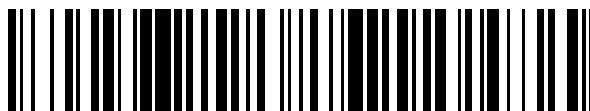


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 778 076**

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 15175006 (4)**

97 Fecha y número de publicación de la concesión europea: **25.12.2019 EP 2945158**

54 Título: **Método y disposición para suavizar ruido estacionario de fondo**

30 Prioridad:

**05.03.2007 US 892994 P**

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

**07.08.2020**

73 Titular/es:

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

72 Inventor/es:

**BRUHN, STEFAN**

74 Agente/Representante:

**LINAGE GONZÁLEZ, Rafael**

**ES 2 778 076 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

Método y disposición para suavizar ruido estacionario de fondo

5 **Campo técnico**

La presente invención se refiere a la codificación de voz en sistemas de telecomunicaciones en general, especialmente a métodos y disposiciones para suavizar el ruido estacionario de fondo en tales sistemas.

10 **Antecedentes**

La codificación de voz es el proceso de obtener una representación compacta de señales de voz para una transmisión eficiente a través de canales y/o almacenamiento por cable e inalámbrico de banda limitada. Hoy en día, los codificadores de voz se han convertido en componentes esenciales en las telecomunicaciones y en la infraestructura multimedia. Los sistemas comerciales que dependen de una codificación de voz eficiente incluyen comunicación celular, protocolo de voz sobre Internet (VOIP), videoconferencia, juguetes electrónicos, archivo y voz y datos digitales simultáneos (DSVD), así como numerosos juegos basados en PC y aplicaciones multimedia.

Al ser una señal de tiempo continuo, la voz puede representarse digitalmente a través de un proceso de muestreo y cuantificación. Las muestras de voz se cuantifican típicamente usando cuantificación de 16 bits u 8 bits. Al igual que muchas otras señales, una señal de voz contiene una gran cantidad de información que es redundante (información mutua no nula entre muestras sucesivas en la señal) o perceptualmente irrelevante (información que no es percibida por los oyentes humanos). La mayoría de los codificadores de telecomunicaciones tienen pérdidas, lo que significa que el discurso sintetizado es perceptivamente similar al original pero puede ser físicamente diferente.

Un codificador de voz convierte una señal de voz digitalizada en una representación codificada, que generalmente se transmite en tramas. En consecuencia, un decodificador 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). Algunos ejemplos de tales codificadores son: los códecs de voz 3GPP FR, EFR, AMR y AMR-WB, los códecs de voz 3GPP2 EVRC, SMV y EVRC-WB, y varios códecs de ITU-T como G.728, G723, G. 729, etc.

Todos estos codificadores utilizan un concepto de filtro de síntesis en el proceso de generación de señal. El filtro se usa para modelar el espectro a corto plazo de la señal que se va a reproducir, mientras que se supone que la entrada al filtro maneja todas las demás variaciones de señal.

Una característica común de estos modelos de filtro de síntesis es que la señal que se va a reproducir está representada por parámetros que definen el filtro de síntesis. El término "predictivo lineal" se refiere a una clase de métodos usados a menudo para estimar los parámetros del filtro. En codificadores basados en LPC, la señal de voz se ve como la salida de un sistema lineal invariante en el tiempo (LTI) cuya entrada es la señal de excitación al filtro. Por lo tanto, la señal que se va a reproducir está representada en parte por un conjunto de parámetros de filtro y en parte por la señal de excitación que acciona el filtro. La ventaja de este concepto de codificación surge del hecho de que tanto el filtro como su señal de excitación de accionamiento pueden describirse eficientemente con relativamente pocos bits.

Una clase particular de códecs basados en LPC se basa en el llamado principio de análisis por síntesis (AbS). Estos códecs incorporan una copia local del decodificador en el codificador y encuentran la señal de excitación de accionamiento del filtro de síntesis seleccionando esa señal de excitación entre un conjunto de señales de excitación candidatas que maximiza la similitud de la señal de salida sintetizada con la señal de voz original.

El concepto de utilizar tal codificación predictiva lineal y particularmente la codificación AbS ha demostrado funcionar relativamente bien para señales de voz, incluso a bajas tasas de bits, por ejemplo, 4-12 kbps. Sin embargo, cuando el usuario de un teléfono móvil que usa tal técnica de codificación está en silencio y la señal de entrada comprende los sonidos circundantes, por ejemplo, ruido, los codificadores conocidos actualmente tienen dificultades para hacer frente a esta situación, ya que están optimizados para señales de voz. Un oyente en el lado receptor puede molestarse fácilmente cuando no se pueden reconocer los sonidos de fondo familiares, ya que el codificador los ha "maltratado".

El llamado remolino causa una de las degradaciones de calidad más severas en los sonidos de fondo reproducidos. Este es un fenómeno que ocurre en sonidos de ruido de fondo relativamente estacionario, como el ruido de un automóvil, y es causado por fluctuaciones temporales no naturales de la potencia y el espectro de la señal decodificada. Estas fluctuaciones a su vez son causadas por una estimación y cuantificación inadecuadas de los coeficientes del filtro de síntesis y su señal de excitación. Por lo general, el remolino se vuelve menor cuando aumenta la tasa de bits del códec.

