

Unión Internacional de Telecomunicaciones

UIT-R

Sector de Radiocomunicaciones de la UIT

Recomendación UIT-R BS.1196-2
(03/2010)

**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 el Anexo 1 a 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, 2010

© UIT 2010

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.1196-2* **

Codificación de audio para la radiodifusión digital

(Cuestión UIT-R 19/6)

(1995-2001-2010)

Cometido

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

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 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;
- d) que la evaluación subjetiva de los sistemas de audio de calidad de audio intermedia se trata en la Recomendación UIT-R BS.1534 (MUSHRA);
- e) que la codificación a baja velocidad binaria para el audio de alta calidad ha sido probada por el Sector de Radiocomunicaciones de la UIT;
- f) 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;
- g) que varios 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;
- h) 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;
- j) que los organismos de radiodifusión que no han iniciado aún los servicios deben poder elegir el sistema más adecuado para su aplicación;

* La Comisión de Estado 6 de Radiocomunicaciones efectuó modificaciones de redacción en esta Recomendación en 2003 de conformidad con la Resolución UIT-R 44.

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

- 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»),

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:

- MPEG-4 HE AAC v2 como especifica la Norma ISO/CEI 14496-3:2009;
- E-AC-3 como especifica la Norma ETSI TS 102 366 (2008-08);

NOTA 1 – MPEG-4 HE AAC v2 y E-AC-3 son superconjuntos más flexibles de MPEG-4 AAC-LC y AC-3.

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 (2008-8);

NOTA 1 – La Norma ISO/CEI 11172-3 puede citarse a veces como Norma 13818-3 pues esta especificación incluye la Norma 11172-3 por referencia.

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 1 – 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 la codificación ISO/CEI 11172-3 Capa II con una velocidad binaria de al menos 180 kbits/s 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;

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;

8 que el cumplimiento de lo dispuesto por esta Recomendación sea voluntario. Sin embargo, la Recomendación puede contener ciertas disposiciones obligatorias (por ejemplo, para garantizar la interoperabilidad o la aplicabilidad) y el cumplimiento de la Recomendación se logra cuando estas disposiciones obligatorias se satisfacen. La redacción del texto en el sentido de obligatoriedad y sus equivalentes negativos se emplea para expresar los requisitos. La utilización de esta redacción no debe interpretarse en modo alguno como el cumplimiento parcial o total de esta Recomendación,

recomienda además

1 que se haga referencia a 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.

NOTA 1 – La información sobre los códecs incluida en esta Recomendación figura en los Apéndices 1 a 4.

