

Creación de un Sistema de Audiencia Virtual. Flujo y tecnología en un sistema de información audiovisual sincrónico con un espectáculo en directo

Creation of a Virtual Audience System. Flow and Technology in an Audiovisual Information System Synchronous with a Live Show

Daniel Torras-i-Segura

Cómo citar este artículo.

Torras-i-Segura, Daniel (2024). "Creación de un Sistema de Audiencia Virtual. Flujo y tecnología en un sistema de información audiovisual sincrónico con un espectáculo en directo". *Profesional de la información*, v. 33, n. 4, e330419. <https://doi.org/10.3145/epi.2024.0419>

Artículo recibido el 17-05-2024
Aceptación definitiva: 17-07-2024



Daniel Torras-i-Segura ✉

<https://orcid.org/0000-0002-2639-3821>

Universitat Pompeu abra, TecnoCampus
Av. Ernest Lluch, 32 (edificio TCM1)
08302 Mataró (Barcelona), España
dtorras@tecnocampus.cat

Resumen

Las restricciones de la pandemia forzaron la ideación de un sistema de información que integrase la participación remota del público en un espectáculo televisivo en directo. Los mecanismos existentes entonces eran demasiados rígidos, de poca calidad audiovisual y con un retardo entre emisión y recepción excesivo e insostenible para una interacción televisiva en tiempo real. La creación del *Sistema de Audiencia Virtual* (VAS) planteó la utilización de la más reciente tecnología en ámbitos como *cloud* y *edge computing*, la transmisión de audio y video en baja latencia, la comunicación interactiva en el entorno web, y la creación de sonido binaural inmersivo para superar las limitaciones del momento. Con estos nuevos elementos tecnológicos se desarrolló un nuevo sistema de transmisión y recepción de información audiovisual que permite la adaptación y escalabilidad de la producción mediática al tipo de evento transmitido y su público con una mínima inversión y, al mismo tiempo, genera un nuevo concepto de audiencia propio del siglo XXI. El *Sistema de Audiencia Virtual* posibilita una participación interactiva y sincronizada en programas de entretenimiento en directo y permite a la vez la creación de participantes virtuales en remoto, a la vez que garantiza una experiencia real e inmersiva para el usuario. La tecnología WebRTC, RTMP y SRT utilizada, ya popularizada y con compatibilidad universal; la baja latencia conseguida, por debajo de dos segundos para el ciclo completo, y la sensación inmersiva del público predicen que este sistema puede transformar en breve la metodología de producción televisiva de los espectáculos con participación de público en directo.

Palabras clave

Audiencia virtual, Participación interactiva, Sensación inmersiva, Tiempo real, Emisión mediática, Baja latencia, Sincrónica, Remota, Audiovisual, Televisión, Programas de entretenimiento.

Abstract

The restrictions of the pandemic forced the idea of an information system that integrated remote public participation in a live television show. The existing mechanisms were then too rigid, of poor audiovisual quality, and had an excessive and unsustainable delay between emission and reception for real-time television interaction. The creation of the *Virtual*



Audience System (VAS) proposed the use of the most recent technology in areas such as cloud and edge computing, low-latency audio and video transmission, interactive communication in the web environment, and the creation of immersive binaural sound to overcome the limitations of the moment. With these new technological elements, a new system of transmission and reception of audiovisual information was developed that allows the adaptation and scalability of media production to the type of event broadcasted and its audience with a minimum investment and, at the same time, generates a new concept of audience appropriate for the 21st century. The *Virtual Audience System* enables interactive and synchronized participation in live entertainment shows and, at the same time, allows the creation of remote virtual participants while guaranteeing a real and immersive experience for the user. The WebRTC, RTMP and SRT technology used, which is already popular and has universal compatibility; the low latency achieved, which is below two seconds for the complete cycle; and the immersive sensation of the audience—everything together predicts that this system can soon transform the television production methodology of shows with live audience participation.

Keywords

Virtual Audience, Interactive Participation, Immersive Sensation, Real Time, Broadcasting, Low-Latency, Synchronic, Remote, Audiovisual, Television, Entertainment Programs.

1. Introducción

VAS es el acrónimo de *Virtual Audience System* en inglés que denomina un sistema de gestión de datos audiovisuales y de emisión ideado para programas televisivos de entretenimiento en directo. Este sistema de información de audiencia virtual recrea una experiencia inmersiva para el público online, sobre todo porque consigue una baja latencia en la sincronización entre emisión y recepción. Sin entrar en formulaciones técnicas, el presente artículo detalla su configuración, arquitectura técnica e implementación para después apuntar y destacar las nuevas aportaciones más relevantes de este sistema de información mediática.

El origen del VAS está relacionado con la irrupción de la pandemia de Covid-19, la cual forzó en 2020 medidas restrictivas del contacto social presencial en diversos países. En marzo de ese mismo año, en España se estaba realizando la emisión de la temporada 11 de la producción del concurso de talentos (*talent show*) *Operación Triunfo*. La productora *Gestmusic (Banijay Iberia)* junto con RTVE se vieron obligados a plantear una alternativa para mantener la participación del público —un elemento crucial en este formato de programa televisivo— dado que no era posible su presencia en plató ante la aprobación del Decreto de Alarma (*Gestmusic, 2021*).

A partir de una emisión forzada con participación online de baja calidad televisiva se planteó la invención de un sistema que superase estas limitaciones, con la vista puesta en la posible evolución de la televisión y de la producción de entretenimiento para este medio (*Alsina, 2024*). Por lo tanto, con el objetivo general de mejorar la experiencia de entretenimiento en directo en las condiciones establecidas por la pandemia —es decir, con la única posibilidad de la participación del público online, desde múltiples y diversas ubicaciones en remoto— el sistema VAS plantea un proyecto de arquitectura flexible y escalable, con mejor calidad de imagen, audio inmersivo, interacción sincrónica a distancia y todo con baja latencia (*Alsina, 2024*).

Este estudio pretende explicar la posibilidad técnica actual de mejorar la participación del público virtual en los espectáculos en directo de televisión mediante el análisis del caso del proyecto VAS. Paralelamente, a través de este análisis se apuntan observaciones y tendencias sobre la tipología de interacción del público virtual; la adecuación del sistema VAS a los patrones de consumo televisivo actuales; y, como síntesis, la concreción conceptual de este nuevo sistema de participación televisiva.