El remolino se ha identificado como un problema en la técnica anterior y se han propuesto múltiples soluciones en la literatura. Una de las soluciones propuestas se describe en la patente de los Estados Unidos 5632004 [1]. De acuerdo con esta patente, durante la inactividad de voz, los parámetros del filtro se modifican mediante filtrado de paso bajo o expansión de ancho de banda de modo que se reducen las variaciones espectrales del sonido de fondo sintetizado. Este método fue refinado en la patente de los Estados Unidos 5579432 [2] de tal manera que la técnica anti-remolino descrita solo se aplica al ruido estacionario de fondo detectado.

Otro método que aborda el problema de remolino se describe en la patente de los Estados Unidos 5487087 [3]. Este método usa un esquema de cuantificación de señal modificada que coincide tanto con la señal en sí como con sus variaciones temporales. En particular, se prevé usar un cuantificador de fluctuación tan reducida para los parámetros del filtro LPC y los parámetros de ganancia de señal durante los períodos inactivos de voz.

Las degradaciones de la calidad de la señal causadas por fluctuaciones de potencia no deseadas de la señal sintetizada se abordan mediante otro conjunto de métodos. Uno de ellos se describe en la patente estadounidense 6275798 [4] y también es parte del algoritmo de códec de voz AMR descrito en 3GPP TS 26.090 [5]. De acuerdo con esto, la ganancia de al menos una componente de la señal de excitación del filtro sintetizado, la contribución fija del libro de códigos, se suaviza de forma adaptativa dependiendo de la estacionalidad del espectro LPC a corto plazo. Este método se ha desarrollado en la patente EP 1096476 [6] y la solicitud de patente EP 1688920 [7] donde el suavizado implica además una limitación de la ganancia que se usará en la síntesis de la señal. Un método relacionado para ser usado en codificadores de voz LPC se describe en el documento US 5953697 [8]. De acuerdo con esto, la ganancia de la señal de excitación del filtro de síntesis se controla de modo que la amplitud máxima de la voz sintetizada llegue a la envoltura de forma de onda de voz de entrada.

Sin embargo, una clase adicional de métodos que abordan el problema de remolino funciona como un post-procesador después del decodificador de voz. La patente EP 0665530 [9] describe un método que durante la inactividad de voz detectada reemplaza una porción de la señal de salida del decodificador de voz por un ruido blanco filtrado de paso bajo o una señal de ruido de confort. Se adoptan enfoques similares en varias publicaciones que divulgan métodos relacionados que reemplazan parte de la señal de salida del decodificador de voz con ruido filtrado.

Murashima A. et al. [10] describe una técnica de post-procesamiento para mejorar la calidad de codificación de CELP bajo ruido de fondo. Suaviza de forma adaptativa tanto la envoltura espectral como la energía de la señal de excitación estimada para reducir sus fluctuaciones temporales, que causan la degradación perceptual.

La codificación escalable o embebida, con referencia a la figura 1, es un paradigma 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, cada una encima de la otra, proporcionan alguna mejora en relación con la codificación, que se logra con todas las capas desde el núcleo hasta la capa previa respectiva. Cada capa agrega una tasa de bits adicional. El flujo de bits generado está embebido, lo que significa que el flujo de bits de la codificación de capa inferior está embebido en flujos de bits de capas superiores. Esta propiedad hace posible que en cualquier lugar de la transmisión o en el receptor se caigan los bits que pertenecen a las capas superiores. Tal flujo de bits despojado aún puede decodificarse hasta la capa cuyos bits se retienen.

El algoritmo de compresión de voz escalable más común hoy en día es el códec PCM de logaritmo de ley A/U G.711 de 64 kbps. El códec G.711 muestreado de 8 kHz convierte muestras PCM lineales de 12 bits o 13 bits en muestras logarítmicas de 8 bits. La representación de bits ordenada de las muestras logarítmicas permite robar los bits menos significativos (LSB) en un flujo de bits G.711, haciendo que el codificador G.711 sea prácticamente escalable por SNR entre 48, 56 y 64 kbps. Esta propiedad de escalabilidad del códec G.711 se usa en las redes de comunicación conmutadas por circuito para fines de señalización de control en banda. Un ejemplo reciente del uso de esta propiedad de escalado G.711 es el protocolo 3GPP TFO que permite la configuración y el transporte de voz de banda ancha a través de enlaces PCM heredados de 64 kbps. Inicialmente, se usan ocho kbps del flujo original G.711 de 64 kbps para permitir una configuración de llamada del servicio de voz de banda ancha sin afectar considerablemente la calidad del servicio de banda estrecha. Después de la configuración de la llamada, la voz de banda ancha usará 16 kbps del flujo G.711 de 64 kbps. Otros estándares de codificación de voz más antiguos que soportan escalabilidad de bucle abierto son G.727 (ADPCM embebido) 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 de MPE puede mejorarse mediante la transmisión de información adicional de parámetros de filtro o información adicional de parámetros de innovación. El Sector de Normalización de la Unión Internacional de Telecomunicaciones, UIT-T, ha finalizado recientemente la estandarización de un nuevo códec escalable G.729.1, apodado s G.729. EV. El rango de tasa de bits de este códec de voz escalable es de 8 kbps a 32 kbps. El principal caso de uso de este códec es permitir el intercambio eficiente de un recurso de ancho de banda limitado en las pasarelas domésticas o de oficina, por ejemplo, el enlace ascendente xDSL 64/128 kbps compartido entre varias llamadas VOIP.

