

RECOMENDACIÓN UIT-R BS.1115-1^{*, **}
Codificación de audio a baja velocidad binaria
(Cuestión UIT-R 19/6)

(1994-2005)

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que hay servicios de sonido digital de gran calidad a disposición del consumidor a través de medios tales como el disco compacto (CD), la cinta magnética y la radiodifusión;
- b) que en las Recomendaciones UIT-R BS.774 y UIT-R BO.789 figuran las bases de un sistema de radiodifusión sonora digital;
- c) que en la Recomendación UIT-R BS.1548 se especifican los requisitos del usuario en cuanto a sistemas de codificación de audio para la radiodifusión digital;
- d) que la calidad de la imagen de audio básica y estereofónica necesaria para la radiodifusión sonora digital ha de ser prácticamente indistinguible de la del CD;
- e) que las características de los sistemas de codificación a baja velocidad binaria utilizados en las conexiones de programa (enlaces de contribución, distribución y comentarios) que alimentan los sistemas de radiodifusión sonora digital deben ser tales que se puedan aplicar señales de sonido de óptima calidad al emisor de radiodifusión sonora digital;
- f) que, en el caso de los enlaces de contribución y distribución, esto implica un nivel de calidad superior al del CD debido a los requisitos en cuanto a márgenes de procesamiento y de conexión en cascada y (para el caso de los enlaces de contribución) al margen de sobrecarga u «holgura»;
- g) que, en el caso de los enlaces de comentarios, esto implica un nivel de calidad que sea capaz de dar señales vocales de calidad excelente al oyente, pero que cabe esperar un nivel de calidad inferior cuando esos enlaces cursen programas musicales;
- h) que la codificación a baja velocidad binaria produce generalmente un retardo de varias décimas de milisegundo, lo que puede tener consecuencias en las prácticas operacionales, como la presentación de indicaciones que no salen a antena;
- j) que el Sector de Radiocomunicaciones de la UIT ha probado la codificación a baja velocidad binaria para las señales de sonido de gran calidad y ha obtenido una calidad satisfactoria en diversas aplicaciones (véanse el Apéndice 6 y la Recomendación UIT-R BS.1548);
- k) que los métodos objetivos convencionales (por ejemplo, los que miden la relación señal/ruido y la distorsión) pueden ya no ser adecuados para evaluar la calidad de los sistemas con

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

** Esta Cuestión debe señalarse a la atención de las Comisiones de Estudio 9 y 15 del Sector de Normalización de las Telecomunicaciones, de la Organización Internacional de Normalización/Comisión Electrotécnica Internacional (ISO/CEI) (JTC1/SC29/WG11), de la Unión Europea de Radiodifusión (UER) y de la Audio Engineering Society (AES).

esquemas de codificación del sonido a baja velocidad binaria, y que el método objetivo perceptivo de medición indicado en la Recomendación UIT-R BS.1387 podría complementar o sustituir los métodos de evaluación objetivos convencionales y complementar los métodos subjetivos de evaluación,

recomienda

1 que para las aplicaciones de radiodifusión sonora enumeradas en los *recomienda* 4 a 6, se utilice la codificación del sonido a baja velocidad binaria que se describe en las Normas ISO/CEI 11172-3:1993 (véase la Nota 1), ISO/CEI 13818-7:2003 o AC-3 de acuerdo con el Anexo 2 de la Recomendación UIT-R BS.1196 (en los Apéndices 1, 3 y 4 figuran descripciones breves de estas Normas);

NOTA 1 – Algunas veces se puede hacer referencia al códec ISO/CEI 11172-3 como 13818-3 ya que éste incluye por referencia a la Norma 11172-3.

2 que para todas las aplicaciones, la frecuencia de muestreo sea de 48 kHz o 32 kHz;

3 que las señales de entrada del codificador de sonido de baja velocidad binaria no tengan acentuación y que el codificador no aplique acentuación;

4 que para la emisión de radiodifusión sonora digital (véase la Nota 1), se utilicen los siguientes códecs:

- ISO/CEI 11172-3 de capa II con una velocidad binaria de 128 kbit/s para una señal monoaural y de 256 kbit/s (es decir, 2×128 kbit/s) para una señal estereofónica, con una codificación independiente de las componentes izquierda y derecha de la señal estereofónica (véase la Nota 2);
- codificación ISO/CEI 13818-7 a una velocidad binaria de al menos 144 kbit/s para una señal estereofónica;
- codificación AC-3 a una velocidad binaria de al menos 192 kbit/s para una señal estereofónica.

NOTA 1 – En el Apéndice 1 del Anexo 2 de la Recomendación UIT-R BS.1548 se presenta información acerca de sistemas de codificación que han demostrado cumplir los requisitos de calidad y otros requisitos del usuario para la emisión.

NOTA 2 – Se prevé que los organismos de radiodifusión desearán disponer de cierta capacidad para datos auxiliares dentro de la señal emitida. Pero deberán tener en cuenta que toda reducción de la velocidad binaria de audio afectará en general la calidad del sonido.

