



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 363 346**

51 Int. Cl.:
G10L 19/00 (2006.01)
H03M 7/40 (2006.01)
G10L 19/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **05731220 .9**
96 Fecha de presentación : **21.03.2005**
97 Número de publicación de la solicitud: **1743326**
97 Fecha de publicación de la solicitud: **17.01.2007**

54 Título: **Códec de audio multi-canal sin pérdidas.**

30 Prioridad: **25.03.2004 US 556183 P**
04.08.2004 US 911062
04.08.2004 US 911067

45 Fecha de publicación de la mención BOPI:
01.08.2011

45 Fecha de la publicación del folleto de la patente:
01.08.2011

73 Titular/es: **DTS, Inc.**
5220 Las Virgenes Road
Calabasas, California 91302, US

72 Inventor/es: **Fejzo, Zoran**

74 Agente: **Ungría López, Javier**

ES 2 363 346 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Códec de Audio Multi-Canal sin Pérdidas

5 **Referencias Cruzadas con Solicitudes Relacionadas**

La presente solicitud reivindica el beneficio de prioridad bajo 35 U.S.C. 119(e) para la Solicitud Provisional de los Estados Unidos N° 60/566.183 titulada "Códec de Audio Sin Perdidas Compatible con Versiones Anteriores" presentada el 25 de Marzo de 2004.

10

Antecedentes de la Invención**Campo de la invención**

15 Esta invención se refiere a los códec de audio sin pérdida y más especialmente a un códec de audio multi-canal sin pérdidas con funcionamiento mejorado de la compresión.

Descripción de la Técnica Relacionada

20 Actualmente están en uso varios sistemas de codificación de audio de baja tasa de bits con pérdidas en un amplio intervalo de consumidores y productos y servicios de reproducción de audio profesional. Por ejemplo, el sistema de codificación de audio Dolby AC3 (Dolby digital) es una normativa de nivel internacional para la codificación en estéreo y pistas de sonido de audio de canal 5.1 para Discos Láser, video DVD con codificación NTSC, y ATV, usando tasas de bits de hasta 640 Kbtis/s. Las normativas de codificación de audio MPEG I y MPEG II, se usan
25 ampliamente para la codificación de pistas de sonido estéreo y multicanal para video DVD con codificación PAL, difusión de radio digital terrestre en Europa y difusión por Satélite en los Estados Unidos a tasas de bit de hasta 768 Kbit/s. El sistema de codificación de audio Acústico Coherente (Sistemas de Teatro Digital) DTS se usa frecuentemente para pistas de sonido de audio de canal de calidad de estudio 5.1 para Disco Compacto, video de DVD, Difusión de Satélite en Europa y Disco Láser y tasas de bits de hasta 1536 Kbit/s.

30

Recientemente, muchos consumidores han mostrado interés en los llamados códec "sin pérdidas". Los códec "sin pérdidas" descansan en algoritmos que comprimen los datos sin descartar ninguna información y producen una señal decodificada que es idéntica que la señal fuente (digitalizada). Este funcionamiento tiene un coste: tales códec típicamente requieren más ancho de banda que los códec con pérdidas, y comprimen los datos en un menor grado.

35

La Figura 1 es una representación de un diagrama de bloques de las operaciones involucradas en la compresión sin pérdidas de un único canal de audio. Aunque los canales en el audio multi-canal no son generalmente independientes, la dependencia a menudo es débil y difícil de tener en cuenta. Por lo tanto, los canales se comprimen típicamente de forma separada. Sin embargo, algunos codificadores intentarán eliminar la correlación formando una señal residual única y codificando (Canal1, Canal1-Canal2). Enfoques más sofisticados toman, por ejemplo, varias etapas sucesivas de proyección ortogonal sobre la dimensión de canal. Todas las técnicas están basadas en el principio de eliminar en primer lugar la redundancia de la señal y a continuación codificar la señal resultante con un esquema de codificación digital eficaz. Los códec sin pérdida incluyen el MPL (DVD de Audio), audio de Monkey (aplicaciones de ordenadores), sin pérdidas de Appel, Windows Media Pro sin pérdidas, AudioPack, DVD, LTAC, MUSICcompress, OggSquish, Philips, Shorten, Sonarc y WA. Una revisión de muchos de estos códec se proporciona en el documento de "Compresión sin Pérdidas de Audio Digital" de Mat Hans, Ronald Schafer, Hewlett Packard, 1999.

40

El entramado 10 se introduce para proporcionar la facilidad de edición, el volumen total de datos prohíbe la descompresión repetitiva de toda la señal que precede a la región a editar. La señal de audio se divide en tramas independientes de igual tiempo de duración. Esta duración debería no ser demasiado corta, ya que puede resultar un control significativo de la cabecera que se fija para cada una de las tramas. Por el contrario, la duración de trama no debería ser demasiado larga, ya que esto limitaría la adaptabilidad temporal y haría la edición más difícil. En muchas aplicaciones, el tamaño de la trama está restringido por la tasa de bit de pico del medio sobre el cual se transfiere el audio, la capacidad de almacenamiento intermedio del decodificador y el deseo de que cada una de las tramas sea decodificable independientemente.

50

La des-correlación intra-canal 12 elimina la redundancia des-correlando las muestras de audio en cada uno de los canales dentro de una trama. La mayor parte de los algoritmos eliminan la redundancia por algún tipo de modelado predictivo lineal de la señal. En este enfoque, se aplica un predictor lineal a las muestras de audio en cada una de las tramas dando por resultado una secuencia de predicción de muestras de error. Un segundo enfoque, menos común, es obtener una representación cuantizada de baja tasa de bits o una representación con pérdidas de la señal, y a continuación comprimir sin pérdidas la diferencia entre la versión con pérdidas y la versión original. La codificación de entropía 14 elimina la redundancia de error de la señal residual sin perder ninguna información. Métodos típicos incluyen la codificación de Huffman, la codificación de longitud de carrera y la codificación Rice. La salida es una señal comprimida que se puede reconstruir sin pérdidas.

60

65

La especificación de DVD existente y la especificación preliminar de DVD HD fija un límite hardware sobre el tamaño de una unidad de acceso de datos, que representa una parte del flujo de audio que una vez extraída puede decodificarse totalmente y enviarse las muestras de audio reconstruidas a las memorias intermedias de salida. Lo que esto significa para un flujo sin pérdidas es que la cantidad de tiempo que cada una de las unidades de acceso puede representar tiene que ser suficientemente pequeña para que en el peor caso de la tasa de bit de pico, la carga de datos codificados no exceda el límite hardware. El tiempo de duración debe ser también reducido para tasas de muestreo incrementadas y un número de canales incrementado, lo cual aumenta la tasa de bits de pico.

Para asegurar la compatibilidad, estos códec existentes tendrán que fijar la duración de toda la trama para que sea lo suficientemente corta para que no exceda el límite hardware en el peor caso de configuración de canal /frecuencia de muestreo /ancho de bit. En la mayor parte de las configuraciones, esto será excesivo y puede degradar el funcionamiento de la compresión. Además, este enfoque del peor caso no se adapta bien con canales adicionales.

Un artículo de LIEBCHEN T y otros titulado "MPEG-4 ALS: una normativa emergente para la codificación de audio sin pérdidas" (CONFERENCIA DE COMPRESIÓN DE DATOS, 2004. PROCEDIMIENTOS. DCC 2004 SNOWBIRD, UT, ESTADOS UNIDOS, 23-25 de Marzo de 2004, PISCATAWAY, NJ, ESTADOS UNIDOS, IEEE, 23 de Marzo de 2004 (23-03-2004), páginas 439-448, el documento XP010692571 ISBN: 978-0-7695-2082-7) proporciona una breve visión general de una normativa emergente para la codificación de audio sin pérdidas, MPEG-4 ALS. Se realiza una comparación entre la normativa emergente y el estado la técnica de los algoritmos para la compresión de audio sin pérdidas.

El documento WO 00/74038 A revela un método y un sistema para la reducción de las discontinuidades de bloques inducidas por la cuantización que se presentan de la comprensión con pérdidas y la descompresión de señales continuas, especialmente las señales de audio. Una realización abarca un algoritmo de códec de audio eficaz de propósito general, de ultra baja latencia. La invención incluye un método y un aparato para la compresión y descompresión de señales de audio usando un análisis de fronteras y síntesis de estructura para reducir sustancialmente la trama inducida por cuantización o discontinuidad de bloque; una transformada de paquetes de coseno adaptativo (ACPT) como la transformada de elegir capturar eficazmente las características de audio de entrada; un clasificador de señal residual para separar los grupos de señal fuerte del ruido y las componentes de señal débil (llamadas colectivamente residuo); un algoritmo de cuantización de vectores dispersos adaptativa (ASVQ) para componentes de señal, un modelo de ruido estocástico para el residuo; y un algoritmo de control de tasa asociado. La invención incluye además las implementaciones del programa de ordenador correspondiente de estos y otros algoritmos.

Otro artículo de LIEBCHEN T se titula "Codificación de Audio Sin Perdidas usando Predicción Multicanal Adaptativa" (CITACIÓN DE INTERNET, [En Línea] del 5 de Octubre de 2002 (05-10-2002), documento XP002466533). Enseña cómo puede extenderse la predicción convencional (mono), para la codificación de audio sin pérdidas a la predicción estéreo y multicanal para mejorar la eficacia de la compresión. Se dan los resultados para las grabaciones en estéreo y multicanal.

Sumario de la Invención

La presente invención de acuerdo con la reivindicación 1 proporciona un códec de audio sin pérdidas en el cual se optimiza el funcionamiento de la compresión para una restricción del tamaño máximo sobre cada una de las unidades de datos decodificable independientemente.

El códec de la señal de audio sin perdidas segmenta los datos de audio dentro de cada una de las tramas para mejorar el funcionamiento de la compresión sujeta a la restricción de que cada uno de los segmentos debe ser totalmente decodificable y menor que un tamaño máximo. Para cada una de las tramas, el códec selecciona la duración del segmento y los parámetros de codificación, por ejemplo, un codificador de entropía particular y sus parámetros para cada uno de los segmentos, que minimiza la carga de datos codificada para toda la trama sujeta a las restricciones. Pueden seleccionarse distintos conjuntos parámetros de codificación para cada uno de los canales o puede seleccionarse un conjunto global de parámetros de codificación para todos los canales. El funcionamiento de la compresión puede mejorarse adicionalmente formando M/2 canales de des-correlación para M canales de audio. El triplete de canales (base, correlado, descorrelado) proporciona dos posibles pares de combinaciones (base, correlado) y (base, descorrelado) que pueden considerarse durante la segmentación y la optimización de la codificación de entropía para mejorar adicionalmente el funcionamiento de la compresión. Los pares de canales pueden especificarse por segmentos o por trama.

En una realización de ejemplo, el codificador encuadra los datos de audio y a continuación extrae pares de canales ordenados incluyendo un canal base y un canal correlado y genera un canal descorrelado para formar al menos un triplete (base, correlado, descorrelado). Si el número de canales es impar, se procesa un canal base extra. La predicción de polinomios fijos o adaptativos se aplica a cada uno de los canales para formar las señales residuales.

El codificador determina la duración del segmento, los pares de canales (base, correlado) o (base, descorrelado)

para la trama y los conjuntos de parámetros de codificación (selección de código de entropía y parámetros) para cada uno de los segmentos por la primera partición de la trama en un número máximo de segmentos de duración mínima. Los parámetros de codificación óptima para la partición actual se determinan calculando los parámetros para uno o más codificadores de entropía (Binario, Rice, Huffman, etc.) y seleccionando el codificador y los parámetros con la carga de datos codificados más pequeña para cada uno de los canales (base, correlado, descorrelado) para cada uno de los segmentos. Para cada uno de los tripletes, se selecciona el par de canales (base, correlado) o (base, descorrelado) con la carga de datos codificados más pequeña. Usando el par de canales seleccionado, puede determinarse un conjunto global de parámetros de codificación para cada uno de los segmentos sobre todos los canales. El codificador selecciona el conjunto global o conjuntos distintos de parámetros de codificación en base a los cuales tiene la carga de datos codificados total más pequeña (cabecera y datos de audio).

Una vez que se ha determinado el conjunto óptimo de parámetros de codificación y los pares de canales para la partición actual, el codificador calcula la carga de datos codificados en cada uno de los segmentos a través de todos los canales. Asumiendo la restricción de que se satisface la restricción sobre el tamaño de segmento máximo, el codificador determina si la carga de datos codificados total para toda la trama para la partición actual es menor que la óptima actual para una partición anterior. Si es cierto, el conjunto actual de parámetros de codificación y carga de datos codificados se almacena y se aumenta la duración del segmento. Este proceso se repite hasta que el tamaño del segmento viola la restricción del tamaño máximo o la duración del segmento aumenta a la duración de trama. El codificador de entropía codifica (usando el codificador de entropía y los parámetros seleccionados) las señales residuales en cada uno de los canales de audio de los pares de canales seleccionados y todos los canales no pareados.

