

Unión Internacional de Telecomunicaciones

UIT-R

Sector de Radiocomunicaciones de la UIT

Recomendación UIT-R BS.1196-8
(10/2019)

**Codificación de audio para
la radiodifusión digital**

Serie BS
Servicio de radiodifusión
(sonora)



Unión
Internacional de
Telecomunicaciones

Prólogo

El Sector de Radiocomunicaciones tiene como cometido garantizar la utilización racional, equitativa, eficaz y económica del espectro de frecuencias radioeléctricas por todos los servicios de radiocomunicaciones, incluidos los servicios por satélite, y realizar, sin limitación de gamas de frecuencias, estudios que sirvan de base para la adopción de las Recomendaciones UIT-R.

Las Conferencias Mundiales y Regionales de Radiocomunicaciones y las Asambleas de Radiocomunicaciones, con la colaboración de las Comisiones de Estudio, cumplen las funciones reglamentarias y políticas del Sector de Radiocomunicaciones.

Política sobre Derechos de Propiedad Intelectual (IPR)

La política del UIT-R sobre Derechos de Propiedad Intelectual se describe en la Política Común de Patentes UIT-T/UIT-R/ISO/CEI a la que se hace referencia en la Resolución UIT-R 1. Los formularios que deben utilizarse en la declaración sobre patentes y utilización de patentes por los titulares de las mismas figuran en la dirección web <http://www.itu.int/ITU-R/go/patents/es>, donde también aparecen las Directrices para la implementación de la Política Común de Patentes UIT-T/UIT-R/ISO/CEI y la base de datos sobre información de patentes del UIT-R sobre este asunto.

Series de las Recomendaciones UIT-R

(También disponible en línea en <http://www.itu.int/publ/R-REC/es>)

Series	Título
BO	Distribución por satélite
BR	Registro para producción, archivo y reproducción; películas en televisión
BS	Servicio de radiodifusión (sonora)
BT	Servicio de radiodifusión (televisión)
F	Servicio fijo
M	Servicios móviles, de radiodeterminación, de aficionados y otros servicios por satélite conexos
P	Propagación de las ondas radioeléctricas
RA	Radio astronomía
RS	Sistemas de detección a distancia
S	Servicio fijo por satélite
SA	Aplicaciones espaciales y meteorología
SF	Compartición de frecuencias y coordinación entre los sistemas del servicio fijo por satélite y del servicio fijo
SM	Gestión del espectro
SNG	Periodismo electrónico por satélite
TF	Emisiones de frecuencias patrón y señales horarias
V	Vocabulario y cuestiones afines

Nota: Esta Recomendación UIT-R fue aprobada en inglés conforme al procedimiento detallado en la Resolución UIT-R 1.

Publicación electrónica
Ginebra, 2020

© UIT 2020

Reservados todos los derechos. Ninguna parte de esta publicación puede reproducirse por ningún procedimiento sin previa autorización escrita por parte de la UIT.

RECOMENDACIÓN UIT-R BS.1196-8*

Codificación de audio para la radiodifusión digital

(Cuestión UIT-R 19-1/6)

(1995-2001-2010-2012-02/2015-10/2015-2017-01/2019-10/2019)

Cometido

Esta Recomendación especifica los sistemas de codificaciones de la fuente de audio aplicables a la radiodifusión sonora y de televisión digital. Específica, además, un sistema aplicable a la mejora multicanal con compatibilidad hacia atrás de los sistemas de radiodifusión sonora y de televisión digital.

Palabras clave

Audio, códec, codificación audio, digital, radiodifundir, radiodifusión, sonido, televisión

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que los requisitos de usuario para los sistemas de codificación de audio para la radiodifusión digital se especifican en la Recomendación UIT-R BS.1548;
- b) que los sistemas de sonido multicanal con y sin acompañamiento de imagen son objeto de la Recomendación UIT-R BS.775 y que un sistema de sonido multicanal de alta calidad que utilice una reducción de la velocidad binaria eficaz es esencial en un sistema de radiodifusión digital;
- c) que el sistema de sonido avanzado especificado en la Recomendación UIT-R BS.2051 consiste en configuraciones de canal tridimensionales y utiliza metadatos estáticos o dinámicos para controlar señales basadas en objetos, en escenas o en canales;
- d) que la evaluación subjetiva de los sistemas de audio con pequeñas degradaciones, incluidos los sistemas de sonido multicanal, es objeto de la Recomendación UIT-R BS.1116;
- e) que la evaluación subjetiva de los sistemas de audio de calidad de audio intermedia se trata en la Recomendación UIT-R BS.1534 (MUSHRA);
- f) que la codificación a baja velocidad binaria para el audio de alta calidad ha sido probada por el Sector de Radiocomunicaciones de la UIT;
- g) que los elementos comunes en los métodos de codificación de fuente de audio entre los diferentes servicios pueden proporcionar una mayor flexibilidad al sistema y lograr un menor coste del receptor;
- h) que numerosos servicios de radiodifusión ya utilizan o han especificado la utilización de los códecs de audio de las familias MPEG-1, MPEG-2, MPEG-4, AC-3 y E-AC-3;
- i) que la Recomendación UIT-R BS.1548 indica los códecs que han demostrado que satisfacen los requisitos del organismo de radiodifusión para la contribución, la distribución y la emisión;

* Esta Recomendación debe señalarse a la atención de la Organización Internacional de Normalización (ISO) y de la Comisión Electrotécnica Internacional (CEI).

- j) que los organismos de radiodifusión que no han iniciado aún los servicios deben poder elegir el sistema más adecuado para su aplicación;
- k) que puede que los organismos de radiodifusión necesiten considerar la compatibilidad con los sistemas y equipos de radiodifusión establecidos cuando elijan un sistema;
- l) que al introducir un sistema de sonido multicanal, deben considerarse los receptores monofónicos y estereofónicos existentes;
- m) que una extensión multicanal compatible hacia atrás a un sistema de codificación de audio existente puede proporcionar una mayor eficacia de la velocidad binaria que la difusión simultánea («simulcast»);
- n) que, preferiblemente, un sistema de codificación de audio debería poder codificar tanto la voz como la música con la misma alta fidelidad,

recomienda

1 que para las nuevas aplicaciones de emisión de radiodifusión sonora y de televisión digital, en las que no es necesaria la compatibilidad con las transmisiones y equipos establecidos, se utilice uno de los siguientes sistemas de codificación de audio de baja velocidad binaria:

- HE AAC ampliado como especifica la Norma ISO/CEI 23003-3:2012;
- E-AC-3 como especifica la Norma ETSI TS 102 366 (2014-08);
- AC-4 como especifican las Normas ETSI TS 103 190-1 v1.3.1 (2018-02) y ETSI TS 103 190-2 v1.2.1 (2018-02);
- perfil de baja complejidad de audio MPEG-H 3D como especifica la Norma ISO/CEI 23008-3:2019;
- DTS-UHD como especifica la norma ETSI TS 103 491 v.1.1.1 (2017-04).

NOTA 1 – HE AAC ampliado es un superconjunto más flexible de MPEG-4 HE AAC v2, HE AAC y AAC LC, e incluye codificación unificada de la voz y el sonido (USAC) MPEG-D.

NOTA 2 – E-AC-3 es un superconjunto más flexible de AC-3.

NOTA 3 – Las especificaciones relacionadas con AC-4, el perfil de baja complejidad de audio MPEG-H 3D y DTS-UHD incluyen capacidades aptas para dar soporte al sistema de sonido avanzado que figura en la Recomendación UIT-R BS.2051. A fin de obtener información relacionada con el nivel de cumplimiento del códec, los usuarios pueden consultar la Recomendación UIT-R BS.1548;

2 que para aplicaciones de emisión de radiodifusión sonora y de televisión digital, en las que no es necesaria la compatibilidad con las transmisiones y los equipos establecidos, se utilice uno de los siguientes sistemas de codificación de baja velocidad binaria:

- MPEG-1 Capa II como especifica la Norma ISO/CEI 11172-3:1993.
- MPEG-2 Capa II con velocidad de muestreo mitad, como especifica la Norma ISO/CEI 13818-3:1998.
- MPEG-2 AAC-LC o MPEG-2 AAC-LC con SBR como especifica la Norma ISO/CEI 13818-7:2006.
- MPEG-4 AAC-LC como especifica la Norma ISO/CEI 14496-3:2009.
- MPEG-4 HE AAC v2 como especifica la Norma ISO/CEI 14496-3:2009.
- AC-3 como especifica la Norma ETSI TS 102 366 (2014-08).

NOTA 4 – La Norma ISO/CEI 11172-3 puede citarse a veces como Norma 13818-3 pues esta especificación incluye la Norma 11172-3 por referencia.

NOTA 5 – Se alienta a apoyar HE AAC ampliado, como especifica la Norma ISO/CEI 23003-3:2012. Incluye todas las versiones AAC anteriormente mencionadas, garantizando de este modo la compatibilidad tanto con los futuros nuevos sistemas de radiodifusión como con los establecidos en todo el mundo, con la misma aplicación de decodificador simple.

3 que para una extensión multicanal compatible hacia atrás de los sistemas de radiodifusión sonora y de televisión digital, se empleen las extensiones de audio multicanal descritas en la Norma ISO/CEI 23003-1:2007;

NOTA 6 – Como la tecnología MPEG Surround descrita en la Norma ISO/CEI 23003-1:2007 es independiente de la tecnología de compresión (codificador básico) utilizada para la transmisión de la señal compatible hacia atrás, las herramientas de mejora multicanal descritas pueden utilizarse en combinación con cualquiera de los sistemas de codificación indicados en los *recomienda* 1 y 2.

4 que para los enlaces de distribución y contribución, se utilice una de las codificaciones siguientes con una velocidad binaria mínima de entre las siguientes velocidades binarias por cada señal de audio (es decir, por cada señal monofónica o por cada componente de una señal estereofónica con codificación independiente) excluyendo los datos auxiliares:

- MPEG-1 Capa II, según se define en ISO/CEI 11172-3, con una velocidad binaria de al menos 180 kbit/s por cada señal monofónica;
- MPEG-4 AAC, según se define en ISO/CEI 14496-3, con una velocidad binaria de al menos 144 kbit/s por cada señal monofónica en caso de hasta cinco cascadas;
- AC-4 según se define en ETSI TS 103 190-1 v1.3.1 (2018-02) y ETSI TS 103 190-2 v1.2.1 (2018-02), con una velocidad binaria de al menos 128 kbit/s por cada señal monofónica en caso de hasta cinco cascadas;
- MPEG-H 3D Audio, según se define en ISO/CEI 23008-3, con una velocidad binaria de al menos 144 kbit/s por cada señal monofónica en caso de hasta cinco cascadas;

5 que para los enlaces de comentarios, se utilice la codificación ISO/CEI 11172-3 Capa III a una velocidad binaria de al menos 60 kbit/s excluyendo los datos auxiliares para las señales monofónicas y de al menos 120 kbits/s excluyendo los datos auxiliares para las señales estereofónicas, empleando codificación estereofónica conjunta;

6 que para las aplicaciones de alta calidad la frecuencia de muestreo sea de 48 kHz;

7 que la señal de entrada al codificador de baja velocidad binaria no lleve acentuación ni aplique acentuación el codificador,

recomienda además

1 que se haga referencia a la versión más reciente de la Recomendación UIT-R BS.1548 para obtener información sobre las configuraciones del sistema de codificación que han demostrado satisfacer la calidad y otros requisitos de usuario para la contribución, la distribución y la emisión;

2 que se realicen estudios adicionales sobre los requisitos para el sistema de sonido avanzado especificado en la versión más reciente de la Recomendación UIT-R BS.2051 y que se actualice esta Recomendación una vez concluidos los estudios.

NOTA – La información sobre los códecs incluida en esta Recomendación figura en los Anexos 1 a 8.

Anexo 1 (informativo)

Audio MPEG-1 y MPEG-2, Capa II y Capa III

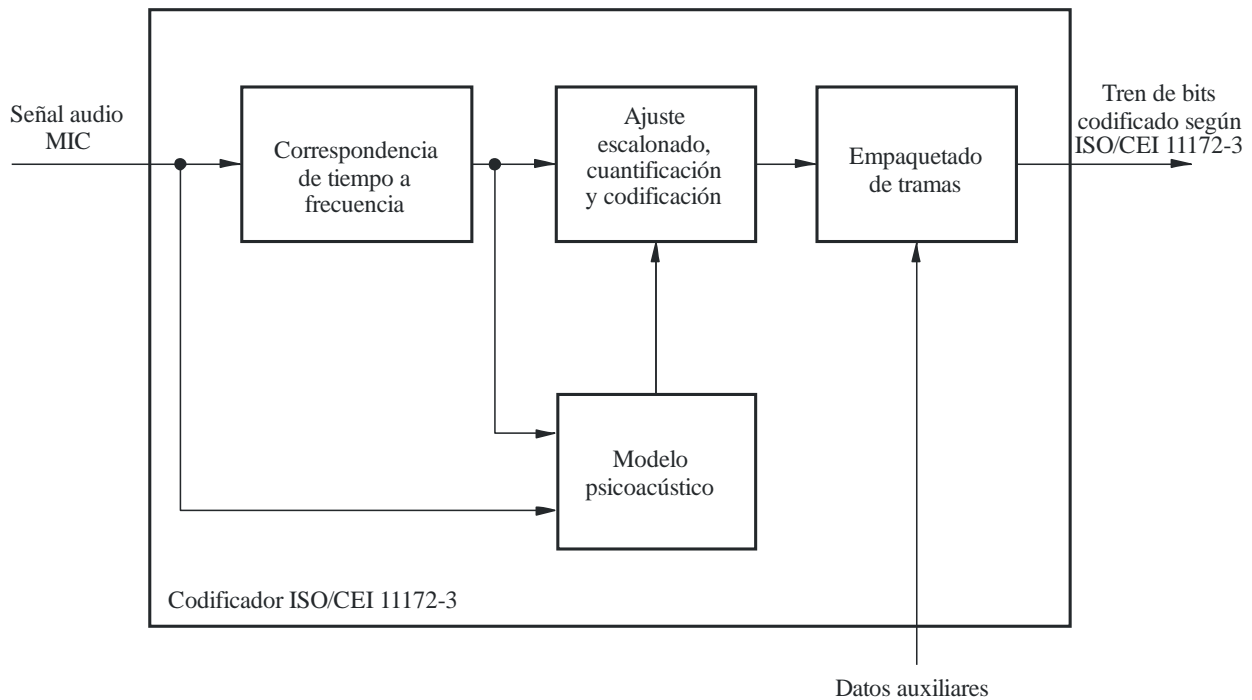
1 Codificación

El codificador procesa la señal de sonido digital y produce el tren binario comprimido. El algoritmo del codificador no está normalizado y puede utilizar diversos medios de codificación tales como el de estimación de umbral de enmascaramiento del auditorio, la cuantificación y el ajuste escalonado (*scaling*) (véase la Nota 1). No obstante, la salida del codificador debe ser tal que un decodificador conforme a esta Recomendación produzca una señal de audio adecuada para la aplicación prevista.

NOTA 1 – Un codificador que cumpla la descripción de los Anexos C y D de la Norma ISO/CEI 11172-3, 1993, dará un nivel mínimo de calidad satisfactorio.

La descripción que sigue es la de un codificador típico como el indicado en la Fig. 1. Las muestras de audio de entrada se aplican al codificador. La correspondencia entre tiempo y frecuencia crea una representación filtrada y submuestreada del tren de audio de entrada. Las muestras sometidas a la correspondencia pueden denominarse muestras de subbandas (como en las Capas I o II, según se indica más adelante) o muestras de subbandas transformadas (como en la Capa III). Un modelo psicoacústico que utiliza una transformada rápida de Fourier, en paralelo con la correspondencia entre tiempo y frecuencia de la señal de audio crea un grupo de datos para controlar la cuantificación y la codificación. Estos datos difieren según la realización concreta del codificador. Una posibilidad consiste en utilizar una estimación del umbral de enmascaramiento para controlar el cuantificador. El bloque de ajuste escalonado, cuantificación y codificación crea un conjunto de símbolos de codificación a partir de las muestras de entrada con correspondencia. También en este caso, la función de transferencia de este bloque puede depender del sistema de codificación. El bloque «empaquetamiento de tramas» ensambla el tren binario real a partir de los datos de salida de los otros bloques (por ejemplo, datos de asignación de bits, factores de ajuste, muestras de subbanda codificadas), y añade, si es necesario, otra información en el campo de datos auxiliares (por ejemplo, la protección contra errores).

