

INFORME UIT-R BT.2044

**Retardo de tiempo admisible de ida y vuelta para las inserciones en programas
radiofónicos y de radiodifusión de televisión – Contexto y fundamentos**

(Cuestión UIT-R 35/6)

(2004)

ÍNDICE

Página

1	Ámbito	2
2	Referencias Bibliográficas y Bibliografía	2
3	Términos y definiciones	3
4	Abreviaturas y siglas	3
5	Causas del retardo y el eco	4
5.1	Contexto.....	4
5.2	Antecedentes históricos	4
5.3	Diseño del sistema y factores de configuración frente a las prácticas de funcionamiento	5
5.4	Diseño del sistema y factores de configuración que causan retardo y eco	6
5.4.1	Procesos de codificación/decodificación.....	6
5.4.2	Retardo de propagación	6
5.5	Magnitud del retardo en varios medios de transmisión	7
5.6	Magnitud del retardo en los códecs de vídeo.....	7
5.7	Magnitud del retardo en los códecs de audio.....	8
5.8	Simetría del circuito.....	10
5.9	Prácticas operativas que causan eco	10
5.10	Retardo entre el sonido y la imagen	10
6	Efectos del retardo y el eco	10
6.1	Retardo de audio y retardo de vídeo	10
6.2	Variables que afectan la perturbación.....	11
6.3	Ecos y fusión.....	11
6.4	Tipos de perturbación y sus efectos	12
6.5	Adaptación.....	12
7	Resumen	12

1 Cometido

Este Informe tiene por objeto examinar los efectos del retardo y el nivel del eco en un bucle de repliegue de audio en el contexto de producción de radiodifusión. También estudia los efectos del retardo de audio-vídeo. No se contemplan los efectos de reverberación y ruido adicionales en el entorno de escucha y se supone que no hay una pérdida significativa en la calidad de la señal del sistema de comprobación técnica.

2 Referencias Bibliográficas y Bibliografía

- [1] TECHNICAL REPORT ETR 250 [julio de 1996] Transmission and Multiplexing (TM). Speech communication quality from mouth to ear for 3.1 kHz handset telephony across networks European Telecommunications Standards Institute, <http://pda.etsi.org/pda/queryform.asp>.
- [2] ETSI Technical report ETR 262 [enero de 1996] Broadband integrated services digital network (B-ISDN); Asynchronous transfer mode (ATM); Video on demand (VOD) network aspects. European Telecommunications Standards Institute.
- [3] TECHNICAL REPORT ETR 275 [abril de 1996] Transmission and Multiplexing (TM). Considerations on transmission delay and transmission delay values for components on connections supporting speech communication over evolving digital networks. European Telecommunications Standards Institute, <http://pda.etsi.org/pda/queryform.asp>.
- [4] Recomendación UIT-R BT.1359 [noviembre de 1998] Temporización relativa del sonido y la imagen para la radiodifusión. Unión Internacional de Telecomunicaciones.
- [5] ETSI/GSM Recommendation 06.10 Version 3.2.0 [febrero de 1992] GSM Full Rate Speech Transcoding European Telecommunications Standards Institute, <http://pda.etsi.org/pda/queryform.asp>.
- [6] DAVIS, P. [1995] A tutorial on MPEG audio compression. *IEEE Multimedia*, p 60-74.
- [7] ONVURAL, R. [1994] *Asynchronous transfer mode networks: performance issues*, Boston, Artech House.
- [8] Technical Report TR 100 815 V1.1.1 [febrero de 1999] Digital video broadcasting; guidelines for the handling of Asynchronous transfer mode signals in DVB systems. European Telecommunications Standards Institute.
- [9] C.S0014-0 Version 1.0 Enhanced Variable Rate Codec (EVRC) (CDMA2000 specification). Third Generation Partnership Project 2 (3GPP2) EIA/TIA, diciembre de 1999.
- [10] SIU-WAH WONG [1991] An Evaluation of 6.4 kbit/s Speech Codecs for Inmarsat-M System. *IEEE*, p. 629-632.
- [11] Recomendación UIT-R BT.1377 [1998] Marcado del retardo (procesamiento) del caudal en los aparatos de vídeo y audio. Unión Internacional de Telecomunicaciones.
- [12] EVEREST, F.A. [1994] *The Master Handbook of Acoustics*. TAB Books/McGraw-Hill.
- [13] Recomendación UIT-T G.114 (2003) – Tiempo de transmisión en un sentido.
- [14] Recomendación UIT-T G.113 (2001) – Degradaciones de la transmisión debido al tratamiento de las señales vocales.
- [15] ES 200 677 V1.2.1 Public Switched Telephone Network (PSTN) [marzo de 1998] Requirements for handset telephony. European Telecommunications Standards Institute.

- [16] LOCHNER, J.P.A. & BURGER, J.F. [1958] The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech. *Acustica* Vol. 8, 1 p. 1-10.
- [17] MEYER, E. & SCHODDER, G.R. [1952] Über den Einfluß von Schallrückwürfen auf Richtungslokalisation und Lautstärke bei Sprache. *Nachr. Akad. Wiss.*, Göttingen, 6.
- [18] CHURCH, S. On beer & audio coding – why something called AAC is cooler than a pilsner & how it got to be that way. <http://www.broadcastpapers.com/audio/TelosAAC08.htm>.

