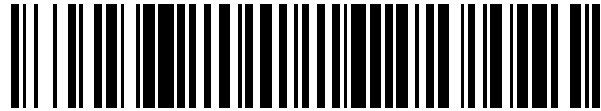


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 529 025**

51 Int. Cl.:

G10L 19/26 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.02.2012 E 12704258 (8)**

97 Fecha y número de publicación de la concesión europea: **03.12.2014 EP 2676268**

54 Título: **Aparato y método para procesar una señal de audio decodificada en un dominio espectral**

30 Prioridad:

14.02.2011 US 201161442632 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
16.02.2015

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**FUCHS, GUILLAUME;
GEIGER, RALF;
SCHNELL, MARKUS;
RAVELLI, EMMANUEL y
DOEHLA, STEFAN**

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 529 025 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

Aparato y método para procesar una señal de audio decodificada en un dominio espectral

5
[0001] La presente invención se relaciona con el procesamiento de audio y, en particular, con el procesamiento de una señal de audio decodificada con el propósito de realzar la calidad.

10
[0002] Recientemente, se alcanzaron otros avances con respecto a los códecs de audio conmutados. Un códec de audio conmutado de alta calidad y baja velocidad de transmisión de bits es el concepto de codificación unificada de voz y audio (concepto USAC). Existe un pre/post-procesamiento común que consiste en una unidad funcional de MPEG envolvente (MPEGs) para manejar un procesamiento estéreo o multicanal y una unidad SBR intensificada (eSBR) que maneja la representación paramétrica de las frecuencias de audio más elevadas en la señal de entrada. Seguidamente hay dos ramas, una que consiste en un tramo de una herramienta de codificación de audio avanzada (AAC) y la otra que consiste en un tramo basado en la codificación por predicción lineal (dominio LP o LPC) que, a su vez, presenta una representación en el dominio frecuencial o una representación en el dominio temporal de la LPC residual. Todos los espectros transmitidos correspondientes tanto a AAC como a LPC están representados en el dominio MDCT siguiendo la cuantización y la codificación aritmética. La representación en el dominio temporal utiliza un esquema de codificación por excitación ACELP. Los diagramas de bloques del codificador y el decodificador están expuestos en la Fig. 1.1 y en la Fig. 1.2 de ISO/IEC CD 23003-3.

20
[0003] Un ejemplo adicional de códec de audio conmutado es el códec de banda ancha adaptativa de múltiples velocidades (AMR-WB+) descrito en 3GPP TS 26.290 V10.0.0 (2011-3). EL códec de audio AMR-WB+ procesa cuadros de entrada iguales a 2048 muestras a una frecuencia de muestreo interno F_s . Las frecuencias de muestreo interno están restringidas al rango de 12800 a 38400 Hz. Los cuadros de 2048 muestras se dividen en dos bandas de frecuencia iguales de muestreo crítico. Esto da lugar a dos súper cuadros de 1024 muestras correspondientes a la banda de baja frecuencia (LF) y de alta frecuencia (HF). Cada súper cuadro está dividido en cuatro cuadros de 256 muestras. El muestreo a la velocidad de muestreo interno se obtiene utilizando un esquema de conversión de muestreo variable que remuestrea la señal de entrada. Las señales de LF y HF son codificadas a continuación utilizando dos estrategias diferentes: la LF es codificada y decodificada utilizando un codificador/decodificador de "núcleo", sobre la base de ACELP conmutada y excitación por códigos de transformadas (TCX). En el modo ACELP se utiliza el códec AMR-WB standard. La señal HF se codifica con una cantidad relativamente baja de bits (16 bits por cuadro) utilizando un método de extensión de ancho de banda (BWE). El codificador AMR-WB incluye una funcionalidad de pre-procesamiento, un análisis de LPC, una funcionalidad de búsqueda de lazo abierto, una funcionalidad de búsqueda de libros de códigos adaptativos, una funcionalidad de búsqueda de libros de códigos innovadores y actualización de memorias. El decodificador ACELP comprende varias funcionalidades, como por ejemplo la decodificación de libros de códigos adaptativos, ganancias de decodificación, la decodificación de los libros de códigos innovativos, ISP de decodificación, un filtro de predicción de largo plazo (filtro LTP), la funcionalidad de excitación de construcción, una interpolación de ISP correspondiente a cuatro subcuadros, un post-procesamiento, un filtro de síntesis, un bloque de desénfasis y un bloque de aumento de número de muestras para obtener, en última instancia, la porción de banda inferior de la salida de voz. La porción más alta de la banda de la salida de voz se genera por el escalamiento de las ganancias utilizando un índice de ganancias HB, una bandera VAD y una excitación aleatoria de 16 kHz. Más aun, se utiliza un filtro de síntesis HB seguida por un filtro pasa banda. Se presentan más detalles en la Fig. 3 de G.722.2.

30
[0004] Se ha mejorado este esquema en la AMR-WB+ mediante la ejecución de un post-procesamiento de la señal mono de banda baja. Se hace referencia a las Figs. 7, 8 y 9 que ilustran la funcionalidad en AMR-WB+. La Fig. 7 ilustra un intensificador de tono 700, un filtro paso bajo 702, un filtro paso alto 704, una etapa de rastreo de tono 706 y un sumador 708. Los bloques están conectados de la manera ilustrada en la Fig. 7 y son alimentados por la señal decodificada.

40
[0005] En la intensificación de tono de baja frecuencia, se utiliza una descomposición en dos bandas y se aplica el filtrado adaptativo sólo a la banda inferior. Esto da lugar a un post-procesamiento total que se dirige mayormente a frecuencias cercanas a los primeros armónicos de la señal de voz sintetizada. La Fig. 7 ilustra el diagrama de bloques del intensificador de tono de dos bandas. En la rama más alta la señal decodificada es filtrada por el filtro paso alto 704 para producir las señales de banda más alta s_H . En la rama más baja, se procesa en primer lugar la señal decodificada por medio del intensificador de tono adaptativo 700 y luego se la filtra por medio del filtro paso bajo 702 para obtener la señal de post-procesamiento de banda inferior (s_{LEE}). La señal decodificada de post-procesamiento se obtiene sumando la señal de post-procesamiento de banda inferior y la señal de banda superior. El objetivo del intensificador de tono consiste en reducir el ruido interarmónico en la señal decodificada que se obtiene por medio de un filtro lineal de tiempo variable con una función de transferencia H_E indicada en la primera línea de la Fig. 9 y descrita por la ecuación de la segunda línea de la Fig. 9. α es un coeficiente que controla la atenuación interarmónica. T es el período de tono de la señal de entrada $\hat{S}(n)$ y $s_{LE}(n)$ es la señal de salida del intensificador de tono. Los parámetros T y α varían con el tiempo y están dados por el módulo de rastreo de tono 706 con un valor de $\alpha = 1$, la ganancia del filtro descrito por la ecuación de la segunda línea de la Fig. 9 es exactamente cero a las frecuencias $1/(2T)$, $3/(2T)$, $5/(2T)$, etc., es decir, en el punto medio entre la DC (0 Hz) y las

frecuencias armónicas $1/T$, $3/T$, $5/T$, etc. Cuando α se aproxima a cero, la atenuación entre los armónicos producidos por el filtro definida en la segunda línea de la Fig. 9 se reduce. Cuando α es cero, el filtro no tiene efecto y es un pasa todo. Para confinar el post-procesamiento a la región de baja frecuencia, la señal intensificada s_{LE} es filtrada por paso bajo para producir la señal s_{LEF} que se suma a la señal de filtro paso alto s_H para obtener la señal de síntesis post-procesamiento s_E .

[0006] En la Fig. 8 se ilustra otra configuración equivalente a la ilustración de la Fig. 7 y la configuración de la Fig. 8 elimina la necesidad de filtrado de paso alto. Esto se explica con respecto a la tercera ecuación correspondiente a s_E en la Fig. 9. El $h_{LP}(n)$ es la respuesta impulsional del filtro paso bajo y $h_{HP}(n)$ es la respuesta impulsional del filtro paso alto complementario. A continuación, la señal post-proceso $s_{E(n)}$ está dada por la tercera ecuación de la Fig. 9. Por consiguiente, el post procesamiento es equivalente a la sustracción de la señal de error a largo plazo filtrada por paso bajo y escalada $\alpha \cdot e_{LT}(n)$ de la señal de síntesis $\hat{s}(n)$. La función de transferencia del filtro de predicción a largo plazo está dada de acuerdo con lo indicado en la última línea de la Fig. 9. Esta configuración alternativa de post-procesamiento está ilustrada en la Fig. 8. El valor T está dado por el retardo de tono de bucle cerrado recibido en cada subcuadro (el retardo de tono fraccionado redondeado al entero más próximo). Se ejecuta un simple rastreo para verificar la duplicación de tonos. Si la correlación de tono normalizado en el retardo $T/2$ es mayor que 0,95, luego se utiliza el valor $T/2$ como nuevo retardo de tono correspondiente al post-procesamiento. El factor α está dado por $\alpha = 0,5g_p$, que se limita a α mayor o igual a cero y menor o igual a 0,5. g_p es la ganancia de tono decodificada limitada entre 0 y 1. En el modo TCX, se ajusta el valor de α a cero. Se utiliza el filtro paso bajo FIR de fase lineal con 25 coeficientes con la frecuencia de corte de aproximadamente 500 Hz. El retardo del filtro es de 12 muestras). La rama superior necesita introducir un retardo que corresponde al retardo del procesamiento en la rama inferior para mantener las señales de las dos ramas alineadas antes de ejecutar la sustracción. En AMR-WB+ $F_s=2x$ la velocidad de muestreo del núcleo. La velocidad de muestreo del núcleo es igual a 12800 Hz. Por lo tanto, la frecuencia de corte es igual a 500Hz.

