



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 364 471**

51 Int. Cl.:
H04S 3/00 (2006.01)
H04S 7/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **99900039 .1**
96 Fecha de presentación : **06.01.1999**
97 Número de publicación de la solicitud: **1072089**
97 Fecha de publicación de la solicitud: **31.01.2001**

54 Título: **Procedimiento y aparato de procesamiento de señales de audio.**

30 Prioridad: **25.03.1998 AU PP2595**
31.03.1998 AU PP2710
31.03.1998 AU PP2709
12.11.1998 AU PP7164

45 Fecha de publicación de la mención BOPI:
05.09.2011

45 Fecha de la publicación del folleto de la patente:
05.09.2011

73 Titular/es:
DOLBY LABORATORIES LICENSING Corp.
100 Portrero Avenue
San Francisco, California 94103, US

72 Inventor/es: **McGrath, David, Stanley;**
McKeag, Adam, Richard;
Dickens, Glenn, Norman;
Cartwright, Richard, James y
Reilly, Andrew, Peter

74 Agente: **Sugrañes Moliné, Pedro**

ES 2 364 471 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

Procedimiento y aparato de procesamiento de señales de audio

5 Campo de la invención

La presente invención se refiere al campo del procesamiento de señales de audio y, en particular, desvela procedimientos de convolución eficaces para la convolución de señales de audio de entrada con funciones de respuesta de impulso o similares.

10

Antecedentes de la invención

15 La solicitud PCT internacional número PCT/AU93/00330 titulada "*Digital Filter Having High Accuracy and Efficiency*", presentada por el presente solicitante, desvela un proceso de convolución que tiene una latencia extremadamente baja además de permitir una larga convolución eficaz de funciones detalladas de respuesta de impulso.

20 Se conoce la utilización de la convolución de funciones de respuesta de impulso para añadir "color" a señales de audio de manera que cuando, por ejemplo, se reproducen en auriculares, las señales proporcionan una experiencia de escucha "fuera de la cabeza". Desafortunadamente, el proceso de convolución, aunque utiliza avanzas técnicas algorítmicas tales como la transformada rápida de Fourier (FFT, *Fast Fourier Transform*), requiere frecuentemente un excesivo tiempo computacional. Los requisitos computacionales aumentan normalmente cuando múltiples canales deben convolucionarse de manera independiente, como es normalmente el caso cuando se requieren capacidades de sonido totalmente envolventes. Normalmente, los procesadores DSP de los módems no pueden proporcionar los recursos para una convolución total de las señales, especialmente cuando hay restricciones de tiempo real en la latencia de la convolución.

25

30 La patente estadounidense 5491754 a nombre de Jot et al., describe una espacialización artificial de señales de audio utilizando una cola de reverberación generada artificialmente. En particular, se describe un procedimiento para una espacialización artificial de señales digitales de audio, haciendo posible que actúe sobre señales elementales, réplicas de la señal digital de audio, donde diferentes retardos crean señales elementales retardadas sumadas después de ponderarse con la señal digital de audio con el fin de crear la señal digital de audio espacializada. Una pluralidad de combinaciones lineales de las señales como señales elementales retardadas y combinadas se suma con las señales elementales para simular una reverberación posterior. Las combinaciones lineales se llevan a cabo mediante un bucle unitario y una atenuación. Con cada retardo se lleva a cabo una función monotonía decreciente de un tiempo de reverberación que ha de simularse y que es proporcional al retardo. Se realiza una corrección espectral antes de la suma ponderada. Este procedimiento requiere además una gran capacidad de procesamiento.

35

40 El documento EP 0 762 803, titulado "*Headphone device*", desvela un dispositivo de auriculares que detecta un ángulo de giro de la cabeza para realizar una localización de imagen de sonido.

40

El documento US 5 802 180, titulado "*Method and Apparatus for efficient presentation of high-quality three-dimensional audio including ambient effects*", desvela una espacialización de campos de sonido como respuesta a la ubicación de la fuente de sonido y a la posición del oyente.

45 El documento WO 95/31881, titulado "*Three-dimensional virtual audio display employing reduced complexity imaging filters*", se refiere a una visualización virtual de audio utilizando funciones alisadas de transferencia relativas a la cabeza.

50 El documento US 5 438 623, titulado "*multi-channel spatialization system for audio signals*", desvela la utilización de funciones sintéticas de transferencia relativas a la cabeza para la espacialización de campos sonoros.

El documento US 5 371 799, titulado "*stereo headphone sound source localization system*", se refiere a la espacialización de campos sonoros procesando por separado sonido directo, reflexiones anteriores y reverberación.

55 Por lo tanto, existe una necesidad general de reducir los requisitos de procesamiento de un sistema de convolución total manteniendo sustancialmente al mismo tiempo la calidad global del proceso de convolución.

Resumen de la invención

60 Según un primer aspecto de la presente invención, se proporciona un procedimiento para procesar una serie de señales de audio de entrada que representan una serie de fuentes de audio virtuales situadas en posiciones predeterminadas alrededor de un oyente para generar un conjunto reducido de señales de salida de audio para su reproducción en dispositivos de altavoz situados alrededor de un oyente, comprendiendo el procedimiento las etapas de:

(a) para cada una de dichas señales de audio de entrada y para cada una de dichas señales de salida de audio:

5 (i) convolucionar dichas señales de audio de entrada con una parte de cabecera inicial de una respuesta de impulso correspondiente que correlaciona sustancialmente el sonido inicial y reflexiones anteriores para una respuesta de impulso de una fuente de audio virtual correspondiente con un dispositivo de altavoz correspondiente para formar una serie de respuestas iniciales;

10 (b) para cada una de dichas señales de audio de entrada y para cada una de dichas señales de salida de audio:

(i) formar una mezcla combinada a partir de dichas señales de entrada de audio; y

(ii) determinar una única cola de convolución;

15 (iii) convolucionar dicha mezcla combinada con dicha única cola de convolución para formar una respuesta de cola combinada;

(c) para cada una de dichas señales de salida de audio:

20 (i) combinar una serie correspondiente de respuestas iniciales y una respuesta de cola combinada correspondiente para formar dicha señal de salida de audio, caracterizado porque dicha única cola de convolución se forma combinando las colas de dichas respuestas de impulso correspondientes.

25 Las señales de audio de entrada se convierten preferentemente al dominio de frecuencia y una convolución se lleva a cabo en el dominio de frecuencia. Las respuestas de impulsos pueden simplificarse en el dominio de frecuencia poniendo a cero los coeficientes de mayor frecuencia y eliminando etapas de multiplicación donde se utilizan preferentemente los coeficientes de mayor frecuencia puestos a cero.

30 Las convoluciones se llevan a cabo preferentemente utilizando un proceso de convolución de baja latencia. El proceso de convolución de baja latencia puede incluir preferentemente las etapas de: transformar primeras partes de las señales de audio de entrada con un tamaño de bloque predeterminado en bloques de coeficientes de entrada de dominio de frecuencia correspondientes; transformar segundas partes de las señales de respuesta de impulso con un tamaño de bloque predeterminado en bloques de coeficientes de impulso de dominio de frecuencia correspondientes; combinar cada uno de los bloques de coeficientes de entrada de dominio de frecuencia con bloques predeterminados de los bloques de coeficientes de impulso de dominio de frecuencia correspondientes de una manera predeterminada para generar bloques de salida combinados; agrupar bloques predeterminados de los bloques de salida combinados para generar respuestas de salida de dominio de frecuencia para cada una de las señales de salida de audio; transformar las respuestas de salida de dominio de frecuencia en señales de salida de audio de dominio de tiempo correspondientes; proporcionar las señales de salida de audio de dominio de tiempo.

40 La presente invención puede implementarse de varias formas diferentes, por ejemplo, utilizando una unidad de procesamiento de protección contra saltos ubicada en una unidad de reproducción de CD-ROM, utilizando un circuito integrado dedicado que comprende una forma modificada de un convertidor de digital a analógico, utilizando un procesador de señales digitales dedicado o programable, o utilizando un procesador DSP interconectado entre un convertidor de analógico a digital y un convertidor de digital a analógico. Como alternativa, la invención puede implementarse utilizando un dispositivo externo extraíble aparte conectado entre un generador de señales de salida sonoras y un par de auriculares, proporcionándose las señales de salida sonoras en una forma digital para su procesamiento por el dispositivo externo.

50 Modificaciones adicionales pueden incluir la utilización de un control variable para alterar las funciones de respuesta de impulso de una manera predeterminada.

55 Según un segundo aspecto de la presente invención, se proporciona un aparato que incluye un procesador y un dispositivo de almacenamiento configurado con instrucciones que cuando se ejecutan en el procesador hacen que se implemente un procedimiento según el primer aspecto de la invención.

Breve descripción de los dibujos

60 A continuación se describirán formas preferidas de la invención, solamente a modo de ejemplo, con referencia a los dibujos adjuntos, en los que:

La Fig. 1 ilustra esquemáticamente el procesamiento de convolución global para correlacionar una serie de señales con dos canales de salida de auricular.

La Fig. 2 ilustra el proceso FFT tradicional de solapar y guardar.

La Fig. 3 ilustra el proceso de baja latencia utilizado en la realización preferida.

5 La Fig. 4 es una ilustración esquemática del proceso genérico de convolución de dominio de frecuencia.

La Fig. 5 ilustra una primera simplificación del proceso de la Fig. 4.

10 La Fig. 6 ilustra el procesamiento idealizado de una serie de señales de entrada para el oído izquierdo de un auricular.

La Fig. 7 ilustra una primera simplificación de los requisitos de procesamiento de la Fig. 6.

15 La Fig. 8 ilustra en mayor detalle una implementación en el dominio de frecuencia de la disposición de la Fig. 7 utilizando una convolución de baja latencia.

La Fig. 9 ilustra el proceso de síntesis estándar para obtener coeficientes de dominio de frecuencia.

20 La Fig. 10 ilustra una forma modificada de generación de coeficientes de dominio de frecuencia.

La Fig. 11 ilustra la extensión de la realización preferida a datos de audio de mayor frecuencia.

25 La Fig. 12 ilustra una implementación mediante la cual el sistema de circuitos de procesamiento de audio se utiliza en lugar de la función de protección contra saltos de un reproductor de CD existente.

La Fig. 13 ilustra una implementación mediante la cual el sistema de circuitos de procesamiento de audio se utiliza en la misma cápsula de CI como un convertidor de digital a analógico.

30 La Fig. 14 ilustra una implementación mediante la cual el sistema de circuitos de procesamiento de audio se utiliza en la cadena de señales antes del convertidor de digital a analógico.

La Fig. 15 ilustra una implementación mediante la cual el sistema de circuitos de procesamiento de audio se utiliza en una configuración junto con convertidores de analógico a digital y de digital a analógico.

35 La Fig. 16 ilustra una extensión con respecto al circuito de la Fig. 17 para incluir una entrada digital opcional.

La Fig. 17 ilustra una pluralidad de posibles realizaciones físicas de la presente invención.

Descripción de realizaciones preferidas y de otras realizaciones

40 En la realización preferida se desea aproximar la larga convolución total de una serie de señales de entrada con funciones de respuesta de impulso para cada oído de manera que las salidas puedan sumarse para el oído izquierdo y el oído derecho para su reproducción en auriculares.

45 Haciendo referencia a la Fig. 1, se ilustra el proceso de convolución total para un conjunto de señales de sonido envolvente Dolby de 6 entradas que comprende un canal delantero izquierdo, un canal delantero central, un canal delantero derecho, un canal envolvente izquierdo, un canal envolvente derecho y un canal de efectos de baja frecuencia, cada uno indicado de manera genérica 2. Para cada canal se aplica una función de respuesta de impulso izquierda y derecha. Por lo tanto, para el canal delantero izquierdo 3, una función de respuesta de impulso delantero izquierdo correspondiente 4 se convoluciona 6 con la señal izquierda. La función de respuesta de impulso delantero izquierdo 4 es una respuesta de impulso que se recibiría mediante el oído izquierdo para una salida idealizada de sobreimpulsos de un altavoz de canal delantero izquierdo colocado en una posición idealizada. La salida 7 se suma 10 a la señal de canal izquierdo para los auriculares.

55 Asimismo, la respuesta de impulso correspondiente 5 para el oído derecho para un altavoz de canal izquierdo se convoluciona 8 con la señal delantera izquierda para producir una salida 9 que se suma 11 al canal derecho. Un proceso similar se produce para cada una de las demás señales.

60 Por lo tanto, la disposición de la Fig. 1 requerirá aproximadamente 12 etapas de convolución para las seis señales de entrada. Es probable que este elevado número de convoluciones sea bastante pesado para un chip de DSP, especialmente cuando se utiliza la convolución larga deseada.

Haciendo referencia ahora a la Fig. 2, se ilustra el proceso de convolución estándar de "solapar y guardar", el cual se describe completamente en documentos de normalización tales como "Digital Signal Processing" de John Proakis y

Dimitis Manolakis, *McMillan Publishing Company*, 1992.

