad no es así.

Sin embargo, deshabilitar completamente la caché también tiene efectos negativos ya que puede ir en detrimento de la evaluación de las opciones de velocidad más bajas: Un usuario que, en una misma sesión por el portal fotográfico, quisiera volver atrás en su historial para volver a ver la misma imagen que momentos antes tenía cargada, deberá volver a pasar por todo el proceso de descarga de contenidos, lo que influye negativamente en su valoración.

La experiencia real de navegación se sitúa en un punto intermedio. Por ello, desde la contribución se solicita que se especifique más adecuadamente la acción a tomar al respecto de la caché del navegador.

Temporización

Muchos de los participantes, en conversaciones posteriores a la finalización de sus respectivas pruebas, expusieron su queja al respecto de la extensa duración del experimento. En muchos casos, y a pesar del descanso de 10 minutos, varios participantes mostraron cierta fatiga y hastío al tener que volver a repetir todo el procedimiento en una segunda fase, especialmente después de haberse recorrido gran parte de los contenidos del LCS. Por este motivo, se tomó la determinación de realizar una prueba de menor duración con la mitad de los usuarios expertos que se sometieron al test ([tabla 6.1](#)).

Los resultados arrojados destacan claramente que existe más linealidad en la evaluación de las condiciones cuando la duración del test es más corta. Dicho de otra manera, los resultados obtenidos son peores (y en especial en la fase 2) cuanto más largo es el test. Por este motivo, también se propone como mejora la reducción de los tiempos de condición de los 180 segundos a, al menos, los 120 que se configuraron en la prueba corta (tabla 7.2).

Tarea	Duración (estándar)	Duración (corta)	Condiciones
Instrucciones	5	5	
Registro de usuario	4	4	
Entrenamiento	9	6	3
Fase 1	36	24	12
Descanso	10	10	
Fase 2	36	24	12
Total	100 min	73 min	27

Tabla 7.2 – Fases del experimento (versión estándar vs. versión corta)

Definición de las tareas del participante

De nuevo, el *draft* [8] no es especialmente preciso con las tareas que debe realizar el participante durante las sesiones de navegación. A priori se busca evaluar su experiencia en el proceso de navegación libre para diferentes condiciones. Sin embargo, una especificación más detallada de las posibles acciones que deben realizar durante el tiempo de cada condición, podría tener gran influencia en los resultados.

Si el usuario tiene que realizar una lista de tareas definidas, puede comprobar hasta qué grado ha logrado satisfacerlas en el tiempo disponible para cada condición. De esta manera, se podría incluir una magnitud objetiva en la evaluación de la QoE que diese información al usuario sobre cuán satisfactoria es la condición en observación, más allá de hacer clic aleatoriamente en los enlaces de las webs.

De hecho, esta propuesta enlaza en cierto modo con la anterior sobre temporización. De la observación de la actitud de los participantes (tanto expertos como no expertos) a lo largo de la realización de las pruebas, se pudo percibir un patrón de actuación común en algunos de ellos. En concreto, en la evaluación de condiciones relativas al contenido fotográfico, varios participantes descubrieron ciertas fotografías de gran tamaño que les servían para extraer conclusiones respecto a la condición en evaluación, únicamente mediante la observación del tiempo que tardaban en cargar.

Si bien es cierto que esto es un indicador de experiencia adquirida por parte de los participantes a lo largo del experimento, no estaban realmente navegando por múltiples páginas durante toda la sesión, sino que su actividad se asemejaba más a la definida en la metodología de la ITU-T G.1030 [45]. Por este motivo, se incluye también en la contribución una breve indicación sobre la idoneidad de definir más adecuadamente las tareas de usuario.

8. CONCLUSIONES

En este capítulo se presentan las conclusiones del estudio realizado para la presente Tesis de Máster, destacando las aportaciones más interesantes e indicando las posibles líneas de continuación del trabajo realizado. Para finalizar, se detallan las contribuciones más importantes en la divulgación del estudio realizado.

8.1. Valoración general del estudio

A lo largo de este documento se ha descrito el estudio realizado sobre la evaluación de la QoE en el Servicio de Navegación Web. Para ello, se ha hecho un análisis de dos *drafts* de la ITU especialmente relacionados con este tema:

- El borrador **P.STMWeb** [8][9], que define una arquitectura de pruebas y una metodología para el estudio de la Calidad Experimentada en el Servicio de Navegación Web, que pretende ir más allá de la aplicación estudio sobre la carga sencilla de una página web, tal y como se define en la recomendación ITU-T G.1030, para ahondar en el concepto de experiencia respecto a una sesión de navegación completa.
- El borrador **G.QoE-Web** [10], que categoriza los distintos factores de influencia que es necesario considerar al realizar un estudio de QoE.