[0007] Se ha encontrado que, en especial en el caso de las aplicaciones de bajo retardo, el retardo de filtro de 12 muestras introducido por el filtro de paso bajo FIR de fase lineal contribuye al retardo total del esquema de codificación/ decodificación. Hay otras fuentes de retardos sistemáticos en otros lugares de la cadena de codificación/ decodificación, y el retardo de filtro FIR se acumula con las demás fuentes.

[0008] Uno de los objetivos de la presente invención es presentar un concepto de procesamiento mejorado de señales de audio que se adecúa más a las aplicaciones en tiempo real o situaciones de comunicaciones de dos vías tales como las situaciones de teléfonos móviles.

[0009] Este objetivo se logra mediante un aparato para procesar una señal de audio decodificada de acuerdo con la reivindicación 1 o un método para procesar una señal de audio decodificada de acuerdo con la reivindicación 15 o un programa de computación de acuerdo con la reivindicación 16.

[00010] La presente invención se basa en el hallazgo de que la contribución del filtro paso bajo en el post filtrado de bajos de la señal decodificada al retardo total es problemático y tiene que reducirse. Para este fin, no se filtra por paso bajo la señal de audio filtrada en el dominio temporal sino que se la filtra por paso bajo en el dominio espectral, como por ejemplo un dominio QMF o cualquier otro dominio espectral, como por ejemplo un dominio MDCT, un dominio FFT, etc. Se ha descubierto que la transformada del dominio espectral al dominio frecuencial y, por ejemplo, a un dominio frecuencial de baja resolución tal como un dominio QMF se puede ejecutar con bajo retardo y la selectividad por frecuencias del filtro a implementar en el dominio espectral puede ser implementada simplemente ponderando las señales de subbandas individuales de la representación en el dominio frecuencial de la señal de audio filtrada. Esta "impresión" de la característica seleccionada según la frecuencia se ejecuta, por lo tanto, sin retardo sistemático alguno, puesto que una operación de multiplicación o ponderación con una señal de subbanda no incurre en ningún retardo. La sustracción de la señal de audio filtrada y la señal de audio original es ejecutada también en el dominio espectral. Por añadidura, es preferible ejecutar operaciones adicionales que de todas maneras son necesarias, como por ejemplo se ejecuta una decodificación por replicación de bandas espectrales o una decodificación estéreo o multicanal en el mismísimo dominio QMF. Se realiza una conversión de frecuencia-tiempo sólo al final de la cadena de decodificación a fin de llevar la señal de audio producida en última instancia al dominio temporal. Por ende, dependiendo de la aplicación, la señal de audio generada como resultado por el sustractor se puede convertir de regreso al dominio temporal como ocurre cuando ya no se necesitan más operaciones de procesamiento en el dominio QMF. Sin embargo, cuando el algoritmo de decodificación tiene operaciones adicionales de procesamiento en el dominio QMF, en ese caso no se conecta el convertidor de frecuencia-tiempo a la salida del sustractor, sino que se conecta a la salida del último dispositivo de procesamiento en el dominio frecuencial.

[00011] De preferencia, el filtro para filtrar la señal de audio decodificada es un filtro de predicción a largo plazo. Más aun, es preferible que la representación espectral sea una representación QMF y también es preferible que la selectividad por frecuencia sea una característica de paso bajo.

[00012] Sin embargo, se puede utilizar cualquier otro filtro diferente de un filtro de predicción a largo plazo, cualquier otra representación espectral diferente de una representación QMF o cualquier otra selectividad de frecuencia

diferente de una característica de paso bajo para obtener un post-procesamiento de bajo retardo de una señal de audio decodificada.

[00013] A continuación se describen las realizaciones preferidas de la presente invención con respecto a los dibujos que la acompañan, en los cuales:

La Fig. 1a es un diagrama de bloques de un aparato para procesar una señal de audio decodificada de acuerdo con una realización;

La Fig. 1b es un diagrama de bloques de una forma de realización preferida del aparato para procesar una señal de audio decodificada;

La Fig. 2a ilustra una característica selectiva en frecuencia a título de ejemplo como característica de paso bajo;

La Fig. 2b ilustra coeficientes de ponderación y las subbandas asociadas;

La Fig. 2c ilustra una cascada del convertidor de tiempo/ espectral y un ponderador conectado subsiguientemente para aplicar coeficientes de ponderación a cada señal de subbanda individual;

La Fig. 3 ilustra una respuesta impulsional en la respuesta frecuencial del filtro paso bajo en AMR-WB+ expuesto en la Fig. 8;

La Fig. 4 ilustra una respuesta impulsional y la respuesta frecuencial transformada al dominio QMF;

La Fig. 5 ilustra factores de ponderación para los ponderadores correspondientes al ejemplo de 32 subbandas QMF;

La Fig. 6 ilustra la respuesta frecuencial correspondiente a 16 bandas QMF y los 16 factores de ponderación asociados;

La Fig. 7 ilustra un diagrama de bloques del intensificador de tono de baja frecuencia de AMR-WB+;

La Fig. 8 ilustra una configuración implementada de post-procesamiento de AMR-WB+;

La Fig. 9 ilustra una derivación de la implementación de la Fig. 8; y

La Fig. 10 ilustra una implementación de bajo retardo del filtro del filtro de predicción a largo plazo de acuerdo con una realización.

[00014] La Fig. 1a ilustra un aparato para procesar una señal de audio decodificada en la línea 100. La señal de audio decodificada en la línea 100 es suministrada como entrada al filtro 102 para filtrar la señal de audio decodificada a fin de obtener una señal de audio filtrada en la línea 104. El filtro 102 está conectado a una etapa de conversión de tiempo a espectral 106 ilustrada en forma de dos convertidores de tiempo a espectral individuales 106a en el caso de la señal de audio filtrada y 106b en el de la señal de audio decodificada en la línea 100. La etapa de conversión de tiempo a espectral está configurada para convertir la señal de audio y la señal de audio filtrada a una representación espectral correspondiente cada una de las cuales tiene una pluralidad de señales de subbanda. Esto está indicado por líneas dobles en la Fig. 1a, lo que indica que la salida de los bloques 106a, 106b comprende una pluralidad de señales de subbanda individuales en lugar de una señal única ilustrada en conexión con la entrada a los bloques 106a, 106b.

[00015] El aparato para procesar comprende además un ponderador 108 para ejecutar una ponderación selectiva según la frecuencia de la señal de audio filtrada producida como salida del bloque 106a mediante la multiplicación de las señales de subbanda individuales por respectivos coeficientes de ponderación a fin de obtener una señal de audio filtrada y ponderada en la línea 110.

[00016] Además, se incluye un sustractor 112. El sustractor está configurado para ejecutar una sustracción por subbandas entre la señal de audio filtrada y ponderada y la representación espectral de la señal de audio generada por el bloque 106b.

[00017] Por añadidura se incluye un convertidor de espectral a tiempo 114. La conversión de espectral a tiempo ejecutada por el bloque 114 es tal que la señal de audio generada como resultado por el sustractor 112 o una señal derivada de la señal de audio así obtenida se convierte en una representación en el dominio temporal a fin de obtener la señal de audio decodificada y procesada en la línea 116.

[00018] Si bien la Fig. 1a indica que el retardo por la conversión de tiempo a espectral y la ponderación es significativamente menor que el retardo producido por el filtrado FIR, esto no es indispensable en todos los casos, ya que en situaciones en las cuales el QMF es absolutamente necesario se evita la acumulación de retardos del filtrado FIR y QMF. Por ende, la presente invención también es ventajosa cuando el retardo por la ponderación y conversión de tiempo a espectral es aun mayor que el retardo de un filtro FIR para el post filtrado de bajos.