Estas y otras características y ventajas de la invención serán evidentes para los especialistas en la técnica a partir de la siguiente descripción detallada de las realizaciones preferidas, tomadas junto con los dibujos adjuntos, en los que:

Breve Descripción de los Dibujos

La FIG. 1, como se ha descrito anteriormente, es un diagrama de bloques para un codificador de audio normalizado sin pérdidas;
 las FIG. 2a y 2b son diagramas de bloques de un codificador y decodificador de audio sin pérdidas, respectivamente, de acuerdo con la presente invención;
 la FIG. 3 es un diagrama de la información de cabecera como se relaciona con la segmentación y la selección de código de entropía;
 las FIG. 4a y 4b son diagramas de bloques del procesamiento de análisis de ventana y el procesamiento de análisis de ventana inverso;
 la FIG. 5 es un diagrama de flujo de la des-correlación cruzada de canales;
 las FIG. 6a y 6b son diagramas de bloque del análisis y procesamiento de la predicción adaptativa y el procesamiento de la predicción adaptativa inversa;
 las FIG. 7a y 7b son diagramas de flujo de una segmentación óptima y la selección del código de entropía;
 las FIG. 8a y 8b son diagramas de flujo de la selección de código de entropía para un conjunto de canales; y
 las FIG. 9a y 9b son diagramas de bloques de un núcleo más un códec de extensión sin pérdidas.

Descripción Detallada de la Invención

La presente invención proporciona un códec de audio sin pérdidas en el cual el funcionamiento de la compresión está optimizado sometido a la restricción de tamaño máximo sobre cada una de las unidades de datos decodificable independientemente. El codificador de audio se adapta a medida que continúa creciendo el número de canales en una señal de audio multicanal.

CÓDEC DE AUDIO SIN PÉRDIDAS

Como se muestra en las Figuras 2a y 2b, los bloques funcionales esenciales son similares a los codificadores y los decodificadores sin pérdidas existentes con la excepción de la selección de la segmentación y del código de entropía. La señal de audio de PCM multicanal 20 está sujeta al procesamiento de análisis de ventana 22, que bloquea los datos en tramas de una duración constante y elimina la redundancia des-correlando las muestras de audio en cada uno de los canales dentro de una trama. En lugar de codificar la entropía de las señales residuales directamente, la presente invención realiza una segmentación óptima y un proceso de selección de código de entropía 24 que segmenta los datos en una pluralidad de segmentos y determina la duración del segmento y los parámetros de codificación, por ejemplo, la selección de un codificador de entropía particular y sus parámetros, para cada uno de los segmentos que minimiza la carga de datos codificados para toda la trama sujeta a la restricción de que cada una de los segmentos debe ser totalmente decodificable y menor que el tamaño máximo. Los conjuntos de parámetros de codificación están optimizados para cada uno de los distintos canales y pueden optimizarse para un conjunto global de parámetros de codificación. Cada uno de los segmentos se codifica de entropía a continuación 26 de acuerdo con su conjunto particular de parámetros de codificación. Los datos codificados y la información de

cabecera se empaquetan 28 en un flujo de bits 30.

Como se muestra en la Figura 3, la cabecera 32 incluye información adicional además de la que se proporciona normalmente para un códec sin pérdidas para implementar la segmentación y selección de código de entropía. Más específicamente, la cabecera incluye una información de cabecera común 34 tal como el número de segmentos (NúmSegmentos) y el número de muestras en cada uno de los segmentos.

El (NúmMuestrEnSegmento), la información de cabecera del conjunto de canales 36 tal como los coeficientes de des-correlación cuantizados (CuantCanalDecorrCoef []) y la información de cabecera de segmento 38 tal como el número de bytes en el segmento actual para el conjunto de canales (CanalConjByteCons), un indicador de optimización global (TodosCanalesIguarParámInd) y los indicadores del codificador de entropía (RiceCodifInd [], CodifParám []) que indica si se usa una codificación Rice o Binaria y los parámetros de codificación.

Como se muestra en la Figura 2b, para realizar la operación de decodificación el flujo de bits 30 se desempaqueta 40 para extraer la información de cabecera y los datos codificados. Una decodificación de entropía 42 se realiza sobre cada uno de los segmentos de cada uno de los canales de acuerdo con los parámetros de codificación asignados para reconstruir sin pérdidas las señales residuales. Estas señales se sujetan a continuación al procesamiento de análisis de ventana inverso 44, que realiza la predicción inversa para reconstruir sin pérdidas la señal de audio PCM original 20.

PROCESAMIENTO DE ANÁLISIS DE VENTANA

Como se muestra en las Figuras 4a y 4b, una realización de ejemplo de un procesamiento de análisis de ventana 22 selecciona bien la predicción adaptativa 46 o la predicción de polinomios fijos 48 para descorrelar cada uno de los canales, que es un enfoque bastante común. Como se describirá en detalle con referencia a la Figura 6, se estima el orden del predictor óptimo para cada uno de los canales. Si el orden es mayor que cero, se aplica la predicción adaptativa. De otro modo se usa la predicción de polinomios fijos más simple. De forma similar, en el decodificador el procesamiento de análisis de ventana inverso 44 selecciona bien la predicción adaptativa inversa 50 o la predicción de polinomios fijos inversa 52 para reconstruir la señal de audio de PCM a partir de las señales residuales. Los órdenes del predictor adaptativo y los índices de coeficientes de predicción adaptativa y los órdenes del predictor fijados se empaquetan 53 en la información de cabecera del conjunto de canales.

Des-correlación Cruzada de Canales

De acuerdo con la presente invención, el funcionamiento de la descompresión puede mejorarse adicionalmente implementando la des-correlación cruzada de los canales 54, que ordena los M canales de entrada en pares de canales de acuerdo con la medida de la correlación entre los canales. Uno de los canales se designa como el canal "base" y el otro se designa como canal "correlado". Se genera un canal descorrelado para cada uno de los pares de canales para formar un "tripleto" (base, correlado, descorrelado). La formación del tripleto proporciona dos posibles pares de combinaciones (base, correlado) y (base, descorrelado) que pueden considerarse durante la segmentación y la optimización de la codificación de entropía para un funcionamiento de la compresión mejorado adicionalmente (véase la Figura 8a). Un enfoque más simple pero menos efectivo sería reemplazar el canal correlado con el canal descorrelado si, por ejemplo, su varianza fuese más pequeña.

El PCM original de M canales y el PCM de M/2 canales descorrelados 56 se retransmiten ambos para la predicción adaptativa y las operaciones de predicción de polinomios fijos, lo cual genera señales residuales para cada uno de los canales. Como se muestra en la Figura 3, los índices (OrigCanalOrden []) que indican el orden original de los canales antes de la clasificación realizada durante el proceso de des-correlación orientada a pares y el indicador PWCanalDescorrInd [] para cada uno de los pares de canales indicando la presencia de un código para los coeficientes de des-correlación cuantizada se almacenan en la cabecera del conjunto de canales 36 en la Figura 3.

Como se muestra en la Figura 4b, para realizar la operación de decodificación del procesamiento de análisis de ventana inverso 44 la información de la cabecera se desempaqueta 58 y los residuos se pasan a través de la predicción de polinomios fijos inversa 52 o la predicción adaptativa inversa 50 de acuerdo con la información de cabecera, a saber los órdenes del predictor adaptativo y fijo para cada uno de los canales. La señal de audio de PCM de M canales descorrelados (M/2 canales se descartan durante la segmentación) se pasa a través de la des-correlación cruzada de los canales inversa 60, que lee los índices OrigCanalOrden[] y el indicador PWCanalDescorrInd [] desde la cabecera del conjunto de canales y reconstruye sin pérdidas la señal de audio de PCM de M canales 20.

Un proceso de ejemplo para la realización de la des-correlación cruzada de canales 54 se ilustra en la Figura 5. A modo de ejemplo, la señal de audio de PCM se proporciona como M=6 canales distintos, L, R, C, Ls, Rs, y LFE, que también corresponden directamente a una configuración del conjunto de canales almacenada en la trama. Otros conjuntos de canales pueden ser, por ejemplo, el entorno posterior a la izquierda del centro y el entorno posterior a la derecha del centro para producir la señal de audio del entorno 7.1. El proceso comienza arrancando un bucle de trama y arrancando un bucle del conjunto de canales (etapa 70). Se calculan la estimación de la auto-correlación de

5 retardo cero para cada uno de los canales (etapa 72) y la estimación de correlación cruzada de retardo cero para todas las combinaciones posibles de pares de canales en el conjunto de canales (etapa 74). A continuación, se estiman los coeficientes de correlación orientada a pares de canales CORCOEF como la estimación de correlación cruzada de retardo cero divididos por el producto de las estimaciones de auto-correlación de retardo cero para los canales involucrados en el par (etapa 76). Los CORCOEF se almacenan desde el mayor valor absoluto al más pequeño y se almacenan en una tabla (etapa 78). Comenzando desde la parte superior de la tabla, se extraen los índices de pares de canales correspondientes hasta que se han configurado todos los pares (etapa 80). Por ejemplo, los 6 canales pueden parearse en base a sus CORCOEF como (L, R), (Ls, Rs) y (C, LFE).

10 El proceso arranca un bucle de pares de canales (etapa 82), y selecciona un canal "base" como el canal con la estimación más pequeña de auto-correlación de retardo cero, que es indicativa de una energía más baja (etapa 84). En este ejemplo, los canales L, Ls y C forman los canales base. El coeficiente de des-correlación del par de canales (CanalParDescorrCoef) se calcula como la estimación de correlación cruzada de retardo cero dividida por la estimación de auto-correlación de retardo cero del canal base (etapa 86). El canal des-correlado se genera multiplicando las muestras del canal base por el CanalParDescorrCoef y restando el resultado de las muestras correspondientes del canal correlado (etapa 88). Los pares de canales y su canales des-correlados asociados definen "tripletes" (L, R, R-CanalParDescorrCoef[1]*L), (Ls, Rs, Rs-CanalParDescorrCoef[2]*Ls), (C, LFE, LFE-CanalParDescorrCoef[3]*C) (etapa 89). El CanalParDescorrCoef [] para cada uno de los pares de canales (y cada uno de los conjuntos de canales) y los índices de canal que definen la configuración de pares se almacenan en la información de la cabecera del conjunto de canales (etapa 90). Este proceso se repite para cada uno de los conjuntos de canales en una trama y a continuación para cada una de las tramas en la señal de audio de PCM en la ventana (etapa 92).

25 Predicción adaptativa

25 **Análisis de la Predicción Adaptativa y la Generación de Residuos**

30 La predicción lineal intenta eliminar la correlación entre las muestras de una señal de audio. El principio básico de la predicción lineal es predecir un valor de una muestra $s(n)$ usando las muestras anteriores $s(n-1)$, $s(n-2)$, ... y restar el valor predicho $\hat{s}(n)$ de la muestra original $s(n)$. La señal residual resultante $e(n) = s(n) - \hat{s}(n)$ idealmente estará des-correlada y consecuentemente tendrá un espectro de frecuencias plano. Además, la señal residual tendrá una varianza más pequeña implicando entonces que la señal original que serán necesarios menos bits para su representación digital.

35 En una realización de ejemplo de un códec de audio, se describe un modelo predictor de FIR por la siguiente ecuación:

$$e(n) = s(n) - Q\left\{\sum_{k=1}^M a_k * s(n-k)\right\}$$

40 donde $Q\{\}$ denota la operación de cuantización, M denota el orden del predictor y a_k son coeficientes de la predicción cuantizada. Una cuantización particular $Q\{\}$ es necesaria para una compresión con pérdidas ya que la señal original se reconstruye sobre el lado de decodificación, usando diversas arquitecturas de procesador de precisión finita. La definición de $Q\{\}$ está disponible tanto para el codificador como el decodificador y la reconstrucción de la señal original se obtienen simplemente por:

$$s(n) = e(n) - Q\left\{\sum_{k=1}^M a_k * s(n-k)\right\}$$

45 donde se asume que están disponibles los mismos coeficientes de predicción cuantizada a_k tanto para el codificador como el decodificador. Se transmite un nuevo conjunto de parámetros del predictor por cada una de las ventanas de análisis (trama) permitiendo al predictor adaptarse a la estructura de la señal de audio variable con el tiempo.

50 Los coeficientes de predicción están diseñados para minimizar los residuos de la predicción de mínimos cuadrados. La cuantización $Q\{\}$ hace del predictor un predictor no lineal. Sin embargo en la realización de ejemplo la cuantización se hace con una precisión de 24 bits y es razonable asumir que los efectos no lineales resultantes pueden ignorarse durante la optimización de coeficientes del predictor. Ignorando la cuantización $Q\{\}$, el problema de optimización subyacente puede representarse como un conjunto de ecuaciones lineales que involucran los retardos de la secuencia de auto-correlación de la señal y los coeficientes del predictor desconocidos. Este conjunto de ecuaciones lineales puede resolverse de forma eficaz usando el algoritmo de Levinson-Durbin (LD).

60 Los coeficientes de predicción lineal resultantes (LPC) necesitan cuantizarse, de este modo pueden transmitirse de forma eficaz en un flujo codificado. Desafortunadamente la cuantización directa de los LPC no es el enfoque más

eficaz ya que los pequeños errores de cuantización pueden causar grandes errores espectrales. Una representación alternativa de los LPC es la representación de coeficientes de reflexión (RC), que exhibe menos sensibilidad a los errores de cuantización. Esta representación también puede obtenerse a partir del algoritmo LD. Por definición del algoritmo LD, está garantizado que los RC tienen una magnitud ≤ 1 (ignorando los errores numéricos). Cuando el valor absoluto de los RC está próximo a 1 la sensibilidad de la predicción lineal para los errores de cuantización presentes en los RC cuantizados se hacen altos. La solución es realizar una cuantización no uniforme de los RC con etapas de cuantización más finas alrededor de la unidad. Esto se consigue en dos etapas:

1) transformar los RC a una representación de proporciones de log de área (LAR) por medio de la función de mapeo

$$LAR = \log \frac{1+RC}{1-RC}$$

donde log representa el logaritmo en base natural.

