

Impacto de Parámetros de QoS en Aspectos de QoE: Análisis desde el Punto de Vista de la Sincronización Multimedia

Mario Montagud^{1,2}, Fernando Boronat², Marc Martínez², Jordi Belda², Pablo Cesar¹

¹Centrum Wiskunde & Informatica (CWI), Amsterdam

²Universitat Politècnica de València (UPV) – Campus de Gandia

montagud@cwi.nl; {fboronat@dcom., marmarc3@upvnet., jorbelva@epsg.}.upv.es; P.S.Cesar@cwi.nl

Resumen- La sincronización multimedia ha sido un área de investigación clave desde los inicios de los sistemas multimedia. En este artículo se ofrecen una visión general y un análisis sobre el impacto de varios parámetros de QoS en diferentes aspectos de la QoE, desde el punto de vista de la sincronización multimedia. En primer lugar, se presentan los diferentes tipos de sincronización multimedia y su relevancia para garantizar una QoE satisfactoria. En segundo lugar, se muestra que la magnitud de los retardos y su variabilidad en las redes actuales es bastante superior a los umbrales permisibles por los usuarios en diferentes tipos y ejemplos de sincronización multimedia, reflejando así su necesidad. En tercer lugar, se describe el impacto del ancho de banda y la tasa de pérdidas sobre la sincronización multimedia. Por último, se argumenta la influencia del uso de diferentes alternativas para conseguir la sincronización multimedia sobre varios factores de QoS y de QoE.

Palabras Clave- IDMS, QoS, QoE, Sincronización Híbrida, Sincronización Multimedia

I. INTRODUCCIÓN

Las redes actuales de conmutación de paquetes no ofrecen garantías ni una respuesta estable y óptima para la distribución de contenidos multimedia. Esto se convierte en un desafío, más aún cuando existen requisitos estrictos en cuanto a retardo, interactividad, tasa de pérdidas, ancho de banda y/o escalabilidad. En los últimos años se ha estado investigando con el objetivo de mejorar las prestaciones tanto de las tecnologías de red como de los sistemas finales y de interconexión, para poder ofrecer servicios multimedia sobre redes IP de manera satisfactoria. Algunos de los avances más relevantes que se pueden destacar son: redes de mayor alcance y capacidad, mayores velocidades y capacidades de computación, códecs más avanzados y eficientes, nuevas tecnologías de transmisión adaptativas (p.ej., Hypertext Transfer Protocol o HTTP streaming, Datagram Congestion Control Protocol o DCCP...), políticas de admisión, reserva de recursos y asignación de prioridades, mecanismos de encaminamiento avanzados, sistemas de replicación de contenido y caching, etc.

En las comunicaciones telemáticas y, especialmente, en los servicios multimedia interactivos, los parámetros de calidad de servicio (Quality of Service o QoS), como pueden ser la disponibilidad, el retardo y su variabilidad (ya sea para un mismo flujo de datos, es decir el jitter, como entre flujos y entre receptores), la tasa de pérdidas y el ancho de banda disponible, tienen un impacto muy claro sobre varios aspectos de la calidad de experiencia (Quality of Experience

o QoE) percibida por los usuarios, como son la satisfacción, la perceptibilidad, la molestia, la inmervisidad, el *engagement*, la conectividad virtual (*networked togetherness*), la igualdad de condiciones (*fairness*), etc. Así pues, es necesario optimizar los aspectos tecnológicos, estrechamente relacionados con la QoS, así como los aspectos operacionales, con tal de brindar a los usuarios la mejor QoE posible, pues ello va a determinar el éxito de los servicios multimedia existentes y de los que se puedan desarrollar en el futuro.

En este artículo se ofrecen una visión general y un análisis sobre la influencia de varios parámetros de QoS, centrándose especialmente en los retardos y su variabilidad, en diversos aspectos de QoE, desde el punto de vista de la sincronización multimedia. Se van a presentar los diferentes tipos de sincronización multimedia, destacando algunos de los casos de uso más relevantes, y se van a identificar los desafíos tecnológicos, ligados a parámetros de QoS, que pueden frustrar la QoE percibida por los usuarios. En particular, se van a identificar umbrales tolerables para diferentes tipos y casos de uso de sincronización multimedia, así como también se van a describir algunas implicaciones con respecto al uso de varias alternativas para conseguir la sincronización multimedia. Asimismo, a lo largo del artículo se va a identificar la necesidad de seguir investigando en ciertos aspectos, tanto para determinar los umbrales permisibles como para diseñar soluciones que mejoren la QoE.

Este artículo se ha estructurado de la siguiente manera. En la Sección II se presentan los diferentes tipos de sincronización multimedia, en sus diferentes variantes, incluyendo algunos ejemplos relevantes. En la Sección III se enumeran los diferentes factores que pueden contribuir al incremento de los retardos y su variabilidad en las comunicaciones multimedia. Asimismo, se recogen resultados de medidas de retardos en diferentes escenarios reales. En la Sección IV se presentan los umbrales tolerables por la percepción humana en cuanto a errores (esto es, asincronías) en diferentes tipos y ejemplos de sincronización multimedia. En la Sección V se presenta brevemente la influencia del esquema de control para la sincronización utilizado en varios factores de QoS y aspectos de QoE. En la Sección VI se discute el impacto del uso de varias estrategias para la selección de un flujo o componente multimedia como la referencia temporal para la sincronización. El artículo finaliza con las conclusiones en la Sección VII.

II. TIPOS DE SINCRONIZACIÓN MULTIMEDIA

Se pueden distinguir varios tipos y variantes de sincronización multimedia (véase la Fig. 1), en función del número de componentes multimedia, de emisores, flujos y receptores involucrados. Primero, la sincronización intra-media (también conocida como sincronización intra-flujo) es necesaria para mantener las relaciones temporales originales entre las unidades de datos, Media Units o MUs (p.ej., tramas de video, muestras de voz...), que forman parte de cada componente multimedia (p.ej., video, audio...). Estas relaciones temporales se suelen especificar en los procesos de captura o emisión y pueden verse alteradas durante la cadena de distribución multimedia extremo-a-extremo, debido principalmente al jitter y las pérdidas. Es por ello que se necesitan mecanismos precisos y adaptativos para mantenerlas (y restablecerlas, si es necesario) de manera que el contenido multimedia se reproduzca de acuerdo a los patrones temporales apropiados. Típicamente, se utilizan buffers de reproducción con tal de compensar el efecto del jitter y de las pérdidas. Si no se incluyen mecanismos de sincronización intra-media, existirán distorsiones temporales en la reproducción individual de cada componente multimedia (p.ej., chasquidos en el audio, pausas y acelerones en el video...), y esto puede llegar a ser muy molesto para los usuarios (es decir, se corresponderá con una mala QoE).