VAS plantea una arquitectura flexible y escalable, con buena calidad de imagen, audio inmersivo, interacción sincrónica a distancia y todo con baja latencia

2. Estado de la cuestión

En el proyecto analizado en este artículo son centrales e intervienen diferentes tecnologías y conceptos como el *cloud computing*, el *edge computing*, el sonido binaural inmersivo, la tecnología Ambisonics, y los protocolos SRT, WebRTC y RTMP, principalmente. En este apartado revisaremos sus definiciones básicas y qué se ha teorizado al respecto.

2.1. Tecnología ‘cloud computing’

La tecnología *cloud computing* o tareas y servicios de computación en la nube se basa en utilizar bajo previa demanda equipos y recursos remotos a través de la red de internet o bien ubicados en esta. La finalidad de esta tecnología es evitar el coste elevado de la adquisición de equipos, su mantenimiento y, al mismo tiempo, la ocupación de espacio físico y la inversión de tiempo en su instalación (*Nigro, 2022*). También, como en el caso de proyecto que se estudia en este artículo, la tecnología *cloud computing* permite más flexibilidad y posibilidades de adaptación tanto en la diversidad de emisores como de la audiencia receptora (*Mell; Grance, 2011*). En resumen, el *cloud computing* permite a las

empresas evitar el aprovisionamiento, la configuración y la gestión de equipos y servicios informáticos y, en cambio, contratar y pagar únicamente aquellos que realmente necesitan y finalmente utilizan.

Hoy en día, el *cloud computing* se ha convertido en un estándar tecnológico y existen gran variedad de servicios *cloud computing* basados en software (SaaS), plataformas (PaaS), infraestructuras (IaaS), nubes públicas, privadas, híbridas o comunitarias, entre otras variaciones (Joyanes Aguilar, 2018; Nigro, 2022).

2.2. Tecnología ‘edge computing’

La tecnología *edge computing* o computación en el borde (o en la frontera) es un nuevo paradigma de la arquitectura computacional aparecido alrededor de 2002 que se basa en procesar los datos del internet de las cosas (o *IoT*) en la periferia de la red, dentro de la nube; es decir, procesar los datos dentro de la misma fuente que los genera o tan cerca de ella como sea posible (Mela *et al.*, 2021). Tal y como comenta Rebato (2022), el *edge computing* “consiste en acercar el poder de procesamiento lo más cerca posible de donde los datos están siendo generados”, o lo que sería lo mismo, “consiste en acercar la nube hasta el usuario, hasta el borde mismo (*edge*, en inglés) de la red”.

2.3. Audio inmersivo binaural

El término binaural hace referencia a las diferencias entre los parámetros acústicos en las señales captadas y procesados por los dos oídos en un ser humano. Según la localización de la fuente sonora, un mismo sonido puede recibirse con un cierto retraso en el oído del lado opuesto a la fuente sonora (oído contralateral). Esta mínima diferencia de tiempo interaural (DTI) representa una variación de fase que el cerebro interpreta como la localización espacial de la fuente. También puede procesarse una diferencia de energía sonora entre los dos oídos ipsilateral y contralateral, denominada diferencia de nivel interaural o ILD (De-Muynke, 2024).

Una recreación de las condiciones naturales de recepción binaural –con diferencias DTI e ILD– a través de transductores de reproducción sonora (auriculares o altavoces) emula la experiencia de estar presente en el lugar del evento, simulando la captación directa de la localización espacial y de las diferentes intensidades de las fuentes sonoras presentes, como si el oyente estuviera inmerso en el entorno sonoro real. De ahí la denominación de audio inmersivo para aquel sistema que presenta al espectador una esfera sonora envolvente construida con diferencias DTI e ILD. Igualmente, otros factores como la reverberación (DRR) también añaden características del entorno (De-Muynke, 2024).

Una recreación mediante transductores de las condiciones naturales de recepción binaural emula la experiencia de estar presente en el lugar del evento

2.4. Tecnología Ambisonics

La tecnología Ambisonics produce un formato de audio tridimensional que permite localizar fuentes sonoras esféricamente, es decir, con posiciones arriba, debajo y detrás del oyente además de las ya introducidas en el plano horizontal por formatos anteriores (Narbutt *et al.*, 2020). Ambisonics se basa en una descripción y codificación matemática de un campo sonoro en tres dimensiones que convierte una señal de audio en una serie de armónicos esféricos (Boren, 2017). Por esta razón, El efecto inmersivo de la tecnología Ambisonics es independiente del número y del tamaño de los altavoces utilizados (Heller; Benjamin, 2018).

El proyecto VAS desarrolla un decodificador (*decoder*) propio para el contenido Ambisonics con la finalidad de optimizar su uso con auriculares, aunque manteniendo también un funcionamiento posible y correcto con altavoces (Alsina, 2024). La función principal de un decodificador es crear la sensación perceptiva de que el escenario sonoro emulado se está reproduciendo lo más precisamente posible (Heller; Benjamin, 2018). La opción de un decodificador pensado especialmente para auriculares encaja con el perfil virtual de la audiencia y sus hábitos de consumo audiovisual, principalmente, a través de internet y con dispositivos móviles y personales (Bull, 2010; Andersson Schwarz, 2016).

2.5. Protocolo SRT

SRT, de sus siglas en inglés *Secure Reliable Transport*, es un protocolo –reglas establecidas para transmitir datos entre dispositivos– de código abierto, creado por la empresa canadiense Haivision en 2013 y orientado al transporte de señal de vídeo y audio con ultra baja latencia (Carrillo, n.d.). Su característica principal es que puede gestionar “una gran calidad de imagen en redes no gestionadas e impredecibles” (Carrillo, n.d.). Al poder implementarse sin licencia, no representa ningún coste de compra para las empresas que lo utilizan (Lee; Cha, 2023).

SRT es independiente del sistema codificador-decodificador (códec) y garantiza un envío sin pérdidas, ya que dispone de instrucciones de recuperación de paquetes perdidos en baja latencia. La seguridad del enlace se realiza bajo encriptación y puede atravesar el *firewall* o barrera de seguridad en toda su transmisión, al especificar modos de envío y de recepción (Anitua, n.d.).

SRT es capaz de detectar la latencia producida y adaptarse al ancho de banda disponible y a las condiciones de transmisión para mantener un enlace óptimo. Por lo tanto, con el protocolo SRT “se puede realizar con éxito una transmisión y distribución de vídeo de alta definición segura y confiable en entornos de Internet comunes y entre múltiples ubicaciones” (Gestmusic, 2021).