En el procedimiento tradicional de solapar y guardar ilustrado como 20 en la Fig. 2, la señal de entrada 21 se digitaliza y se divide en N bloques de muestra 22, siendo N normalmente una potencia de 2. Asimismo, una respuesta de impulso 23 de longitud N se determina normalmente tomando mediciones deseadas del entorno, rellenándose con ceros hasta alcanzar una longitud $2N$. Una primera correlación 24 se aplica a los bloques de longitud $2N$ de la respuesta de impulso 23 para formar N números complejos que presentan coeficientes reales e imaginarios. Después se aplica la FFT para generar N coeficientes de frecuencia. La etapa 24 puede llevarse a cabo una vez antes de que comience el procesamiento y los coeficientes de dominio de frecuencia correspondientes 25 se almacenan para utilizarse posteriormente.

Después, se toman bloques de longitud $2N$ del audio de entrada y se aplica de nuevo una transformada rápida de Fourier para determinar datos de dominio de frecuencia correspondientes 28 correspondientes a los $2N$ valores de entrada reales. Después, los dos conjuntos de datos se multiplican elemento a elemento 30 para generar datos de dominio de frecuencia 31. Después se aplica una transformada inversa de Fourier para generar $2N$ valores reales, descartándose el primer bloque 34 de longitud N y convirtiéndose el segundo bloque 35 de longitud N en los valores de salida 36 para el audio de salida. El proceso ilustrado en la Fig. 2 se conoce ampliamente como un proceso de convolución estándar en el dominio de frecuencia. Sin embargo, desafortunadamente, debido al requisito de que los datos de entrada deben recopilarse en bloques y debido a que el proceso FFT tiene una cierta duración, que depende del valor de N (siendo el tiempo de procesamiento $O(N \log N)$), se produce una determinada latencia o retardo desde el momento en que $2N$ valores de entrada iniciales se introducen en la primera FFT 27 y hasta el tiempo de la salida posterior desde la FFT inversa 32. Normalmente, este retardo o latencia es muy poco deseable, especialmente cuando deben cumplirse requisitos de tiempo real.

En la solicitud PCT mencionada anteriormente número PCT/AU93/00330, se desveló un procedimiento para proporcionar un proceso de convolución de latencia extremadamente baja adecuado para su utilización en tiempo real. Aunque se insta al lector que acuda a la memoria descriptiva PCT mencionada anteriormente, a continuación se expondrá una breve exposición del proceso de baja latencia con referencia a la Fig. 3, la cual ilustra como 40 las etapas básicas del proceso de baja latencia en el que, en primer lugar, una entrada de audio se convierte al dominio de frecuencia mediante un proceso de solapamiento de dominio de frecuencia FFT 41. Los datos de dominio de frecuencia se almacenan 42 y después se proporcionan a bloques de almacenamiento posteriores, por ejemplo 43, 44, después de cada "ciclo" de convolución. En primer lugar, los datos de dominio de frecuencia, por ejemplo 42, se multiplican elemento a elemento 50 por coeficientes de dominio de frecuencia correspondientes 51, correspondiendo los coeficientes de dominio 51 a una primera parte de la función de respuesta de impulso.

Simultáneamente, los datos de dominio de frecuencia previamente retardados 43 se multiplican 54 con los coeficientes de dominio de frecuencia 53 correspondientes a una parte posterior de la función de respuesta de impulso. Esta etapa también se repite para el resto de la función de respuesta de impulso. Las salidas se suman elemento a elemento 56 para generar datos de dominio de frecuencia globales que se transforman mediante la transformada inversa y rápida de Fourier, descartándose 57 la mitad de los datos para generar una salida de audio 58. La disposición de la Fig. 3 permite llevar a cabo una convolución extremadamente larga con una baja latencia.

El proceso general de la Fig. 3 puede utilizarse para llevar a cabo el proceso global equivalente de la Fig. 1. Esto se ilustra en la Fig. 4, en la que, en primer lugar, cada uno de los seis canales de entrada, denotados por ejemplo como 60, se convierte al dominio de frecuencia utilizando un proceso de solapamiento FFT 61. Después, cada canal se combina 62 en el dominio de frecuencia con coeficientes de dominio de frecuencia correspondientes a las funciones de respuesta de impulso. El proceso de dominio de frecuencia 62 también puede incluir la agrupación de los componentes de frecuencia para formar las salidas izquierda y derecha 64, 65. Finalmente se aplica un proceso inverso de descarte de dominio de frecuencia 66, 67 para generar las salidas de canal izquierda y derecha.

La cantidad requerida de cálculo puede reducirse sustancialmente simplificando el número de convoluciones requeridas.

Haciendo referencia ahora a la Fig. 5, se ilustra una forma de simplificación llevada a cabo en la realización preferida, donde el canal central 70 se multiplica por un factor de ganancia 71 y se suma 72, 73 a los canales izquierdo y derecho, respectivamente. Asimismo, partes del canal de efectos de baja frecuencia 71 se suman a cada uno de los otros canales.

Después, las señales se someten a un procesador de solapamiento de transformada de Fourier 75, 78, reduciéndose de 6 a 4 el número de canales sometidos al proceso de transformada de Fourier de gran carga computacional. Después se aplica el proceso de dominio de Fourier 84 para generar salidas 79, 80 a partir de las cuales se aplica un proceso de descarte y de transformada inversa de Fourier 82, 83 para salir de los canales izquierdo y derecho.

Haciendo referencia ahora a la Fig. 6, se ilustra esquemáticamente el resultado final global idealizado del proceso de la Fig. 5 para el oído izquierdo, donde las cuatro señales de entrada 90 se someten cada una a un filtro correspondiente 91 a 94 de respuesta de impulso finita de longitud total antes de sumarse 95 para formar una señal de salida correspondiente 96. Desafortunadamente, para conseguir un alto nivel de realismo, normalmente es necesario utilizar funciones de respuesta de impulso bastante largas. Por ejemplo, funciones de respuesta de impulso con una longitud de pulsos de 7.000 pulsos aproximadamente son habituales con señales de audio estándar de 48 kHz. De nuevo, con la disposición de la Fig. 6, excesivos requisitos computacionales pueden dar como resultado filtros de longitud extendida.

Un análisis de los detalles de los coeficientes de respuesta de impulso, y algunos experimentos, ha mostrado que todos los indicios necesarios para una localización precisa de las fuentes de sonido están incluidos en el tiempo de las reflexiones directas y en algunas primeras órdenes de reflexión, y que el resto de la respuesta de impulso solo se requiere para resaltar el "tamaño" y la "vivacidad" del entorno acústico. Esta observación puede utilizarse para separar la parte direccional o de "cabecera" de cada una de las respuestas (digamos, los primeros 1024 pulsos) con respecto a la parte de reverberación o de "cola". Todas las partes de "cola" pueden agruparse y el filtro resultante puede excitarse con la suma de las señales de entrada individuales. Esta implementación simplificada se muestra esquemáticamente como 100 en la Fig. 7. Los filtros de cabecera 101 a 104 puede ser filtros cortos de 1024 pulsos y las señales se suman 105 y se introducen en el filtro de cola extendido, el cual puede comprender 6.000 pulsos aproximadamente, sumándose 109 los resultados para proporcionarse a continuación. Este proceso se repite para el oído derecho. La utilización de la cola combinada que presenta una única función de respuesta de cola reduce los requisitos de cálculo de dos maneras. En primer lugar, existe una reducción obvia en el número de términos de las sumas de convolución que deben calcularse en tiempo real. Esta reducción viene dada por un factor del número de canales de entrada. En segundo lugar, la latencia de cálculo del cálculo de filtro de cola solo necesita ser lo bastante corta como para alinear el primer pulso del filtro de cola con el último pulso de cada uno de los filtros de cabecera. Cuando se utilizan técnicas de implementación de filtrado de bloques tales como solapar/sumar, solapar/guardar o el algoritmo de convolución de baja latencia de la solicitud PCT mencionada anteriormente, esto implica que, opcionalmente, pueden utilizarse bloques más grandes para implementar la cola que los utilizados para las cabeceras, a una velocidad de trama inferior.

Haciendo referencia ahora a la Fig. 8, se ilustra en detalle un diagrama de flujo global del procesamiento en el dominio de frecuencia cuando se implementa el sistema de cola combinada de la Fig. 7. La disposición 110 de la Fig. 8 está destinada a funcionar como o un proceso en el dominio de frecuencia, tal como el 84 de la Fig. 5. El sistema global incluye sumas 111 y 112 que se proporcionan a los canales izquierdo y derecho, respectivamente. Las cuatro entradas son la entrada delantera izquierda, la entrada delantera derecha, la entrada trasera izquierda y la entrada trasera derecha, las cuales se tratan de manera simétrica almacenándose la primera entrada en un bloque de almacenamiento de retardo 113 y además multiplicándose 114 con coeficientes de dominio de frecuencia obtenidos a partir de una primera parte de la respuesta de impulso 115, proporcionándose la salida a un sumador 11. El canal derecho se trata de manera simétrica para generar una salida de canal derecho y no se analizará en mayor detalle en este documento.

Después de un "ciclo", el bloque de retardo 113 se envía al bloque de retardo 120. Para los expertos en la materia de la programación DSP les resultará evidente que esto puede comprender simplemente una nueva correlación de punteros de bloques de datos. Durante el siguiente ciclo, los coeficientes 121 se multiplican 122 con los datos del bloque 120, enviándose la salida al sumador de canal izquierdo 111. Los dos conjuntos de coeficientes 115 y 121 corresponden a la parte de cabecera de la función de respuesta de impulso. Cada canal tendrá funciones de cabecera individualizadas para los canales de salida izquierdo y derecho.

Las salidas de los bloques de retardo 120, 125, 126 y 127 se envían a un sumador 130 y la suma se almacena en un bloque de retardo 131. El bloque de retardo 131 y bloques de retardo posteriores, por ejemplo 132, 133, implementan el filtro de cola combinada, multiplicándose 136 un primer segmento almacenado en el bloque de retardo 131 con coeficientes 137 para enviarse a la suma de canal izquierdo III. En el siguiente ciclo, el bloque de retardo 131 se envía al bloque 132 y un proceso similar se lleva a cabo, realizándose para cada bloque de retardo restante, por ejemplo 133. De nuevo, el canal derecho se trata de manera simétrica.

A partir del análisis anterior resultará evidente que hay una pluralidad de funciones de respuesta de impulso o partes de las mismas utilizadas en la construcción de la realización preferida. A continuación se describirá una optimización computacional adicional del proceso de síntesis de los bloques de coeficientes de dominio de frecuencia haciendo referencia inicialmente a la Fig. 9. Con el fin de determinar los coeficientes de dominio de frecuencia requeridos, una respuesta de impulso 140 se divide en una pluralidad de segmentos 141 de longitud N. Cada segmento se rellena con N valores de datos adicionales 142 de valor cero antes de aplicarse 143 una correlación FFT con respecto a N datos complejos para convertir los valores en N coeficientes de dominio de frecuencia 144. Este proceso puede repetirse para obtener coeficientes de dominio de frecuencia posteriores 145, 146, 147.

La utilización del proceso de segmentación de la Fig. 9 puede dar lugar frecuentemente a componentes de dominio

de frecuencia artificialmente alta. Esto es una consecuencia directa del proceso de segmentación y de su interacción con la transformada rápida de Fourier, que debe dar como resultado componentes de frecuencia que aproximen la discontinuidad en los valores de datos finales (siendo la FFT un módulo periódico del tamaño de los datos). Normalmente, la FFT resultante tiene componentes de frecuencia significativamente alta que están presentes sustancialmente como resultado de esta discontinuidad. En la realización preferida, se lleva a cabo preferentemente un proceso para reducir los componentes de alta frecuencia hasta un punto en que una cantidad significativa del cálculo puede descartarse debido a un conjunto de componentes de dominio de frecuencia nula. A continuación se describirá este proceso de formación de bloques de coeficientes de dominio de frecuencia limitados por banda con referencia a la Fig. 10.

La respuesta de impulso inicial 150 se divide de nuevo en segmentos 151 de longitud N . Después, cada segmento se rellena hasta formar segmentos 152 de longitud $2N$. Después, los datos 152 se multiplican 154 con una función de "división en ventanas" 153 que incluye partes de extremo graduadas 156, 157. Las dos partes de extremo están diseñadas para correlacionar los extremos de la secuencia de datos 151 con valores de magnitud cero conservando al mismo tiempo la información entre los mismos. La salida resultante 159 contiene valores de cero en los puntos 160, 161. Después, la salida 159 se somete a un proceso de FFT real para generar coeficientes de dominio de frecuencia 165 que presentan una pluralidad de coeficientes mayores 167 en el dominio de frecuencia inferior de la transformada de Fourier además de una pluralidad de componentes despreciables 166 que, por tanto, pueden descartarse. De este modo, un conjunto parcial final de componentes de dominio de frecuencia 169 se utiliza como los componentes de dominio de frecuencia que representan la parte correspondiente de los datos de respuesta de impulso.

El descarte de los componentes 166 significa que, durante el proceso de convolución, solamente necesita llevarse a cabo una forma restringida del procesamiento de convolución y que no es necesario multiplicar el conjunto total de N coeficientes complejos, ya que un número sustancial de los mismos tienen valor 0. De nuevo, esto da lugar a mayores ganancias de eficacia ya que el requisito computacional del proceso de convolución se reduce. Además, reducciones significativas en los requisitos de memoria del algoritmo son posibles aprovechando el hecho de que tanto el almacenamiento de datos como de coeficientes pueden reducirse como resultado de este descarte de coeficientes.