Segundo, la sincronización inter-media (también conocida como sincronización inter-flujo) se ocupa de preservar las dependencias temporales entre las MUs de diferentes componentes multimedia (p.ej., la sincronización labial o *lip-sync*). La falta de sincronización inter-media puede causar confusión a los usuarios (p.ej., comentarios de audio o subtítulos que no se corresponden con las escenas que se visualizan), por lo que puede llegar a ser irritante (mala QoE). Se pueden distinguir dos situaciones diferentes en cuanto a la sincronización inter-media. La primera de ellas es cuando los diferentes componentes multimedia se multiplexan en un único flujo de datos, tal y como se puede observar en la Fig.2.a.1 (en este caso los términos sincronización inter-media e inter-flujo no se pueden utilizar como sinónimos, al existir un único flujo). La segunda de ellas es cuando cada componente multimedia se envía en un flujo de datos independiente, tal y como se puede observar en la Fig.2.a.2. En el primer caso, conseguir la sincronización inter-media es más fácil, ya que los componentes multimedia están contenidos en el mismo flujo y, por tanto, han sufrido prácticamente el mismo retardo, así como también es más fácil identificarlos y enlazarlos. Sin embargo, el segundo caso otorga más flexibilidad y eficiencia con respecto al uso de ancho de banda. Por ejemplo, un usuario podría recibir sólo los componentes multimedia que le interesan, o cesar temporalmente la recepción de algún componente multimedia (considerado menos prioritario) en situaciones de congestión. La primera alternativa se suele utilizar en tecnologías MPEG (p.ej., Moving Picture Experts Group 2 Transport Stream o MPEG2-TS, Dynamic Adaptive Streaming over HTTP o DASH...), mientras que la segunda es más común cuando se utilizan los protocolos Real-Time Transport Protocol / RTP Control Protocol (RTP/RTCP) [1] para la entrega de contenidos multimedia.

Un sub-tipo específico de la sincronización inter-media es la sincronización inter-emisor (también conocida como sincronización multi-fuente), y se ocupa de sincronizar la reproducción de varios componentes multimedia enviados por diferentes emisores o fuentes (véase la Fig. 2.b). También puede suceder que los diferentes componentes multimedia se envíen mediante diferentes protocolos o incluso a través de diferentes redes, de la misma o diferente naturaleza. En este último caso, se suele llamar sincronización híbrida (*hybrid synchronization*) [2]. Un ejemplo muy relevante en el mercado de TV actual es la sincronización entre flujos broadcast y broadband, como pueden ser flujos DVB (Digital Video Broadcasting) y flujos obtenidos a través de HTTP streaming (p.ej., DASH, HTTP Live Streaming o HLS...). Esta es una de las posibilidades que puede ofrecer el estándar HbbTV (Hybrid Broadcast Broadband TV) [3]. A través de un uso simultáneo y coordinado de las redes broadband y broadcast, la sincronización híbrida permite ofrecer servicios multimedia enriquecidos y personalizados a los usuarios, y por tanto una mejor QoE. El contenido ofrecido por la red broadcast se puede complementar con contenido adicional relacionado ofrecido por redes broadband. Escenarios típicos son la selección de videos multi-cámara, la selección del canal de audio o emisora de radio favorito/a, el acceso a contenido web complementario, todo ello a través de redes broadband. Esto permite una mayor flexibilidad, un mayor abanico de posibilidades y un ahorro significativo de costes, debido al alto precio del ancho de banda broadcast (licencias para el uso del espectro).

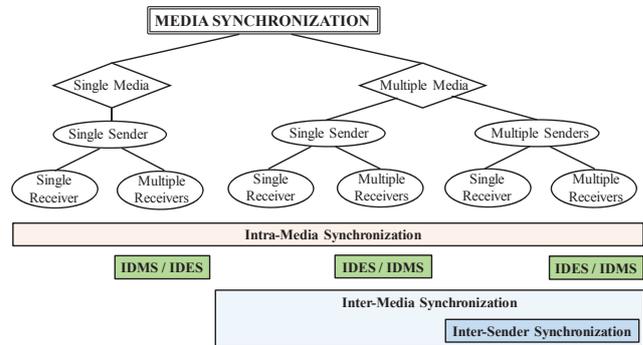


Fig. 1. Visión general del área de investigación relativa a la sincronización multimedia.

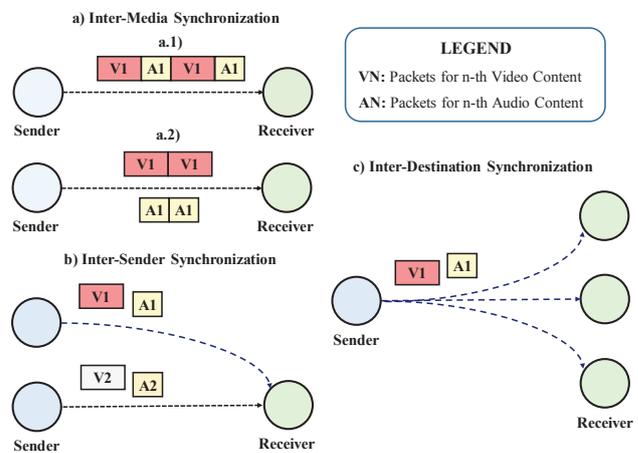


Fig. 2. Diferentes tipos y variantes de sincronización multimedia.