2.6. Protocolo WebRTC

WebRTC es el acrónimo y reducción de *Web Real-Time Communications*, en referencia a las instrucciones para la comunicación en tiempo real, en directo, a través de la plataforma web (Anitua, n.d.). WebRTC es un protocolo de código abierto con API (*Interfaz de Programación de Aplicaciones*, un conjunto de parámetros y protocolos que permiten que dos componentes de software se comuniquen entre ellos) de JavaScript, desarrollado en 2011 para compartir videos entre navegadores o aplicaciones móviles. Puede utilizarse en páginas web o en comunicaciones entre usuarios sin necesidad de instalar software o aplicaciones (Loreto; Romano, 2012).

Al mismo tiempo, WebRTC permite la transmisión de voz y video con muy baja latencia, casi instantánea – con una latencia de entrega de menos de 500 milisegundos– con cualquier tipo de navegador en el envío y en la recepción (Anitua, n.d.). También trabaja con cualquier tipo de códec por lo que se le califica de “códec agnóstico” (Blum *et al.*, 2021). Actualmente, es un estándar imprescindible de la comunicación por redes en tiempo real.

WebRTC es un protocolo de código abierto para compartir videos entre navegadores o aplicaciones móviles que permite la transmisión de voz y video con muy baja latencia

2.7. Protocolo RTMP

El protocolo de mensajería en tiempo real o *Real-Time Messaging Protocol* (RTMP) fue lanzado por *Macromedia* (Adobe) para la transmisión de muy baja latencia en tiempo real o *streaming* de video, audio y datos entre un servidor y un reproductor Flash (Anitua, n.d.). Para este propósito, RTMP fue estable y seguro y aun siendo uno de los protocolos más antiguos del mercado todavía se mantiene como protocolo en la transmisión en vivo, aunque su formato original, el Flash, está cada vez más en desuso (Anitua, n.d.).

RTMP está fundamentado en el TCP o *Transmission Control Protocol* original de Internet y se constituye como una tecnología de transmisión continua con comunicación constante entre el cliente y el servidor (Nurrohman; Abdurohman, 2018). Tal y como explica Anitua (n.d.), “una vez recibida una secuencia de paquetes se notifica al servidor. Si hay paquetes perdidos dentro de esa secuencia, se retransmitirá la secuencia completa de paquetes”. Por esta razón, la latencia “puede aumentar notablemente en una transmisión en directo” y es uno de sus aspectos mejorables. Aun así, aplicando ciertas tecnologías adicionales RTMP puede conseguir una latencia inferior a 150 ms en la transmisión de video (Nurrohman; Abdurohman, 2018).

3. Metodología

Para la elaboración de este estudio, en primer lugar, se han examinado críticamente las memorias y documentación del proyecto VAS que corresponderían a las referencias Alsina (2024); Alsina *et al.* (2024); y Gestmusic (2021). Después de este escrutinio, los conceptos centrales que incluye esta documentación se han contrastado y documentado a través de la literatura científica más reciente, validada por fuentes científicas y del sector tecnológico. En esta fase se sigue mayormente la propuesta de Templier y Paré (2015) para el análisis crítico de literatura científica. Posteriormente, se ha entrevistado a los responsables técnicos del proyecto mediante la técnica de una entrevista estructurada para aclarar y contrastar ideas, propósitos y conceptos.

En una fase posterior, se han analizado los datos experimentales del proyecto obtenidos de la documentación y las entrevistas y contrastados por la investigación mediante las técnicas Scrum y *Agile Testing* (Gestmusic, 2021). Igualmente, se han comparado los resultados con la planificación y las intenciones iniciales, así como con las aportaciones de otros académicos del área. Finalmente, se han observado los resultados audiovisuales de una transmisión en directo focalizando la atención en la velocidad de la interacción entre los participantes, el flujo dinámico del espectáculo, la inteligibilidad del contenido y la calidad técnica audiovisual.

4. Resultados

El proyecto VAS se proponía conseguir un audio y video de calidad, sin *delay* o retraso de audio ni de video, con una ingesta de datos multipunto sólida y eficaz, con seguridad y protección del contenido, y con una experiencia inmersiva lo más parecida al directo en plató para el participante remoto y usuario virtual. El proyecto planificó materializar estas características en un prototipo –inicialmente para uso interno– cien por cien ubicado en la nube y basado en el formato *Platform-as-Service* (PaaS).

El ensamble y desarrollo de este sistema de información audiovisual se realizó dividiendo el proyecto en tres ámbitos: el estudio de grabación o el evento, los módulos en la nube, y, finalmente, la solución tecnológica del espectador.

En la idea inicial, en el estudio o plató se configuró una “caja negra” [sic.] (Gestmusic, 2021) con un número adecuado de señales audiovisuales SDI/NDI de entrada. La caja negra realiza la función de codificador/decodificador traduciendo las señales SDI a un código compatible con la emisión por *streaming* y viceversa. Es el enlace conversor entre los mensajes de internet de la audiencia y el plató (C. Flamarique, comunicación personal, 18 de abril de 2024) (Ver Figura 1. Esquema inicial del sistema VAS).

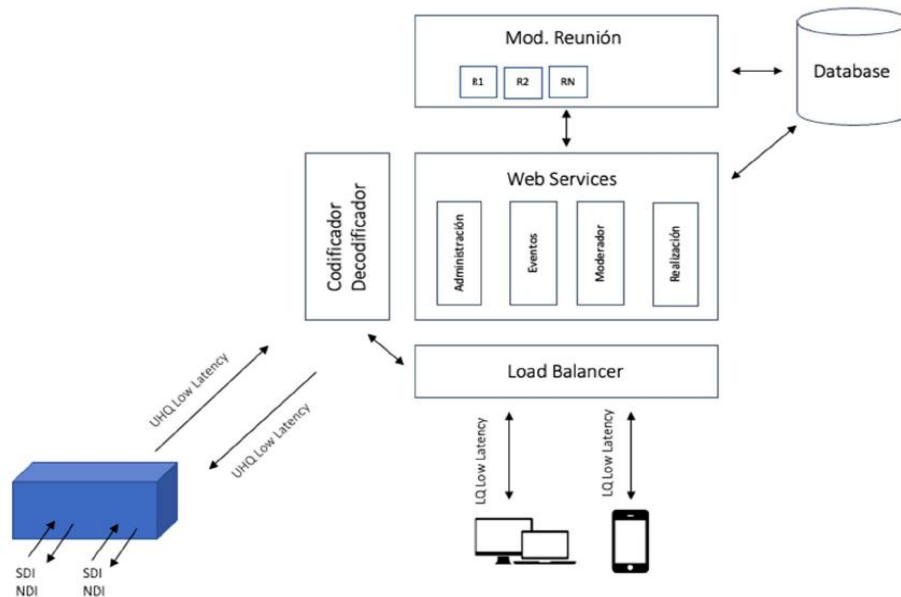


Figura 1: Esquema Inicial del Sistema VAS. Fuente: Alsina (2024).