FIGURA 1
Esquema de bloques de un codificador típico



BS.1196-01

2 Capas

De acuerdo con la aplicación, pueden utilizarse distintas capas del sistema de codificación, con complejidad y calidad crecientes del codificador.

Capa I: Esta capa contiene la correspondencia básica de la entrada audio digital a una segmentación fija de 32 subbandas para formatear los datos en bloques, un modelo psicoacústico para determinar la asignación de bits adaptativa, y una cuantificación con compresión-expansión y formateado de bloques. Una trama de la Capa I representa 384 muestras por canal.

Capa II: Esta capa ofrece una codificación adicional para la adjudicación de bits, los factores de ajuste y las muestras. Una trama de Capa II representa $3 \times 384 = 1\,152$ muestras por canal.

Capa III: Esta capa introduce un aumento de la resolución de frecuencia basado en un banco de filtros híbridos (banco de filtros de 32 subbandas con transformada de coseno discreto modificada de longitud variable). Añade un cuantificador no uniforme, una segmentación adaptativa y una codificación de entropía de los valores cuantificados. Una trama de Capa III representa 1 152 muestras por canal.

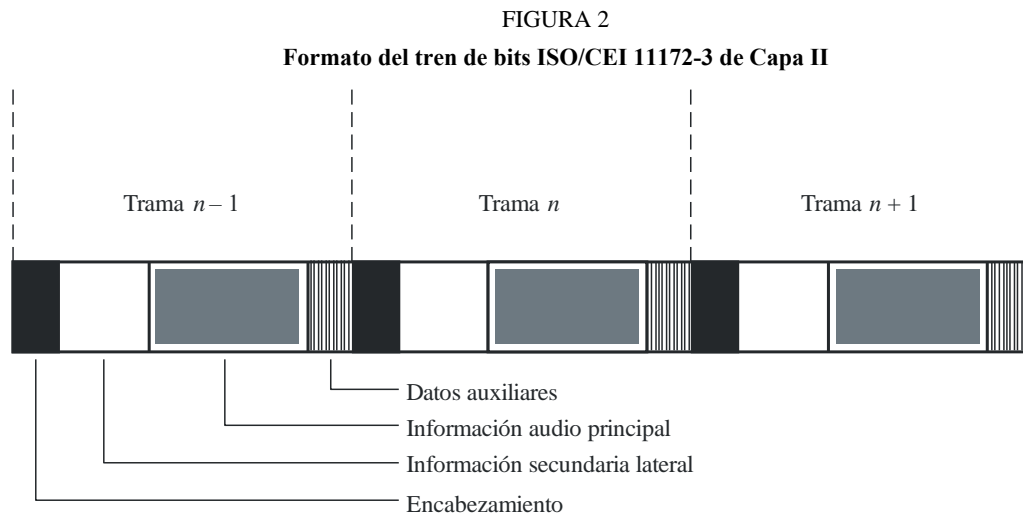
Hay cuatro modos distintos posibles para cualquier capa:

- de canal simple;
- de canal doble (dos señales de audio independientes codificadas en un tren binario para, por ejemplo, aplicaciones bilingües;
- de estereofonía (señales izquierda y derecha de un par estereofónico codificado en un tren binario); y

- de estereofonía mixta (señales izquierda y derecha de un par estereofónico codificadas en un tren binario, aprovechando la irrelevancia y la redundancia estereofónicas). El modo estereofónico puede servir para mejorar la calidad de audio a baja velocidad binaria y/o reducir la velocidad binaria de las señales estereofónicas.

3 Formato del tren de bits codificado

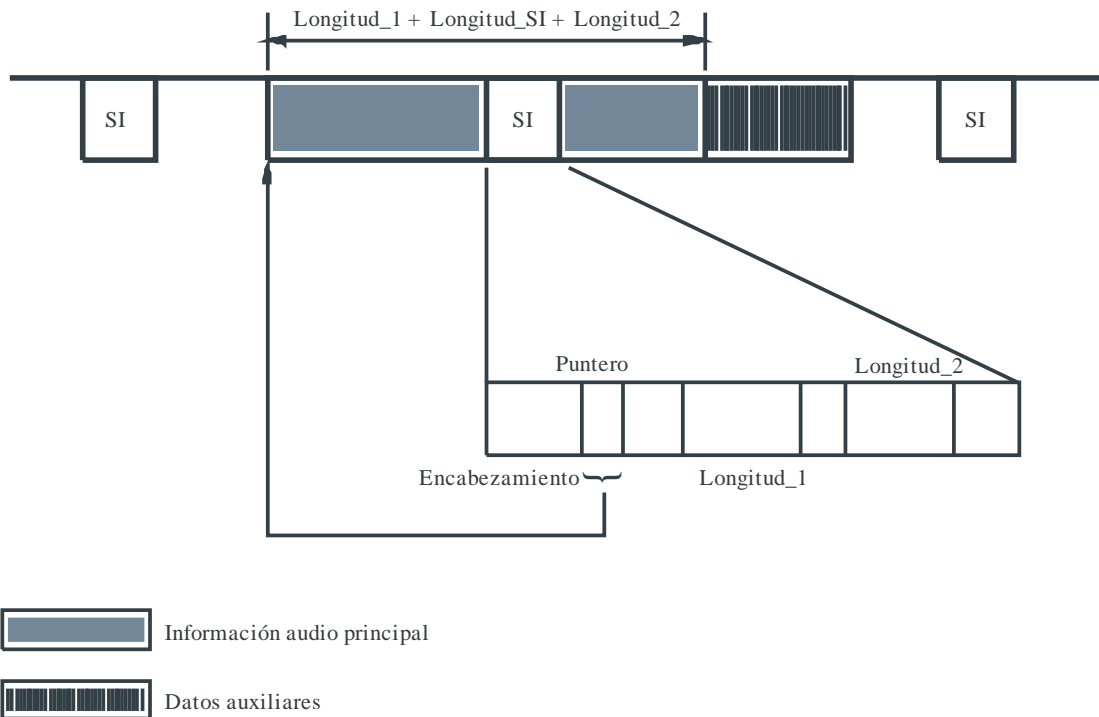
La Fig. 2 ofrece una panorámica del tren de bits ISO/CEI 11172-3 para la Capa II, y la Fig. 3, para la Capa III. Un tren de bits codificado se compone de tramas consecutivas. Según la capa, una trama contiene los campos siguientes:



Capa II:

Encabezamiento:	parte del tren de bits que contiene información de sincronismo y estado
Información secundaria lateral:	parte del tren de bits que contiene información sobre la adjudicación de bits y el factor de ajuste
Información audio principal:	parte del tren de bits que contiene muestras codificadas en subbanda
Datos auxiliares:	parte del tren de bits que contiene datos definibles por el usuario

FIGURA 3
Formato del tren de bits de la Capa III ISO/CEI 11172-3



Capa III:

Información secundaria lateral (SI):	parte del tren de bits que contiene el encabezamiento, el puntero, la longitud_1 y la longitud_2, así como información del factor de escala, etc.
Encabezamiento:	parte del tren de bits que contiene información de sincronismo y estado
Puntero:	indica el comienzo de la información audio principal
Longitud_1:	longitud de la primera parte de la información audio principal
Longitud_2:	longitud de la segunda parte de la información audio principal
Información audio principal:	parte del tren de bits que contiene la señal de audio codificada
Datos auxiliares:	parte del tren de bits que contiene datos definibles por el usuario

BS.1196-03

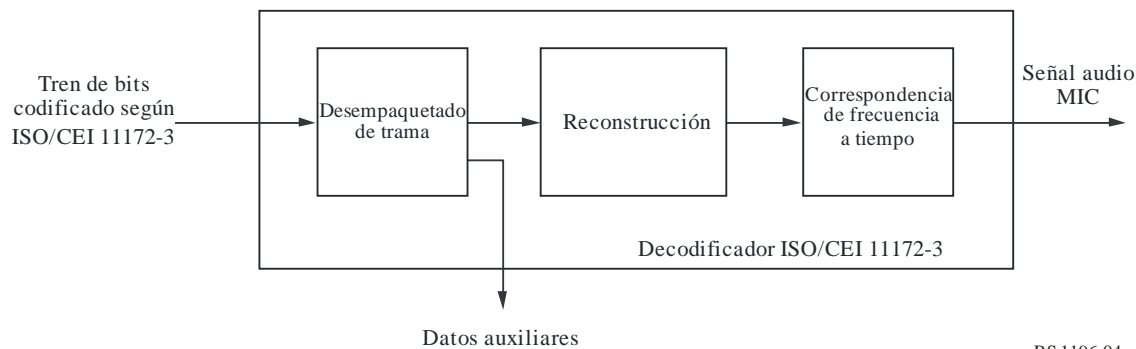
4 Decodificación

El decodificador acepta el tren binario de audio comprimido en la sintaxis que define la ISO/CEI 11172-3, decodifica los elementos de datos y utiliza la información para producir una salida de audio digital.

El tren binario audio codificado se aplica al decodificador. En el proceso de desempaquetado y de decodificación del tren binario se efectúa facultativamente una detección de errores si se aplica una verificación de errores en el codificador. Los datos del tren de bits se desempaquetan para recuperar los diversos elementos de información, como por ejemplo el encabezamiento de trama audio, la adjudicación de bits, los factores de ajuste, las muestras con correspondencia y, opcionalmente, los datos auxiliares. El proceso de reconstrucción recompone la versión cuantificada del conjunto de muestras con correspondencia. La correspondencia de frecuencias a tiempo vuelve a convertir estas muestras con correspondencia en muestras de audio MIC lineales.

FIGURA 4

Diagrama de bloques del decodificador



BS.1196-04

Anexo 2 (informativo)

Audio MPEG-2 y MPEG-4 AAC

1 Introducción

En ISO/CEI 13818-7 se describen las Normas MPEG-2 de audio no compatibles hacia atrás denominadas codificación de audio avanzada (AAC, *advanced audio coding*) de MPEG-2, una norma multicanal de calidad mayor que la que se podría lograr con un desarrollo que garantizara la compatibilidad hacia atrás con MPEG-1.

El sistema AAC está compuesto por tres perfiles que permiten el balance entre la memoria y potencia de procesamiento requeridas y la calidad del audio:

– *Perfil principal*

El perfil principal ofrece la mejor calidad de audio para cualquier velocidad de transmisión de datos en concreto. Salvo por el control de ganancia, se pueden emplear todas las herramientas con el fin de lograr una alta calidad de audio. La memoria y potencia de procesamiento que se requieren son mayores que las del perfil de baja complejidad (LC). Un decodificador de perfil principal puede decodificar un tren de bits codificado con un perfil LC.

– *Perfil de baja complejidad (LC, low complexity)*

Los requerimientos en cuanto a potencia de procesamiento y memoria del perfil LC son menores que los del perfil principal, aunque se mantiene una alta característica de calidad. El perfil LC no emplea predictor ni una herramienta de control de ganancia, pero tiene limitaciones en cuanto al orden de conformación de ruido temporal (TNS, *temporal noise shaping*).

– *Perfil de velocidad de muestreo escalable (SSR, scalable sampling rate)*

El perfil SSR puede ofrecer una señal escalable en frecuencia empleando la herramienta de control de ganancia. Puede escoger las bandas de frecuencia a decodificar, por lo que el decodificador necesita menos hardware. Por ejemplo, al decodificar sólo la banda de frecuencias más baja a la frecuencia de muestreo de 48 kHz, el decodificador puede reproducir una señal de audio de una anchura de banda de 6 kHz con una mínima complejidad en la decodificación.

El sistema AAC funciona con 12 tipos de frecuencias de muestreo que van desde 8 kHz hasta 96 kHz, tal y como se muestra en el Cuadro 1, y hasta con 48 canales de audio. En el Cuadro 2 se muestran las configuraciones de canal por defecto que incluyen la monocanal, la bicanal, la pentacanal (tres canales al frente y dos atrás) y la pentacanal con canal de efectos de baja frecuencia (LFE, *low frequency effect*) (con anchura de banda < 200 Hz) etc. Adicionalmente a las configuraciones por defecto, es posible especificar el número de altavoces en cada posición (al frente, a los lados, y atrás), lo que permite una disposición flexible de los altavoces multicanal. También se tiene capacidad para la separación de canales. El usuario puede definir un coeficiente para separar señales de audio multicanal en señales bicanal. Se puede, por lo tanto, controlar la calidad del sonido mediante el uso de un dispositivo de reproducción de sólo dos canales.

CUADRO 1

Frecuencias de muestreo permitidas

Frecuencia de muestreo (Hz)
96 000
88 200
64 000
48 000
44 100
32 000
24 000
22 050
16 000
12 000
11 025
8 000

CUADRO 2

Configuraciones del canal por defecto⁽¹⁾

Valor⁽²⁾	Nº de altavoces	Elementos audiosintácticos, en el orden en que se reciben	Correspondencia entre el elemento por defecto y el altavoz⁽³⁾	Nombre del canal especificado en las Recomendaciones UIT-R BS.775 o BS.2051⁽⁴⁾
1	1	single_channel_element	M+000	Mono
2	2	channel_pair_element	M+030, M-030	Izquierdo, derecho
3	3	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Izquierdo, derecho
4	4	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Izquierdo, derecho
		single_channel_element()	M+180	Mono envolvente
5	5	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Izquierdo, derecho
		channel_pair_element()	M+110, M-110	Izquierdo envolvente, derecho envolvente

CUADRO 2 (continuación)

Valor ⁽²⁾	Nº de altavoces	Elementos audiosintácticos, en el orden en que se reciben	Correspondencia entre el elemento por defecto y el altavoz ⁽³⁾	Nombre del canal especificado en las Recomendaciones UIT-R BS.775 o BS.2051 ⁽⁴⁾
6	5 + 1	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Izquierdo, derecho
		channel_pair_element()	M+110, M-110	Izquierdo envolvente, derecho envolvente
		lfe_element()	LFE1	Efectos de baja frecuencia
7	7 + 1 Frontal	single_channel_element()	M+000	N/A *2
		channel_pair_element()	M+030, M-030	
		channel_pair_element()	M+045, M-045	
		channel_pair_element()	M+110, M-110	
		lfe_element()	LFE1	
8-10	–	–	Reservado	–
11	6 + 1	single_channel_element()	M+000	N/A
		channel_pair_element()	M+030, M-030	
		channel_pair_element()	M+110, M-110	
		single_channel_element()	M+180	
		lfe_element()	LFE1	
12	7 + 1 Posterior	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Izquierdo, derecho
		channel_pair_element()	M+090, M-090	Lateral izquierdo envolvente, lateral derecho envolvente
		channel_pair_element()	M+135, M-135	Posterior izquierdo envolvente, posterior derecho envolvente
		lfe_element()	LFE1	Efecto de baja frecuencia
13	22 + 2	single_channel_element()	M+000	Central frontal
		channel_pair_element()	M+030, M-030	Central izquierdo frontal, central derecho frontal
		channel_pair_element()	M+060, M-060	Izquierdo frontal, derecho frontal
		channel_pair_element()	M+090, M-090	Izquierdo lateral, derecho lateral
		channel_pair_element()	M+135, M-135	Izquierdo posterior, derecho posterior
		single_channel_element()	M+180	Central posterior
		lfe_element()	LFE1	Efectos de baja frecuencia-1
		lfe_element()	LFE2	Efectos de baja frecuencia-2
		single_channel_element()	U+000	Central frontal superior
		channel_pair_element()	U+045, U-045	Izquierdo frontal superior, derecho frontal superior
		channel_pair_element()	U+090, U-090	Izquierdo lateral superior, derecho lateral superior
		single_channel_element()	T+000	Central superior
		channel_pair_element()	U+135, U-135	Izquierdo posterior superior, derecho posterior superior
		single_channel_element()	U+180	Central posterior superior
single_channel_element()	B+000	Central frontal inferior		
channel_pair_element()	B+045, U-045	Izquierdo frontal inferior, derecho frontal inferior		

CUADRO 2 (*fin*)

Valor ⁽²⁾	Nº de altavoces	Elementos audiosintácticos, en el orden en que se reciben	Correspondencia entre el elemento por defecto y el altavoz ⁽³⁾	Nombre del canal especificado en las Recomendaciones UIT-R BS.775 o BS.2051 ⁽⁴⁾
14	7 + 1 superior	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Izquierdo, derecho
		channel_pair_element()	M+110, M-110	Izquierdo envolvente, derecho envolvente
		lfe_element()	LFE1	Efectos de baja frecuencia
		channel_pair_element()	U+030, U-030	Izquierdo frontal superior, derecho frontal superior
15	–	–	Reservado	–

⁽¹⁾ La lista se reproduce del Cuadro 1.19 de ISO/CEI 14496-3:2009/Amd.4:2013.