5 que para los enlaces de distribución y contribución se utilice la codificación ISO/CEI 11172-3 de capa II con una velocidad binaria de 180 kbit/s, al menos, por señal de sonido (es decir, por señal monoaural o por componente de una señal estereofónica codificada independientemente), excluyendo los datos auxiliares (véanse el Apéndice 2 y la Nota 2), pero en el caso especial de un solo enlace de distribución sin otras conexiones en tándem, se podrá utilizar una velocidad binaria de al menos 120 kbit/s por señal audio (véase la Nota 3);

NOTA 1 – En el Apéndice 1 del Anexo 1 a la Recomendación UIT-R BS.1548 se presenta información acerca de sistemas de codificación que han demostrado cumplir los requisitos de calidad y otros requisitos del usuario para la contribución y la distribución.

NOTA 2 – Al aplicar esta Recomendación debe tenerse en cuenta que algunas redes de telecomunicaciones no permiten la utilización sin restricción de los canales a 64 kbit/s.

NOTA 3 – Cuando, por ejemplo, un transmisor analógico es alimentado por un solo enlace de distribución, o cuando una señal codificada según la norma de emisión final del estudio se envía al transmisor para su emisión sin ninguna codificación o decodificación ulterior.

6 que para los enlaces de comentarios, se utilice la codificación ISO/CEI 11172-3 de Capa III con una velocidad binaria de 60 kbit/s, al menos, excluyendo los datos auxiliares para las señales

monoaurales (véase la Nota 1), y 120 kbit/s al menos, excluyendo los datos auxiliares para las señales estereofónicas y utilizando codificación estereofónica mixta (véanse el Apéndice 2 y la Nota 2);

NOTA 1 – Las pruebas realizadas en 1993 con cascos de escucha para evaluar la calidad de un códec de comentarios monoaural con codificación de capa III a 60 kbit/s indicaron que no se consigue la calidad requerida con una señal vocal. Esta Nota se mantendrá en la Recomendación hasta que la calidad del códec de comentarios satisfaga los requisitos.

NOTA 2 – Al aplicar esta Recomendación se debe tener en cuenta que algunas redes de telecomunicación no permiten la utilización sin restricción de los canales a 64 kbit/s.

7 que el retardo se reduzca al mínimo en la implementación de los codificadores y decodificadores.

NOTA 1 – Las Normas ISO/CEI 11172-3 (MPEG-1 audio) y 13818-7 (MPEG-2 AAC) se encuentran disponibles en versión electrónica en la siguiente dirección: <http://www.iso.org/itu>.

Apéndice 1

Descripción del método de codificación ISO/CEI 1172-3 Codificación de imágenes en movimiento y de su correspondiente audio para medios de almacenamiento digital de hasta unos 1,5 Mbit/s – Parte 3: Audio