La consideración conjunta de ambos *drafts* para el presente estudio, unido a varios conceptos de interés extraídos del análisis previo del estado del arte, han facilitado el desarrollo de una plataforma de validación, que no sólo ha arrojado datos interesantes al respecto de la percepción y experiencia de los usuarios que participaron en el experimento, sino que también ha supuesto un avance en el desarrollo de QoXcloud, la plataforma para la que el presente trabajo sirve de prototipo.

Aportaciones del estado del arte a la realización del estudio

Fundamentalmente, dos son las recomendaciones de referencia del estudio:

- Por una parte, la recomendación **ITU-T P.800** [48], que pese no estar enfocada a la tipología de servicio que se aborda en este trabajo, define la que es probablemente la metodología de referencia en el estudio de la calidad experimentada. Múltiples trabajos directa o indirectamente relacionados con la percepción del usuario y su satisfacción, siguen un esquema metodológico análogo al de la ITU-T P.800.
- Por otra parte, se encuentra la recomendación **ITU-T G.1030** [45], una de las primeras recomendaciones en realizar un acercamiento subjetivo al servicio de navegación web, que ha permitido, incluso, inferir formulaciones matemáticas que relacionan la valoración MOS con parámetros intrínsecos de QoS. La intención de completar esta recomendación con un experimento de navegación web completo es lo que motiva el borrador P.STMWeb y, por extensión, este trabajo.

Consideraciones sobre la evaluación de QoE objetiva

Además de la vertiente de estudio subjetiva de la QoE, se ha analizado el acercamiento objetivo. La percepción personal del autor de este trabajo al respecto del análisis objetivo de la QoE es que éste requiere de una determinación muy precisa de los factores a analizar, no siempre abordable, así como de la existencia de métricas capaces de relacionar variaciones objetivas en la calidad o rendimiento con la percepción subjetiva de los usuarios. Generalmente, las métricas de QoE objetiva requieren haber sido validadas mediante pruebas de QoE subjetiva que permita establecer la relación, lo que implica la necesidad de aplicar metodologías subjetivas como las de la ITU, al menos en una primera fase.

La ventaja de las mediciones de QoE objetiva es que, en aquellos tests en los que se puedan aplicar, no se requerirá de la presencia de usuarios validando el sistema o servicio mediante su opinión, sino que basta con recurrir a dichas métricas, reduciendo tiempo y esfuerzo por parte del equipo investigador. En cualquier caso, y dependiendo del servicio o aplicación a analizar, no siempre se disponen de estas métricas o, si existen, dadas las características del servicio en cuestión, no siempre son aplicables, lo que implica la validación clásica por medio de encuestas de percepción y satisfacción.

Validación de la recomendación ITU-T P.1501

Gracias a la utilización de encuestas de evaluación subjetiva, se han podido obtener las conclusiones probablemente más destacables de este trabajo, al respecto de la validación del borrador P.STMWeb [8][9], **ya convertido en la recomendación ITU-T P.1501 [11]**. Las propuestas aportadas desde este estudio a través de la contribución presentada al grupo de trabajo SG12 de la ITU [85], han sido el detonante principal de la aprobación de la recomendación de la ITU-T anteriormente mencionada.

En concreto, desde la exposición de los resultados obtenidos, se sugería la revisión de los tiempos de condición (150 segundos de navegación + 30 segundos de evaluación) que, en la mayoría de los casos experimentados, resultaban excesivos. La propia experiencia adquirida por los participantes a lo largo del experimento les permitía tomar la decisión sobre la votación posterior, prácticamente en los primeros segundos de carga de contenidos. Los editores de la recomendación ITU-T P.1501 tuvieron este apunte en consideración y finalmente, se han establecido tiempos de condición incluso más reducidos a los que desde este trabajo se sugirieron. Los nuevos tiempos de condición han pasado de 150 a 60 segundos para la navegación, manteniéndose los 30 segundos de evaluación, lo que supone una reducción de 40 minutos sobre los 100 minutos de la versión inicial.

