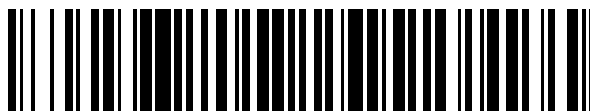


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 548 010**

51 Int. Cl.:

G10L 19/08 (2013.01)
G10L 19/04 (2013.01)
G10L 19/12 (2013.01)
G10L 19/02 (2013.01)
G10L 21/00 (2013.01)
G10L 19/26 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.02.2008 E 08712799 (9)**

97 Fecha y número de publicación de la concesión europea: **22.07.2015 EP 2132731**

54 Título: **Procedimiento y dispositivo para suavizar ruido de fondo estacionario**

30 Prioridad:

05.03.2007 US 892994 P

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

13.10.2015

73 Titular/es:

TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)
(100.0%)
164 83 Stockholm, SE

72 Inventor/es:

BRUHN, STEFAN

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 548 010 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 dispositivo para suavizar ruido de fondo estacionario

Sector técnico

5 La presente invención se refiere a la codificación de voz en sistemas de telecomunicaciones en general, especialmente a procedimientos y dispositivos para suavizar ruido de fondo estacionario en dichos sistemas.

Antecedentes

10 La codificación de la voz es el proceso de obtención de una representación compacta de señales de voz para su transmisión eficiente sobre canales cableados e inalámbricos de banda limitada, y/o para su almacenamiento. Actualmente, los codificadores de voz se han convertido en componentes esenciales en las telecomunicaciones y en la infraestructura multimedia. Los sistemas comerciales que se basan en una codificación eficiente de la voz incluyen comunicación celular, voz sobre protocolo de internet (VOIP, voice over internet protocol), videoconferencia, juguetes electrónicos, archivo, y voz y datos digitales simultáneos (DSVD, simultaneous voice and data), así como numerosos juegos basados en PC y aplicaciones multimedia.

15 Al ser una señal continua en el tiempo, la voz se puede representar digitalmente por medio de un proceso de muestreo y cuantificación. Las muestras de voz se cuantifican habitualmente utilizando cuantificación de 16 bits o bien de 8 bits. Como muchas otras señales, una señal de voz comprende una gran cantidad de información que es redundante (información mutua diferente de cero entre muestras sucesivas en la señal) o bien irrelevante perceptualmente (información que no es percibida por los oyentes humanos). La mayoría de los codificadores de telecomunicaciones tienen pérdidas, lo que significa que la voz sintetizada es perceptualmente similar a la original pero puede ser diferente físicamente.

20 Un codificador de voz transforma una señal de voz digitalizada en una representación codificada, que normalmente se transmite en tramas. Correspondientemente, un descodificador de voz recibe tramas codificadas y sintetiza la voz reconstruida. Muchos codificadores de voz modernos pertenecen a una gran clase de codificadores de voz conocidos como LPC (codificadores predictivos lineales). Unos pocos ejemplos de dichos codificadores son los códec de voz 3GPP FR, EFR, AMR y AMR-WB, los códec de voz 3GPP2 EVRC, SMV y EVRC-WB, y varios códec ITU-T, tales como G.728, G723, G.729, etc.

Todos estos codificadores utilizan un concepto de filtro de síntesis en el proceso de generación de la señal. El filtro se utiliza para modelizar el espectro de tiempo reducido de la señal que se debe reproducir, mientras que se asume que la entrada al filtro trata todas las demás variaciones de la señal.

30 Una característica común de estos modelos de filtro de síntesis es que la señal que se debe reproducir se representa mediante parámetros que definen el filtro de síntesis. El término "predictivo lineal" se refiere a una clase de procedimientos utilizados a menudo para estimar los parámetros del filtro. En los codificadores basados en LPC, la señal de voz se ve como la salida de un sistema lineal invariante en el tiempo (LTI, linear time-invariant) cuya entrada es la señal de excitación al filtro. Por lo tanto, la señal que se debe reproducir se representa parcialmente por un conjunto de parámetros del filtro y parcialmente por la señal de excitación que activa el filtro. La ventaja de dicho concepto de codificación se deriva del hecho de que tanto el filtro la señal de excitación de activación se pueden describir eficientemente con relativamente pocos bits.

35 Una clase particular de códec basados en LPC, se basa en el denominado principio de análisis por síntesis (AbS). Estos códec incorporan una copia local del descodificador en el codificador, y encuentran la señal de excitación de activación del filtro de síntesis seleccionando aquella señal de excitación entre un conjunto de señales de excitación candidatas, que maximiza la similitud de la señal de salida con la señal de voz original.

40 El concepto de utilizar dicha codificación predictiva lineal y particularmente codificación AbS ha demostrado que funciona relativamente bien para señales de voz, incluso a tasas de bits de, por ejemplo, 4 a 12 kbps. Sin embargo, cuando el usuario de un teléfono móvil que utiliza dicha técnica de codificación está en silencio y la señal de entrada comprende los sonidos circundantes, por ejemplo ruido, los codificadores actualmente conocidos tienen dificultades para tratar con esta situación, dado que están optimizados para señales de voz. Un oyente en el lado de recepción se puede incomodar cuando hay sonidos de fondo familiares que no se pueden reconocer debido a que han sido "maltratados" por el codificador.

45 La denominada turbulencia provoca una de las degradaciones de calidad más importantes en los sonidos de fondo reproducidos. Éste es un fenómeno que se produce en sonidos de ruido de fondo relativamente estacionarios, tales como ruido de coches, y está provocado por fluctuaciones temporales no naturales de la potencia y del espectro de la señal descodificada. A su vez, estas fluctuaciones están provocadas por una estimación y una cuantificación inadecuadas de los coeficientes del filtro de síntesis y su señal de excitación. Normalmente, la turbulencia disminuye cuando aumenta la tasa de bits del códec.

La turbulencia se ha identificado como un problema en la técnica anterior, y se han propuesto múltiples soluciones a la misma en la bibliografía. Una de las soluciones propuestas se describe en la patente U.S.A. 5632004 [1]. Según esta patente, durante la inactividad de la voz, se modifican los parámetros del filtro por medio de filtrado de paso bajo o expansión del ancho de banda, de tal modo que se reducen las variaciones espectrales del sonido de fondo sintetizado. Este procedimiento se ha perfeccionado en la patente U.S.A. 5579432 [2], de manera que la técnica anti-turbulencia descrita se aplica solamente sobre la parte estacionaria detectada del ruido de fondo.

Otro procedimiento que trata el problema de la turbulencia se describe en la patente U.S.A. 5487087 [3]. Este procedimiento utiliza un esquema modificado de cuantificación de señales, que se adapta tanto a la propia señal como a sus variaciones temporales. Específicamente, se contempla utilizar dicho cuantificador de fluctuación reducida para parámetros del filtro LPC y parámetros de ganancia de señal durante periodos de inactividad de la voz.

Las degradaciones de la calidad de la señal provocadas por fluctuaciones de potencia no deseadas de la señal sintetizada son tratadas por otro conjunto de procedimientos. Uno de estos es el que se describe en la patente U.S.A. 6275798 [4], y asimismo una parte del algoritmo del códec de voz AMR descrito en el documento 3GPP TS 26.090 [5]. Según estos, la ganancia de por lo menos un componente de la señal de excitación del filtro sintetizada, la contribución fija del libro de código, se suaviza de manera adaptativa en función del carácter estacionario del espectro a corto plazo LPC. Este procedimiento ha evolucionado en la patente EP 1096476 [6] y en la solicitud de patente EP 1688920 [7], donde el suavizado involucra además una limitación de la ganancia que se debe utilizar en la síntesis de la señal. Se describe un procedimiento relacionado para su utilización en codificadores de voz LPC en el documento US 5953697 [8]. Según éste, la ganancia de la señal de excitación del filtro de síntesis se controla de tal modo que la amplitud máxima de la voz sintetizada alcanza exactamente la envolvente de la forma de onda de voz de entrada.