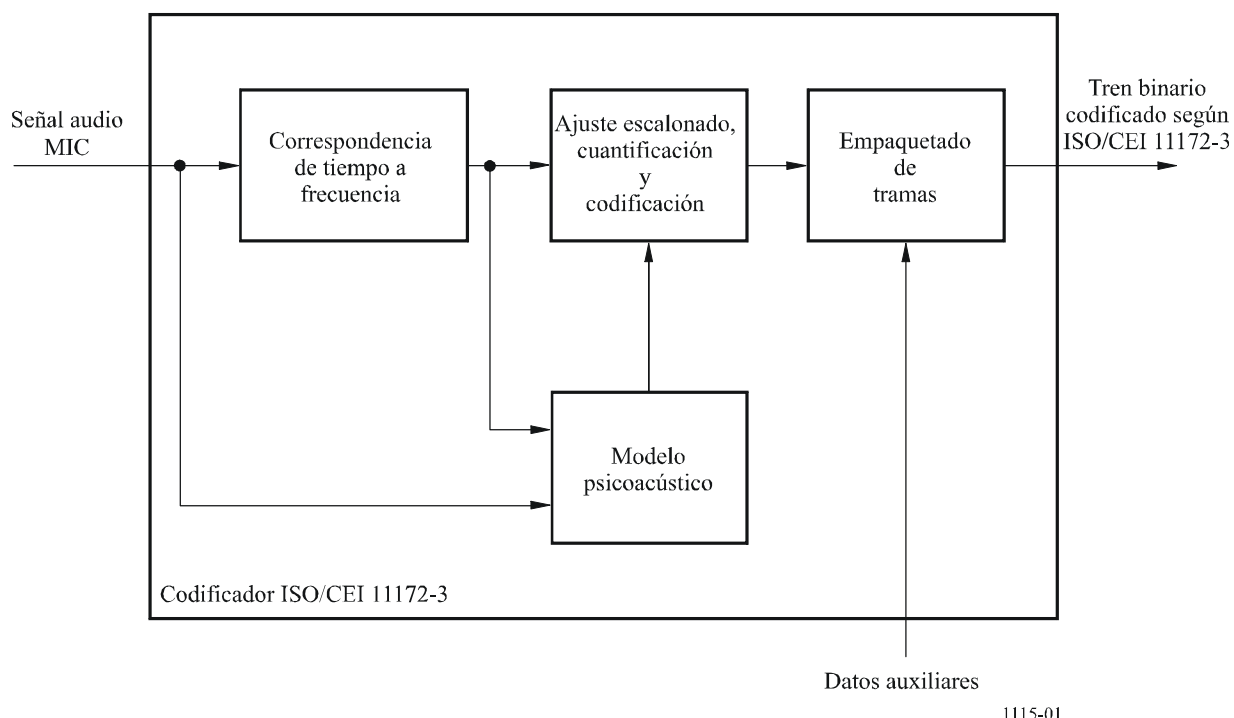
1 Codificación

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



1115-01

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 binario codificado

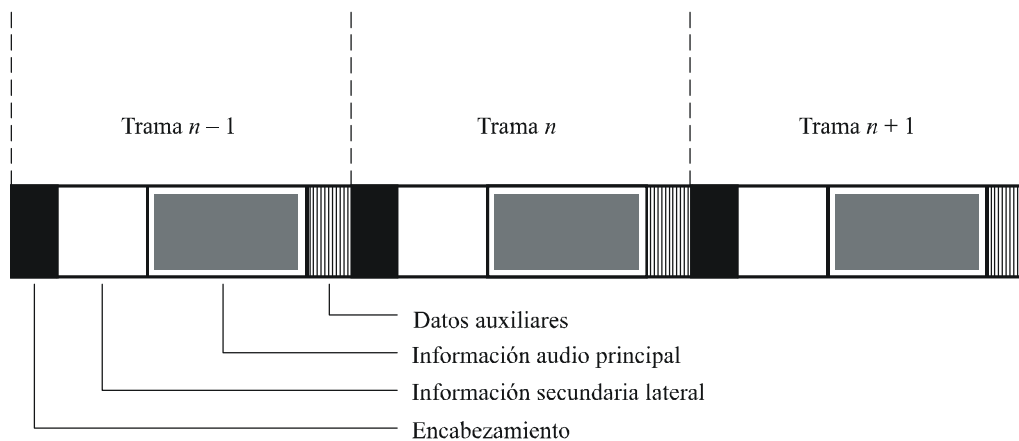
La Fig. 2 ofrece una panorámica del tren binario ISO/CEI 11172-3 para la Capa II, y la Fig. 3, para la Capa III. Un tren binario codificado se compone de tramas consecutivas. Según la capa, una trama contiene los campos siguientes:

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 binario 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.

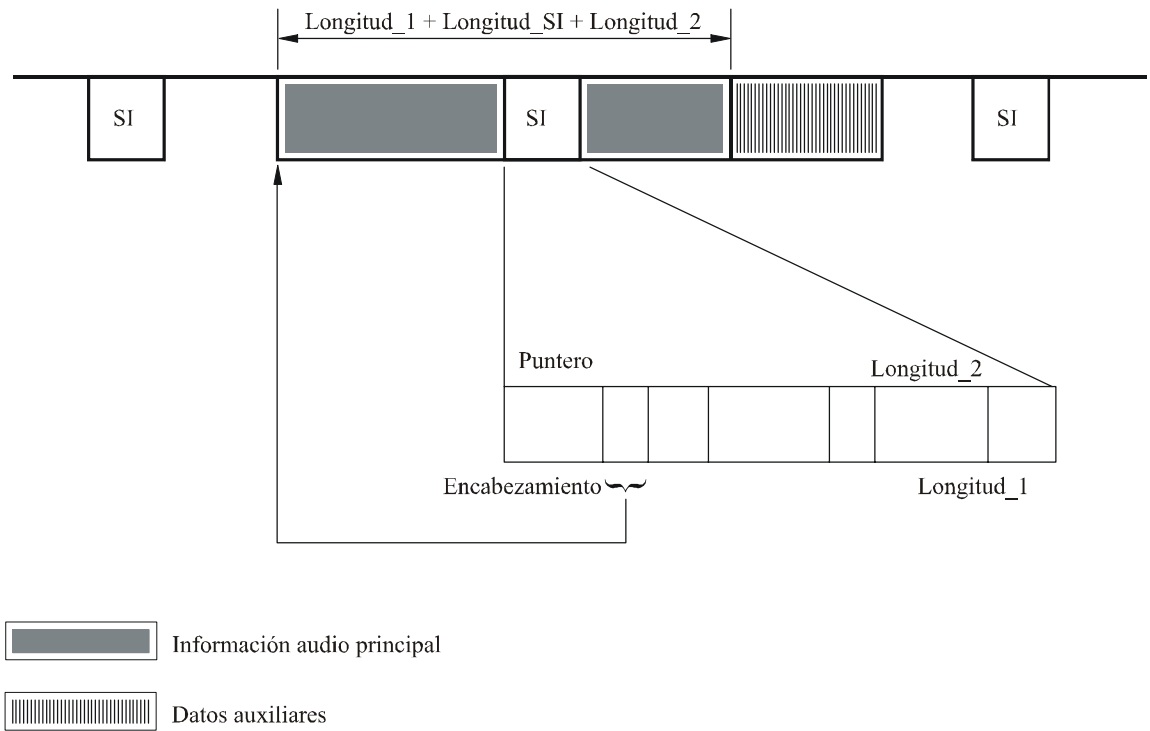
FIGURA 2
Formato del tren binario ISO/CEI 11172-3 de Capa II



Capa II:

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

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



 Información audio principal

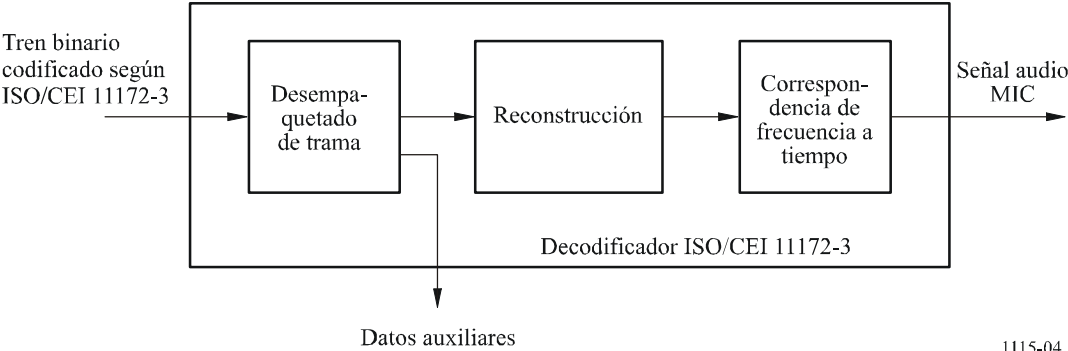
 Datos auxiliares

Capa III:

Información secundaria lateral (SI):	parte del tren binario 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 binario 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 binario que contiene la señal de audio codificada
Datos auxiliares:	parte del tren binario que contiene datos definibles por el usuario

1115-03

FIGURA 4
Diagrama de bloques del decodificador



1115-04

Apéndice 2

Directrices para la utilización de códecs ISO/CEI 11172-3 a velocidades binarias específicas

La intención de este Apéndice es suministrar algunas directrices con respecto a las aplicaciones en las que sea necesario utilizar un códec ISO/CEI 11172-3 a una velocidad binaria específica no incluida explícitamente en el cuadro de velocidades binarias de la Norma ISO/CEI (véase el Cuadro 1). Para elegir esas velocidades binarias específicas puede ser necesario tener en cuenta la velocidad binaria disponible en el canal de transmisión y la correspondiente velocidad binaria mínima indicada en la Recomendación.

Aparece un análisis más completo en la Recomendación UIT-T J.52 sobre transmisión digital de señales de radiodifusión sonora de alta calidad con uno, dos o tres canales a 64 kbit/s por señal monoaural (y hasta seis por señal estereofónica).

Las siguientes velocidades binarias están incluidas explícitamente en la Norma ISO/CEI:

CUADRO 1

Capa II (kbit/s)	Capa III (kbit/s)
32	32
48	40
56	48
64	56
80	64
96	80
112	96
128	112
160	128
192	160
224	192
256	224
320	256
384	320

Las cifras del Cuadro 1 indican la velocidad binaria total por programa audio, independientemente del modo. Existe un formato adicional, el formato libre, que puede utilizarse con velocidades binarias fijas no incluidas en el Cuadro, con un máximo de 384 kbit/s para la Capa II y de 320 kbit/s para la Capa III.

Una trama audio ISO/CEI 11172-3 de Capa II o Capa III siempre corresponde a 1 152 muestras de entrada MIC audio por canal. La duración de la trama es de 24 ms con una velocidad de muestreo de 48 kHz, y de 36 ms con una velocidad de muestreo de 32 kHz. Una trama siempre consiste en un número entero de bytes. El número de bytes se puede calcular multiplicando la duración de la trama por la velocidad binaria y dividiendo el resultado por 8. La velocidad binaria y la frecuencia de muestreo se indican en el encabezamiento que va al comienzo de la trama de audio.

Pueden utilizarse los tres métodos siguientes para obtener una velocidad binaria no enumerada en el Cuadro.

1 Formato libre

La condición de formato libre tiene que indicarse en el encabezamiento. En este caso, al comenzar a funcionar, el decodificador mide la distancia entre palabras de sincronización consecutivas tras lo cual puede utilizarse un procedimiento de sincronización por volante, como en el caso de una velocidad binaria definida. La longitud de la trama en bytes puede calcularse mediante la fórmula anterior. El codificador tendrá en cuenta esa longitud.

Por ejemplo, si se desea una velocidad binaria de 240 kbit/s para un programa estereofónico con una frecuencia de muestreo de 48 kHz, la longitud de la trama será $0,024 \times 240\,000/8 = 720$ bytes. En el caso de velocidades binarias que arrojen un número no entero de bytes en la trama, la velocidad binaria requerida puede implementarse mediante relleno. Los detalles de este procedimiento aparecen en el § 2.4.2.3 de la Norma ISO/CEI 11172-3.

Este método puede utilizarse para las Capas II y III.

2 Utilización del campo de datos auxiliares (para la Capa II)

Puede elegirse una de las velocidades binarias enumeradas (superior o igual a la requerida). Si se limita la asignación de bits, puede ajustarse el codificador para que reserve una cierta cantidad de bits para los datos auxiliares, de forma que la velocidad binaria de audio real sea igual a la requerida. Los datos auxiliares de la Capa II vienen al final de cada trama audio.

Los datos auxiliares se pueden suprimir o sobrescribir antes de la transmisión. Si se los suprime, hay que insertar la cantidad correspondiente de bits ficticios antes de la decodificación.