2) cuantizar uniformemente las LAR

La transformación de RC \rightarrow LAR deforma la escala de amplitud de los parámetros de modo que el resultado de las etapas 1 y 2 es equivalente a la cuantización no uniforme con etapas de cuantización más finas alrededor de la unidad.

Como se muestra en la Figura 6a, en una realización de ejemplo de un análisis de predicción adaptativa los parámetros de LAR cuantizados se usan para representar parámetros de predictor adaptativo y transmitidos en el flujo de bits codificados. Las muestras en cada uno de los canales de entrada se procesan independientemente entre sí y en consecuencia la descripción sólo considerará el procesamiento en un canal único.

La primera etapa es calcular la secuencia de auto-correlación sobre la duración de la ventana de análisis (trama) (etapa 100). Para minimizar los efectos de bloqueo que se causan por las discontinuidades en las fronteras de trama, los datos en primer lugar se ajustan en ventanas. La secuencia de autocorrección para un número especificado (igual a un orden de LP máximo +1) de retardos se estima a partir del bloque de datos ajustado a ventana.

El algoritmo de Levinson-Durbin (LD) se aplica al conjunto de retardos de auto-correlación estimados y se calcula el conjunto de coeficientes de reflexión (RC), hasta el orden de LP máximo, (etapa 102). Un resultado intermedio del algoritmo de (LD) es un conjunto de varianzas estimadas de residuos de predicción para cada uno de los órdenes de predicción lineal hasta el orden de LP máximo. En el siguiente bloque, usando este conjunto de varianzas residuales, se selecciona el orden del predictor lineal (PrOr) (etapa 104).

Para el orden del predictor seleccionado se transforma el conjunto de coeficientes de reflexión (RC), al conjunto de parámetros de proporción de área-log (LAR) usando la función de mapeo establecida anteriormente (etapa 106). Se introduce una limitación de RC antes de la transformación para impedir la división por cero:

$$RC = \begin{cases} Umbral & \forall RC > Umbral \\ -1 & \forall RC < -1 \\ RC & \text{En otro caso} \end{cases}$$

donde Umbral denota un número próximo pero menor que 1. Los parámetros de LAR se cuantizan (etapa 108) de acuerdo con la siguiente norma:

$$QLARInd = \begin{cases} \left\lfloor \frac{LAR}{q} \right\rfloor & \forall LAR \geq 0 \\ -\left\lfloor \frac{-LAR}{q} \right\rfloor & \forall LAR < 0 \end{cases}$$

donde QLARInd denota los índices LAR cuantizados, [x] indica la operación de encontrar el mayor valor de número entero menor o igual que x y q denota el tamaño de la etapa de cuantización. En la realización de ejemplo, la región [-8 a 8] se codifica usando 8 bits, es decir $q = (2 * 8) / 2^8$ y en consecuencia QLARInd se limita de acuerdo con:

$$QLARInd = \begin{cases} 127 & \forall QLARInd > 127 \\ -127 & \forall QLARInd < -127 \\ QLARInd & \text{En otro caso} \end{cases}$$

Antes de empaquetar (etapa 110), $QLARInd$ se traduce de valores con signo a valores sin signo usando el siguiente mapeo

$$PackLARInd = \begin{cases} 2 * QLARInd & \forall QLARInd \geq 0 \\ 2 * (-QLARInd) - 1 & \forall QLARInd < 0 \end{cases}$$

5 En el bloque "RC LUT", se realizan una cuantización inversa de parámetros LAR y una traducción a parámetros RC en una etapa simple usando una tabla de búsqueda (etapa 112). La tabla de búsqueda consiste de valores cuantizados del mapeo inverso de $RC \rightarrow LAR$, es decir el mapeo $LAR \rightarrow RC$ dado por:

$$RC = \frac{e^{LAR} - 1}{e^{LAR} + 1}$$

10 La tabla de búsqueda se calcula en valores cuantizados de LAR iguales a 0, 1,5*q, 2,5*q,... 127,5*q. Los valores correspondientes de RC, después de poner a escala por 2^{16} , se redondean a números enteros sin signo de 16 bits y se almacenan como números de punto fijo sin signo Q16 en una tabla de 128 entradas.

15 Los parámetros cuantizados de RC se calculan a partir de la tabla y los índices LAR de cuantización $QLARInd$ como

$$QRC = \begin{cases} TABLA [QLARInd] & \forall QLARInd \geq 0 \\ - TABLA [-QLARInd] & \forall QLARInd < 0 \end{cases}$$

Los parámetros de RC cuantizados QRC_{ord} para $ord = 1, \dots, PrOr$ se traducen a los parámetros de predicción lineal cuantizados (LP_{ord} para $ord = 1, \dots, PrOr$) de acuerdo con el siguiente algoritmo (etapa 114):

```

20 Para ord = 0 a PrOr - 1 hacer
    Para m = 1 a ord hacer
         $C_{ord+1, m} = C_{ord, m} + (QRC_{ord+1} * C_{ord, ord+1-m} + (1 \ll 15)) \gg 16$ 
    fin
25  $C_{ord+1, ord+1} = QRC_{ord+1}$ 
    fin
    Para ord = 0 a PrOr - 1 hacer
         $LP_{ord+1} = C_{PrOr, ord+1}$ 
    fin
30

```

35 Como los coeficientes de RC cuantizados se representaron en un formato de punto fijo con signo Q16 el algoritmo anterior generará coeficientes LP también en el formato de punto fijo con signo Q16. La trayectoria de cálculo del decodificador sin pérdidas se diseña para soportar resultados intermedios de hasta 24 bits. Por lo tanto es necesario realizar una comprobación de saturación después de que se calcula cada $C_{ord+1, m}$. Si se produce la saturación en cualquier etapa del algoritmo, se fija el indicador de saturación y el orden del predictor adaptativo PrOr, para un canal particular se pone a cero (etapa 116). Para este canal particular con $PrOr = 0$ se realizará una predicción de coeficientes fijos en lugar de una predicción adaptativa (véase la Predicción de Coeficientes Fijos). Obsérvese que los índices de cuantización de LAR sin signo ($PackLARInd[n]$ para $n = 1, \dots, PrOr[Ch]$) se empaquetan dentro del flujo codificado sólo para los canales con $PrOr[Canal] > 0$.

40 Finalmente se realiza la predicción lineal adaptativa para cada uno de los canales con $PrOr > 0$ y se calculan los residuos de la predicción $e(n)$ de acuerdo con las siguientes ecuaciones (etapa 118):

$$s(\overline{n}) = \left[\left\{ \sum_{k=1}^{PrOr} LP_k * s(n-k) \right\} + (1 \ll 15) \right] \gg 16$$

Límite $\overline{s(n)}$ a 24 - intervalo de bits $(-2^{23} \text{ a } 2^{23} - 1)$

$$e(n) = s(n) + \overline{s(n)}$$

Límite $e(n)$ para 24 - intervalo de bits $(-2^{23} \text{ a } 2^{23} - 1)$
para $n = PrOr + 1, \dots, \text{NúmMuestEnTrama}$

55 Como el objetivo del diseño en la realización de ejemplo es que cada trama sea un "punto de acceso aleatorio", la historia de las muestras no se transporta entre las tramas. En cambio la predicción se ocupa sólo en la muestra $PrOr + 1$ en la trama.

Los residuos de la predicción adaptativa $e(n)$ se codifican además en entropía y se empaquetan dentro del flujo de bits codificados.

Predicción Adaptativa Inversa del Lado del Decodificador

5 Del lado de la decodificación, la primera etapa en la realización de la predicción adaptativa inversa es desempaquetar la información de la cabecera y extraer los órdenes de predicción adaptativa $PrOr[Canal]$ para cada uno de los canales $Canal=1, \dots, N\acute{u}mCanales$ (etapa 120). A continuación para los canales con $PrOr[Canal]>0$, se extrae la versión sin signo de los índices de cuantización de LAR ($PackLARInd[n]$ para $n=1, \dots, PrOr[Canal]$). Para cada uno de los canales con orden de predicción $PrOr[Canal]>0$ se mapean los $PackLARInd[n]$ a valores con signo $QLARInd[n]$ usando el siguiente mapeo:

$$15 \quad QLARInd[n] = \begin{cases} PackLARInd[n] \gg 1 & \forall \text{ número par de } PackLARInd[n] \\ -(PackLARInd[n] \gg 1) - 1 & \forall \text{ número impar de } PackLARInd[n] \end{cases}$$

para $n = 1, \dots, PrOr[Ch]$

donde el símbolo \gg denota una operación de desplazamiento a la derecha de un número entero.

20 Se realiza una cuantización inversa de los parámetros LAR y una traducción a los parámetros RC en una única etapa usando una cuantificación de RC LUT (etapa 122). Esta es la misma tabla de búsqueda que $TABLA\{}$ como se define en el lado del codificador. Los coeficientes de reflexión cuantizados para cada uno de los canales ($QRC[n]$ para $n=1, \dots, PrOr[Canal]$) se calculan a partir de la $TABLA\{}$ y los índices LAR de cuantización $QLARInd[n]$, como

$$25 \quad QRC[n] = \begin{cases} TABLA[QLARInd[n]] & \forall QLARInd[n] \geq 0 \\ -TABLA[-QLARInd[n]] & \forall QLARInd[n] < 0 \end{cases}$$

para $n = 1, \dots, PrOr[Canal]$

30 Para cada uno de los canales, los parámetros de RC cuantizados QRC_{ord} para $ord = 1, \dots, PrOr[Canal]$ se traducen a parámetros de predicción lineal cuantizados (LP_{ord} para $ord = 1, \dots, PrOr[Canal]$) de acuerdo con el siguiente algoritmo (etapa 124):

```
35 Para ord = 0 a PrOr - 1 hacer
    Para m = 1 a ord hacer
         $C_{ord+1, m} = C_{ord, m} + (QRC_{ord+1} * C_{ord, ord+1-m} + (1 \ll 15)) \gg 16$ 
    fin
     $C_{ord+1, ord+1} = QRC_{ord+1}$ 
40 fin
    Para ord = 0 a PrOr - 1 hacer
         $LP_{ord+1} = C_{PrOr, ord+1}$ 
    fin
```

45 Cualquier posibilidad de saturación de los resultados intermedios se elimina sobre el lado de la codificación. Por lo tanto sobre el lado de la decodificación no hay ninguna necesidad de realizar la comprobación de saturación después del cálculo de cada uno de los $C_{ord+1, m}$.

50 Finalmente para cada uno de los canales con $PrOr[Canal] > 0$ se realiza una predicción lineal adaptativa inversa (etapa 126). Asumiendo que los residuos de predicción $e(n)$ se extraen previamente y se decodifica la entropía, las señales originales reconstruidas $s(n)$ se calculan de acuerdo con las siguientes ecuaciones:

$$\overline{s(n)} = \left[\left\{ \sum_{k=1}^{PrOr[Ch]} LP_k * s(n-k) \right\} + (1 \ll 15) \right] \gg 16$$

Límite $\overline{s(n)}$ a $24 - \text{rango de bits} (-2^{23} \text{ a } 2^{23} - 1)$

$$e(n) = s(n) - \overline{s(n)}$$

para $n = PrOr[Canal] + 1, \dots, N\acute{u}nMuestEnTrama$

55 Como la historia de las muestras no se mantiene dentro de las tramas la predicción adaptativa inversa comenzará desde la muestra ($PrOr[Canal] + 1$) en la trama.

Predicción de coeficientes fijos

Una forma muy simple de coeficientes fijos del predictor lineal se ha encontrado que es útil. Los coeficientes de la predicción fija se deducen de acuerdo con un método de aproximación de polinomios muy simple propuesto en primer lugar por Shorten (T. Robinson. SHORTEN: Simple comprensión de la forma de onda sin pérdidas y casi sin pérdidas. Informe técnico 156. Universidad de Cambridge Departamento de Ingeniería, Calle Trumpington, Cambridge CB2 1PZ, Reino Unido, Diciembre de 1994). En este caso los coeficientes de predicción son los especificados fijando un polinomio de orden p para los últimos p puntos de datos. Expandiendo sobre cuatro aproximaciones

$$\hat{s}_0[n] = 0$$

$$\hat{s}_1[n] = s[n-1]$$

$$\hat{s}_2[n] = 2s[n-1] - s[n-2]$$

$$\hat{s}_3[n] = 3s[n-1] - 3s[n-2] + s[n-3]$$

Una propiedad interesante de estas aproximaciones de polinomios es que la señal residual resultante $e_k[n] = s[n] - \hat{s}_k[n]$ puede implementarse de modo eficaz en el siguiente modo recursivo.

$$e_0[n] = s[n]$$

$$e_1[n] = e_0[n] - e_0[n-1]$$

$$e_2[n] = e_1[n] - e_1[n-1]$$

$$e_3[n] = e_2[n] - e_2[n-1]$$

El análisis de predicción de coeficientes fijos se aplica sobre la base de trama por trama y no descansa sobre muestras calculadas anterior ($e_k[-1] = 0$). El conjunto residual con la magnitud de la suma más pequeña sobre toda la trama se define como la mejor aproximación. El orden residual óptimo se calcula para cada uno de los canales separadamente y se empaqueta dentro del flujo como el Orden de Predicción Fijo (FPO[Canal]). Los residuos $e_{FPO}[\text{Canal}]$ en la trama actual se codifica la entropía adicionalmente y se empaqueta dentro del flujo.