Otra clase más de procedimientos que tratan el problema de la turbulencia funcionan como un posprocesador después del descodificador de voz. El documento EP 0665530 [9] describe un procedimiento que, durante inactividad de voz detectada, sustituye una parte de la señal de salida del descodificador de voz por un ruido blanco filtrado en paso bajo o señal de ruido cómoda. Se adoptan enfoques similares en varias publicaciones que dan a conocer procedimientos relacionados que sustituyen parte de la señal de salida del descodificador de voz con ruido filtrado.

La codificación escalable o incorporada, haciendo referencia a la figura 1, es un parámetro de codificación en el que la codificación se realiza en capas. Una capa base o central codifica la señal a una tasa de bits baja, mientras que las capas adicionales, una sobre otra, proporcionan alguna mejora con respecto a la codificación, que se consigue con todas las capas desde la central hasta la respectiva capa anterior. Cada capa aumenta en alguna medida la tasa de bits. El flujo de bits generado es incorporado, lo que significa que el flujo de bits de la codificación de las capas inferiores se incorpora en los flujos de bits de las capas superiores. Esta propiedad hace posible, en cualquier momento en la transmisión o en el receptor, desechar los bits pertenecientes a las capas superiores. Dicho flujo de bits básico puede seguir siendo descodificado hasta la capa cuyos bits se han retenido.

El algoritmo de compresión de voz escalable más común actualmente es el códec PCM logarítmico de ley A/U G.711 de 64 kbps. El códec G.711 muestreado a 8 kHz transforma muestras PCM lineales de 12 bits o 13 bits en muestras logarítmicas de 8 bits. La representación de bits ordenados de las muestras logarítmicas permite extraer los bits menos significativos (LSBs, Least Significant Bits) en un flujo de bits G.711, haciendo el codificador G.711 prácticamente escalable en SNR entre 48, 56 y 64 kbps. La propiedad de escalabilidad del códec G.711 se utiliza en las redes de comunicación de conmutación de circuitos con propósitos de señalización de control en banda. Un ejemplo reciente de utilización de esta propiedad de escalamiento de G.711 es el protocolo 3GPP TFO, que permite el establecimiento y transporte de voz de banda ancha sobre conexiones PCM de 64 kbps heredadas. Ocho kbps del flujo original G.711 de 64 kbps se utilizan inicialmente para permitir un establecimiento de llamada del servicio de voz de banda ancha sin afectar considerablemente a la calidad del servicio de banda estrecha. Después del establecimiento de llamada, la voz de banda ancha utilizará 16 kbps del flujo de G.711 de 64 kbps. Otros estándares de codificación de voz más antiguos que soportan escalabilidad en circuito abierto son G.727 (ADPCM incorporado) y en cierta medida G.722 (ADPCM de sub-banda).

Un avance más reciente en la tecnología de codificación de voz escalable es el estándar MPEG-4 que proporciona extensiones de escalabilidad para MPEG4-CELP. La capa base MPE se puede mejorar mediante la transmisión de información adicional de parámetros de filtro o información adicional de parámetros de innovación. El Sector de Estandarización de la Unión Internacional de Telecomunicaciones, ITU-T, ha acabado recientemente la estandarización de un nuevo códec escalable G.729.1, apodado G.729.EV. El intervalo de tasas de bits de este códec de voz escalable es desde 8 kbps hasta 32 kbps. El caso principal de utilización para este códec es para permitir la compartición eficiente de un recurso de ancho de banda limitado en pasarelas domésticas o de oficinas, por ejemplo un enlace ascendente compartido xDSL de 64/128 kbps entre varias llamadas VOIP.

Una tendencia reciente en la codificación de voz escalable es dotar a las capas superiores de soporte para la codificación de las señales de audio no de voz, tales como música. En dichos códec, las capas inferiores utilizan simplemente codificación de voz convencional, por ejemplo según el paradigma de análisis por síntesis del que

CELP es un ejemplo destacado. Dado que dicha codificación es muy adecuada para voz solamente, pero no tanto para señales de audio no de voz, tales como música, las capas superiores funcionan según un paradigma de codificación que se utiliza en códecs de audio. En este caso, habitualmente la codificación de las capas superiores trabaja sobre el error de codificación, de la codificación de las capas inferiores.

5 Otro procedimiento relevante en relación con los códecs de voz es la denominada compensación de inclinación espectral, que se realiza en el contexto del filtrado posterior adaptativo de voz descodificada. El problema resuelto por éste es compensar la inclinación espectral introducida mediante filtros posteriores por formantes o de corto plazo. Dichas técnicas son parte, por ejemplo, del códec AMR y del códec SMV, y están dirigidas principalmente al rendimiento del códec con voz antes que a su rendimiento con ruido de fondo. El códec SMV aplica esta
10 compensación de inclinación en el dominio residual ponderado antes del filtrado de síntesis, aunque no en respuesta a un análisis LPC de los residuos.

El problema de los procedimientos de US 5632004, US 5579432 y US 5487087 descritos anteriormente es que asumen que la excitación del filtro de síntesis LPC tiene un espectro blanco (es decir, plano) y que todas las fluctuaciones espectrales que provocan el problema de la turbulencia están relacionadas con las fluctuaciones de los
15 espectros del filtro de síntesis LPC. Sin embargo, no es éste el caso, y sobre todo si la señal de excitación está cuantificada solamente de manera tosca. En este caso, las fluctuaciones espectrales de la señal de excitación tienen un efecto similar a las fluctuaciones del filtro LPC y, por lo tanto, es necesario evitarlas.

El problema de los procedimientos que tratan las fluctuaciones de potencia no deseadas de la señal sintetizada es que están tratando solamente una parte del problema de la turbulencia, pero no proporcionan una solución
20 relacionada con las fluctuaciones espectrales. Las simulaciones muestran que incluso en combinación con los procedimientos citados que tratan las fluctuaciones espectrales, todavía no se pueden evitar todas las degradaciones de la calidad de la señal relacionadas con la turbulencia durante sonidos de fondo estacionarios.

Un problema de los procedimientos que funcionan como un posprocesador después del descodificador de voz es que sustituyen solamente una parte de la señal de salida descodificada de voz con una señal de ruido suavizada. De
25 este modo, el problema de la turbulencia no está resuelto en la parte de señal restante que se origina en el descodificador de voz y, por lo tanto, la señal de salida final no está conformada utilizando el mismo filtro de síntesis LPC que la señal de salida del descodificador de voz. Esto puede conducir a posibles discontinuidades del sonido, especialmente durante transiciones de inactividad a voz activa. Además, dichos procedimientos de posprocesamiento son desfavorables, debido a que requieren una complejidad computacional relativamente
30 elevada.

Ninguno de los procedimientos anteriores proporciona una solución al problema de que una de las razones de la turbulencia reside en las fluctuaciones espectrales de la señal de excitación del filtro de síntesis LPC. Este problema se agrava especialmente si la señal de excitación está representada con muy pocos bits, cual es el caso habitualmente para códecs de voz que funcionan a tasas de bits de 12 kbps o menores.