Además de la sincronización de componentes multimedia en cada receptor específico (p.ej., TV, PC...), también es necesaria la sincronización simultánea de los procesos de reproducción de componentes multimedia en diferentes receptores o dispositivos (véase la Fig.2.c). Esto se conoce como sincronización inter-dispositivo (Inter Device Synchronization o IDES), y también como sincronización inter-destinatario (Inter-Destination Synchronization o IDMS), especialmente cuando los dispositivos están separados geográficamente (p.ej., en Social TV, servicios de multi-conferencia, juegos en red multi-jugador...) [4]. La sincronización IDMS es muy relevante para posibilitar una gran variedad de experiencias multimedia compartidas entre usuarios remotos (hasta 20 casos de uso se describen en [5]). Su inexistencia puede afectar negativamente a la QoE, por ejemplo, causando interacciones incoherentes o malentendidos en las comunicaciones en grupo, desigualdad de condiciones en escenarios de ámbito competitivo (p.ej., juegos en red), lo que puede ser muy desesperante para los usuarios. El ejemplo típico es el de un usuario viendo un partido de fútbol online con sus amigos de forma remota, y percibir un gol a través de los canales de chat, o de los gritos de sus vecinos, antes de ver el gol en su propia pantalla.

III. VARIABILIDAD DE RETARDOS EN LAS COMUNICACIONES

Se ha comentado previamente que los retardos y su variabilidad, ya sea para un mismo flujo (jitter), entre diferentes flujos o entre diferentes receptores, son los factores más determinantes por los que se van a necesitar soluciones de sincronización multimedia. En primer lugar, en esta sección se presentan los diferentes factores a lo largo de la cadena de distribución multimedia que contribuyen al incremento de la magnitud de los retardos y de su variabilidad. En segundo lugar, se proporcionan datos de medidas realizadas en estudios de investigación recientes que reflejan la magnitud de los retardos extremo-a-extremo y de su variabilidad en redes de distribución actuales.

Los retardos y su variabilidad no son un problema serio cuando se consume contenido multimedia almacenado de manera aislada, pero sí se convierten en una barrera cuando se consume contenido multimedia en vivo y/o diferentes usuarios están interactuando de manera remota en el contexto de un consumo conjunto de contenido multimedia.

A. Factores que Contribuyen a los Retardos y a su Variabilidad

Cuando se distribuye contenido multimedia por Internet, cada uno de los componentes multimedia involucrados (p.ej., audio, video...) se estructura, durante el proceso de captura o lectura, en una secuencia de MUs (p.ej., tramas de video o muestras de voz) que, a su vez, se encapsularán en uno o varios paquetes de datos para su transmisión por las redes IP. Cada uno de estos paquetes puede seguir la misma o diferente ruta que los demás paquetes del mismo o de diferentes componentes multimedia, así como también puede verse afectado de manera diferente por el jitter.

Tal y como se puede apreciar en la Fig. 3, a lo largo de la cadena de distribución multimedia extremo-a-extremo varios factores pueden contribuir al incremento de la magnitud de los retardos y de su variabilidad [4,6], pudiendo ambos oscilar desde unos pocos milisegundos hasta varios segundos.

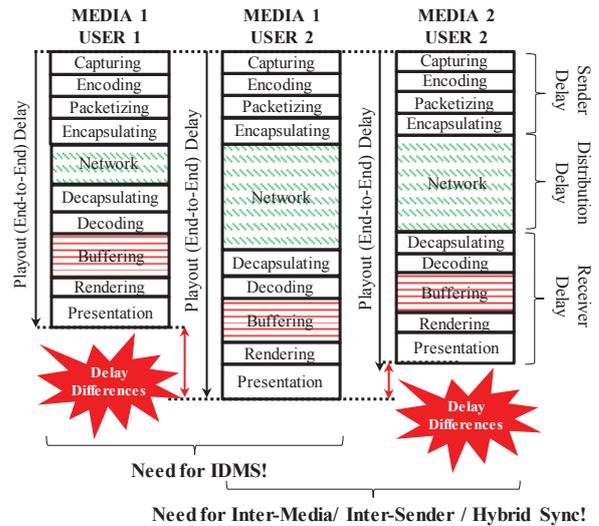


Fig. 3. Variabilidad de retardos: Necesidad de sincronización multimedia.

En el lado del emisor, procesos como la captura, codificación, cifrado, buffering, paquetización y encapsulamiento, entre otros, pueden contribuir al incremento de los retardos y de su variabilidad. Por ejemplo, el estudio en [7] demuestra que los tiempos de captura y codificación pueden diferir significativamente, en el orden de decenas de milisegundos, para diferentes flujos de audio y video, así como para diferentes procesadores, dependiendo también de su carga puntual de proceso.

En la red de distribución, varios factores pueden afectar al incremento de los retardos, así como de su variabilidad, como son los procesos de propagación y transmisión de paquetes, tanto a través de los enlaces de comunicaciones como de los dispositivos intermedios y finales involucrados, los procesos de buffering y control en los routers intermedios (p.ej., decisiones de encaminamiento, políticas de QoS...), así como otros procesos avanzados como son la fragmentación y reensamblado de paquetes, trans-codificación, conversión de formatos, etc. Además, la carga instantánea de la red va a influir significativamente en la magnitud de estos retardos y de su variabilidad, tanto para un mismo flujo multimedia (es decir, el jitter), como entre diferentes flujos multimedia recibidos por el mismo o diferentes dispositivos.

En el lado del receptor, varios factores también contribuirán al incremento de los retardos y de su variabilidad, como son los procesos de buffering, descifrado, decodificación, de-paquetización o procesado en los diferentes niveles de la pila de protocolos, técnicas de detección y corrección de errores, retardos de entrega (*rendering delay*), etc. Quizás, de entre todos ellos, los retardos de buffering y decodificación sean los más relevantes. Estos retardos pueden fluctuar con respecto al tiempo, así como también pueden diferir entre diferentes receptores, debido a las diferentes capacidades de procesado, a la carga instantánea de CPU (Central Processing Unit), a la respuesta en tiempo real del sistema operativo en cuestión, al software y hardware disponible, así como a la carga instantánea de la red (pues determinará la tasa a la que se recibe el contenido multimedia).