Inicialmente, en la nube se planificó un *web services*, un *load balancer*, un *cloud database* y un módulo denominado ‘reunión’. El *web services* permite intercambiar datos entre aplicaciones y se compone del módulo administración (para configuración y gestión de los eventos), el módulo gestión del evento (gestionar espectadores, parrilla, espectador en modo participante, etc.), el módulo moderador (animación y comunicación con el público virtual) y el módulo realización (realizar la muestra de distintos planos de video). El *load balancer* por su parte distribuye la carga de trabajo entre distintos servidores y aplicaciones. El *cloud database* es una base de datos en la nube y que se ejecuta desde esta. Y, por último, el módulo reunión gestiona la conexión con la ‘caja negra’ del estudio, con los espectadores y con otros puntos (Gestmusic, 2021).

Según el esquema inicial, el cliente o espectador solamente recibe la señal del evento en su dispositivo con una alta calidad de imagen y sonido (HQ). Se previó que el sistema VAS fuera compatible con dispositivos de escritorio (PC/Mac) y con aplicaciones de Android, iOS i Windows. La señal que envía el cliente al programa varía si este es espectador (baja calidad o LQ) o si es participante (calidad media o MQ) (Gestmusic, 2021).

Después de la implementación y la fase de pruebas, el esquema final tiene como epicentro la aplicación web VAS donde se reciben y se realizan los videos de los participantes y desde donde se envía la señal a la mesa de realización de plató. El video realizado desde plató o el evento se envía y distribuye directamente al público virtual, sin procesarlo de nuevo en la aplicación web (ver Figura 2. Esquema definitivo del sistema VAS).

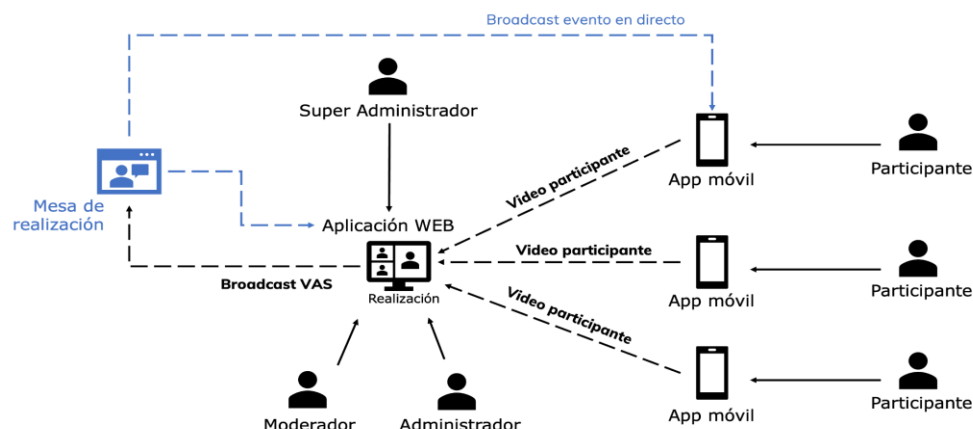


Figura 2: Esquema Definitivo del Sistema VAS. Fuente: Alsina et al. (2024).

Las diferentes pruebas técnicas realizadas validan el uso de la tecnología WebRTC, RTMP y SRT, así como la mediación de la plataforma de emisión de video a baja latencia *Vonage*. Esta plataforma es un servicio con coste de suscripción que agrupa emisores o *streams*, ofrece la herramienta *OpenTok Playground* para configurar las sesiones en *Vonage*, y, al mismo tiempo, permite realizar la imagen de los diferentes participantes virtuales. Por otro lado, finamente se descarta el uso del formato LL-HLS y de dispositivos como el NDI (Alsina et al., 2024).

El esquema del flujo de la emisión del plató (o evento) a los participantes (ver Figura 3. *Flujo de evento a participantes*) queda configurado con la utilización del protocolo WebRTC entre la emisión de plató y la plataforma *Vonage*, así como la utilización de este mismo protocolo entre *Vonage* la aplicación web VAS y los clientes o participantes. Por otro lado, se incluye un servidor RTMP para enlazar el flujo de vuelta de *Vonage* a Plató a través del protocolo RTMP (Alsina et al., 2024)

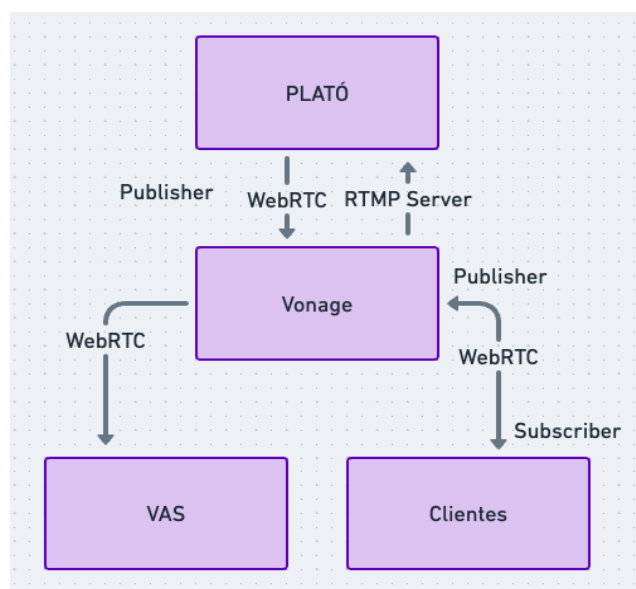


Figura 3: Flujo de Evento A Participantes. Fuente: Alsina et al. (2024).

La latencia desde la emisión de la señal del programa hasta la recepción por parte del público virtual se encuentra por debajo de un segundo en la prueba final (ver Tabla 1. *Pruebas, tecnologías y latencias*). Sin embargo, al considerar el ciclo completo, es decir, incluyendo el retorno a plató o realización de la respuesta en video de los participantes remotos y su devolución a los usuarios, la latencia final se sitúa entre 1 y 2 segundos (Alsina et al., 2024).

Tabla 1: Pruebas, Tecnologías y Latencias. Fuente: Elaboración propia.