Otra de las aportaciones, la referida a la definición fija de los valores de condición de ancho de banda, ha sido también considerada. Si bien se mantienen como ejemplo orientativo los valores originales entre 64 y 2048 kbps, la recién publicada recomendación sí establece que dichos valores podrán ser acomodados a las necesidades específicas del estudio. Gracias a la incorporación de esta propuesta, la vigencia de la recomendación está garantizada más allá de tecnologías y velocidades existentes en el momento de su publicación.

No se entra a valorar el procedimiento más adecuado de selección de los peldaños de esa escala de condiciones. Sin embargo, desde este trabajo se insiste en realizar una selección de valores acorde a los anchos de banda más extendidos en el territorio donde se realice el experimento, por los motivos que ya se destacaron en el capítulo anterior.

Por último, en cuanto a las consideraciones sobre el cacheado de información, se establece que debe hacerse un uso normal del navegador para poder disponer de una experiencia más cercana al caso real posible. Por ello, para que los efectos de condiciones anteriores no influyan en evaluaciones subsiguientes, debe limpiarse o resetearse la caché en cada cambio de condición.

Validación de la recomendación ITU-T G.1031

A partir de las aportaciones anteriormente mencionadas, que han apoyado la aprobación de la recomendación ITU-T P.1501 [11], surgen también nuevas aportaciones para la validación del borrador G.QoE-Web [10] como nueva recomendación ITU-T G.1031 [12]. Dada la estrecha relación entre ambas recomendaciones, no se concibe la validación del borrador P.STMWeb [8][9] sin hacer extensible ese mismo resultado al draft G.QoE-Web [10], al menos de manera indirecta.

Recordemos una vez más que el *draft* P.STMWeb define una arquitectura y una metodología para la evaluación de la QoE en el servicio de navegación web, mientras que G.QoE-Web establece una categorización de los factores de influencia que pueden afectar a la evaluación de la QoE, por ejemplo, en la arquitectura que define P.STMWeb.

Precisamente, el proceso de identificación de los posibles factores de influencia presentes en la arquitectura adoptada en el presente estudio, ha sido de vital importancia para detectar la necesidad de definir un plan de acción sobre aquellos factores con verdadera capacidad de distorsionar los resultados. El caso de la caché del navegador es el perfecto ejemplo de ello, aunque no es el único.

A lo largo de la descripción de la arquitectura se ha hecho referencia también a los factores de influencia de la red. De producirse una segunda iteración en presente estudio, sería muy necesario disponer de un esquema de red en el laboratorio capaz de soportar el tráfico agregado en el caso de mayor ancho de banda.

Las pruebas realizadas en este estudio no requerían de mayores consideraciones puesto que el tráfico agregado de tres conexiones a 2048 Kbps es muy inferior a la capacidad de la LAN que lo soporta (100Mbps). Sin embargo, ante una condición de 50Mbps, la misma red LAN sólo podría gestionar dos UD a máxima tasa de transferencia, lo cual es un cuello de botella a tener muy en cuenta si se requiere más de dos equipos de usuario realizando el experimento en paralelo.

Queda por tanto claro que la aplicación de la recomendación ITU-T G.1031 [12] es clave en el proceso de definición de cualquier arquitectura de pruebas de evaluación de QoE y que el estudio realizado da buena cuenta de ello.

8.2. Líneas futuras

Como líneas futuras tras la realización del presente estudio, se puede plantear, a corto plazo, una nueva iteración en el ciclo de vida definido al comienzo del capítulo 6. Con dicha nueva iteración, podríamos obtener nuevos datos de QoE con un nuevo grupo de participantes, partiendo de los cambios introducidos en la recomendación ITU-T P.1501 [11]. Al tratarse de anchos de banda más coherentes con los valores domésticos contratados, y tiempos de condición con menos desgaste para el participante, es previsible obtener unos resultados más reales con los que quizá, para una muestra de población suficientemente grande, poder inferir una formulación que permita relacionar la experiencia de los usuarios en términos de MOS con parámetros puramente intrínsecos de QoS como es la medición de ancho de banda real que recibe el usuario respecto a su valor nominal contratado.

En el momento de redacción de estas líneas, de hecho, se han implementado los cambios aportados y se han considerado los nuevos factores de influencia derivados de dicha actualización de velocidades, aunque no se ha establecido en un período cercano de tiempo una nueva fase de experimentación.

En este momento, se está abogando por continuar con el desarrollo de QoXcloud, cuyos avances preliminares pudieron ser presentados en el congreso ITU Kaleidoscope 2014 celebrado en junio en San Petersburgo.