3 Términos y definiciones

Latencia:	Retardo en los sistemas con conmutación de paquetes debido al encaminamiento y colas de los paquetes.
Fluctuación de fase:	Variación del retardo durante una transmisión en un sistema con conmutación de paquetes.
Trama-I:	Trama inicial en un grupo de imágenes MPEG.
Trama-B:	Trama interpolada bidireccional entre una trama I y una trama P en un grupo de imágenes MPEG.
Trama-P:	Trama interpolada en un grupo de imágenes MPEG.
Efecto local:	Parte atenuada de la señal de audio transmitida que vuelve al origen. Puede ser intencional en los teléfonos. Está provocado por un desequilibrio en el sistema híbrido de 2 a 4 hilos o por una terminación de línea incorrecta.
Sistema de 2 hilos:	Sistema en el que se transmiten las señales en ambos sentidos por la misma línea de transmisión.
Sistema de 4 hilos:	Sistema en el que se transmiten las señales en líneas de transmisión distintas para cada sentido.
Híbrido de audio:	Circuito de audio diferencial utilizado para realizar la conversión entre un sistema de 2 hilos y un sistema de 4 hilos. Normalmente utiliza un transformador de audio con un arrollamiento dividido en una configuración tipo puente o un equivalente activo.

4 Abreviaturas y siglas

ms:	Milisegundo (10^{-3} s)
μs:	Microsegundo (10^{-6} s)
ns:	Nanosegundo (10^{-9} s)
códec:	Codificador/decodificador
ETSI:	Instituto Europeo de Normas de Telecomunicación (<i>European telecommunication standards Institute</i>)
UIT:	Unión Internacional de Telecomunicaciones
MPEG:	Grupo de Expertos de imágenes en movimiento
tps:	Tramas

5 Causas del retardo y el eco

5.1 Contexto

Los sistemas de comunicaciones en bucle se emplean normalmente en la producción de radiodifusión donde se utilizan emplazamientos múltiples o distantes, especialmente en la producción de programas en directo. El bucle puede servir para conectar un entrevistador situado en el estudio con un invitado que se encuentra a distancia o también para conectar el estudio a un presentador que se encuentre en un lugar distante. Idealmente, cada tren separado de audio y vídeo debe desplazarse únicamente en un sentido en este bucle. En la práctica, es bastante común que la señal de audio retorne por el bucle hasta su origen, especialmente cuando los teléfonos forman parte del circuito. Debido a la distancia y/o la necesidad de realizar los procesos de codificación/decodificación para comprimir la señal digitalizada en la transmisión y restaurarla nuevamente en recepción, la señal puede sufrir un retardo muy significativo. El retardo y la fuga de señal en torno a un bucle de audio se combinan para producir eco.

Como las señales de vídeo han necesitado normalmente una sincronización mucho más estricta que las de audio, no ha sido muy común transmitir vídeo a lo largo de bucles de gran longitud. Con la llegada del vídeo digital, la sincronización se ha simplificado y los equipos han reducido su tamaño, por lo que las transmisiones de vídeo en bucles de gran longitud pueden convertirse en una práctica más habitual. Sin embargo, en el instante de redactar estas líneas aún no es muy común que la señal de vídeo se desplace en torno a un bucle completo para retornar a su fuente. Sin embargo, la necesidad de sincronizar el audio con el vídeo puede dar lugar a que el retardo del vídeo sea el factor limitativo en un bucle de comunicaciones audio-vídeo.

Cuando se utilizan bucles de gran longitud, los retardos y el eco pueden provocar importantes dificultades, tanto a la hora de mantener la calidad técnica de la señal como para evitar perturbaciones a los participantes. Ambos efectos interfieren con las operaciones de producción.

El retardo y el eco son problemas distintos pero relacionados. El retardo se debe a varios factores; principalmente el procesamiento de codificación/decodificación y la longitud del trayecto desde el origen al destino¹. El eco se debe a las fugas que se producen en torno a un bucle. A medida que aumenta el retardo, el eco asociado cada vez es más apreciable y más perturbador.

5.2 Antecedentes históricos

Las molestias causadas por el retardo y los ecos del sonido son familiares en salas de gran tamaño y en los estadios deportivos y, desde luego, no es un problema nuevo. En las comunicaciones electrónicas, el problema apareció en primer lugar en la telefonía y la mayoría de los conocimientos que se poseen sobre las causas y efectos del retardo en la radiodifusión se derivan de las experiencias obtenidas en esta área. Por consiguiente, es útil tener algunos conocimientos sobre las causas y efectos del retardo en telefonía y las estrategias que se han adoptado para controlar este problema. La sincronización del sonido y la imagen tampoco es un problema nuevo y los conocimientos que se tienen en esta área se basan fundamentalmente en las experiencias con el cine.

Cuando aparecieron por primera vez los sistemas de telefonía, cada línea telefónica requería un par de hilos separado, no solamente en la central local sino también entre centrales. En consecuencia, el coste de los pares de hilos suponía una gran parte del coste total de la infraestructura. La utilización

¹ La longitud de un trayecto de señal puede ser considerablemente mayor que la distancia directa o geodésica entre dos puntos. La diferencia entre la longitud del trayecto de la señal y la distancia geodésica es particularmente importante cuando se utilizan satélites geoestacionarios.

de un transformador «híbrido» de audio (fundamentalmente un sistema en puente para combinar y separar circuitos equilibrados) permitió que un solo par de hilos (sistema de dos hilos) transmitiese señales en ambos sentidos, logrando así reducir a la mitad los costes de transmisión de un sistema de cuatro hilos.