El proceso de predicción de coeficientes fijos inverso, del lado del decodificador, se define por una fórmula recursiva de orden k para el cálculo del residuo de orden k en el ejemplo de muestreo n :

$$e_k[n] = e_{k+1}[n] + e_k[n-1]$$

donde la señal original deseada $s[n]$ viene dada por

$$s[n] = e_0[n]$$

y donde para cada uno de los residuos de orden k $e_k[-1] = 0$

como ejemplo se presentan las repeticiones para la predicción de coeficientes fijos de tercer orden donde los residuos $e_3[n]$ se codifican, se transmiten en el flujo y se desempaquetan del lado del decodificador:

$$e_2[n] = e_3[n] + e_2[n-1]$$

$$e_1[n] = e_2[n] + e_1[n-1]$$

$$e_0[n] = e_1[n] + e_0[n-1]$$

$$s[n] = e_0[n]$$

SEGMENTACIÓN Y SELECCIÓN DEL CÓDIGO DE ENTROPÍA

Una realización de ejemplo de la segmentación y la selección del código de entropía 24 se ilustra en las Figuras 7 y 8. Para establecer la duración óptima de un segmento, los parámetros de codificación (selección del código de entropía y parámetros) y pares de canales, se determinan los parámetros de codificación y los pares de canales para una pluralidad de duraciones de segmentos diferentes y de entre esos candidatos se selecciona el que tiene la carga de datos codificados mínima por trama que satisface las restricciones que cada uno de los segmentos debe ser decodificable independientemente y no exceder un tamaño máximo. La segmentación "óptima", los parámetros de codificación y los pares de canales son por supuesto objeto de las restricciones del proceso de codificación así como las restricciones sobre el tamaño del segmento. Por ejemplo, en el proceso de ejemplo, el tiempo de duración de todos los segmentos en la trama es igual, la búsqueda de la duración óptima se realiza sobre una rejilla diádica, y la selección de pares de canales es válida sobre toda la trama. Al coste de la complejidad adicional del codificador y

de bits de control, el tiempo de duración puede permitirse que varíe dentro de una trama, la búsqueda de la duración óptima podría resolverse de forma más fina y la selección del par de canales podría realizarse sobre la base de un segmento.

5 El proceso de ejemplo comienza inicializando los parámetros del segmento (etapa 150) tal como el número mínimo de muestras en un segmento, el tamaño máximo permitido de un segmento, el número máximo de segmentos y el número máximo de particiones. Por lo tanto, el procesamiento comienza un bucle de partición que se indexa desde 0 hasta el número máximo de particiones menos uno (etapa 152) e inicializa los parámetros de partición incluyendo el número de segmentos, el número de muestras en un segmento y el número de bytes consumidos en una partición (etapa 154). En esta realización particular, los segmentos son del mismo tiempo de duración y el número de segmentos escala como una potencia de dos con cada iteración de partición. El número de segmentos preferiblemente se inicializa al máximo, y por lo tanto el mínimo tiempo de duración. Sin embargo, el proceso podría usar segmentos de tiempo de duración variable, lo cual podría proporcionar una mejor compresión de los datos de audio pero a expensas de un control adicional. Además, el número de segmentos no tiene que limitarse a las potencias de dos o buscarse a partir de un mínimo a la duración máxima.

Una vez inicializado, el proceso arranca un bucle de un conjunto de canales (etapa 156) y determina los parámetros de codificación de entropía óptima y la selección del par de canales para cada uno de los segmentos y el consumo de los bytes correspondientes (etapa 158). Los parámetros de codificación `PWCanalDecorrInd` [], `TodosCanalesIgualesParámlnd` [], `RiceCodificInd` [], `CodifParám` [] y `CanalConjByteCons` [] se almacenan (etapa 160). Esto se repite para cada uno del conjunto de canales hasta que termina el bucle del conjunto de canales (etapa 162).

El proceso comienza un bucle de segmento (etapa 164) y calcula el consumo de bytes (`SegmByteCons`) en cada uno de los segmentos sobre todos los conjuntos de canales (etapa 166) y actualiza el consumo de bytes (`ByteConsEnPart`) (etapa 168). En este punto, el tamaño del segmento se compara con la restricción del tamaño máximo (etapa 170). Si se viola la restricción se descarta la partición actual. Además, como el proceso comienza con la duración de tiempo más pequeña, una vez que el tamaño del segmento es demasiado grande el bucle de partición termina (etapa 172) y la mejor solución (tiempo de duración, pares de canales, parámetros de codificación) para ese punto se empaqueta dentro de la cabecera (etapa 174) y el proceso se mueve a la siguiente trama. Si la restricción falla sobre el tamaño mínimo del segmento (etapa 176), entonces el proceso termina e informa de un error (etapa 178) porque la restricción del tamaño máximo no puede satisfacerse. Asumiendo que la restricción se satisface, este proceso se repite para cada uno de los segmentos en la partición actual hasta que el bucle de segmento termina (etapa 180).

Una vez que el bucle de segmento se ha completado y se ha calculado el consumo de bytes para toda la trama como se representa por `ByteConsEnPart`, esta carga de datos se compara con la carga de datos mínima actual (`MinByteEnPart`) a partir de una iteración de partición anterior (182). Si la partición actual representa una mejora entonces la partición actual (`PartInd`) se almacena como la partición óptima (`OptPartInd`) y la carga de datos mínima se actualiza (etapa 184). Estos parámetros y los parámetros de codificación almacenados se almacenan a continuación como la solución óptima actual (etapa 186). Esto se repite hasta que el bucle de partición termina (etapa 172), en cuyo momento la información de segmentación y los parámetros de codificación se empaquetan en la cabecera (etapa 150) como se muestra en la Figura 3.

Una realización de ejemplo para la determinación de los parámetros de codificación óptimos y el consumo de bits asociado para un conjunto de canales para la partición actual (etapa 158) se ilustra en las Figuras 8a y 8b. El proceso arranca un bucle de segmento (etapa 190) y el bucle de canal (etapa 192) en el cual los canales para nuestro ejemplo son:

50 Canal 1: L;
 Canal 2: R
 Canal 3: R – CanalParDecorrCoef [1] * L
 Canal 4: Ls
 Canal 5: Rs
 55 Canal 6: Rs – CanalParDecorrCoef [2] * Ls
 Canal 7: C
 Canal 8: LFE
 Canal 9: LFE – CanalParDecorrCoef[3] * C

60 El proceso determina el tipo de código de entropía, el parámetro de codificación correspondiente y el consumo de bits correspondiente para los canales base y correlado (etapa 194). En este ejemplo, el proceso calcula los parámetros de codificación óptima para un código binario y un código Rice y a continuación selecciona el código con el menor consumo de bits para el canal y cada uno de los segmentos (etapa 196). En general, la optimización puede realizarse para uno, dos o más códigos de entropía posibles. Para los códigos binarios el número de bits se calcula a partir del valor absoluto máximo de todas las muestras en el segmento del canal actual. El parámetro de codificación Rice se calcula a partir del valor absoluto promedio de todas las muestras en el segmento del canal

actual. En base a la selección, se fija el RiceCodifInd, se fija BitCons y se fija CodifParám para el NúmBitsBinario o el RiceKParám (etapa 198).

5 Si el canal actual que se está procesando es un canal correlado (etapa 200) entonces se repite la misma optimización para el canal descorrelado correspondiente (etapa 202), se selecciona el mejor código de entropía (etapa 204) y se fijan los parámetros de codificación (etapa 206). El proceso se repite hasta que el bucle de canal termina (etapa 208) y el bucle de segmento termina (etapa 210).

10 En este punto, se han determinado los parámetros de codificación óptima para cada uno de los segmentos y para cada uno de los canales. Estos parámetros de codificación y las cargas de datos podrían devolverse para los pares de canales (base, correlado) desde la señal de audio de PCM original. Sin embargo, el funcionamiento de la compresión puede mejorarse seleccionando entre los canales (base, correlado) y (base, descorrelado) en los tripletes.

15 Para determinar qué pares de canales (base, correlado) o (base, descorrelado) para los tres tripletes, se arranca un bucle de pares de canales (etapa 211) y se calcula la contribución de cada uno de los canales correlados (Canal 2, Canal 5 y Canal 8) y cada uno de los canales descorrelados (Canal 3, Canal 6 y Canal 9) para el consumo de bits de la trama global (etapa 212). Las contribuciones del consumo de la trama para cada uno de los canales correlados se compara frente a las contribuciones de consumo de la trama para los canales descorrelados correspondientes, es decir, Canal 2 para el Canal 3, Canal 5 para el Canal 6, y Canal 8 para el Canal 9 (etapa 214). Si la contribución para el canal descorrelado es mayor que el canal correlado, se fija PWCanalDecorrInd a falso (etapa 216). En caso contrario, el canal correlado se reemplaza por el canal descorrelado (etapa 218) y el PWCanalDecorrInd se fija verdadero y se configuran los pares de canales como (base, descorrelado) (etapa 220).

25 En base a estas comparaciones el algoritmo seleccionará:

1. El Canal 2 ó el Canal 3 como el canal que se emparejará con el canal base correspondiente Canal 1;
2. El Canal 5 ó el Canal 6 como el canal que se emparejará con el canal base correspondiente Canal 4; y
3. El Canal 8 ó el Canal 9 como el canal que se emparejará con el canal base correspondiente Canal 7;

30

Estas etapas se repiten para todos los pares de canales hasta que termina el bucle (etapa 222).

35 En este punto, se han determinado, los parámetros de codificación óptimos para cada uno de los segmentos y cada uno de los distintos canales y los pares de canales óptimos. Estos parámetros de codificación para cada uno de los distintos pares de canales y cargas de datos podrían devolverse al bucle de partición. Sin embargo, puede estar disponible un funcionamiento de compresión adicional calculando un conjunto de parámetros de codificación global para cada uno de los segmentos a través de todos los canales. A lo más, la porción de datos codificados de la carga de datos será del mismo tamaño que los parámetros de codificación optimizados para cada uno de los canales y lo más probable algo mayor. Sin embargo, la reducción en los bits de control puede más que la desviación de la eficacia de codificación de los datos.

40

45 Usando los mismos pares de canales, el proceso comienza un bucle de segmento (etapa 230), calcula los consumos de bits (CanalConjByteCons [seg]) por segmento para todos los canales usando los distintos conjuntos de parámetros de codificación (etapa 232) y almacena CanalConjByteCons [seg] (etapa 234). A continuación se determina un conjunto global de parámetros de codificación (selección del código de entropía y parámetros) para el segmento a través de todos los canales (etapa 236) usando los mismos cálculos del código binario y el código Rice que anteriormente excepto a través de todos los canales. Se seleccionan los mejores parámetros y se calcula el consumo de bytes (SegmByteCons) (etapa 238). El SegmByteCons se compara con el CanalConjByteCons [seg] (etapa 240). Si el uso de los parámetros globales no reduce el consumo de bits, se fija el indicador TodosCanalesIsgualParámInd [seg] a falso (etapa 242). De otro modo, el indicador TodosCanalesIsgualParámInd se fija a verdadero (etapa 244) y los parámetros de codificación global y el consumo de bits correspondiente por segmento se almacenan (etapa 246). Este proceso se repite hasta que se alcanza el fin del bucle del segmento (etapa 248). Todo el proceso se repite hasta que termina el bucle del conjunto de canales (etapa 250).

50

55 El proceso de codificación está estructurado de modo que las diferentes funcionalidades pueden deshabilitarse por el control de unos pocos indicadores. Por ejemplo un indicador único controla si el análisis de des-correlación del canal orientado a pares se va a realizar o no. Otro indicador controla si se realizará o no el análisis de la predicción adaptativa (otro indicador más para la predicción fija). Además un indicador único controla si se realizará o no la búsqueda de parámetros globales sobre todos los canales. La segmentación es también controlable fijando el número de particiones y la duración mínima del segmento (en la forma más simple puede ser una única partición con una duración de segmento predeterminada). En esencia, fijando unos pocos indicadores en el codificador, el codificador puede colapsar a una codificación simple de entramado y entropía.

60

65 **CÓDEC DE AUDIO SIN PÉRDIDAS COMPATIBLE CON VERSIONES ANTERIORES**

El códec sin pérdidas puede usarse como un "codificador de extensión" en combinación con un codificador de

núcleo con pérdidas. Un flujo del codificador de núcleo "con pérdidas" se empaqueta como un flujo de bits del núcleo y una señal de diferencia codificada sin pérdidas se empaqueta como un flujo de bits de extensión separado. Una vez realizada la decodificación en un decodificador extendido con características sin pérdidas, los flujos con pérdidas y sin pérdidas se combinan para construir una señal reconstruida sin pérdidas. En un decodificador de una generación anterior, el flujo sin pérdidas se ignora, y flujo "con pérdidas" del núcleo se decodifica para proporcionar una señal de audio multicanal de alta calidad con el ancho de banda y la proporción de señal a ruido característica del flujo del núcleo.

La Figura 9 muestra una vista a nivel de sistema de un codificador sin pérdidas compatible con versiones anteriores 400 para un canal de una señal multicanal. Se proporcionan a la entrada como señal de audio digitalizada, muestras de audio PCM de M bits adecuadas 402. Preferiblemente, la señal de audio digitalizada tiene una tasa de muestreo y un ancho de banda que excede al de un codificador de núcleo con pérdidas modificado 404. En una realización, la tasa de muestreo de la señal de audio digitalizada es de 96 KHz (correspondiente a un ancho de banda de 48 KHz para la señal de audio muestreada). Debería entenderse también que la señal de audio de entrada puede ser, y preferiblemente es, una señal multicanal en donde cada canal se muestrea a 96 KHz. La discusión que sigue se concentrará en el procesamiento de un canal único, pero la extensión para múltiples canales es sencilla. La señal de entrada se duplica en el nodo 406 y se maneja en ramas paralelas. En una primera rama de la trayectoria de la señal, un codificador con pérdidas modificado, de banda ancha 404 codifica la señal. El codificador de núcleo modificado 404, que se describe con detalle más adelante, produce un flujo de bits de núcleo codificados 408 que se dirige a un empaquetador o multiplexor 410. El flujo de bits de núcleo 408 se comunica también a un decodificador del núcleo modificado 412, que produce como salida una señal de núcleo reconstruida modificada 414.