Una tendencia reciente en la codificación de voz escalable es proporcionar capas superiores con soporte para la codificación de señales de audio que no son de voz, como la música. En tales códecs, las capas inferiores emplean una mera codificación de voz convencional, por ejemplo, de acuerdo con el paradigma de análisis por síntesis del cual CELP es un ejemplo destacado. Como tal codificación es muy adecuada solo para voz pero no tanto para señales de audio que no son de voz, como la música, las capas superiores funcionan de acuerdo con un paradigma de codificación, que se usa en códecs de audio. Aquí, típicamente la codificación de la capa superior funciona en el error de codificación de la codificación de la capa inferior.

Otro método relevante con respecto a los códecs de voz es la llamada compensación de inclinación espectral, que se realiza en el contexto del filtrado posterior adaptativo de la voz decodificada. El problema resuelto por esto es compensar la inclinación espectral introducida por los filtros posteriores a corto plazo o formantes. Tales técnicas son parte de, por ejemplo, el códec AMR y el códec SMV y se dirigen principalmente al rendimiento del códec durante la voz en lugar de su rendimiento de ruido de fondo. El códec SMV aplica esta compensación de inclinación en el dominio residual ponderado antes del filtrado de síntesis, aunque no en respuesta a un análisis LPC del residuo.

El problema con los métodos descritos anteriormente de los documentos US 5632004, US 5579432 y US 5487087 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 causan el problema de remolino están relacionadas con las fluctuaciones de los espectros de filtro de síntesis LPC. Sin embargo, este no es el caso y especialmente si la señal de excitación solo se cuantifica de manera aproximada. En ese caso, las fluctuaciones espectrales de la señal de excitación tienen un efecto similar a las fluctuaciones del filtro LPC y, por lo tanto, deben evitarse.

El problema con los métodos que abordan las fluctuaciones de potencia no deseadas de la señal sintetizada es que solo abordan una parte del problema de remolino, pero no proporcionan una solución relacionada con las fluctuaciones espectrales.

Las simulaciones muestran que, incluso en combinación con los métodos citados que abordan las fluctuaciones espectrales, no se pueden evitar todas las degradaciones de calidad de la señal relacionadas con remolinos durante los sonidos de fondo estacionarios.

Un problema con los métodos que funcionan como post-procesador después del decodificador de voz es que reemplazan solo una porción de la señal de salida decodificada de voz con una señal de ruido suavizada. Por lo tanto, el problema de remolino no se resuelve en la porción de señal restante que se origina desde el decodificador de voz y, por lo tanto, la señal de salida final no se forma usando el mismo filtro de síntesis LPC que la señal de salida del decodificador de voz. Esto puede conducir a posibles discontinuidades de sonido, especialmente durante las transiciones de inactividad a la voz activa. Además, tales métodos de post-procesamiento son desventajosos, ya que requieren una complejidad computacional relativamente alta.

Ninguno de los métodos existentes proporciona una solución al problema de que una de las razones para el remolino radica en las fluctuaciones espectrales de la señal de excitación del filtro de síntesis de LPC. Este problema se vuelve grave, especialmente si la señal de excitación se representa con muy pocos bits, que es típicamente el caso de los códecs de voz que funcionan a tasas de bits de 12 kbps o inferiores.

En consecuencia, existe la necesidad de métodos y disposiciones para aliviar los problemas de remolino descritos anteriormente causados por el ruido estacionario de fondo durante los períodos de inactividad de voz.

## Sumario

Un objeto de la presente invención es proporcionar calidad mejorada de señales de voz en un sistema de telecomunicaciones.

Un objeto adicional es proporcionar calidad mejorada de una señal de salida de decodificador de voz durante períodos de inactividad de voz con ruido estacionario de fondo.

Las ventajas de la presente invención comprenden:

habilitar una señal mejorada de salida de decodificador de voz;

habilitar una señal suave de salida de decodificador de voz.

## Breve descripción de los dibujos

La invención, junto con otros objetos y ventajas de la misma, puede entenderse mejor haciendo referencia a la siguiente descripción tomada junto con los dibujos que se acompañan, en los que:

la figura 1 es un esquema de bloques de un códec de audio y voz escalable;

la figura 2 es un diagrama de flujo que ilustra una realización de un método de acuerdo con la presente invención;

5 la figura 3 es un diagrama de flujo de una realización adicional de un método de acuerdo con la presente invención;

la figura 4 es un diagrama de bloques que ilustra realizaciones de un método de acuerdo con la presente invención;

10 la figura 5 es una ilustración de una realización de una disposición de acuerdo con la presente invención.