[00019] La Fig. 1b ilustra una forma de realización preferida de la presente invención en el contexto del decodificador USAC o el decodificador AMR-WB+. El aparato ilustrado en la Fig. 1b comprende una etapa de decodificador ACELP 120, una etapa de decodificador TCX 122 y un punto de conexión 124 en el cual se conectan las salidas de los decodificadores 120, 122. El punto de conexión 124 da comienzo a dos ramas individuales. La primera rama comprende el filtro 102 que está configurado, preferentemente, como filtro de predicción a largo plazo que está establecido por el retardo de tono T seguido por un amplificador 129 de una ganancia adaptativa α . Más aun, la primera rama comprende el convertidor de tiempo a espectral 106a que está implementado preferentemente en forma de banco de filtros de análisis QMF. Por añadidura, la primera rama comprende el ponderador 108 que está configurado para ponderar las señales de subbanda generadas por el banco de filtros de análisis QMF 106a.

[00020] En la segunda rama, la señal de audio decodificada se convierte al dominio espectral en el banco de filtros de análisis QMF 106b.

[00021] Si bien se ilustran los bloques individuales de QMF 106a, 106b en forma de dos elementos separados, se debe notar que, para analizar la señal de audio filtrada y la señal de audio, no es indispensable necesariamente tener dos bancos de filtros de análisis QMF. Por el contrario, basta con un único banco de filtros de análisis QMF y una memoria cuando las señales se transforman una tras otra. Sin embargo, en el caso de las implementaciones con muy bajo retardo, es preferible utilizar los bancos de filtros de análisis QMF individuales por cada señal por lo que el bloque QMF único no forma el cuello de botella del algoritmo.

[00022] De preferencia, la conversión al dominio espectral y de regreso al dominio temporal es ejecutada por un algoritmo que tiene un retardo por el hecho de que la transformada directa e inversa es menor que el retardo del filtrado en el dominio temporal con la característica selectiva según la frecuencia. Por consiguiente, las transformadas deben tener un retardo total menor que el retardo del filtro en cuestión. Son de particular utilidad las transformadas de baja resolución tal como las transformadas basadas en QMF, puesto que la resolución de baja frecuencia da lugar a la necesidad de una ventana de transformación pequeña, es decir a un retardo sistemático reducido. Las aplicaciones preferidas sólo requieren una transformada de baja resolución que descomponga la señal en menos de 40 subbandas, como por ejemplo en 32 o sólo 16 subbandas. Sin embargo, aun en aplicaciones en que la conversión de tiempo a espectral y ponderación introducen un retardo mayor que el filtro paso bajo, se obtiene una ventaja debido al hecho de que se evita una acumulación de los retardos correspondientes al filtro paso bajo y la conversión de tiempo a espectral necesaria de todas maneras para otros procedimientos.

[00023] Sin embargo, en el caso de aplicaciones que requieren de todas maneras una conversión de tiempo frecuencia debido a otras operaciones de procesamiento tales como remuestreo, SBR o MPS, se obtiene una reducción del retardo independientemente del retardo en que se incurre por la conversión de tiempo-frecuencia o frecuencia-tiempo, puesto que la "inclusión" de la implementación de filtro en el dominio espectral, se evita por completo el retardo por el filtro en el dominio temporal debido a que se ejecuta una ponderación por subbandas sin retardo sistemático alguno.

[00024] El amplificador adaptativo 129 está controlado por un controlador 130. El controlador 130 está configurado para ajustar la ganancia α del amplificador 129 a cero, cuando la señal de entrada es una señal decodificada por TCX. Por lo general, en los códecs de audio con conmutación tales como USAC o AMR-WB+, la señal decodificada en el punto de conexión 124 es por lo general el decodificador TCX 122 o del decodificador ACELP 120. Por ende, hay un multiplex de tiempo de señales de salida decodificadas de los dos decodificadores 120, 122. El controlador 130 está configurado para determinar, con respecto a un instante de tiempo actual, si la señal de salida es de una señal decodificada por TCX o una señal decodificada por ACELP. Cuando se determina que hay una señal TCX, luego se ajusta a cero la ganancia adaptativa α de manera que la primera rama que consiste en los elementos 102, 129, 106a, 108 no tiene significancia alguna. Esto se debe al hecho de que el tipo específico de post filtrado utilizado en AMR-WB+ o USAC sólo es necesario para la señal codificada por ACELP. Sin embargo, cuando se ejecutan otras implementaciones de post filtrado aparte del filtrado de armónicos o la intensificación de tono, luego se puede ajustar una ganancia variable α de manera diferente dependiendo de las necesidades.

[00025] Sin embargo, cuando el controlador 130 determina que la señal disponible en el momento es una señal decodificada por ACELP, luego se ajusta valor del amplificador 129 al valor correcto para α que por lo general es de entre 0 y 0,5. En este caso, la primera rama es significativa y la señal de salida del sustractor 112 es sustancialmente diferente de la señal de audio codificada originalmente en el punto de conexión 124.

[00026] La información de tono (retardo de tono y ganancia alfa) utilizada en el filtro 120 y el amplificador 128 puede provenir del decodificador y/o de un rastreador de tono especializado. De preferencia, la información procede del

decodificador y luego es reprocesada (refinada) por medio de un rastreador de tonos especializado / análisis de predicción a largo plazo de la señal decodificada.

5 **[00027]** La señal de audio así obtenida generada por el sustractor 112 que ejecuta la sustracción por banda o por subbanda no se convierte de inmediato de regreso al dominio temporal. Por el contrario, la señal es enviada a un módulo decodificador SBR 128. El módulo 128 está conectado a un decodificador mono-estéreo o mono-multicanal tal como un decodificador MPS 131, donde MPS significa MPEG envolvente.

10 **[00028]** Por lo general, el número de bandas es incrementado por el decodificador de replicación de ancho de banda espectral que está indicado por las tres líneas adicionales 132 a la salida del bloque 128.

15 **[00029]** Además, el número de salidas también es incrementado por el bloque 131. El bloque 131 genera, a partir de la señal mono a la salida del bloque 129, por ejemplo, una señal de 5 canales o cualquier otra señal que tenga dos o más canales. A título de ejemplo, se ilustra una situación de 5 canales que consta de un canal izquierdo L, un canal derecho R, un canal central C, un canal envolvente izquierdo L_s y un canal envolvente derecho R_s . El convertidor espectral a tiempo 114 existe, por lo menos, para cada uno de los canales individuales, es decir que existe cinco veces en la Fig. 1b para convertir cada señal de canal individual del dominio espectral que, en el ejemplo de la Fig. 1b, es el dominio QMF, de regreso al dominio temporal a la salida del bloque 114. Una vez más, no es necesario que haya una pluralidad de convertidores espectral a tiempo individuales. Puede haber uno solo también, que procese las conversiones una tras otra. Sin embargo, cuando es necesaria una implementación con muy bajo retardo, es preferible emplear un convertidor individual de espectral a tiempo por cada canal.

20 **[00030]** La presente invención es ventajosa por el hecho de que se reduce el retardo introducido por el post filtrado de bajos y, específicamente, por la implementación del filtro paso bajo FIR. Por ende, ningún tipo de filtrado selectivo por la frecuencia introduce un retardo adicional con respecto al retardo necesario para el QMF o, en términos generales, la transformada de tiempo/frecuencia.

25 **[00031]** La presente invención es particularmente ventajosa cuando se necesita una QMF o, en general, una transformada de tiempo-frecuencia de todas maneras, como por ejemplo en el caso de la Fig. 1b, donde la funcionalidad SBR y la funcionalidad MPS son ejecutadas de todas maneras en el dominio espectral. Una implementación alternativa en la cual se necesita un QMF es cuando se lleva a cabo el remuestreo de la señal decodificada, y cuando, para el remuestreo, es necesario un banco de filtros de análisis QMF y un banco de filtros de síntesis QMF con un número diferente de canales de bancos de filtros.

30 **[00032]** Por añadidura, se mantiene un encuadre constante entre ACELP y TCX debido a que ambas señales, es decir, TCX y ACELP, tienen ahora el mismo retardo.

35 **[00033]** Se describe la funcionalidad de un decodificador de extensión de ancho de banda 129 en forma detallada en la sección 6.5 de ISO/IEC CD 23003-3. La funcionalidad del decodificador multicanal 131 ha sido descrita en detalle, por ejemplo en la sección 6.11 de ISO/IEC CD 23003-3. Las funcionalidades detrás del decodificador TCX y el decodificador ACELP han sido descritas en detalle en los bloques 6.12 a 6.17 de ISO/IEC CD 23003-3.

40 **[00034]** A continuación se describen las Figs. 2a a 2c para ilustrar un ejemplo esquemático. La Fig. 2a ilustra una respuesta frecuencial selectiva en frecuencia de un filtro paso bajo esquemático.