Las funciones de combinación y separación en este sistema no funcionaban adecuadamente pues exigían una buena adaptación de impedancias con la línea, en el extremo emisor y una línea con una terminación adecuada en el extremo receptor. Las características eléctricas de las líneas de transmisión por hilos varían con la frecuencia, la distancia y las condiciones atmosféricas (los primeros teléfonos, al igual que el telégrafo, utilizaban líneas áreas) y los primeros circuitos híbridos sólo podían adaptarse para unas características de línea fijas. Por lo tanto, la fuga a través del híbrido en el microteléfono, conocida como efecto local, es inevitable debido a la desadaptación local y al eco en la línea. La amplitud y fase del efecto local no es uniforme con la frecuencia y la señal de sonido que llega al auricular de un circuito híbrido con una ligera desadaptación no es agradable ni inspira confianza. Permitiendo una pequeña cantidad de fuga de banda ancha local a fin de suprimir la fuga por adaptación no uniforme, se logra una respuesta más uniforme de la voz del hablante en el auricular y también se le da cierta confianza de que el aparato está funcionando correctamente, al menos en el extremo cercado. Por esa razón, es una práctica habitual desde hace mucho tiempo permitir deliberadamente una cierta fuga entre las mitades de un bucle de audio en los sistemas de 2 hilos.

El eco procedente del extremo lejano de una línea desadaptada era un problema relativamente menor en los primeros sistemas telefónicos, pues las longitudes de las líneas rara vez superaban los 30 km. Cuando aparecieron los sistemas de amplificación y se hizo posible la telefonía a larga distancia, el problema del eco aumentó pues no sólo había más puntos en los que la línea podía estar desadaptada², sino que además el retardo del eco era notablemente mayor y ello aumentaba su audibilidad. La supresión del eco se convirtió en un problema importante para las compañías telefónicas y continúa siendo fundamental a la hora de mantener la calidad de las líneas de larga distancia.

Con la aparición de la multiplexación y la digitalización, los modernos sistemas telefónicos son sistemas a 4 hilos entre centrales y en muchos casos³, esta estructura se mantiene hasta el propio microteléfono. El eco en la línea constituirá entonces únicamente un problema en los teléfonos con altavoz.

En la radiodifusión, la utilización de sistemas de comunicaciones a 2 hilos se ha limitado a los canales de servicio y a los programas de línea abierta a los oyentes. La radiodifusión en exteriores ha utilizado normalmente líneas independientes para el control del sonido y la componente del programa. El eco es relativamente fácil de controlar en estas situaciones⁴ y suele ser el retardo, y no el eco, el problema fundamental para los diseñadores del sistema.

5.3 Diseño del sistema y factores de configuración frente a las prácticas de funcionamiento

El retardo y el eco pueden ser debidos al diseño del sistema, como en el caso de los sistemas de telefonía, a la configuración del sistema, tales como el grupo de imágenes establecido en un codificador de vídeo MPEG o a un inadecuado funcionamiento o práctica técnica, tales como

² Una línea telefónica analógica de larga distancia requiere repetidores cada 30 km, aproximadamente. En cada uno de estos puntos la línea puede estar desadaptada y dar lugar a ecos.

³ Por ejemplo, sistemas PABX digitales conectados de forma digital a la central local.

⁴ No hay garantías sin embargo de que sea controlado.

altavoces de control o auriculares con fugas en micrófonos abiertos. Para controlar el retardo y el eco en los bucles es necesario un adecuado diseño del sistema, o una configuración apropiada del sistema y una buena práctica de funcionamiento.

5.4 Diseño del sistema y factores de configuración que causan retardo y eco

Existen dos causas principales del retardo en los sistemas de audio/vídeo: los procesos de codificación/decodificación y la propagación de la señal.

5.4.1 Procesos de codificación/decodificación

Los procesos de codificación/decodificación en los sistemas digitales constan de dos etapas. La primera etapa de codificación es la conversión analógica a digital y la segunda etapa es la compresión de los datos. En la decodificación, el proceso es inverso: en primer lugar se descomprime la señal y a continuación se convierte de digital a analógica.

La conversión analógica a digital y la conversión digital a analógica son procesos relativamente rápidos y normalmente se llevan a cabo en el intervalo entre muestras. Para la radiodifusión de audio este intervalo es de 1/48 kHz, o aproximadamente 21 μ s. Para el audio de telefonía este intervalo es de aproximadamente 1/8 kHz, o unos 125 μ s. Para el vídeo de radiodifusión de definición normalizada es de 1/13 MHz, aproximadamente 77 ns.

La compresión de datos se utiliza para disminuir la necesidad de anchura de banda y se emplea casi universalmente en la radiodifusión para la transmisión a larga distancia (entre edificios, localidades, ciudades o países). El audio y el vídeo en la cadena de radiodifusión se transmiten normalmente en forma digital comprimida durante la producción, intercambio, distribución y emisión de los programas. La compresión de datos se utiliza casi también de forma universal para la grabación, aunque algunas de las normas de compresión para la grabación, son distintas de las de transmisión.

Los retardos debidos a la compresión de los datos son algo más largos que los causados por las conversiones analógico a digital y digital a analógico. Para lograr elevados factores de compresión, se disminuye la redundancia entre tramas de vídeo comparando las tramas con una trama patrón y codificando únicamente las diferencias. Por consiguiente, la codificación de vídeo se basa en un grupo de imágenes en lugar de realizarse trama a trama. El proceso de comparación de trama en la etapa de codificación puede utilizar bloques de hasta 15 tramas con subgrupos de hasta 4 tramas. Un tamaño de grupo de imagen de 15 imágenes da lugar a un mínimo retardo de 4 tramas. El retardo de codificación total incluido el tiempo de procesamiento puede ser considerablemente más elevado que el mínimo número de tramas intermedias y normalmente es, al menos, de 7 u 8 tramas para esta configuración. En la etapa de decodificación se producen retardos similares.