⁽²⁾ La configuración del canal audio de salida se indica mediante un campo de cuatro bits que contiene el valor de la configuración del canal como se define en ISO/CEI 23001-8:2013, «Puntos código independientes de la codificación». MPEG-2 es aplicable a los valores de la configuración de canal hasta. MPEG-4 AAC es aplicable a valores de configuración de canal hasta 15.

⁽³⁾ Identificación de los altavoces por etiquetas de conformidad con la Recomendación UIT-R BS.2051.

⁽⁴⁾ Téngase en cuenta que las etiquetas y nombres de los canales dependen de la configuración real del canal.

⁽⁵⁾ N/A: no aplicable; la configuración no está disponible en la Recomendación UIT-R BS.2051 ni en la Recomendación UIT-R BS.775.

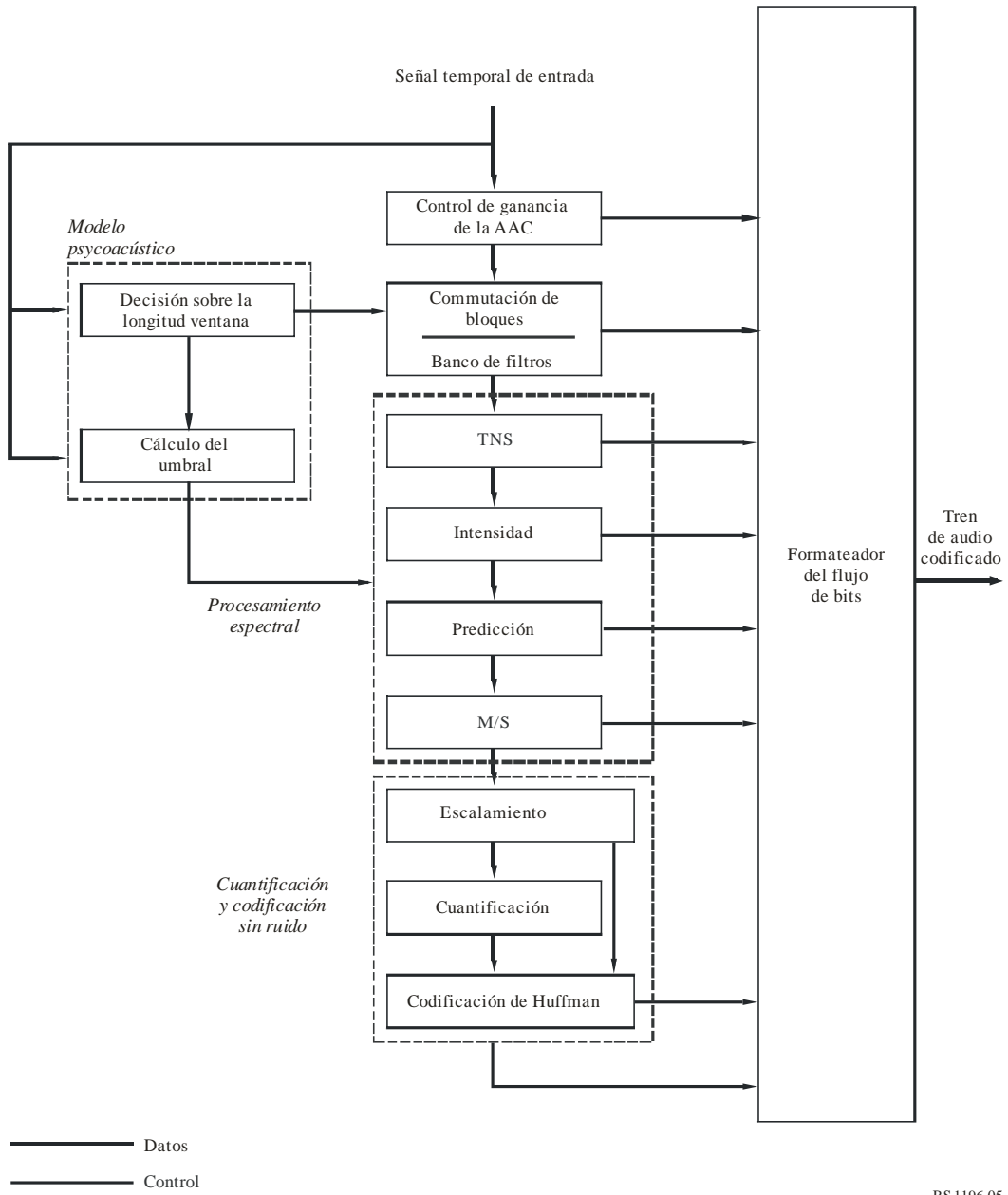
2 Codificación

La estructura básica del codificador MPEG-2 AAC se muestra en le Fig. 5. El sistema AAC consiste en las siguientes herramientas de codificación:

- Control de ganancia: El control de ganancia divide la señal de entrada en cuatro bandas de frecuencia espaciada de manera uniforme. El control de ganancia se emplea para el perfil SSR.
- Banco de filtros: Un banco de filtros (MDCT, *modified discrete cosine transform* – transformada de coseno discreta modificada) descompone la señal de entrada en subcomponentes espectrales muestreadas con una resolución de frecuencia de 23 Hz y una resolución en el tiempo de 21,3 ms (128 componentes espectrales) o con una resolución de frecuencia de 187 Hz y una resolución en el tiempo de 2,6 ms (1 024 componentes espectrales) a un muestreo de 48 kHz. La forma de la ventana se selecciona de entre dos formas de ventana alternativas.
- Conformación del ruido temporal (TNS): Tras el banco de filtros de análisis, se lleva a cabo la operación TNS (*temporal noise shaping*). La técnica de TNS le permite al codificador tener control sobre la estructura fina temporal del ruido de cuantificación.
- Codificación estereofónica media/lateral (M/S) y codificación estereofónica de la intensidad: En las señales de audio multicanal se puede emplear la codificación estereofónica de la intensidad y la codificación estereofónica M/S. En la codificación estereofónica de la intensidad sólo se transmite la envolvente de energía con el fin de disminuir la cantidad de información direccional emitida. En la codificación estereofónica M/S, en vez de las señales originales izquierda y derecha, se puede transmitir las señales de suma normalizada (M proviene de la palabra en inglés «middle», que significa media) y de diferencia (S proviene de la palabra en inglés «side», que significa lateral).

- Predicción: Para disminuir la redundancia en señales estacionarias, se lleva a cabo la predicción en el dominio del tiempo entre subcomponentes espectrales muestreadas de tramas contiguas.
- Cuantificación y codificación sin ruido: Se emplea un cuantificador no uniforme en la herramienta de cuantificación con tamaños de paso de 1,5 dB. Se aplica codificación de Huffman en el espectro cuantificado, para los diversos factores de escala y para la información direccional.
- Formateador del tren de bits: Finalmente se emplea un formateador del tren de bits con el fin de multiplexar dicho tren, que consiste en los coeficientes espectrales codificados y cuantificados e información adicional de cada herramienta.
- Modelo psicoacústico: El umbral actual de enmascaramiento se calcula empleando un modelo psicoacústico a partir de la señal de entrada. Se emplea un modelo psicoacústico similar al modelo psicoacústico 2 de la Norma ISO/CEI 11172-3. Durante el proceso de cuantificación, con el fin de disminuir el ruido de cuantificación audible y adicionalmente para la selección de una herramienta de codificación adecuada, se emplea una relación señal/máscara, que se calcula a partir del umbral de enmascaramiento y el nivel de la señal de entrada.

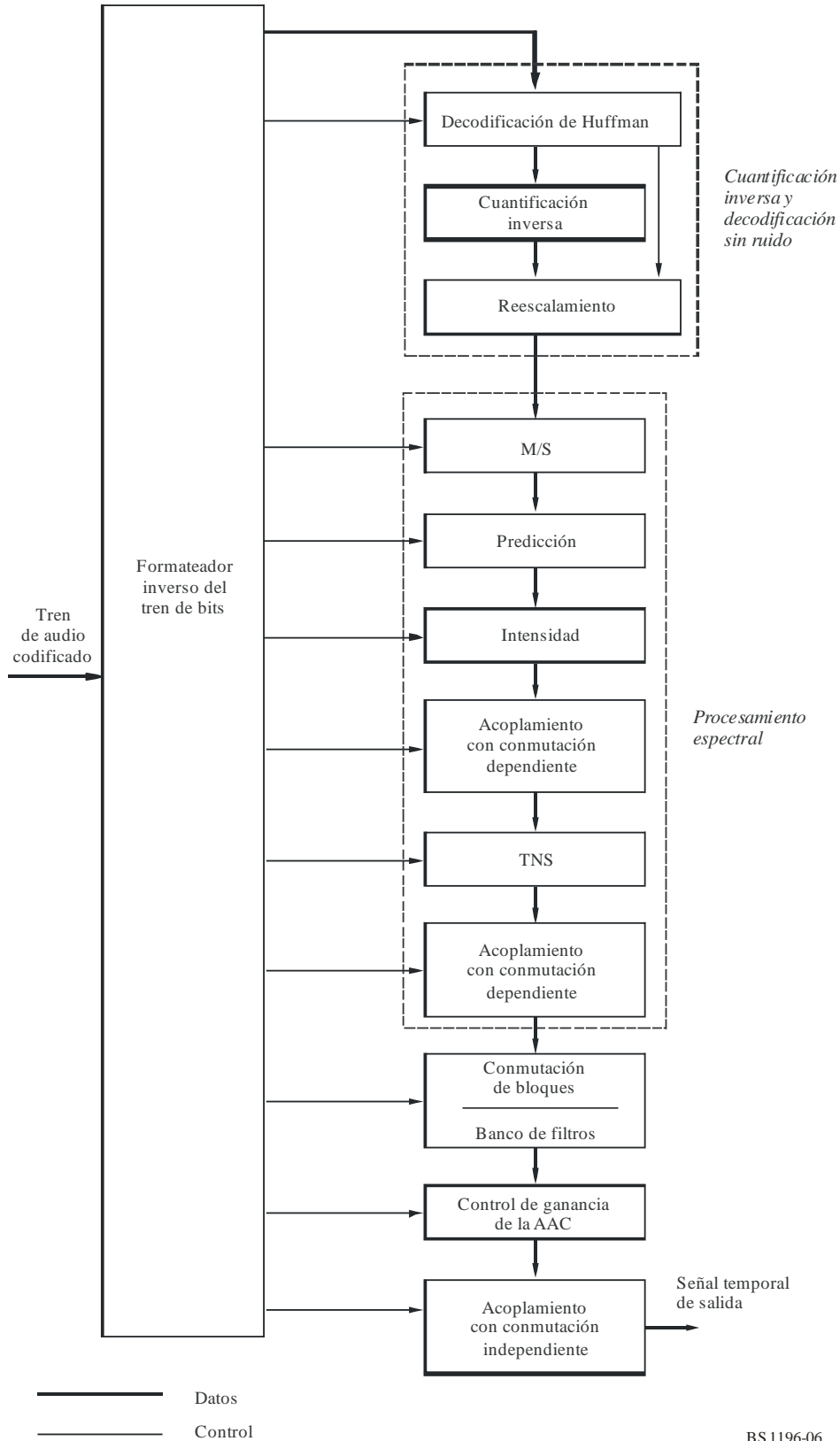
FIGURA 5
 Diagrama de bloques del codificador AAC con MPEG-2



3 Decodificación

En la Fig. 6 se muestra la estructura básica del decodificador AAC con MPEG-2. El proceso de decodificación es básicamente el inverso del proceso de codificación.

FIGURA 6
Diagrama de bloques del decodificador AAC con MPEG-2

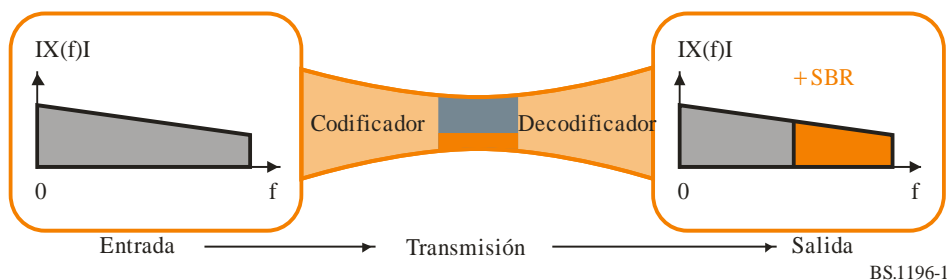


Las funciones del decodificador consisten en encontrar la descripción del espectro de audio cuantificado en el tren de bits, decodificar los valores cuantificados y otra información de reconstrucción, reconstruir el espectro cuantificado empleando las herramientas que estén activas en el tren de bits con el fin de llegar al espectro de la señal real y que se describe en el tren de bits de entrada, y finalmente convertir el espectro que está en el dominio de la frecuencia al dominio en el tiempo, utilizando opcionalmente una herramienta de control de ganancia. Tras la reconstrucción inicial y el escalamiento de la reconstrucción del espectro, se pueden emplear muchas herramientas opcionales que modifican uno o más de los espectros con el fin de lograr una codificación más eficaz. En cada una de las herramientas opcionales que operan en el dominio espectral, se reserva el uso de la opción de «dejar pasar», y en todos los casos en que se omite una operación espectral, el espectro se pasa directamente desde la entrada, a través de la herramienta sin ninguna modificación.

4 ACC de alta eficacia y réplica de banda espectral

La ACC de alta eficacia (HE ACC) introduce replicación de banda espectral (SBR) que es un método para lograr una codificación altamente eficiente de las altas frecuencias en los algoritmos de compresión de audio. Ofrece calidad mejorada de los códecs de audio y voz de baja velocidad binaria ya sea aumentando la anchura de banda de audio a una velocidad binaria determinada o mejorando la eficacia de la codificación para un nivel de calidad concreto.

Sólo se codifica y transmite la parte inferior del espectro. Se trata de la porción de espectro a la que es más sensible el oído humano. En vez de transmitir la parte alta del espectro, la SBR se emplea como un proceso posterior a la codificación para reconstruir las frecuencias más elevadas basándose en un análisis de las frecuencias más bajas transmitidas. La reconstrucción precisa se garantiza transmitiendo los parámetros relativos a la SBR en el tren de bits codificado a una velocidad binaria muy baja.



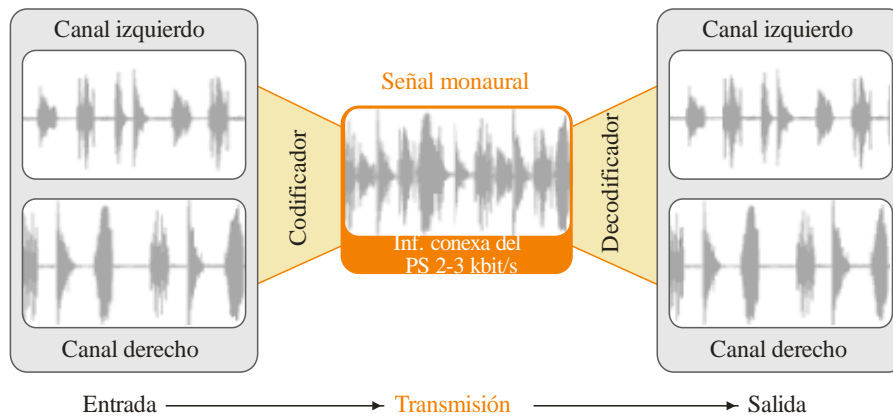
BS.1196-1

El tren de bits HE ACC es una mejora del tren de bits de audio AAC. Los datos de SBR adicionales se incorporan en el elemento de relleno AAC asegurando de esa forma la compatibilidad con la norma AAC. La tecnología HE ACC es un sistema de velocidad doble. El tren de bits de audio AAC con compatibilidad hacia atrás directa se transmite a la mitad de la velocidad de muestreo de la mejora SBR; de esa forma, un decodificador AAC, que no es capaz de decodificar los datos de mejora SBR, producirá una señal temporal de salida a la mitad de la velocidad de muestreo que la producida por un decodificador HE ACC.