Entre tanto, la señal de audio digitalizada de entrada 402 en la trayectoria en paralelo se somete a un retardo de compensación 416, sustancialmente igual que el retardo introducido en el flujo de audio reconstruido (por el codificador modificado y los decodificadores modificados), para producir un flujo de audio digitalizado retardado. El flujo de audio 400 se resta del flujo de audio digitalizado retardado 414 en el nodo sumador 420. El nodo sumador 420 produce una señal diferencia 422 que representa la señal original y la señal del núcleo reconstruida. Para cumplir la codificación puramente "sin pérdidas", es necesario codificar y transmitir la señal de diferencia con técnicas de codificación sin pérdidas. Por consiguiente, la señal de diferencia 422 se codifica con un codificador sin pérdidas 424, y el flujo de bits de extensión 426 se empaqueta con el flujo de bits de núcleo 408 en el empaquetador 410 para producir un flujo de bits de salida 428.

Obsérvese que la codificación sin pérdidas produce un flujo de bits de extensión 426 que es a una tasa de bits variable, para acomodarse a las necesidades del codificador sin pérdidas. El flujo empaquetado se sujeta a continuación opcionalmente a capas adicionales de codificación incluyendo la codificación de canal, y a continuación se transmite o se graba. Obsérvese que para los propósitos de esta revelación, la grabación puede considerarse como una transmisión a través de un canal.

El codificador de núcleo 404 se describe como "modificado" porque en una realización capaz de manejar el ancho de banda extendido el codificador del núcleo requeriría modificación. Un banco de filtros de análisis de 64 bandas 430 dentro del codificador rechaza la mitad de sus datos de salida 432 y el codificador de sub-bandas del núcleo 434 codifica sólo las 32 bandas de frecuencia inferiores. Esta información descartada no es concerniente a decodificadores heredados que serían incapaces de reconstruir la mitad superior del espectro de la señal en ningún caso. La información restante se codifica como por un codificador no modificado para formar un flujo de salida del núcleo compatible con versiones anteriores. Sin embargo, en otra realización que funciona a una tasa de muestreo de 48 KHz o inferior, el codificador del núcleo podría ser una versión sustancialmente no modificada de un codificador del núcleo anterior. De forma similar, para el funcionamiento por encima de la tasa de muestreo de los decodificadores heredados, el decodificador de núcleo modificado 412 incluye un decodificador de sub-banda de núcleo 436 que decodifica las muestras en las 32 sub-bandas inferiores. El decodificador de núcleo modificado toma las muestras de las sub-bandas a partir de las 32 sub-bandas inferiores y los ceros fuera de las muestras de las sub-bandas no transmitidas para las 32 bandas superiores 438 y reconstruye todas las 64 bandas usando un filtro de síntesis de QAF de 64 bandas 440. Para el funcionamiento a la tasa de muestreo convencional (por ejemplo, 48 KHz y por debajo) el decodificador del núcleo podría ser una versión sustancialmente sin modificar de un decodificador del núcleo anterior o equivalente. En algunas realizaciones la elección de la tasa de muestreo podría realizarse en el momento de la codificación, y los módulos de codificación y decodificación reconfigurados en ese momento por software como se desee.

Como el codificador sin pérdidas se está usando para codificar la señal diferencia, puede parecer que sería suficiente un código de entropía simple. Sin embargo, debido a las limitaciones de la tasa de bit sobre los códec del núcleo con pérdidas existentes, se requiere una cantidad considerable de bits totales para proporcionar un flujo de bits sin pérdidas aún restante. Además, debido a las limitaciones del ancho de banda del códec del núcleo el contenido de información por encima de 24 KHz en la señal diferencia está aún correlada. Por ejemplo una abundancia de componentes armónicos incluyendo trompeta, guitarra, triángulo,... que alcanzan más allá de 30 KHz. Por lo tanto los códec sin pérdidas más sofisticados que mejoran el funcionamiento de la compresión dan un valor añadido. Además, en algunas aplicaciones los flujos de bits del núcleo y la extensión deben satisfacer aún la restricción de que las unidades decodificables no deben exceder un tamaño máximo. Los códec sin pérdidas de la

presente invención proporcionan tanto un funcionamiento de la compresión mejorado como una flexibilidad mejorada para satisfacer estas restricciones.

5 A modo de ejemplo, 8 canales de una señal de audio de PCM a 96 KHz de 24 bits requieren 18,5 Mbps. La compresión sin pérdidas puede reducir esto a aproximadamente 9 Mbps, La Acústica Coherente DTS codificaría el núcleo a 1,5 Mbps, dejando una señal diferencia de 7,5 Mbps. Para un tamaño de segmento máximo de 2 KBytes, la duración del segmento promedio es de $2048 * 8 / 7500000 = 2,18$ mseg o aproximadamente 209 muestras a 96 KHz. Un tamaño de trama típico para el núcleo con pérdidas para satisfacer el tamaño máximo está entre 10 y 20 mseg.

10 A nivel de sistema, el códec sin pérdidas y el códec sin pérdidas compatible con las versiones anteriores pueden combinarse para una codificación sin pérdidas de canales extra de audio en un ancho de banda extendido mientras que se mantiene la compatibilidad con las versiones anteriores de los códec con pérdidas existentes. Por ejemplo, 8 canales de señal de audio de 96 KHz a 18,5 MHz pueden codificarse sin pérdidas para incluir canales 5.1 de señal de audio de 48 KHz a 1,5 Mbps. El núcleo más el codificador sin pérdidas se usaría para codificar los canales 5.1. El
15 codificador sin pérdidas se usará para codificar las señales de diferencia en los canales 5.1. Los restantes 2 canales se codifican en un conjunto de canales separados usando un codificador sin pérdidas. Como todos los conjuntos de canales necesitan considerarse cuando se intenta optimizar la duración del segmento, se usarán todas las herramientas de codificación de un modo o de otro. Un decodificador compatible decodificaría todos los 8 canales y reconstruirá sin pérdidas la señal de audio de 18,5 Mbps de 96 KHz. Un decodificador más antiguo decodificaría sólo
20 los canales 5.1 y reconstruiría la señal de 1,5 MHz a 48 KHz.

En general, pueden proporcionarse más de un canal puro sin pérdidas para los propósitos de adaptar la complejidad del decodificador. Por ejemplo, para una mezcla de 10,2 original los conjuntos de canales podrían organizarse de modo que:

25 - CANALCONJ 1 transporta 5.1 (con 10.2 incorporado para una mezcla inferior de 5.1) y se codifica usando núcleo + sin pérdidas
- CANALCONJ 1 y CANALCONJ 2 transportan 7.1 (con 10.2 incorporado para una mezcla inferior de 7.1) donde CANALCONJ 2 codifica 2 canales usando sin pérdidas
30 - CANALCONJ 1 + CANALCONJ 2 + a CANALCONJ 3 transportan la mezcla discreta total de 10.2 donde CANALCONJ 3 codifica los canales 3.1 restantes usando sólo sin pérdidas.

Un decodificador que es capaz de decodificar sólo 5.1 decodificará el CANALCONJ 1 e ignorará los otros conjuntos de canales. Un decodificador que es capaz de decodificar sólo 7.1 decodificará CANALCONJ 1 y CANALCONJ 2 e
35 ignorará todos los otros conjuntos de canales, ...

Además, la señal con pérdidas más el núcleo sin pérdidas no está limitado a 5.1. Las implementaciones actuales soportan hasta 6.1 usando la señal con pérdidas (núcleo + XCh) y sin pérdidas y puede soportar unos canales genéricos m.n organizados en cualquier número de conjuntos de canales. La codificación con pérdidas tendrá un
40 núcleo compatible con las versiones anteriores de 5.1 y todos los demás canales que se codifican con el códec con pérdidas irán dentro de la versión XXCh. Esto proporciona el códec sin pérdidas global con una flexibilidad de diseño considerable para mantener la compatibilidad hacia atrás con los decodificadores existentes mientras que se soportan canales adicionales.

45 Aunque se han mostrado y descrito varias realizaciones ilustrativas de la invención, se ocurrirán numerosas variaciones y realizaciones alternativas a los especialistas en la técnica. Tales variaciones y realizaciones alternativas se contemplan, y pueden realizarse sin apartarse del ámbito de la invención como se define en las reivindicaciones adjuntas.

50

REIVINDICACIONES

1. Un método para codificar sin pérdidas datos de audio de PCM, que comprende:

5 poner en bloques la señal de audio multicanal en tramas de igual tiempo de duración;
 procesar la señal de audio multicanal para ordenar los canales en pares incluyendo un canal base y un canal
 correlado; en el que los dos canales más correlados forman un primer par y así sucesivamente hasta que se
 agotan los canales, si queda un canal impar este forma un canal base;
 10 determinar una estimación de correlación cruzada de retardo cero para los pares de canales;
 determinar una estimación de auto-correlación de retardo cero del canal base;
 procesar el orden de los pares de canales para determinar un coeficiente de des-correlación dividiendo la
 estimación de correlación cruzada de retardo cero por la estimación de auto-correlación de retardo cero del
 canal base;
 15 generar un canal descorrelado para cada uno de los pares de canales para formar al menos un triplete (base,
 correlado, descorrelado), en el que el canal descorrelado se genera multiplicando el canal base por el
 coeficiente de des-correlación y restando el resultado del canal correlado;
 seleccionar los parámetros de codificación en base a las combinaciones posibles de pares de canales de
 dichos canales base y correlado y dichos canales base y descorrelado;
 20 seleccionar pares de canales (base, correlado) o (base, descorrelado) extraídos de cada uno de dichos
 tripletes;
 codificar la entropía de cada uno de los canales en los pares seleccionados de acuerdo con los parámetros de
 codificación; y
 empaquetar los datos de audio codificados en un flujo de bits.

25 2. El método de la reivindicación 1, en el que en cada uno de los pares el canal que tiene la estimación de auto-
 correlación de retardo cero más pequeña es el canal base.

30 3. El método de la reivindicación 1, en el que la etapa de procesamiento incluye el procesamiento de la señal de
 audio multicanal para crear pares de canales incluyendo el canal base y el canal correlado, y el método comprende
 además:

segmentar cada una de las tramas en una pluralidad de segmentos de un tiempo de duración predeterminado,
 y en el que la etapa de seleccionar pares de canales minimiza una carga de datos codificados de la trama
 sujeta a la restricción de que cada uno de los segmentos debe ser decodificable y menor de un tamaño
 35 máximo; y
 en el que la etapa de codificación de entropía incluye codificar la entropía de cada uno de los segmentos de
 cada uno de los canales en los pares seleccionados de acuerdo con los parámetros de codificación.

40 4. El método de la reivindicación 3, en el que la duración predeterminada de los segmentos se determina en parte
 seleccionando uno de la pluralidad de codificadores de entropía y sus parámetros de codificación.

45 5. El método de la reivindicación 3, en el que a cada uno de los canales se asigna un conjunto de parámetros de
 codificación incluyendo el codificador de entropía seleccionado y sus parámetros, la duración del segmento se
 determina en parte seleccionando bien un conjunto distinto de parámetros de codificación para cada uno de los
 canales o un conjunto global de parámetros de codificación para dicha pluralidad de canales.

6. El método de la reivindicación 3, en el que la duración predeterminada es la misma para cada segmento en una
 trama.

50 7. El método de la reivindicación 3, en el que la duración predeterminada está determinada para cada una de las
 tramas y varía sobre la secuencia de tramas.

8. El método de la reivindicación 3, en el que la duración predeterminada se determina,

55 a) realizando una partición de la trama en un número de segmentos de una duración determinada;
 b) determinando un conjunto de parámetros de codificación y carga de datos codificados para cada uno de los
 segmentos en cada uno de los canales;
 c) calculando las cargas de datos codificados para cada uno de los segmentos a través de todos los canales;
 d) si la carga de datos codificados a través de todos los canales para cualquier segmento excede el tamaño
 60 máximo, descartando el conjunto de parámetros de codificación;
 e) si la carga de datos codificados para la trama para la partición actual es menor de una carga de datos
 codificados mínima para particiones anteriores, almacenando el conjunto actual de parámetros de codificación
 y actualizando la carga de datos codificados mínima; y
 f) repitiendo las etapas desde a) hasta e) para la pluralidad de segmentos de una duración diferente.

65 9. El método de la reivindicación 8, en el que la duración del segmento se fija a una duración mínima inicialmente y

se aumenta en cada iteración de partición.