Otro factor adicional a tener en cuenta cuando se utilizan TV digitales es el retardo de display (*display lag*), que es la diferencia entre el instante en el que una imagen (procedente

de una señal de video) se entrega a un display y el instante en el que dicha imagen se visualiza. Este retardo se origina debido principalmente a rutinas de procesamiento de la imagen, como es el escalado. El *display lag* puede originar una asincronía entre la señal de video y la de audio, afectando, por tanto, a la sincronización inter-media (si no se compensa de algún modo). El *display lag* en TVs de alta definición (High Definition or HD TV) puede oscilar entre 30 y 90ms, dependiendo del modelo y del tipo de señal de video de entrada [8]. En este contexto, el estudio en [9] analizó el impacto de diferentes modos de TV (concretamente, el modo "Dinámico" y el modo "Juego") sobre el retardo de video y su variabilidad. Para la TV bajo test, se observó que el retardo en modo "Juego" fue casi 90ms menor y un poco menos variable (a costa de una ligera degradación de calidad, según la documentación de la TV, aunque fue inapreciable para los autores de dicho estudio). En concreto, en el escenario bajo test, para el modo "Juego", el retardo extremo-a-extremo medio fue de 357ms (con una desviación estándar de 20ms), mientras que para el modo "Dinámico", fue de 442ms (con una desviación estándar de 36ms). Por tanto, las características y configuración de los dispositivos finales pueden tener un impacto significativo sobre el retardo y su variabilidad.

Otro factor muy importante que puede contribuir a las diferencias de retardos, tanto para un mismo flujo como entre varios flujos, y tanto para un mismo receptor como entre diferentes receptores, son las imperfecciones de los relojes, pues van a acarrear desviaciones (lineales y/o aleatorias) en las tasas de generación, transmisión y reproducción de los contenidos multimedia.

Un análisis más exhaustivo sobre la contribución e impacto de cada uno de los factores mencionados en esta sección sobre la magnitud de los retardos y de su variabilidad se puede consultar en [6].

B. Magnitud de los Retardos en Redes Actuales

Después de presentar los diferentes factores que contribuyen a los retardos y a su variabilidad, esta sección proporciona datos sobre su magnitud en escenarios reales, tanto para un mismo flujo, como para diferentes flujos enviados a un mismo o diferentes receptores, de acuerdo a medidas realizadas en estudios relacionados.

En primer lugar, el estándar International Telecommunication Union (ITU) G.1050 [10] indica que los valores de retardos típicos en Internet pueden oscilar entre 20 y 500ms, mientras que los valores de jitter pueden oscilar entre 0 y 500ms. Por otro lado, el estándar ITU-T (ITU Telecommunication Standardization Sector) G.114 [11] indica que se requieren retardos inferiores a 150ms para aplicaciones de conferencia sobre Internet, mientras que retardos superiores a 400ms son típicamente inaceptables.

En [12], se indica que los retardos extremo-a-extremo cuando se utilizan herramientas de videoconferencia convencionales en un entorno LAN (Local Area Network) pueden oscilar entre 99ms, para Google Talk (Gtalk), y 312ms (con una desviación estándar de 67ms), para Skype. Asimismo, retardos extremo-a-extremo en torno a 350ms (también con una desviación estándar de 67ms) se obtuvieron cuando se utilizó una herramienta de videoconferencia ad-hoc en un escenario de red distribuido entre Bélgica y el Reino Unido.

En [13], se realizaron medidas de retardos en diferentes escenarios de red. A partir de estas medidas, se hizo un análisis de los mejores y peores casos para diferentes factores de retardo a lo largo de la cadena de distribución multimedia, cuyo resumen se puede consultar en la Tabla I. Se puede apreciar que los procesos de codificación, transmisión y buffering son los que más retardos introducen. Además, se comprobó en este estudio que las diferencias de retardo extremo-a-extremo entre diferentes receptores en un escenario IPTV pueden llegar a ser superiores a 6s (con retardos que oscilan entre 250ms y 6500ms).

El trabajo de investigación en [8] proporciona medidas sobre diferencias entre tiempos de reproducción cuando se recibe el mismo contenido multimedia utilizando diferentes tecnologías de TV, como son diferentes variantes del estándar DVB, redes de cable analógicas, IPTV y TV vía Web. Se indica que las diferencias de retardos entre tecnologías broadcast pueden llegar a ser en torno a 5s, incluso hasta 8s si se comparan con tecnologías de TV vía Web.

De manera similar, el estudio en [14] proporciona medidas sobre la magnitud de las diferencias de retardos para diferentes tecnologías de TV. Se indica que las diferencias de retardos entre tecnologías TV broadcast en un entorno nacional (Holanda) pueden llegar a ser del orden de 5s, mientras que en un entorno internacional (entre Holanda y el Reino Unido) pueden llegar a ser del orden de 6s. Asimismo, estas medidas reflejan que las tecnologías broadcast analógicas tienen un retardo significativamente menor que las digitales y que, en general, las señales HD tienen un retardo ligeramente superior a las de definición estándar (Standard Definition o SD). Además, se comprobó que las diferencias de retardos entre tecnologías de TV broadcast y tecnologías de TV basadas en Web (broadband), como DASH, pueden llegar a ser superiores a 1 minuto. Por último, en [14] también se percibieron diferencias de retardos entre receptores muy significativas cuando se utiliza exactamente el mismo dispositivo de consumo, la misma tecnología de TV y la misma configuración (p.ej., mismo tipo de suscripción y misma calidad de señal) en cada uno de ellos. Sin embargo, no se proporcionan magnitudes para estas diferencias, debido a la falta de resultados concluyentes, ya que, según los autores, no se disponía de suficientes medidas para diferentes combinaciones y números de usuarios distribuidos geográficamente, utilizando diferentes tipos de tecnologías y configuraciones.

Tabla I
FACTORES QUE CONTRIBUYEN AL RETARDO Y A SU VARIABILIDAD [13]

	Factores	Rango de Retardos (ms)
Emisión	Captura de Video	17 - 40
	Codificación de Video	50 - 2000
	Cifrado	0 - 50
	Protección frente a errores	0 - 100
	Buffer de transmisión	50 - 500
Red	Transmisión de subida	10 - 300
	Trans-codificación	0 - 2000
	Transmisión de bajada	10 - 300
Recepción	Jitter buffer	50 - 500
	Protección frente a errores	0 - 100
	Descifrado	0 - 50
	Decodificación de Video	50 - 500
	Display buffer	0 - 50
TOTAL		250 - 6500

IV. UMBRALES TOLERABLES PARA LA PERCEPCIÓN HUMANA

Conocida ya la magnitud de los retardos y de su variabilidad en escenarios reales, en esta sección se presentan los umbrales tolerables en cuando a diferencias de retardos en diferentes tipos y ejemplos de sincronización multimedia, basándose en las conclusiones de trabajos de investigación relacionados. Asimismo, se va a argumentar brevemente y de manera cualitativa el impacto de otros parámetros de QoS, como son el ancho de banda y la tasa de pérdidas, sobre la percepción humana (QoE), desde el punto de vista de la sincronización multimedia.