Prueba	Tecnología	Latencia Obtenida	Observaciones
Previa	Zoom / Kit SDK, HD-SDI	Alta: >5s.	Baja calidad de imagen y sonido
Prueba 1	LL-HLS, WebRTC, Vonage	Alta: >5s.	+ OBS Studio para una cámara virtual
Prueba 2	RTMP, WebRTC, Vonage	Media: >3s.	Se requiere un servidor RTMP
Prueba 3	RTMP, WebRTC, Vonage	Baja: <2s. <1s. en un trayecto parcial	Se requiere un servidor RTMP

La solución final incluye entonces los protocolos WebRTC y RTMP, un servidor RTMP y el uso de la plataforma comercializada *Vonage*. Con esta configuración se obtiene una latencia en el ciclo completo inferior a dos segundos y una mayor calidad de imagen y sonido.

5. Discusión

El seguimiento del proceso de diseño e implementación del sistema de información VAS constata que aún no existe ninguna tecnología accesible y aplicable a esta arquitectura que pueda rebajar la latencia del ciclo completo de emisión-recepción por debajo de 1 segundo. Tal y como acertadamente indica **Anitua** (n.d.), el espectador contemporáneo demanda “inmediatez en un mundo digital”. Y se podría añadir que la demanda mediante una conexión audiovisual (Arjona-Martín; Victoria-Mas, 2018). Sin embargo, aún no se ha establecido o divulgado la tecnología que permita una conexión instantánea o inmediata –sin latencia efectiva para la percepción humana– entre una emisión mediática y sus espectadores remotos, la cual podría situarse por debajo de los 250 ms (Blum et al., 2021).

Específicamente, el proyecto VAS ha utilizado en gran medida tecnologías ya existentes, de código abierto algunas de ellas (como WebRTC, SRT y el reproductor VLC, por ejemplo), que se han ido perfeccionando colectivamente en la última década, y otros tan veteranos como RTMP, pero con una penetración

El proyecto VAS ha utilizado en gran medida tecnologías ya existentes, de código abierto algunas de ellas, que se han ido perfeccionando colectivamente en la última década

destacable (un 33%, según **Anitua** (n.d.). Sin poder ofrecer una solución cien por cien abierta y gratuita –por la inclusión de la plataforma *Vonage*, principalmente–, el sistema VAS sí que optimiza la conexión audiovisual en directo, con mayor calidad de video y audio y con una latencia inferior a dos segundos, aceptable para una interacción televisiva (**Alsina et al.**, 2024), al máximo de las posibilidades de la tecnología actual.

Cabe destacar que, si bien no es un sistema que funcione autónoma y totalmente por internet, ya que este requiere una mínima infraestructura en plató, VAS sí que aprovecha el potencial de las tecnologías libres de la red. La idea detrás de esta inclusión técnica es obtener el máximo de compatibilidad con redes, dispositivos, aplicaciones y softwares de forma que el sistema sea más accesible para cualquier tipo de usuario o espectador (**Das et al.**, 2021). VAS tiene vocación universal. Esta decisión se suma a la corriente actual de convergencia tecnológica que parece resurgir para facilitar la conexión y actividad sobre todo a los usuarios finales, una idea que ya apuntaba prematuramente **Silverstone** (2013).

La idea es obtener el máximo de compatibilidad con redes, dispositivos, aplicaciones y softwares de forma que el sistema sea accesible para cualquier tipo de usuario

Un aspecto singular del proyecto VAS es el énfasis en recrear una experiencia el máximo de similar a la participación del público desde el plató. Este objetivo se persigue, mayormente, mediante la aplicación de audio inmersivo con la tecnología Ambisonics y el desarrollo e inclusión de un decodificador propio (**Alsina**, 2024). Esta acción apunta la importancia del audio en el procesamiento del mensaje audiovisual ya que la *espacialización* 3D de las fuentes sonoras puede emular un escenario diferente del entorno físico donde se encuentra el participante remoto (**De-Muynke**, 2024).

Igualmente, la utilización de Ambisonics encaja con la idea de una tecnología el máximo de abierta y universal, ya que la codificación de esta es totalmente independiente de cualquier otro sistema de reproducción. Esta independencia entre origen codificado y sistema de recepción hace de Ambisonics, según **Boren** (2017), un formato de producción y distribución muy flexible para audio 3D.

El entorno sonoro de recepción será, por lo tanto, también determinante en la interpretación e incluso en la calidad de la participación interactiva del espectador (**Wenzel et al.**, 2017). La producción del evento, consiguientemente, deberá realizar una labor intensa de comunicación, concienciación y convicción de los participantes en el evento sobre su espacio/entorno de conexión con la transmisión general. Y, de igual manera, la optimización o neutralización del espacio de recepción mediante uso de una tecnología específica a tal efecto puede ser uno de los futuros retos de este tipo de comunicaciones, sus dispositivos, softwares y aplicaciones.

El entorno sonoro de recepción será determinante en la interpretación e incluso en la calidad de la participación interactiva del espectador

De esta manera, aunque autores como **Choueiri** (2017) defienden que la recepción del audio binaural 3D es más natural –más fácilmente relacionada con la distribución de las fuentes sonoras originales– mediante el uso de altavoces (debido a la excesiva proximidad del tímpano con el transductor reproductor, en el caso de auriculares), la fórmula idónea y más habitual parece ser, en cambio, la recepción mediante auriculares. Estos últimos dispositivos permiten el aislamiento sonoro, al menos parcial, y más concentración en la emisión del evento, al mismo tiempo que ofrecen la posibilidad de audiovisionar la transmisión estando en movimiento (**Bull**, 2010; **Detry**, 2016).

Del análisis, se destaca también la excesiva dependencia del resultado audiovisual de la calidad de la conexión doméstica, por wifi o con cable. La calidad de la red doméstica todavía es muy determinante. Este hándicap remarca sin duda la heterogeneidad de los espectadores y participantes remotos y, al mismo tiempo, destaca la complejidad de este sistema, ya que este debe buscar protocolos y elementos universales, o cuanto más compatibles con cualquier dispositivo y software mejor, para funcionar eficientemente. La latencia –esencia del problema comunicativo de este sistema– y también la calidad audiovisual dependen al final en gran medida del ancho de banda y del tipo de conexión de cada usuario, por lo que es difícil configurar y prever un estándar general en esta emisión-recepción. WebRTC y RTMP cumplen de momento este requisito. La evolución de la tecnología reductora de latencia avanza inevitablemente en paralelo y a la par que la tecnología del wifi o de la calidad de conexión (**Joyanes Aguilar**, 2018; **Rebato**, 2022).