- 5 10. El método de la reivindicación 9, en el que la duración del segmento se fija inicialmente a una potencia de dos y se dobla en cada una de las iteraciones de partición.
11. El método de la reivindicación 9, en el que si la carga de datos codificados a través de todos los canales para cualquier segmento excede el tamaño máximo, la iteración de partición termina.
- 10 12. El método de la reivindicación 8, en el que el conjunto de parámetros de codificación incluye una selección de un codificador de entropía y sus parámetros.
13. El método de la reivindicación 12, en el que el codificador de entropía y sus parámetros se seleccionan para minimizar la carga de datos codificados para ese segmento en ese canal.
- 15 14. El método de la reivindicación 8, que comprende además generar un canal descorrelado para pares de canales para formar un triplete (base, correlado, descorrelado), seleccionando bien el par de canales (base, correlado) o el par de canales (base, descorrelado), y codificar la entropía de los canales en los pares de canales seleccionados.
- 20 15. El método de la reivindicación 8, en el que el conjunto determinado de parámetros de codificación es bien distinto para cada uno de los canales o global para todos los canales en base a lo cual produce una carga de datos codificados más pequeña incluyendo tanto el control como los datos de audio para la trama.
- 25 16. El método de la reivindicación 3, en el que la duración predeterminada del segmento se determina para minimizar la carga de datos codificados de cada una de las tramas.
17. El método de la reivindicación 3, en el que la duración predeterminada del segmento se determina en parte seleccionando un conjunto de parámetros de codificación incluyendo uno de la pluralidad de codificadores de entropía y sus parámetros de codificación para cada uno de los segmentos.
- 30 18. El método de la reivindicación 17, en el que la duración predeterminada del segmento se determina en parte seleccionando bien un conjunto distinto de parámetros de codificación para cada uno de los canales o un conjunto global de parámetros de codificación para dicha pluralidad de canales.
- 35 19. El método de la reivindicación 17, en el que los conjuntos de parámetros de codificación se calculan para diferentes duraciones de segmentos y se selecciona la duración correspondiente al conjunto que tiene la carga de datos codificados más pequeña que satisface la restricción sobre el segmento máximo.
- 40 20. El método de la reivindicación 3, que comprende además generar un canal descorrelado para pares de canales para formar al menos un triplete (base, correlado, descorrelado), la duración predeterminada del segmento se determina en parte seleccionando bien un par de canales (base, correlado) o un par de canales (base, descorrelado) para cada uno de dichos tripletes para la codificación de entropía.
- 45 21. El método de la reivindicación 20, en el que los pares de canales se seleccionan determinando si el canal descorrelado o correlado contribuye con el menor número de bits para la carga de datos codificados.
22. El método de la reivindicación 20, en el que dos canales más correlados forman parte de un primer par y así sucesivamente hasta que se agotan los canales, si queda un número impar de canales se forma un canal base.
- 50 23. El método de la reivindicación 22, en el que en cada uno de los pares, el canal que tiene la estimación de auto-correlación de retardo cero más pequeña es el canal base.
24. El método de la reivindicación 23, en el que el canal descorrelado se genera multiplicando el canal base por un coeficiente de des-correlación y restando el resultado del canal correlado.

55

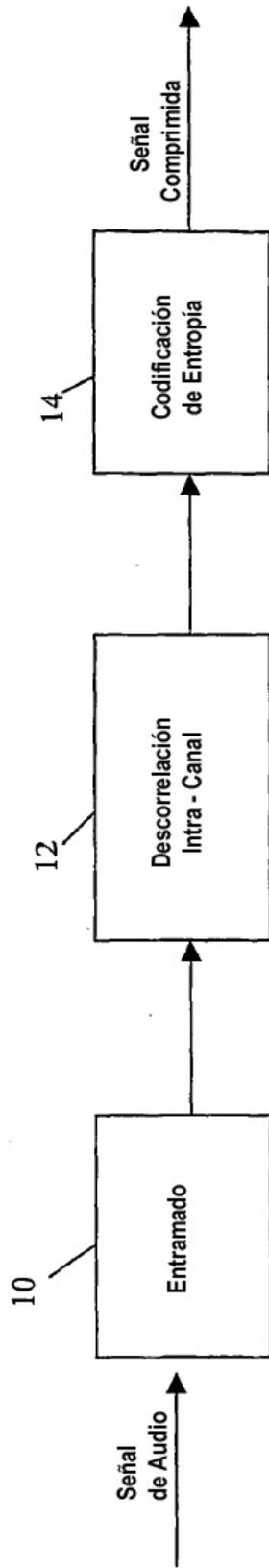


Fig. 1 (Técnica Anterior)