Por ejemplo, si se desea una velocidad binaria de 240 kbit/s para un programa estereofónico con una frecuencia de muestreo de 48 kHz, puede elegirse la velocidad de 256 kbit/s y, según la Norma ISO/CEI, indicarse en el encabezamiento de la trama. Hay que ajustar el codificador para que reserve 48 bytes por trama para datos auxiliares, lo que corresponde a una capacidad de datos auxiliares de 16 kbit/s.

3 Conmutación dinámica de la velocidad binaria (para la Capa III)

En la Capa III puede emplearse una conmutación dinámica de la velocidad binaria para obtener velocidades binarias no especificadas en el Cuadro 1. Por ejemplo, si se requiere una velocidad binaria de 120 kbit/s, un codificador de Capa III puede funcionar en un modo en el que se seleccionen en forma alternada las velocidades binarias de 128 kbit/s y 112 kbit/s, definidas explícitamente en la Norma ISO/CEI. Cuando el codificador funciona en este modo se lo debe fijar a un número de asignación de bits correspondiente al promedio de esas velocidades binarias. La técnica de reserva de bits de la Capa III mantiene una velocidad binaria constante para el proceso de codificación. Con este método, la distancia en bits entre palabras de sincronización toma dos valores en forma alternada.

Apéndice 3

Descripción del método de codificación de ISO/CEI 13818-7 Codificación genérica de imágenes en movimiento y de su correspondiente información de audio – Parte 7: Codificación de audio avanzada (AAC)

Véase la Norma ISO/CEI 13818-7.

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 2, y hasta con 48 canales de audio. En el Cuadro 3 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 2

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 3

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.
- *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