Sin embargo, la audiencia no recibe ni debe compartir su emisión, su señal audiovisual, con todos los participantes remotos. Cada participante tan solo recibe la señal única del evento enviado desde plató (**C. Flamarique**, comunicación personal, 18 de abril de 2024). Es el plató quien recoge todas las señales agrupadas y realizadas previamente en la aplicación web y, posteriormente, realiza de nuevo esta señal para integrarla a conveniencia en la presentación del evento (**Alsina et al.**, 2024). Se podría hablar de una doble realización, en referencia al contenido interactivo del participante remoto. Por lo tanto, la calidad de la conexión debe asegurar una

La evolución de la tecnología reductora de latencia avanza inevitablemente en paralelo y a la par que la tecnología del wifi o de la calidad de conexión

latencia y una calidad audiovisual sobre todo en el envío de la intervención interactiva en remoto hacia plató (**Arjona-Martín; Victoria-Mas**, 2018). Aunque, por supuesto, también es relevante un retorno-respuesta ágil del plató al participante remoto, en este itinerario la calidad audiovisual puede ser más secundaria siempre y cuando garantice como mínimo una comprensión completa del mensaje –en el sentido de que mientras se entienda el contenido y sea a tiempo razonable para generar una respuesta, si la imagen es más o menos pixelada o precisa, es secundario, ya que la señal la recibe y ve sólo el participante conectado (**Izquierdo-Castillo; Latorre-Lázaro**, 2022).

La priorización de la diversidad de dispositivos en la recepción, la relevancia del entorno del espectador, la incidencia de la calidad técnica de su conexión y la búsqueda de protocolos y aplicaciones de uso universal y abierto encaja –todos estos aspectos– con la tendencia actual de consumo televisivo anticipada por **Cebrián Herreros** (2004). Esta tendencia de recepción mediática, aparte de salir del flujo televisivo (**Majó**, 2009) mediante la elección de un horario de consumo personal para cada usuario (característica esta que no encaja en la tipología de programas en directo que analizamos), se distingue por la atención mediática en espacios personales (domésticos, laborales...), a través de dispositivos propios y personalizados (*smartphones*, tabletas, *laptops*...), con dispositivos multifunciones y multimedia (**Majó**, 2009; **Andersson Schwarz**, 2016), y con la posibilidad de recepción *in itinere*, es decir, en movimiento (**Joyanes Aguilar**, 2018).

En este sentido, el sistema VAS se adapta perfectamente a la filosofía y requerimientos del paradigma de consumo televisivo actual –diverso, personal y móvil, se podría resumir de acuerdo con **Joyanes Aguilar** (2018)– excepto por la citada importancia del entorno para el procesamiento sonoro y por la hora síncrona, única, de emisión y participación (**Izquierdo-Castillo; Latorre-Lázaro**, 2022). Sin embargo, se podría considerar que estos últimos aspectos son más propios del formato del contenido (el *talent show*) y de las exigencias sociales del medio, que no tanto del sistema técnico VAS en sí; aunque el mismo sistema sí que podría plantearse el reto de solucionar en un futuro alguna de estas disfunciones.

Anitua (n.d.) comenta que “la transmisión en vivo tendrá un valor de más de 180 mil millones de dólares en 2027”. La aplicación de este sistema de información, al mismo tiempo que cambia y facilita la forma de realización de los *shows* y espectáculos emitidos en directo, es entonces del todo pertinente. La dependencia de infraestructura, hardware y equipos físicos se reduce considerablemente y el sistema permite más flexibilidad y adaptación según las dimensiones del evento y la tipología de receptores, incluso según el formato o género del programa o contenido emitido. El sistema VAS facilita enormemente y puede normalizar la interacción audiovisual síncrona en remoto.

En este sentido, podría considerarse que el sistema VAS formaliza e impulsa una nueva categoría conceptual de audiencia o público: la audiencia virtual. Este tipo de espectador se sitúa aislado en remoto en su espacio personal y con sus dispositivos de comunicación, pero, no obstante, ejerce y disfruta de una participación interactiva equivalente o casi igual a la que muestra el público televisivo tradicional que asiste presencialmente al plató para seguir la realización del espectáculo en directo. Si bien la transformación de la audiencia televisiva se basa en la interacción a distancia (**Scolari**, 2008; **van Es**, 2016; **Wohn; Na**, 2011), la diferencia de esta nueva audiencia es una interacción audiovisual en tiempo real que al mismo tiempo forma parte audiovisualmente –con imagen en movimiento y sonido– del evento transmitido. Este modelo podría tener alguna similitud e interrelaciones con la plataforma *Twitch* (**Lozano; Naranjo**, 2020). Por otro lado, la audiencia del sistema VAS, técnicamente, de momento contiene el límite de incluir como máximo unos 50 o 55 participantes remotos (**Alsina et al.**, 2024), mientras que en plató se han llegado a albergar hasta 500 personas (**Gestmusic**, 2021).

Por lo tanto, la audiencia virtual generada por el sistema VAS es diferente a una conexión por redes sociales o internet –por su agrupación, límite de participantes e interacción audiovisual en baja latencia– y, al mismo tiempo, es más que la asistencia de público a plató –en cuanto a comodidad, intimidad y personalización de la comunicación. Es una nueva categoría que merece ser teorizada en futuras investigaciones para indagar, al mismo tiempo, si esta no arrastra también con ella un nuevo concepto de televisión (**Scolari**, 2008).

La diferencia de la nueva audiencia generada por VAS es una interacción audiovisual en tiempo real que al mismo tiempo forma parte audiovisualmente del evento transmitido

Los avances que aporta el proyecto VAS podrían representar una mejora sustancial de la producción, emisión y recepción del entretenimiento en directo respecto los sistema y usos tecnológicos actuales, con clara aplicación en la industria televisiva y musical (**Gestmusic**, 2022). Para las empresas productoras, el VAS podría representar una liberación del hardware y los equipos físicos –a parte de los estrictamente necesarios para la realización– y una mayor adaptabilidad en la producción segura de eventos con participación interactiva de público virtual. Al mismo tiempo, para los espectadores la novedad sería la incursión en una experiencia inmersiva muy similar a un entorno real con mayor calidad de audio y video (**Alsina**, 2024).

En general pues, se puede afirmar que el sistema de información audiovisual VAS, generador de la idea de una audiencia virtual síncrona efectiva, es novedoso, beneficioso para la producción y la calidad audiovisual de los espectadores (participantes o no), y aporta una solución asequible con la tecnología actual al problema de la latencia en la interacción televisiva en vivo y en espectáculos en directo, en general. La televisión, con VAS, se expande más por las redes y viceversa (**Scolari**, 2008), gestando así un nuevo espacio y ente comunicativo.