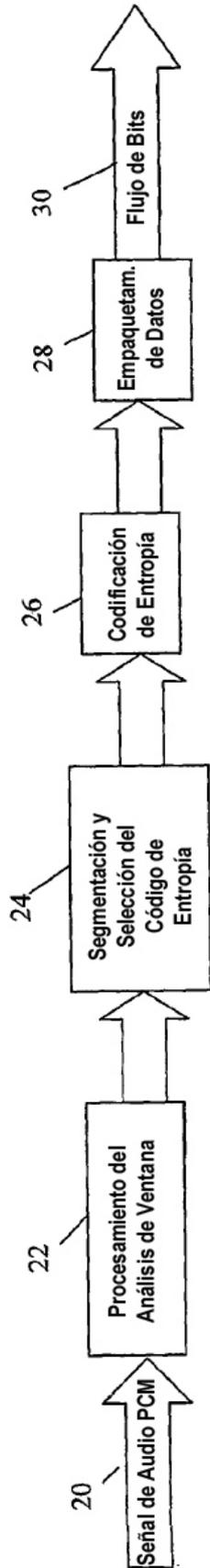


Fig. 2a

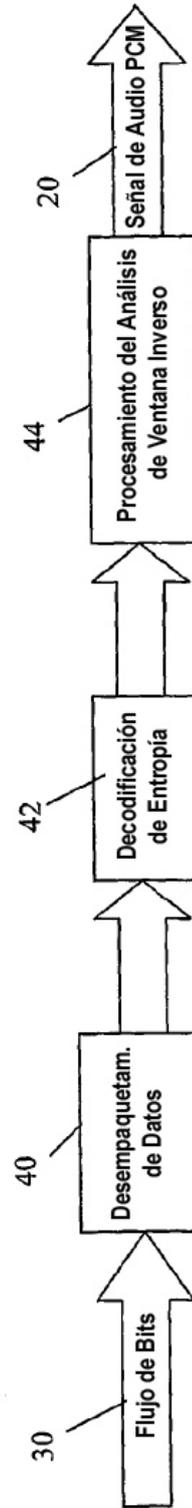


Fig. 2b

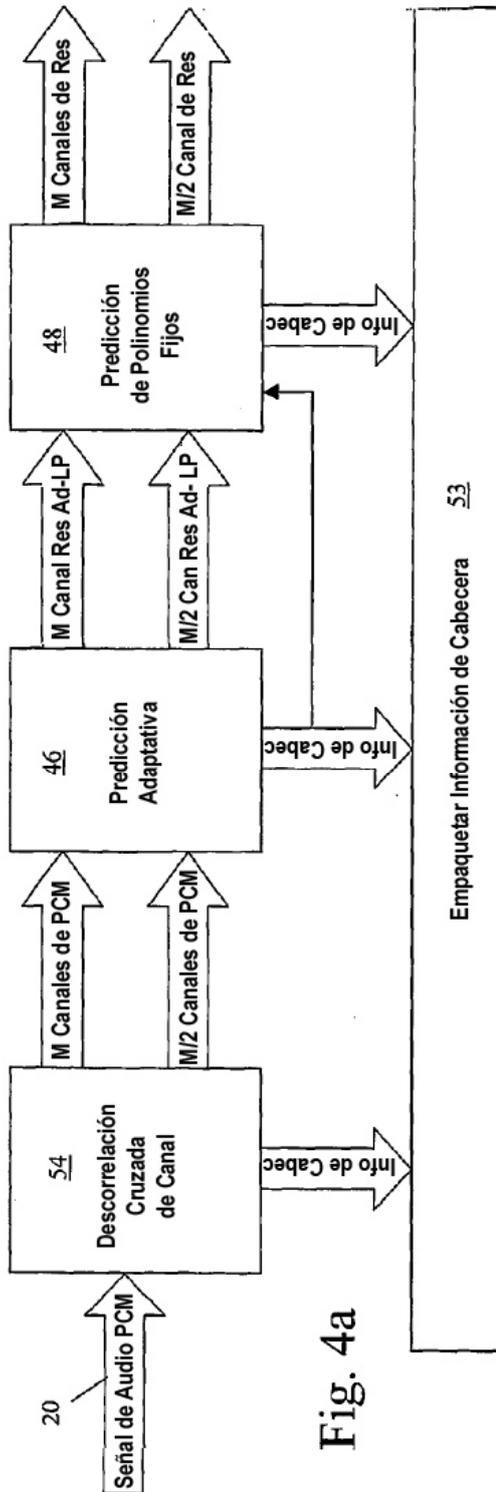


Fig. 4a

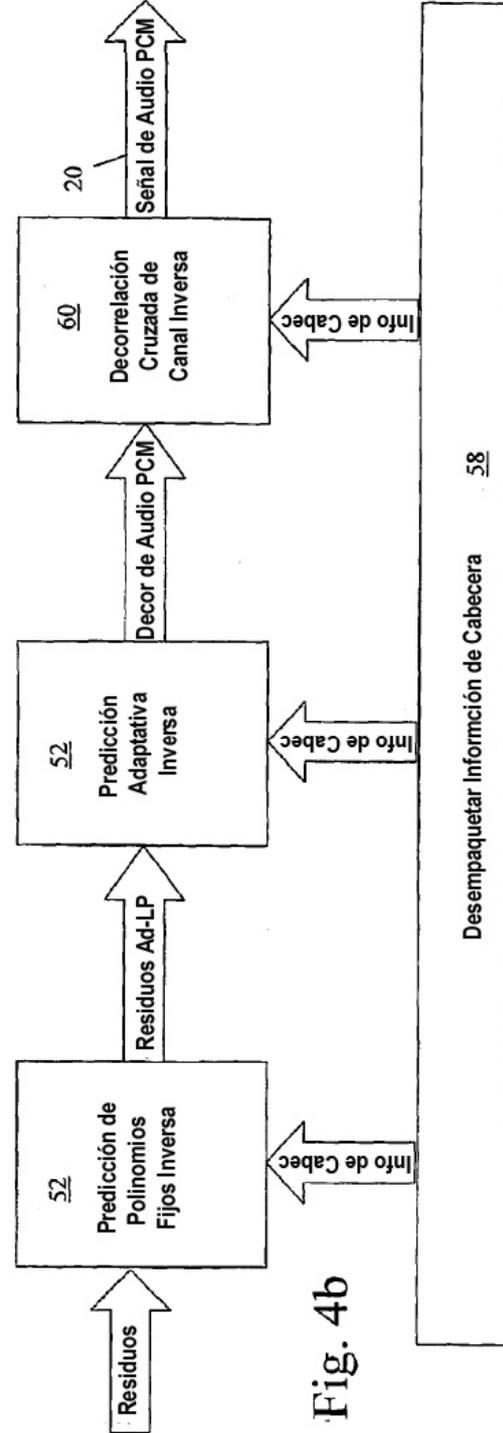


Fig. 4b

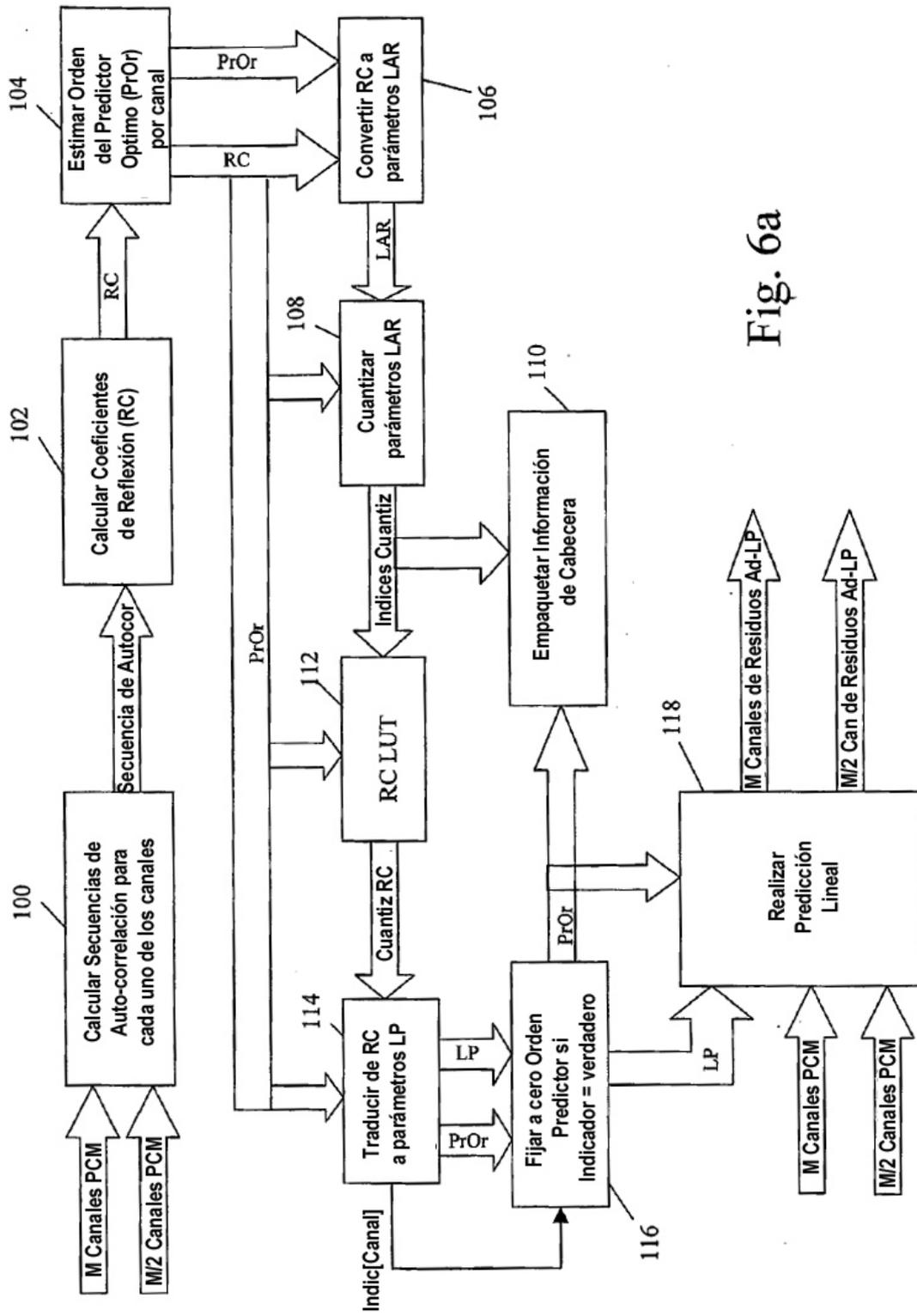


Fig. 6a

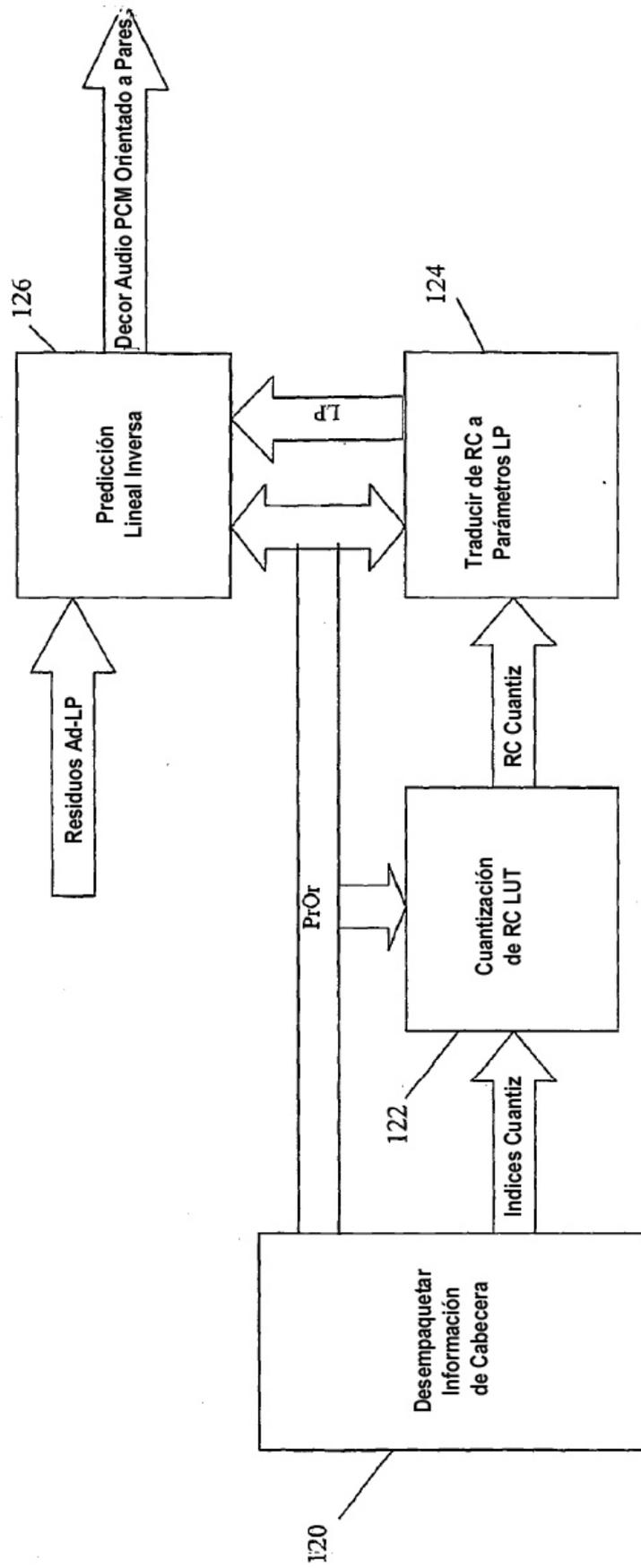


Fig. 6b

