

RECOMENDACIÓN UIT-R BS.1115<sup>\*,\*\*</sup>**Codificación del sonido a baja velocidad binaria**

(1994)

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 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;
- d) 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;
- e) 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»;
- f) 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;
- g) 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;
- h) que el Sector de Radiocomunicaciones de la UIT ha ensayado la codificación a baja velocidad binaria para las señales de sonido de gran calidad y ha obtenido una calidad satisfactoria en una serie de aplicaciones (véase el Apéndice 4);

---

\* 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 la Comisión de Estudio 11 de Radiocomunicaciones, de las Comisiones de Estudio 9 y 15 de la Normalización de las Telecomunicaciones, de la Organización Internacional de Unificación de Normas/Comisión Electrotécnica Internacional (ISO/CEI) (JTC1/SC29/WP11), de la Unión Europea de Radiodifusión (UER) y de la Audio Engineering Society (AES).

j) que los métodos objetivos convencionales (por ejemplo, los que miden la relación señal/ruido y la distorsión) pueden no ser ya adecuados para evaluar la calidad de los sistemas con codificación del sonido a baja velocidad binaria, y que los métodos objetivos perceptivos de evaluación actualmente en estudio podrían complementar o sustituir a los métodos de evaluación objetivos convencionales y complementar los métodos subjetivos de evaluación,

*recomienda*

**1** que para las aplicaciones (véase la Nota 1) de radiodifusión sonora enumeradas en los § 4 a 6, se utilice la codificación del sonido a baja velocidad binaria descrita en [ISO/CEI 11172-3, 1993] (véase la Nota 2) (en el Apéndice 1 figura una descripción breve de esta Norma);

NOTA 1 – La aplicabilidad de la presente Recomendación a los sistemas multicanal podrá ser revisada después de haber considerado los sistemas multicanal que se están desarrollando.

NOTA 2 – Conforme a la política del UIT-T, los proponentes de la técnica de codificación están dispuestos a revelar todos los detalles pertinentes y aceptan que se utilice su propiedad intelectual de forma gratuita o mediante el pago de un canon equitativo. El Anexo H de ISO/CEI 11172-3 contiene más información.

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

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

**4** que para la emisión de radiodifusión sonora digital, se utilice la codificación 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,  $2 \times 128$  kbit/s) para una señal estereofónica, con una codificación independiente de las componentes izquierda y derecha de la señal estereofónica (véase la Nota 1);

NOTA 1 – 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 1), 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 2);

NOTA 1 – Al aplicar esta Recomendación habrá que tener en cuenta que algunas redes de telecomunicación no permiten la utilización sin restricciones de los canales a 64 kbit/s.

NOTA 2 – 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 habrá que tener en cuenta que algunas redes de telecomunicación no permiten la utilización sin restricciones de los canales a 64 kbit/s.

7 que se siga estudiando los métodos objetivos perceptivos (véase el Apéndice 3) para evaluar los sistemas de codificación de audio a baja velocidad binaria;