En una realización preferida, N es igual a 512, los filtros de cabecera tienen una longitud de 1024 pulsos y los filtros de cola tienen una longitud de 6144 pulsos. Por tanto, cada filtro de cabecera está compuesto por dos bloques de coeficientes (tal y como se ilustra en la Fig. 8) y cada filtro de cola está compuesto por doce bloques de coeficientes. En esta realización preferida, todos los filtros de cabecera y los 4 primeros bloques de cada filtro de cola se implementan utilizando conjuntos completos de transformadas de coeficientes de Fourier, los 4 bloques siguientes de cada filtro de cola se implementan utilizando bloques de coeficientes en los que solamente menos de la mitad de los componentes de frecuencia están presentes, y los 4 últimos bloques de cada filtro de cola se implementan utilizando bloques de coeficientes en los que solamente menos de una cuarta parte de los componentes de frecuencia están presentes.

La realización preferida puede extenderse a una situación en la que se utilizan entradas de audio de mayor frecuencia, pero donde es deseable mantener un requisito computacional de baja frecuencia. Por ejemplo, actualmente es habitual en la industria adoptar una frecuencia de muestreo de 96 kHz para muestras digitales y, por tanto, es deseable proporcionar una convolución de respuestas de impulso muestreadas también a esta frecuencia. Haciendo referencia a la Fig. 11, se ilustra una forma de extensión que utiliza una frecuencia de muestreo de respuestas de impulsos inferior. En esta disposición, la señal de entrada a una frecuencia 170 de 96 kHz se envía a un almacenamiento intermedio de retardo 171. La señal de entrada también se filtra paso bajo 172 y después se diezma 173 en un factor de 2 hasta una frecuencia de 48 kHz, sometiéndose después a un filtrado FIR según el sistema mostrado anteriormente antes de la que frecuencia de muestreo se duplique 175, pasando posteriormente por un filtro paso bajo 176 y sumándose 177 posteriormente a la señal de entrada retardada originalmente enviada desde el almacenamiento intermedio de retardo 171, la cual se ha multiplicado previamente 178 por un factor de ganancia A . La salida del sumador 177 forma la salida digital convolucionada.

Normalmente, si la respuesta de impulso deseada de 96 kHz se denota como $h_{96}(t)$, entonces los coeficientes FIR de 48 kHz, denotados como $h_{48}(t)$ pueden obtenerse a partir de PasoBajo[$h_{96}(t)$], donde, en este caso, la notación indica que la respuesta de impulso original $h_{96}(t)$ se ha filtrado paso bajo. Sin embargo, en el procedimiento mejorado de la Fig. 11, si la respuesta deseada se denota como $h_{96}(t)$ y la respuesta de impulso retardada se denota como $A \cdot \delta(t - \tau)$, entonces los coeficientes FIR de 48 kHz denotados como $h_{48}(t)$ pueden obtenerse a partir de PasoBajo[$h_{96}(t) - A \cdot \delta(t - \tau)$]. La elección del factor de retardo τ y del factor de ganancia A se realiza de manera que la señal resultante del elemento de ganancia 178 tenga el tiempo de llegada correcto y una amplitud para generar los componentes de alta frecuencia que se requieren en la parte de llegada directa de la respuesta de impulso acústica de 96 kHz. Además, en lugar de utilizar una disposición de retardo y ganancia, es posible utilizar una FIR dispersa para crear múltiples ecos conformados según una banda ancha y una frecuencia.

Por lo tanto, puede observarse que la realización preferida proporciona un sistema de convolución de cálculo reducido manteniendo un número sustancial de las características del sistema de convolución total.

5 La realización preferida toma una señal de entrada digital multicanal o señal de entrada de sonido envolvente tal como Dolby Prologic, Dolby Digital (AC-3) y DTS, y utiliza uno o más conjuntos de auriculares para su salida. La señal de entrada se procesa de manera binaural utilizando la técnica mencionada anteriormente para mejorar las experiencias de escucha a través de los auriculares en una gran variedad de material fuente, haciendo que suene de este modo "fuera de la cabeza", o para proporcionar una mejor escucha de sonido envolvente.

10 Dada una técnica de procesamiento de este tipo para producir un efecto fuera de la cabeza, un sistema que lleva a cabo este procesamiento puede proporcionarse utilizando una pluralidad de diferentes realizaciones. Por ejemplo, muchas posibles realizaciones físicas diferentes son posibles, y el resultado final puede implementarse utilizando técnicas analógicas o digitales de procesamiento de señales o una combinación de ambas.

15 En una implementación puramente digital, se supone que los datos de entrada se obtienen en una forma digital muestreada en el tiempo. Si la realización se implementa como parte de un dispositivo de audio digital tal como un disco compacto (CD, *compact disc*), un MiniDisc, un disco de vídeo digital (DVD, *digital video disc*) o una cinta de audio digital (DAT, *digital audio tape*), los datos de entrada ya estarán disponibles en esta forma. Si la unidad se implementa como un dispositivo físico por sí mismo, puede incluir un receptor digital (SPDIF o similar, ya sea óptico o eléctrico). Si la invención se implementa de manera que solamente esté disponible una señal de entrada analógica, esta señal analógica debe digitalizarse utilizando un convertidor de analógico a digital (ADC).

20 Después, esta señal de entrada digital se procesa mediante un procesador de señales digitales (DSP) de alguna forma. Ejemplos de DSP que pueden utilizarse son:

- 25
1. Un circuito integrado semipersonalizado o totalmente personalizado diseñado como un DSP dedicado a la tarea.
 2. Un chip DSP programable, por ejemplo el Motorola DSP56002.
 3. Uno o más dispositivos lógicos programables.

30 En caso de que la realización se utilice con un conjunto específico de auriculares, el filtrado de las funciones de respuesta de impulso puede aplicarse para compensar cualquier característica de respuesta de frecuencia no deseada de dichos auriculares.

35 Después del procesamiento, las señales de salida digitales estéreo se convierten en señales analógicas utilizando convertidores de digital a analógico (DAC), amplificándose si fuera necesario, y se encaminan hacia las salidas de auricular estéreo, quizá a través de otro sistema de circuitos. Esta fase final puede tener lugar o bien en el dispositivo de audio en caso de que una realización esté incorporada, o como parte de un dispositivo distinto en caso de que una realización se implemente de esta manera.

40 Los ADC y/o los DAC también pueden estar incorporados en el mismo circuito integrado al igual que el procesador. Una realización también puede implementarse de manera que parte o todo el procesamiento se realice en el dominio analógico. Las realizaciones presentan preferentemente algún procedimiento para activar y desactivar el efecto "binauralizador" y pueden incorporar un procedimiento para conmutar entre ajustes de ecualizador para diferentes conjuntos de auriculares o para controlar otras variaciones en el procesamiento realizado incluyendo, quizás, el volumen de salida.

45 En una primera realización ilustrada en la Fig. 12, las etapas de procesamiento están incorporadas en un reproductor portátil de CD o DVD en sustitución de un CI de protección contra saltos. Muchos reproductores de CD actualmente disponibles incorporan una característica de "protección contra saltos" que almacena de manera intermedia los datos leídos desde el CD en una memoria de acceso aleatorio (RAM). Si se detecta un "salto", es decir, el flujo de audio se interrumpe porque el mecanismo de la unidad recibe un golpe haciendo que salga de la pista, la unidad puede volver a leer los datos del CD mientras reproduce los datos desde la RAM. Esta protección contra saltos está implementada frecuentemente como un DSP dedicado, ya sea con una RAM en el chip o fuera del chip.

50 Esta realización se implementa de manera que pueda utilizarse en sustitución del procesador de protección contra saltos con unas mínimas modificaciones en los diseños existentes. Esta implementación puede implementarse con mayor probabilidad como un circuito integrado totalmente personalizado, satisfaciendo la función de los procesadores de protección contra saltos existentes y la implementación del procesamiento "fuera de la cabeza".

55 Una parte de la RAM ya incluida para la protección contra saltos puede utilizarse para ejecutar el algoritmo "fuera de la cabeza" para el procesamiento de tipo HRTF. Muchos de los bloques de construcción de un procesador de protección contra saltos también son útiles para el procesamiento descrito para esta invención. Un ejemplo de una disposición de este tipo se ilustra en la Fig. 12.

60

En esta realización de ejemplo, el DSP personalizado 200 se proporciona como una sustitución del DSP de protección contra saltos en un reproductor de CD o DVD 202. El DSP personalizado 200 toma los datos de entrada del disco y proporciona señales estéreo a un convertidor de digital a analógico 201 que proporciona salidas analógicas que se amplifican 204, 205 para proporcionar las salidas de auricular izquierda y derecha. El DSP personalizado puede incluir una RAM incorporada 206 o, como alternativa, una RAM externa 207 según los requisitos. Un conmutador binauralizador 28 puede proporcionarse para activar y desactivar el efecto binauralizador.

En una segunda realización ilustrada en la Fig. 13, el procesamiento está incorporado en un dispositivo de audio digital 210 (tal como un reproductor de CD, de MiniDisc, de DVD o de DAT) en sustitución del DAC. En esta implementación, el procesamiento de señales se lleva a cabo mediante un circuito integrado dedicado 211 que incorpora un DAC. Éste puede incorporarse fácilmente en un dispositivo de audio digital con pequeñas modificaciones en los diseños existentes ya que el circuito integrado tiene patillas prácticamente compatibles con los DAC existentes.

El CI personalizado 211 incluye un núcleo DSP incorporado 212 y la utilidad habitual de conversión de digital a analógico 213. El CI personalizado toma la salida habitual de datos digitales y lleva a cabo el procesamiento a través del DSP 212 y una conversión de digital a analógico 213 para proporcionar salidas estéreo. De nuevo, un conmutador binauralizador 214 puede proporcionarse para controlar los efectos de binauralización si fuera necesario.

En una tercera realización, ilustrada en la Fig. 14, el procesamiento está incorporado en un dispositivo de audio digital 220 (tal como un reproductor de CD, de MiniDisc, de DVD o de DAT) como una fase adicional 221 en la cadena de señales digitales. En esta implementación, el procesamiento de señales se lleva a cabo por un DSP dedicado o programable 221 montado en un dispositivo de audio digital e insertado en la cadena de señales digitales estéreo antes del DAC 222.

En una cuarta realización, ilustrada en la Fig. 15, el procesamiento está incorporado en un dispositivo de audio (tal como un reproductor de casete personal o un receptor de radio estéreo 230) como una fase adicional en la cadena de señales analógicas. Esta realización utiliza un ADC 232 para hacer uso de las señales de entrada analógicas. Esta realización puede fabricarse con mayor probabilidad en un único circuito integrado 231 que incorpora un ADC 232, un DSP 233 y un DAC 234. También puede incorporar algún procesamiento analógico. Esto puede añadirse fácilmente en la cadena de señales analógicas de los diseños existentes de reproductores de casete y dispositivos similares.

En una quinta realización, ilustrada en la Fig. 16, el procesamiento se implementa como un dispositivo externo para utilizarse con entradas estéreo en forma digital. La realización puede ser una unidad física por sí misma o integrarse en un conjunto de auriculares como se ha descrito anteriormente. Puede alimentarse mediante baterías con la opción de recibir energía desde una fuente de suministro externa con un grupo de enchufes de CC. El dispositivo toma una entrada estéreo digital en una forma óptica o eléctrica, tal y como está disponible en algunos reproductores de CD y DVD o similares. Los formatos de entrada pueden ser SPDIF o similares, y la unidad puede soportar formatos de sonido envolvente tales como Dolby Digital AC-3 o DTS. También puede tener entradas analógicas tal y como se describirá posteriormente. El procesamiento se realiza mediante un núcleo DSP 241 contenido en el CI personalizado 242. Un DAC 243 está dispuesto a continuación. Si esta DAC no puede activar directamente los auriculares, amplificadores adicionales 246, 247 se añaden después del DAC. Esta realización de la invención puede implementarse en un circuito integrado personalizado 242 que incorpora un DSP, un DAC y posiblemente un amplificador de auriculares.

Como alternativa, la realización puede implementarse como una unidad física por sí misma o integrarse en un conjunto de auriculares. Puede alimentarse mediante baterías con la opción de recibir energía desde una fuente de suministro externa con un grupo de enchufes de CC. El dispositivo toma una entrada estéreo analógica que se convierte en datos digitales a través de un ADC. Después, estos datos se procesan utilizando un DSP y vuelven a convertirse a analógico a través de un DAC. Parte de o todo el procesamiento puede realizarse, en cambio, en el dominio analógico. Esta implementación puede fabricarse en un circuito integrado personalizado que incorpora un ADC, un DSP, un DAC y posiblemente un amplificador de auriculares así como cualquier sistema de circuitos requerido de procesamiento analógico.

Las realizaciones pueden incorporar un control a distancia o "zoom" que permite al oyente modificar la distancia percibida o entorno de la fuente de sonido.