El objetivo es el de poder disponer de una plataforma en la nube, capaz de servir en la validación de recomendaciones y nuevas tecnologías, y ofrecer los resultados obtenidos a toda la comunidad de las telecomunicaciones.

A este respecto, se pretende mantener la colaboración con grupos de trabajo de la ITU, y abrir esta pretensión a otros círculos de interés, como el grupo de Conformidad e Interoperabilidad de la ITU [87], con la intención de avanzar en la definición de la plataforma en la nube de la ITU en la que se basa la propia arquitectura de QoXcloud.

A más largo plazo, quizá como proyecto de tesis doctoral, podría plantearse un estudio interdisciplinar que abordara los factores psicológicos de los participantes, de cara a obtener un acercamiento más profundo al concepto de QoE. Habida cuenta de la tendencia actual hacia la consideración del factor humano, puede ser de gran interés a la comunidad científica (tanto tecnológica como psicológica) la definición de un modelo metodológico que permita estudiar fehacientemente sobre un escenario incluso similar al propuesto en este estudio, la caracterización de aquellos factores cognitivos, perceptivos, sensoriales y/o físicos de la persona que influyen directamente en la QoE.

8.3. Publicaciones destacadas

A continuación se recogen las contribuciones más importantes en relación con el trabajo realizado en esta Tesis de Máster:

- **Título:** "A Cloud Platform for QoE Evaluation: QoXcloud"
Autores: Eduardo SAIZ, Eva IBARROLA, Leire CRISTOBO, Ianire TABOADA
Congreso: ITU Kaleidoscope 2014: Living in a converged world – impossible without standards?
Publicación: Proceedings
Lugar: San Petersburgo, Rusia
Fecha: 3-5 Junio 2014

- **Título:** "A new global quality of service model: QoXphere"
Autores: Eva IBARROLA, Eduardo SAIZ, Luis ZABALA, Leire CRISTOBO, Jin XIAO
Revista: IEE Communications Magazine
Publicación: Volumen 52, Número 1. Páginas 193-199
Fecha: Enero 2014

- **Título:** "Test results and new proposals for enhancement of Draft P.STMWeb"
(COM 12 – C 124 – E)
Autores: Eva IBARROLA, Eduardo SAIZ
Lugar: Reunión del SG-12 de la ITU. Ginebra, Suiza
Fecha: Diciembre 2013

- **Título:** "QoXphere: A new QoS Framework for Future Networks"
Autores: Eva IBARROLA, Eduardo SAIZ, Jin XIAO, Luis ZABALA, Leire CRISTOBO
Congreso: ITU Kaleidoscope 2013: Building Sustainable Communities
Publicación: Proceedings
Lugar: Kyoto, Japón
Fecha: 22-24 Abril 2013

9. GLOSARIO

9.1. Terminología

Calidad de Servicio y métricas

QoS	Quality of Service	Calidad de Servicio
QoS_D	QoS Delivered	QoS Entregada
QoS_O	QoS Offered	Calidad Ofrecida
QoS_R	QoS Required	Calidad Requerida / Requerimientos de QoS
QoS_P	QoS Perceived	Calidad Percibida
QoE	Quality of Experience	Calidad de Experiencia
QoUE	Quality of User Experience	Calidad de Experiencia de Usuario
QoCE	Quality of Customer Experience	Calidad de Experiencia de Cliente
QoBiz	Quality of Business	Calidad de Negocio
QoX	Global Quality of Service	Calidad Global de Servicio ¹²
FR	Full Reference (Metrics)	(Métricas de) Referencia Completa
NR	No Reference (Metrics)	(Métricas) Sin Referencia
RR	Reduced Reference (Metrics)	(Métricas de) Referencia Reducida

Equipos y módulos de la arquitectura de pruebas

EMU	Network Emulator	Emulador de Red
LCS	Local Content Server	Servidor Local de Contenidos
OP	Operator Front-end	Interfaz de Operador
PHM	Page Health Monitor	Monitor de Salud de las Páginas
SEQ	Sequencer	Secuenciador
UD	User Device	Dispositivo de Usuario

Otros términos

CSAT	Customer Satisfaction Model	Modelo de Satisfacción de Usuario
HF	Human Factor	Factor Humano
ISP	Internet Service Provider	Proveedor de Servicios de Internet
KPI	Key Performance Indicator	Indicador Clave de Desempeño
KQI	Key Quality Indicator	Indicador Clave de Calidad
LAN	Local Area Network	Red de Área Local
MOS	Mean Opinion Score	Escala Opinión Media
NGN	Next Generation Network	Red de Nueva Generación
NP	Network Performance	Rendimiento de la Red
NRA	National Regulation Agency	Agencia de Regulación Nacional
SLA	Service Level Agreement	Acuerdo de Nivel de Servicio
SSIM	Structural Similarity Index Metric	Métrica de Índice de Similitud Estructural
VQM	Video Quality Metric	Métrica de Calidad de Vídeo

¹² Definición del autor. QoX como tal no tiene una definición 'oficializada'.
La 'X' permite englobar todos los aspectos de la QoS.