A. Niveles de asincronías permisibles

Desde los inicios de las comunicaciones multimedia, la problemática del jitter ha sido palpable, pues repercute significativamente en los niveles de fidelidad e inteligibilidad de las señales multimedia, incluso en entornos locales. De hecho, ya se pueden encontrar en la década de los 70 estudios que analizaron el impacto del jitter en la sincronización intra-media. Por ejemplo, en [15] se defiende que la máxima asincronía o diferencia de retardos, también referida en algunas publicaciones como desviación (*skew*) o error, para la sincronización intra-media en audio de alta calidad es de unos 200ns.

Asimismo, con la introducción de los primeros servicios de contenido bajo demanda, CoD (Content on Demand), y de audio/video conferencia, la sincronización inter-media (principalmente *lip-sync*) se identificó como un problema a solucionar, por lo que varios estudios han tratado de determinar los umbrales tolerables para la percepción humana (QoE) en estos casos. Por ejemplo, los estudios subjetivos realizados en [16] reflejan que asincronías inferiores a 80ms para *lip-sync* son indetectables, mientras que asincronías superiores a 160ms son inaceptables. En [17], se indica que asincronías entre -80ms (audio atrasado con respecto al video) y +80ms (audio adelantado con respecto al video) son perceptibles, pero tolerables para la mayoría de los usuarios, mientras que niveles de asincronía superiores a -240ms y +160ms son intolerables. En este trabajo, además, se clasifican diferentes niveles de QoS en cuanto a límites de asincronía, dependiendo del tipo de información multimedia, sus parámetros y la aplicación específica, oscilando entre requisitos de sincronización muy estrictos para audio/audio (+/-11µs) hasta requisitos menos estrictos para audio/puntero (-500ms to 750ms).

De manera similar, en [18] se clasifican los umbrales para *lip-sync* en 3 rangos: imperceptibilidad (desde -95ms hasta +25ms), perceptibilidad (desde -125 hasta +45ms) y aceptabilidad (desde -185 hasta +90ms). Estos umbrales se ilustran en la Fig.4. El eje vertical de esta figura representa una diferencia constante entre el umbral perceptible y el aceptable (45ms para audio adelantado y 60ms para audio atrasado). El área roja representa los rangos intolerables para audio adelantado y atrasado, mientras que el área verde representa los límites de imperceptibilidad.

En [19], se recomiendan niveles más estrictos para *lip-sync*, estableciendo los niveles aceptables entre +15ms para audio adelantado y -45ms para audio atrasado.

Normalmente, en los estudios sobre perceptibilidad relacionados con *lip-sync* se concluye que los usuarios son más sensibles a asincronías cuando el audio está adelantado

con respecto al video, que al revés. Esto se puede explicar debido a que la velocidad de la luz es mucho mayor que la velocidad del sonido y, por tanto, en condiciones normales (es decir, asumiendo la misma distancia a la pantalla y a los altavoces), la señal de audio llega más tarde a los usuarios que la señal de video [20].

En [21] se analizaron los niveles de asincronía permisibles entre varias señales acústicas (estéreo) para un array de micrófonos. Se concluyó que una asincronía de 17ms entre las señales de audio es perceptible y que es preferible que la asincronía no supere los 11ms.

En [22] se evaluó la calidad de sincronización inter-media entre los componentes de video en videos 3D estereoscópicos, con una tasa de 25 tramas por segundo (*fps*, *frames per seconds*) y con una resolución espacial de 1920x1080. Se indica que una asincronía inferior a 80ms garantiza una buena calidad visual 3D, mientras que la QoE es inaceptable si la asincronía es superior a 200ms.

En [23] se evaluaron subjetivamente los niveles de asincronía aceptables entre un flujo de video y un flujo de datos multimedia provenientes de dispositivos que incluyen información sobre el tacto (*haptics*). Se demuestra que asincronías en torno a 40-80ms son apenas perceptibles, mientras que asincronías superiores a 300ms son molestas.

En [24] se realizaron estudios subjetivos para evaluar los niveles de asincronía tolerables entre aromas y contenido audiovisual, asumiendo una *lip-sync* perfecta. La conclusión fue que los umbrales de sincronización son de 30s cuando los aromas se presentan antes que el contenido audiovisual y de 20s cuando se presentan después.

Combinando los dos estudios anteriores, en [25] se investigaron los límites permisibles para la sincronización entre aromas y haptic, estableciendo los umbrales permisibles entre 1s y 3s.

Con respecto a la sincronización de subtítulos con contenido audiovisual, no se dispone de conocimiento sobre la existencia de estudios que hayan tratado de identificar los umbrales de sincronización permisibles. Esto podría ser debido a que los subtítulos normalmente se enviaban multiplexados con el contenido audiovisual en los sistemas multimedia tradicionales. Sin embargo, en los sistemas multimedia modernos, es posible la generación automatizada y en vivo de los subtítulos, posiblemente por fuentes independientes, así como su personalización (en cuanto a formato, idioma...) y su presentación en entornos multi-pantalla. Es por ello que la sincronización de subtítulos con contenido audiovisual, ya sea en la pantalla principal como en pantallas secundarias, es un problema de investigación actual e interesante.

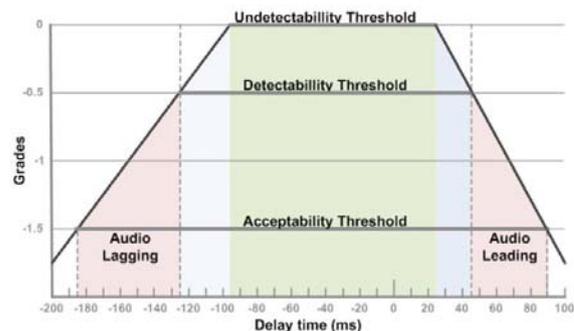


Fig. 4. Umbrales de percepción para lip-sync [19].

Con respecto a la sincronización IDES, varios umbrales de asincronía permisibles se indican en [26]: +/- 10µs para audio altamente acoplado; 2ms para audio en tiempo real; 15ms para audio adelantado y 45ms para video retrasado en *lip-sync*; y +/-80ms para animaciones de video.