El retardo debido a la compresión de datos del audio digitalizado es de un orden similar aunque las técnicas de compresión son algo distintas de las que se utilizan para el vídeo.

5.4.2 Retardo de propagación

El retardo de propagación puede dividirse en dos factores: retardo de distancia y retardo de conmutación o latencia.

Los retardos de propagación debido a la distancia son del orden de 5 μ s/km de longitud de trayecto en un guíaondas, en un cable o en una fibra óptica y del orden de 3,3 μ s/km para las señales radiadas. En ambos casos, la longitud del trayecto puede ser similar a la distancia directa punto a punto (geodésica) o puede ser considerablemente mayor, como sucede en los casos de transmisión por satélite. Para largas distancias, este retardo puede ser comparable a los retardos de codificación/decodificación. Una señal que se desplace con propagación terrenal al lado opuesto del globo terráqueo presenta un retardo de distancia de unos 100 ms. Por satélite, el retardo de distancia para el mismo destino es de unos 280 ms, suponiendo que en la retransmisión intervenga un solo satélite geostacionario (situado a una altitud de 36 000 km sobre la superficie de la Tierra o 42 000 km desde el horizonte).

Por lo general, los retardos de conmutación no son perceptualmente significativos en los sistemas de portadoras analógicas o en los sistemas digitales síncronos. Los sistemas digitales con conmutación de paquetes, tales como el modo de transmisión asíncrono (ATM), presentan retardos en cada nodo de conmutación pues allí se retiene cada paquete y se lee su encabezamiento antes de retransmitirle al siguiente espacio disponible a través de una línea adecuada. Por tanto, se produce un retardo debido al proceso de lectura/asignación de trayecto y debido a la espera en cola. Este retardo se conoce como latencia. Como el trayecto de transmisión se asigna dinámicamente para cada paquete en los sistemas con conmutación de paquetes, también hay una cierta variabilidad en el retardo de transmisión, conocida como fluctuación de fase. Los retardos de conmutación de paquetes, sin embargo, son pequeños (aproximadamente unos pocos milisegundos) comparados con el retardo de distancia para grandes distancias. Los retardos por conmutación de paquetes también son bastante coherentes para un trayecto constante. La mayor parte de la fluctuación de fase proviene generalmente de la variación en la longitud del trayecto.

5.5 Magnitud del retardo en varios medios de transmisión

Existen dos causas principales del retardo: retardo de transmisión y retardo de codificación/decodificación. El retardo de transmisión varía con el modo de transmisión, dependiendo de que la transmisión sea terrenal o por satélite y continua o por paquetes.

- Retardo de transmisión
 - Trayecto por satélite: para los satélites geoestacionarios situados a una altitud de unos 36000 km por encima de la superficie de la Tierra, tales como los satélites de radiodifusión, el retardo de trayecto debido a la distancia varía desde unos 239 ms, cuando el satélite se encuentra directamente sobre la fuente y el receptor, a unos 281 ms cuando el satélite se encuentra en el horizonte de la fuente y el receptor.
 - Para satélites en órbita baja tales como los que se utilizan en el sistema de telefonía móvil Iridium, la altitud es de unos 780 km desde la superficie de la Tierra y el retardo es proporcionalmente inferior. Como los satélites en órbita baja no son geoestacionarios, el retardo desde estos satélites varía con su posición. Admitiendo un radio de cobertura de 1800 km para cada satélite (el sistema Iridium cuenta con 66 satélites) la máxima distancia al satélite es de 2000 km, lo que produce un retardo de propagación en un solo sentido de unos 6,6 ms. Retardos adicionales debidos a los enlaces entre satélites y al proceso a bordo aumentarían el retardo global en el sistema Iridium.
 - Analógico y RDSI: el retardo de transmisión en los sistemas de portadoras analógicas terrenales y RDSI terrenal es de unos $5 \mu\text{s}/\text{km}$, basándose en una velocidad de propagación en la fibra óptica o en el guiaoondas de cobre de aproximadamente $2/3$ la velocidad de la luz. Ello produce un retardo de 100 ms para un trayecto hasta el lado opuesto del globo terráqueo.
 - ATM: además del factor de distancia de $5 \mu\text{s}/\text{km}$, los sistemas ATM presentan una latencia de conmutación debida al método de encaminamiento de los datos en paquetes. Cada nodo tiene una latencia de hasta 150 μs y puede haber hasta 40 nodos en un trayecto de transmisión largo. En la práctica, la latencia suele ser de 20-40 μs por nodo.

5.6 Magnitud del retardo en los códecs de vídeo

La codificación de vídeo de radiodifusión utiliza casi universalmente la compresión MPEG para la transmisión a larga distancia. El retardo de codificación de vídeo en MPEG depende del parámetro de grupo de imágenes que en la mayoría de los códecs puede fijarse entre 1 y 15 tramas. El retardo de codificación de vídeo no depende de la longitud global del grupo de imágenes sino de la longitud

del subgrupo que contiene la primera imagen (la trama I) y la primera trama P. El grupo de imágenes normalmente se fija a 15 para maximizar la compresión de datos. Con ello se obtiene una longitud del subgrupo inicial de 4 tramas con un mínimo retardo de codificación de 133 ms a 30 tps o 160 ms para 25 tps y un retardo de codificación típico de aproximadamente el doble de estos valores. Si se fijan unos grupos de imágenes más cortos puede reducirse el mínimo retardo hasta 33 ms y 40 ms, respectivamente, a costa de una menor compresión de los datos (mayor velocidad binaria) o de una calidad de imagen inferior a velocidad binaria constante.