5 AAC versión 2 de alta eficacia y estereofonía paramétrica

HE AAC v2 es una extensión de HE ACC e introduce la estereofonía paramétrica (PS) a fin de mejorar la eficacia de la compresión de audio en las señales estereofónicas de baja velocidad binaria.

El codificador analiza la señal de audio estereofónica y construye una representación paramétrica de la imagen estereofónica. Ya no hay necesidad de transmitir ambos canales y sólo se codifica una representación monoaural de la señal estereofónica original. Esta señal se transmite junto con los parámetros necesarios para la reconstrucción de la imagen estereofónica.



Como resultado, la calidad de audio percibida de un tren de bits de audio a baja velocidad binaria (por ejemplo, 24 kbit/s) que incorpora estereofonía paramétrica es mucho más elevada que la calidad de un tren de bits similar sin estereofonía paramétrica.

El tren de bits HE AAC v2 se construye sobre el tren de bits HE ACC. Los datos en estereofonía paramétrica adicionales se incluyen en el elemento de extensión SBR de un tren HE ACC monofónico, garantizando de esa forma la compatibilidad con la HE ACC así como con la ACC.

Un decodificador HE ACC, que no es capaz de decodificar la mejora estereofonía paramétrica, produce una señal de salida monofónica en toda la anchura de banda. Un simple decodificador ACC, que no es capaz de decodificar los datos de mejora SBR, produce una señal temporal de salida monofónica a la mitad de la velocidad de muestreo.

Anexo 3 (informativo)

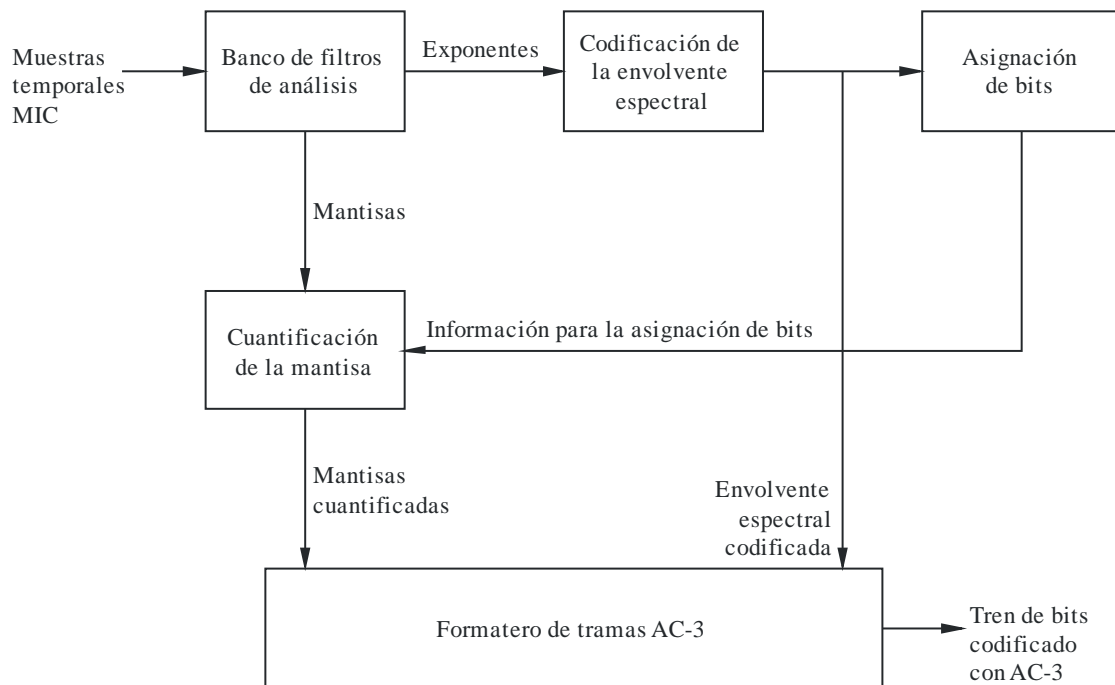
Audio AC-3 y E-AC-3

1 Codificación

El algoritmo de compresión digital AC-3 puede codificar de 1 a 5,1 canales de audio original en representación MIC (modulación por impulsos codificados) en un tren de bits serie con una velocidad binaria en la gama de 32 kbit/s a 640 kbit/s. El algoritmo AC-3 logra una alta ganancia de codificación (relación entre velocidad binaria de entrada y velocidad binaria de salida) mediante la cuantificación gruesa de una representación en el dominio de la frecuencia de la señal de audio. En la Fig. 7 se muestra un diagrama de bloques de este proceso. El primer paso en el proceso de codificación consiste en transformar la representación del audio de una secuencia de muestras temporales MIC a una secuencia de bloques de coeficientes de frecuencia. Esto se hace en el banco de filtros del análisis. Los bloques superpuestos de 512 muestras temporales se multiplican por una ventana temporal y se transforman al dominio de la frecuencia. Debido a la superposición de bloques, cada muestra de

entrada MIC está representada en dos bloques secuenciales transformados. Se puede disminuir la representación en el dominio de la frecuencia por un factor de dos, de manera tal que cada bloque contenga 256 coeficientes de frecuencia. Los coeficientes de frecuencia individual se representan en una notación exponencial binaria como un exponente binario y una mantisa. El conjunto de exponentes se codifica en una representación gruesa del espectro de señal conocido como envolvente espectral. Esta envolvente espectral se emplea en la rutina de asignación de bits de base, que determina la cantidad de bits que se deben emplear para codificar cada mantisa. La envolvente espectral y las mantisas con codificación gruesa para 6 bloques de audio (1 536 muestras de audio) se formatean para formar una trama AC-3. El tren de bits de AC-3 es una secuencia de tramas AC-3.

FIGURA 7
Codificador AC-3



BS.1196-07

El codificador AC-3 real es más complejo de lo que muestra la Fig. 7. También incluye las siguientes funciones no indicadas anteriormente:

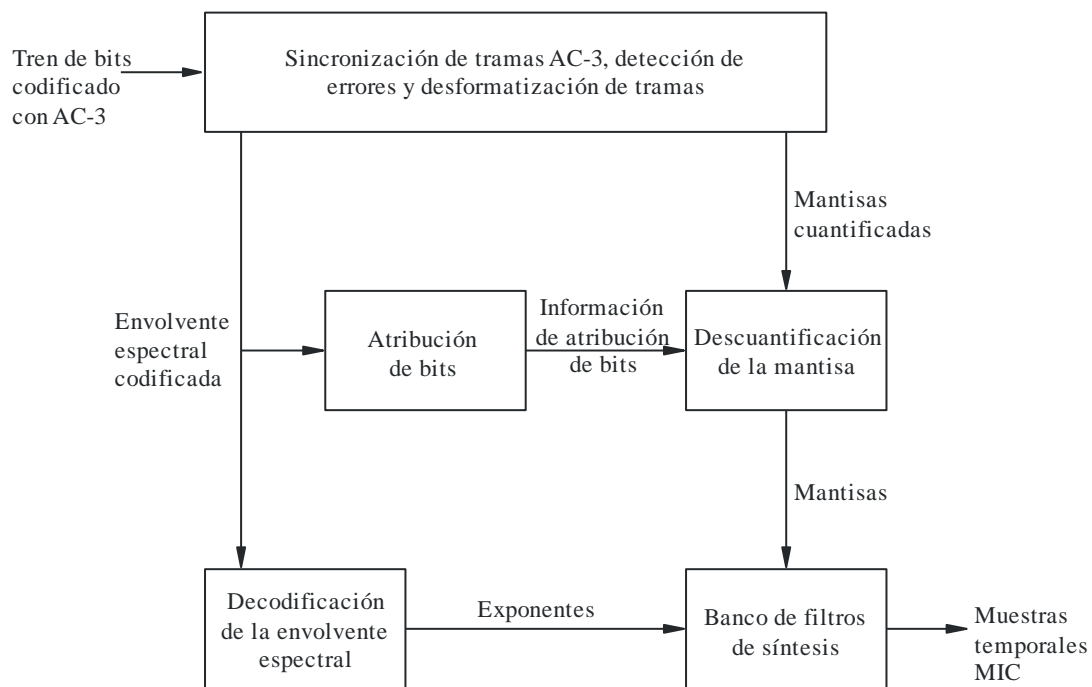
- se anexa un encabezamiento de trama que contiene información (velocidad binaria, velocidad de muestreo, número de canales codificados, etc.) necesaria para sincronizar y decodificar el tren de bits codificado;
- se insertan códigos de detección de error que le permiten al decodificador verificar que una trama recibida de datos no tenga errores;
- se puede modificar dinámicamente la resolución espectral del banco de filtro de análisis con el fin de lograr una mejor aproximación a la característica tiempo/frecuencia de cada bloque de audio;
- se puede codificar la envolvente espectral con una resolución variable de tiempo/frecuencia;
- con el fin de lograr una asignación de bits óptima, se puede realizar una asignación de bits más compleja, y modificar los parámetros de la rutina de asignación de bits base;
- se pueden acoplar entre sí los canales a altas frecuencias con el fin de lograr una mayor ganancia de codificación para el funcionamiento a velocidades binarias menores;

- en el modo bicanal se puede llevar a cabo de manera selectiva un proceso de reelaboración de las matrices de sonido, con el fin de lograr una ganancia de codificación adicional, y para obtener mejores resultados si la señal bicanal se decodifica mediante un decodificador de entorno matricial.

2 Decodificación

El proceso de decodificación es básicamente el inverso del proceso de codificación. El decodificador, que se muestra en la Fig. 8, debe sincronizarse con respecto al tren de bits codificado, hay que verificar los errores, y realizar el proceso inverso de formateo de los diferentes tipos de datos tales como la envolvente espectral codificada y las mantisas cuantificadas. Se ejecuta la rutina de asignación de bits y los resultados se emplean para desempaquetar y descuantificar las mantisas. Se decodifica la envolvente espectral con el fin de obtener los exponentes. Para producir las muestras temporales MIC decodificadas se transforman de nuevo al dominio del tiempo los exponentes y las mantisas.

FIGURA 8
Decodificador AC-3



BS.1196-08

El decodificador AC-3, que es en realidad más complejo que el que se muestra en la Fig. 8, incluye las siguientes funciones no indicadas anteriormente:

- se puede aplicar la ocultación de errores o el silenciamiento en los casos en que se detecte un error;
- se deben desacoplar los canales que hayan tenido acoplados sus contenidos de alta frecuencia;
- se debe llevar a cabo el proceso de elaboración inversa de matrices (en modo bicanal) si los canales han pasado por el proceso de reelaboración de matrices;

- la resolución del banco de filtros de síntesis se debe alterar de manera dinámica en la misma forma que se hizo con el banco de filtros de análisis del codificador durante el proceso de codificación.

3 E-AC-3

El AC-3 mejorado (E-AC-3) añade varias características y herramientas de codificación adicionales al códec AC-3 básico descrito anteriormente. Las herramientas de codificación adicionales proporcionan una mejora en la eficacia de la codificación permitiendo el funcionamiento a velocidades binarias inferiores, mientras que las características adicionales ofrecen flexibilidad de aplicación adicional.

Herramientas de codificación adicionales:

- Transformada híbrida adaptativa – Capa adicional aplicada en el banco de filtros de análisis/síntesis para proporcionar una resolución espectral más detallada (1/6 de AC-3).
- Procesamiento de preruido transitorio – Herramienta adicional para disminuir el preruido transitorio.
- Extensión espectral – Síntesis del decodificador de las componentes de frecuencias más elevadas basándose en la información conexas creada por el decodificador.
- Acoplamiento mejorado – Tratamiento de la fase y la amplitud en el acoplamiento de canal.

Características adicionales:

- Granularidad en la velocidad de datos más detallada.
- Máxima velocidad de datos más elevada (3 Mbit/s).
- Los subtrenes pueden incorporar canales de audio adicionales, por ejemplo, 7,1 canales o pistas de comentarios.

Anexo 4 (informativo)

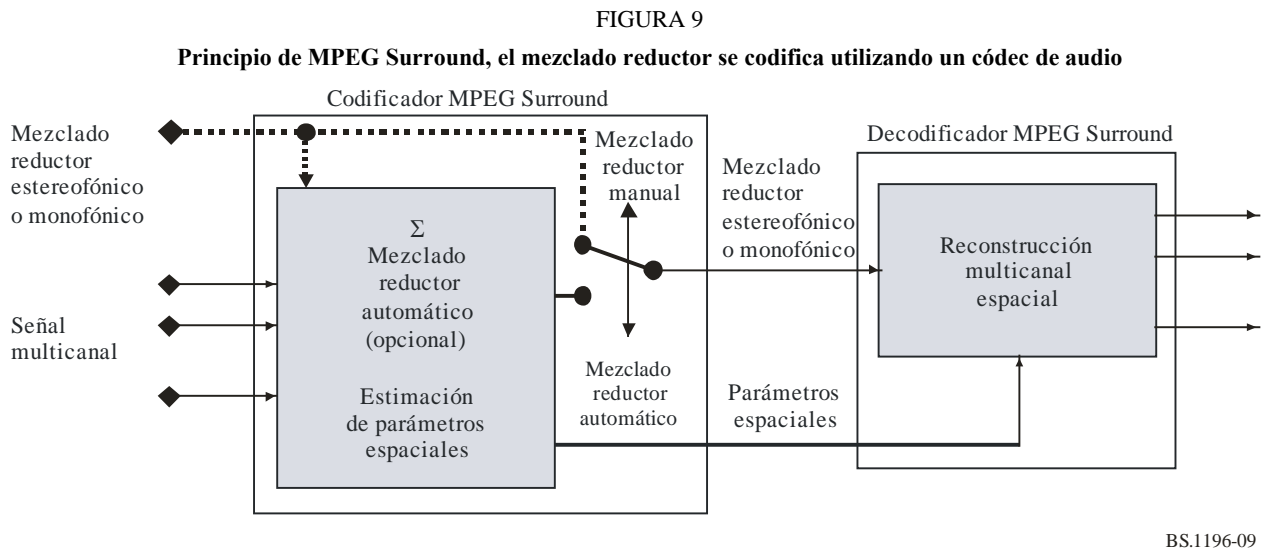
MPEG Surround

1 Introducción

La tecnología ISO/CEI 23003-1 o MPEG Surround (envolvente) proporciona un método extremadamente eficaz para la codificación del sonido multicanal y permite la transmisión del sonido ambiente a velocidades binarias normalmente utilizadas para la codificación del sonido monofónico o estereofónico. Es capaz de representar una señal de audio multicanal de N canales basada en un mezclado reductor de canal $M < N$ y datos de control adicional. En los modos de funcionamiento preferidos, un codificador MPEG Surround crea un mezclado reductor monofónico o estereofónico a partir de la señal de entrada de audio multicanal. El mezclado reductor se codifica empleando un códec de audio básico normalizado; por ejemplo, uno de los sistemas de codificación indicados en los *recomienda* 1 y 2. Además del mezclado reductor, MPEG Surround genera una descripción del parámetro de imagen espacial del audio multicanal que se añade como un tren de datos auxiliar al códec de audio básico de manera compatible hacia atrás. Los decodificadores monofónicos o estereofónicos establecidos ignorarán los datos auxiliares y reproducirán la señal de audio con

mezclado reductor monofónico o estereofónico. Los decodificadores de MPEG Surround decodificarán en primer lugar el mezclado reductor monofónico o estereofónico y a continuación utilizarán los parámetros de imagen espacial extraídos del tren de datos auxiliares para generar una señal de audio multicanal de alta calidad.

La Fig. 9 ilustra el principio de MPEG Surround.



Utilizando MPEG Surround, los servicios existentes pueden mejorarse fácilmente para proporcionar sonido ambiente de manera compatible hacia atrás. Si bien un decodificador estereofónico en un dispositivo de usuario existente ignora los datos de MPEG Surround y reproduce la señal estereofónica sin ninguna degradación de calidad, un decodificador MPEG Surround proporcionará un audio multicanal de alta calidad.