### Abreviaturas

AbS	Análisis por síntesis
ADPCM	PCM diferencial adaptativo
AMR-WB	Banda ancha multifrecuencia adaptativa
EVRC-WB	Códec de banda ancha de frecuencia variable mejorada
CELP	Predicción lineal excitada por código
ISP	Par espectral de inmitancia
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 de código de pulso
SMV	Codificador de voz de modo seleccionable
VAD	Detector de actividad de voz

### 15 Descripción detallada

Todos los casos siguientes de la palabra "realización/realizaciones", si se refieren a combinaciones futuras diferentes de las definidas por las reivindicaciones independientes, se refieren a ejemplos que se solicitaron originalmente que no representan realizaciones de la presente invención reivindicada; esos ejemplos todavía se muestran solo con fines ilustrativos.

20

La presente invención se describirá en el contexto de una sesión de voz, por ejemplo, llamada telefónica, en un sistema general de telecomunicaciones. Típicamente, los métodos y disposiciones se implementarán en un decodificador adecuado para la síntesis de voz. Sin embargo, es igualmente posible que los métodos y disposiciones se implementen en un nodo intermediario en la red y posteriormente se transmitan a un usuario de destino. El sistema de telecomunicaciones puede ser tanto inalámbrico como por cable.

25

En consecuencia, la presente invención permite métodos y disposiciones para aliviar los problemas conocidos anteriormente descritos con remolinos causados por el ruido estacionario de fondo durante los períodos de inactividad de voz en una sesión de voz telefónica. Específicamente, la presente invención permite mejorar la calidad de una señal de salida del decodificador de voz durante períodos de inactividad de voz con ruido estacionario de fondo.

30

Dentro de esta divulgación, el término sesión de voz debe interpretarse como cualquier intercambio de señales vocales a través de un sistema de telecomunicaciones. En consecuencia, una señal de sesión de voz puede describirse como que comprende una parte activa y una parte de fondo. La parte activa es la señal de voz real de la sesión. La parte de fondo es el ruido que rodea al usuario, también conocido como ruido de fondo. Un período de inactividad se define como un período de tiempo dentro de una sesión de voz donde no hay una parte activa, solo una parte de fondo, por ejemplo, la parte de voz de la sesión está inactiva.

35

De acuerdo con una realización básica, la presente invención permite mejorar la calidad de una sesión de voz al reducir las variaciones de potencia y las fluctuaciones espectrales de la señal de excitación del filtro de síntesis LPC durante la detección de períodos de inactividad de voz.

40

De acuerdo con una realización adicional, la señal de salida se mejora aún más combinando la modificación de la señal de excitación con una operación de suavizado de parámetros LPC.

45

Con referencia al diagrama de flujo de la figura 2, una realización de un método de acuerdo 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 señal de voz activa y/o una componente de ruido estacionario de fondo). Posteriormente, se determina un conjunto de parámetros LPC S20 para la señal recibida. Además, se determina una señal de excitación S30 para la señal recibida. Se sintetiza una señal de salida y la salida S40 se basa en los parámetros LPC determinados y la señal de excitación determinada. De acuerdo con la presente invención, la señal de excitación se mejora o modifica S35 reduciendo la potencia y las fluctuaciones espectrales de la señal de excitación para proporcionar una señal de salida suavizada.

Con referencia al diagrama de flujo de la figura 3, se describirá una realización adicional de un método de acuerdo con la presente invención. Los pasos correspondientes conservan los mismos números de referencia que los de la figura 2. Además del paso de modificar 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, el suavizado de parámetros LPC.

El suavizado de parámetros LPC S25 de acuerdo con una realización adicional de la presente invención, con referencia a la figura 4, comprende realizar el suavizado de parámetros LPC de tal manera que el grado de suavizado esté controlado por algún factor  $\beta$ , que a su vez se deriva de un parámetro denominado factor de ruidosidad.

En un primer paso, se calcula S20 un conjunto filtrado de paso bajo de parámetros LPC. Preferiblemente, esto se realiza mediante filtrado autorregresivo de primer orden de acuerdo con:

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

Aquí  $\tilde{a}(n)$  representa el vector de parámetros LPC filtrado de paso bajo obtenido para una trama  $n$  presente,  $a(n)$  es el vector de parámetros LPC decodificado para la trama  $n$ , y  $\lambda$  es un factor de ponderación que controla el grado de suavizado. Una opción adecuada para  $\lambda$  es 0,9.

En un segundo paso S25, se calcula una combinación ponderada del vector de parámetros LPC filtrado de paso bajo  $\tilde{a}(n)$  y el parámetro decodificado LPC vector  $a(n)$  usando el factor de control de suavizado  $\beta$ , de acuerdo con:

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

Los parámetros LPC pueden estar en cualquier representación adecuada para el filtrado y la interpolación, y preferiblemente se pueden representar como frecuencias espectrales de línea (LSF) o pares espectrales de inmitancia (ISP).

Típicamente, el decodificador de voz puede interpolar los parámetros LPC a través de subtramas en las que preferiblemente también los parámetros LPC filtrados de paso bajo se interpolan en consecuencia. En una realización particular, el decodificador de voz funciona con tramas de 20 ms de longitud y 4 subtramas de 5 ms cada una dentro de una trama. Si el decodificador de voz calcula originalmente los 4 vectores de parámetros LPC de subtrama interpolando entre un vector de parámetros LPC  $a(n-1)$  de la trama final de la trama anterior, un vector de parámetros LPC  $a_m(n)$  de la trama media y un vector de parámetros LPC  $a(n)$  de la trama final de la trama actual, entonces la combinación ponderada de los vectores de parámetros LPC filtrados de paso bajo y los vectores de parámetros LPC decodificados se calcula de la siguiente manera:

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

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

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

Posteriormente, estos vectores de parámetros LPC suavizados se usan para la interpolación en subtrama, en lugar de los vectores de parámetros LPC decodificados originales  $a(n-1)$ ,  $a_m(n)$ , y  $a(n)$ .