Apéndice 1

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

1 Codificación

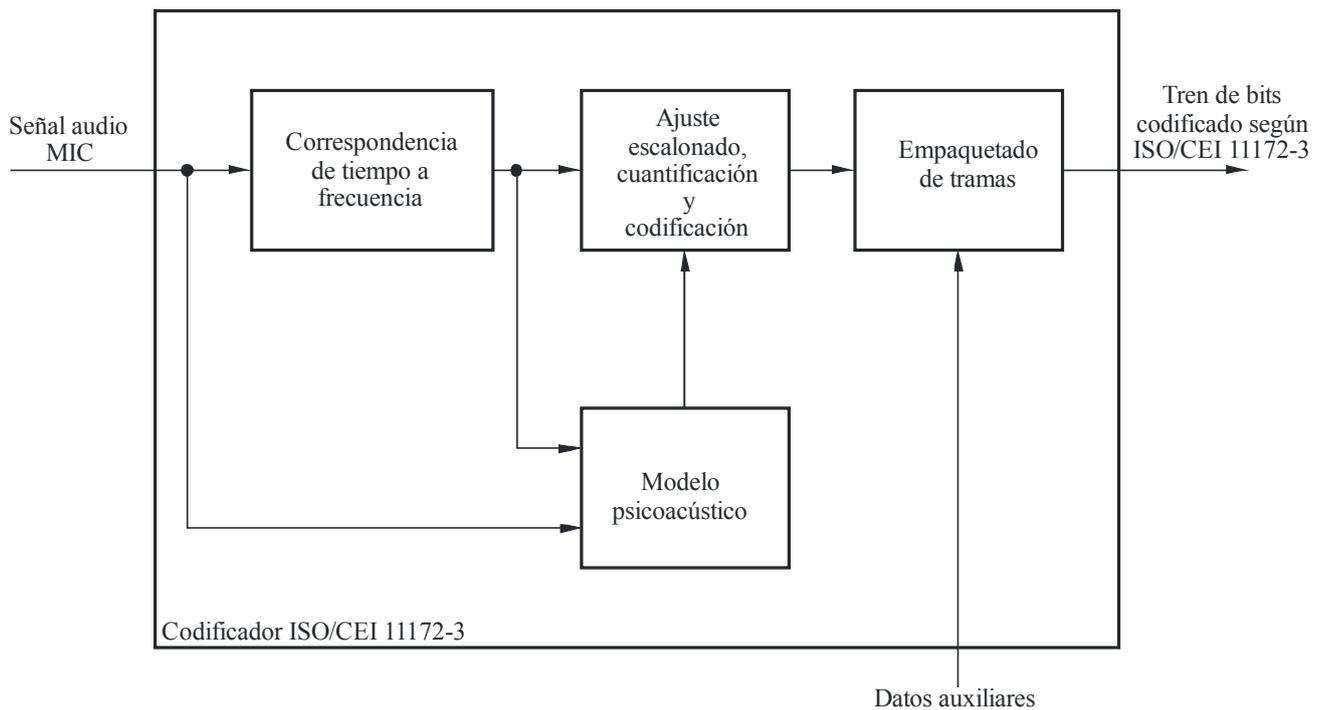
El codificador procesa la señal de sonido digital y produce el tren binario comprimido. El algoritmo del codificador no está normalizado y puede utilizar diversos medios de codificación tales como el de estimación de umbral de enmascaramiento del auditorio, la cuantificación y el ajuste escalonado (scaling) (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 submustrada del tren de audio de entrada. Las muestras sometidas a la correspondencia pueden denominarse muestras de sub-bandas (como en las Capas I o II, según se indica más adelante) o muestras de sub-bandas 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 sub-banda 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



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 sub-bandas 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 sub-bandas 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.

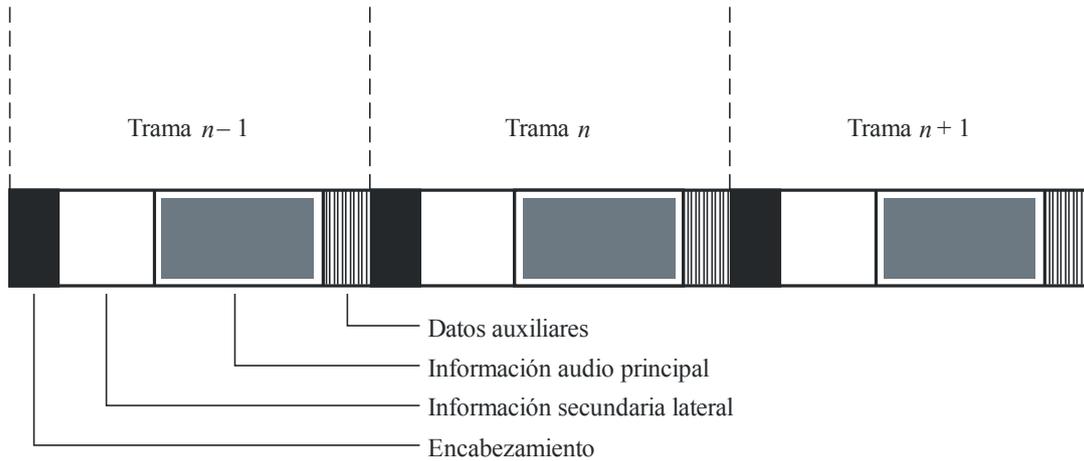
Hay cuatro modos distintos posibles para cualquier capa:

- 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:

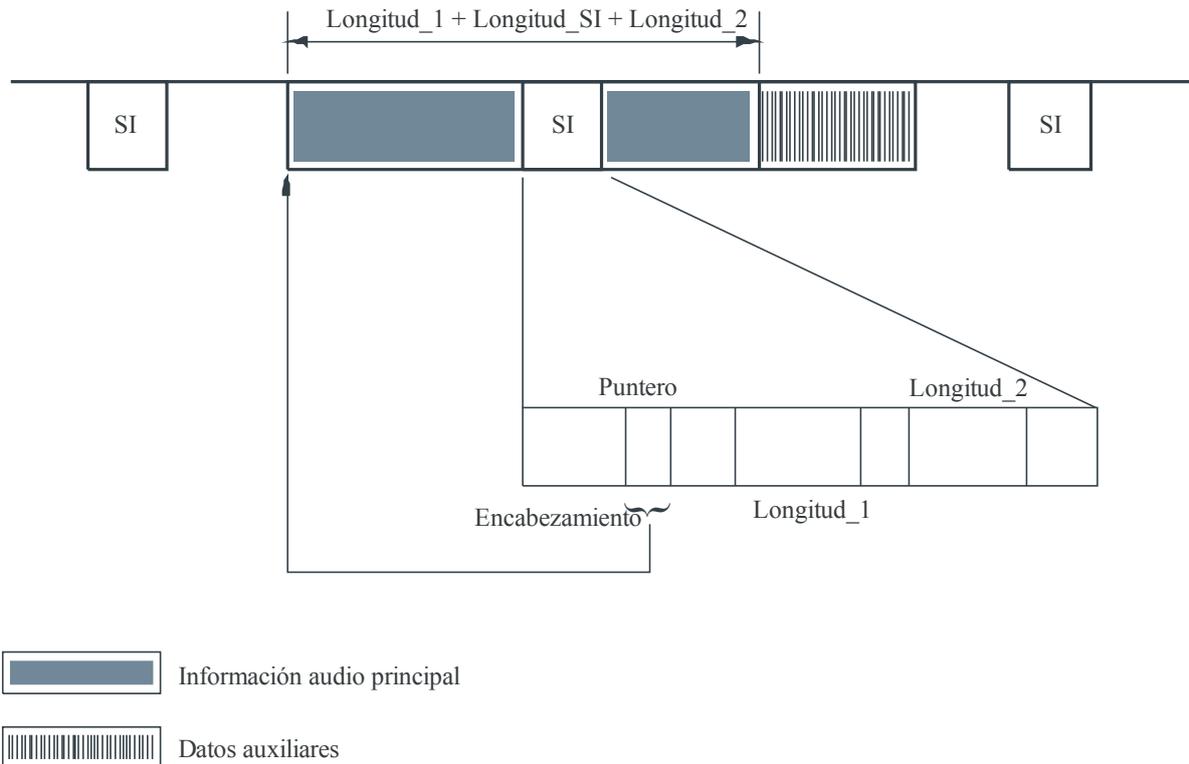
FIGURA 2
Formato del tren de bits ISO/CEI 11172-3 de Capa II



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 sub-banda
- 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.

1196-03

4 Decodificación

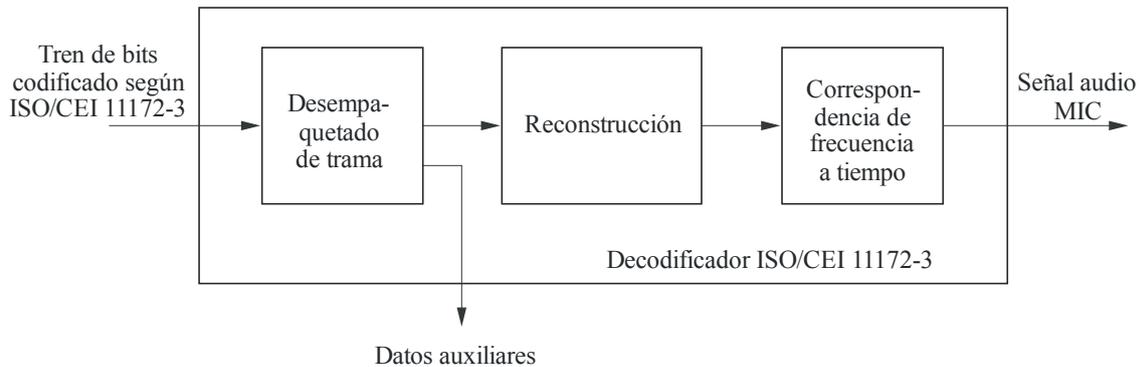
El decodificador acepta el tren binario de audio comprimido en la sintaxis que define la ISO/CEI 11172-3, decodifica los elementos de datos y utiliza la información para producir una salida de audio digital.