2 Codificación

El objetivo del codificador MPEG Surround es representar una señal de entrada multicanal como una señal monofónica o estereofónica compatible hacia atrás, combinada con parámetros espaciales que permiten la reconstrucción de una salida multicanal que se asemeja a las señales de entrada multicanal originales desde un punto de vista perceptual. Además del mezclado reductor generado automáticamente, puede utilizarse un mezclado reductor creado externamente («mezclado reductor artístico»). El mezclado reductor deberá preservar las características espaciales del sonido de entrada.

MPEG Surround se basa en la tecnología de estereofonía paramétrica que ha sido combinada con HE-AAC, dando lugar a la especificación HE AAC v2. Combinando los módulos estereofónicos paramétricos múltiples y otros módulos recientemente desarrollados, se han definido varias estructuras que soportan distintas combinaciones de números de canales de salida y con mezclado reductor. Como ejemplo, para una señal de entrada de 5,1 multicanal dispone de tres configuraciones distintas; una configuración para los sistemas basados en mezclado reductor estereofónico (configuración 525) y dos configuraciones diferentes para los sistemas basados en mezclado reductor monofónico (una configuración 515₁ y 515₂ que emplea una concatenación distinta de recuadros).

MPEG Surround incorpora un cierto número de herramientas con características que permiten una amplia aplicación de la norma. Una característica esencial de MPEG Surround es la capacidad de incrementar gradualmente la calidad de la imagen espacial desde una tara espacial muy baja hasta la transparencia. Otra característica esencial es que la entrada del decodificador puede hacerse compatible con las tecnologías existentes de entornos matriciales.

Éstas y otras características se obtienen mediante las siguientes herramientas de codificación principales:

- Codificación residual: Además de los parámetros espaciales, también pueden transportarse señales residuales utilizando una técnica de codificación híbrida. Estas señales sustituyen parte de las señales descorreladas (que forman parte de los recuadros estereofónicos paramétricos). Las señales residuales se codifican transformando las señales en el dominio QMF en señales en el dominio MDCT, tras lo cual se codifican los coeficientes MDCT empleando AAC.
- Compatibilidad matricial: Opcionalmente, el mezclado reductor estereofónico puede someterse a un preprocesamiento para que sea compatible con las tecnologías existentes de entorno matricial a fin de asegurar la compatibilidad hacia atrás con los decodificadores que sólo pueden decodificar el tren de bits estereofónico pero van equipados con un decodificador de entorno matricial.
- Señales con mezclado reductor arbitrario: El sistema MPEG Surround es capaz de manejar no sólo las mezclas reductoras generadas por el codificador sino también el mezclado reductor artístico suministrado al codificador además de la señal original multicanal.
- MPEG Surround sobre MIC: Normalmente, los parámetros espaciales del sistema MPEG Surround se cursan en la porción de datos auxiliares del esquema de compresión de audio subyacente. Para aplicaciones donde el mezclado reductor se transmite como MIC, MPEG Surround también soporta un método que permite el transporte de los parámetros espaciales a través de los canales de audio sin comprimir. La tecnología subyacente se denomina datos enterrados.

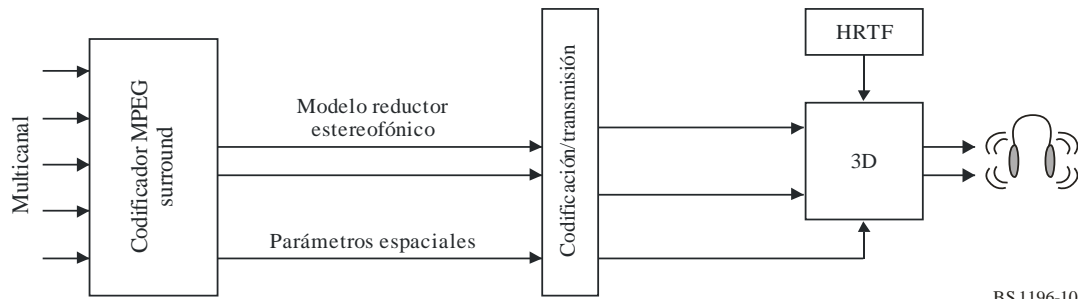
3 Decodificación

Además de ofrecer una salida multicanal, un decodificador MPEG Surround también soporta configuraciones de salida alternativas:

- Entorno virtual: El sistema MPEG Surround puede explotar los parámetros espaciales para entregar el mezclado reductor a una salida de entorno virtual estereofónica para su reproducción a través de auriculares tradicionales. La norma no especifica la Función de transferencia relativa a la cabeza (HRTF), sino simplemente la interfaz a esta HRTF permitiendo libertad en la implementación dependiendo del caso. El procesamiento de entorno virtual puede aplicarse tanto en el decodificador como en el codificador; en este último caso se ofrece la posibilidad de una experiencia de entorno virtual en el mezclado reductor que no requiere un decodificador MPEG Surround. Un decodificador MPEG Surround puede, sin embargo, deshacer el procesamiento de entorno virtual del mezclado reductor y volver a aplicar el entorno virtual alternativo. El principio básico se representa en la Fig. 10.

FIGURA 10

Decodificación del entorno virtual del MPEG Surround



BS.1196-10

- Modo de matriz mejorada: En el caso de contenido estereofónico tradicional, donde no hay información conexas espacial, el MPEG Surround puede estimar la información conexas espacial a partir del mezclado reductor y crear de esa forma la salida multicanal ofreciendo al mismo tiempo una calidad que va más allá de los sistemas de entorno matricial convencionales.
- Poda: Como resultado de la estructura subyacente, un decodificador MPEG Surround puede adaptar su salida a configuraciones de canal donde el número de canales es inferior al número de canales en la entrada multicanal del codificador.

4 Perfiles y niveles

El decodificador MPEG Surround puede implementarse como una versión de alta calidad y una versión de baja calidad. Ambas versiones funcionan sobre el mismo tren de datos aunque con distintas señales de salida.

El perfil de línea de base MPEG Surround define seis niveles jerárquicos diferentes que permiten distintos números de canales de entrada y salida, para diferentes gamas de velocidad de muestreo y para diversas anchuras de banda de la decodificación de señal residual. El nivel del decodificador debe ser igual o mayor que el nivel del tren de bits para garantizar una decodificación adecuada. Además, los decodificadores de nivel 1, 2 y 3 son capaces de decodificar todos los trenes de bits de nivel 2, 3 y 4, aunque posiblemente con una calidad ligeramente reducida debido a las limitaciones del decodificador. La calidad y el formato de la salida de un decodificador MPEG Surround dependen, además, de la configuración específica del propio decodificador. No obstante, los aspectos de configuración del decodificador son completamente ortogonales a los diferentes niveles de este perfil.

5 Interconexión con códecs de audio

MPEG Surround funciona como una extensión de preprocesamiento y postprocesamiento en los esquemas de codificación de audio existentes. Por tanto, está equipado con medios para acomodar virtualmente cualquier codificador de audio básico. La configuración de trama del MPEG Surround es extremadamente flexible a fin de asegurar la sincronía con una amplia gama de codificadores y se proporcionan los medios para optimizar la conexión con codificadores que ya utilicen herramientas paramétricas (por ejemplo, replicación de banda espectral) MPEG Surround funciona como una extensión de preprocesamiento y postprocesamiento en los esquemas de codificación de audio existentes. Por tanto, está equipado con medios para acomodar virtualmente cualquier codificador de audio básico. La configuración de trama del MPEG Surround es extremadamente flexible a fin de asegurar la sincronía con una amplia gama de codificadores y se proporcionan los medios para optimizar la conexión con codificadores que ya utilicen herramientas paramétricas (por ejemplo, replicación de banda espectral).

Anexo 5 (informativo)

Codificación de audio avanzada de gran eficacia ampliado (HE AAC ampliado)

1 Introducción

El perfil HE AAC ampliado se especifica en la Norma ISO/CEI 23003-3 sobre codificación unificada de la voz y el sonido (USAC) MPEG-D. USAC es una norma de codificación de audio que permite codificar voz, sonido o una combinación de voz y sonido con una calidad de audio coherente para todo el material sonoro en una amplia variedad de velocidades binarias. Apoya la codificación monocanal y multicanal a velocidades binarias elevadas, ofreciendo una calidad perceptualmente transparente. Al mismo tiempo, propicia una codificación muy eficaz a velocidades binarias muy bajas conservando todo el ancho de banda de audio.

Mientras que los códecs de audio anteriores presentaban ventajas e inconvenientes en la codificación de voz o sonido, USAC es capaz de codificar todo el contenido con una alta fidelidad similar, con independencia del tipo de contenido de que se trate.

Para obtener la misma calidad en la codificación de voz y sonido, USAC emplea la probada transformada discreta de coseno modificada (MCDT), basada en técnicas de codificación conocidas a partir de audio MPEG-4 (MPEG-4 AAC, HE AAC, HE AAC v2), y las combina con elementos especializados del codificador de voz, como la predicción lineal con excitación por código algebraico (ACELP). Distintas herramientas de codificación paramétricas, como la replicación de banda espectral (SBR) MPEG-4 y el MPEG-D MPEG Surround, se ven mejoradas y estrechamente incorporadas al códec. Fruto de ello es una codificación altamente eficaz y que funciona hasta en las velocidades binarias más bajas.

En la actualidad, la norma USAC especifica dos perfiles:

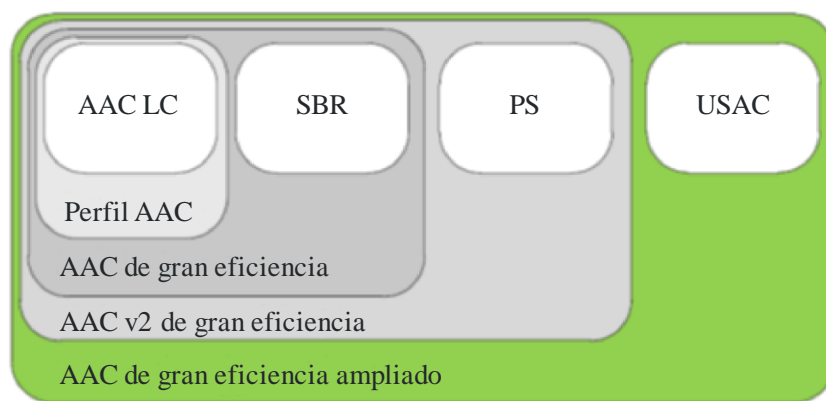
– *Perfil de línea de base de USAC*

El perfil de línea de base USAC incluye todas las funciones de la norma USAC y mantiene la complejidad computacional global en unos niveles bajos. Quedan excluidas aquellas herramientas que exigen una memoria o capacidad de tratamiento excesivas.

– *Perfil HE AAC ampliado*

Dirigido específicamente a las aplicaciones que deben mantener la compatibilidad con la familia de perfiles AAC existentes (AAC, HE AAC y HE AAC v2), este perfil añade distintas capacidades USAB para ampliar el perfil HE AAC v2 existente. Incluye el nivel 2 del *perfil de línea de base USAC*. En consecuencia, los decodificadores HE AAC ampliado pueden descodificar todos los trenes de datos HE AAC v2, así como los trenes de datos USAC (hasta dos canales).

FIGURA 11
Estructura de HE AAC ampliado



BS.1196-11

USAC apoya frecuencias de muestreo que van desde 7,35 kHz hasta 96 kHz y ha demostrado que ofrece una buena calidad de audio para un abanico de velocidades binarias que van desde los 8 kbit/s hasta aquellas en las que se obtiene la transparencia perceptual. Así quedó probado en la prueba de verificación (Documento MPEG2011/N12232) del JTC 1 de la ISO/CEI, la CE 29 y el GT 11 que se adjunta al Documento 6B/286(Rev. 2).

La elección de la configuración de canales es libre. Pueden señalarse eficazmente 13 configuraciones de canales por defecto distintas para las aplicaciones más comunes. Estas configuraciones por defecto incluyen todas las configuraciones de canal MPEG-4, como mono, estéreo, 5.0 y 5.1 Surround, o incluso las configuraciones de altavoces 7.1 ó 22.2.

2 Codificación

Comúnmente utilizada en la normalización de MPEG, la Norma ISO/CEI 23003-3 solamente especifica el proceso de decodificación de archivos y trenes de datos USAC MPEG-D. Normativamente, no especifica el proceso de codificación.

En la Fig. 12 se muestra una típica estructura de codificador posible.

El codificador se compone de las siguientes herramientas de codificación:

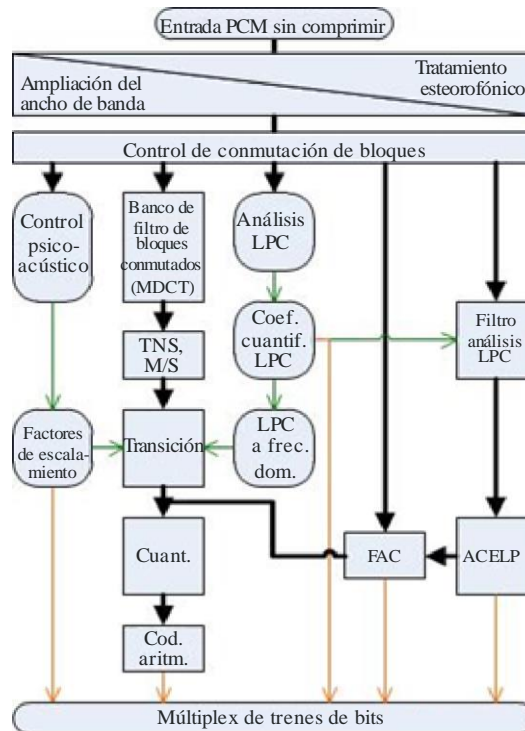
- **Procesamiento estereofónico:** En velocidades binarias bajas/intermedias, USAC utiliza tecnologías de codificación estereofónica paramétrica. En principio, son similares a la herramienta PS descrita en el Apéndice 2.5 de la Norma ISO/CEI 23003-3 aunque en este caso se basan en MPEG Surround, tal y como se describe en el Anexo 4 y, por lo tanto, reciben el nombre de MPEG Surround 2-1-2 (MPS 2-1-2). El codificador extrae una representación paramétrica de gran eficiencia de la imagen estereofónica a partir de la señal de audio entrante. Estos parámetros se transmiten en el tren de datos junto con una señal submezclada monoaural. Opcionalmente, el codificador puede optar por transmitir una señal residual que modifique el proceso de reconstrucción de la señal estereofónica en el decodificador. El mecanismo de codificación residual permite la armoniosa transición de una codificación estereofónica del canal totalmente paramétrica a otra totalmente discreta. La herramienta MPS 2-1-2 es un elemento intrínseco del códec USAC. A velocidades binarias superiores, si la codificación paramétrica y ACELP no están normalmente activas, cabe la posibilidad de llevar a cabo la codificación estereofónica únicamente en el dominio MDCT mediante una predicción estereofónica de valor complejo. Este método recibe el nombre de

codificación estereofónica de predicción compleja. Puede verse como una generalización de la codificación estereofónica M/S tradicional.

- Ampliación del ancho de banda: La ampliación del ancho de banda paramétrica es una versión mejorada múltiple de la replicación de banda espectral (SBR) MPEG-4 descrita en el Apéndice 2.4 de la Norma ISO/CEI 23003-3. El codificador estima la envolvente espectral y la tonalidad de las bandas de audiofrecuencia superiores y transmite los parámetros correspondientes al decodificador. El codificador puede elegir entre dos tipos de reemisores distintos (armonioso o copia) y tres factores de transposición (1:2, 3:8, 1:4). La herramienta SBR ampliada es un elemento intrínseco del códec USAC.
- Banco de filtros, conmutación de bloques: Un banco de filtros basado en MDCT constituye la base del codificador básico. En función del mecanismo de conformación de ruido cuantificado que se aplique, se podrá elegir una u otra resolución de la transformada (1 024, 512, 256, o 128 líneas espectrales). Combinado con el factor de transposición SBR 3:8, es posible modificar la resolución de las alternativas anteriormente mencionadas a un valor de $\frac{3}{4}$, mejorando así la granularidad temporal incluso a velocidades de muestreo inferiores.
- Conformación de ruido temporal (TNS), codificación estereofónica M/S, cuantificación: Estas herramientas se han adoptado a partir de AAC y se emplean de un modo similar al descrito en el Apéndice 2.2 de la Norma ISO/CEI 23003-3.
- Codificador aritmético adaptable al contexto: La codificación sin ruido (es decir, entropía) de los coeficientes espectrales MDCT se lleva a cabo a través de un codificador aritmético que elige sus cuadros de probabilidad a partir de líneas espectrales previamente codificadas.
- Control psicoacústico, transición del factor de escala: El modelo psicoacústico basado en el factor de escala es similar al empleado en AAC; véase el Apéndice 2.2 de la Norma ISO/CEI 23003-3.
- Transición basada en parámetros de codificación predictiva lineal (LPC): Esta herramienta de conformación del ruido espectral es una alternativa a la transición del factor de escala anteriormente mencionada. La versión ponderada de la representación de una frecuencia del conjunto de coeficientes de un filtro LPC se aplica a los coeficientes espectrales MDCT antes de llevar a cabo la cuantificación y la codificación.
- ACELP: La herramienta codificadora de predicción lineal con excitación por código algebraico (ACELP) emplea la probada representación de excitación por libro de códigos adaptable/innovadora obtenida de los códecs más avanzados de señal vocal.
- Múltiplex de trenes de bits: El último tren de bits se compone de varios elementos producidos por las herramientas de codificación.
- FAC: La herramienta de corrección por solapamiento en recepción (FAC) proporciona un mecanismo para una transición armoniosa de una codificación basada en MDCT afectada por el solapamiento a una codificación ACELP basada en el dominio del tiempo.