Con respecto a la sincronización híbrida, también queda claro que existirán diferentes niveles de asincronía tolerables, en función del caso de uso. Por ejemplo, cuando se presenta contenido web relacionado con el flujo broadcast, los niveles de sincronización necesarios no son muy estrictos, pero si se reproducen videos multi-vista, ya sea en modo mosaico en un mismo dispositivo o en escenarios multi-pantalla, se requiere sincronización a nivel de trama (*frame-accurate synchronization*). Este concepto es relevante, porque va a vincular los niveles temporales de sincronización necesarios a la tasa de envío/reproducción y, por tanto, al ancho de banda consumido por cada componente de video. Por ejemplo, para videos con una tasa de de 25fps, la asincronía no deberá superar los 40ms (1/25 ms). Así pues, se puede deducir que a mayor tasa de tramas, menores serán los niveles de asincronía permisibles. Sin embargo, el concepto de sincronización a nivel de trama es un poco ambiguo, puesto que una asincronía tan pequeña como 1ms ya podría provocar que se estuviesen reproduciendo tramas diferentes de cada flujo durante ese periodo.

Por último, en [20] se recopilan y analizan hasta 20 casos de uso diferentes en los que IDMS es necesaria o ventajosa. Además, dichos casos de uso se clasifican, de manera cualitativa y aproximada, en función de los niveles de sincronización que precisan, según datos recogidos en publicaciones relacionadas. El resumen de esta clasificación se puede consultar en la Tabla II. Se puede apreciar que los umbrales de asincronías permisibles para los sistemas de audio distribuidos son muy estrictos (incluso por debajo del milisegundo). Asimismo, las aplicaciones de multi-conferencia y juegos en red también requieren niveles de sincronización bastante estrictos y, ya en el nivel inferior, aplicaciones como Social TV son más flexibles en este aspecto. Por ejemplo, las evaluaciones subjetivas realizadas en [27] mostraron que 500ms es el nivel inferior de asincronía que empiezan a detectar algunos usuarios en un entorno Social TV distribuido, que asincronías de 1s ya suelen ser perceptibles (más para los usuarios que chatean mediante servicios de audio-conferencia que para los que lo hacen mediante texto) y que asincronías de 2s ya son muy molestas para la gran mayoría de usuarios.

Sin embargo, a pesar de dicha clasificación preliminar, se necesita más investigación con tal de determinar, de manera subjetiva y rigurosa, los límites de asincronías permisibles en diferentes escenarios y aplicaciones IDMS, en función del canal de interacción (texto, audio y video chat) utilizado.

De manera general, se puede apreciar que los niveles de asincronía permisibles para los diferentes tipos de sincronización y tipos de datos multimedia (excepto para aromas) son bastante menores que las magnitudes de diferencias de retardos existentes en las redes actuales. Por tanto, se van a necesitar soluciones precisas y adaptativas para la sincronización multimedia, con el objetivo final de proporcionar una QoE satisfactoria, dados unos niveles de QoS ofrecidos por el entorno de red.

Tabla II
UMBRALES DE ASINCRONÍA PERMISIBLES PARA IDMS

Nivel de sincronización	Requisitos temporales aproximados	Casos de Uso Relevantes
Muy Alto	~ 10 µs – 10 ms	- Altavoces estéreo en red - Array de transductores acústicos en fase - Video walls en red
Alto	~ 10 – 100 ms	- Tele-orquesta distribuida - Juegos de preguntas en red - Juego en red multi-jugador - Conferencia en grupo - Sistemas de conferencia local con amplificación de audio - Aplicaciones multi-pantalla
Medio	~ 100 – 500 ms	- Educación a distancia - Trabajo en grupo en red - Juego de baja interactividad - Votaciones online - Conmutación de la reproducción entre dispositivos
Bajo	~ 500 – 2000 ms	- Compartición de comandos de navegación en TV - TV Social

B. Impacto del ancho de banda

Se ha comentado previamente que algunos tipos y ejemplos de sincronización multimedia requieren niveles de sincronización a nivel de trama. Así pues, la tasa de tramas de los flujos multimedia involucrados y, por tanto, el ancho de banda consumido, influirá en los umbrales de asincronía permisibles.

Asimismo, el envío de datos multimedia con codificación escalable o multi-capa, así como audio multi-canal o espacial, influirá tanto en el ancho de banda necesario como en los niveles de sincronización que se precisan, ya que todos los componentes de las señales multimedia se deberán reproducir de forma sincronizada.

También se intuye que el tamaño y formato de video, el tipo de la pantalla y de dispositivo de consumo pueden influir en la percepción humana. Sin embargo, no existen otros estudios que hayan determinado la influencia de dichos factores sobre los umbrales de asincronía permisibles. Únicamente, en [28] se observa que, en dispositivos móviles, el error de sincronización es más tolerable cuando menor es el tamaño de la imagen (para tasas de tramas reducidas), pero no se proporcionan umbrales.

C. Impacto de la tasa de pérdidas

Es obvio que la pérdida de paquetes tiene un claro impacto sobre la fluidez de la reproducción de los flujos multimedia y, por tanto, sobre los diferentes tipos de sincronización multimedia. Por lo tanto, es interesante determinar el impacto de la pérdida de paquetes y la degradación de video y audio, para diferentes formatos de codificación, sobre la sincronización multimedia y, en definitiva, sobre la percepción de los usuarios (QoE), en diferentes tipos de servicios multimedia.

En este sentido, cabe destacar que en el caso de que los diferentes componentes multimedia se multiplexen en un flujo agregado, las pérdidas de paquetes afectarán a todos los componentes, mientras que esto podría no ocurrir si se envían en flujos separados. Además, las pérdidas de paquetes pueden tener mayor impacto cuando se incluyen datos multimedia discretos (e.g., aromas), porque la pérdida de un

único paquete podría provocar que la MU en cuestión no se pueda presentar al usuario, privando de la inmersividad deseada. Por tanto, en estos casos, el diseño de mecanismos eficientes de recuperación frente a errores es muy deseable.