5.7 Magnitud del retardo en los códecs de audio

Los códecs de audio pueden presentar características muy distintas dependiendo de la aplicación a la que se destinen. En la radiodifusión, los teléfonos fijos y móvil se utilizan frecuentemente para la componente del programa, especialmente en programas de noticias y programas de línea abierta a los oyentes. Los trenes de audio en Internet también se están convirtiendo en una parte cada vez más común de los servicios de radiodifusión, pero no se utilizan aún ampliamente para la componente del programa. El audio de los programas de alta calidad también puede codificarse para la transmisión con un códec de calidad de radiodifusión.

Retardos de transcodificación típicos para una gama de sistemas son:

- Telefonía móvil GSM [5]: Retardo de transcodificación recomendado: mínimo 20 ms y máximo 30 ms.
- Telefonía móvil AMDC [10]: Máximo retardo de codificación recomendado: 20 ms; máximo retardo de decodificación recomendado: 3 ms.
- Sistemas de telefonía por satélite: Existe un cierto número de sistemas en competencia incluidos el ICO, Iridium, AceS, AMSC-TMI e Inmarsat. Estos sistemas pueden clasificarse como de órbita terrestre baja (LEO) (ICO, Globalstar, Iridium y Teledesic) u órbita terrestre geostacionaria (GEO) (Inmarsat, Satphone, ASC, Thuraya, APMT, EAST). Un sistema GEO típico, Inmarsat Mini-M, utiliza una codificación de excitación multibanda avanzada (AMBE) a 4,8 kbit/s con retardos de codificación/decodificación comparables a los de los sistemas de telefonía móvil terrenal GSM y AMDC. Sin embargo, en el caso GEO el retardo de distancia tiende a hacer insignificantes los retardos de codificación/decodificación.
- Códec MPEG de estudio.

El audio para la radiodifusión puede estar englobado en el vídeo utilizando codificación MPEG-2 o puede enviarse en líneas RDSI/ATM separadas. Si está englobado en el vídeo, el audio estará sincronizado con la señal de vídeo. Si se envía por separado, el audio puede estar sin comprimir, produciendo un retardo de codificación/decodificación despreciable, o puede estar comprimido. Los algoritmos de compresión varían según los fabricantes de manera que se citan los retardos mínimo y típico.

Debido a la naturaleza de las señales de audio una mínima velocidad binaria para una determinada calidad de audio se desviaría un factor de 10 o más a lo largo del tiempo. Por regla general, los intervalos de tiempo con amplia demanda a la velocidad binaria local son muy breves y están rodeados por intervalos de tiempo con velocidad binaria muy baja. Por consiguiente, la mayoría de los esquemas de codificación de audio modernos realizan una cierta promediación de la velocidad binaria a lo largo del tiempo para proporcionar una velocidad binaria constante. Dependiendo de si la memoria tampón para realizar esta promediación se encuentra en el lado del codificador o del decodificador, el retardo de codificación o decodificación es más grande.

Existen varias opciones para la compresión de audio MPEG-1: Capa I, Capa II y Capa III. La Capa III también se conoce como MP3. MPEG-2 además de Capa I, II y III ofrece codificación de audio avanzada (AAC). En MPEG-4 se eligió AAC como códec básico para el audio natural y se amplió mediante nuevas herramientas y funcionalidades. En el contexto del retardo de ida y vuelta la versión de bajo retardo de la codificación AAC es la extensión más importante.

Mínimos retardos para MPEG-1 Capa I, II, III, según [6], [18]

Capa	Velocidad binaria objetivo (kbit/s)	Relación de compresión	Mínimo retardo teórico (ms)
AAC-LD	64	12:1	20
Capa I	192	4:1	19
Capa II	128	6:1	35
Capa II	64	12:1	59

Codificación	Velocidad binaria (kbit/s)	Velocidad de muestreo (kHz)	Retardo típico (ms)
AAC-LD estereofónica	128	48	60
AAC-LD monofónica	64s	48	50
AAC estereofónica	128	48	172
Capa III estereofónica	128	48	326
Capa II estereofónica	128	48	224
Capa II estereofónica	128	24 (semimodo)	398
G.722	64	48	10

Retardo de extremo a extremo (incluido el canal RDSI) para un códec de audio disponible en el comercio [18].

Códec	Velocidad binaria (kbit/s)	Retardo algorítmico con la reserva de bits puesta a cero (ms)	Carga de trabajo al 100%, transmisión por ráfaga (ms)	Carga de trabajo al 100%, transmisión continua (ms)	Carga de trabajo al 30%, transmisión por ráfagas (ms)	Carga de trabajo al 30%, transmisión continua (ms)
Capa 2	192	34	No disponible	No disponible	No disponible	No disponible
Capa 3	128	54	118	142	107	131
MPEG-4 AAC	96	55	82	211	63	192
MPEG-4 HE AAC	56	129	184	361	145	322
MPEG-4 AAC-LD	128	20	33	44	24	35

Retardo de extremo a extremo de los actuales esquemas de codificación de audio.