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



1196-04

Apéndice 2

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 monocal, 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 de canal por defecto

Número de altavoces	Elementos audiosintácticos, listados en el orden en que se reciben	Correspondencia entre el elemento por defecto y el altavoz
1	single_channel_element	Altavoz frontal central
2	channel_pair_element	Altavoces frontales izquierdo y derecho
3	single_channel_element()	Altavoz frontal central
	channel_pair_element()	Altavoces frontales izquierdo y derecho
4	single_channel_element()	Altavoz frontal central
	channel_pair_element()	Altavoces frontales izquierdo y derecho
	single_channel_element()	Altavoz de ambiente posterior
5	single_channel_element()	Altavoz frontal central
	channel_pair_element()	Altavoces frontales izquierdo y derecho
	channel_pair_element()	Altavoces traseros de ambiente izquierdo y de ambiente derecho
5+1	single_channel_element()	Altavoz frontal central
	channel_pair_element()	Altavoces frontales izquierdo y derecho
	channel_pair_element()	Altavoces traseros de ambiente izquierdo y de ambiente derecho
	Lfe_element()	Altavoz de efectos de baja frecuencia
7+1	single_channel_element()	Altavoz frontal central
	channel_pair_element()	Altavoces frontales centrales izquierdo y derecho
	channel_pair_element()	Altavoces frontales externos izquierdo y derecho
	channel_pair_element()	Altavoces traseros de ambiente izquierdo y de ambiente derecho
	lfe_element()	Altavoz de efectos de baja frecuencia

2 Codificación

En la Fig. 5 se muestra la estructura básica del codificador AAC con MPEG-2. El sistema AAC consta de 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.

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.

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.