9.2. Organizaciones y grupos

ETSI	European Telecommunications Standards Institute Instituto Europeo de Normas de Telecomunicaciones
ETSI STF #	ETSI Specialist Task Force #(Number) Grupo de Trabajo de Especialistas #(Número) de la ETSI
FTW	Forschungszentrum Telekommunikation Wien Centro de Investigación de las Telecomunicaciones de Viena
IEEE	Institute of Electrical and Electronics Engineers Instituto de Ingenieros Eléctricos y Electrónicos
ITU	International Telecommunication Union Unión Internacional de Telecomunicaciones
ITU-T	ITU – Telecommunication Standardization Sector ITU – Sector de Normalización de las Telecomunicaciones
ITU-R	ITU – Radiocommunication Standardization Sector ITU – Sector de Normalización de las Radiocomunicaciones
ISO	International Organization for Standardization Organización Internacional de Normalización
TMF	Telemanagement Forum
VQEG	Video Quality Experts Group Grupo de Expertos de Calidad de Vídeo

10. REFERENCIAS

- [1] ITU-T, G.1000: "Communications quality of service: A framework and definitions", 2001.
- [2] ITU-T, P.10/G.100 (2006) Amendment 2 (07/08): "New definitios for inclusion in Recommendation ITU-T P.10/G.100", 2008.
- [3] ITU-T. M.3050.1: "Enhanced Telecom Operations Map (eTOM) – The business process framework", 2007.
- [4] R. Stankiewicz et al, "QoX: What is it really?". Communications Magazine, IEEE, vol 49, pp. 148-158, 2010.
- [5] A.v. Moorsel, "Metrics for the Internet Age: Quality of Experience and Quality of Business", at 5th Performability Workshop. Erlangen, Germany, 2001.
- [6] K. Kilkki, "Quality of Experience in Communications Ecosystem", Journal of Universal Computer Science, vol. 14, pp. 615-624, 2008.
- [7] E. Ibarrola, E. Saiz, J. Xiao, L. Zabala & L. Cristobo, "QoXPHERE: A new QoS framework for future networks", in ITU Kaleidoscope 2013: Building Sustainable Communities (K-2013), 2003 Proceedings of, 2013, pp. 1-7.
- [8] ITU-T, SG-12. COM 12 – C 337 – E: "Draft P.STMWeb", A1 Telekom Austria AG. March 2013.
- [9] ITU-T, SG-12. COM 12 – C 46 – E: "Draft Test Plan for P.STMWeb". A1 Telekom Austria AG. March 2013.
- [10] ITU-T, SG-12. COM 12 – C 34 – E: "G.QoE-Web: Relevant factors and Use Cases for Web QoE", A1 Telekom Austria AG. March 2013.
- [11] ITU-T, P.1501: "Subjective testing methodology for web browsing", 2014.
- [12] ITU-T, G.1031: "QoE factors in web-browsing", 2014.
- [13] Networking Quality and Security (NQaS) Research Group – UPV/EHU
<http://det.bi.ehu.es/NQAS>
- [14] E. Saiz, E. Ibarrola, L. Cristobo and I. Taboada, "A cloud platform for QoE evaluation: QoXCLOUD", in ITU Kaleidoscope 2014: Impossible without standards? (K-2014). Saint Petersburg, Russian Federation, June 2014.
- [15] TMF GB935, "Business Metrics Concepts and Principles", Release 7.1.1.
- [16] L. Cristobo, "Definición de un modelo para la gestión de la calidad de negocio (QoBiz) en los servicios de telecomunicación", Tesis de Máster TICRM, UPV/EHU, Curso académico 2012/2013.
- [17] R. Partearroyo et al, "QoSmeter: Generic quality of service measurement infrastructure", at IFIP Networking 2006 workshop 'Towards the QoS Internet' (To-QoS'2006), Coimbra, Portugal, 2006.
- [18] L. Zabala et al, "LabQoS: A platform for network test environments", in ITU-T Kaleidoscope 2011: The fully networked human? – Innovations for future networks and services. Cape Town, South Africa, 2011.
- [19] E. Ibarrola, "Propuesta de una metodología para la aplicación de un modelo para la gestión global de la QoS en un ISP en el marco de la UIT-T G.1000". Tesis doctoral, UPV/EHU, 2012.
- [20] ETSI EG 202 057-4: "Speech processing, transmission and Quality aspects (STQ); User related QoS parameter definitions and measurements; Part 4". V1.2.1, 2008.
- [21] ITU-T, E.800: "Terms and definitions related to quality of service and network performance including dependability", 1994.