Como anteriormente, un elemento importante de la presente invención es la reducción de las fluctuaciones de potencia y espectro de la señal de excitación del filtro LPC durante los períodos de inactividad de voz. De acuerdo con una realización preferida de la invención, la modificación se realiza de modo que la señal de excitación tenga

menos fluctuaciones en la inclinación espectral y que se compense esencialmente una inclinación espectral existente.

5 En consecuencia, los inventores tienen en cuenta y reconocen que muchos códecs de voz (y códecs AbS en particular) no producen necesariamente señales de excitación blancas o sin inclinación. Más bien, optimizan la excitación con el objetivo para que coincida la señal de entrada original con la señal sintetizada, lo que especialmente en el caso de codificadores de voz de tasa baja puede provocar fluctuaciones significativas de la inclinación espectral de la señal de excitación de trama a trama.

10 La compensación de inclinación se puede hacer con un filtro de compensación de inclinación (o filtro de blanqueamiento)  $H(z)$  de acuerdo con:

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

15 Los coeficientes de este filtro  $a_i$  se calculan fácilmente como coeficientes LPC de la señal de excitación original. Una elección adecuada del orden indicador  $P$  es 1, en cuyo caso esencialmente se realiza una compensación de inclinación en lugar de un blanqueamiento. En ese caso, el coeficiente 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 cero y primero de autocorrelación de la señal de excitación del filtro de síntesis LPC original.

25 La operación de compensación de inclinación o blanqueamiento descrita se realiza preferiblemente al menos una vez para cada trama o una vez para cada subtrama.

De acuerdo con una realización particular alternativa, la potencia y las fluctuaciones espectrales de la señal de excitación también pueden reducirse reemplazando una parte de la señal de excitación con una señal de ruido blanco. Para este fin, primero se genera una secuencia aleatoria adecuadamente escalada. El escalado se realiza de modo que su potencia sea igual a la potencia de la señal de excitación o la potencia suavizada de la señal de excitación. Se prefiere el último caso y el suavizado se puede hacer mediante filtrado de paso bajo de estimaciones de la potencia de la señal de excitación o un factor de ganancia de excitación derivado de ella. En consecuencia, un factor de ganancia no suavizado  $g(n)$  se calcula como la raíz cuadrada de la potencia de la señal de excitación. Luego se realiza el filtrado de paso bajo, preferiblemente mediante filtrado autorregresivo de primer orden de acuerdo con:

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

40 Aquí  $\tilde{g}(n)$  representa el factor de ganancia filtrado de paso bajo obtenido para la trama actual  $n$  y  $\kappa$  es un factor de ponderación que controla el grado de suavizado. Una elección adecuada para  $\kappa$  es 0.9. Si la secuencia aleatoria original tiene una potencia normalizada (varianza) de 1, después de escalar la señal de ruido  $r$ , su potencia corresponde a la potencia de la señal de excitación o de la potencia suavizada de la señal de excitación. Se observa que la operación de suavizado del factor de ganancia también podría realizarse en el dominio logarítmico de acuerdo con

45

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

50 En el siguiente paso, la señal de excitación se combina con la señal de ruido. Para este fin, la señal de excitación  $e$  se escala por algún factor  $\alpha$ , la señal de ruido  $r$  se escala con algún factor  $\beta$  y luego se agregan las dos señales escaladas:

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

55 El factor  $\beta$  puede pero no necesariamente corresponde al factor de control  $\beta$  usado para el suavizado de parámetros LPC. Puede derivarse nuevamente de un parámetro denominado factor de ruidosidad. De acuerdo con una realización preferida, el factor  $\beta$  se elige como  $1 - \alpha$ . En ese caso, una elección adecuada para  $\alpha$  es 0,5 o mayor, aunque menor o igual a 1. Sin embargo, a menos que  $\alpha$  sea igual a 1, se observa que la señal  $\hat{e}'$  tiene una potencia menor que la señal de excitación  $e$ . Este efecto a su vez puede causar discontinuidades indeseables en la señal de

salida sintetizada en las transiciones entre la inactividad y la voz activa. Para resolver este problema, se debe considerar que  $e$  y  $r$  generalmente son secuencias aleatorias estadísticamente independientes. En consecuencia, la potencia de la señal de excitación modificada depende del factor  $\alpha$  y las potencias de la señal de excitación  $e$  y la señal de ruido  $r$ , de la siguiente manera:

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

Por lo tanto, para garantizar que la señal de excitación modificada tenga una potencia adecuada, debe ser ampliada por un factor  $\gamma$ :

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

Bajo el supuesto simplificado (ignorando el suavizado de potencia de la señal de ruido descrita 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  $\gamma$  tiene que ser elegido de la siguiente manera:

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

Una aproximación adecuada es escalar solo la señal de excitación con un factor  $\gamma$  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 preferiblemente una vez para cada trama, pero también se puede hacer una vez para cada subtrama.