6. Conclusiones

El proyecto VAS se plantea para aportar una solución específica al impedimento de la presencia de público en plató, en un formato en el cual el público es indispensable. Desde un principio, se busca una solución basada en tecnología existente, abierta o asequible, y el máximo de compatible con cualquier otra tecnología de recepción.

Los problemas iniciales para superar son, principalmente, la alta latencia en la transmisión y recepción de audio y vídeo y, al mismo tiempo, una baja calidad de estos. Manteniendo una calidad audiovisual aceptable para una emisión televisiva, el proyecto VAS consigue una latencia por debajo de dos segundos en el circuito completo de interacción con los participantes virtuales remotos. Se constata que actualmente, con la tecnología existente, es difícil superar este hito.

Como solución, se planifica el uso de RTMP, WebRTC y SRT, así como de Ambisonics y otros softwares y protocolos accesibles. Estos formaran parte de la solución final, añadiendo también una suscripción con cuota económica a la plataforma *Vonage*, la cual actúa como receptora y editora de todos los perfiles de participantes remotos. Se descartan otros formatos como el LL-HLS, MPEG-DASH o dispositivos como NDI por sus limitaciones de universalidad o bien por no contribuir suficientemente a la reducción de la latencia.

Aunque la solución no es ni cien por cien en la nube (se necesita un equipo mínimo en plató y, al mismo tiempo, servidores remotos), ni cien por cien gratuita, sí que consigue el máximo de compatibilidad con cualquier tipo de dispositivo receptor, software o protocolo.

La solución final permite mucha más flexibilidad, escalabilidad y facilidades en la producción televisiva de un evento que requiera intervenciones interactivas de participantes virtuales. No obstante, el número posible de participantes virtuales se reduce considerablemente –hasta 10 veces– respecto las posibilidades de presencia en plató. Aun así, VAS representa el primer proyecto de audiencia virtual efectivo y comercializable.

Este nuevo tipo de audiencia presenta todavía una gran dependencia de la calidad de la banda de conexión y wifi propio de cada participante remoto. En este aspecto determinante se observa claramente la complejidad que representa la amplia diversidad tecnológica aún existente, así como la heterogeneidad social y económica sobre todo de los usuarios finales. Estas variables más incontrolables pueden ser un obstáculo para la producción y esta debe minimizarlas en la selección de participantes.

Con todo, el proyecto VAS de audiencia virtual representa un gran salto cualitativo en la flexibilidad y calidad de la producción y emisión de programas con público, especialmente del formato *talent show*. La tecnología VAS es aplicable también a otros eventos mediáticos y culturales, como conciertos, por ejemplo.

La posibilidad de disponer de un conjunto de participantes virtuales remotos que interactúe con una muy baja latencia, con una sensación de participación casi en tiempo real, genera una nueva categoría conceptual de audiencia –la cual no es ni virtual, ni presencial, estrictamente– y, por extensión, un nuevo concepto de televisión. Estas nuevas tipologías de participación mediática merecen un análisis más profundo en futuras investigaciones.

7. Financiación

El proyecto VAS, impulsado por el departamento de desarrollo e investigación de la empresa *Gestmusic* (Banijay Ibérica), recibió una ayuda de los fondos europeos *Next Generation* distribuidos por el Ministerio de Asuntos Económicos y Transformación Digital en la convocatoria competitiva de 2021. Como parte del plan de acción y difusión de este proyecto subvencionado se incluye el asesoramiento del grupo de investigación *So, Silenci, Imatge i Tecnologia (SSIT)* para la elaboración de este artículo científico.

Referencias

- Alsina, Lluís.** (2024). *Memoria final del proyecto VAS* [Manuscrito en preparación]. Gestmusic - Banijay Ibérica.
- Alsina, Lluís; Flamarique, Carles; Flamarique, Germán.** (2024). *Pruebas técnicas VAS* [Manuscrito en preparación]. Gestmusic - Banijay Ibérica.
- Andersson Schwarz, Jonas.** (2016). "Public Service Broadcasting and Data-Driven Personalization: A View from Sweden". *Television & New Media*, v. 17, n. 2, pp. 124-141. <https://doi.org/10.1177/1527476415616193>
- Anitua, Asier.** (n.d.). "Transmisión de vídeo y audio: protocolos de streaming". Telefónica. Servicios Audiovisuales. <https://www.telefonicaserviciosaudiovisuales.com/articulos-de-divulgacion/protocolos-de-streaming>
- Arjona-Martín, José-Borja; Victoria-Mas, Juan-Salvador.** (2018). "Web-based video services. A quantitative and descriptive study". *Profesional de la información*, v. 27, n. 1, pp. 75-85. <https://doi.org/10.3145/epi.2018.ene.07>
- Blum, Niklas; Lachapelle, Serge; Alvestrand, Harald.** (2021). "WebRTC: Real-Time Communication for the Open Web Platform". *Communications of the ACM*, v. 64, n. 8, pp. 50-54. <https://doi.org/10.1145/3453182>
- Boren, Braxton.** (2017). "History of 3D Sound". En: *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Roginska, Agnieszka; Geluso, Paul (Eds.), pp. 40-62. Routledge. <https://doi.org/10.4324/9781315707525-3>
- Bull, Michael.** (2010). "iPod: un mundo sonoro personalizado para sus consumidores". *Comunicar*, v. 34, n. XVII, pp. 55-63. <https://doi.org/10.3916/C34-2010-02-05>