45 **[00035]** La Fig. 2b ilustra los índices de ponderación correspondientes a los números de subbandas o las subbandas indicadas en la Fig. 2a. En el caso esquemático de la Fig. 2a, las subbandas 1 a 6 tienen coeficientes de ponderación iguales a 1, es decir, sin ponderación y las bandas 7 a 10 tienen coeficientes decrecientes de ponderación y las bandas 11 a 14 tienen ceros.

50 **[00036]** En la Fig. 2c se ilustra una implementación correspondiente de una cascada de un convertidor de tiempo a espectral tal como 106a y el ponderador conector subsiguiente 108. Cada subbanda 1, 2 ..., 14 es ingresada a un bloque de ponderación individual indicado por W_1, W_2, \dots, W_{14} . El ponderador 108 aplica el factor de ponderación de la tabla de la Fig. 2b a cada señal de subbanda individual mediante la multiplicación de cada muestreo de la señal de subbanda por el coeficiente de ponderación. A continuación, a la salida del ponderador, existen señales de subbanda ponderadas que luego son ingresadas al sustractor 112 de la Fig. 1^a, que ejecuta además una sustracción en el dominio espectral.

55 **[00037]** La Fig. 3 ilustra la respuesta impulsional y la respuesta en la frecuencia del filtro paso bajo de la Fig. 8 del codificador AMR-WB+. El filtro paso bajo $h_{LP}(n)$ en el dominio temporal se define en AMR-WB+ por los siguientes coeficientes.

$$a[13] = [0.088250, 0.086410, 0.081074, 0.072768, 0.062294, 0.050623, 0.038774, 0.027692, 0.018130, 0.010578, 0.005221, 0.001946, 0.000385];$$

$$65 \quad h_{LP}(n) = a[13-n] \text{ para } n \text{ de } 1 \text{ a } 12$$

$h_{LP}(n)=a(n-12)$ para n de 13 a 25

5 [00038] La respuesta impulsional y la respuesta frecuencial ilustradas en la Fig. 3 corresponden a una situación en que el filtro se aplica a la muestra de una señal en el dominio temporal de 12.8 kHz. El retardo generado es entonces un retardo de 12 muestras, es decir, 0,9375 ms.

10 [00039] El filtro ilustrado en la Fig. 3 tiene una respuesta en la frecuencia en el dominio QMF, donde cada QMF tiene una resolución de 400 Hz. 32 bandas QMF cubren el ancho de banda de la señal muestreada a 12,8 kHz. La respuesta frecuencial y el dominio QMF están ilustrados en la Fig. 4.

[00040] La amplitud de la respuesta frecuencial con una resolución de 400 Hz forma las ponderaciones empleadas al aplicar el filtro paso bajo en el dominio QMF. Las ponderaciones provistas al ponderador 108 son, respecto de los parámetros ilustrativos antes citados, las esbozadas en la Fig. 5.

15 [00041] Esas ponderaciones se pueden calcular de la siguiente manera:
 $W = \text{abs}(DFT(h_{LP}(n), 64))$, donde $DFT(x, N)$ se refiere a la Transformada Discreta de Fourier de longitud N de la señal x . Si x es más corta que N , la señal se rellena con un tamaño N de x ceros. La longitud N de la DFT corresponde al doble del número de subbandas QMF. Dado que $h_{LP}(n)$ es una señal de coeficientes reales, W exhibe una simetría hermítica y $N/2$ coeficientes de frecuencia entre la frecuencia 0 y la frecuencia de Nyquist.

20 [00042] Mediante el análisis de la respuesta en la frecuencia de los coeficientes de filtro, esto corresponde aproximadamente a una frecuencia de corte de $2 \cdot \pi \cdot 10 / 256$. Esto se utiliza para diseñar el filtro. A continuación se cuantizan los coeficientes para escribirlos en 14 bits para ahorrar cierto consumo de ROM y en vista de una implementación de punto fijo.

25 [00043] A continuación se ejecuta el filtrado en el dominio QMF de la siguiente manera:

30 $Y =$ señal post-procesada en el dominio QMF
 $X =$ señal decodificada en la señal QMF procedente del codificador de núcleo
 $E =$ ruido inter-armónico generado en TD para extraer de X

$Y(k) = X(k) - W(k) \cdot E(k)$ para k de 1 a 32

35 [00044] La Fig. 6 ilustra un ejemplo adicional en que el QMF tiene una resolución de 800 Hz, por lo que 16 bandas cubren la totalidad del ancho de banda de la señal muestreada a 12,8 kHz. Los coeficientes W son entonces los indicados en la Fig. 6 debajo del trazado. El filtrado se lleva a cabo de manera igual a la descrita con respecto a la Fig. 6, aunque k sólo abarca de 1 a 16.

40 [00045] La respuesta frecuencial del filtro en las 16 bandas QMF está representada de acuerdo con lo ilustrado en la Fig. 6.

[00046] La Fig. 10 ilustra otra intensificación del filtro de predicción a largo plazo ilustrado en 102 en la Fig. 1b.

45 [00047] En particular, en el caso de una implementación de bajo retardo, el término $\hat{s}(n+T)$ de la tercera a la última línea de la Fig. 9 es problemático. Esto se debe al hecho de que las T muestras están en el futuro con respecto al tiempo real n . Por lo tanto, para enfrentar situaciones en que, debido a la implementación de bajo retardo, aún no están disponibles los valores futuros, $\hat{s}(n+T)$ es reemplazado por \hat{s} como se indica en la Fig. 10. A continuación, el filtro de predicción a largo plazo se aproxima a la predicción a largo plazo del estado de la técnica, aunque con menos o cero retardo. Se ha encontrado que la aproximación es suficientemente buena y que la ganancia con respecto al retardo reducido es más ventajosa que la ligera pérdida de intensificación de tono.

50 [00048] Si bien se han descrito algunos aspectos en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del método correspondiente, en el cual un bloque o dispositivo corresponde a un paso del método o a una característica de un paso del método. De manera análoga, los aspectos descritos en el contexto de un paso del método también representan una descripción de un bloque o ítem correspondiente o de una característica de un aparato correspondiente.

55 [00049] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo un disco blando, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema de computación programable de tal manera que se ejecute el método respectivo.

[00050] Algunas realizaciones de acuerdo con la invención comprenden un transportador de datos no transitorio que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema de computación programable de tal manera que se ejecute uno de los métodos descritos en la presente.

5 **[00051]** En general, las realizaciones de la presente invención pueden ser implementadas en forma de producto programa de computación con un código de programa, donde el código de programa cumple la función de ejecutar uno de los métodos al ejecutarse el programa de computación en una computadora. El código de programa puede ser almacenado, por ejemplo, en un portador legible por una máquina.

10 **[00052]** Otras realizaciones comprenden el programa de computación para ejecutar uno de los métodos aquí descritos, almacenado en un portador legible por una máquina.

15 **[00053]** En otras palabras, una realización del método de la invención consiste, por lo tanto, en un programa de computación que consta de un código de programa para realizar uno de los métodos aquí descritos al ejecutarse el programa de computación en una computadora.

20 **[00054]** Otra forma de realización de los métodos de la presente invención consiste, por lo tanto, en un portador de datos (o medio de almacenamiento digital, o medio legible por computadora) que comprende, grabado en el mismo, el programa de computación para ejecutar uno de los métodos aquí descritos.

25 **[00055]** Otra forma de realización del método de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa de computación para ejecutar uno de los métodos aquí descritos. El flujo de datos o la secuencia de señales puede estar configurado, por ejemplo, para ser transferido por medio de una conexión de comunicación de datos, por ejemplo por internet.

30 **[00056]** Otra de las realizaciones comprende un medio de procesamiento, por ejemplo una computadora, un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los métodos aquí descritos.

35 **[00057]** Otra de las realizaciones comprende una computadora en la que se ha instalado el programa de computación para ejecutar uno de los métodos aquí descritos.

40 **[00058]** En algunas formas de realización, se puede utilizar un dispositivo lógico programable (por ejemplo una matriz de puertas programables por campo) para ejecutar algunas o todas las funcionalidades de los métodos aquí descritos. En algunas realizaciones, una matriz de puertas programables por campo puede cooperar con un microprocesador para ejecutar uno de los métodos aquí descritos. Por lo general, los métodos son ejecutados preferentemente por cualquier aparato de hardware.

45 **[00059]** Las realizaciones precedentemente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y detalles aquí descritos han de ser evidentes para las personas con capacitación en la técnica. Por lo tanto, sólo es intención limitarse al alcance de las siguientes reivindicaciones de patente y no a los detalles específicos presentados a manera de descripción y explicación de las realizaciones aquí presentadas.