En una realización preferida, este control se implementa como un control deslizante. Cuando este control está al mínimo, el sonido parece provenir desde muy cerca de los oídos y, de hecho, puede ser un sonido estéreo plano no binauralizado. En el ajuste máximo del control, el sonido se percibe como si viniera de lejos. El control puede modificarse entre estos extremos para controlar la percepción "fuera de la cabeza" del sonido. Empezando con el control en la posición mínima y deslizándolo hacia su posición máxima, el usuario podrá adaptarse más rápidamente

a la experiencia binaural que con un simple conmutador binaural de activación/desactivación. La implementación de un control de este tipo puede comprender la utilización de diferentes conjuntos de respuestas de filtro para diferentes distancias.

5 Implementaciones de ejemplo que presentan mecanismos deslizantes similares se muestran en la Fig. 17. Como alternativa, pueden proporcionarse controles adicionales para la conmutación del entorno de audio.

10 Como una alternativa adicional, una realización puede implementarse como una solución de circuito integrado genérico adecuada para una amplia gama de aplicaciones incluyendo las mencionadas anteriormente. Este mismo circuito integrado puede incorporarse prácticamente en cualquier pieza de un equipo de audio con salida de auriculares. También será el bloque de construcción principal de cualquier unidad física producida específicamente como una implementación de la invención.

15 Un circuito integrado de este tipo incluirá algunos de o todos los componentes ADC, DSP, DAC, entrada de audio digital estéreo PS de memoria, entrada de audio digital S/PDIF, amplificador de auriculares, así como patillas de control para permitir que el dispositivo funcione en diferentes modos (por ejemplo de entrada analógica o digital).

20 Los expertos en la materia apreciarán que numerosas variaciones y/o modificaciones pueden realizarse en la presente invención mostrada en las realizaciones específicas descritas. Por lo tanto, las presentes realizaciones deben considerarse en todos los aspectos como ilustrativas y no como restrictivas.

REIVINDICACIONES

- 5 1. Un procedimiento para procesar una serie de señales de audio de entrada (60) que representan una serie de fuentes de audio virtuales situadas en posiciones predeterminadas alrededor de un oyente para generar un conjunto reducido de señales de salida de audio para su reproducción en dispositivos de altavoz situados alrededor de un oyente, comprendiendo el procedimiento las etapas:
- (a) para cada una de dichas señales de audio de entrada y para cada una de dichas señales de salida de audio:
- 10 (i) convolucionar dichas señales de audio de entrada (60) con una parte de cabecera inicial (101, 102, 103, 104) de una respuesta de impulso correspondiente que correlaciona sustancialmente el sonido inicial y reflexiones anteriores para una respuesta de impulso de una fuente de audio virtual correspondiente con un dispositivo de altavoz correspondiente para formar una serie de respuestas iniciales;
- 15 (b) para cada una de dichas señales de audio de entrada (60) y para cada una de dichas señales de salida de audio:
- (i) formar una mezcla combinada a partir de dichas señales de entrada de audio (60); y
(ii) determinar una única cola de convolución;
20 (iii) convolucionar dicha mezcla combinada con dicha única cola de convolución para formar una respuesta de cola combinada;
- (c) para cada una de dichas señales de salida de audio:
- 25 (i) combinar una serie correspondiente de respuestas iniciales y una respuesta de cola combinada correspondiente para formar dicha señal de salida de audio,
- caracterizado porque** dicha única cola de convolución se forma combinando las colas de dichas respuestas de impulso correspondientes.
- 30 2. Un procedimiento según la reivindicación 1, que comprende además la etapa de procesar previamente dichas respuestas de impulso:
- (a) construyendo un conjunto de funciones de respuesta de impulso correspondientes (140);
35 (b) dividiendo dichas funciones de respuesta de impulso (140) en una pluralidad de segmentos (141);
(c) para un número predeterminado de dichos segmentos (141), reduciendo los valores de respuesta de impulso en los extremos de dichos segmentos (141).
3. Un procedimiento según cualquier reivindicación anterior, en el que las señales de audio de entrada (60) se convierten al dominio de frecuencia y dicha convolución se lleva a cabo en el dominio de frecuencia.
- 40 4. Un procedimiento según la reivindicación 3, en el que las respuestas de impulsos se simplifican en el dominio de frecuencia poniendo a cero los coeficientes de mayor frecuencia y eliminando las etapas de multiplicación donde se utilizan los coeficientes de mayor frecuencia puestos a cero.
- 45 5. Un procedimiento según cualquier reivindicación anterior, en el que dichas convoluciones se llevan a cabo utilizando un proceso de convolución de baja latencia.
- 50 6. Un procedimiento según la reivindicación 5, en el que el proceso de convolución de baja latencia incluye las etapas de:
- transformar primeras partes de las señales de audio de entrada con un tamaño de bloque de solapamiento predeterminado en bloques de coeficientes de entrada de dominio de frecuencia correspondientes;
transformar segundas partes de las funciones de respuesta de impulso con un tamaño de bloque predeterminado en bloques de coeficientes de impulso de dominio de frecuencia correspondientes;
55 combinar cada uno de los bloques de coeficientes de entrada de dominio de frecuencia con bloques predeterminados de los bloques de coeficientes de impulso de dominio de frecuencia correspondientes de una manera predeterminada para producir bloques de salida combinados;
agrupar bloques predeterminados de los bloques de salida combinados para generar respuestas de salida de dominio de frecuencia para cada dispositivo de altavoz;
60 transformar las respuestas de salida de dominio de frecuencia en señales de salida de audio de dominio de tiempo correspondientes;
descartar parte de las señales de salida de audio de dominio de tiempo;
proporcionar la parte restante de las señales de salida de audio de dominio de tiempo a los dispositivos de altavoz.

7. Un procedimiento según cualquier reivindicación anterior, en el que el conjunto de señales de audio de entrada (60) incluye una señal de canal delantero izquierdo, una señal de canal delantero derecho, una señal de canal delantero central, una señal de canal trasero izquierdo y una señal de canal trasero derecho.
- 5 8. Un procedimiento según cualquier reivindicación anterior, en el que los dispositivos de altavoz son uno o más conjuntos de auriculares.
9. Un procedimiento según cualquier reivindicación anterior, en el que el procedimiento se lleva a cabo utilizando una unidad de procesamiento de protección contra saltos ubicada en una unidad de reproducción de CD-ROM.
- 10 10. Un procedimiento según una de las reivindicaciones 1 a 8, en el que el procedimiento se lleva a cabo utilizando un circuito integrado dedicado que comprende una forma modificada de un convertidor de digital a analógico.
- 15 11. Un procedimiento según una de las reivindicaciones 1 a 8, en el que el procedimiento se lleva a cabo utilizando un procesador de señales digitales dedicado o programable.
12. Un procedimiento según una de las reivindicaciones 1 a 8, en el que el procedimiento se lleva a cabo sobre entradas analógicas mediante un procesador DSP interconectado entre un convertidor de analógico a digital y un convertidor de digital a analógico.
- 20 13. Un procedimiento según una de las reivindicaciones 1 a 8, en el que los dispositivos de altavoz son un par de auriculares, comprendiendo además el procedimiento la formación de señales de salida sonoras en forma digital para su procesamiento por un dispositivo externo extraíble aparte conectado entre un generador de señales de salida sonoras y el par de auriculares.
- 25 14. Un procedimiento según cualquier reivindicación anterior, que comprende además utilizar un control variable para alterar las funciones de respuesta de impulso de una manera predeterminada.
- 30 15. Un aparato que incluye un procesador y un dispositivo de almacenamiento configurado con instrucciones que cuando se ejecutan en el procesador hacen que se implemente el procedimiento indicado en cualquiera de las reivindicaciones de procedimiento 1 a 14.