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

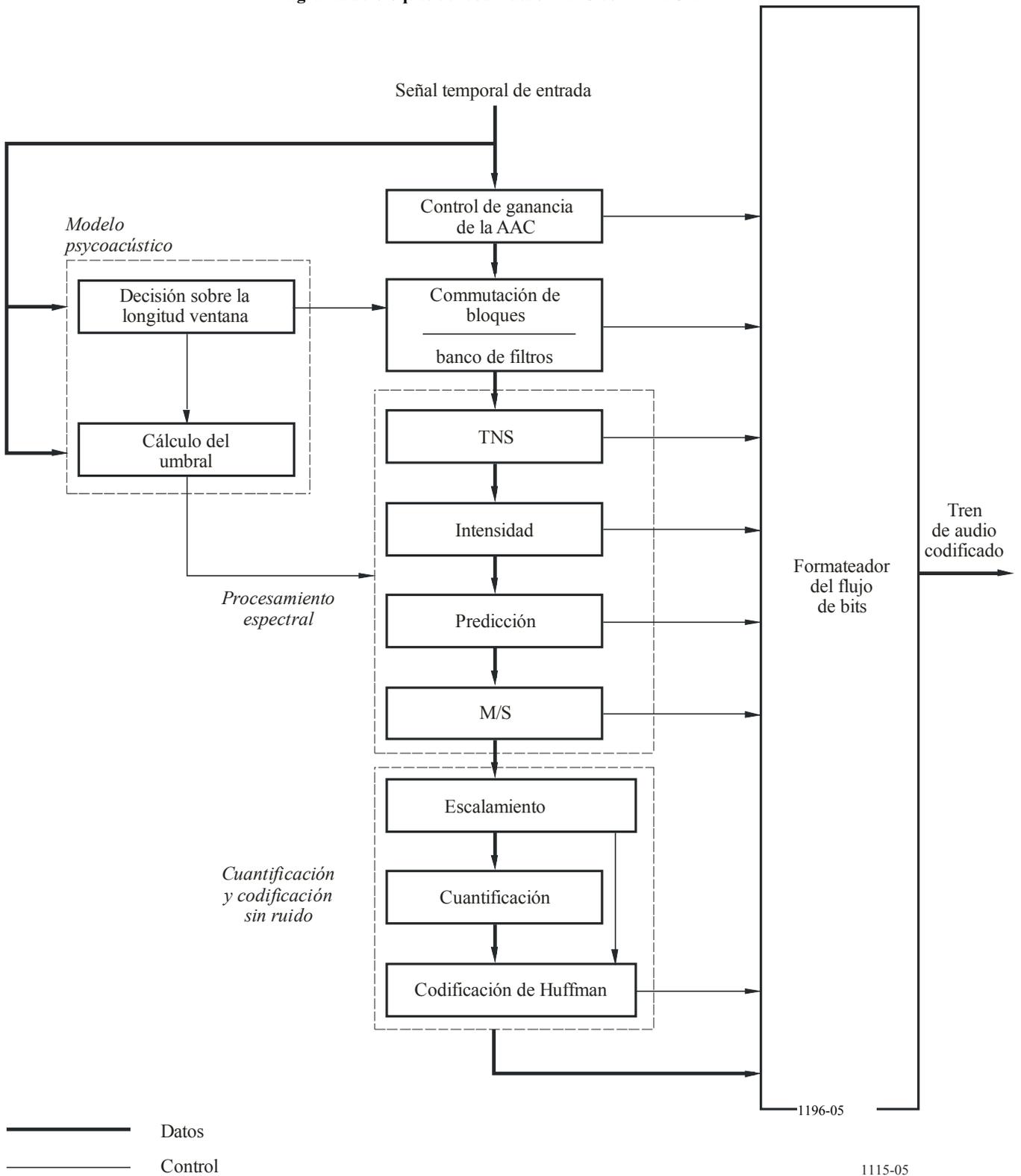
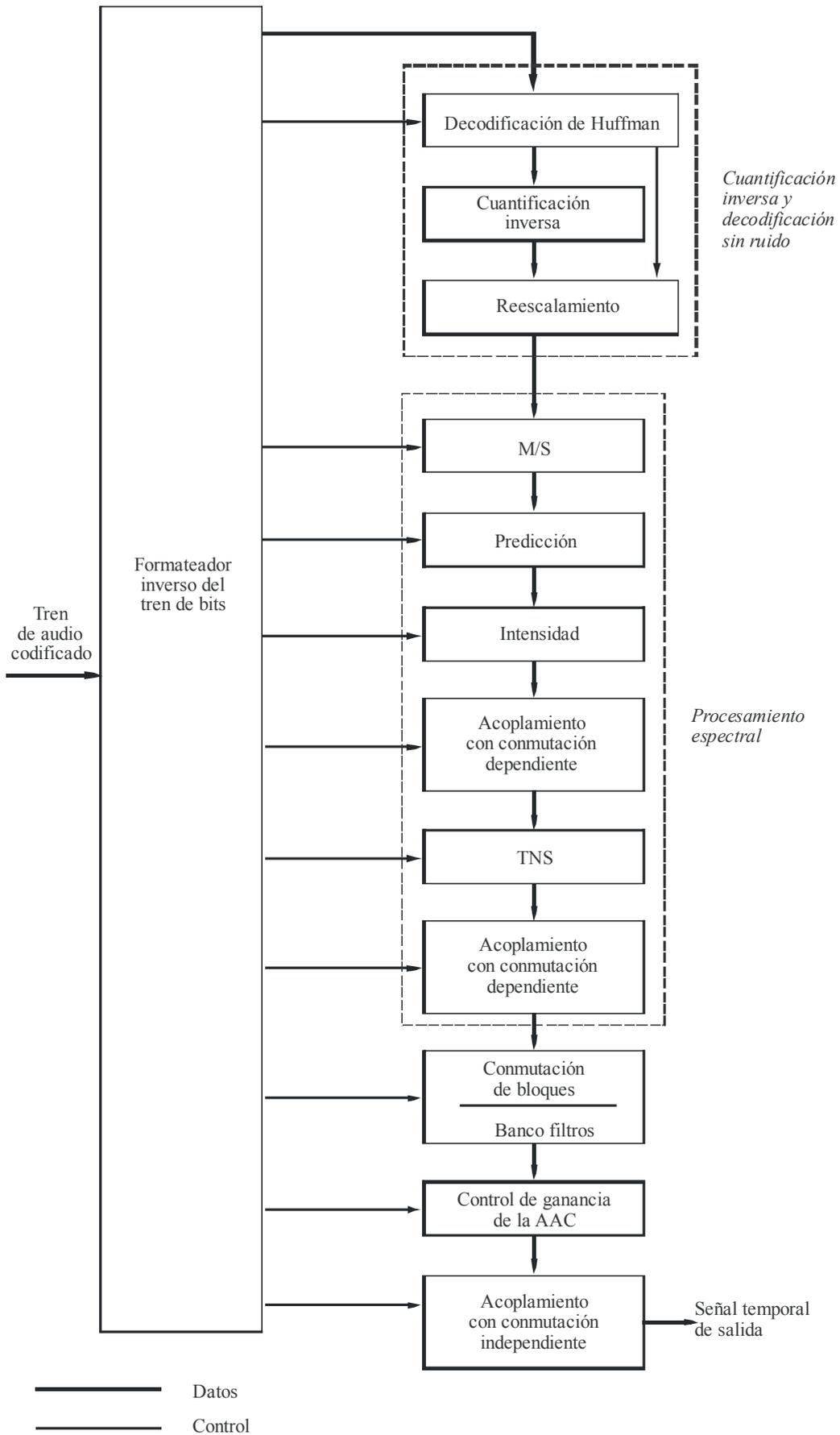