- Carrillo, Jesús.** (n.d.). "SRT, el estándar de facto en streaming". Telefónica Servicios Audiovisuales. <https://www.telefonicaserviciosaudiovisuales.com/articulos-de-divulgacion/srt-el-estandar-de-facto-en-streaming>
- Cebrián Herreros, Mariano.** (2004). *Modelos de televisión: generalista, temática y convergente con Internet*. Barcelona: Paidós. <https://www.uoc.edu/uocpapers/dt/esp/cebrian.html>
- Choueiri, Edgar.** (2017). "Binaural Audio Through Loudspeakers". En: *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Roginska, Agnieszka; Geluso, Paul (Eds.), pp. 124-179. Routledge. <https://doi.org/10.4324/9781315707525-6>
- Das, Souvik; Sarkar, Partha; Hussain, Alfarid.** (2021). "Perception Analysis of TV Reality Shows: Perspective of Viewers' and Entertainment Industry Professionals". *International Journal of Media, Journalism and Mass Communications (IJMJMCM)*, v. 7, n. 2, pp. 22-31. <https://doi.org/10.20431/2454-9479.0702003>
- De-Muynke, Julien.** (2024). "Virtual Reconstruction and Auralisation of Historically-informed Soundscapes: Application to Audio-Augmented Visits of Heritage Sites." Doctoral thesis not published, Sorbonne Université
- Detry, Lionel.** (2016). "Mobile Music Listening: The Users and their Art of Making Do Lionel Detry". *Networking Knowledge: Journal of the MeCCSA Postgraduate Network*, v. 9, n. 2, pp. 1-13. <https://doi.org/10.31165/nk.2016.92.429>
- Gestmusic.** (2021). *Proyecto VAS. Memoria del proyecto. Convocatoria de ayudas 2021 destinadas a proyectos de investigación y desarrollo en inteligencia artificial y otras tecnologías digitales y su integración en las cadenas de valor* [Proyecto núm. C005/21-ED]. [Manuscrito no publicado].
- Gestmusic.** (2022). "Gestmusic recibe una subvención de los fondos europeos Next Generation para desarrollar su sistema de audiencia virtual VAS". <https://www.gestmusic.es/gestmusic-recibe-subvencion-fondos-europeos-next-generation-para-desarrollar-sistema-audiencia-virtual-vas>
- Heller, Aaron J; Benjamin, Eric M.** (2018). "Design and Implementation of Filters for Ambisonic Decoders". En: *Proceedings of the 1st International Faust Conference (IFC-18)*. <https://hal.science/hal-03162951>
- Izquierdo-Castillo, Jessica; Latorre-Lázaro, Teresa.** (2022). "Oferta de contenidos de las plataformas audiovisuales. Hacia una necesaria conceptualización de la programación streaming". *Profesional de la información*, v. 31, n. 2, pp. e310218. <https://doi.org/10.3145/epi.2022.mar.18>
- Joyanes Aguilar, Luis.** (2018). "Computación en la nube: Notas para una estrategia española en cloud computing". *Revista del Instituto Español de Estudios Estratégicos*, pp. 89-112. <https://revista.ieee.es/article/view/406>
- Lee, Jeong-Hwan; Cha, Jihun.** (2023). "Optimization of Video Streaming Using SRT Protocol in Mobile Communication Network". En: *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*. pp. 1802-1804. IEEE. <https://doi.org/10.1109/ICTC58733.2023.10392477>
- Loreto, Salvatore; Romano, Simon Pietro.** (2012). "Real-Time Communications in the Web: Issues, Achievements, and Ongoing Standardization Efforts". *IEEE Internet Computing*, v. 16, n. 5, pp. 68-73. <https://doi.org/10.1109/MIC.2012.115>
- Lozano, Juan Francisco Gutiérrez; Naranjo, Antonio Cuartero.** (2020). "El auge de Twitch: nuevas ofertas audiovisuales y cambios del consumo televisivo entre la audiencia juvenil". *Ámbitos. Revista Internacional de Comunicación*, n. 50, pp. 159-175. <https://doi.org/10.12795/Ambitos.2020.i50.11>
- Majó, Joan.** (2009). "Tendencias de futuro en el audiovisual". *Quaderns del CAC. Convergencia tecnológica y audiovisual*, n. 31-32, pp. 27-32. https://www.cac.cat/sites/default/files/migrate/quaderns_cac/Q31-32_ES.pdf
- Mela, José-Luís; Cedeño, Gloris-Denisse; Cedeño-Herrera, Edwin.** (2021). "Edge Computing: Aplicaciones y desafíos actuales". *Visión Antataura*, v. 5, n. 1, pp. 75-91. <https://revistas.up.ac.pa/index.php/antataura/article/view/2199>
- Mell, Peter; Grance, Tim.** (2011). *The NIST Definition of Cloud Computing (Publication No. NIST SP 800-145)*. National Institute of Standards and Technology - Computer Security Resource Center. <https://doi.org/10.6028/NIST.SP.800-145>
- Narbutt, Mirosław; Skoglund, Jan; Allen, Andrew; Chinen, Michael; Barry, Dan; Hines, Andrew.** (2020). "AMBIQUAL: Towards a Quality Metric for Headphone Rendered Compressed Ambisonic Spatial Audio". *Applied Sciences*, v. 10, n. 9, pp. 3188. <https://doi.org/10.3390/app10093188>
- Nigro, Hector.** (2022). "Cloud computing: retos y oportunidades". *Revista Ingeniería, Matemáticas y Ciencias de la Información*, v. 9, n. 18, pp. 11-16. <https://doi.org/10.21017/rimci.2022.v9.n18.a116>
- Nurrohman, Anif; Abdurrohman, Maman.** (2018). "High Performance Streaming Based on H264 and Real Time Messaging Protocol (RTMP)". En: *2018 6th International Conference on Information and Communication Technology (ICoICT)*. pp. 174-177. IEEE. <https://doi.org/10.1109/ICoICT.2018.8528770>
- Rebato, Carlos.** (2022). "Qué es Edge Computing, explicado de manera sencilla". Telefónica Tech. <https://telefonicatech.com/blog/edge-computing-que-es>
- Scolari, Carlos Alberto.** (2008). "Hacia la hipertelevisión: los primeros síntomas de una nueva configuración del dispositivo televisivo". *Diálogos de la comunicación*, v. 77, pp. 1-9. <http://hdl.handle.net/10230/25464>
- Silverstone, Roger.** (2013). *Televisión y vida cotidiana*. Madrid: Amorrortu. <https://www.amorrortueditores.com/Papel/9789505186457/Televisi%C3%B3n+y+vida+cotidiana>
- Templier, Mathieu; Paré, Guy.** (2015). "A Framework for Guiding and Evaluating Literature Reviews". *Communications of the Association for Information Systems*, v. 37, n. 1, pp. 6. <https://doi.org/10.17705/1CAIS.03706>
- van Es, Karin.** (2016). "Social TV and the Participation Dilemma in NBC's The Voice". *Television & New Media*, v. 17, n. 2, pp. 108-123. <https://doi.org/10.1177/1527476415616191>

- Wenzel, Elizabeth M; Begault, Durand R; Godfroy-Cooper, Martine.** (2017). "Perception of Spatial Sound". En: *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Roginska, A.; Geluso, P. (Eds.), pp. 5-39. New York: Routledge. <https://doi.org/10.4324/9781315707525-2>
- Wohn, D Yvette; Na, Eun-Kyung.** (2011). "Tweeting about TV: Sharing television viewing experiences via social media message streams". *First Monday*, v. 16, n. 3, <https://doi.org/10.5210/fm.v16i3.3368>