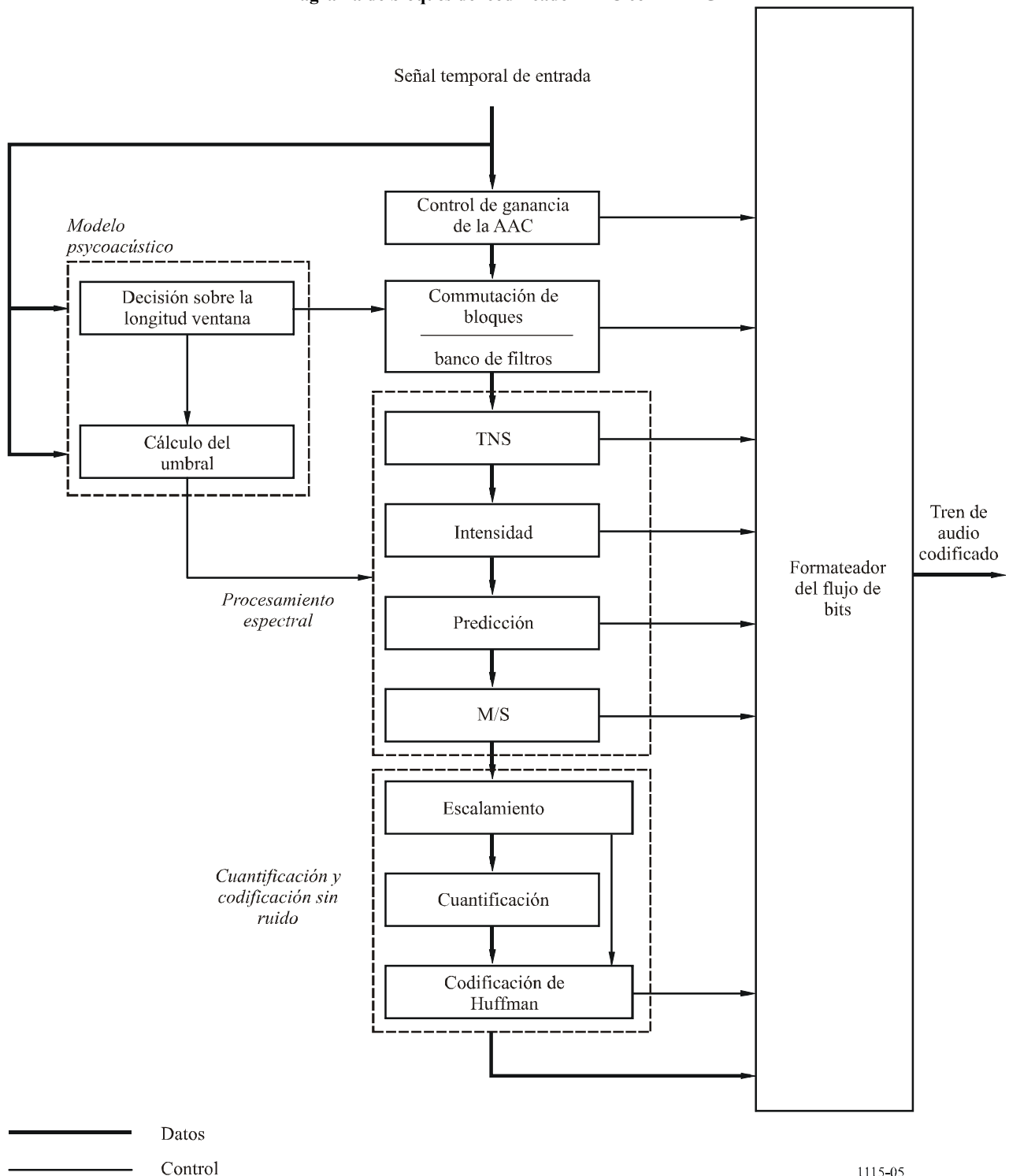
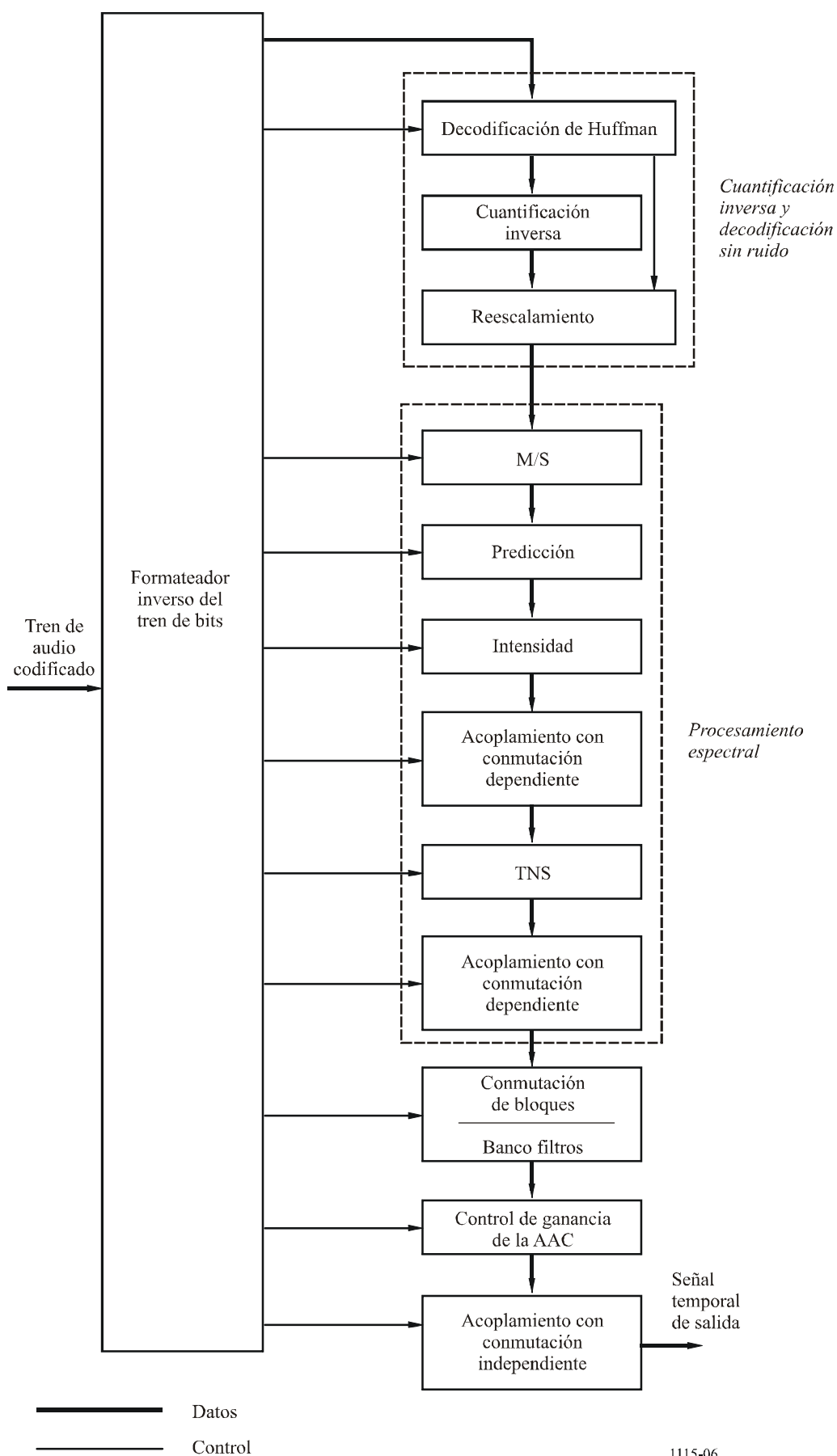