REIVINDICACIONES

1. Aparato para procesar una señal de audio decodificada (100) que comprende:
 un filtro (102) para filtrar la señal de audio decodificada para obtener una señal de audio filtrada (104);
 el aparato estando caracterizado porque comprende:
 una etapa de conversión de tiempo a espectral (106) para convertir la señal de audio decodificada y la señal de audio filtrada en correspondientes representaciones espectrales, en donde cada representación espectral tiene una pluralidad de señales de subbanda;
 un ponderador (108) para ejecutar una ponderación selectiva en frecuencia de la representación espectral de la señal de audio filtrada mediante la multiplicación de señales de subbanda con respectivos coeficientes de ponderación para obtener una señal de audio filtrada y ponderada;
 un sustractor (112) para ejecutar un sustracción por subbandas entre la señal de audio filtrada y ponderada y la representación espectral de la señal de audio decodificada para obtener una señal de audio resultante;
 y
 un convertidor de espectral a tiempo (114) para convertir la señal de audio resultante o una señal derivada de la señal de audio resultante a una representación en el dominio temporal para obtener una señal de audio decodificada y procesada (116).
2. El aparato según la reivindicación 1, que además comprende un decodificador de intensificación de ancho de banda (129) o un decodificador mono–estéreo o mono–multicanal (131) para calcular la señal derivada de la señal de audio resultante,
 en donde el convertidor de espectral a tiempo (114) está configurado para no convertir la señal de audio resultante sino la señal derivada de la señal de audio resultante al dominio temporal de manera que todo el procesamiento por el decodificador de intensificación de ancho de banda (129) o el decodificador mono–estéreo o mono–multicanal (131) se ejecuta en el mismo dominio espectral definido por la etapa de conversión de tiempo a espectral (106).
3. El aparato según la reivindicación 1 o 2,
 en donde la señal de audio decodificada es una señal de salida decodificada por ACELP, y
 en donde el filtro (102) es un filtro de predicción a largo plazo controlado por la información de tono.
4. El aparato según cualquiera de las reivindicaciones precedentes,
 en donde el ponderador (108) está configurado para ponderar la señal de audio filtrada de manera que las subbandas de frecuencia inferior estén menos atenuadas o no atenuadas que las subbandas de frecuencia superior de modo que la ponderación selectiva en frecuencia imprime una característica de paso bajo a la señal de audio filtrada.
5. El aparato según cualquiera de las reivindicaciones precedentes,
 en donde la etapa de conversión de tiempo a espectral (106) y el convertidor de espectral a tiempo (114) están configurados para implementar un banco de filtros de análisis QMF y un banco de filtros de síntesis QMF, respectivamente.
6. El aparato según cualquiera de las reivindicaciones precedentes,
 en el cual el sustractor (112) está configurado para sustraer una señal de subbanda de la señal de audio filtrada y ponderada de la correspondiente señal de subbanda de la señal de audio para obtener una subbanda de la señal de audio resultante, en donde las subbandas pertenecen al mismo canal de banco de filtros.
7. El aparato según cualquiera de las reivindicaciones precedentes,
 en donde el filtro (102) está configurado para ejecutar una combinación ponderada de la señal de audio y por lo menos la señal de audio desplazada en el tiempo por un período de tono.
8. El aparato según la reivindicación 7,
 en donde el filtro (102) está configurado para ejecutar la combinación ponderada sólo combinando la señal de audio y la señal de audio existente en instantes anteriores.
9. El aparato según cualquiera de las reivindicaciones precedentes,
 en donde el convertidor de espectral a tiempo (114) tiene un número diferente de canales de entrada con respecto a la etapa de conversión de tiempo a espectral (106) de manera que se obtiene una conversión de velocidad de muestreo, en donde se obtiene un aumento del número de muestras, cuando el número de canales de entrada al convertidor de espectral a tiempo es mayor que el número de canales de salida de la etapa de conversión de tiempo a espectral y en donde se ejecuta una reducción del número de muestras, cuando el número de canales de entrada al convertidor de espectral a tiempo es menor que el número de canales de salida de la etapa de conversión de tiempo a espectral.
10. El aparato según cualquiera de las reivindicaciones precedentes, que además comprende:

- un primer decodificador (120) para producir la señal de audio decodificada en una primera porción de tiempo;
 un segundo decodificador (122) para producir otra señal de audio decodificada en una segunda porción de tiempo diferente;
- 5 una primera rama de procesamiento conectada al primer decodificador (120) y al segundo decodificador (122);
 una segunda rama de procesamiento conectada al primer decodificador (120) y al segundo decodificador (122),
 en donde la segunda rama de procesamiento comprende el filtro (102) y el ponderador (108) y comprende
 10 asimismo una etapa de ganancia controlable (129) y un controlador (130), en donde el controlador (130) está configurado para ajustar una ganancia de la etapa de ganancia (129) a un primer valor para la primera porción de tiempo y a un segundo valor o a cero para la segunda porción de tiempo, que es menor que el primer valor.
- 15 11. El aparato según cualquiera de las reivindicaciones precedentes, que además comprende un rastreador de tono para obtener un retardo de tono y para ajustar el filtro (102) sobre la base del retardo de tono como la información de tono.
- 20 12. El aparato según la reivindicación 10 o 11, en donde el primer decodificador (120) está configurado para suministrar la información de tono o una parte de la información de tono para ajustar el filtro (102).
13. El aparato según la reivindicación 10, 11 o 12, en donde una salida de la primera rama de procesamiento y una salida de la segunda rama de procesamiento están conectadas a las entradas del sustractor (112).
- 25 14. El aparato según cualquiera de las reivindicaciones precedentes, en donde la señal de audio decodificada es proporcionada por un decodificador ACELP (120) incluido en el aparato, y en donde el aparato comprende además un decodificador adicional (122) implementado en forma de decodificador TCX.
- 30 15. Método para procesar una señal de audio decodificada (100), que comprende:
 filtrar (102) la señal de audio decodificada para obtener una señal de audio filtrada (104);
 el método estando caracterizado porque comprende:
 35 convertir (106) la señal de audio decodificada y la señal de audio filtrada en sendas representaciones espectrales, en donde cada representación espectral tiene una pluralidad de señales de subbanda;
 ejecutar (108) una ponderación selectiva en frecuencia de la señal de audio filtrada mediante la multiplicación de señales de subbanda con respectivos coeficientes de ponderación para obtener una señal de audio filtrada y ponderada;
 40 ejecutar (112) una sustracción por subbandas entre la señal de audio filtrada y ponderada y la representación espectral de la señal de audio decodificada para obtener una señal de audio resultante; y
 convertir (114) la señal de audio resultante o una señal derivada de la señal de audio resultante en una representación en el dominio temporal a fin de obtener una señal de audio decodificada y procesada (116).
- 45 16. Programa de computación que contiene un código para implementar, al ejecutarse en una computadora, el método para procesar una señal de audio decodificada según la reivindicación 15.