- [22] ITU-T, I.350: "General aspects of quality of service and network performance in digital networks, including ISDNs", 1993.
- [23] ETSI ETR 003: "Network Aspects (NA); General aspects of Quality of Service (QoS) and Network Performance (NP)", 1994.
- [24] M.J. Bitner, B.H. Booms & L.A. Mohr, "Critical Service Encounters: The Employee's Viewpoint". *Journal of Marketing*, 58 (Author), 95-106. 1994.
- [25] F. Liberal, A. Ferro & J.O. Fajardo, "Application of a PQoS Based Quality Management Model to Identify Relative Importance of the Agents", at the IEEE ICIS (Bangkok), ISBN: 0-7803-9283-3. Bangkok, 2005.
- [26] A. Bouch & M.A. Sasse, "It ain't what you charge, it's the way that you do it: a user perspective of network QoS and pricing", at Integrated Network Management. Distributed Management for the Networked Millennium. Proceedings of the Sixth IFIP/IEEE International Symposium. Boston, MA, USA, 1999.
- [27] A. Bouch & M.A. Sasse, "Network Quality of Service – An Integrated Perspective", at the 4th International Distributed Conference (IDC'99). Madrid, 1999.
- [28] ITU-T, E.800: "Definitions of terms related to Quality of Service", 2008.
- [29] ISO 9000: "Quality management systems – Fundamentals and vocabulary", 2005.
- [30] ETSI TS 102 250-1: "Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 1: Identification of Quality of Service criteria", V2.2.1, 2011.
- [31] European Network on Quality of Expertise in Multimedia Systems and Services, Qualinet – ICT Action COST IC1003. <http://www.qualinet.eu>
- [32] European Cooperation in Science and Technology, COST. <http://www.cost.eu>
- [33] Qualinet (COST Action IC 1003), "Qualinet White Paper on Definitions of Quality of Experience (2012)". Lausanne, Switzerland, Version 1.2, March 2013
- [34] L.A. Pervin & O.P. John, "Handbook of Personality theory and research", The Guilford Press, pp. 102-138, 2001. Cited from Carducci, "The Psychology of Personality", John Willey & Sons, 2009.
- [35] ITU-T, E.860: "Framework of a service level agreement", 2002.
- [36] K. Laghari, N. Crespi & K. Conelly, "Toward total Quality of Experience: A QoE model in a Communication Ecosystem", *Communications Magazine*, IEEE, vol 50, issue 4, pp. 58 – 65. 2012.
- [37] M. Varela, J.P. Laulajainen, "QoE-driven mobility management integrating the users' Quality Perception into network-level decision making", 3rd International Workshop on Quality of Multimedia Experience (QoMEX), Belgium. 2011
- [38] T. Hoßfeld, R. Scharz, M. Seufert, M. Hirth, T. Zinner & P. Tran-Gia, "Quantification of YouTube QoE via Crowdsourcing", IEEE International Workshop on Multimedia Quality of Experience – Modeling, Evaluation and Directions (MQoE 2011), Proceedings of, 2011.
- [39] J. Xiao & R. Boutaba, "Assessing Network Service Profitability: Modeling From Market Science Perspective", *Networking*, IEEE/ACM Transactions on, vol. 15, no. 6, pp. 1307-1320. 2007.
- [40] F. Liberal, "Propuesta de un modelo y una metodología para la gestión de la calidad en los servicios de telecomunicación". Tesis doctoral, UPV/EHU, 2005.
- [41] E.W. Anderson & M.W. Sullivan, "The Antecedents and Consequences of Customer Satisfaction for Firms", *Marketing Science*, vol. 12, no. 2, pp. 125-143. Spring, 1993.