El audio MPEG codificado con vídeo MPEG tendrá generalmente el mismo retardo que el vídeo.

5.8 Simetría del circuito

No debe de suponerse automáticamente que los dos semibucles en un bucle de la componente del programa serán idénticos en cuanto al retardo de procesamiento, retardo de trayecto, retardo global o fugas. En la radiodifusión es habitual utilizar un circuito de alta calidad en la línea de los programas de radiodifusión en exteriores únicamente y un circuito de calidad inferior en la línea de retorno. También es habitual utilizar un retorno sólo de audio, incluso en los bucles de la componente de vídeo. Ambos factores afectarán el retardo de bucle global y la perturbación perceptual debida al retardo y al eco.

5.9 Prácticas operativas que causan eco

Si bien es inevitable cierto retardo en un circuito de audio, el eco normalmente es evitable. El eco puede ser causado por el diseño del sistema o por unas inadecuadas prácticas operativas. Aunque unas prácticas de funcionamiento adecuadas no garantizan automáticamente un bucle sin ecos, lo minimizarán para un diseño del sistema determinado.

La eficacia de una buena práctica operativa a la hora de controlar el eco depende en gran medida del eco inherente en un bucle. En un circuito de 2 hilos, el eco inherente viene determinado por el diseño del sistema. En un circuito de 4 hilos el eco causado por la fuga de audio de las líneas de retorno a las líneas de programa puede minimizarse:

- utilizando «mix-minus» o «clean feed» para limitar; y/o
- silenciando adecuadamente el limitador cuando los micrófonos están abiertos, mediante un sistema de bloqueo si se utiliza un botón de conversación manual o empleando un circuito silenciador activado por la voz; y/o
- utilizando una comprobación de casco cerrado o una comprobación de auricular para minimizar las fugas en los micrófonos abiertos.

5.10 Retardo entre el sonido y la imagen

El retardo entre el sonido y la imagen también ha sido bien estudiado y se han establecido normas [4] por algún tiempo. El principal problema perceptual con este tipo de retardo es la pérdida de sincronización de los labios en la conversación. La norma de la UIT especifica actualmente un retardo aceptable entre el sonido y la imagen de +25 ms (el sonido precede a la imagen) a -100 ms (la imagen precede al sonido) medidos en el elemento de selección de fuente de programa final (normalmente la sala de control principal de vídeo).

La resincronización por separado del sonido y la imagen con retardo variable es conceptualmente una tarea sencilla, aunque los equipos que lleven a cabo esta labor pueden ser muy costosos. Si la señal de sonido precede a la señal de imagen, puede utilizarse para corregir la centralización una sencilla unidad de retardo de audio de bajo coste. Si la señal de imagen precede a la señal de sonido, la imagen debe retrasarse: este equipo normalmente es más costoso aunque un dispositivo de esas características puede estar incorporado en un servidor de vídeo existente.

6 Efectos del retardo y el eco

6.1 Retardo de audio y retardo de vídeo

Los bucles de comunicaciones de vídeo ya existen desde hace algunos años en las videoconferencias y en la radiodifusión pero son una tecnología relativamente nueva comparada con los bucles de comunicaciones de audio. Por este motivo, el efecto subjetivo de un retardo importante en un bucle de comunicaciones de vídeo (> 1 trama) no ha sido tan bien estudiado como el efecto subjetivo de un retardo largo en un bucle de audio. Debido a ello hay relativamente poca

información sobre los efectos subjetivos del retardo en los bucles de vídeo en comparación con esa misma información referente a los bucles de audio. Por consiguiente, esta contribución se basa fundamentalmente en la información disponible sobre el efecto subjetivo del retardo y el eco en un bucle de audio. Hasta que se disponga de información más detallada sobre dichos efectos subjetivos del retardo en los bucles de vídeo, se supone que el retardo y el eco de audio son la causa fundamental de las perturbaciones, independientemente de si se trata del factor limitativo en el retardo global del bucle audio-vídeo.