35 El artículo "A post-processing technique to improve coding of CELP under background noise", Codificación de voz, 2000. Proceedings. 2000 IEEE Workshop, 2000, páginas 102 a 104, de Murashima y otros, da a conocer un procedimiento de posprocesamiento para mejorar la calidad de codificación de CELP bajo ruido de fondo. Éste suaviza adaptativamente tanto la envolvente espectral como la energía de la señal de excitación estimada para reducir sus fluctuaciones temporales, que pueden provocar la degradación perceptual. La señal de excitación se
40 calcula utilizando la señal sintetizada y los parámetros espectrales proporcionados por el descodificador. El suavizado se aplica solamente en periodos no de voz y la intensidad de suavización se controla en función de las características de la señal sintetizada, para evitar la degradación en periodos de voz y de ruido no estacionario.

Sin embargo, existe la necesidad de procedimientos y dispositivos para mitigar los problemas descritos anteriormente con la turbulencia provocada por ruido de fondo estacionario durante periodos de inactividad de voz.

45 **Compendio**

Un objetivo de la presente invención es proporcionar una calidad mejorada de las señales de voz en un sistema de telecomunicación.

Otro objetivo es proporcionar una calidad incrementada de una señal de salida de descodificador de voz durante periodos de inactividad de voz con ruido de fondo estacionario.

50 La presente invención da a conocer procedimientos y dispositivos de suavizado del ruido de fondo en una sesión de voz de telecomunicaciones. Básicamente, el procedimiento acorde con la invención comprende las etapas de recibir y descodificar S10 una señal representativa de una sesión de voz, comprendiendo dicha señal tanto un componente de voz como un componente de ruido de fondo. A continuación, determinar parámetros LPC S20 y una señal de excitación S30 para la señal recibida. Después, sintetizar y emitir (S40) una señal de salida en base a los
55 parámetros LPC y la señal de excitación determinados. Además, anteriormente a la etapa de síntesis, modificar S35 la señal de excitación determinada reduciendo las fluctuaciones de potencia y espectrales de la señal de excitación para proporcionar una señal de salida suavizada. El procedimiento está caracterizado por una etapa S25 de

modificar el conjunto determinado de parámetros LPC proporcionando un conjunto filtrado en paso bajo de parámetros LPC, y determinar una combinación ponderada del conjunto filtrado en paso bajo y del conjunto determinado de parámetros LPC, y llevar a cabo la etapa de síntesis y emisión en base al conjunto modificado de parámetros LPC para proporcionar una señal de salida suavizada.

- 5 Las ventajas de la presente invención comprenden:
- permitir una señal de salida de descodificador de voz mejorada;
 - permitir una señal de salida de descodificador de voz suave.

Breve descripción de los dibujos

- 10 La invención, junto con objetivos y ventajas adicionales de la misma, se comprenderá mejor haciendo referencia a la siguiente descripción, tomada junto con los dibujos adjuntos, en los cuales:
- la figura 1 es un diagrama de bloques de un códec de voz y audio escalable;
 - la figura 2 es un diagrama de flujo que muestra una realización de un procedimiento acorde con la presente invención;
 - la figura 3 es un diagrama de flujo de otra realización de un procedimiento acorde con la presente invención;
 - 15 la figura 4 es un diagrama de bloques que muestra realizaciones de un procedimiento acorde con la presente invención;
 - la figura 5 es una ilustración de una realización de un dispositivo acorde con la presente invención.

Abreviaturas

- | | | |
|----|---------|--|
| | AbS | Análisis por síntesis |
| 20 | ADPCM | PCM diferencial adaptativo |
| | AMR-WB | Banda ancha multi-tasa adaptativa |
| | EVRC-WB | Códec de banda ancha de tasa variable mejorada |
| | CELP | Predicción lineal con excitación por código |
| | ISP | Par espectral de inmitancia |
| 25 | ITU-T | Unión internacional de telecomunicaciones |
| | LPC | Codificadores predictivos lineales |
| | LSF | Frecuencia espectral de línea |
| | MPEG | Grupo de expertos en imágenes en movimiento |
| | PCM | Modulación por impulsos codificados |
| 30 | SMV | Codificador de voz con modos seleccionables |
| | VAD | Detector de actividad de voz |

Descripción detallada

- 35 La presente invención se describirá en el contexto de una sesión de voz, por ejemplo una llamada de teléfono, en un sistema de telecomunicación general. Habitualmente, los procedimientos y los dispositivos se implementarán en un descodificador adecuado para síntesis de voz. Sin embargo, es igualmente posible que los procedimientos y dispositivos se implementen en un nodo intermedio en la red y se transmitan a continuación a un usuario objetivo. El sistema de telecomunicación puede ser tanto inalámbrico como cableado.

- 40 Por consiguiente, la presente invención facilita procedimientos y dispositivos para mitigar los problemas conocidos descritos anteriormente con la turbulencia provocada por ruido de fondo estacionario durante periodos de inactividad de voz en una sesión telefónica de voz. Específicamente, la presente invención permite mejorar la calidad de una señal de salida de descodificador de voz durante periodos de inactividad de voz con ruido de fondo estacionario.

Dentro de esta descripción, el término sesión de voz se interpreta como cualquier intercambio de señales vocales sobre un sistema de telecomunicación. Por consiguiente, una señal de sesión de voz se puede describir como

comprendiendo una parte activa y una parte de fondo. La parte activa es la propia señal de voz de la sesión. La parte de fondo es el ruido circundante en la posición del usuario, denominado asimismo ruido de fondo. Un período de inactividad se define como un periodo de tiempo dentro de una sesión de voz en la que no hay parte activa, solamente una parte de fondo, por ejemplo la parte de voz de la sesión está inactiva.

- 5 De acuerdo con una realización básica, la presente invención permite mejorar la calidad una sesión de voz reduciendo las variaciones de potencia y las fluctuaciones espectrales de la señal de excitación de filtro de síntesis LPC durante la detección de periodos de inactividad de voz.

De acuerdo con otra realización, la señal de salida se mejora adicionalmente combinando la modificación de la señal de excitación con una operación de suavizado de los parámetros LPC.

- 10 Haciendo referencia al diagrama de flujo de la figura 2, una realización de un procedimiento acorde con la presente invención comprende recibir y decodificar S10 una señal representativa de una sesión de voz (es decir, que comprende una componente de voz en forma de una señal de voz activa y/o un componente de ruido de fondo estacionario). A continuación, se determina S20 un conjunto de parámetros LPC para la señal recibida. Además, se determina S30 una señal de excitación para la señal recibida. Una señal de salida es sintetizada y emitida S40 en base a los parámetros LPC determinados y la señal de excitación determinada. De acuerdo con la presente invención, la señal de excitación es mejorada o modificada S35 reduciendo las fluctuaciones de potencia y espectrales de la señal de excitación para proporcionar una señal de salida suavizada.

- 20 Haciendo referencia al diagrama de flujo de la figura 3, se describirá otra realización de un procedimiento acorde con la presente invención. Las etapas correspondientes mantienen los mismos números de referencia que los de la figura 2. Además de la etapa de modificación de la señal de excitación de la realización descrita anteriormente, también el conjunto determinado de parámetros LPC se somete a una operación de modificación S25, por ejemplo, suavizado de los parámetros LPC.