- [42] B. Schindlholzer, "The confirmation/disconfirmation paradigm: Why satisfied customers are not always satisfied" - April, 2008. <http://www.diametrics.io/the-confirmationdisconfirmation-paradigm-why-satisfied-customers-are-not-always-satisfied.html>
- [43] ITU-T, SG-12. COM 12 – C 30 – E: "Proposal on enhancement of G.1000 framework". March 2013.
- [44] E. Ibarrola, E. Saiz, J. Xiao, L. Zabala & L.Cristobo, "A new global Quality of Service model: QoXPHERE". Communications Magazine, IEEE, vol 52, issue 1, pp. 193-199. January 2014.
- [45] ITU-T, G.1030: "Estimating end-to-end performance in IP networks for data applications", 2014.
- [46] P. Brooks & B. Hestnes, "User measures of Quality of Experience: Why being objective and quantitative is important", Network, IEEE, vol. 24, issue 2, pp. 8-13. 2010.
- [47] M. Fiedler, T. Hossfeld & P. Tran-Gia, "A generic quantitative relationship between Quality of Experience and Quality of Service", Network, IEEE, vol. 24, issue 2, pp. 36-41. 2010.
- [48] ITU-T, P.800: "Methods for subjective determination of transmission quality", 1996.
- [49] ITU-T, P.800.1: "Mean Opinion Score (MOS) terminology", 2006.
- [50] ITU-T, G.1040: "Network contribution to transaction time", 2006.
- [51] ITU-T, G.1010: "End-user multimedia QoS categories", 2001.
- [52] E. Ibarrola, F. Liberal, I. Taboada & R. Ortega: "Web QoE Evaluation in Multi-Agent Networks: Validation of ITU-T G.1030", Fifth International Conference on Autonomic and Autonomous Systems (ICAS'09), Proceedings of, pp. 289-294. 2009.
- [53] ETSI TR 102 643: "Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services", V.1.0.2, 2010.
- [54] T. Zinner, O. Hohlfeld, O. Abboud & T. Hossfeld, "Impact of Frame Rate and Resolution on Objective QoS Metrics", Second International Workshop on Quality of Multimedia Experience (QoMEX), Proceedings of, pp. 29-34. 2010.
- [55] Z. Wang, A. Bovik, H. Sheikh & E. Simoncelli, "Image quality assessment: From error visibility to structural similarity", IEEE Transactions on Image Processing, Vol 13, pp. 600-612, 2004.
- [56] Z. Wang, L. Lu & A. Bovik, "Video quality assessment base don structural distortion measurement", Signal Porcessing: Image Communication, 19(1), January 2014.
- [57] ITU-T, J.144: "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference", 2004.
- [58] ITU-R, BT.1683: "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference", 2004.
- [59] L. Fallon & D. O'Sullivan, "Towards determining web browsing Quality of Experience", 12th Irish Information Technology & Telecommunications (IT&T) Conference, Proceedings of, pp. 117-124. 2013.
- [60] G. Maier, A. Feldman, V. Paxson & M. Allman, "On dominant characteristics of residential broadband internet traffic", ACM SIGCOMM conference on Internet measurement conference (IMC'09), Proceedings of, pp. 90-102. New York, USA. 2009.
- [61] iMacros for Firefox
<https://addons.mozilla.org/es/firefox/addon/imacros-for-firefox/>
- [62] HttpFox. HTTP analyzer addon for Firefox
<https://addons.mozilla.org/es/firefox/addon/httpfox/>
- [63] S. Ihm & V.S. Pai, "Towards understanding modern web traffic", ACM SIGCOMM conference on Internet measurement conference (IMC'11), Proceedings of, pp. 295-312. New York, USA. 2011.