FIGURA 12

Diagrama de bloque del codificador MPEG-D USAC



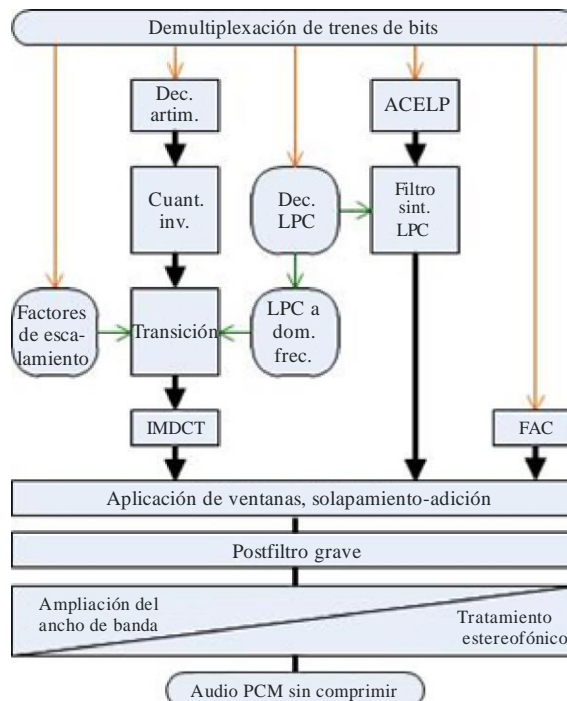
BS.1196-12

3 Decodificación

En la Fig. 13 se muestra la estructura básica del decodificador USAC MPEG-D. El proceso de decodificación suele seguir el camino inverso al descrito en el proceso de codificación.

FIGURA 13

Diagrama de bloques del decodificador MPEG-D USAC



BS.1196-13

El proceso de decodificación puede definirse a grandes rasgos de la manera siguiente:

- *Demultiplexación de trenes de bits:* El decodificador encuentra toda la información relacionada con las herramientas en el tren de datos y la transmite a los respectivos módulos del decodificador.
- *Decodificación básica:* En función del contenido del tren de datos, el decodificador:
 - decodifica y cuantifica inversamente los coeficientes espectrales MDCT, aplica una transición basada en la información del factor de escala o en la información del coeficiente LPC, y aplica nuevas (adicionales) herramientas basadas en MDCT si existen y son aplicables. Por último, se aplica MDCT inverso para obtener la señal en el dominio del tiempo correspondiente; o
 - decodifica la información relativa a ACELP, provoca una señal por excitación y sintetiza una señal de salida con la ayuda de un filtro LPC.
- *Aplicación de ventanas, solapamiento-adición:* Las tramas subsiguientes del codificador básico se concatenan o fusionan en el proceso habitual de solapamiento-adición, tal y como se conoce a partir de AAC. Las transiciones entre la codificación basada en ACELP y la codificación basada en MDCT se obtienen fusionando los datos FAC decodificados.
- *Posfiltro grave:* Puede aplicarse un filtro facultativo de mejora del tono para mejorar la calidad de la señal vocal.
- *Ampliación del ancho de banda, tratamiento estereofónico:* Por último, se aplican las herramientas de codificación paramétrica para ampliar el ancho de banda y las herramientas de codificación estereofónica para reconstruir la señal estereofónica discreta en todo el ancho de banda.

Para cada una de las herramientas optativas, se conserva la opción de «paso a través», y en todos los casos en los que se omite una operación, los datos pasan directamente por la herramienta sin ser modificados.

4 Perfiles y niveles

En la actualidad, MPEG define dos perfiles que emplean el códec USAC.

- *Perfil de línea de base de USAC*

El perfil de línea de base de USAC contiene todo el códec USAC a excepción de un pequeño número de herramientas que muestran una complejidad computacional en el peor de los casos excesiva. Estas herramientas no se han descrito anteriormente. Este perfil proporciona un perfil autónomo claro para aplicaciones y posibilidades de uso en los que no es pertinente la capacidad de soportar los perfiles de la familia AAC (perfil AAC, perfiles HE AAC, perfil HE AAC v2).

- *Perfil HE AAC ampliado*

El perfil de gran eficacia HE AAC ampliado se compone de todas las herramientas del perfil de gran eficiencia AAC v2 y, como tal, puede decodificar todos los trenes de los perfiles de la familia AAC. Además, el perfil incorpora la posibilidad mono/estéreo del perfil de línea de base de USAC. En consecuencia, este perfil es la evolución natural del perfil HE AAC v2, ya que la parte mono/estéreo de USAC (cuando funciona a baja velocidad) ofrece un valor adicional de rendimiento coherente para todos los tipos de contenidos a velocidades bajas.

Anexo 6 (informativo)

Puntos código independientes de la codificación (CICP) para la codificación MPEG

1 Introducción

La norma ISO/CEI 23001-8:2013 describe los aspectos relativos a la codificación de los programas de audio que son independientes de la representación codificada, comprendida la posición y la disposición de los sistemas de altavoces. Las configuraciones de canales por defecto incluyen las especificadas en las Recomendaciones UIT-R BS.775 o UIT-R BS.2051. Todas las configuraciones de canal se muestran en el Cuadro 3.

CUADRO 3

Configuraciones de canales y disposición de altavoces (Nota 1)

Valor de la configuración de canales ^{*1} (Nota 1)	Nº de altavoces (Nota 2)	Correspondencia entre el elemento por defecto y el altavoz (Nota 3)	Nombre del canal especificado en las Recomendaciones UIT-R BS.775 o BS.2051 (Nota 4)
0		Cualquier configuración	
1	1/0.0 (0+1+0)	M+000	Mono
2	2/0.0 (0+2+0)	M+030	Izquierdo
		M-030	Derecho
3	3/0.0 (0+3+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
4	3/1.0 (0+4+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
		M+180	Mono envolvente
5	3/2.0 (0+5+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
		M+110	Izquierdo envolvente
		M-110	Derecho envolvente
6	3/2.1 (0+5+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
		M+110	Izquierdo envolvente
		M-110	Derecho envolvente
		LFE1	Efectos de baja frecuencia

CUADRO 3 (continuación)

Valor de la configuración de canales ^{*1} (Nota 1)	Nº de altavoces (Nota 2)	Correspondencia entre el elemento por defecto y el altavoz (Nota 3)	Nombre del canal especificado en las Recomendaciones UIT-R BS.775 o BS.2051 (Nota 4)
7	5/2.1 (0+7+0)	M+000	N/A ^{*2}
		M+030	
		M-030	
		M+045	
		M-045	
		M+110	
		M-110	
		LFE1	
8	1+1	Canal 1	N/A
		Canal 2	
9	2/1.0 (0+3+0)	M+030	Izquierdo
		M-030	Derecho
		M+180	Mono envolvente
10	2/2.0 (0+4+0)	M+030	Izquierdo
		M-030	Derecho
		M+110	Izquierdo envolvente
		M-110	Derecho envolvente
11	3/3.1 (0+6+0)	M+000	N/A
		M+030	
		M-030	
		M+110	
		M-110	
		M+180	
		LFE1	
12	3/4.1 (0+7+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
		M+090	Lateral izquierdo envolvente
		M-090	Lateral derecho envolvente
		M+135	Posterior izquierdo envolvente
		M-135	Posterior derecho envolvente
		LFE1	Efectos de baja frecuencia

CUADRO 3 (continuación)

Valor de la configuración de canales ^{*1} (Nota 1)	Nº de altavoces (Nota 2)	Correspondencia entre el elemento por defecto y el altavoz (Nota 3)	Nombre del canal especificado en las Recomendaciones UIT-R BS.775 o BS.2051 (Nota 4)
13	11/11.2 (9+10+3)	M+000	Central frontal
		M+030	Central izquierdo frontal
		M-030	Central derecho frontal
		M+060	Izquierdo frontal
		M-060	Derecho frontal
		M+090	Izquierdo lateral
		M-090	Derecho lateral
		M+135	Izquierdo posterior
		M-135	Derecho posterior
		M+180	Central posterior
		LFE1	Efectos de baja frecuencia-1
		LFE2	Efectos de baja frecuencia-2
		U+000	Central frontal superior
		U+045	Izquierdo frontal superior
		U-045	Derecho frontal superior
		U+090	Izquierdo lateral superior
		U-090	Derecho lateral superior
		T+000	Central superior
		U+135	Izquierdo posterior superior
		U-135	Derecho posterior superior
		U+180	Central posterior superior
		B+000	Central frontal inferior
B+045	Izquierdo frontal inferior		
U-045	Derecho frontal inferior		
14	5/2.1 (2+5+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
		M+110	Izquierdo envolvente
		M-110	Derecho envolvente
		LFE1	Efectos de baja frecuencia
		U+030	Izquierdo frontal superior
		U-030	Derecho frontal superior
15	5/5.2 (3+7+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
		M+090	Lateral izquierdo
		M-090	Lateral derecho
		M+135	Posterior izquierdo
		M-135	Posterior derecho
		U+045	Superior izquierdo
		U-045	Superior derecho
		UH+180	Superior central
		LFE1	Efectos de baja frecuencia izquierdos
		LFE2	Efectos de baja frecuencia derechos

CUADRO 3 (continuación)

Valor de la configuración de canales ^{*1} (Nota 1)	Nº de altavoces (Nota 2)	Correspondencia entre el elemento por defecto y el altavoz (Nota 3)	Nombre del canal especificado en las Recomendaciones UIT-R BS.775 o BS.2051 (Nota 4)
16	5/4.1 (4+5+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
		M+110	Izquierdo envolvente
		M-110	Derecho envolvente
		LFE1	Efectos de baja frecuencia
		U+030	Izquierdo frontal superior
		U-030	Derecho frontal superior
		U+110	Posterior izquierdo superior
		U-110	Posterior derecho superior
17	6/5.1 (6+5+0)	M+000	N/A
		M+030	
		M-030	
		M+110	
		M-110	
		LFE1	
		U+000	
		U+030	
		U-030	
		U+110	
		U-110	
		T+000	
		18	
M+030			
M-030			
M+110			
M-110			
M+150			
M-150			
LFE1			
U+000			
U+030			
U-030			
U+110			
U-110			
T+000			

CUADRO 3 (fin)

Valor de la configuración de canales ^{*1} (Nota 1)	Nº de altavoces (Nota 2)	Correspondencia entre el elemento por defecto y el altavoz (Nota 3)	Nombre del canal especificado en las Recomendaciones UIT-R BS.775 o BS.2051 (Nota 4)
19	5/6.1 (4+7+0)	M+000	Central
		M+030	Izquierdo
		M-030	Derecho
		M+090	Lateral izquierdo envolvente
		M-090	Lateral derecho envolvente
		M+135	Posterior izquierdo envolvente
		M-135	Posterior derecho envolvente
		LFE	Efectos de baja frecuencia
		U+045	Frontal izquierdo superior
		U-045	Frontal derecho superior
		U+135	Posterior izquierdo superior
		U-135	Posterior derecho superior
20	7/6.1 (4+9+0)	M+000	Central
		M+SC	Pantalla izquierda
		M-SC	Pantalla derecha
		M+030	Izquierdo
		M-030	Derecho
		M+090	Lateral izquierdo envolvente
		M-090	Lateral derecho envolvente
		M+135	Posterior izquierdo envolvente
		M-135	Posterior derecho envolvente
		LFE	Efectos de baja frecuencia
		U+045	Frontal izquierdo superior
		U-045	Frontal derecho superior
U+135	Posterior izquierdo superior		
U-135	Posterior derecho superior		
21-63	Reservado		

*1 La configuración del canal audio de salida se indica mediante un campo de seis bits que contiene el valor de la configuración del canal como se define en ISO/IEC 23001-8:2013, «Puntos código independientes de la codificación».

*2 N/A: no aplicable; la configuración no está disponible en la Recomendación UIT-R BS.2051 ni en la Recomendación UIT-R BS.775.

NOTA 1 – La lista se ha obtenido del Cuadro 8 de la norma ISO/CEI 23001-8:2013/Amd.1:2015.

NOTA 2 – El concepto de número de altavoz se describe en el convenio de «*altavoces frontales/altavoces envolventes. Altavoces LFE*» y entre paréntesis como «*Altavoces en la capa superior + altavoces en la capa media + altavoces en la capa inferior*» donde se excluyen los altavoces LFE.

NOTA 3 – Identificación de altavoces mediante etiquetas con arreglo a la Recomendación UIT-R BS.2051.

NOTA 4 – Las etiquetas y los nombres del canal dependen de la configuración real de canales.

Anexo 7 (informativo)

AC-4

1 Introducción

AC-4 es un sistema de codificación avanzado para la radiodifusión digital, que utiliza un algoritmo de compresión digital y varias herramientas de codificación paramétrica para mejorar la eficacia y la funcionalidad. AC-4 admite de forma nativa formatos de entrada y salida basados en canales y en objetos.

En el Cuadro 4 se enumeran los formatos de canal que AC-4 admite de forma nativa, incluidas las configuraciones de canal requeridas para el tipo de emisión definido en la Recomendación UIT-R BS.1548. AC-4 también admite la codificación de formatos con un cómputo de canales más elevado, que pueden utilizarse para dar soporte a sistemas de audio avanzados en el marco de la Recomendación UIT-R BS.2051.

CUADRO 4

Resumen de los formatos de canal admitidos de forma nativa

Formato	Número de canales	Nota
Monofónico (formato 1/0) ⁽¹⁾	1	
Estereofónico (formato 2/0) ⁽¹⁾	2	
3.0 (formato 3/0) ⁽¹⁾	3	
5.0/5.1 (formato 3/2) ⁽¹⁾	5/6	
7.0/7.1 (sistema I) ⁽²⁾	7/8	Señalización de tres configuraciones de altavoz diferentes
7.0.4/7.1+4 (sistema J) ⁽²⁾	11/12	Señalización de subconjuntos con menos canales
9.0.4/9.1.4	13/14	Señalización de subconjuntos con menos canales
22.2 (sistema H) ⁽²⁾	24	

⁽¹⁾ Especificado en la Recomendación UIT-R BS.775.

⁽²⁾ Especificado en la Recomendación UIT-R BS.2051.

AC-4 puede codificar de 1 a 22,2 canales de audio original en representación MIC (modulación por impulsos codificados) en un tren de bits serie con una velocidad binaria en la gama de 24 kbit/s a 1 536 kbit/s. Además de soportar representaciones basadas en canales, AC-4 también permite la codificación de objetos de audio dinámicos. El documento ETSI TS 103 190-2 contiene una descripción integral de la sintaxis de tren binario de AC-4.

En el Cuadro 5 se proponen velocidades binarias para las distintas configuraciones de canal, que cumplen los requisitos descritos en la Recomendación UIT-R BS.1548.