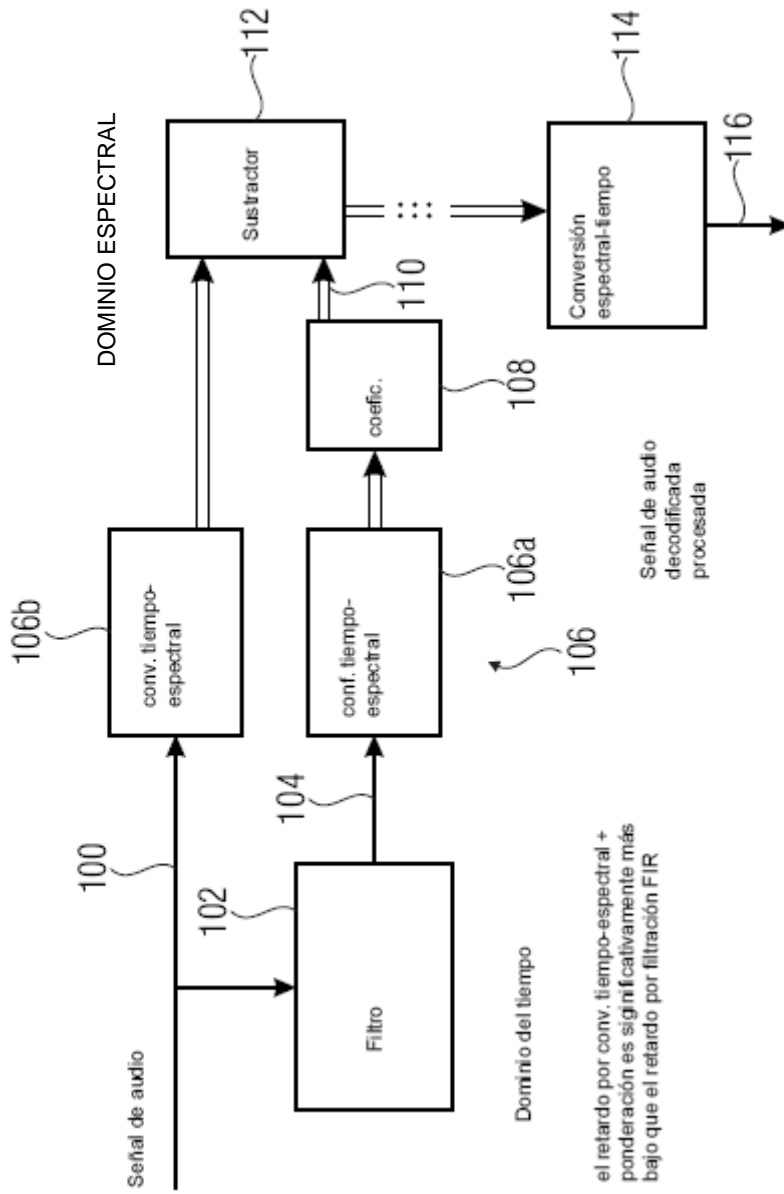


FIG 1A

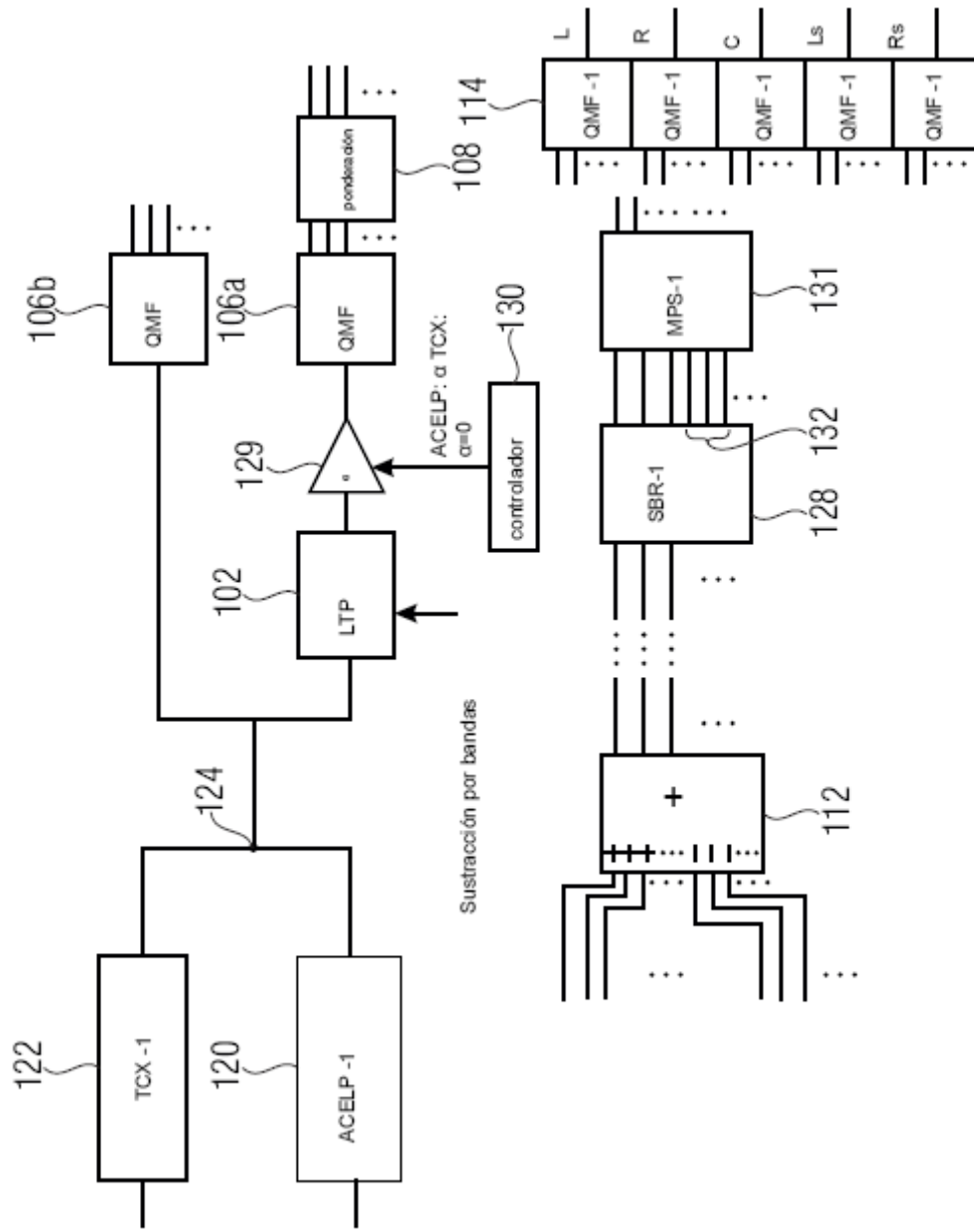


FIG 1B

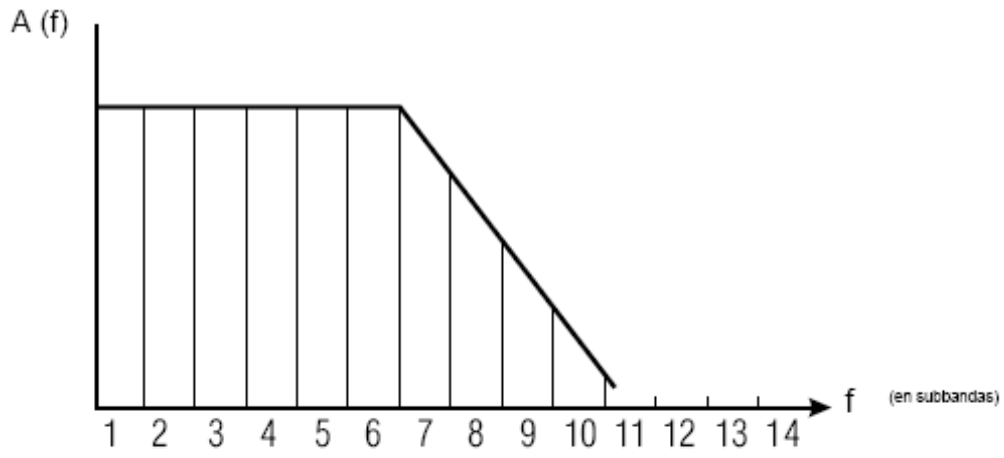


FIG 2A

subbanda N°	coef. de ponderación
1	1
⋮	⋮
6	1
7	0.8
8	0.6
9	0.4
10	0.2
11	0
12	0
13	0
14	0