- [64] H. Cui & . Biersack, "On the Relationship between QoS and QoE for web sessions", EURECOM Department of Networking and Security, France. Research Report RR-12-263. 2012
- [65] A. López Monge, "Aprendiendo a programar con Libpcap". 2005.
<http://www.e-ghost.deusto.es/docs/2005/conferencias/pcap.pdf>
at TCPDump Project Documentation (<http://www.tcpdump.org/#documentation>)
- [66] G. Ortega Castillo, "Clasificación análisis de clústeres: Clustering", Inteligencia Artificial Avanzada, Universidad Técnica Particular de Loja. http://es.slideshare.net/geoc_hhga/clustering
- [67] boxplot. MatWorks Documentantion Center
<http://www.mathworks.es/es/help/stats/boxplot.html>
- [68] L.T. Nguyen, R. Harris & A. Punchihewa, "Assessment of Quality of Experience for web browsing - as function of Quality of Service and Content Factors", 5th International Conference on Ubiquitous and Future Networks (ICUFN), Proceedings of, pp. 764-769. 2013.
- [69] netem (NETwork EMulation). Linux Foundation
<http://www.linuxfoundation.org/collaborate/workgroups/networking/netem>
- [70] S. Egger, T. Hossfeld, R. Schatz & M. Fiedler, "Waiting times in Quality of Experience for web based services", 4th International Workshop on Quality of Multimedia Experience (QoMEX), Proceedings of, pp. 86-96. 2012.
- [71] mod_qos: Quality of Service module for the Apache web server. Opensource
http://opensource.adnovum.ch/mod_qos/
- [72] L.T. Nguyen, R. Harris & J. Jusak, "A pilot study to assess Quality of Experience base don varying network parameters and user behaviour", Computer Science and Information Tecnology, Proceedings of, vol. 5, pp. 30-34. 2010.
- [73] L.T. Nguyen, R. Harris, J. Jusak & A. Punchihewa, "Modelling of Quality of Experience for Web Traffic", 2nd International Conference on Network Applications Protocols and Services (NETAPPS), Proceedings of, pp. 84-89. 2010.
- [74] L.T. Nguyen, R. Harris & J. Jusak, "Analysis of Networking and Application layer derived metrics for web Quality of Experience", Consumer Communications and Networking Conference (CCNC), Proceedings of, pp. 321-325. 2012.
- [75] ETSI EG 202 670: "Human Factors (HF); User Experience Guidelines for real-time communication services expressed in Quality of Service terms", V.1.1.2, 2010.
- [76] S.C. Seow, "Designing and Engineering Time: The Psychology of Time Percepcion in Software". Addison-Wesley Professional ed., 2008.
- [77] G. Antonides, P.C. Verhoef & M. van Aalst, "Consumer perception and evaluation of waiting time: A field experiment", Journal of Consumer Psychology, 12(3):193202, 202.
- [78] S.H. Jeong, "QoS/QoE for Multimedia Applications/Services", ITU Academy, 2012.
http://academy.itu.int/moodle/pluginfile.php/39887/mod_resource/content/1/session13.pdf
- [79] ITU-T, Y.1541: "Network performance objectives for IP-based services", 2011.
- [80] wondershaper: Command-line utility for limiting an adapter's bandwidth. GitHub
<https://github.com/magnifico/wondershaper>
- [81] iptables. Netfilter.org
<http://www.netfilter.org/projects/iptables/>
- [82] cron daemon. Ubuntu Documentation
<https://help.ubuntu.com/community/CronHowto>

- [83]** S. Egger, R. Schatz & B. Gardlo, "Interactive content for subjective studies on web browsing QoE: A Kepler derivative", FTW, 2012. http://docbox.etsi.org/workshop/2012/201211_STQWORKSHOP/S8_QoEandFutureConsideration/FTW_EGGER.pdf
- [84]** Europa Press – Feeds de noticias
<http://www.europapress.es/contenidosrss.aspx>
- [85]** ITU-T, SG-12. COM 12 – C 124 – E: "Test results and new proposals for enhancement of Draft P.STMWeb". University of the Basque Country. December 2013
- [86]** Ministerio de Industria, Energía y Turismo, "Telecomunicaciones y Sociedad de la Información: Calidad de Servicio". <http://www.minetur.gob.es/telecomunicaciones/es-ES/Servicios/CalidadServicio/Paginas/Calidad.aspx>
- [87]** ITU Conformity and Interoperability (C&I)
<http://www.itu.int/es/ITU-T/C-I/Pages/default.aspx>

TESIS DE MÁSTER

TICRM

Máster y Doctorado en Tecnologías de la
Información y Comunicaciones en Redes Móviles

ANEXOS

Estudio de la Calidad Experimentada (QoE) en el Servicio de Navegación Web:
Contribución a la Validación de las Recomendaciones ITU-T P.1501 e ITU-T G.1031

ANEXO I	Draft P.STMWeb & Draft Test Plan P.STMWeb	ANX 01
ANEXO II	Contribución al Draft P.STMWeb	ANX 21
ANEXO III	Diseño de la plataforma de evaluación de QoE.....	ANX 33
ANEXO IV	Instrucciones para la realización del experimento.....	ANX 59