V. INFLUENCIA DE LOS ESQUEMAS DE CONTROL IDMS

En cuanto a la sincronización IDMS, las soluciones existentes se pueden agrupar según el uso de 3 esquemas de control o arquitecturales diferentes para el intercambio de la información necesaria para conseguir la sincronización (véase la Fig.5) [5, 29]: dos de ellos centralizados (M/S Scheme y Synchronization Maestro Scheme o SMS) y otro distribuido (Distributed Control Scheme o DCS). Por un lado, en [5] se proporciona una comparación cualitativa exhaustiva entre estos 3 esquemas de control considerando varios factores relacionados tanto con la QoS (p.ej., robustez, escalabilidad, sobrecarga de tráfico...) como con la QoE (p.ej., igualdad de condiciones, coherencia...), cuyo resumen se puede consultar en la Tabla III. Por otro lado, esta comparación cualitativa se corrobora mediante una evaluación objetiva en entornos simulados en [29].

VI. INFLUENCIA DE LA SELECCIÓN DE LA REFERENCIA TEMPORAL PARA LA SINCRONIZACIÓN

En los diferentes tipos de sincronización multimedia, cuando se detectan situaciones de asincronía, la primera decisión consiste en seleccionar la referencia temporal a la que sincronizarse. En el caso de sistemas incluyendo audio y video (*lip-sync*), el audio se suele seleccionar como el flujo de referencia, ya que los usuarios son más tolerables a ajustes en video que en audio [17]. Sin embargo, en algunos casos, se selecciona el flujo que experimenta mayor retardo como la referencia, puesto que no es posible reproducir el contenido multimedia que todavía no se ha recibido.

Con respecto a la sincronización híbrida, como los flujos broadband normalmente experimentan mayores retardos que los flujos broadcast, ello implica que el contenido broadcast se debe retrasar hasta recibir el contenido broadband asociado, lo que conllevará problemas de interactividad en servicios de TV. Así pues, es necesario y muy relevante minimizar la latencia en las tecnologías de HTTP streaming.

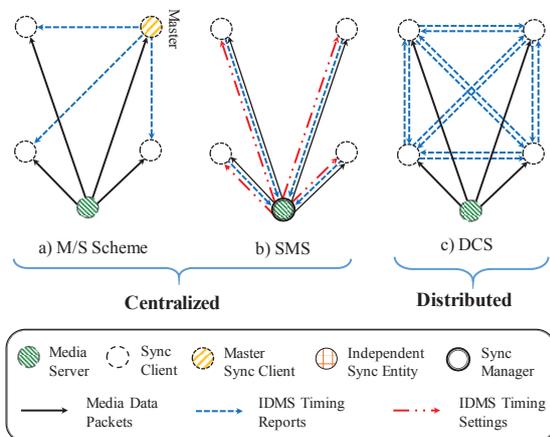


Fig. 5. Esquemas de control IDMS.

Tabla III
COMPARACIÓN ENTRE ESQUEMAS DE CONTROL IDMS

		Factors									
		Robustness	Scalability	Traffic Overhead	Interactivity	Consistency	Causality	Coherence	Flexibility	Fairness	Security
Scheme	M/S	3	1	1	1	2	1	3	3	3	2
	DCS	1	2	3	2	3	3	2	1	1	3
	SMS	2	3	2	3	1	2	1	2	2	1

1: Mejor Esquema; 2: Buen esquema; 3: Peor Esquema

En los sistemas multimedia modernos, que pueden incluir información multimedia multi-sensorial (*multimedia*), como aromas o haptics, de manera adicional a los flujos de audio y video, un tema de investigación muy interesante es determinar las políticas más idóneas para la selección del flujo o el componente multimedia de referencia para la sincronización, tal y como se sugiere en [7].

En el caso de IDMS, si se utilizan los esquemas SMS y DCS, se pueden emplear varias políticas dinámicas para la selección de la referencia temporal para la sincronización [29, 30], con diferentes implicaciones sobre la QoS y la QoE. La selección del receptor más rápido como la referencia puede ser ventajosa para aplicaciones colaborativas, pero puede provocar situaciones de *buffer underflow* si el receptor que marca la referencia es más rápido que ciertos límites permisibles. La selección del receptor más lento como la referencia es apropiada para garantizar igualdad de condiciones (*fairness*) en aplicaciones de ámbito competitivo, pero puede provocar situaciones de *buffer overflow* y retardos excesivos si el receptor que marca la referencia es más lento que ciertos límites permisibles. También se puede seleccionar un punto de reproducción virtual, obtenido como la media de todos los puntos de reproducción involucrados, como la referencia para la sincronización. Esta estrategia permite minimizar el número y la magnitud de los ajustes de sincronización necesarios, pero no puede garantizar que se eviten situaciones de *buffer underflow/overflow*. Otra posibilidad es definir como referencia a un receptor virtual con un punto de reproducción idóneo (sin desviaciones), al que todos los receptores se deberán ajustar. Esta estrategia garantiza, siempre y cuando las condiciones de la red sean estables, una evolución constante del retardo extremo-a-extremo, que es una característica deseable en las comunicaciones multimedia. Finalmente, otras posibilidades son la selección del participante más activo o el que tenga un rol más prioritario (p.ej., el profesor en servicios de *e-learning*) como la referencia para la sincronización, pues seguramente será el foco de atención de la sesión multimedia.

Por último, es importante recalcar que todas estas estrategias de selección de la referencia temporal para la sincronización serán aplicables siempre y cuando tanto los retardos de red como los retardos extremo-a-extremo se mantengan dentro de límites permisibles [6].

VII. CONCLUSIONES

En este artículo se ha descrito la relevancia de los diferentes tipos y ejemplos de sincronización multimedia sobre la QoE, así como su relación con diferentes parámetros de QoS, como son el ancho de banda, la pérdida de paquetes

y, especialmente, los retardos y su variabilidad. Se han proporcionado datos que demuestran que los niveles de asincronías tolerables dependen en gran medida del tipo de sincronización multimedia, del tipo de datos multimedia y del tipo de aplicación bajo estudio. Un conocimiento riguroso y exhaustivo de estos límites es muy importante, pues dará una idea de los niveles de QoS que las redes de distribución de contenidos actuales deberán ofrecer. Asimismo, se ha mostrado que las diferencias de retardos en redes actuales son muy superiores a los niveles de asincronía tolerables, reflejando así la relevancia de la sincronización multimedia. Es por ello que es necesario minimizar, acotar y controlar tanto la magnitud de los retardos como de su variabilidad para proporcionar una QoE satisfactoria. Por último, se ha descrito la influencia de la selección del esquema de control y de la referencia temporal para la sincronización multimedia sobre varios aspectos de QoS y de QoE. La necesidad de más investigación en esta área también se ha identificado y reflejado a lo largo del artículo.