FIG 2B

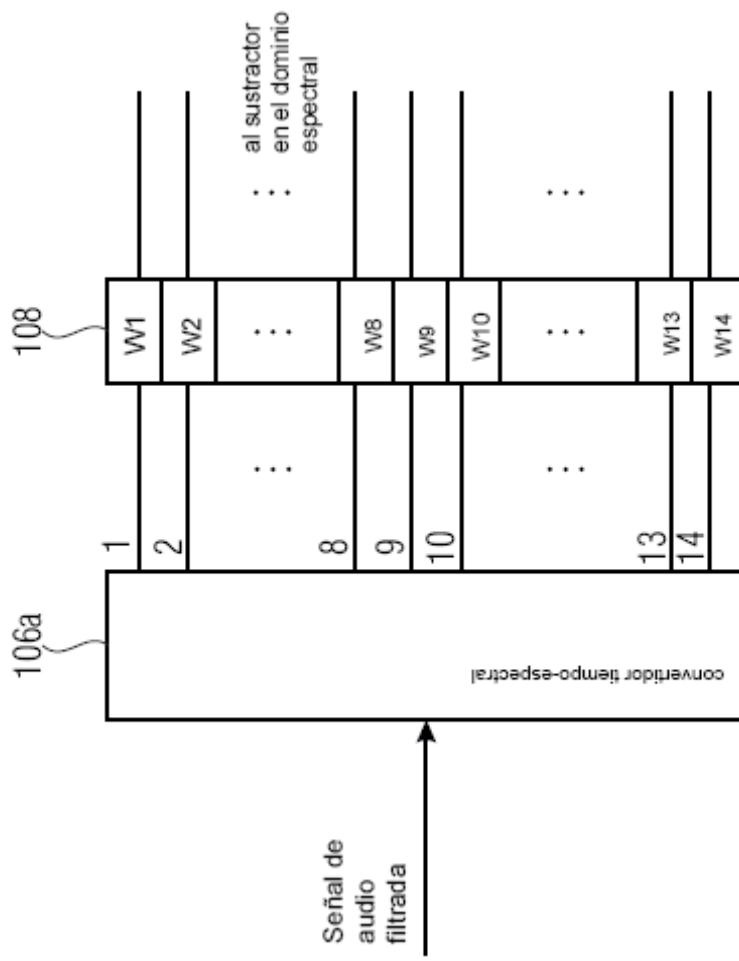
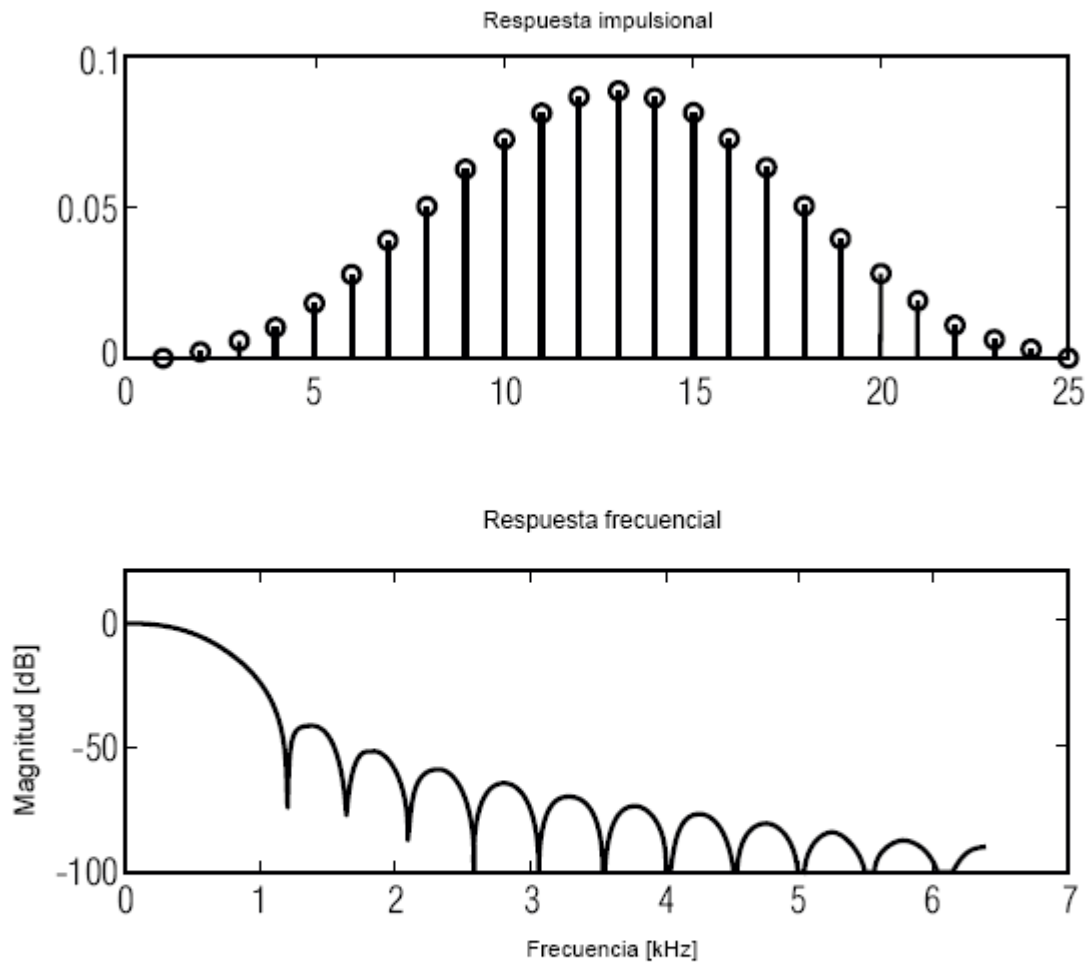


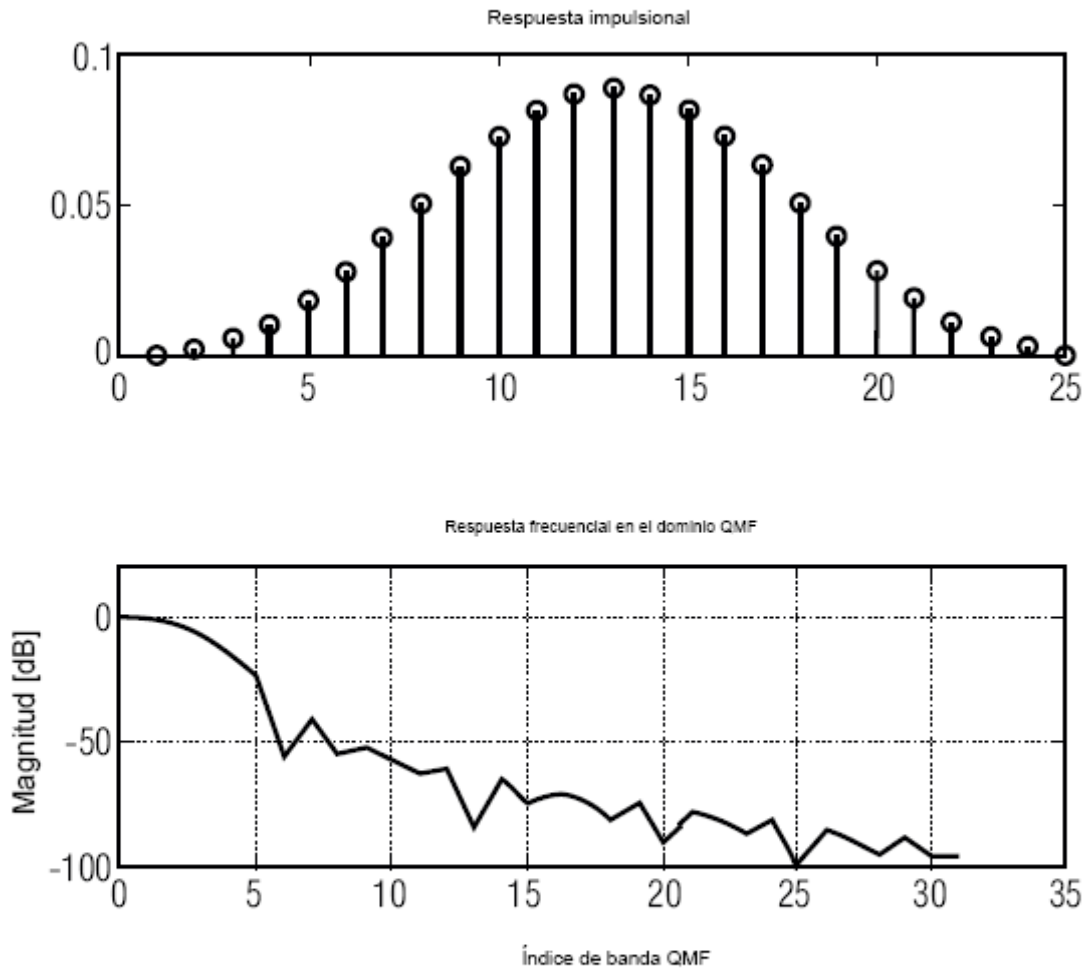
FIG 2C



Respuesta impulsional y frecuencial del filtro de paso bajo en Ar-WB+

FIG 3

Frecuencia de muestreo:
 Resolución de banda
 QMF 32 bandas QMF



Respuesta impulsional y respuesta frecuencial en el dominio QMF

FIG 4

$$W [32] = \{ 1.000040f, 0.917218f, 0.702561f, 0.438018f, 0.208838f, 0.063451f, 0.001553f, \\ 0.008564f, 0.001777f, 0.002472f, 0.001370f, 0.000718f, \\ 0.000926f, 0.000060f, 0.000554f, 0.000182f, 0.000270f, 0.000229f, 0.000080f, 0.000191f, \\ 0.000030f, 0.000122f, 0.000082f, 0.000045f, 0.000081f, \\ 0.000010f, 0.000052f, 0.000035f, 0.000017f, 0.000037f, 0.000016f, 0.000015f \};$$