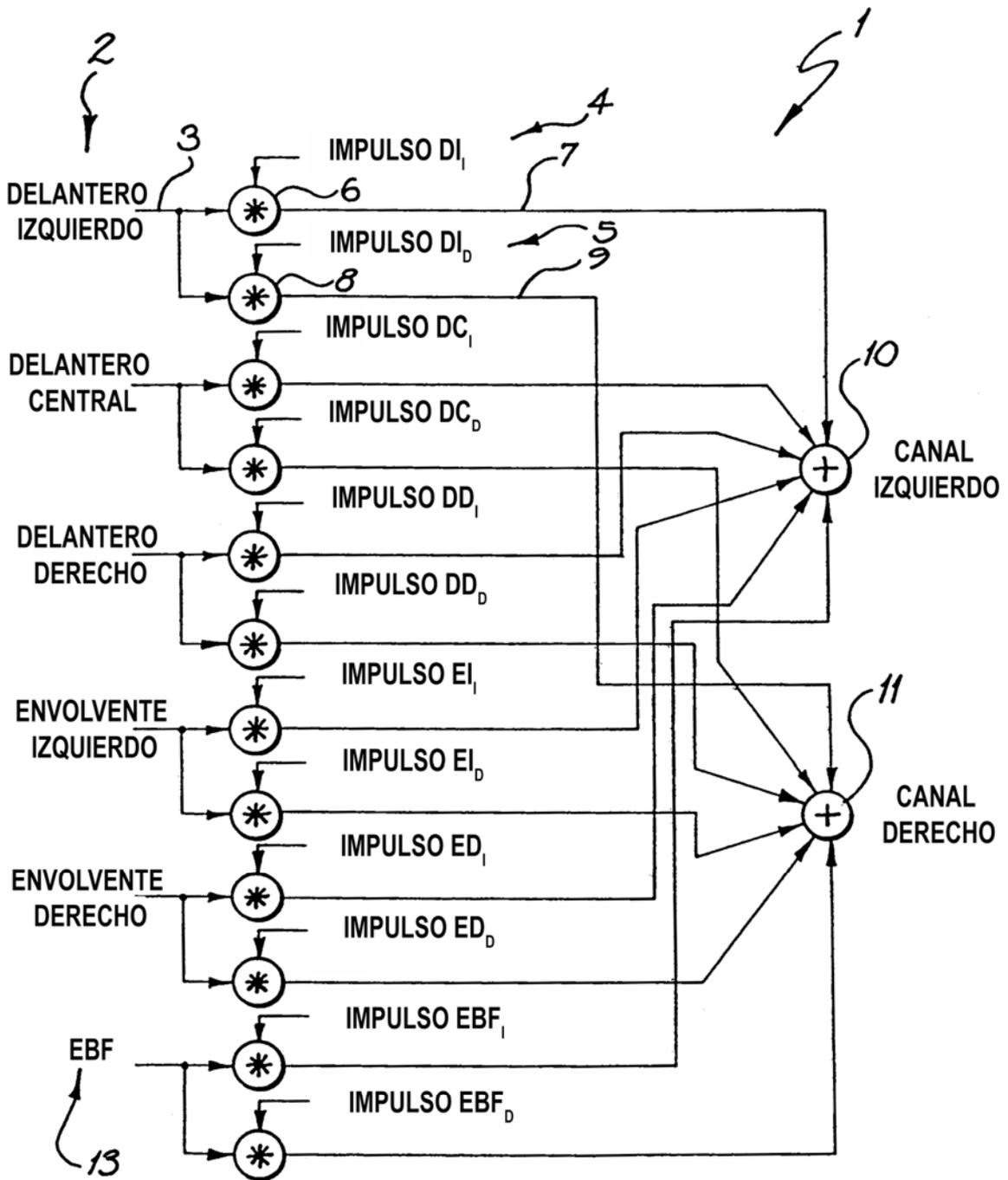


FIG. 1

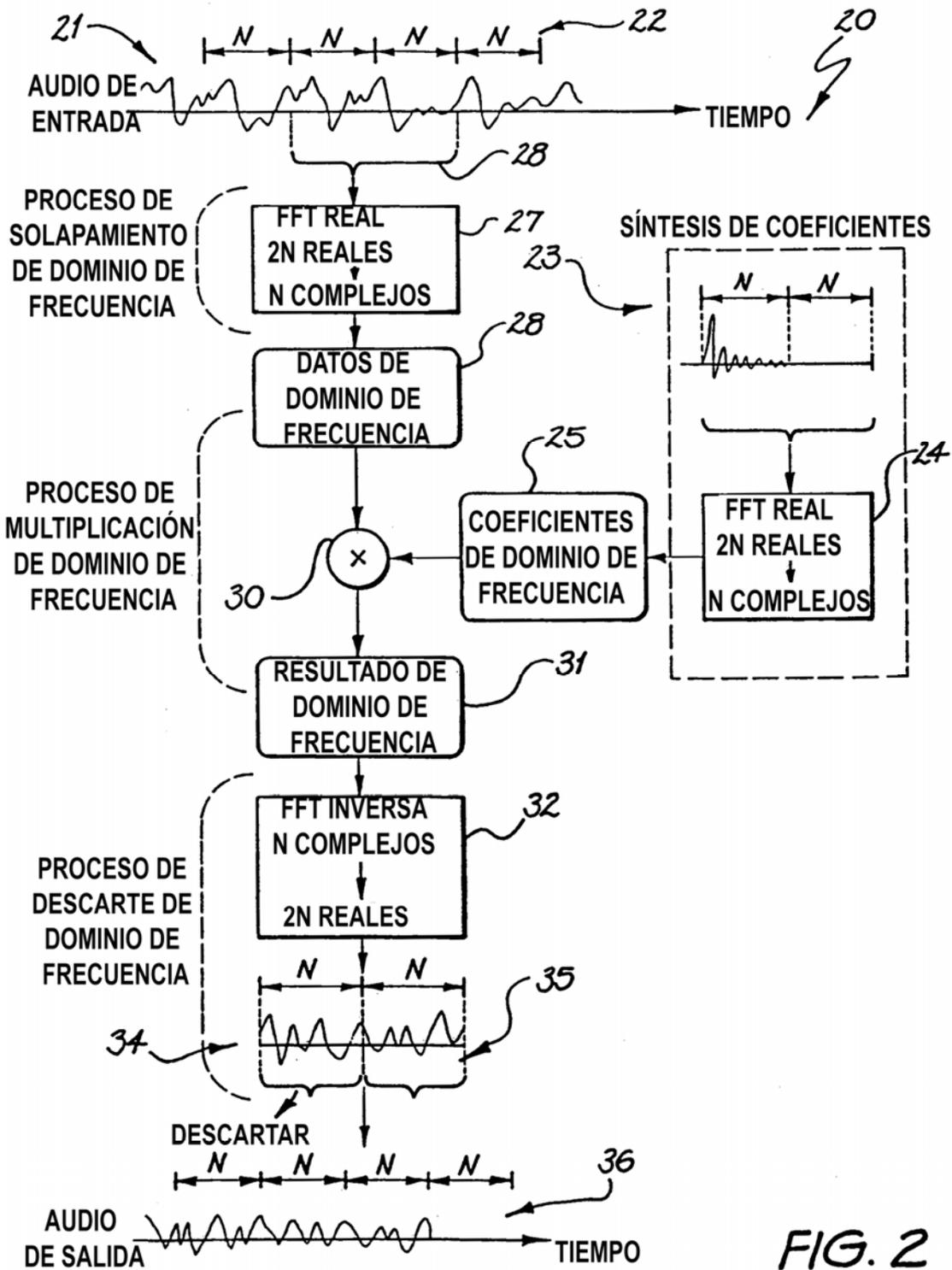


FIG. 2

PROCEDIMIENTO DE "SOLAPAR Y GUARDAR" TRADICIONAL

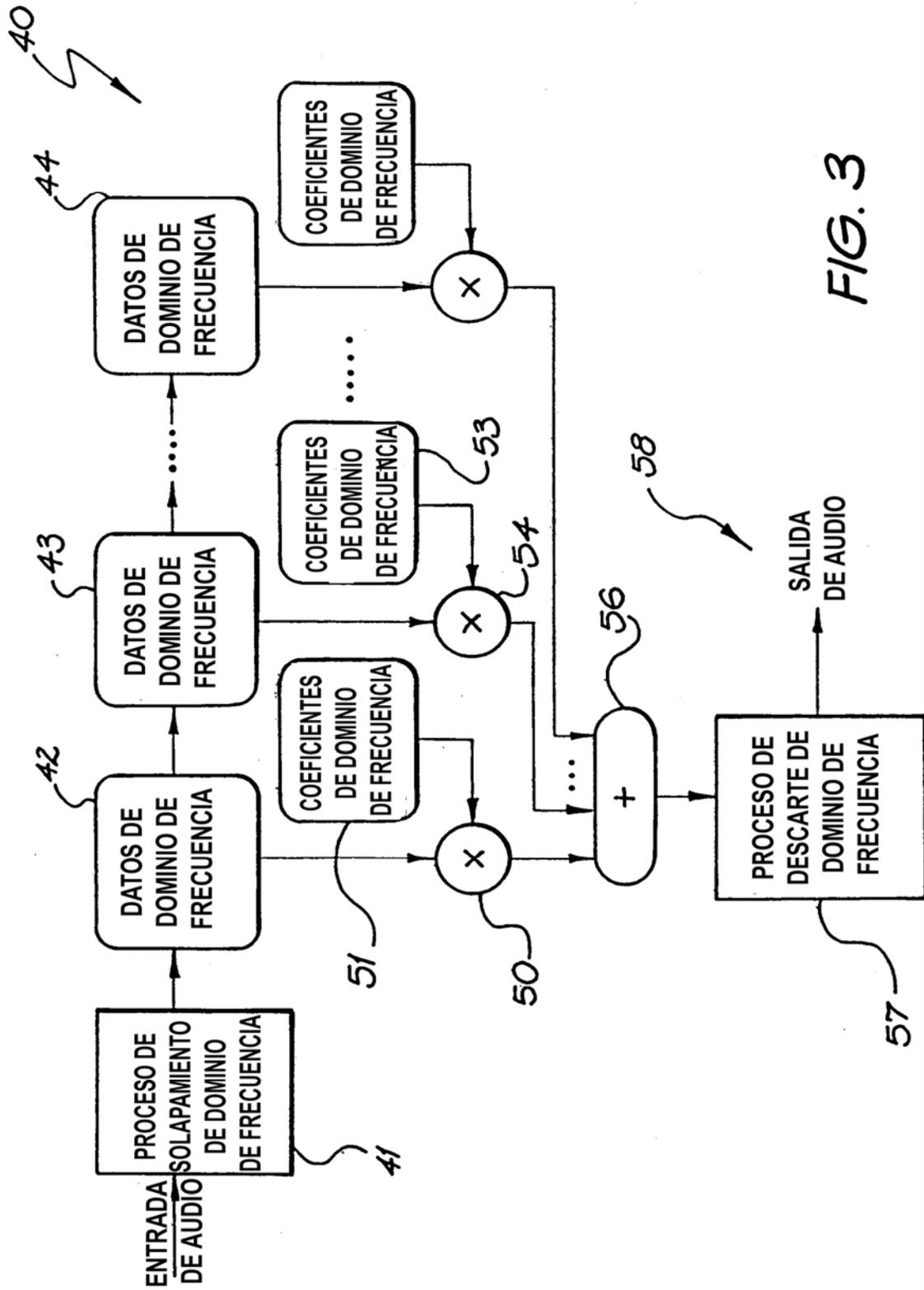


FIG. 3

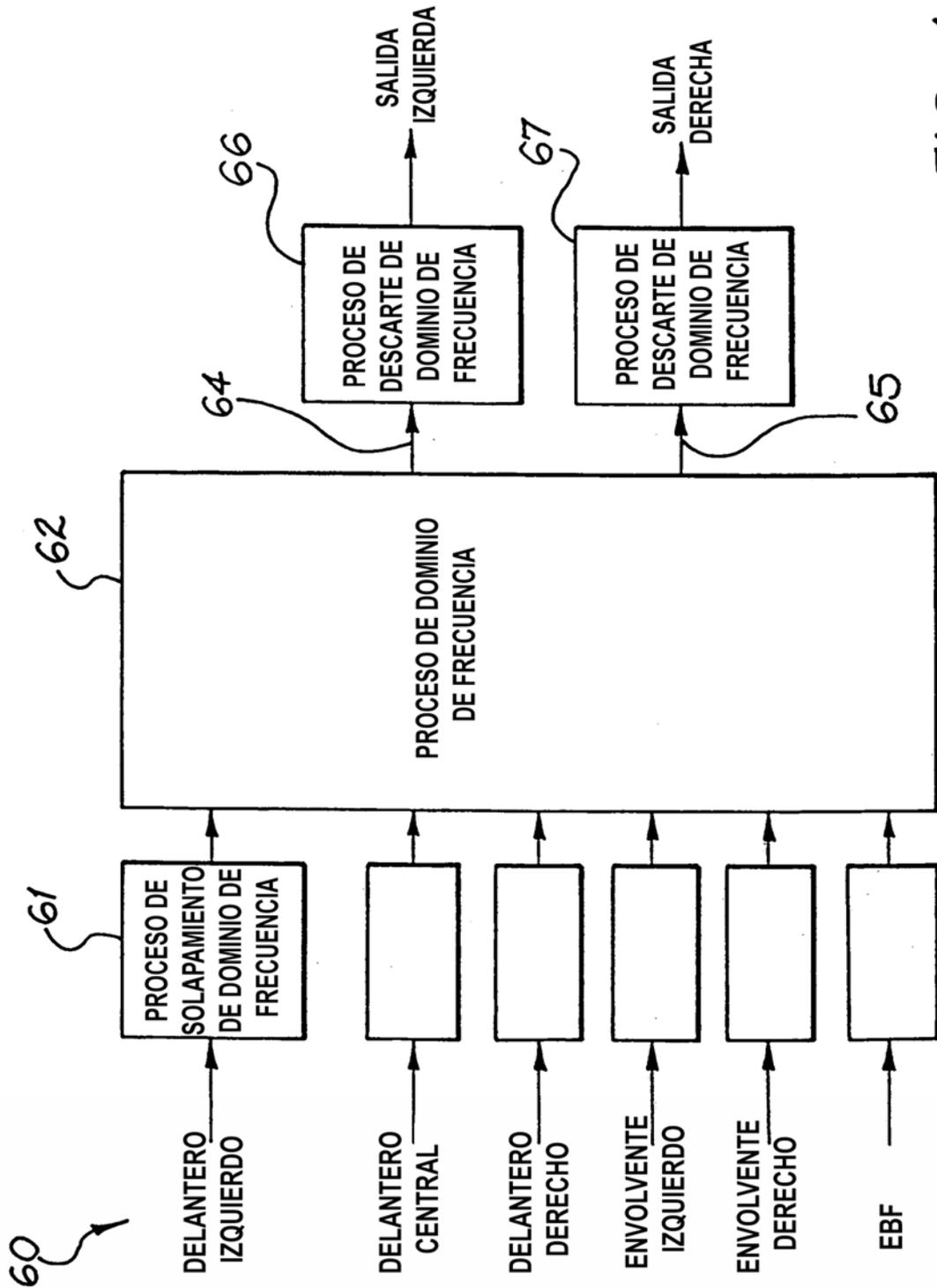


FIG. 4