AGRADECIMIENTOS

El trabajo de Mario Montagud se ha llevado durante el disfrute de un contrato postdoctoral financiado por ERCIM 'Alain Bensoussan' Fellowship Programme. El trabajo de la UPV ha sido financiado, en parte, por el FEDER y por el MINECO, bajo el programa de apoyo a la I+D+i en el proyecto con referencia TEC2013-45492-R.

REFERENCIAS

- [1] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", IETF Internet Standard, RFC 3550, July 2003.
- [2] Hybrid broadcast broadband TV, <http://www.hbbtv.org/>
- [3] C. Concolato, S. Thomas, R. Bouqueau, J. Le Feuvre, "Synchronized Delivery of Multimedia Content over Uncoordinated Broadcast Broadband Networks", MMSys'12, North Carolina (USA), Feb. 2012.
- [4] R. van Brandenburg, H. Stokking, O. Van Deventer, F. Boronat, M. Montagud, K. Gross, "Inter-destination Media Synchronization using the RTP Control Protocol (RTCP)", RFC 7272, June 2014.
- [5] M. Montagud, F. Boronat, H. Stokking, R. van Brandenburg, "Inter-destination multimedia synchronization: schemes, use cases and standardization", MMSJ, 18(6), 459-482, November 2012.
- [6] M. Montagud, "Design, Development and Evaluation of an Adaptive and Standardized RTCP-based IDMS Solution", PhD Thesis, UPV (Spain), Supervisors: F. Boronat (UPV) and P. Cesar (CWI), March 2015.
- [7] Z. Huang, K. Nahrstedt, R. Steinmetz, "Evolution of temporal multimedia synchronization principles: A historical viewpoint", ACM TOMCCAP, 9, 1s, Article 34, 23 pages, October 2013.
- [8] F. Boronat, R. Mekuria, M. Montagud, P. Cesar, "Distributed Media Synchronization for Shared Video Watching: Issues, Challenges, and Examples", Social Media Retrieval, Springer Computer Communications and Networks, ISBN 978-1-4471-4554-7, 2013.
- [9] J. Jansen, D.C.A. Bulterman, "User-centric video delay measurements", ACM NOSSDAV 2013, pp. 37-42, Oslo (Norway), February 2013.
- [10] ITU-T Rec. G.1050: "Network model for evaluating multimedia transmission performance over Internet Protocol", 2007.
- [11] ITU-T Rec. G. 114: "One-way transmission time", 2003.
- [12] J. Jansen, P. Cesar, D.C.A. Bulterman, T. Stevens, I. Kegel, J. Issing, "Enabling Composition-Based Video-Conferencing for the Home", IEEE Transactions on Multimedia, 13(5), pp. 869-881, October 2011.
- [13] M. O. Van Deventer, H. Stokking, O. A. Niamut, F. A. Walraven, V. B. Klos, "Advanced Interactive Television Service Require Synchronization", IWSSIP 2008, Bratislava (Slovak Rep.), June 2008.
- [14] W. Kooij, H. Stokking, R. van Brandenburg, P-T de Boer, "Playout delay of TV signals: measurement system design, validation and results", ACM TVX 2014, Newcastle (UK), June 2014.
- [15] B. Blesser, "Digitization of audio: A comprehensive examination of theory, implementation, and current practice", J. Audio Eng. Soci., 26(10), pp. 739-771, 1978.
- [16] R. Steinmetz, C. Engler, "Human perception of media synchronization", Technical Report 43.9310, IBM European Networking Center Heidelberg, 1993.
- [17] R. Steinmetz, "Human perception of jitter and media synchronization", IEEE JSAC, 14(1), pp. 61-72, January 1996.
- [18] ITU-R BT.1359, International Telecommunication Union Radio communication Sector Relative Timing of sound and Vision for Broadcasting.
- [19] ATSC Implementation Subcommittee Finding: Relative Timing of Sound and Vision for Broadcast Operations, Doc. ID-191, June 2003.
- [20] A. Mason, R. Salmon, "Factors affecting perception of audio video synchronization in television", BBC R&D White Paper WHP176, 2009
- [21] R. Dannenberg, R. Stern, "Experiments concerning the allowable skew of two audio channels operating in the stereo mode". Personal Communications, 1993.
- [22] L. Goldmann, L. Jong-Seok Lee, T. Ebrahimi, "Temporal synchronization in stereoscopic video: Influence on quality of experience and automatic asynchrony detection", IEEE ICIP 2010, Hong Kong, September 2010
- [23] T. Fujimoto, Y. Ishibashi, S. Sugawara, "Influences of Inter-Stream Synchronization Error on Collaborative Work in Haptic and Visual Environments", Symposium on Haptic interfaces for virtual environment and teleoperator systems, March 2008.
- [24] Ghinea, G.; Ademoye, O.A., "Perceived Synchronization of Olfactory Multimedia", Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, vol.40, no.4, pp.657-663, July 2010
- [25] S. Hoshino, Y. Ishibashi, N. Fukushima, S. Sugawara, "QoE assessment in olfactory and haptic media transmission: Influence of inter-stream synchronization error", IEEE CQR Workshop, May 2011.
- [26] Y. Bang, J. Han, K. Lee, J. Yoon, J. Joung, S. Yang, J. Rhee. "Wireless Network Synchronization for Multichannel Multimedia Services", IEEE ICAT 2009, Sarajevo (Bosnia and Herzegovina), February 2009.
- [27] D. Geerts, I.Vaishnavi, R. Mekuria, O. van Deventer, and P. Cesar, "Are we in sync?: synchronization requirements for watching online video together", ACM CHI 2011, Vancouver (Canada), May 2011.
- [28] I. Curcio, M. Lundan, "Human Perception of Lip Synchronization in Mobile Environment", IEEE WoWMoM, Helsinki, June 2007.
- [29] M. Montagud, F. Boronat, H. Stokking, P. Cesar "Design, Development and Assessment of Control Schemes for IDMS in a Standardized RTCP-based Solution", COMNET, 70(9), pp. 240-259, Sept. 2014.
- [30] M. Montagud, F. Boronat, "Enhanced Adaptive RTCP-based Inter-Destination Multimedia Synchronization Approach for Distributed Applications", COMNET, 56(12), pp. 2912-2933, August 2012.