- 25 El suavizado S25 de los parámetros LPC según otra realización de la presente invención, haciendo referencia a la figura 4, comprende llevar a cabo el suavizado de parámetros LPC de tal modo que el grado de suavizado está controlado por algún factor β , que a su vez se obtiene a partir de un parámetro que se denomina factor del nivel de ruido.

En una primera etapa, se calcula S20 un conjunto filtrado en paso bajo de parámetros LPC. Preferentemente, esto se realiza mediante un filtrado autorregresivo de primer orden, según:

$$\tilde{a}(n) = \lambda \cdot \tilde{a}(n-1) + (1 - \lambda) \cdot a(n) \quad (1)$$

- 30 En este caso, $\tilde{a}(n)$ representa el vector de parámetros LPC filtrados en paso bajo, obtenido para una trama actual n , $a(n)$ es el vector de parámetros LPC descodificados para la trama n , y λ es un factor de ponderación que controla el grado de suavizado. Una elección adecuada para λ es 0,9.

En una segunda etapa S25, se calcula una combinación ponderada del vector $\tilde{a}(n)$ de parámetros LPC filtrados en paso bajo y el vector $a(n)$ de parámetros LPC descodificados, utilizando el factor de control de suavizado β , según:

$$\hat{a}(n) = (1 - \beta) \cdot \tilde{a}(n) + \beta \cdot a(n) \quad (2)$$

Los parámetros LPC pueden ser cualquier representación adecuada para filtrado e interpolación, y están representados preferentemente como frecuencias espectrales de línea (LSFs) o pares espectrales de inmitancia (ISPs).

- 40 Habitualmente, el descodificador de voz puede interpolar los parámetros LPC a través de subtramas en las que, preferentemente, se interpolan asimismo correspondientemente los parámetros LPC filtrados en paso bajo. En una realización particular, el descodificador de voz trabaja con tramas de 20 ms de duración y 4 subtramas de 5 ms cada una dentro de una trama. Si el descodificador de voz calcula originalmente los 4 vectores de parámetros LPC de subtrama interpolando entre un vector $a(n-1)$ de parámetros LPC de trama final de la trama anterior, un vector $a_m(n)$ de parámetros LPC de trama intermedia y un vector $a(n)$ de parámetros LPC de trama final de la trama presente, entonces la combinación ponderada de los vectores de parámetros LPC filtrados de paso bajo y los vectores de parámetros LPC descodificados se calcula como sigue:

$$\hat{a}(n-1) = (1 - \beta) \cdot \tilde{a}(n-1) + \beta \cdot a(n-1) \quad (3)$$

$$\hat{a}_m(n-1) = (1 - \beta) \cdot 0.5 \cdot (\tilde{a}(n-1) + \tilde{a}(n)) + \beta \cdot a_m(n-1) \quad (4)$$

$$\hat{a}(n) = (1 - \beta) \cdot \tilde{a}(n) + \beta \cdot a(n) \quad (5)$$

A continuación, estos vectores de parámetros LPC suavizados se utilizan para una interpolación por subtramas, en lugar de los vectores de parámetros LPC descodificados originales $a(n-1)$, $a_m(n)$ y $a(n)$.

- 5 Igual que antes, un elemento importante de la presente invención es la reducción de las fluctuaciones de potencia y del espectro de la señal de excitación del filtro LPC durante periodos de inactividad de voz. De acuerdo con una realización preferida de la invención, la modificación se realiza de tal modo que la señal de excitación tiene menos fluctuaciones en la inclinación espectral y se compensa esencialmente una inclinación espectral existente.

- 10 Por consiguiente, los inventores tienen en cuenta y reconocen que muchos códecs de voz (y códecs AbS en particular) no necesariamente producen señales de excitación sin inclinación o blancas. Por el contrario, estos optimizan la excitación con el objetivo de adaptar la señal de entrada original a la señal sintetizada lo que, especialmente en el caso de codificadores de voz de tasa reducida, puede conducir a fluctuaciones significativas de la inclinación espectral de la señal de excitación entre trama y trama.

La compensación de la inclinación se puede realizar con un filtro de compensación de la inclinación (o filtro blanqueador) $H(z)$, según:

$$H(z) = 1 - \sum_{k=1}^P a_k \cdot z^{-k} \quad (6)$$

- 15 Los coeficientes de este filtro a_k se calculan fácilmente como coeficientes LPC de la señal de excitación original. Una elección adecuada del orden P del predictor es 1, en cuyo caso esencialmente se lleva a cabo solamente compensación de inclinación, más que blanqueado. En ese caso con el coeficiente a_1 se calcula como

$$a_1 = \frac{r_e(1)}{r_e(0)} \quad (7)$$

- 20 donde $r_e(0)$ y $r_e(1)$ son los coeficientes de autocorrelación de orden cero y primero, de la señal de excitación de filtro de síntesis LPC original.

La compensación de inclinación u operación de blanqueado descrita se realiza preferentemente por lo menos una vez para cada trama o una vez para cada subtrama.

- 25 De acuerdo con una realización particular alternativa, las fluctuaciones de potencia y espectrales de la señal de excitación se pueden reducir asimismo sustituyendo una parte de la señal de excitación con una señal de ruido blanco. A este respecto, se genera en primer lugar una secuencia aleatoria escalada adecuadamente. El escalamiento se realiza de tal modo que su potencia sea igual a la potencia de la señal de excitación o a la potencia suavizada de la señal de excitación. El segundo caso es preferible y el suavizado se puede realizar mediante filtrado de paso bajo de la estimación de la potencia de la señal de excitación o un factor de ganancia de excitación obtenido de la misma. Por consiguiente, se calcula un factor de ganancia no suavizado $g(n)$ como la raíz cuadrada de la potencia de la señal de excitación. A continuación se lleva a cabo el filtrado de paso bajo, preferentemente mediante filtrado autorregresivo de primer orden según:

$$\tilde{g}(n) = \kappa \cdot \tilde{g}(n-1) + (1 - \kappa) \cdot g(n) \quad (8)$$

- 35 En este caso $\tilde{g}(n)$ de representa el factor de ganancia del filtrado de paso bajo obtenido para la presente trama n y κ es un factor de ponderación que controla el grado de suavizado. Una elección adecuada para κ es 0,9. Si la secuencia aleatoria original tiene una potencia normalizada (varianza) de 1, entonces después del escalamiento a la señal de ruido r , su potencia corresponde a la potencia de la señal de excitación o a la potencia suavizada de la señal de excitación. Cabe señalar que la operación de suavizado del factor de ganancia se podría realizar asimismo en el dominio logarítmico, según

$$\log(\tilde{g}(n)) = \kappa \cdot \log(\tilde{g}(n-1)) + (1 - \kappa) \cdot \log(g(n)) \quad (9)$$

- 40 En una siguiente etapa, la señal de excitación se combina con la señal de ruido. A este respecto, la señal de excitación e se escala mediante algún factor α , la señal de ruido r se escala con algún factor β y a continuación las dos señales escalada se suman.