FIGURA 6

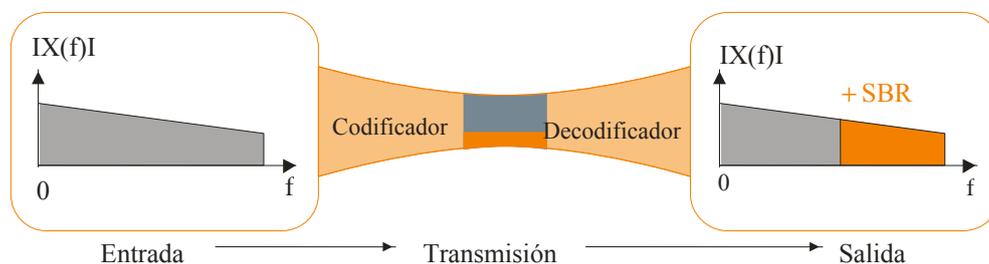
Diagrama de bloques del decodificador AAC con MPEG-2



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.



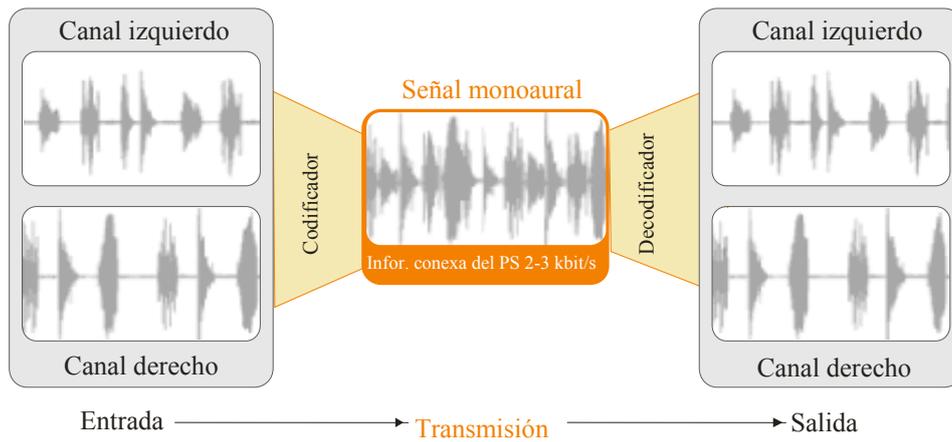
1196

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.



1196

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.