FIGURA 6

Díagrama de bloques del decodificador AAC con MPEG-2



Apéndice 4

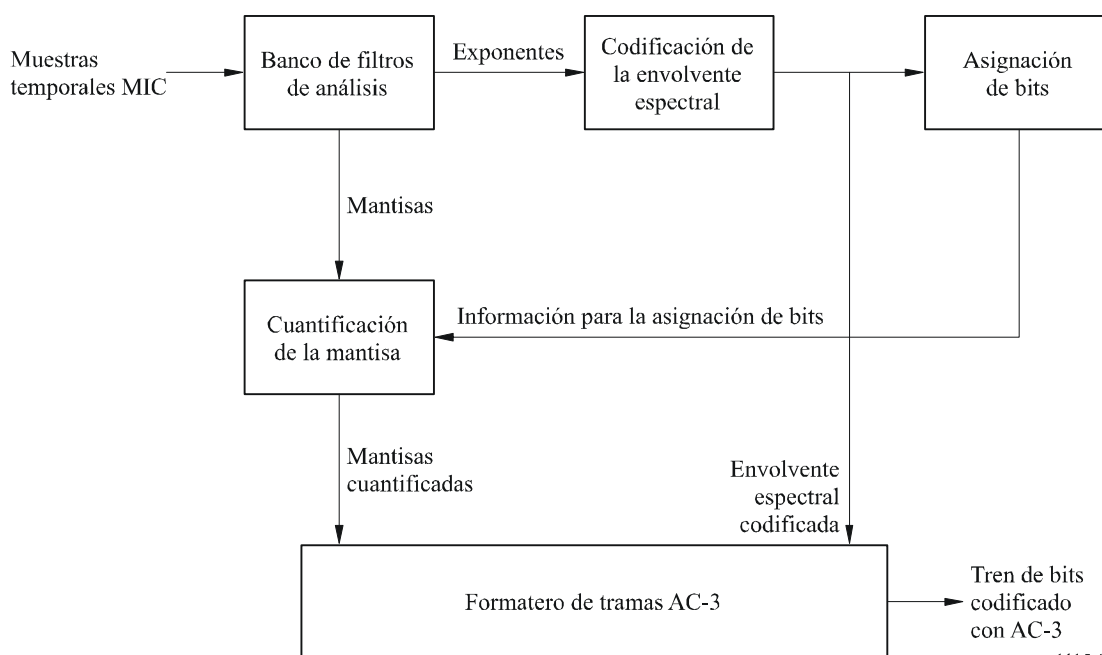
Descripción del método de codificación AC-3

Véase el Anexo 2 a la Recomendación UIT-R BS.1196 y el Anexo B a la Norma ATSC A/53B.

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



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.

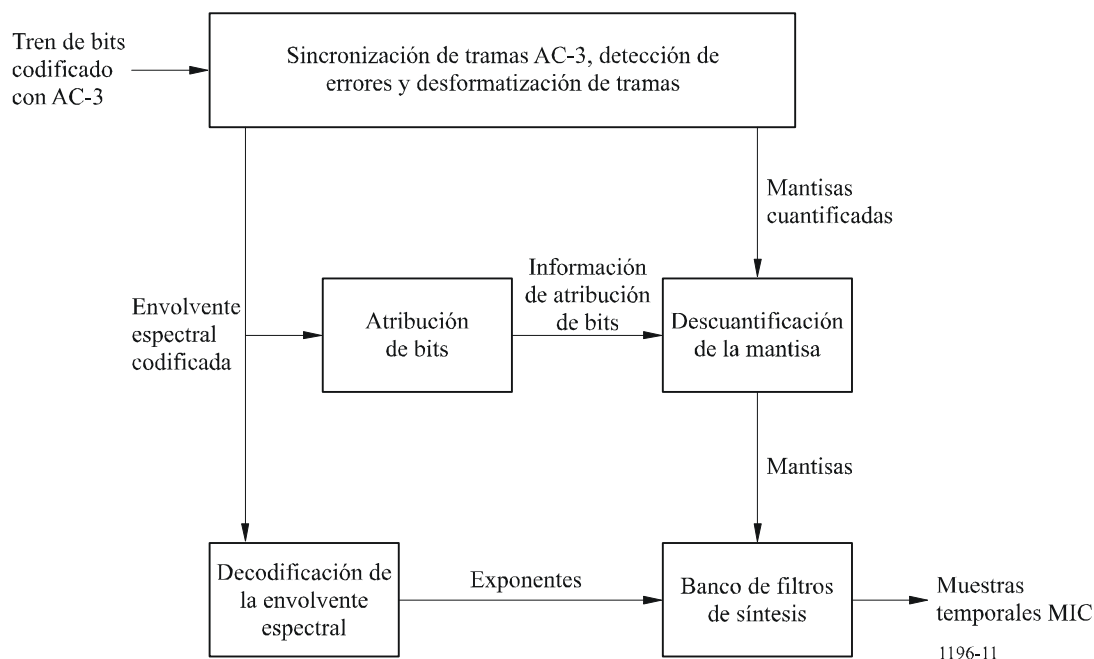
2 Decodificación

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.

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.

FIGURA 8
Decodificador AC-3



Apéndice 5

Métodos objetivos perceptuales

Los métodos objetivos perceptuales se han utilizado experimentalmente para medir la calidad de los códecs, y algunos resultados publicados muestran una buena correlación con los resultados de las pruebas subjetivas.

A continuación se indican los métodos utilizados experimentalmente por el Sector de Radiocomunicaciones:

- relación ruido a enmascaramiento (NMR, *noise to mask ratio*);
- modelo de evaluación perceptiva (PERCEVAL, *perceptual evaluation model*);
- medición perceptiva de la calidad audio (PAQM, *perceptual audio quality measure*);
- modelo objetivo perceptual (POM 620, *perceptual objective model*).

La Recomendación UIT-R BS.1387 especifica un método para llevar a cabo mediciones objetivas de la calidad de audio percibida.

Apéndice 6

Metodología para la selección de los códecs de audio de baja velocidad binaria recomendados en las pruebas realizadas entre 1992 y 1993

Se efectuaron pruebas subjetivas para varias aplicaciones monocanal y aplicaciones convencionales de dos canales, y los resultados fueron comparados con los requisitos. Se están realizando estudios adicionales sobre sistemas de codificación multicanal.