$$\hat{e}' = \alpha \cdot e + \beta \cdot r \quad (10)$$

- 45 El factor β puede, pero no necesariamente tiene por qué, corresponder al factor de control β utilizado para el suavizado de parámetros LPC. De nuevo, éste se puede obtener a partir de un parámetro denominado factor del nivel de ruido. De acuerdo con una realización preferida, el factor β se escoge como $1 - \alpha$. En ese caso, una elección adecuada para α es 0,5 o mayor, aunque menor o igual que 1. Sin embargo, salvo que α sea igual a 1, se observa

que la señal e' tiene una potencia menor que la señal de excitación e . Este efecto, a su vez, puede provocar discontinuidades no deseables en la señal de salida sintetizada en las transiciones entre inactividad y voz activa. Para resolver este problema se ha considerado que e y r sean generalmente secuencias aleatorias estadísticamente independientes. Por consiguiente, la potencia de la señal de excitación modificada depende del factor α y de las potencias de la señal de excitación e y de la señal de ruido r , como sigue:

$$P\{\hat{e}'\} = \alpha^2 \cdot P\{e\} + (1 - \alpha)^2 \cdot P\{r\} \quad (11)$$

Por lo tanto, para asegurar que la señal de excitación modificada tiene una potencia adecuada, ésta tiene que ser escalada adicionalmente mediante un factor γ :

$$\hat{e} = \gamma \cdot \hat{e}' \quad (12)$$

- 10 Con la hipótesis simplificada (ignorando el suavizado de potencia de la señal de ruido que se ha descrito anteriormente) de que la potencia de la señal de ruido y la potencia deseada de la señal de excitación modificada son idénticas a la potencia de la señal de excitación $P\{e\}$, se encuentra que el factor γ se tiene que escoger como sigue:

$$\gamma = \frac{1}{\sqrt{\alpha^2 + (1 - \alpha)^2}} \quad (13)$$

- 15 Una aproximación adecuada es escalar solamente la señal de excitación con un factor γ pero no la señal de ruido:

$$\hat{e} = \gamma \cdot \alpha \cdot e + (1 - \alpha) \cdot r \quad (14)$$

La operación de mezcla de ruido descrita se realiza preferentemente una vez para cada trama, pero podría asimismo realizarse una vez para cada subtrama.

- 20 En el curso de minuciosas investigaciones, se ha encontrado que es preferible que la compensación de inclinación (blanqueado) descrita y la modificación del ruido descrita de la señal de excitación se realicen en combinación. En ese caso, la mejor calidad de la señal de ruido de fondo sintetizada se puede conseguir cuando la modificación del ruido trabaja con la señal de excitación compensada en inclinación en lugar de con la señal de excitación original del decodificador de voz.

- 25 Para hacer el trabajo del procedimiento de manera aún más óptima puede ser necesario garantizar que ni el suavizado de los parámetros LPC ni las modificaciones de la excitación afecten a la señal de voz activa. De acuerdo con una realización básica y haciendo referencia a la figura 4, esto es posible si la operación de suavizado se activa en respuesta a una VAD que indica inactividad de voz S50.

- 30 Otra realización preferida de la invención es su aplicación en un códec de voz escalable. Se puede conseguir un rendimiento global mejorado adicionalmente mediante las etapas de adaptar la operación de suavizado descrita del ruido de fondo estacionario a la tasa de bits a la que se descodifica la señal. Preferentemente, el suavizado se realiza solamente en la descodificación de las capas inferiores de tasa reducida mientras que se desconecta (o se reduce) cuando se descodifica a tasas de bits superiores. La razón es que las capas superiores no acusan normalmente tanta turbulencia y una operación de suavizado podría incluso afectar a la fidelidad con la que el decodificador resintetiza la señal de voz a una tasa de bits superior.

- 35 Haciendo referencia a la figura 5, se describirá un dispositivo 1 en un decodificador que habilita el procedimiento según la presente invención.

- El dispositivo 1 comprende una unidad general de entrada/salida E/S 10 para recibir señales de entrada y transmitir señales de salida desde el dispositivo. La unidad comprende preferentemente cualquier funcionalidad necesaria para recibir y descodificar señales para el dispositivo. Además, el dispositivo 1 comprende una unidad 20 de parámetros LPC para descodificar y determinar parámetros LPC para la señal recibida y descodificada, y una unidad de excitación 30 para descodificar y determinar una señal de excitación para la señal de entrada recibida. Además, el dispositivo 1 comprende una unidad de modificación 35 para modificar la señal de excitación determinada reduciendo las fluctuaciones de potencia y espectrales de la señal de excitación. Finalmente, el dispositivo 1 comprende un filtro o unidad de síntesis LPC 40 para proporcionar una señal de salida de voz sintetizada suavizada en base, por lo menos, a los parámetros LPC determinados y a la señal de excitación determinada modificada.

De acuerdo con otra realización, también haciendo referencia a la figura 5, el dispositivo comprende una unidad de suavizado 25 para suavizar los parámetros LPC determinados procedentes de la unidad 20 de parámetros LPC. Además, la unidad 40 de síntesis de LPC está adaptada para determinar la señal de voz sintetizada en base, por lo menos, a los parámetros LPC suavizados y a la señal de excitación modificada.

5 Finalmente, el dispositivo puede estar dotado de una unidad de detección para detectar si la sesión de voz comprende una parte de voz activa, por ejemplo alguien está realmente hablando, o si existe solamente un ruido de fondo presente, por ejemplo uno de los usuarios está en silencio y el móvil está solamente registrando el ruido de fondo. En tal caso, el dispositivo está adaptado para llevar a cabo solamente las etapas de modificación si existe una parte de voz inactiva de la sesión de voz. En otras palabras, la operación de suavizado de la presente invención (suavizado de parámetros LPC y/o modificación de la señal de excitación) se lleva a cabo solamente durante períodos de inactividad de voz.

Las ventajas de la presente invención comprenden:

10 Con la presente invención es posible mejorar la reconstrucción o la calidad de la señal de voz sintetizada de señales de ruido de fondo estacionario (tal como ruido de coches) durante períodos de inactividad de voz.

Los expertos en la materia comprenderán que se pueden realizar diversos cambios y modificaciones a la presente invención sin apartarse del alcance de la misma, que se define mediante las reivindicaciones adjuntas.

Referencias

- [1] Patente U.S.A. 5632004.
- 15 [2] Patente U.S.A. 5579432.
- [3] Patente U.S.A. 5487087.
- [4] Patente U.S.A. 6275798 B1.
- [5] 3GPP TS 26.090, códec de voz AMR; Funciones de transcodificación..
- [6] EP 1096476.
- 20 [7] EP 1688920
- [8] Patente U.S.A. 5953697
- [9] EP 665530 B1