CUADRO 5

**Resumen de las velocidades binarias para
las distintas configuraciones de canal**

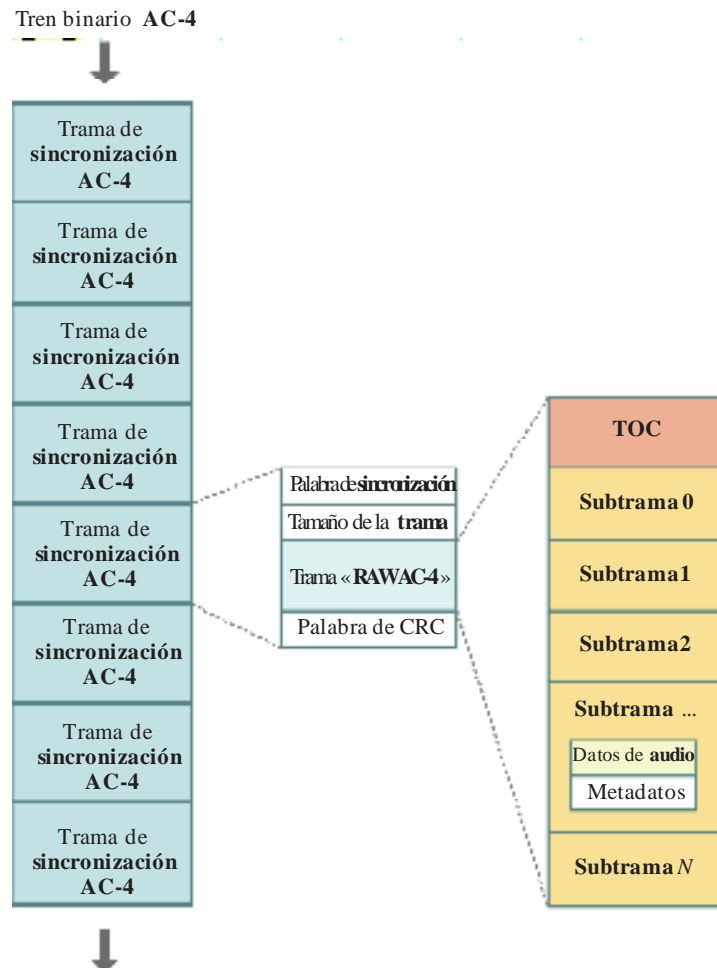
Modo	Velocidad binaria
Estereofónico 2.0	96 kbit/s
Envolvente 5.1	192 kbit/s
Envolvente 22.2	1 536 kbit/s

AC-4 también admite de forma nativa las siguientes características de sistema:

- gestión inteligente de la sonoridad, conforme a las Recomendaciones UIT-R BS.1770 y UIT-R BS.1771, incluida la señalización utilizada para indicar el nivel cumplimiento de la normativa internacional y de varias normas regionales en materia de sonoridad actualmente en vigor;
- soporte para codificar y decodificar representaciones de audio basadas en objetos y canales;
- soporte para entornos auditivos, conforme a lo estipulado en la Recomendación UIT-R BS.1909, en concreto «entornos domésticos» y «entornos móviles», incluido un control de gama dinámica avanzado aplicable a un amplio abanico de dispositivos para entornos «domésticos» y «móviles»;
- mejora del diálogo;
- codificación síncrona de trama de vídeo, que permite alinear las tramas de audio y las de vídeo;
- soporte nativo para el transporte y la señalización de los datos y los metadatos auxiliares mejorados.

Según se indica en la Fig. 14, el tren binario AC-4 se compone de una serie de tramas de sincronización AC-4, que empiezan con una palabra de sincronización y terminan con una palabra de verificación por redundancia cíclica (CRC). Gracias a la palabra de sincronización, el decodificador puede identificar fácilmente la trama AC-4 y empezar a decodificarla, mientras que la palabra CRC le permite detectar errores de tren binario y aplicar las medidas necesarias para ocultar los errores. La trama real del códec o trama «RAW AC-4» se compone de un cuadro de contenidos (TOC) y al menos una subtrama.

FIGURA 14
 Sintaxis de tren binario de alto nivel



BS.119644

Todas las subtramas incluyen tanto los datos de audio codificados, como los metadatos asociados (datos auxiliares). El TOC contiene la información necesaria sobre el modo en que se han de decodificar la o las diversas subtramas.

2 Codificación

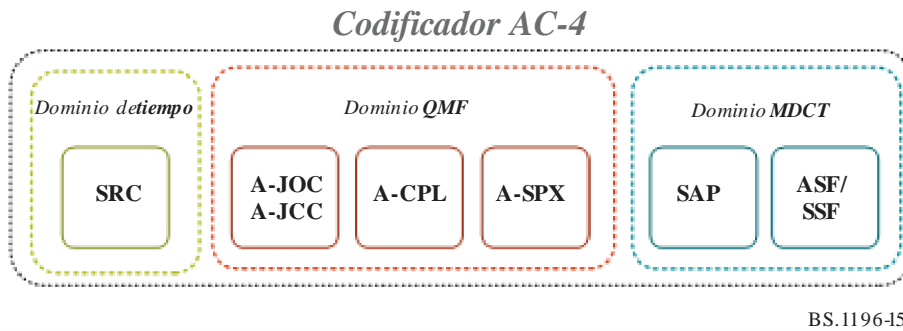
El codificador AC-4 no ha sido objeto de normalización, sin embargo, el hecho de que diversas herramientas de codificación gocen de soporte permite generar trenes binarios codificados compatibles.

AC-4 lleva a cabo la codificación y cuantificación del dominio de la transformada utilizando una transformada de coseno discreta modificada (MDCT) con conmutación de ventana flexible y codificación paramétrica en un banco de filtro de espejo en pseudocuatridadura. AC-4 puede codificar de 1 a 22,2 canales de audio original en representación MIC (modulación por impulsos codificados) en un tren de bits serie con una velocidad binaria en la gama de 24 kbit/s a 1 536 kbit/s. Para la codificación estereofónica de canal 2.0 y 5.1, una velocidad binaria del orden de 96 kbit/s o 192 kbit/s cumple los requisitos de calidad de funcionamiento especificados en la Recomendación UIT-R BS.1548. AC-4 también admite tres modos de atribución de bits: velocidad binaria constante, velocidad binaria promedio y velocidad binaria variable.

El codificador AC-4 puede implementarse junto con diversas herramientas de codificación (véase la Fig. 15), a fin de mejorar la eficacia y la flexibilidad y/o funcionalidad dependiendo del modo de operación y/o aplicación. El orden en que se utilizan las herramientas viene representado de izquierda a derecha, lo que significa que el audio MIC se introducirá en las herramientas situadas a la izquierda en primer lugar y, a continuación, las herramientas situadas a la derecha emitirán un tren binario codificado AC-4. A continuación se describen las herramientas de codificación:

- SRC: El codificador AC-4 requiere un convertidor de velocidad de muestreo (SRC) para ajustar la duración de una trama a la de la trama de vídeo, en los modos de codificación con sincronismo de trama. En función de la velocidad de la trama, la señal de entrada se convierte a una de las velocidades de muestreo internas de 46 034 Hz, 46 080 Hz, 48 000 Hz o 51 200 Hz, utilizadas por las herramientas del codificador basadas en MDCT y en filtros de espejo en cuadratura (QMF).
- A-JCC: La herramienta de codificación de canal conjunta avanzada (A-JCC) realiza una submezcla de la entrada de canal inmersiva (más de 5 canales) en un número menor de canales y codifica los parámetros conexos. Gracias a estos parámetros, el decodificador puede reconstruir todo el canal de entrada.
- A-JOC: La herramienta de codificación de objeto conjunta avanzada (A-JOC) recibe los objetos de audio como elementos de entrada y los codifica espacialmente para producir un número menor de objetos de salida y, de esta forma, reducir el número de señales codificadas con MDCT. Los parámetros codificados con el número reducido de objetos de salida permiten que el decodificador reconstruya los objetos.
- A-CPL: La herramienta de acoplamiento avanzada (A-CPL) realiza una submezcla de 2 canales en 1 y codifica los parámetros asociados, que permiten la reconstrucción de la entrada de los 2 canales originales.
- A-SPX: Las herramientas de extensión espectral avanzada (A-SPX) codifican los parámetros asociados con el contenido de alta frecuencia de la señal de entrada y, a continuación, procede a su ampliación espectral en el decodificador. Los parámetros comprenden medidas relacionadas con el envolvimiento, la tonalidad y el ruido. La resolución espectral y temporal de los parámetros puede adaptarse a las características de la señal de entrada.
- SAP: La herramienta de procesamiento de audio estereofónico (SAP) lleva a cabo una codificación de canal conjunta en el dominio MDCT entre dos o más canales de entrada.
- ASF: La herramienta de extremo frontal espectral de audio (ASF) es una herramienta de cuantificación y codificación basada en MDCT, que utiliza la conmutación de ventanas de transformación. El módulo de control de la conmutación de ventanas selecciona la longitud de transformación óptima para la trama correspondiente, en función del tipo de señal de entrada. Los coeficientes de MDCT y la información de control adicional se almacenan una vez realizada la cuantificación no lineal y la codificación sin ruido del tren binario. La distribución de bits a lo largo del tiempo y la frecuencia se halla bajo el control de una memoria intermedia de bits y un modelo de percepción. El modelo de memoria intermedia de bits también tiene en cuenta los bits utilizados por las demás herramientas de codificación y los metadatos generales.
- SSF: La herramienta de extremo frontal espectral vocal (SSF) es una herramienta alternativa de cuantificación y codificación basada en MDCT, específicamente concebida para la codificación vocal que se ejecuta en transformadas cortas. La SSF lleva a cabo cuantificaciones y codificaciones en el dominio MDCT utilizando un dispositivo de predicción de subbanda. Habida cuenta de que las herramientas SSF y ASF son mutuamente excluyentes, solo podrá utilizarse una de ellas para codificar la MDCT.

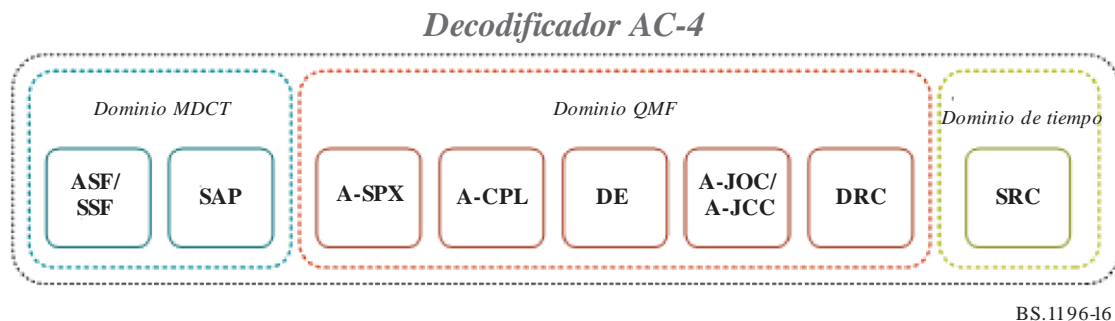
FIGURA 15
Herramientas de codificación disponibles en el sistema AC-4



3 Decodificación

El proceso de decodificación es básicamente el inverso del proceso de codificación y se muestra en la Fig. 16.

FIGURA 16
Herramientas de decodificación



El diseño del decodificador y del tren binario AC-4 admite de forma nativa la implementación de decodificadores más sencillos para dispositivos auxiliares con capacidades limitadas (por ejemplo, teléfonos móviles y/o tabletas).

Para dar soporte a esta capacidad, se utilizan dos métodos:

- 1) Decodificación básica: AC-4 admite dos modos de decodificación, a saber, completa y básica. El decodificador AC-4 puede funcionar en un modo de decodificación básica, en virtud del cual un subconjunto básico de canales (que representa el contenido de todos los canales de entrada) del programa codificado se decodifica permitiendo una reproducción compatible del programa con una complejidad computacional reducida.
- 2) Decodificación escalable: AC-4 también es compatible con la escalabilidad de velocidad de muestreo, en cuyo marco AC-4 puede soportar velocidades de muestreo más altas, concretamente de 96 kHz y 192 kHz, de manera escalable. Los dispositivos que no admiten más que salidas a 48 kHz solo necesitan decodificar la capa básica.

Además de permitir la decodificación del formato de canal soportado de forma nativa, según se indica en el Cuadro 4, AC-4 admite el uso de datos o metadatos auxiliares. Esta particularidad permite llevar a cabo un mezclado descendente de la salida del canal decodificado, pasando de un cómputo de canales más elevado a un número menor de canales de salida, tal y como suelen requerir los dispositivos.

Anexo 8 (informativo)

Perfil de baja complejidad de audio MPEG-H 3D

1 Introducción

La especificación de audio MPEG-H 3D es una norma de codificación de audio elaborada con objeto de permitir la codificación de señales de audio como canales de audio, objetos de audio o ambisonías de alto orden (HOA), y proporciona soluciones para la normalización de la sonoridad y el control de gama dinámica. Todos los contenidos (canales, objetos o HOA) pueden utilizarse de forma independiente o combinada. El uso de grupos de canales de audio, objetos o HOA permite la interactividad o la personalización de un programa, por ejemplo, mediante la selección de distintos idiomas o el ajuste de la ganancia o la posición de los objetos durante su reproducción en el decodificador MPEG-H.

Esta especificación se ha publicado como norma ISO/CEI 23008-3:2015. La Enmienda 3, en la que se describe el perfil de baja complejidad de audio MPEG-H 3D y la tecnología adicional, se publicó a principios de 2017.

El perfil de baja complejidad de audio MPEG-H 3D puede soportar un máximo de 24 altavoces de salida y 56 canales básicos de códec (de los cuales puede decodificar 28 de manera simultánea).

Ejemplos de posibles configuraciones de altavoz objetivo:

- sistema estereofónico 2.0 (formato 2/0 especificado en la Recomendación UIT-R BS.775);
- sistema de audio multicanal 5.1 (formato 3/2 especificado en la Recomendación UIT-R BS.775);
- sistema de audio inmersivo 10.2 (sistema F especificado en la Recomendación UIT-R BS.2051); y
- sistema de audio inmersivo 22.2 (sistema H especificado en la Recomendación UIT-R BS.2051).

La norma puede aplicarse a una amplia variedad de aplicaciones, incluido el almacenamiento y la transmisión de sonido estereofónico y envolvente. Su capacidad para soportar la interactividad y el sonido inmersivo reviste una importancia particular para satisfacer los requisitos en materia de transmisión de medios de la próxima generación, en particular, nuevos sistemas de radiodifusión de televisión (TV), servicios de emisión en continuo con fines lúdicos y servicios y contenido de realidad virtual.

Por ejemplo, en la radiodifusión de TV, los comentarios y los diálogos pueden enviarse como objetos de audio y combinarse con un fondo de canal inmersivo en el decodificador de audio MPEG-H 3D. Ello permite transmitir eficazmente los diálogos en múltiples idiomas y brindar al oyente la posibilidad de ajustar el equilibrio entre el diálogo y otros elementos sonoros según estime conveniente. Este concepto puede extenderse a otros elementos que no suelen formar parte de la radiodifusión, entre ellos, la audiodescripción para personas con discapacidad visual, los comentarios del director o los diálogos de los participantes en eventos deportivos.

El perfil de baja complejidad de audio MPEG-H 3D permite gestionar la sonoridad de conformidad con lo estipulado en las Recomendaciones UIT-R BS.1770 y ITU-R BS.1771, incluida la señalización utilizada para indicar el nivel cumplimiento de la normativa internacional y de varias normas regionales en materia de sonoridad. También es compatible con el control de gama dinámica avanzada (DRC) aplicable a un amplio abanico de dispositivos para entornos «domésticos» y «móviles».

2 Codificación

La arquitectura de códec de audio MPEG-H 3D parte de un códec perceptivo para la compresión de distintas clases de señales de entrada, basado a su vez en la codificación unificada de la voz y el sonido (USAC) MPEG. USAC permite la compresión de señales de audio de uno a múltiples canales a velocidades de al menos 8 kbit/s por canal.

Dados los nuevos requisitos emanados en el contexto del audio 3D, esta tecnología ha sido ampliada por medio de herramientas que permiten explotar especialmente los efectos perceptivos de la reproducción en 3D y, de esta forma, reforzar aún más la eficacia de la codificación. A continuación se facilitan algunos ejemplos en la materia:

- Se proporciona un ruido de relleno mejorado a través de una herramienta inteligente de relleno de espacios (IGF). Esta herramienta restaura paraméricamente segmentos del espectro transmitido utilizando información adecuada de mosaicos espectrales adyacentes en frecuencia y tiempo. La asignación y el procesamiento de estos mosaicos espectrales están bajo el control del codificador y se basan en un análisis de las señales de entrada. De este modo, las lagunas espectrales pueden colmarse con coeficientes espectrales que, en términos perceptivos, encajan mejor que las secuencias de ruido pseudoaleatorio del ruido de relleno convencional.
- Además de mejorar la eficacia de la codificación, la norma USAC-3D básica comprende nuevos mecanismos de señalización para configuraciones de altavoz y/o contenido 3D, y para el tipo de señales presentes en los trenes comprimidos (canal de audio frente a objeto de audio frente a señal HOA).