En el curso de investigaciones minuciosas, se ha encontrado que preferiblemente la compensación de inclinación descrita (blanqueamiento) y la modificación de ruido descrita de la señal de excitación se realizan en combinación. En ese caso, se puede lograr la mejor calidad de la señal de ruido de fondo sintetizada cuando la modificación de ruido funciona con la señal de excitación compensada por inclinación en lugar de la señal de excitación original del decodificador de voz.

Para que el método funcione de manera aún más óptima, puede ser necesario asegurarse de que ni el suavizado del parámetro LPC ni las modificaciones de excitación afecten a la señal de voz activa. De acuerdo con una realización básica y con referencia a la figura 4, esto es posible si la operación de suavizado se activa en respuesta a un VAD que indica la inactividad de voz S50.

Una realización preferida adicional de la invención es su aplicación en un códec de voz escalable. Se puede lograr un rendimiento general mejorado adicional mediante los pasos de adaptar la operación de suavizado descrita del ruido estacionario de fondo a la tasa de bits a la que se decodifica la señal. Preferiblemente, el suavizado solo se realiza en la decodificación de las capas inferiores de tasa baja mientras se apaga (o se reduce) cuando se decodifica a tasas de bits más altas. La razón es que las capas más altas generalmente no sufren tanto el remolino y una operación de suavizado podría incluso afectar la fidelidad a la cual el decodificador vuelve a sintetizar la señal de voz a una tasa de bits más alta.

Con referencia a la figura 5, se describirá una disposición 1 en un decodificador que permite el método de acuerdo con la presente invención.

La disposición 1 comprende una unidad 10 de salida/entrada general I/O para recibir señales de entrada y transmitir señales de salida desde la disposición. La unidad comprende preferiblemente cualquier funcionalidad necesaria para recibir y decodificar señales en la disposición. Además, la disposición 1 comprende una unidad 20 de parámetros LPC para decodificar y determinar parámetros LPC para la señal recibida y decodificada, y una unidad 30 de excitación para decodificar y determinar una señal de excitación para la señal de entrada recibida. Además, la disposición 1 comprende una unidad 35 de modificación para modificar la señal de excitación determinada reduciendo la potencia y las fluctuaciones espectrales de la señal de excitación. Finalmente, la disposición 1 comprende una unidad de síntesis LPC o filtro 40 para proporcionar una señal de salida de voz sintetizada suavizada basada al menos en los parámetros LPC determinados y la señal de excitación determinada modificada.

De acuerdo con otra realización, también con referencia a la figura 5, la disposición comprende una unidad 25 de suavizado para suavizar los parámetros LPC determinados a partir de la unidad 20 de parámetros LPC. Además, la



unidad 40 de síntesis de LPC está adaptada para determinar la señal de voz sintetizada basada en al menos los parámetros de LPC suavizados y la señal de excitación modificada.

5 Finalmente, la disposición se puede proporcionar con una unidad de detección para detectar si la sesión de voz comprende una parte de voz activa, por ejemplo, alguien está hablando realmente, o si solo hay un ruido de fondo presente, por ejemplo, uno de los usuarios está en silencio y el móvil solo registra el ruido de fondo. En ese caso, la disposición está adaptada para realizar solo los pasos de modificación si hay una parte inactiva de voz 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) solo se realiza durante los períodos de inactividad de voz.

10 Las ventajas de la presente invención comprenden:

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

15 Los expertos en la técnica entenderán que pueden realizarse diversas modificaciones y cambios a la presente invención sin apartarse del alcance de la misma, que se define en las reivindicaciones adjuntas.

### 20 Referencias

- [1] Patente de Estados Unidos 5632004.
- [2] Patente de Estados Unidos 5579432.
- 25 [3] Patente de Estados Unidos 5487087.
- [4] Patente de Estados Unidos 6275798 B1.
- [5] 3GPP TS 26.090, Códec de voz AMR; Funciones de transcodificación.
- 30 [6] EP 1096476.
- [7] EP 1688920.
- 35 [8] Patente de Estados Unidos 5953697.
- [9] EP 665530 B1.
- 40 [10] Murashima A. et al. "Una técnica de procesamiento post-procesamiento para mejorar la calidad de codificación de CELP con ruido de fondo", codificación de voz, 2000. Procedimientos. 2000 IEEE workshop del 17 al 20 de septiembre, 2000, Piscataway, NJ, EE.UU., IEEE, págs. 102-104.