Apéndice 3

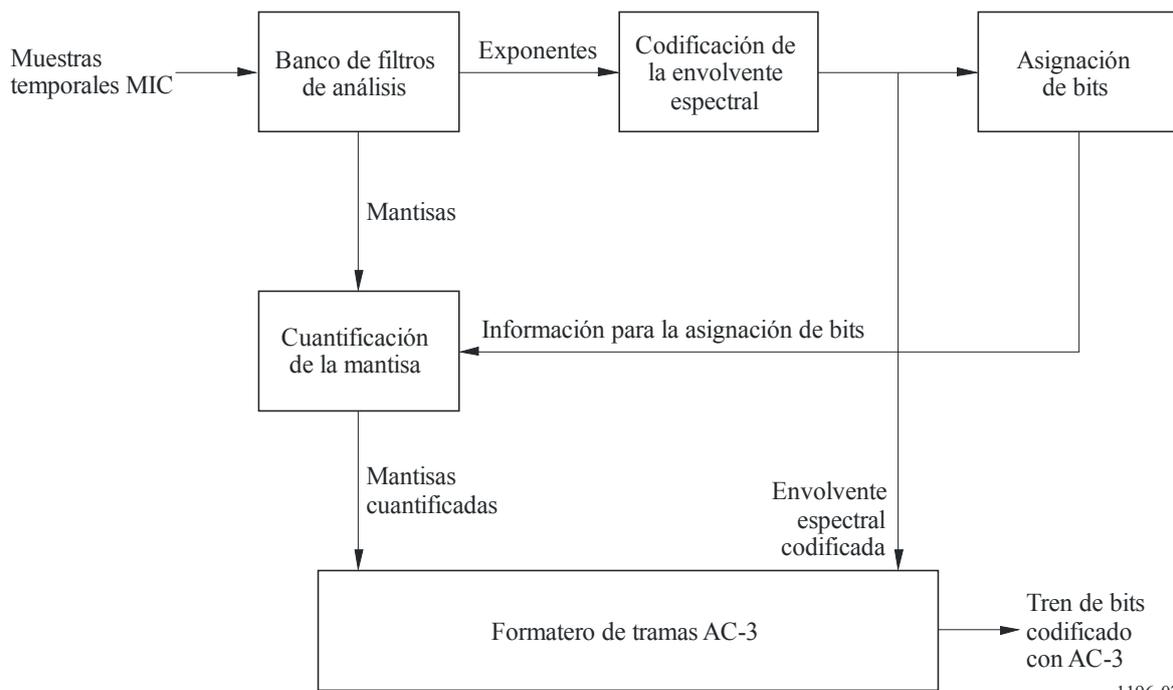
Audio AC-3 y E-AC-3

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



1196-07

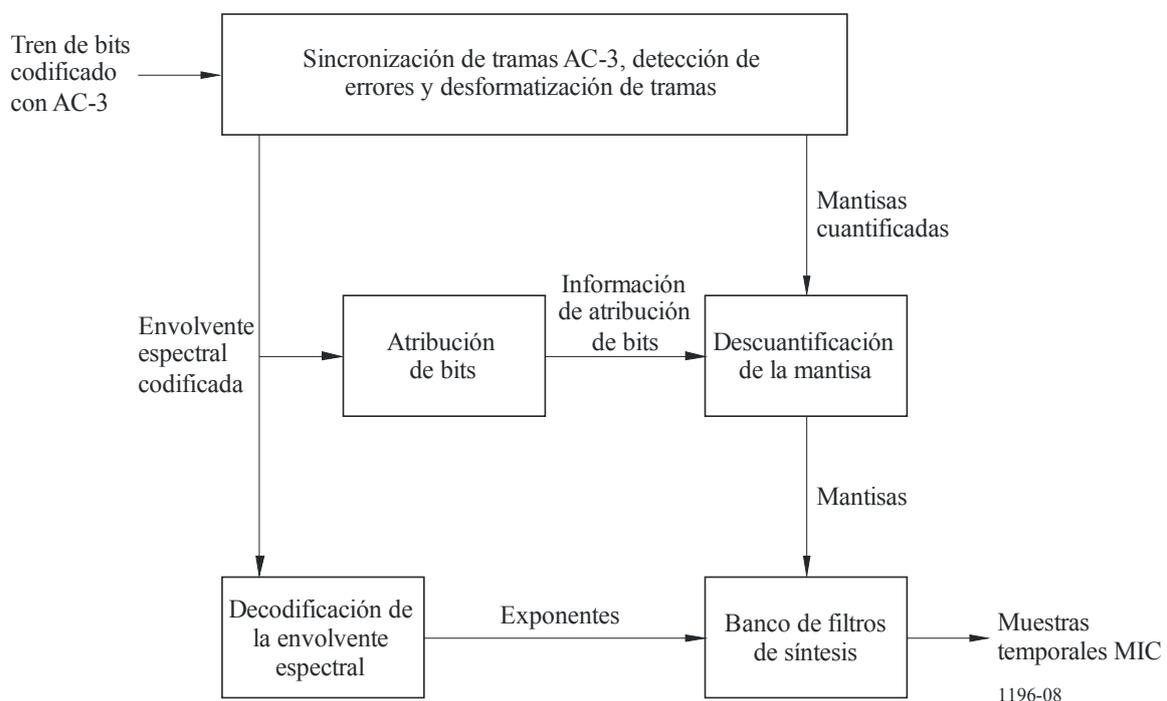
El codificador AC-3 real es más complejo de lo que muestra la Fig. 7. También incluye las siguientes funciones no indicadas anteriormente:

- 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



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.