REIVINDICACIONES

1. Un procedimiento para suavizar ruido de fondo en una sesión de voz de telecomunicaciones, que comprende recibir y descodificar (S10) una señal representativa de una sesión de voz, comprendiendo dicha señal tanto un componente de voz como un componente de ruido de fondo;
- 5 determinar (S20) parámetros LPC para dicha señal recibida;
determinar (S30) una señal de excitación para dicha señal recibida;
modificar (S35) dicha señal de excitación determinada reduciendo las fluctuaciones de potencia y espectrales de la señal de excitación;
- 10 sintetizar y emitir (S40) una señal de salida en base a dichos parámetros LPC y a dicha señal de excitación, **caracterizado por**:
modificar (S25) dicho conjunto determinado de parámetros LPC proporcionando un conjunto filtrado en paso bajo de parámetros LPC, y determinar una combinación ponderada de dicho conjunto filtrado en paso bajo y dicho conjunto determinado de parámetros LPC, y llevar a cabo dicha etapa de síntesis y emisión en base a dicho conjunto modificado de parámetros LPC, para proporcionar una señal de salida suavizada.
- 15 2. El procedimiento según la reivindicación 1, **caracterizado por** llevar a cabo dicho filtrado de paso bajo mediante filtrado autorregresivo de primer orden.
3. El procedimiento según la reivindicación 1, **caracterizado por** dicha etapa de modificar (S35) dicha señal de excitación, que comprende modificar un espectro de dicha señal de excitación compensando una inclinación.
- 20 4. El procedimiento según la reivindicación 1, **caracterizado por** dicha etapa de modificar la señal de excitación que comprende además sustituir por lo menos parte de la señal de excitación con una señal de ruido blanco.
5. El procedimiento según la reivindicación 4, **caracterizado por** las etapas de escalamiento de una potencia de dicha señal de ruido blanco para hacerla igual a la potencia de la señal de excitación determinada o a una representación suavizada de la misma, y combinar linealmente la señal de excitación determinada y la señal de ruido escalada para proporcionar dicha señal de excitación modificada.
- 25 6. El procedimiento según la reivindicación 5, **caracterizado por** llevar a cabo dicha combinación lineal de tal modo que la potencia de la señal de excitación modificada es igual a la potencia de la señal de excitación original.
7. El procedimiento según cualquiera de las reivindicaciones anteriores, **caracterizado por** la etapa adicional (S50) de determinar si dicho componente de voz está activo o inactivo.
- 30 8. El procedimiento según la reivindicación 7, **caracterizado por** llevar a cabo dicha etapa de modificación (S35) solamente si dicho componente de voz está inactivo.
9. Un dispositivo de suavizado, que comprende
medios (10) para recibir y descodificar una señal representativa de una sesión de voz, comprendiendo dicha señal tanto un componente de voz como un componente de ruido de fondo;
medios (20) para determinar parámetros LPC para dicha señal recibida;
- 35 medios (30) para determinar una señal de excitación para dicha señal recibida;
medios (35) para modificar dicha señal de excitación determinada reduciendo las fluctuaciones de potencia y espectrales de la señal de excitación;
- medios (40) para sintetizar una señal de salida en base a dichos parámetros LPC y a dicha señal de excitación, **caracterizado por**:
medios (25) para modificar dicho conjunto determinado de parámetros LPC proporcionando un conjunto filtrado en paso bajo de parámetros LPC, estando adaptados dichos medios (25) para determinar una combinación ponderada de dicho conjunto filtrado en paso bajo y de dicho conjunto determinado de parámetros LPC, y dichos medios de síntesis (40) están adaptados para sintetizar dicha señal de salida en base a dicho conjunto modificado de parámetros LPC con el fin de proporcionar una señal de salida suavizada.
- 45 10. El dispositivo según la reivindicación 9, **caracterizado por** medios adicionales para detectar un estado inactivo de dicho componente de voz.

11. El dispositivo según la reivindicación 10, **caracterizado por que** dichos medios (35) de modificación de la señal de excitación están adaptados para llevar a cabo dicha etapa de modificación en respuesta a un componente de voz inactivo detectado.

5 12. Una unidad de decodificador en un sistema de telecomunicación que comprende un dispositivo de suavizado según cualquiera de las reivindicaciones 9 a 11.