Ponderaciones correspondientes a 31
subbandas de QMF

FIG 5

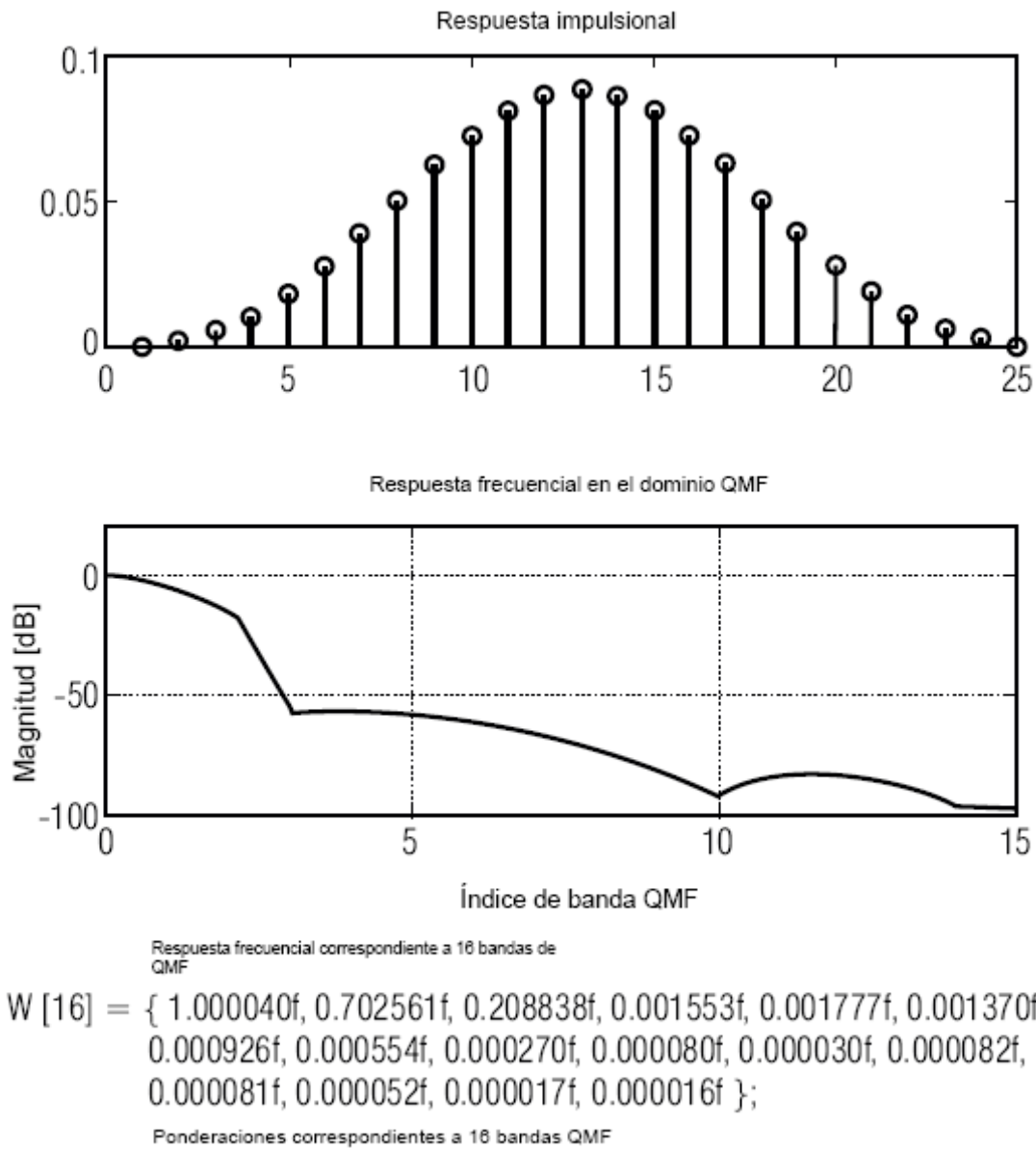


FIG 6

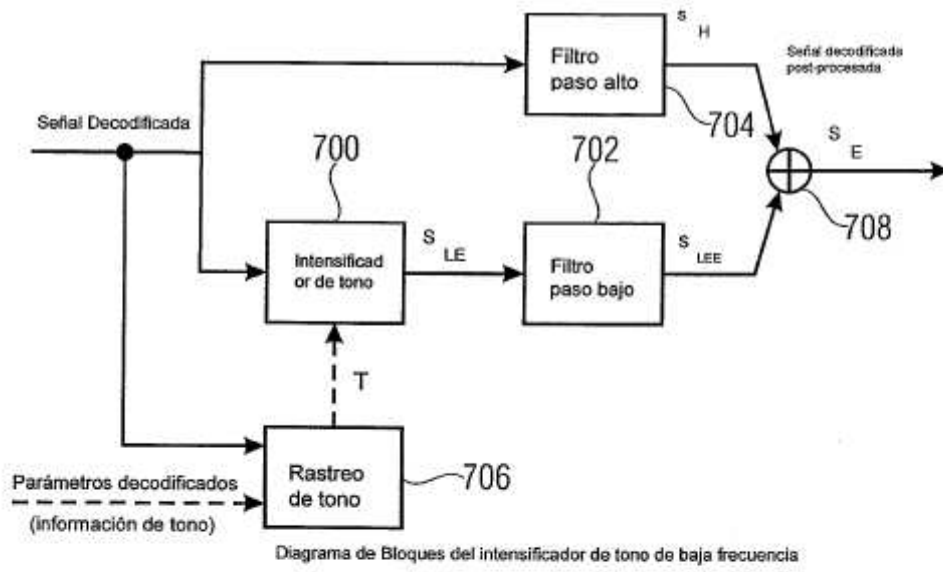


FIG 7

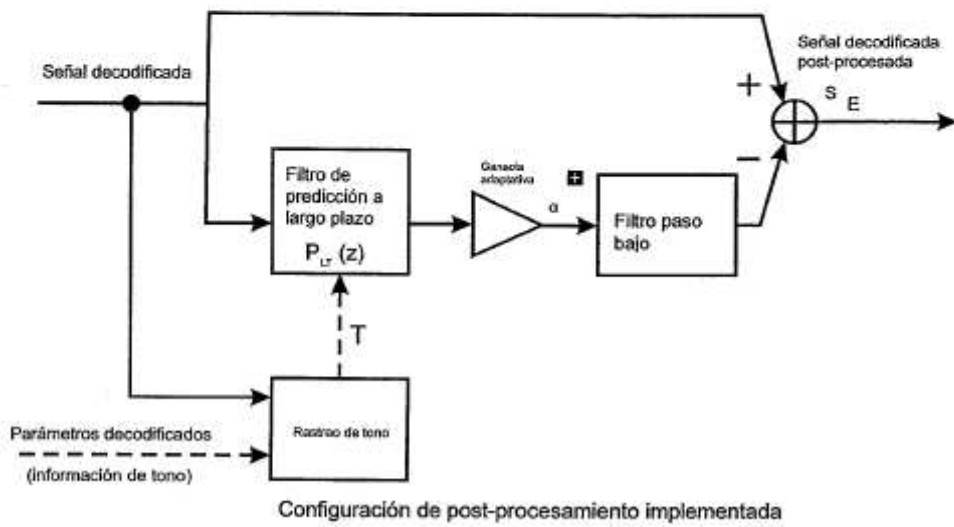


FIG 8

$$H_E(z) = (1-\alpha) + \frac{\alpha}{2} z^T + \frac{\alpha}{2} z^{-T}$$

$$s_{LE}(n) = (1-\alpha)\hat{s}(n) + \frac{\alpha}{2}\hat{s}(n-T) + \frac{\alpha}{2}\hat{s}(n+T)$$

$$s_E(n) = \hat{s}(n) * h_{HP}(n) + s_{LE}(n) * h_{LP}(n)$$

$$= \hat{s}(n) * h_{HP}(n) + \left((1-\alpha)\hat{s}(n) + \frac{\alpha}{2}\hat{s}(n-T) + \frac{\alpha}{2}\hat{s}(n+T) \right) * h_{LP}(n)$$

$$= \hat{s}(n) * h_{HP}(n) + \hat{s}(n) * h_{LP}(n) - \left(\alpha\hat{s}(n) - \frac{\alpha}{2}\hat{s}(n-T) - \frac{\alpha}{2}\hat{s}(n+T) \right) * h_{LP}(n)$$

$$= \hat{s}(n) - \alpha \left(\hat{s}(n) - \frac{1}{2}\hat{s}(n-T) - \frac{1}{2}\hat{s}(n+T) \right) * h_{LP}(n)$$

$$= \hat{s}(n) - \alpha e_{LT}(n) * h_{LP}(n)$$

$$P_{LT}(z) = 1 - 0.5 z^T - 0.5 z^{-T}$$

FIG 9

implementación con bajo retardo del filtro
de predicción a largo plazo:

$\hat{s}(n+T)$ es reemplazado por \hat{s}

$$s_E(n) = \hat{s}(n) - \alpha \left[\hat{s}(n) - \frac{1}{2}\hat{s}(n-T) - \frac{1}{2}\hat{s}(n) \right]$$

$$= \hat{s}(n) - \frac{\alpha}{2} [\hat{s}(n) - \hat{s}(n-T)]$$

$$P_{LT}(z) = 0.5 - 0.5 z^T$$

FIG 10