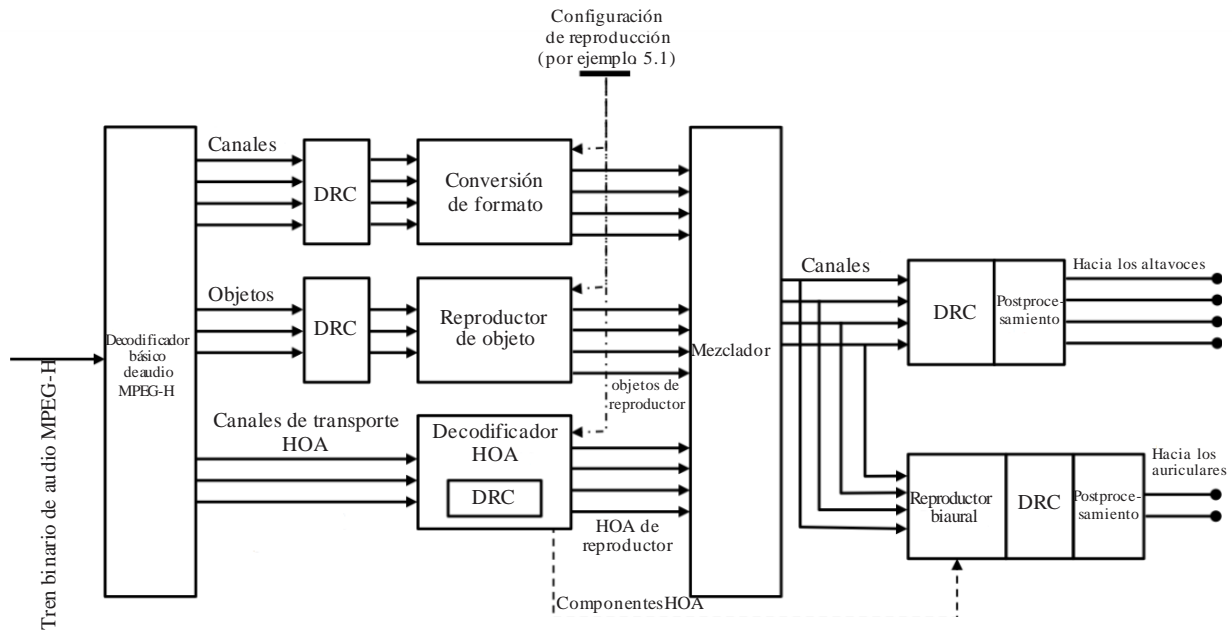
Entre las novedades en materia de diseño de cabida útil de audio comprimida, cabe destacar la mejora de los procedimientos para la conmutación de velocidad instantánea o entrada de señal rápida, tal como figura en el contexto de la emisión en continuo adaptativa dinámica (DASH) MPEG. A tal efecto, se han incluido en la sintaxis las denominadas «tramas de cabida útil inmediata», que permiten una transición sin lagunas entre dos trenes. Esta característica reviste una importancia particular para la emisión en continuo adaptativa a través de redes IP.

3 Decodificación

La Fig. 17 ilustra un diagrama de bloques de nivel superior, que guarda relación con la arquitectura del decodificador de audio del perfil de baja complejidad de audio MPEG-H 3D.

FIGURA 17

Diagrama de bloques de nivel superior de la arquitectura del decodificador de audio del perfil de baja complejidad de audio MPEG-H 3D



BS.119647

Los componentes principales – a saber, un conjunto de reproductores para las distintas clases de señales y un mezclador – integran el denominado decodificador USAC-3D básico. En un primer momento, el decodificador USAC-3D convierte las diferentes señales de base a partir de una representación comprimida de los datos conexos.

A continuación, las diferentes clases de señal (formas de onda en el caso de las señales de canal y de objeto o señales de coeficiente HOA) se introducen en los reproductores asociados, que establecen una correspondencia entre dichas señales y los altavoces, de conformidad con la configuración de reproducción disponible en el lado del receptor. Tan pronto como todas las señales estén disponibles en el formato de reproducción correspondiente, se combinan en una etapa de mezclado para alimentar el altavoz. Si se requiriese una representación binaural, la señal se convertiría en una escena 3D virtual para su reproducción a través de auriculares. Es posible transmitir cualquier combinación de cualesquiera tipos de señales en un único tren MPEG-H, por ejemplo, una combinación de señales de canal con señales de objeto o una escena HOA con objetos.

Los reproductores son:

- Un convertidor de formato, que permite convertir señales de canal de su formato de altavoz de producción a su configuración de altavoz de reproducción.
- Un reproductor de objetos, que permite disponer pistas de objetos estáticos o dinámicos en la configuración de reproducción.
- Un reproductor de HOA, que permite una conversión de la representación de HOA basada en la escena a la configuración de reproducción real.
- Un reproductor binaural, que permite una conversión de la configuración de altavoz virtual a la salida de auriculares.

Además, la reproducción y la realización de las diferentes clases de señal pueden controlarse a través de una interfaz de usuario, si los metadatos correspondientes marcan estas señales como aptas para la interactividad.

Anexo 9 (informativo)

DTS-UHD

1 Introducción

El sistema de codificación DTS-UHD utiliza el motor de codificación de audio (ACE) como método eficiente para comprimir formas de onda de audio. DTS-UHD soporta la codificación de audio por canales (CBA), los campos de sonido ambisónico y el audio por objetos (OBA). DTS-UHD se ha normalizado en ETSI TS 104 491 y forma parte de la especificación del tren de transporte DVB MPEG 2 ETSI TS 101 154. El sistema ofrece una experiencia totalmente inmersiva y, mediante la utilización de objetos, canales, campos sonoros o una combinación de los tres, permite la interacción y la personalización para mejorar la experiencia. También ofrece servicios de accesibilidad para los usuarios con deficiencias visuales y auditivas.

DTS-UHD soporta todas las disposiciones de altavoces especificadas en la Recomendación UIT-R BS.2051.

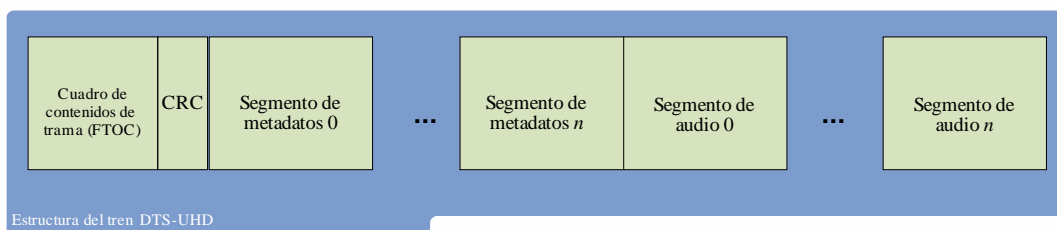
Si bien soporta totalmente el audio por canales, DTS-UHD también puede codificar hasta 224 objetos de audio discretos, que pueden además organizarse en hasta 32 grupos de objetos y 32 presentaciones en un solo tren.

Un tren DTS-UHD es una secuencia de tramas DTS-UHD con tres elementos principales:

- cuadro de contenidos de trama (FTOC) – este elemento permite al decodificador navegar directamente a los elementos de interés de la trama;
- elementos segmento de metadatos;
- elementos segmento de audio.

La estructura se muestra a continuación:

FIGURA 18
Estructura del tren DTS-UHD



BS.1196-18

El FTOC está presente en todas las tramas DTS-UHD e indica si la trama es una trama con sincronismo o sin sincronismo. Las tramas con sincronismo facilitan todos los parámetros necesarios para iniciar una sesión de decodificación.

Los segmentos de audio transportan las formas de onda comprimidas, que se comprimen y transportan de manera óptima en tramas con sincronismo o sin sincronismo. El segmento de audio se organiza en una serie de subtrenes que pueden representar formas de onda monofónicas, formas de onda estereofónicas y formas de onda LFE.

El segmento de metadatos describe por completo un componente de audio, incluido el tipo de componente, qué segmentos de audio se necesitan para reproducir el componente e información sobre

la sonoridad y la compresión de gama dinámica. El segmento de metadatos para un punto de acceso (una trama DTS-UHD con sincronismo) hace referencia a un segmento de audio que también es una trama con sincronismo y contiene el conjunto completo de metadatos necesarios para iniciar la decodificación.

2 Codificador

El codificador ACE DTS-UHD no está normativamente especificado por el ETSI. El codificador implementa un conjunto completo de herramientas para facilitar múltiples trenes binarios para el soporte de distintas configuraciones de altavoz.

La tecnología de codificación DTS-UHD mantiene la calidad de audio y puede producir trenes binarios que cumplen los requisitos de emisión de alta calidad de la Recomendación UIT-R BS.1548 a las siguientes velocidades binarias:

Configuración de altavoz	Velocidad binaria (kbit/s)
0+2+0 (estéreo)	128
0+5+0 (5.1)	192
4+7+0 (7.1.4)	288

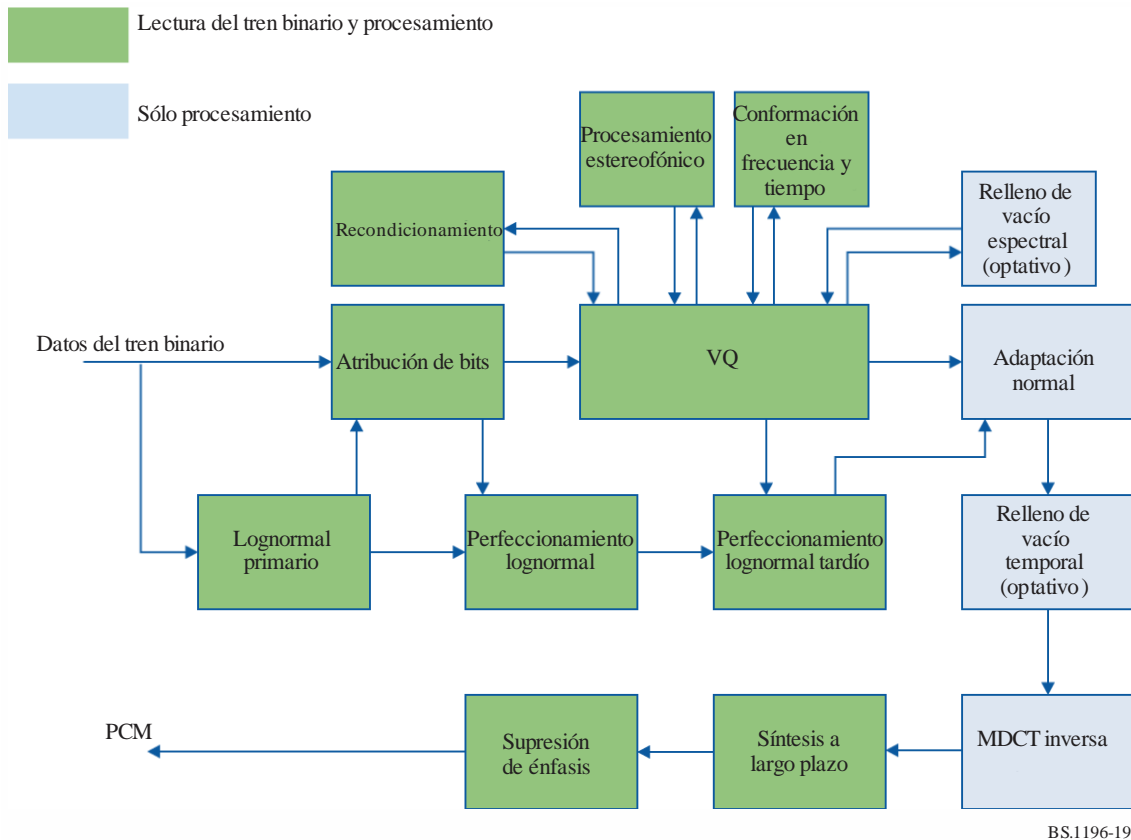
El codificador DTS-UHD permite controlar totalmente la situación de las tramas con sincronismo para permitir al servicio controlar los puntos de entrada en el servicio de audio.

La funcionalidad velocidad binaria variable (VBR) DTS-UHD garantiza una velocidad binaria constante (CBR) en cada grupo de tramas (GoF – intervalo entre dos tramas con sincronismo consecutivas). El codificador configura las tramas con sincronismo y el intervalo de sincronismo puede modificarse dinámicamente durante el proceso de codificación en función de las necesidades. El codificador DTS-UHD también puede configurar la velocidad de cresta VBR instantánea máxima, según sea necesario. El modelo de control de velocidad binaria (BRC) DTS-UHD atribuye dinámicamente bits a cada trama de audio en los objetos, canales y pares de canales dentro de un tren binario. El BRC dirige más bits para codificar el contenido más perceptualmente relevante de cada trama con el objetivo de que la calidad de audio sea igual en todos los objetos y canales de una trama. De este modo se logra una mejor eficacia de codificación con un mayor número de elementos de audio al aprovechar la distribución variable del contenido perceptualmente relevante entre objetos y canales. El codificador también atribuye dinámicamente bits a las tramas de audio con el objetivo de lograr una calidad de audio percibida constante a lo largo del eje temporal (dentro de los límites impuestos por el CMBR FoF y la velocidad binaria de cresta máxima especificada). Téngase en cuenta que la funcionalidad BRC no necesita que se transmita información adicional y que el decodificador ignora la funcionalidad BRC. Cada objeto, canal o par de canales puede decodificarse de manera independiente y «totalmente ejecutable» en el modo BRC.

3 Descodificador

El decodificador ACE se aplica a los trenes ACE estereofónicos, los trenes ACE monofónicos y los trenes ACE LFE. Dentro de una trama ACE los trenes pertenecen a una de esas tres categorías. Los trenes estérofónicos poseen una optimización adicional que aprovecha la información común a ambos canales.

FIGURA 19
Descodificador ACE



BS.1196-19

La descodificación de un tren ACE monofónico o estereofónico se efectúa en las siguientes fases:

Lognormal primario: se extrae del tren binario una primera aproximación de la potencia de banda espectral. Esos valores de potencia se representan en el dominio logarítmico y se denominan valores lognormales.

Atribución de bits: para cada banda espectral se calcula el número de bits disponibles para la descodificación de banda espectral normalizada. Este cálculo depende de la información del tren binario y (potencialmente) de los valores lognormales primarios.

Perfeccionamiento lognormal: se utiliza una pequeña parte de la atribución de bits para perfeccionar los valores lognormales.

VQ: (Cuantización vectorial): se utiliza la mayoría de la atribución de bits para reconstruir los valores de frecuencia espectral normalizados, banda espectral por banda espectral.

Recondicionamiento: en el caso de bandas espectrales originalmente «completas» y valores de atribución de bits pequeños, las bandas dispersas reconstruidas inicialmente se recondicionan para obtener una mejor aproximación del espectro completo.

Procesamiento estereofónico: para los trenes estereofónicos, dos canales reconstruidos se transforman en una representación original (sin mezcla).

Conformación en frecuencia y tiempo: se preprocesan las bandas espectrales con una transformada de Haar simple, reuniendo todas las bandas en la misma representación de tiempo y frecuencia.

Relleno de vacío espectral: las bandas con energía original reconstruidas a todo ceros se rellenan con «ruido» generado a partir de valores de banda reconstruida anteriores.

Perfeccionamiento lognormal tardío: se utilizan todos los bits restantes tras la VQ para efectuar un último perfeccionamiento lognormal.

Adaptación normal: se restauran las bandas espectrales a los valores de potencia correspondientes a los valores lognormales definitivos.

Relleno de vacío temporal: las bandas espectrales todo ceros pueden rellenarse con ruido sintético, en función de la potencia de banda espectral de las tramas anteriores.

MDCT inversa: se transforma la representación espectral en una representación en el dominio temporal (PCM).

Síntesis a largo plazo: se utiliza un filtro de síntesis a largo plazo para reconstruir los componentes de la señal eliminados durante la codificación.