**REIVINDICACIONES**

1.- Un método para suavizar ruido de fondo, comprendiendo el método:

- 5 recibir y decodificar (S10) una señal codificada que comprende tanto una componente de voz como una componente de ruido de fondo;
- determinar (S20) parámetros LPC para dicha señal recibida;
- 10 determinar (S30) una señal de excitación para dicha señal recibida;
- determinar (S50) si dicha componente de voz está activa o inactiva;
- 15 modificar (S35) dicha señal de excitación determinada reduciendo la potencia y las fluctuaciones espectrales de la señal de excitación durante los períodos de inactividad de voz;
- sintetizar y emitir (S40) una señal de salida basada en dichos parámetros LPC y dicha señal de excitación;
- 20 caracterizado por:
- suavizar (S25) dicho conjunto determinado de parámetros LPC proporcionando un conjunto filtrado de paso bajo de parámetros LPC y determinando una combinación ponderada de dicho conjunto filtrado de paso bajo y dicho conjunto determinado de parámetros LPC durante períodos de inactividad de voz; y
- 25 realizar dicha síntesis y emisión (S40) basándose en dicho conjunto suavizado de parámetros LPC y dicha señal de excitación modificada.

2.- El método de acuerdo con la reivindicación 1, en el que dicho filtrado de paso bajo se realiza mediante filtrado autorregresivo de primer orden.

- 30 3.- El método de acuerdo con la reivindicación 1 o 2, en el que dicho paso de modificar dicha señal de excitación comprende realizar una compensación de inclinación de la señal de excitación con un filtro de compensación de inclinación.

- 35 4.- El método de acuerdo con cualquiera de las reivindicaciones 1 a 3, en el que dicho paso de modificar dicha señal de excitación comprende reemplazar al menos parte de la señal de excitación con una señal de ruido blanco.

- 40 5.- El método de acuerdo con la reivindicación 4, que comprende además escalar una potencia de dicha señal de ruido blanco para que sea igual a la potencia de la señal de excitación determinada o un representante suavizado de la misma, y combinar la señal de excitación determinada y la señal de ruido escalada.

6.- Un aparato, que comprende:

- 45 medios (10) para recibir y decodificar una señal codificada que comprende tanto una componente de voz como una componente de ruido de fondo;
- medios (20) para determinar los parámetros LPC para dicha señal recibida;
- 50 medios (30) para determinar una señal de excitación para dicha señal recibida;
- medios (40) para sintetizar una señal de salida basada en dichos parámetros LPC y dicha señal de excitación;
- caracterizado por:
- 55 medios (25) para suavizar dicho conjunto determinado de parámetros LPC proporcionando un conjunto filtrado de paso bajo de parámetros LPC, siendo adaptados dichos medios (25) para determinar una combinación ponderada de dicho conjunto filtrado de paso bajo y dicho conjunto determinado de parámetros LPC durante períodos de inactividad de voz; y siendo dichos medios de síntesis (40) adaptados para sintetizar dicha señal de salida basándose en dicho conjunto modificado de parámetros LPC y dicha señal de excitación modificada.
- 60

7.- El aparato de acuerdo con la reivindicación 6, en el que los medios para modificar la señal de excitación comprenden además medios para realizar la compensación de inclinación de la señal de excitación.

- 65 8.- El aparato de acuerdo con cualquiera de las reivindicaciones 6 o 7, en el que los medios para modificar la excitación comprenden además medios para reemplazar al menos parte de la señal de excitación con una señal de ruido blanco.

9.- Un decodificador de voz que comprende un aparato de acuerdo con cualquiera de las reivindicaciones 6 a 8.

5 10.- Una unidad decodificadora en un sistema de telecomunicaciones que comprende un aparato de acuerdo con cualquiera de las reivindicaciones 6 a 8.