6.2 Variables que afectan la perturbación

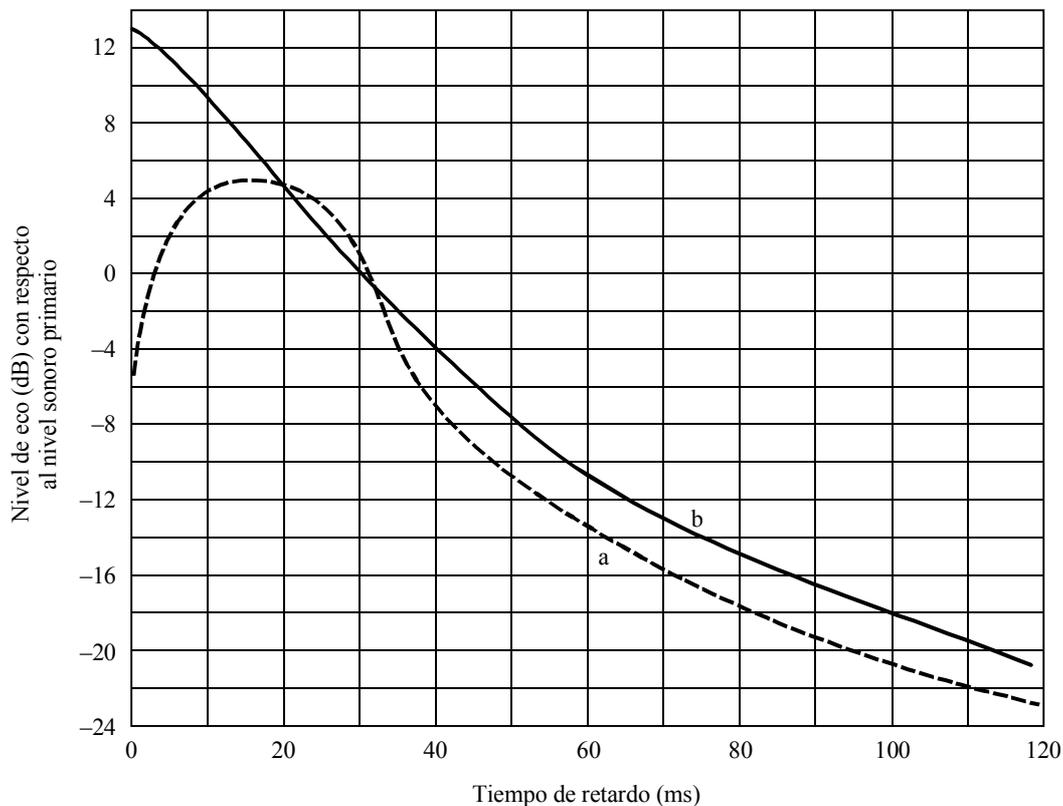
Los retardos de audio tienen un cierto número de efectos perturbadores sobre los sistemas de comunicaciones vocales. El efecto del retardo depende fundamentalmente de dos factores: la magnitud del retardo y las pérdidas de retorno en el bucle [1, 3].

6.3 Ecos y fusión

Para retardos por debajo del umbral en la Fig. 1, el sistema de audición humano no puede distinguir entre dos llegadas en la mayoría de los tipos de señales vocales y señales musicales. Por ello, los dos sonidos parecen «fundirse» y se convierten en un sonido de más volumen. Esto tiene un efecto positivo tanto en la sonoridad percibida como en la inteligibilidad de la señal vocal. Este efecto de fusión es progresivamente más débil a medida que aumenta el retardo, especialmente en el caso de retardos superiores a 30 ms. Cuando no se produce la fusión, el eco es un efecto perturbador.

FIGURA 1

Audibilidad de la señal retardada en presencia de señal directa, en función del retardo y el nivel relativo (extraído de [16])



Comparación de los umbrales de las curvas de percepción.

Curvas a Meyer y Schodder (nivel sonoro primario 55 fones)

b Lochner y Burger (nivel sonoro primario 50 fones)

6.4 Tipos de perturbación y sus efectos

Existen principalmente tres tipos de efectos perturbadores derivados del retardo y el eco:

- Retardo largo sin eco de retorno significativo. Este tipo de retardo puede clasificarse como retardo de semibucle. Puede provocar dificultades en las conversaciones normales haciendo que los interlocutores hablen simultáneamente e interrumpan sus frases cuando se aperciban de esta circunstancia. Ello hace que la conversación se prolongue y sea entrecortada lo que impide la concentración. Los retardos en los circuitos de satélite (unos ~240-280 ms para cada enlace de satélite) son la causa común de este tipo de perturbación.
- Ecos que vuelven al orador. Este tipo de retardo puede clasificarse como un múltiplo entero (1, 2, ... n) de un retardo de bucle completo. Dichos ecos interfieren con el mecanismo normal de realimentación de la palabra, del cerebro a la boca, de allí al oído y del oído al cerebro, causando tartamudeo y dudas. Tanto la duración del retardo como el nivel del eco tienen un efecto significativo sobre la perturbación que aumenta con el retardo a un nivel constante y disminuye con el nivel para un retardo constante. Para retardos grandes (> 240 ms) unas pérdidas de retorno tan bajas como -50 dB pueden causar cierta perturbación.
- Ecos que vuelven al oyente. Pueden clasificarse como un retardo de bucle $1\ 1/2$, $2\ 1/2$, ... etc. Estos ecos perturban al oyente disminuyendo la inteligibilidad de la palabra. Al igual que en el retardo de bucle completo, la perturbación aumenta con la duración del retardo. Se trata de un problema común en los sistemas públicos donde puede ser un eco sencillo en el caso de una sola reflexión o un eco en reverberación en el caso de múltiples reflexiones. La audibilidad de este tipo de retardo se muestra en la Fig. 1, extraída de [16]. Los ecos deben estar por debajo de esta curva para que sean inaudibles como sonidos distintos.

Los efectos del retardo del eco y del nivel del eco en la inteligibilidad de la palabra y la dificultad en la palabra han sido estudiados ampliamente por los organismos de telecomunicaciones [1, 3, 14] y son relativamente bien comprendidos. Sin embargo, los efectos sobre la inteligibilidad de la palabra probablemente son mejor comprendidos que los efectos sobre la dificultad de la palabra.

6.5 Adaptación

La tolerancia aprendida o adquirida a los efectos de un estímulo o irritante se conoce como adaptación. Aunque puede adquirirse cierta adaptación a las señales afectadas por ecos, es generalmente bastante limitada. También puede llevar algún tiempo conseguir la adaptación. La cantidad de adaptación que puede aprenderse y el tiempo necesario para aprenderla probablemente varían considerablemente según las personas.

Como la mayoría del contenido de radiodifusión, especialmente en los programas en directo, lo proporcionan los invitados y no los presentadores, deben formularse directrices generales basándose en la hipótesis de que no se producirá la adaptación en el tiempo disponible.