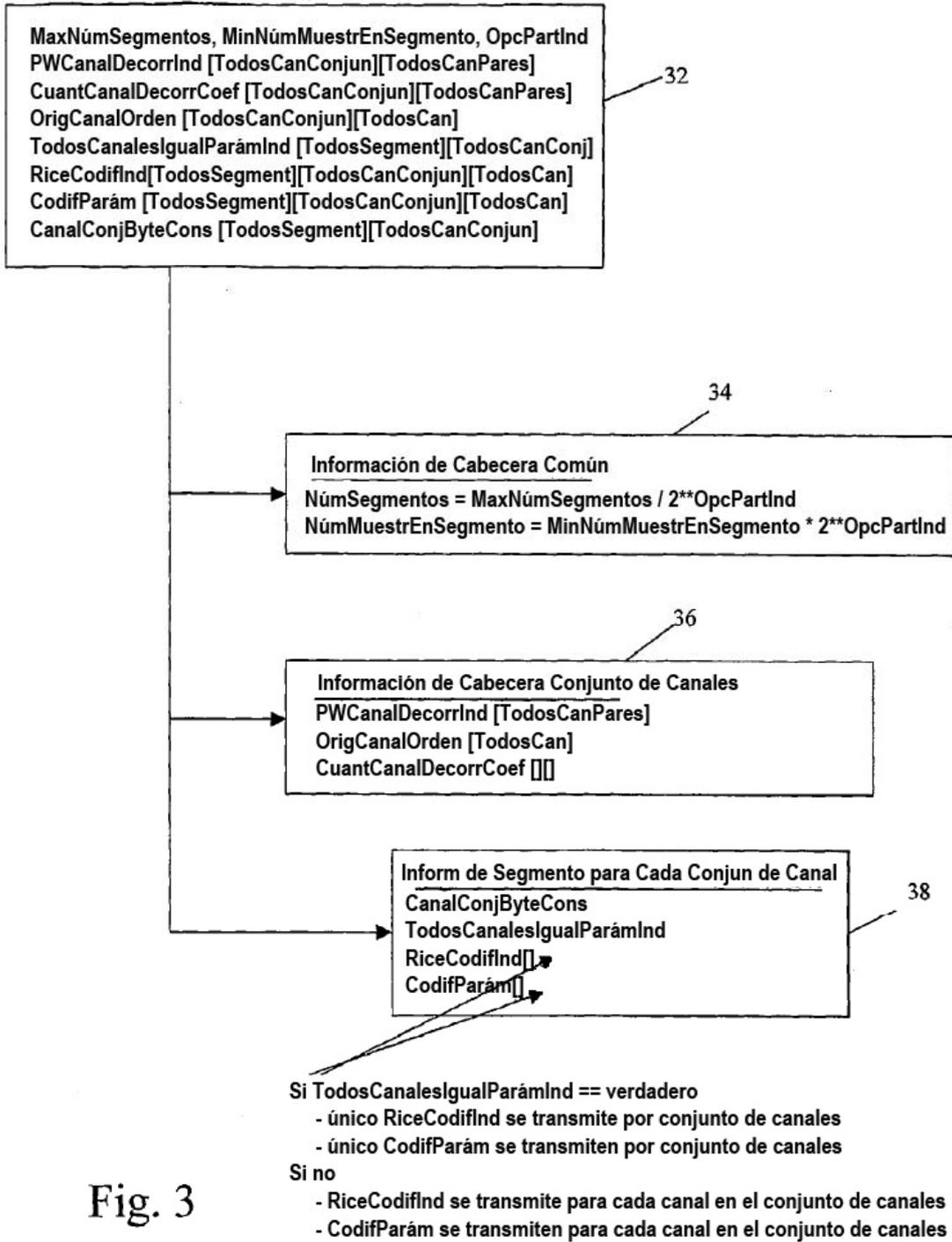


Fig. 3

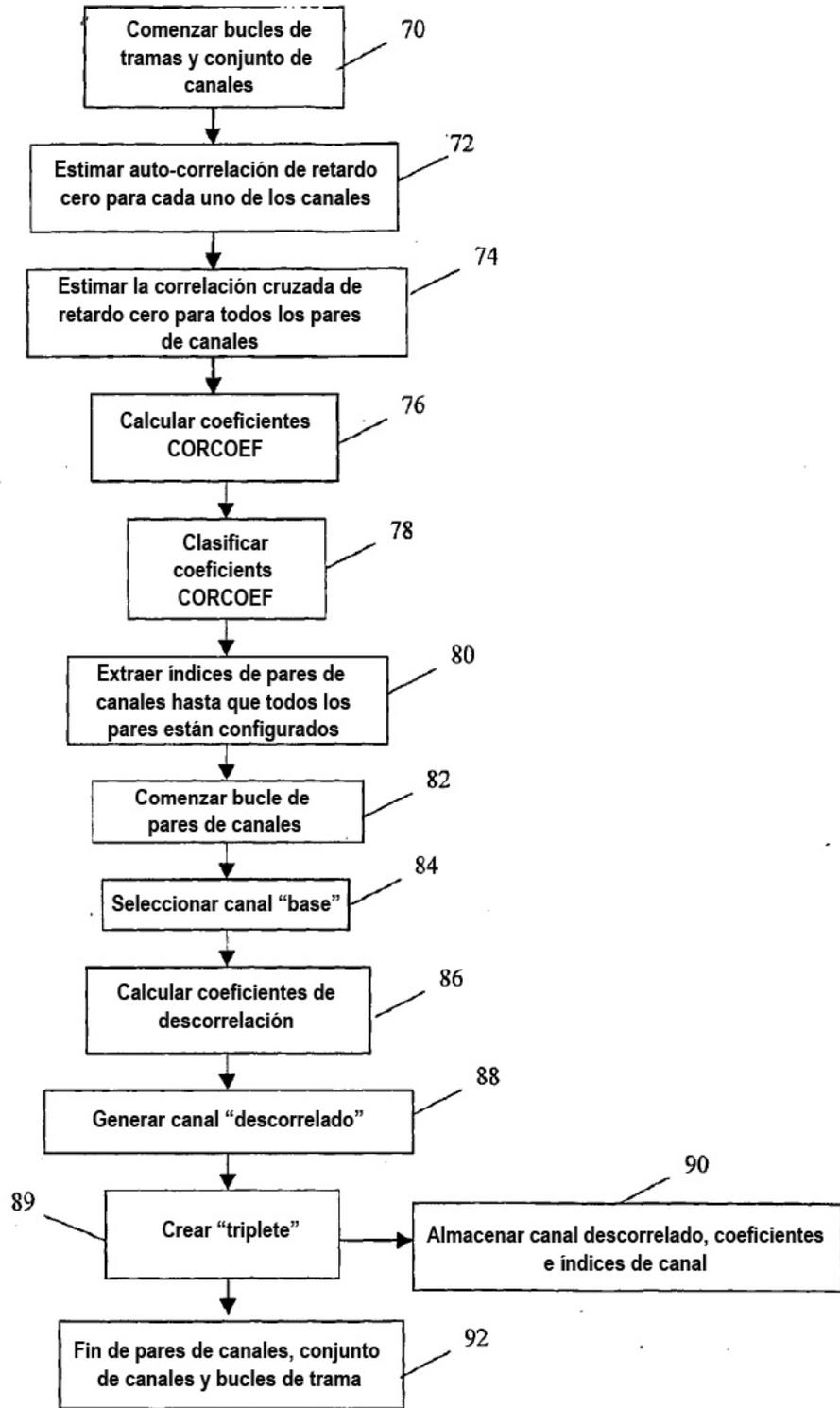


Fig. 5

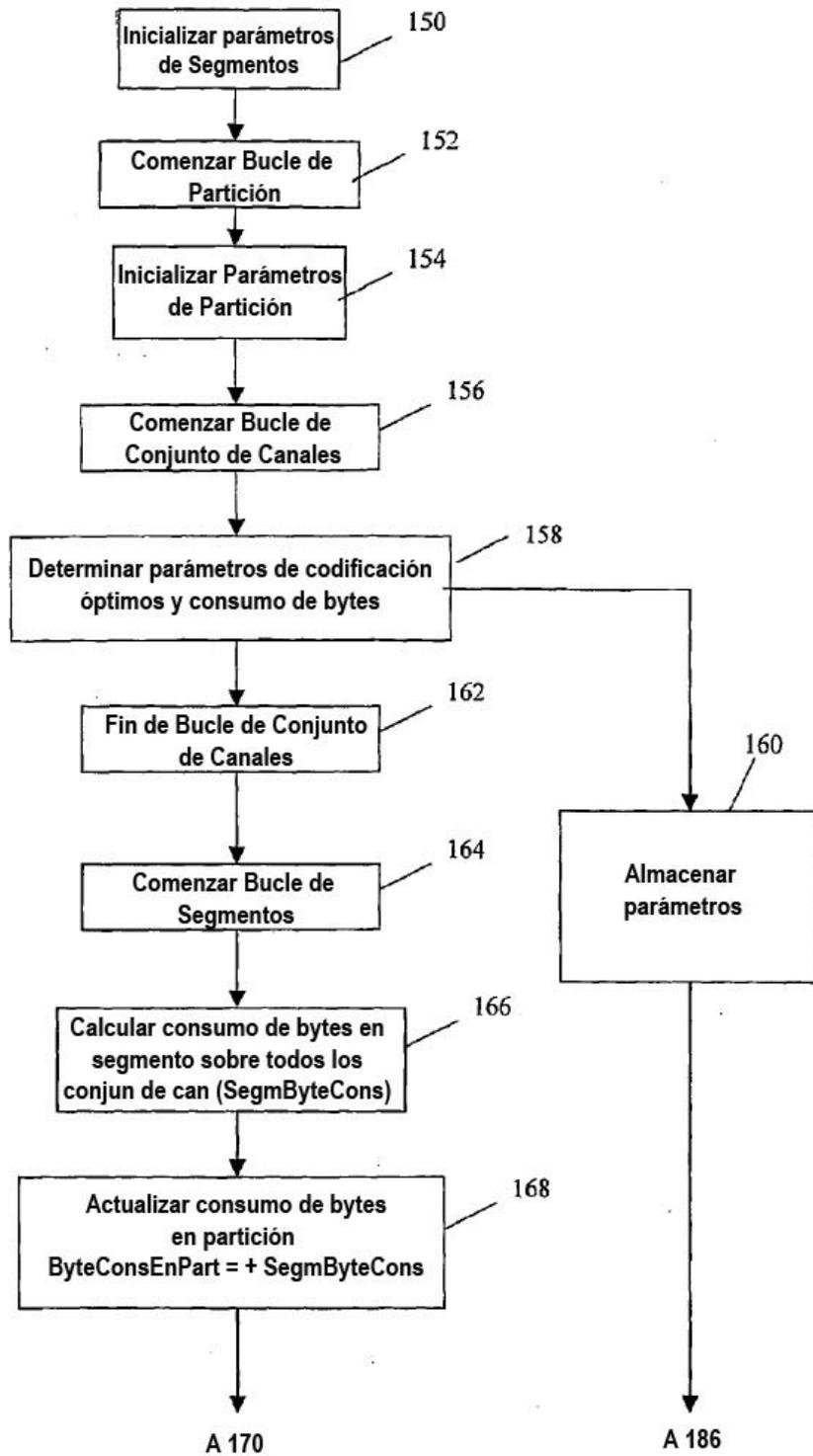


Fig. 7a

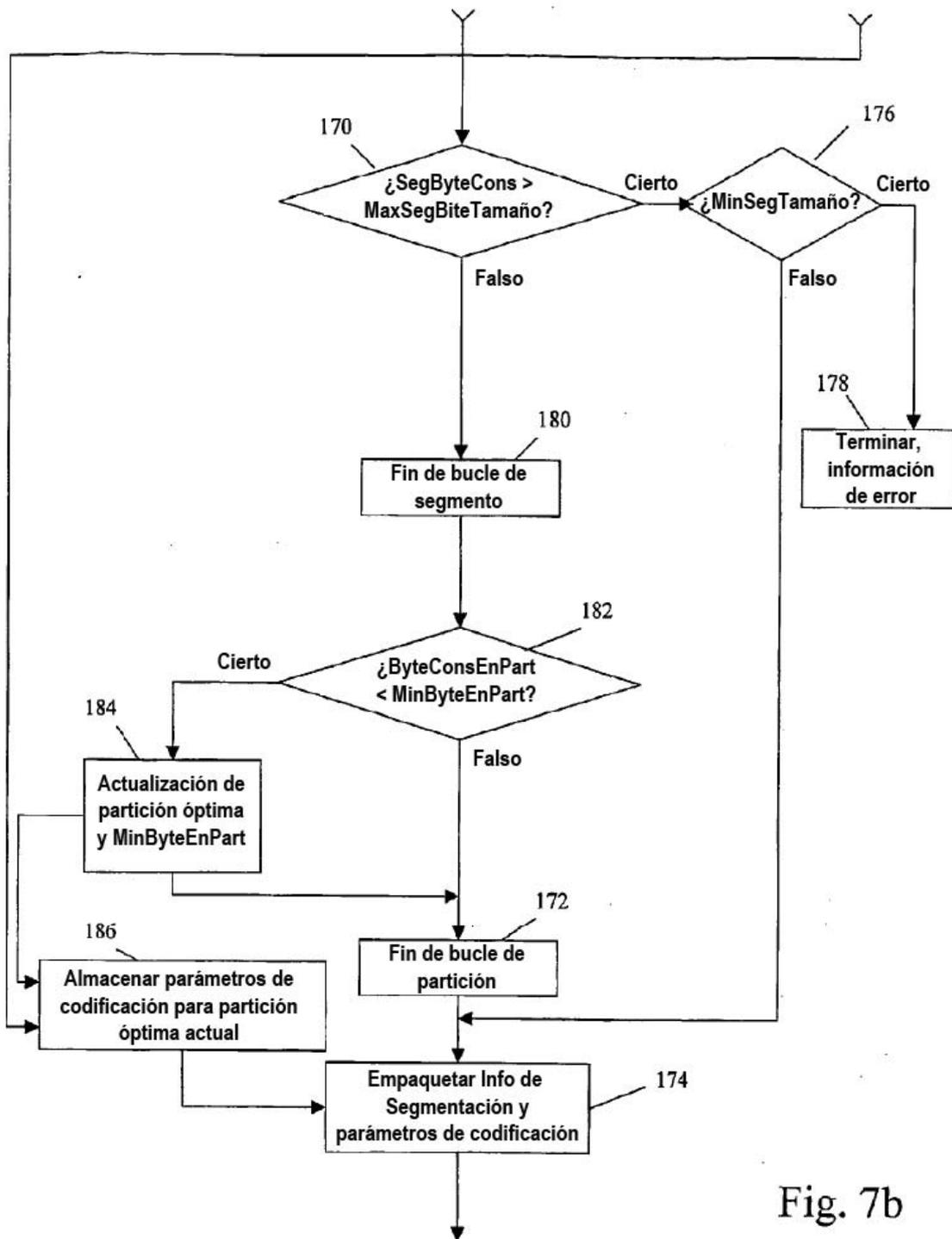


Fig. 7b

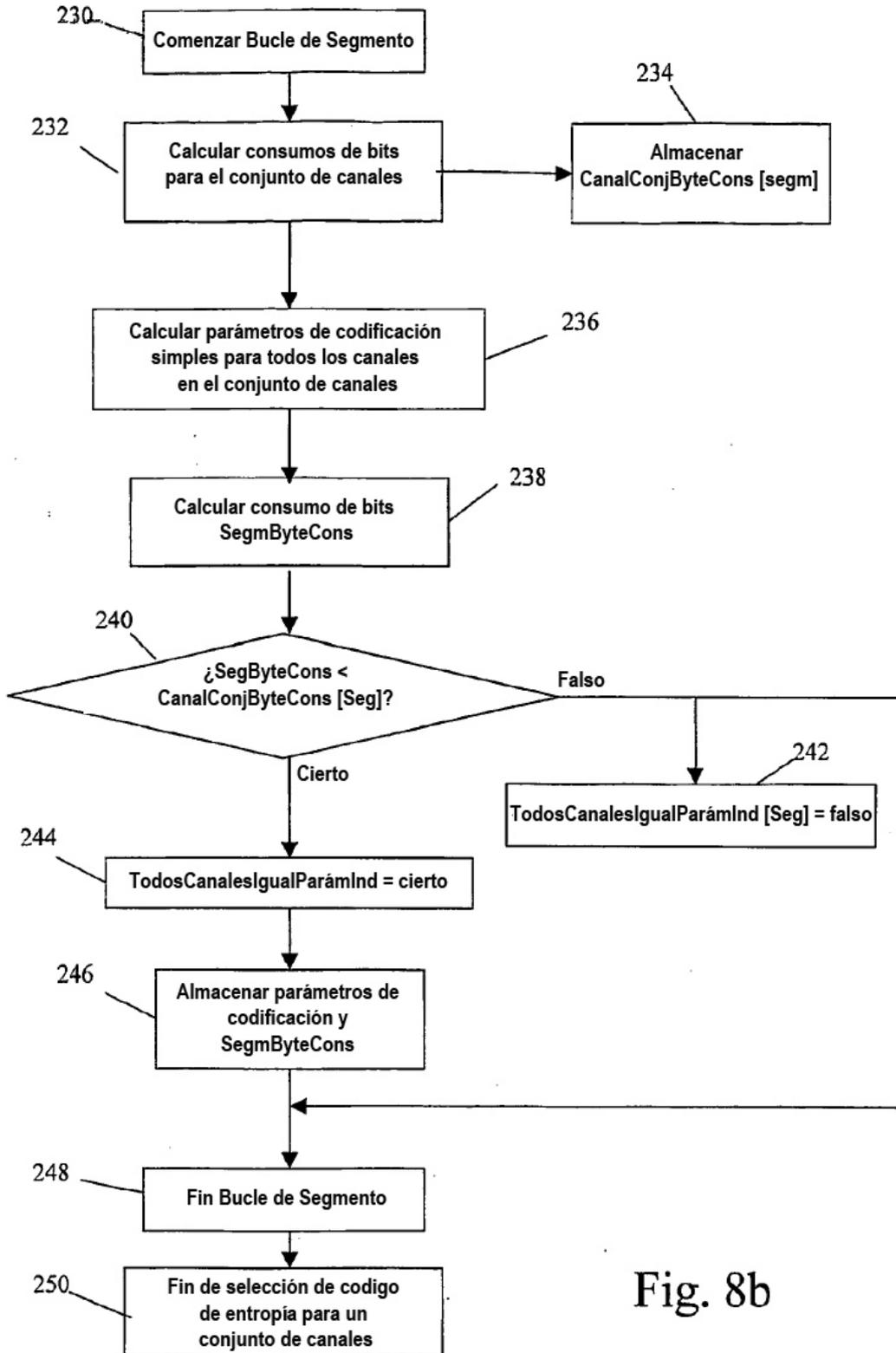


Fig. 8b

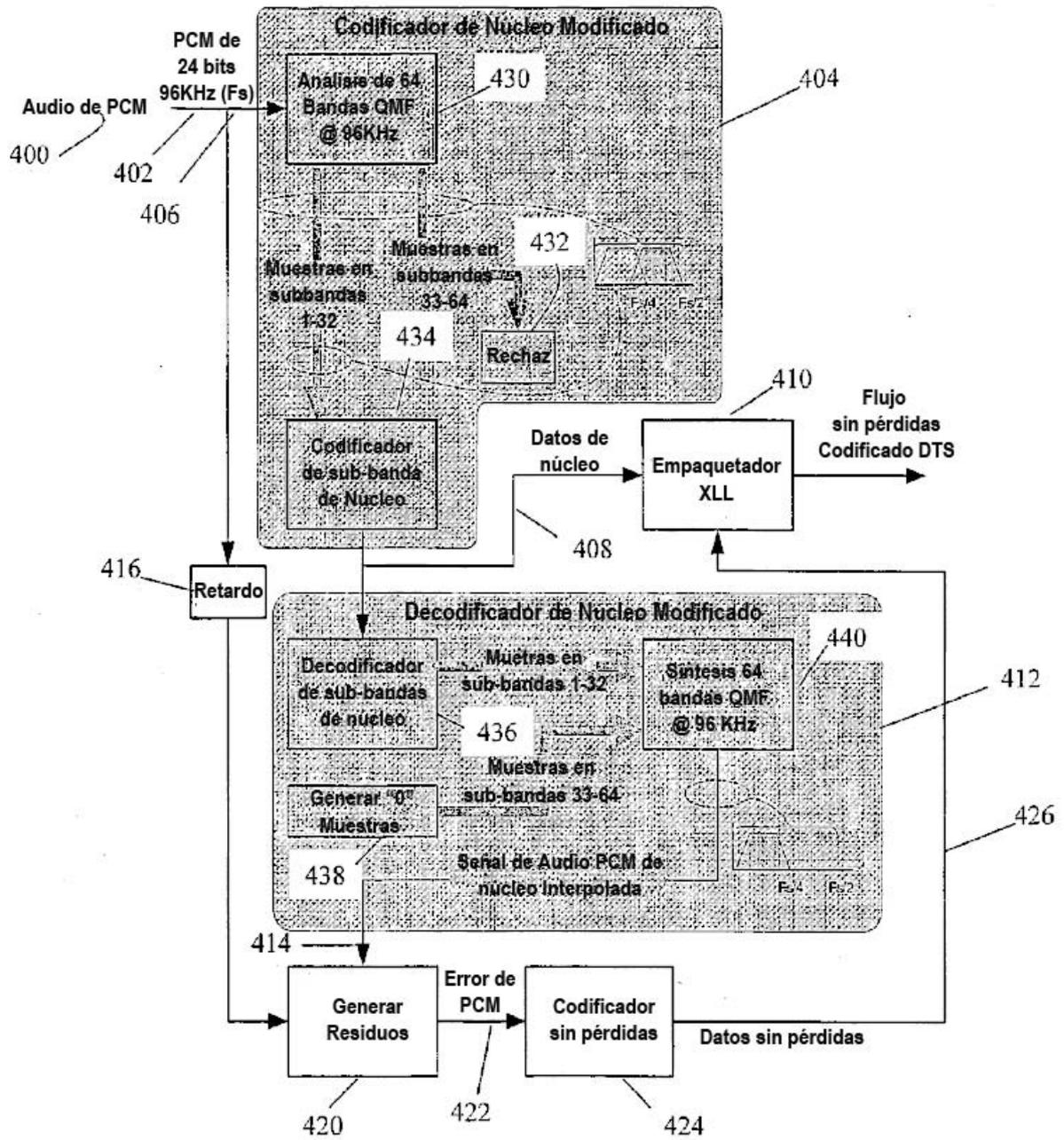


Fig. 9