En las tres aplicaciones de emisión, distribución y contribución, la calidad de audio básica de los códecs reproducida tras la decodificación tiene que ser equivalente a la calidad de disco compacto, es decir, subjetivamente indistinguible de ésta, para la mayoría de los tipos de material de radiodifusión sonora. La calidad de la imagen sonora estereofónica debe mantenerse.

Para la emisión, el material más crítico para los códecs debe ser tal que la degradación sea «perceptible pero no molesta» (nota 4 de degradación). (Las pruebas subjetivas se realizaron utilizando la escala de degradación de cinco notas (Recomendación UIT-R BS.562).) En las pruebas realizadas en 1992, los sistemas indicados a continuación cumplieron estos requisitos:

Dos canales independientes de 128 kbit/s (véase la Nota 1):

- ISO/CEI 11172-3 Capa II
- ISO/CEI 11172-3 Capa III
- Dolby AC-2.

Par estereofónico de 192 kbit/s:

- ISO/CEI 11172-3 Capa II
- ISO/CEI 11172-3 Capa III.

En la distribución, la señal, tras pasar por tres códecs en tándem con 120 kbit/s por canal independiente, debe obtener una nota no más de 0,5 inferior a la de la señal de la fuente original (véase la Nota 1). La repoblación de los códigos binarios se efectuó aplicando una reducción de ganancia de 0,1 dB en el dominio MIC lineal. El códec ISO/CEI 11172-3 de Capa II fue el único sistema que satisfizo los requisitos y, por tanto, constituye la base de la Recomendación para esta aplicación.

NOTA 1 – A esta velocidad binaria los códecs obtuvieron notas de degradación 4,5 o superiores para la mayoría de los elementos críticos. A medida que evolucionen la metodología de prueba y los códecs, los ensayos futuros pueden arrojar notas distintas.

En la contribución, la señal, tras pasar por cinco códecs en tándem con 180 kbit/s por canal independiente, debe ser subjetivamente indistinguible de la señal de la fuente original. La conexión en cascada se realizó utilizando una resolución de 18 bits en MIC lineal. La repoblación de los códigos binarios se efectuó aplicando una resolución de ganancia de 0,1 dB en el dominio MIC lineal. El códec ISO/CEI 11172-3 de Capa II fue el único sistema que satisfizo los requisitos y, por tanto, constituye la base de la Recomendación para esta aplicación.

En los comentarios, la calidad de audio básica de las señales vocales reproducidas tras la decodificación debe ser equivalente a la del original MIC lineal de 14 bits. En las pruebas reales se utilizó un formato MIC lineal de 16 bits. Se admiten niveles perceptibles de degradación para secuencias musicales con calidad de disco compacto. El códec ISO/CEI 11172-3 de Capa III fue el único que actuó satisfactoriamente con señales vocales y fue, en promedio, el mejor para las señales musicales. En las señales vocales, las notas fueron siempre superiores a 4,0 (perceptible pero no molesta) en los modos monoaural (60 kbit/s) y estereofónico (estereofónico mixto de 120 kbit/s).

Este códec constituye, por tanto, la base de la Recomendación para esta aplicación. En las pruebas efectuadas en 1993, los resultados obtenidos con cascos de escucha para evaluar la calidad de un códec de comentarios monoaural con codificación de Capa III a 60 kbit/s indicaron (nota inferior a 4,0) que el mismo no cumple los requisitos de calidad con una señal vocal. Hacen falta más estudios a la velocidad de los 60 kbit/s.

En materia de complejidad, la evaluación indicó que los decodificadores ISO/CEI 11172-3 de Capa II y Dolby AC-2 eran los menos complejos. Basándose en las consideraciones siguientes:

- poca complejidad del decodificador;
- elementos comunes con las aplicaciones de distribución y contribución;
- flexibilidad para mejorar el codificador;
- códec ensayado a ambas velocidades binarias,

se eligieron el formato y el decodificador ISO/CEI 11172-3 de Capa II como base de la Recomendación para la aplicación de emisión a 2×128 kbit/s. En pruebas efectuadas en 1992, el modelo ISO/CEI 11172-3 de Capa II cumplió sólo marginalmente los requisitos a 192 kbit/s. Se decidió que se requieren mejoras a dicha velocidad binaria para cumplir claramente los requisitos de calidad audio básica (como el ISO/CEI 11172-3 de Capa III). Otras pruebas realizadas en 1993 no indicaron mejoras en esta etapa. Para que se vuelva a considerar la utilización de esta velocidad binaria, será preciso realizar más estudios a 192 kbit/s en modo estereofónico. Se solicita a las administraciones que presenten contribuciones sobre este tema.

En 1993 se realizaron pruebas de verificación de la red mediante una cadena de radiodifusión completa, con cinco códecs de contribución en tándem a 180 kbit/s, tres códecs de distribución en tándem a 120 kbit/s y un códec de emisión. Los resultados mostraron que la calidad audio básica en el extremo de dicha cadena no es satisfactoria.

También se realizaron pruebas con ocho códecs en tándem a 180 kbit/s, que demostraron que ésta es una configuración satisfactoria para mantener una calidad aceptable.