7 Resumen

Los sistemas de comunicaciones en bucle se utilizan normalmente en la producción de radiodifusión donde se emplean múltiples emplazamientos o emplazamientos remotos, particularmente en la producción de programas en directo. El bucle puede utilizarse para conectar un entrevistador que se encuentra en el estudio con un invitado a distancia o puede emplearse para conectar el estudio con un presentador que se encuentra en un emplazamiento distante. Idealmente, cada tren de audio y vídeo separado debe desplazarse únicamente en un sentido en este bucle. En la

práctica, es bastante común que el audio retorne a través del bucle hasta su fuente, especialmente cuando los teléfonos forman parte del circuito. Debido a la distancia y/o la necesidad de realizar los procesos de codificación/decodificación para comprimir la señal digitalizada a fin de transmitirla y restaurarla nuevamente en recepción, la señal puede sufrir un retardo significativo.

El retardo y la fuga se combinan para provocar el eco. Tanto el retardo como el eco pueden dar lugar a perturbaciones que afectan negativamente tanto a los participantes como a la audiencia del programa. Existen fundamentalmente tres tipos de efectos perturbadores a causa del retardo y el eco:

- retardo largo sin eco de retorno significativo;
- ecos que vuelven al orador;
- ecos que vuelven al oyente.

El efecto del eco sobre la inteligibilidad de la palabra es relativamente bien comprendido. El efecto del retardo sobre el orador y sobre el oyente no es tan bien comprendido. Por consiguiente, se sugieren aquí algunas directrices basadas en la audibilidad y en los efectos sobre la inteligibilidad de la palabra.

Las molestias causadas por el retardo y los ecos del sonido son familiares en las grandes salas y en los estadios deportivos y, desde luego, no es un problema nuevo. En las comunicaciones electrónicas, el problema apareció por primera vez en la telefonía a 2 hilos y la mayoría de los conocimientos que se tiene sobre las causas y efectos del retardo en la radiodifusión se derivan de las experiencias obtenidas en esta área. Es difícil controlar la fuga entre las mitades de un bucle de audio en sistemas de 2 hilos y es una práctica establecida desde hace largo tiempo permitir deliberadamente cierta fuga controlada en estos sistemas para que la fuga incontrolable sea menos apreciable.

Existen dos causas principales del retardo: retardo de propagación y retardo de codificación/decodificación. El retardo de propagación puede dividirse en dos factores: retardo de distancia y retardo de conmutación o latencia.

Los retardos de propagación debido a la distancia son del orden de $5 \mu\text{s}/\text{km}$ de longitud de trayecto en un guíaondas, cable o fibra óptica y del orden de $3,3 \mu\text{s}/\text{km}$ en las señales radiadas. Una señal que se desplace con propagación terrenal al lado opuesto del globo terráqueo presenta un retardo de distancia de aproximadamente 100 ms. Por satélite, el retardo de distancia para el mismo destino es de unos 280 ms, suponiendo que sólo interviene un satélite geoestacionario en la retransmisión.

La necesidad de sincronización audio-vídeo puede hacer que el retardo de vídeo sea el factor limitativo en un bucle de comunicaciones audio-vídeo. La codificación de vídeo se basa en un grupo de imágenes en vez de tramas individuales. El tamaño de grupo de imágenes de 15 imágenes es el que más habitualmente se utiliza y proporciona la comprensión de datos más elevada. Ello da lugar a un mínimo retardo de cuatro tramas. El retardo de codificación total, incluido el tiempo de procesamiento, puede ser considerablemente mayor que el mínimo número de tramas intermedias y normalmente es de al menos 7 u 8 tramas (230-320 ms) para esta configuración.

Los retardos de transcodificación de audio en la mayoría de los sistemas de telefonía móvil tales como GSM, AMDC y en los sistemas de telefonía por satélite son bajos (3-30 ms). Para codificación de audio de radiodifusión de mayor calidad, tal como MPEG AAC, pueden ser del mismo orden que los retardos de codificación de vídeo (170-400 ms). Sistemas más recientes tales como el MPEG AAC-LD han reducido este retardo a unos 60 ms.

Los sistemas ATM tienen una latencia de conmutación debido al método de encaminamiento de los datos en paquetes. Cada nodo presenta una latencia de hasta $150 \mu\text{s}$ y puede haber hasta 40 nodos en un trayecto de transmisión de gran longitud. En la práctica, la latencia normalmente es de $20\text{-}40 \mu\text{s}$ por nodo, dando lugar a un retardo relativamente pequeño en torno a 1,6 ms.

El eco puede minimizarse realizando un buen diseño del sistema y llevando a cabo prácticas operativas adecuadas.

A continuación se indican tres principios generales que deben observarse en el diseño y configuración de las instalaciones para la componente del programa:

- minimizar el número de etapas de codificación/decodificación;
- mantener el retardo audio-vídeo diferencial entre los límites de +25 ms y –100 ms;
- mantener el límite de fuga de repliegue por debajo del umbral de audibilidad.

Las prácticas operativas que minimizarán el eco incluyen:

- utilizar «mix-minus» o «clean feed» para limitar; y/o
 - utilizar un silenciamiento adecuado en el limitador cuando los micrófonos están abiertos, mediante un bloqueo del sistema si se utiliza un botón de conversación manual o mediante un circuito de silenciamiento activado por voz; y/o
 - utilizar una comprobación de casco cerrado o una comprobación de auricular para minimizar las fugas en los micrófonos abiertos.
-