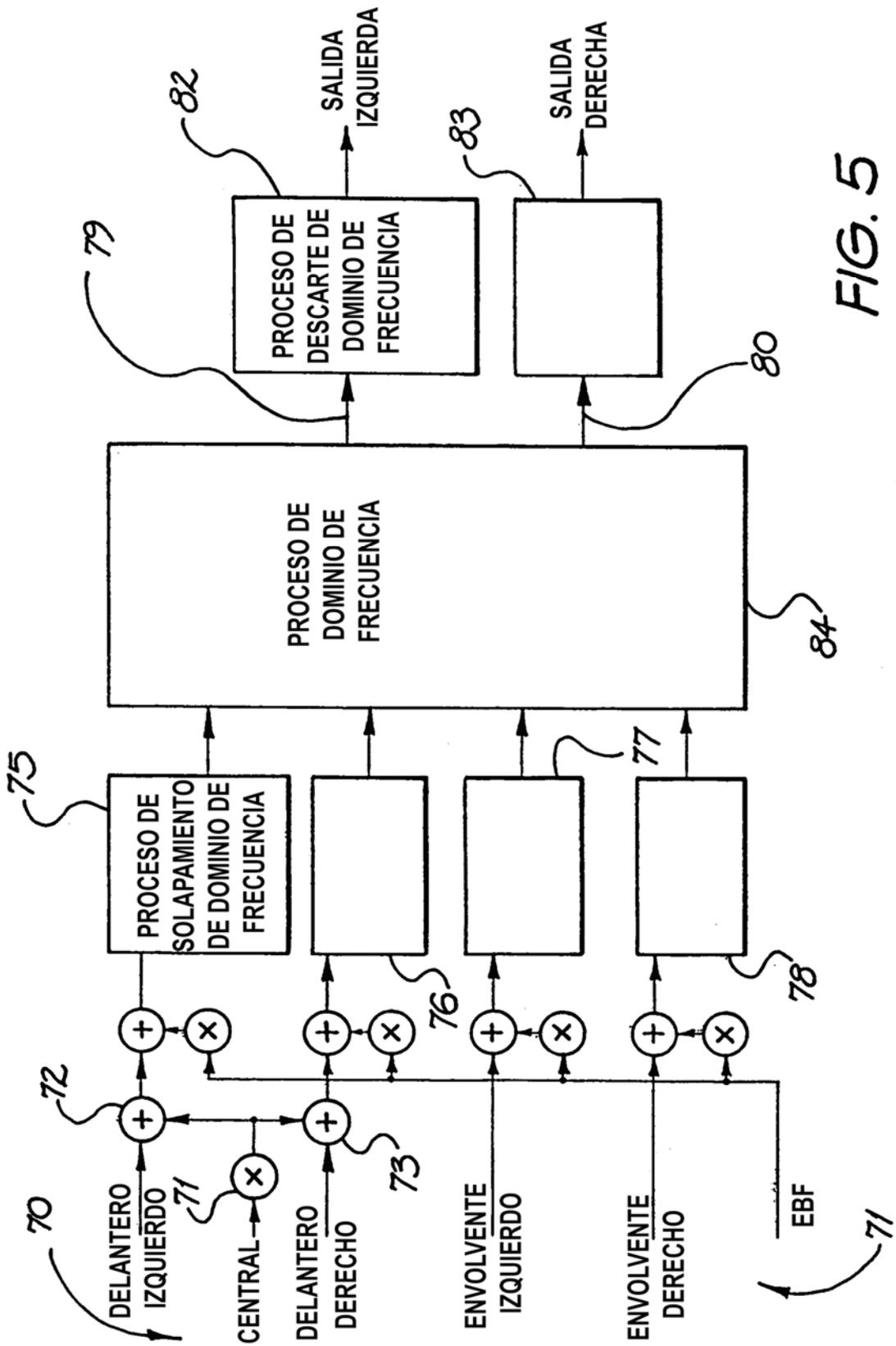
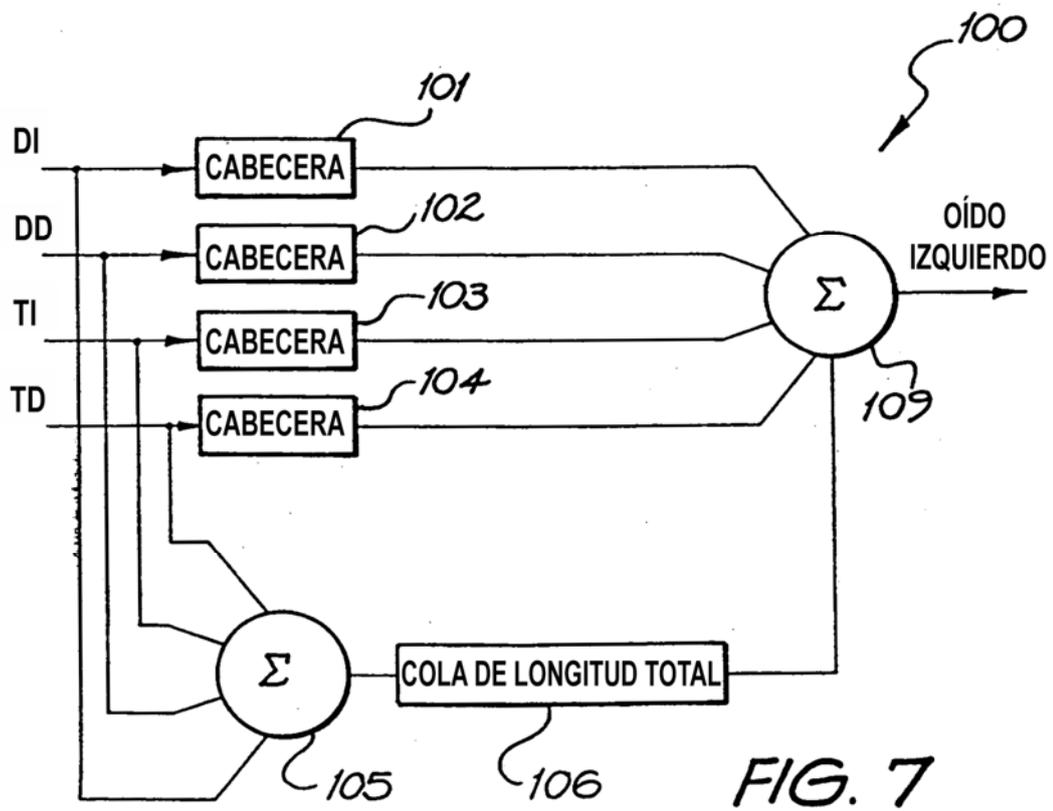
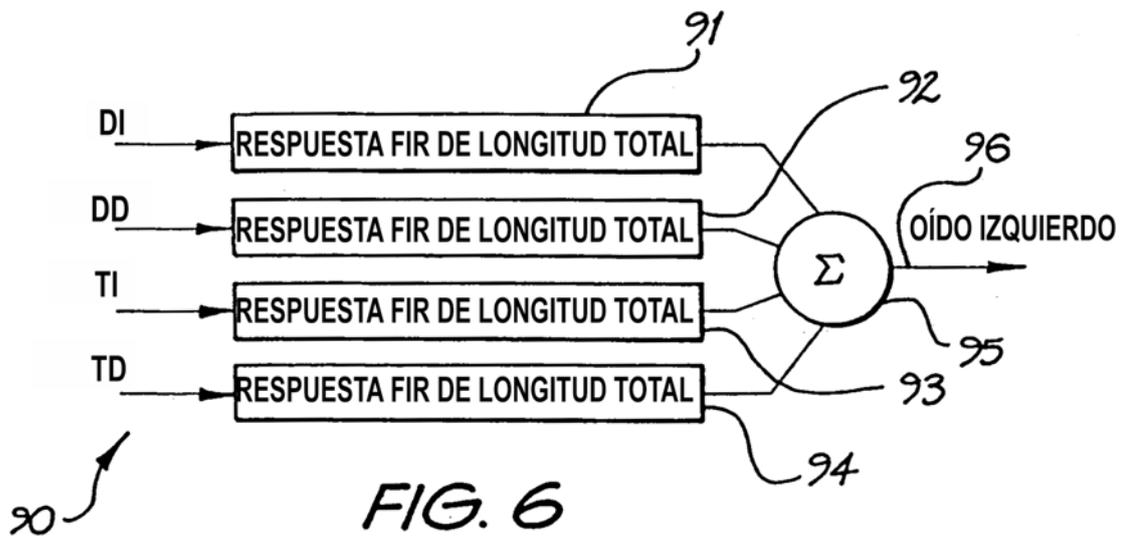


FIG. 5



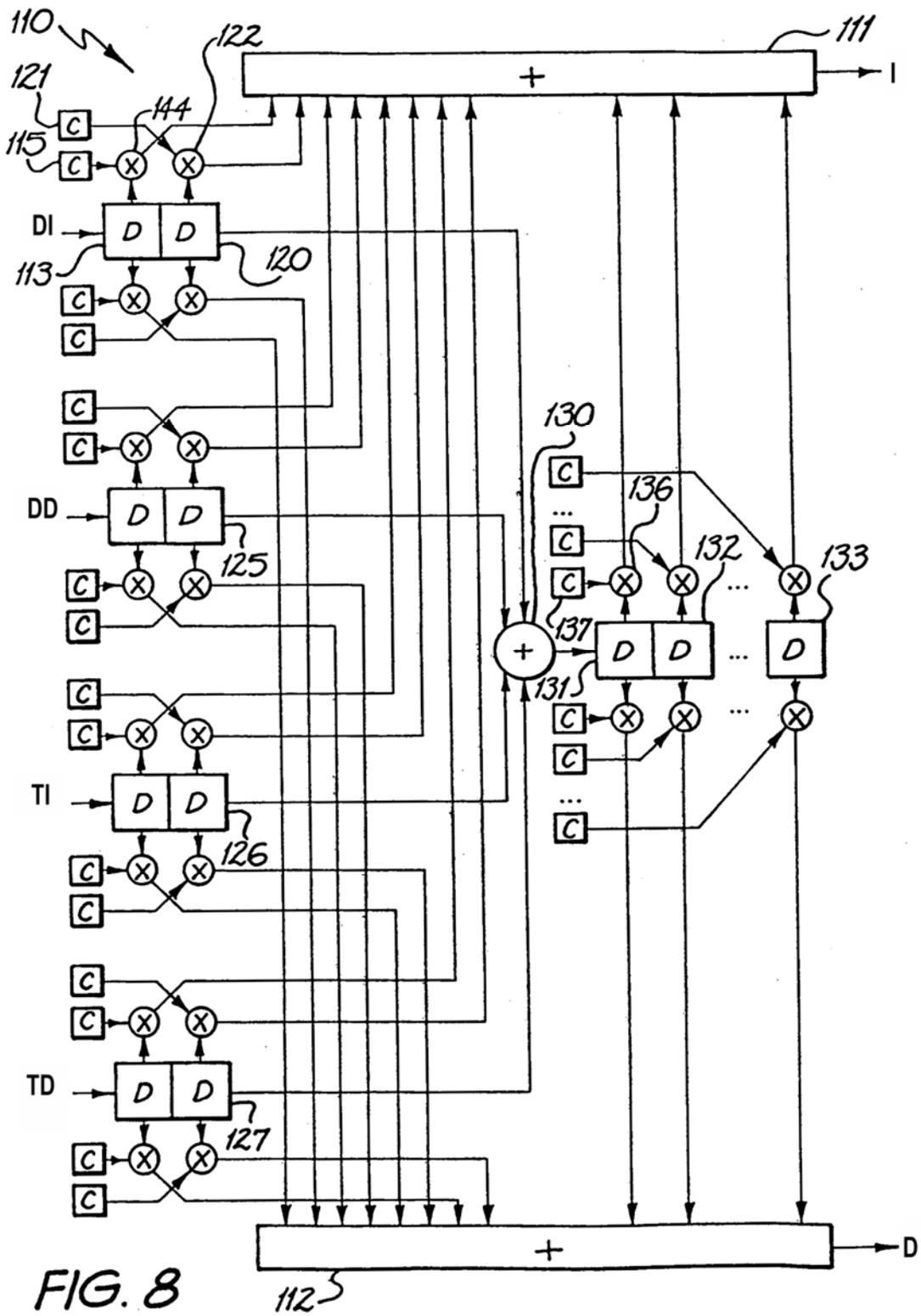


FIG. 8

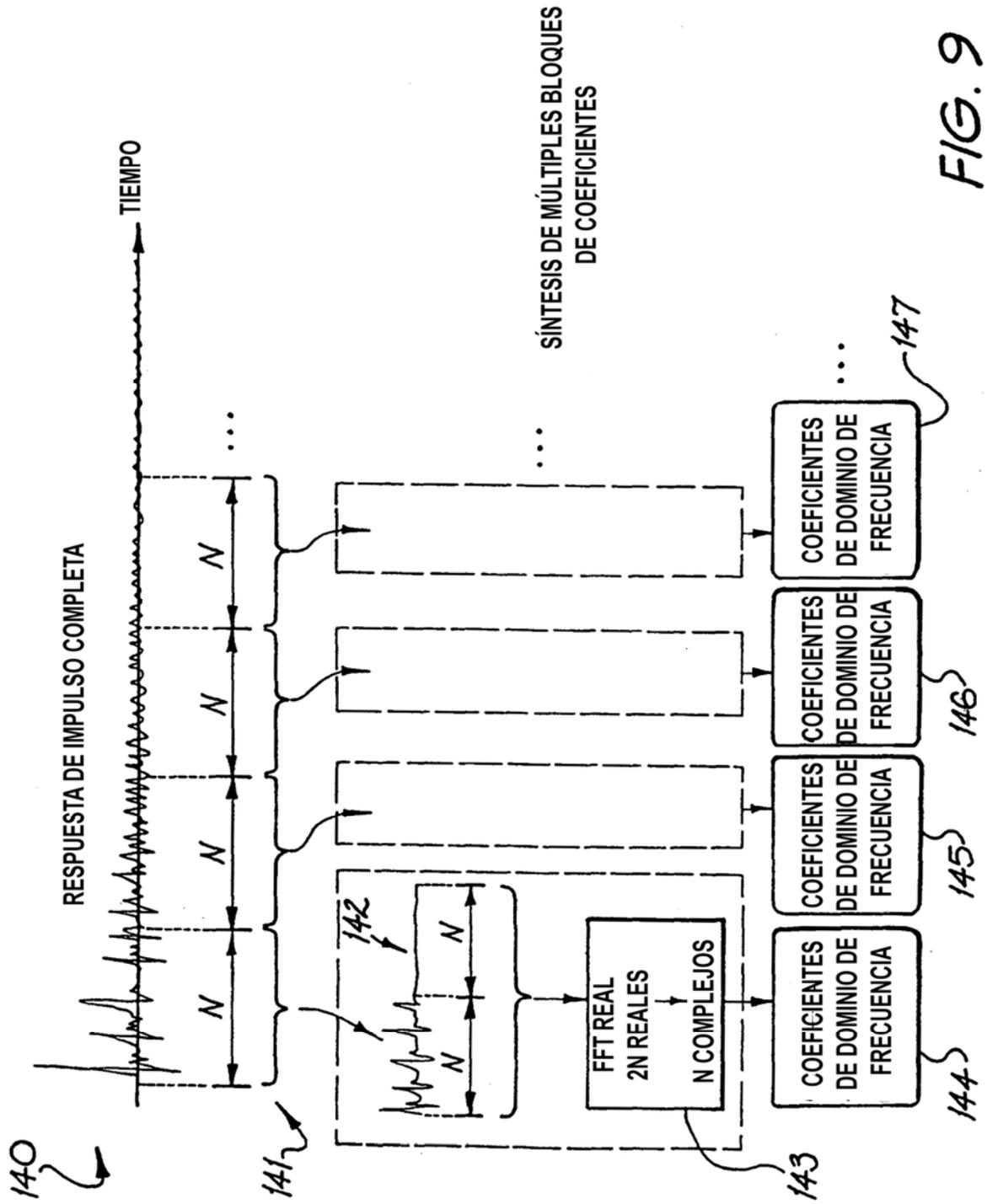
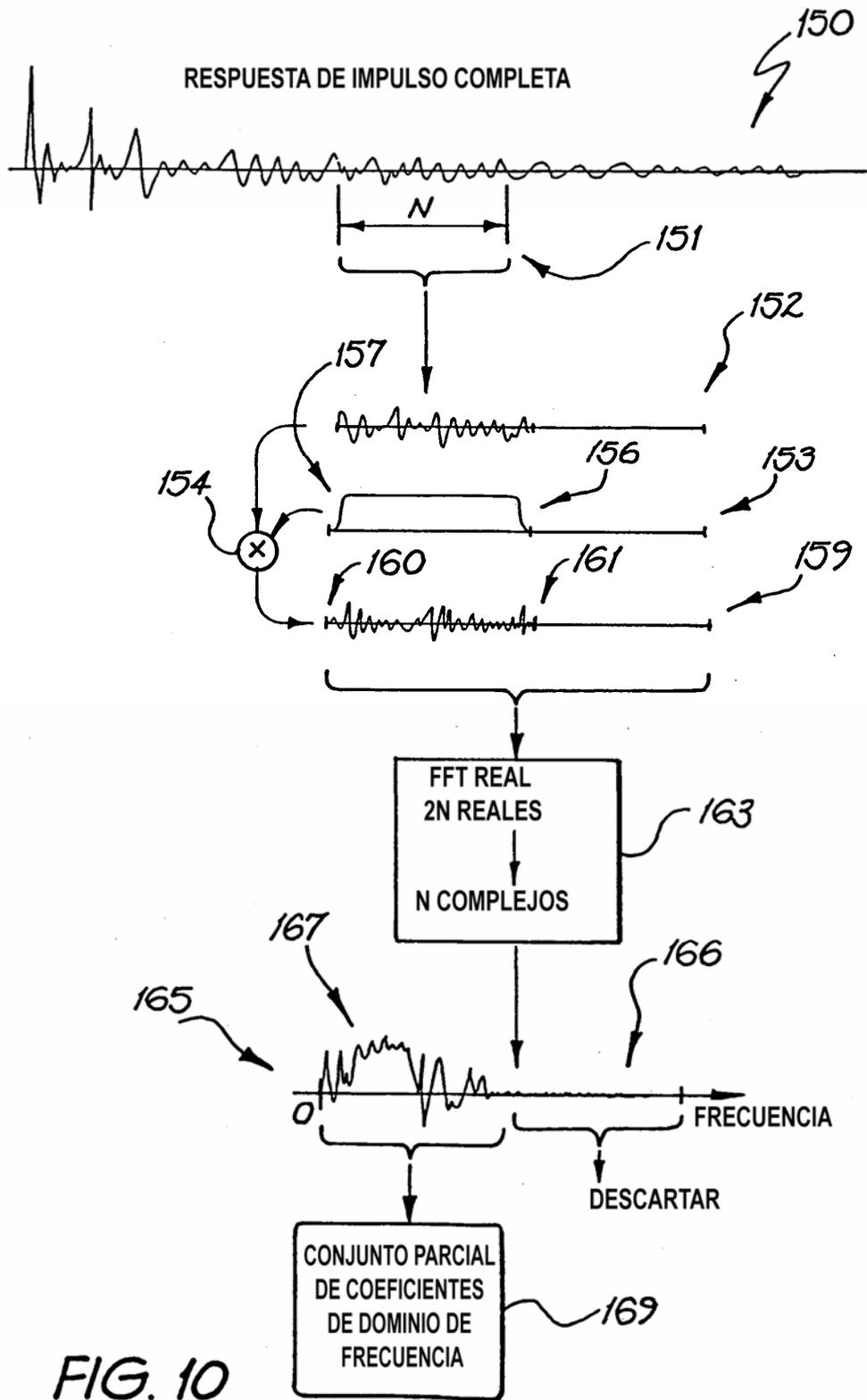