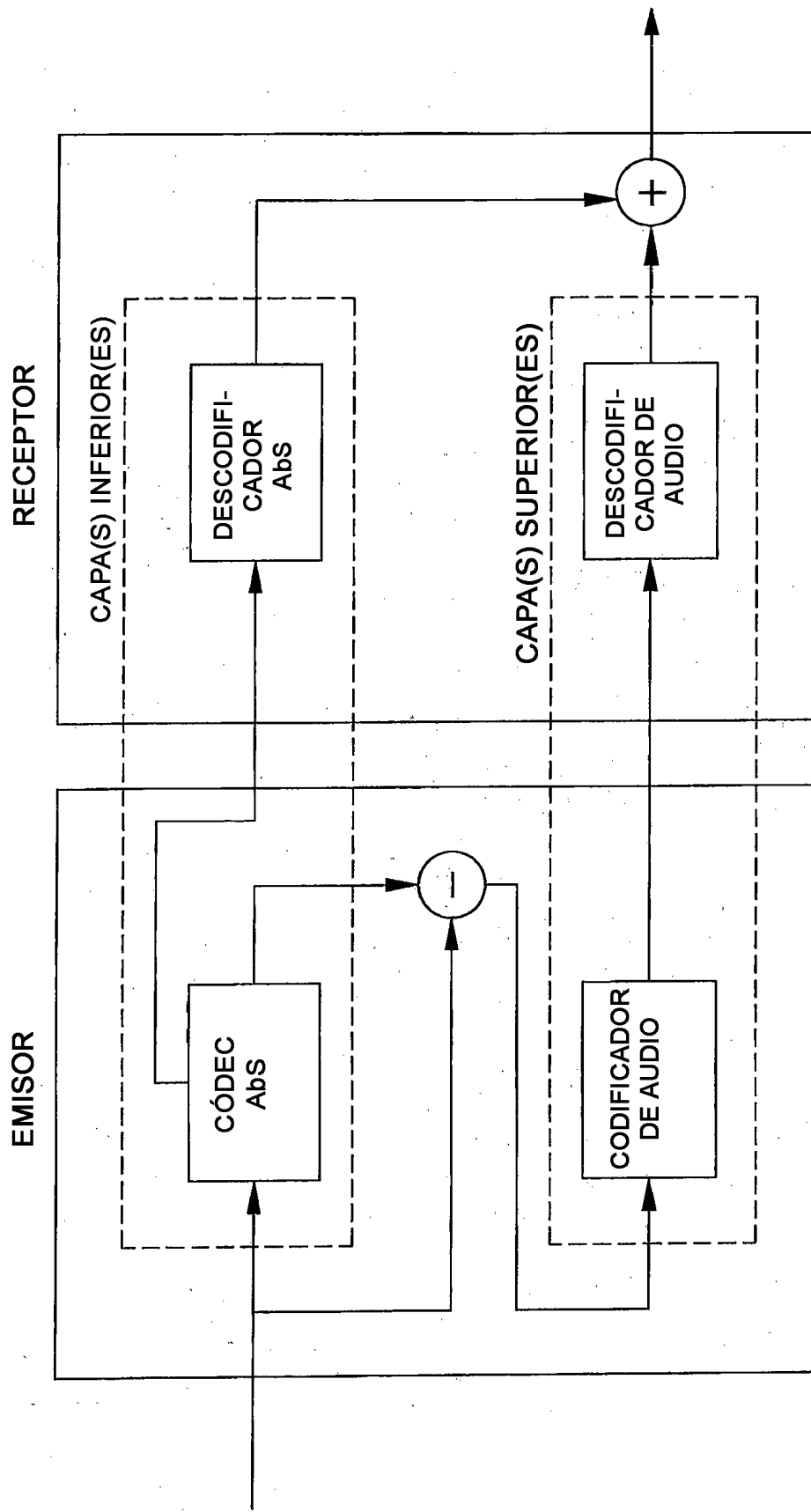


Fig. 1

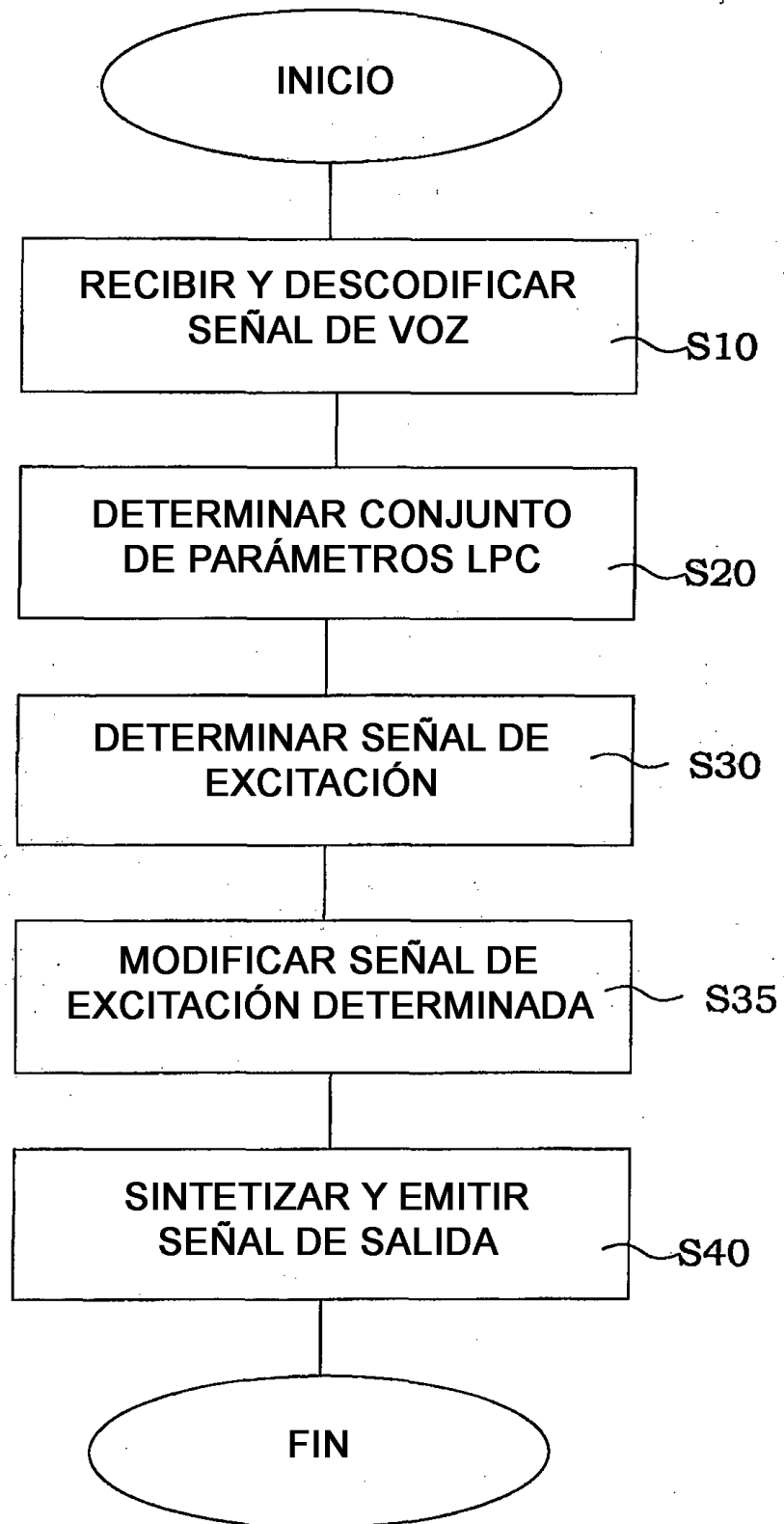


Fig. 2

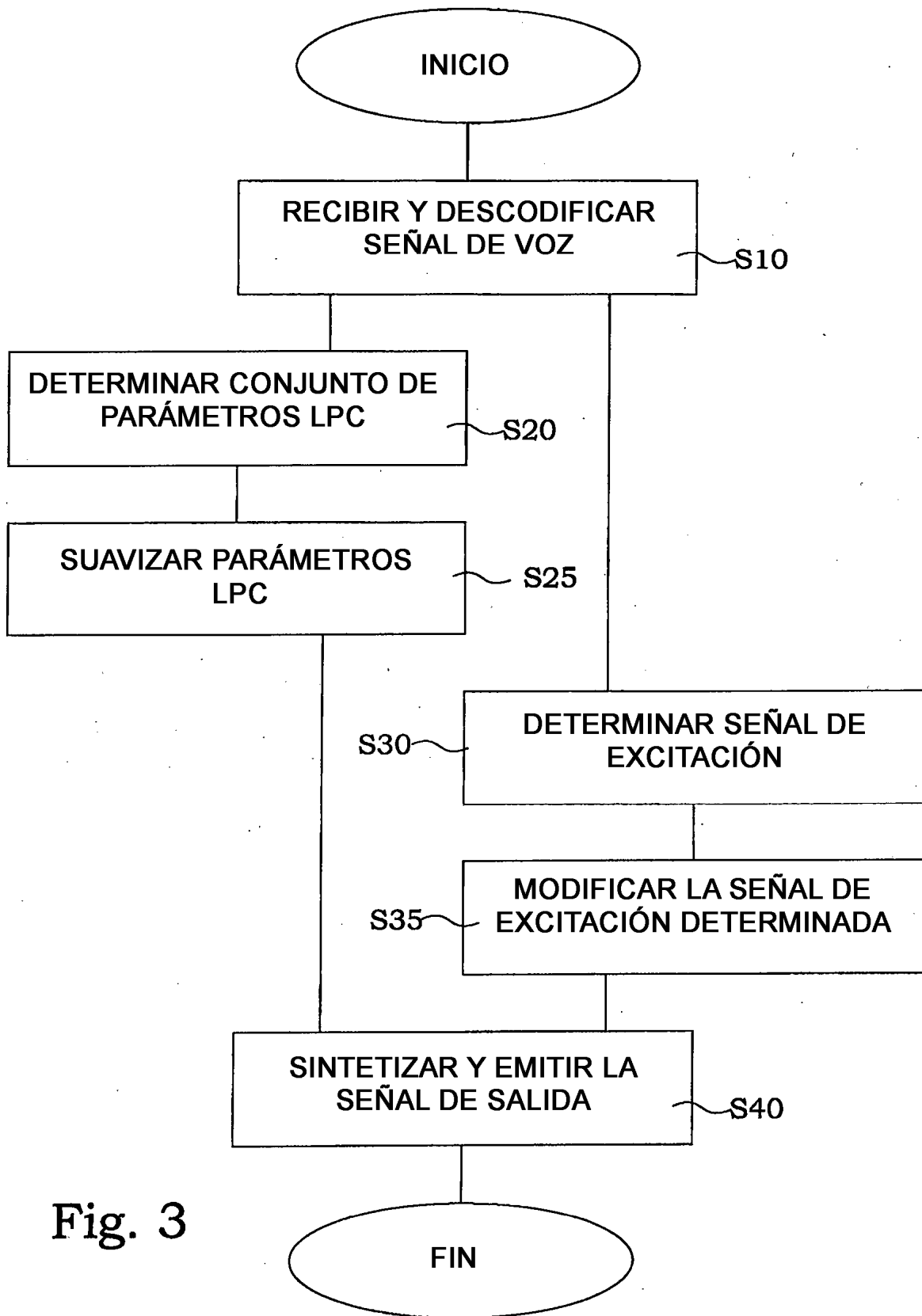