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

## Apéndice 1

### Descripción del método de codificación

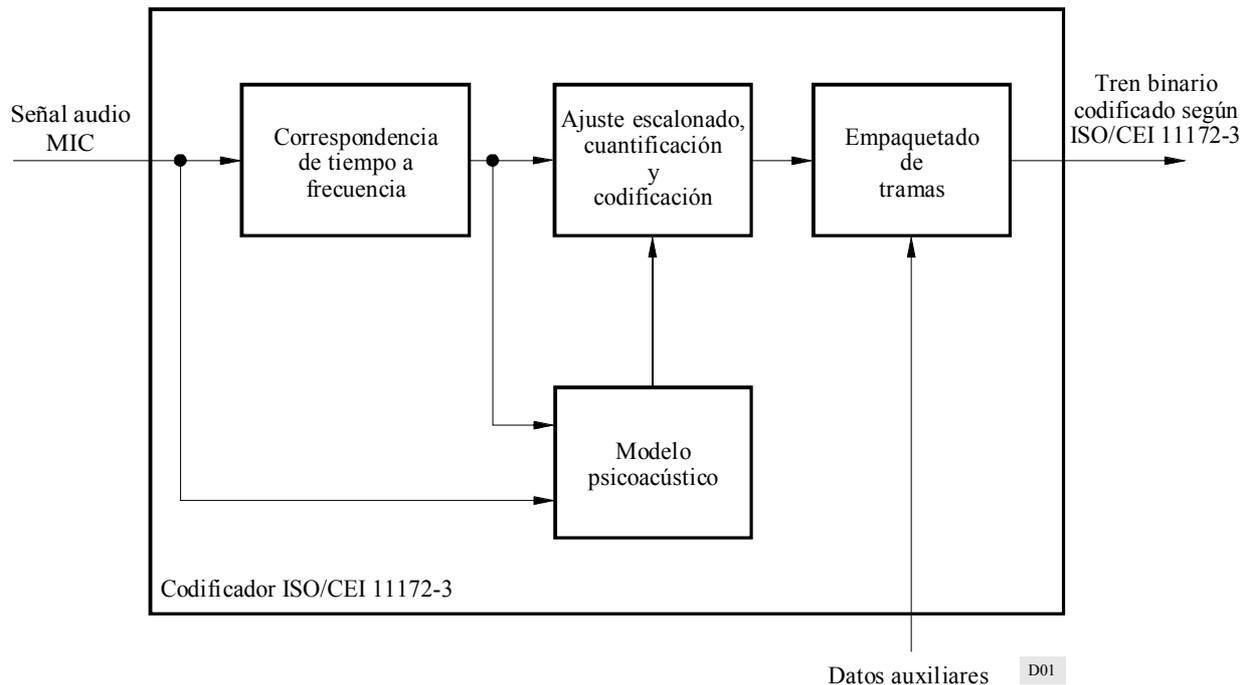
#### 1 Codificación

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 la capa 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



## 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

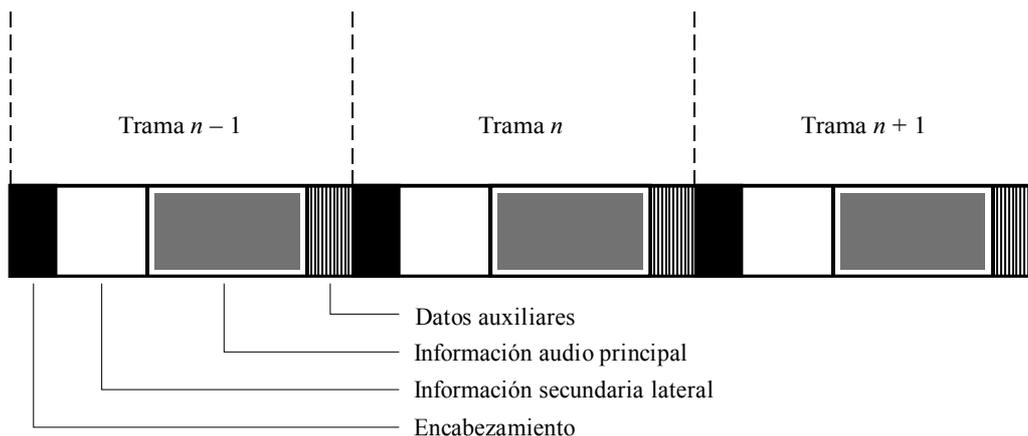
La Fig. 2 ofrece una panorámica del tren binario ISO/CEI 11172-3 para la capa II, y la Fig. 3, para la capa III. Según la capa, una trama contiene los campos siguientes:

### 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.

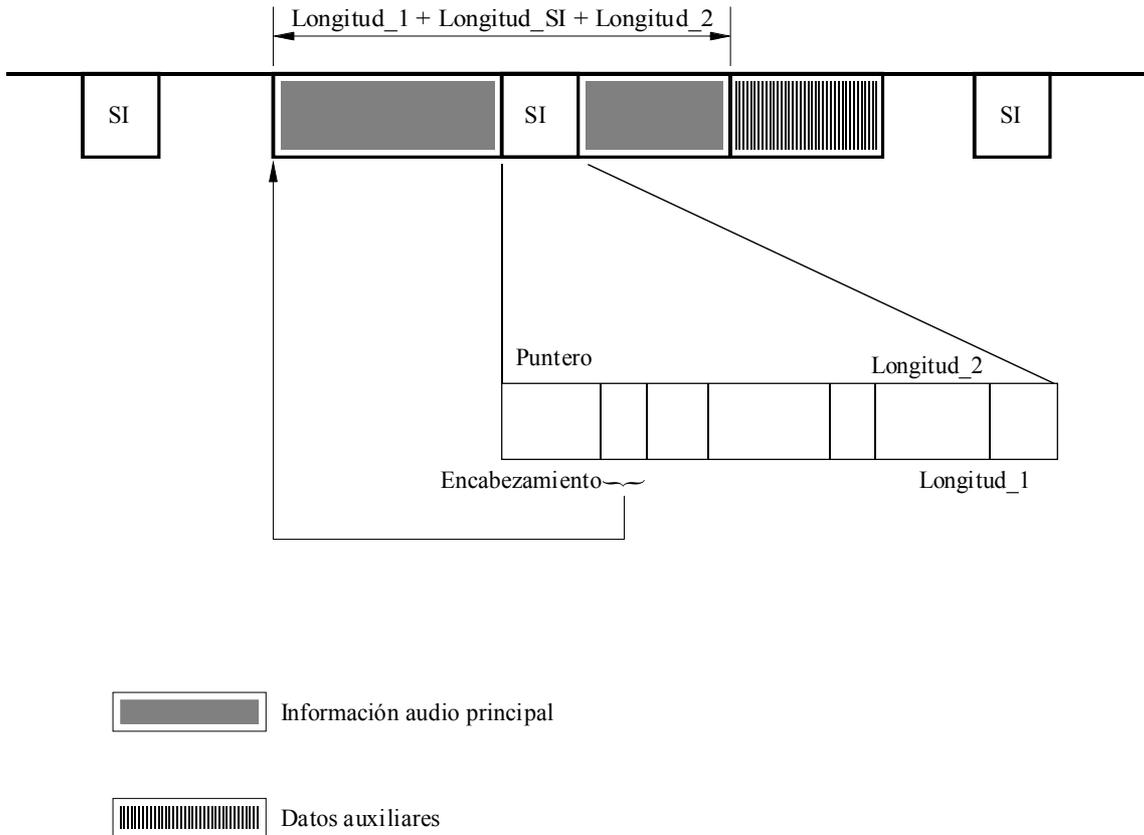
FIGURA 2  
Formato del tren binario ISO/CEI 11172-3 de capa II



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

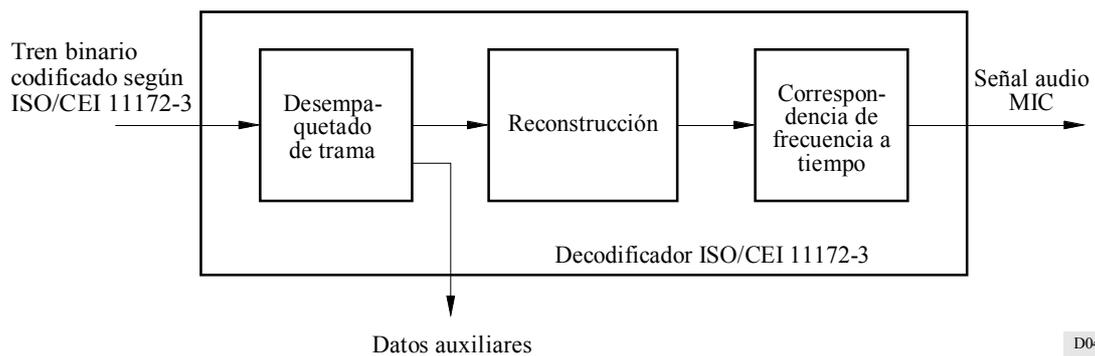


Capa III:

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

D03

FIGURA 4  
Diagrama de bloques del decodificador



D04

## 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 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

### Métodos objetivos perceptivos

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 EVA LUation model);
- medición perceptiva de la calidad audio (PAQM, Perceptual Audio Quality Measure);
- modelo objetivo perceptual (POM, Perceptual Objective Model).

Es preciso realizar otros estudios para determinar el coeficiente de confianza de los métodos objetivos perceptivos.

## Apéndice 4

### Metodología para la selección de los códecs audio de baja velocidad binaria recomendados

Se efectuaron ensayos subjetivos 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  $2 \times 128$  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.

---