FIG. 9



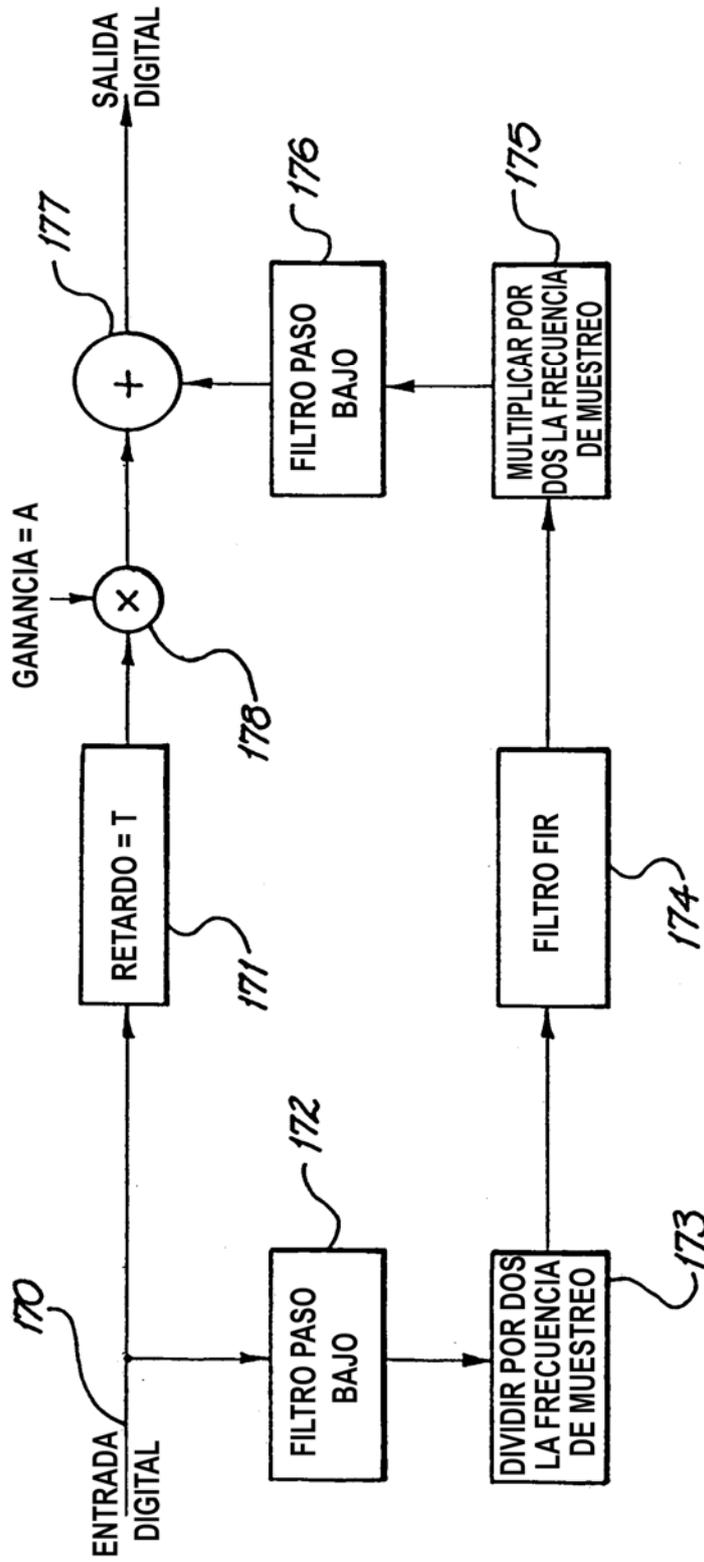


FIG. 11

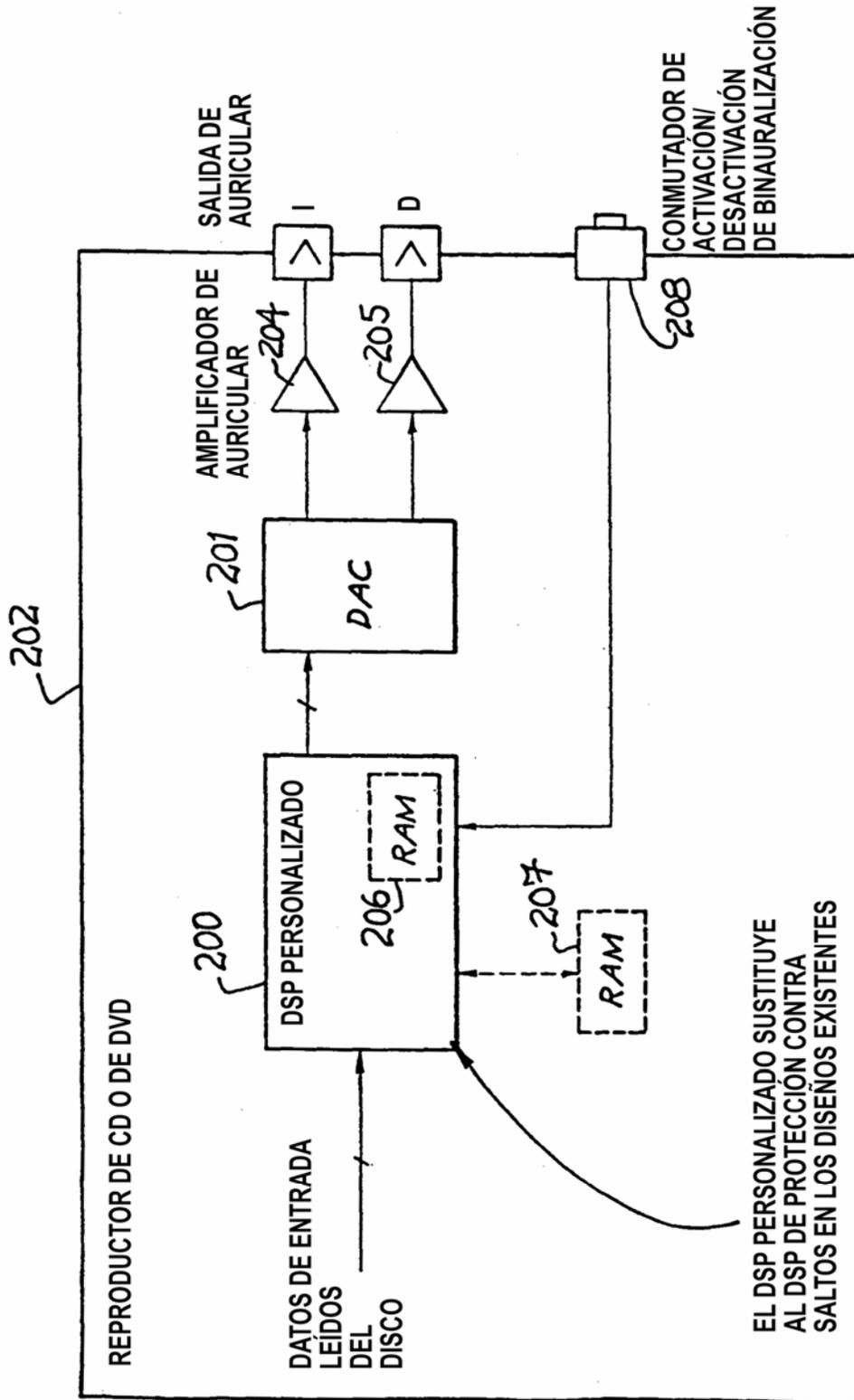


FIG. 12

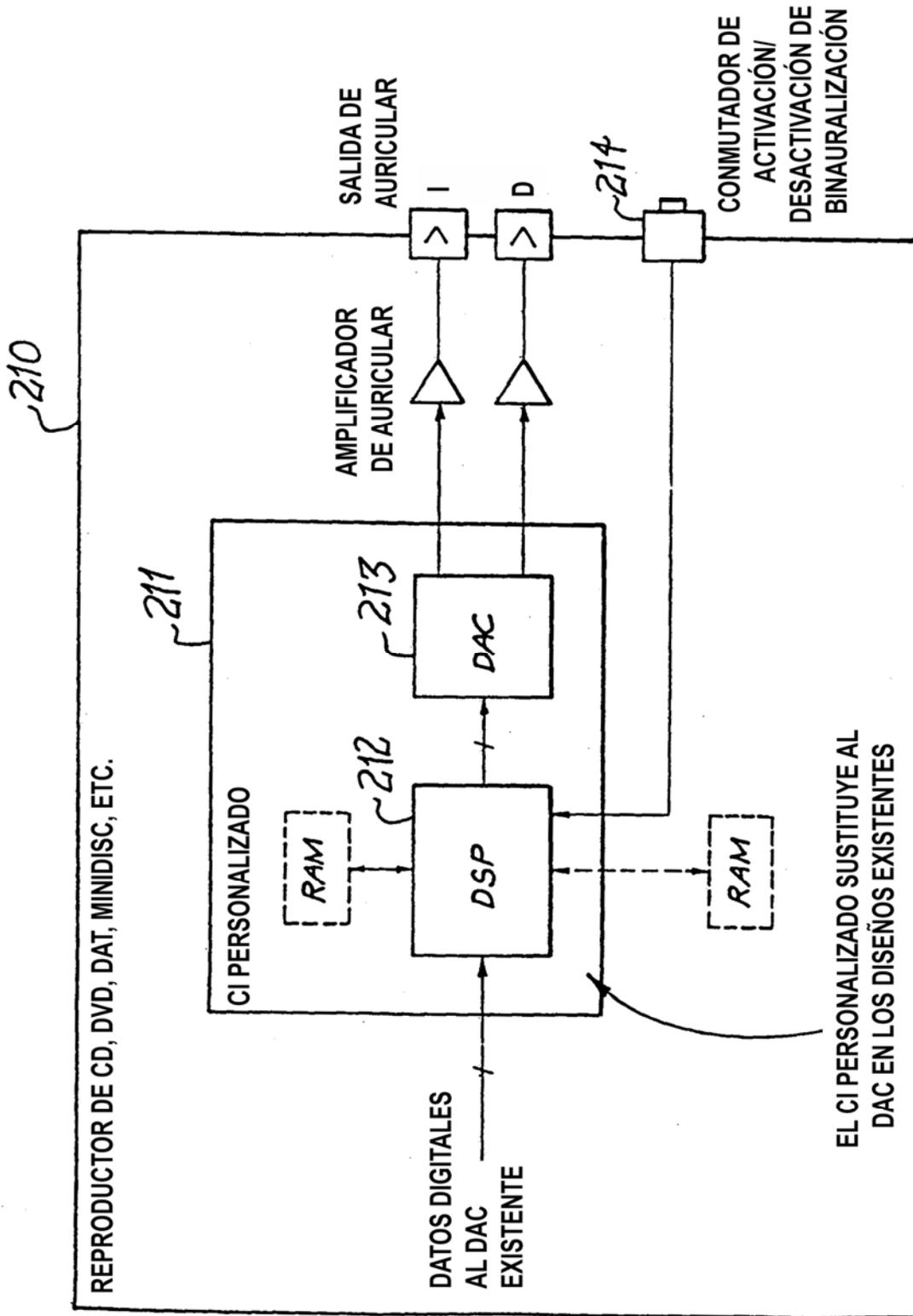