Fig. 3

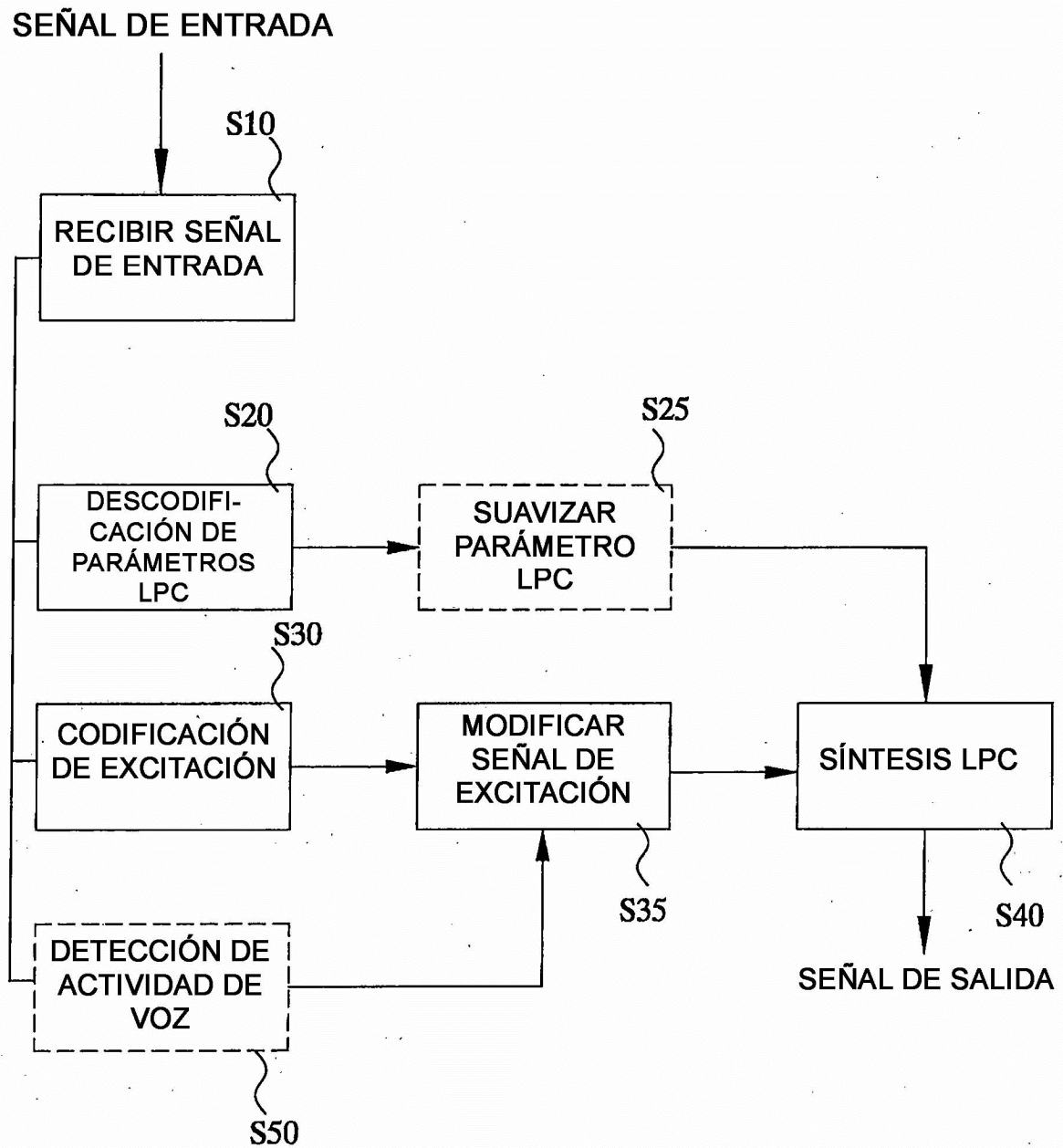


Fig. 4

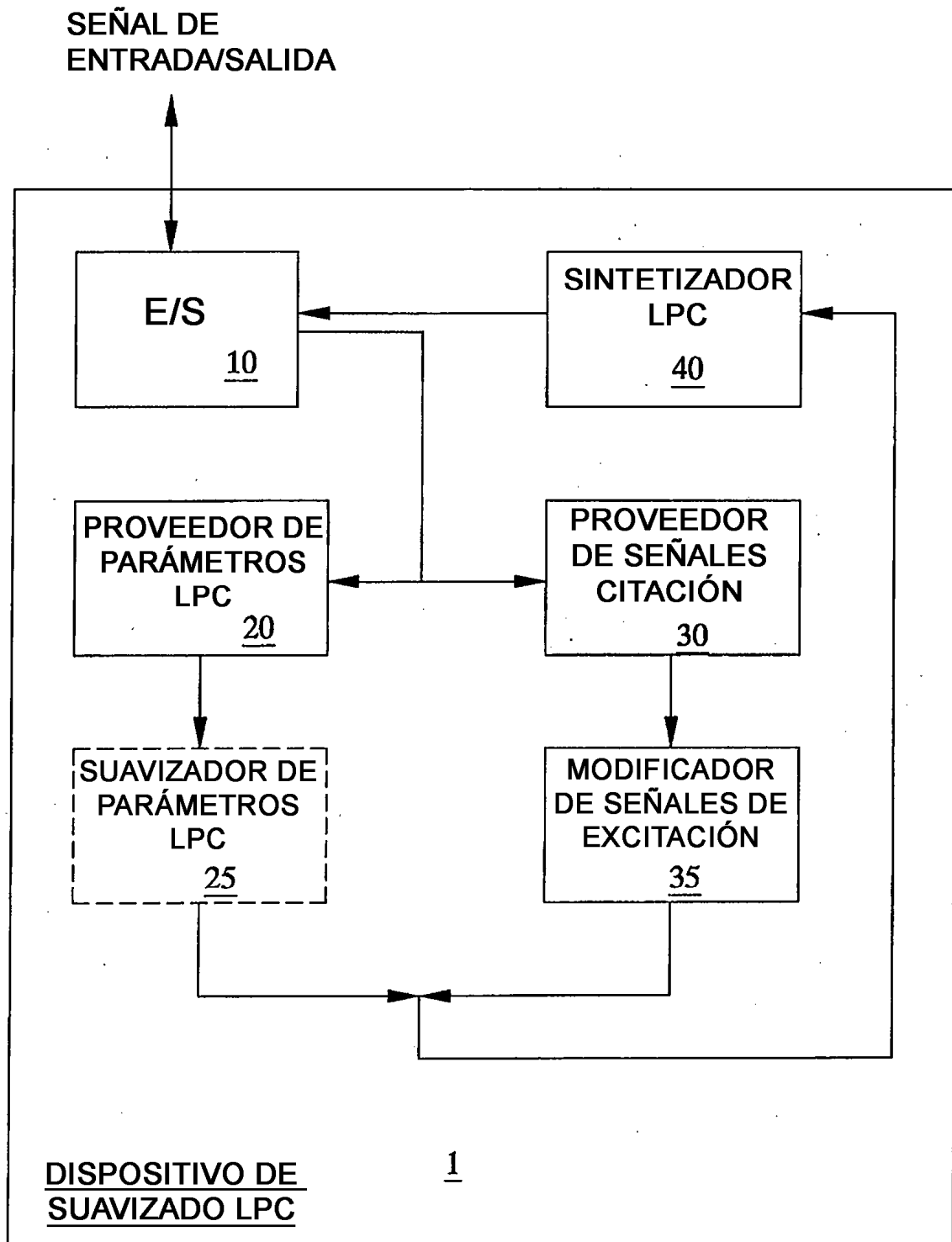


Fig. 5