Apéndice 4

MPEG Surround

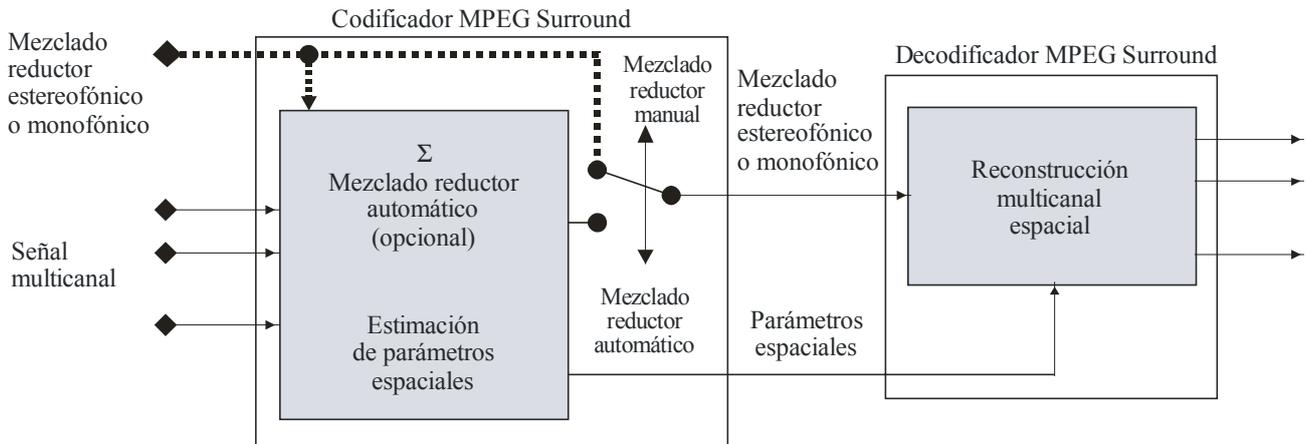
1 Introducción

La tecnología ISO/CEI 23003-1 o MPEG Surround 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.

FIGURA 9

Principio de MPEG Surround, el mezclado reductor se codifica utilizando un códec de audio



1196-09

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 cuadros).

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.

Estas 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

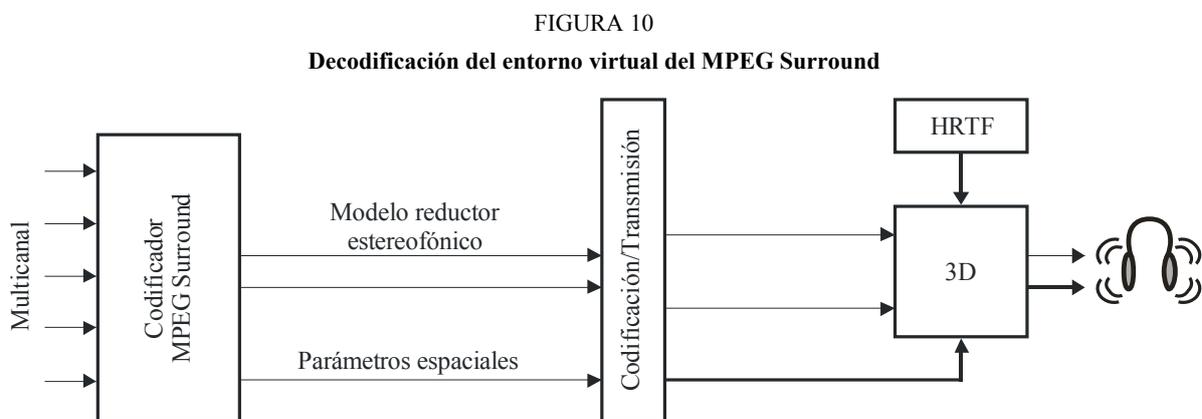
parte de las señales descorreladas (que forman parte de los recuadros estereofónicos paramétricos). Las señales residuales se codifican transformando las señales en el dominio QMF en señales en el dominio MDCT, tras lo cual se codifican los coeficientes MDCT empleando AAC.

- 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.



- 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 y 2 son capaces de decodificar todos los trenes de bits de Nivel 2 y 3, 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).