FIG. 13

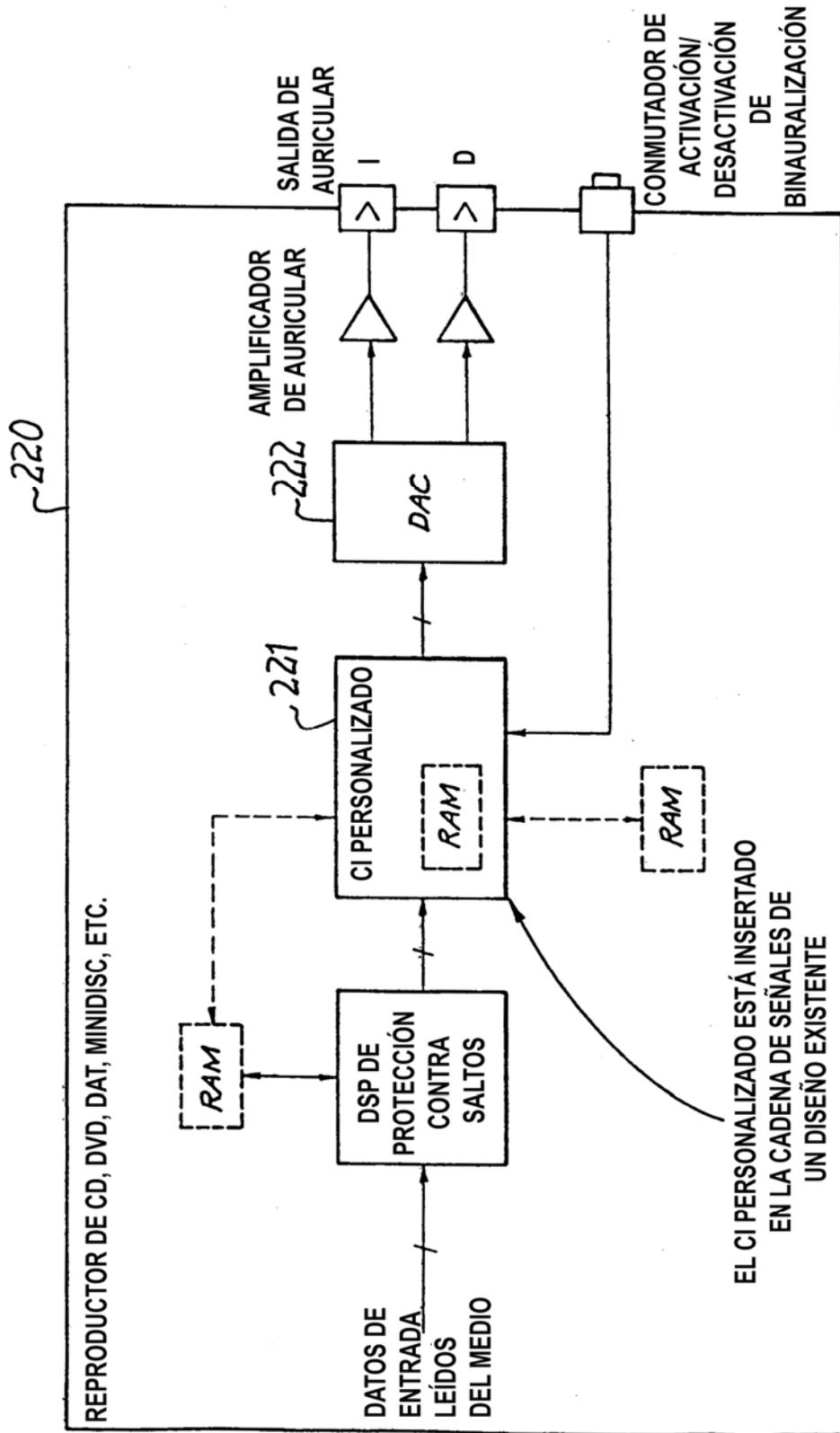


FIG. 14

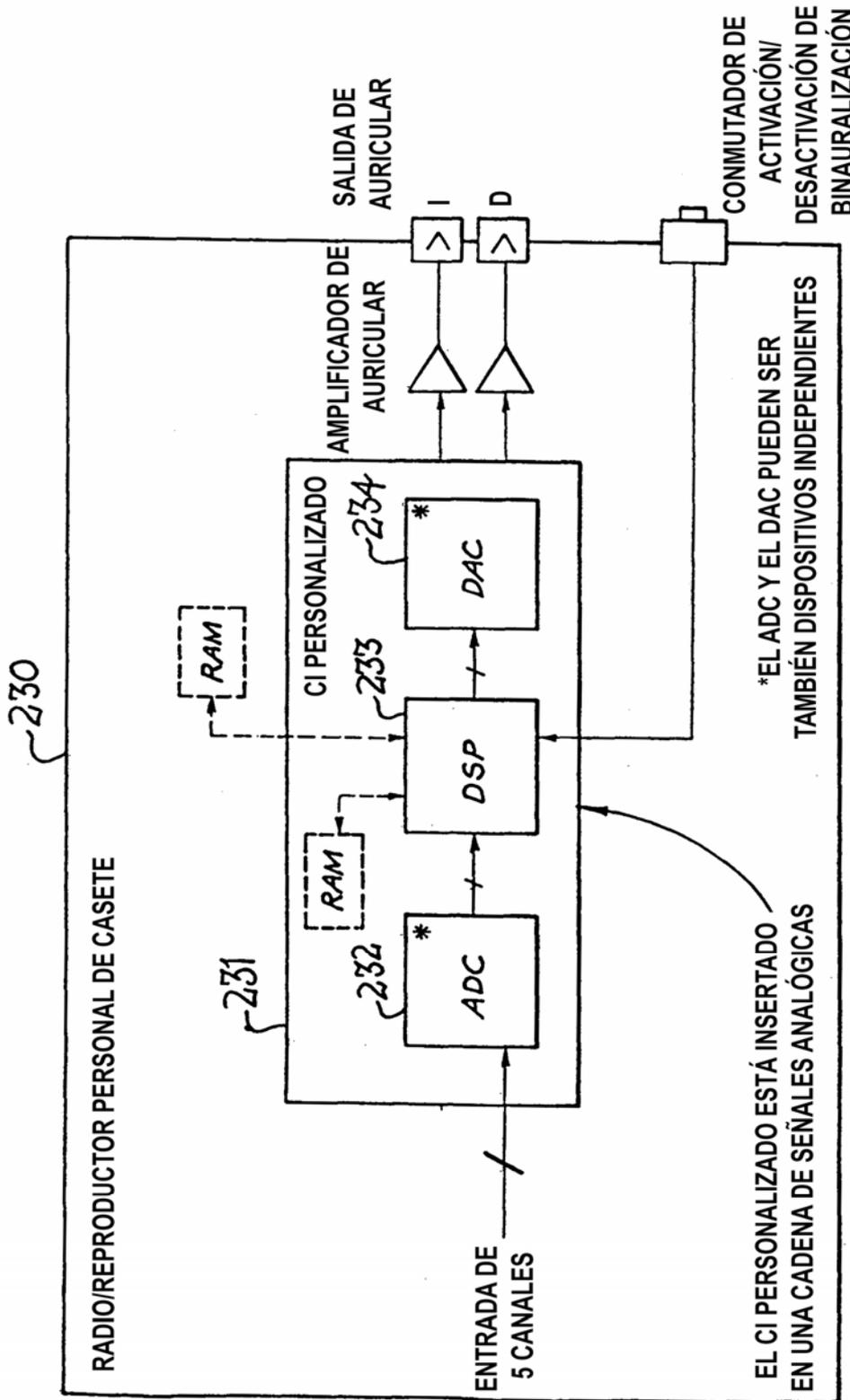


FIG. 15

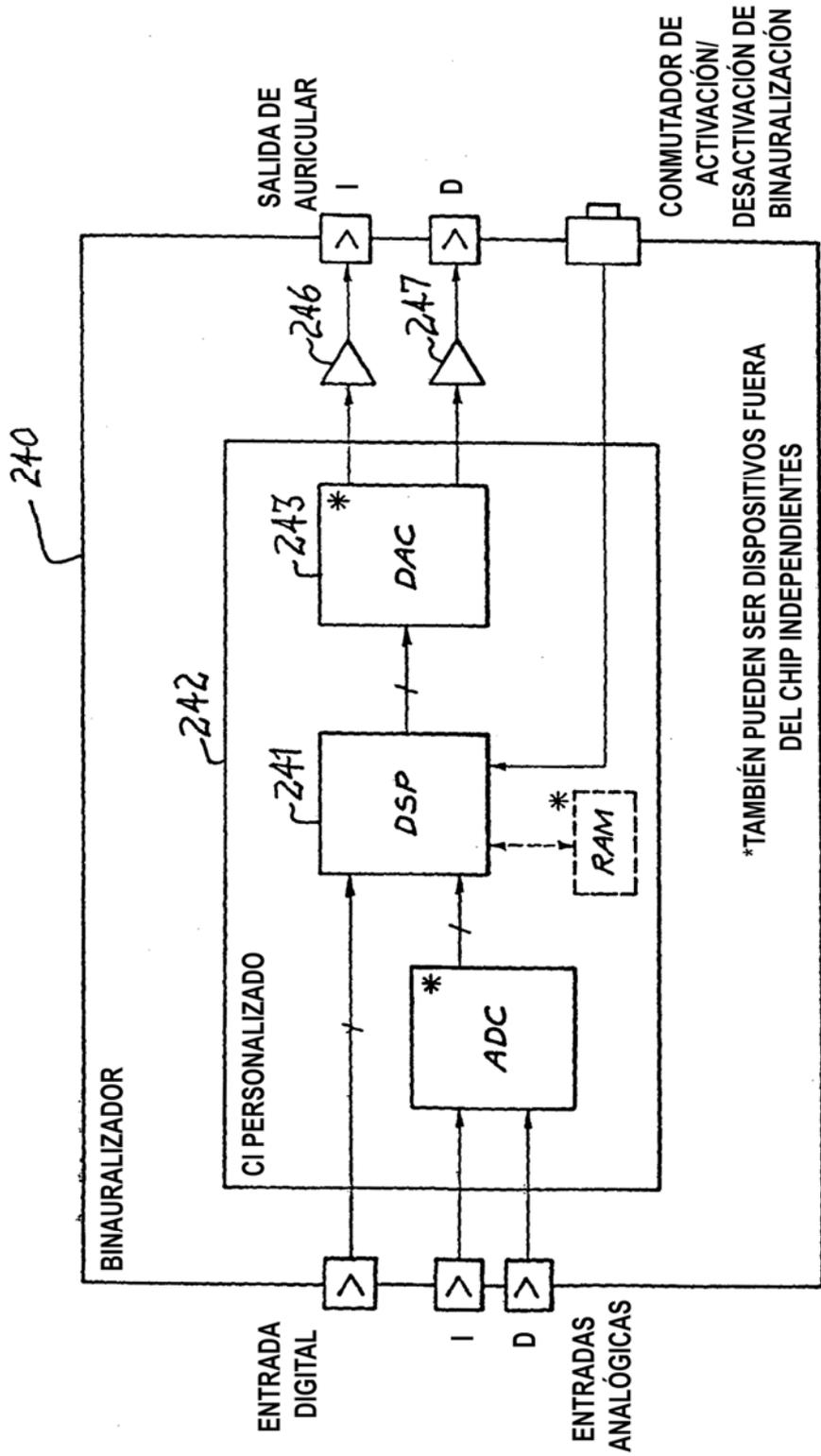


FIG. 16