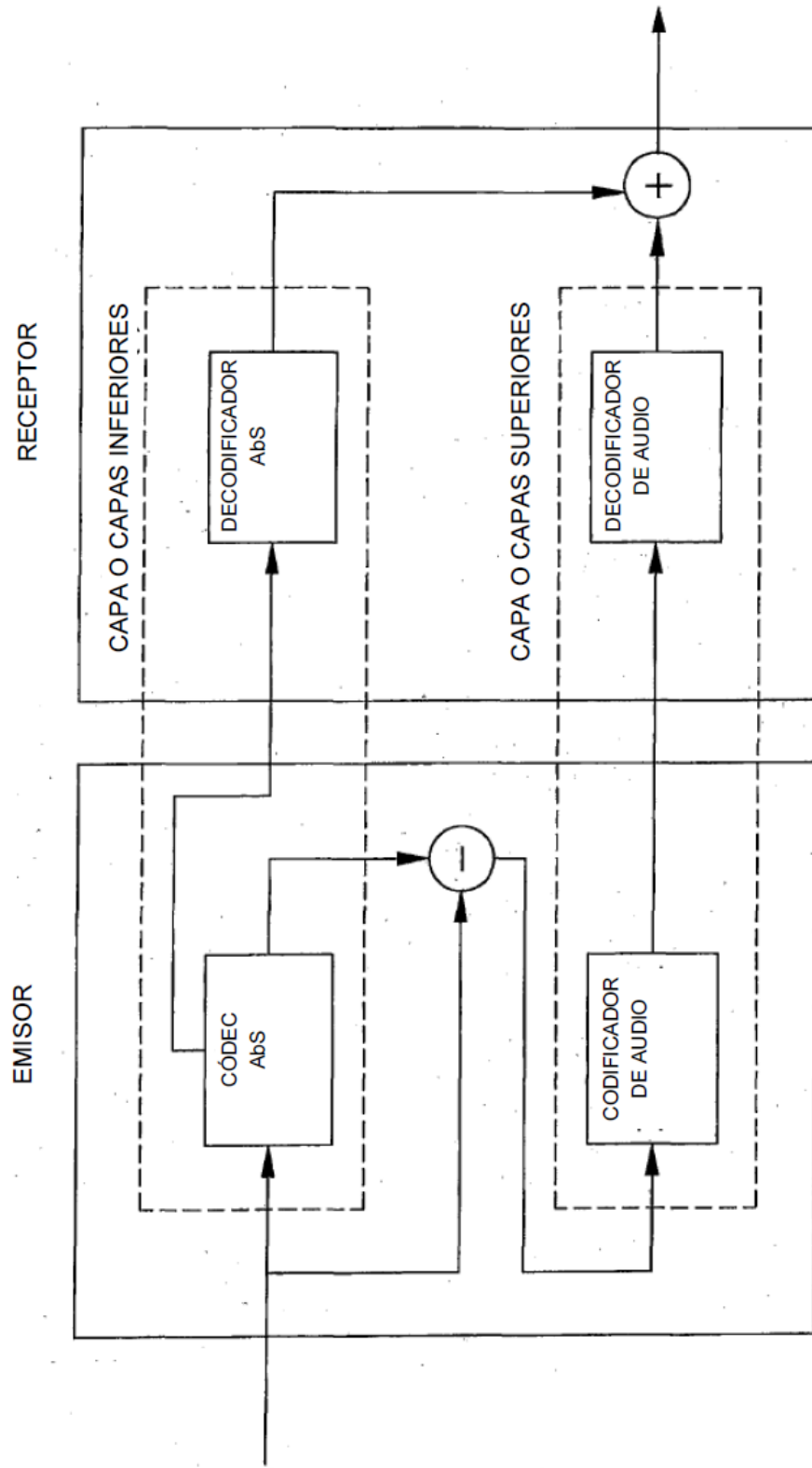


Fig. 1

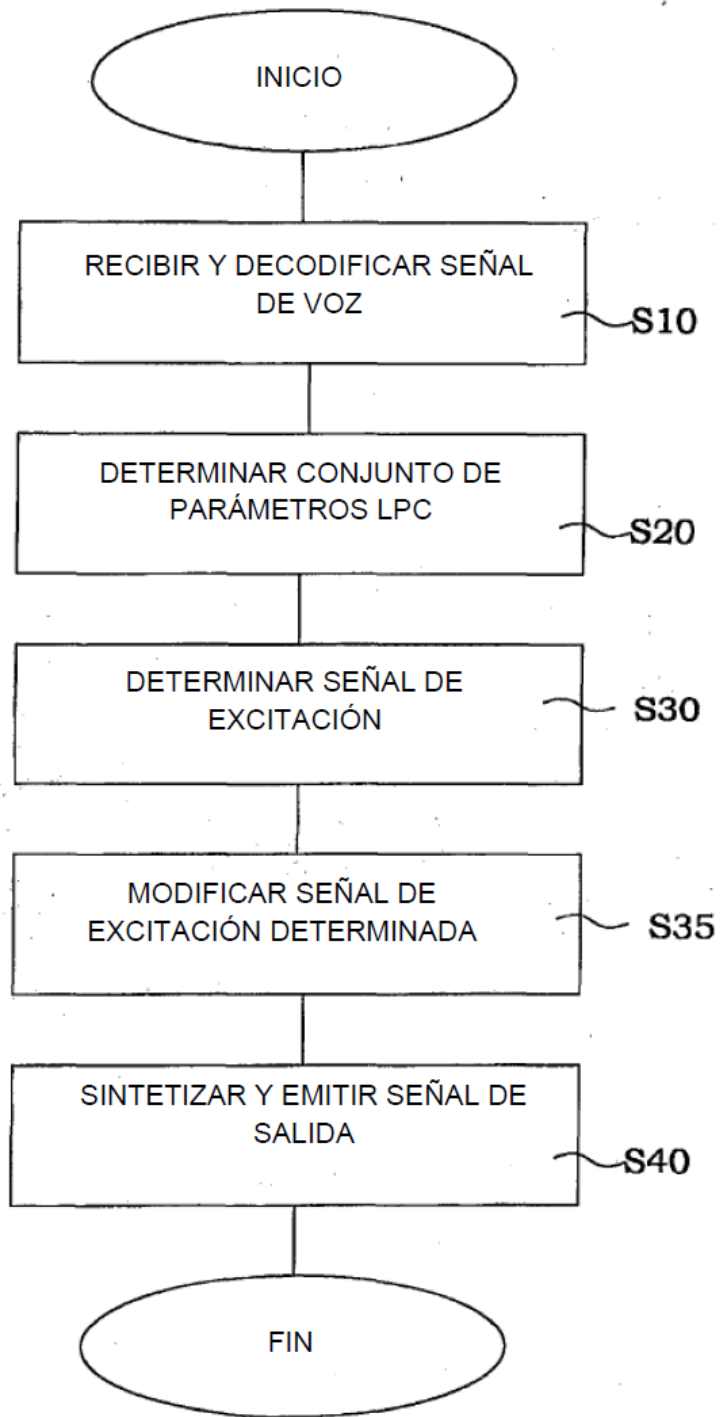


Fig. 2

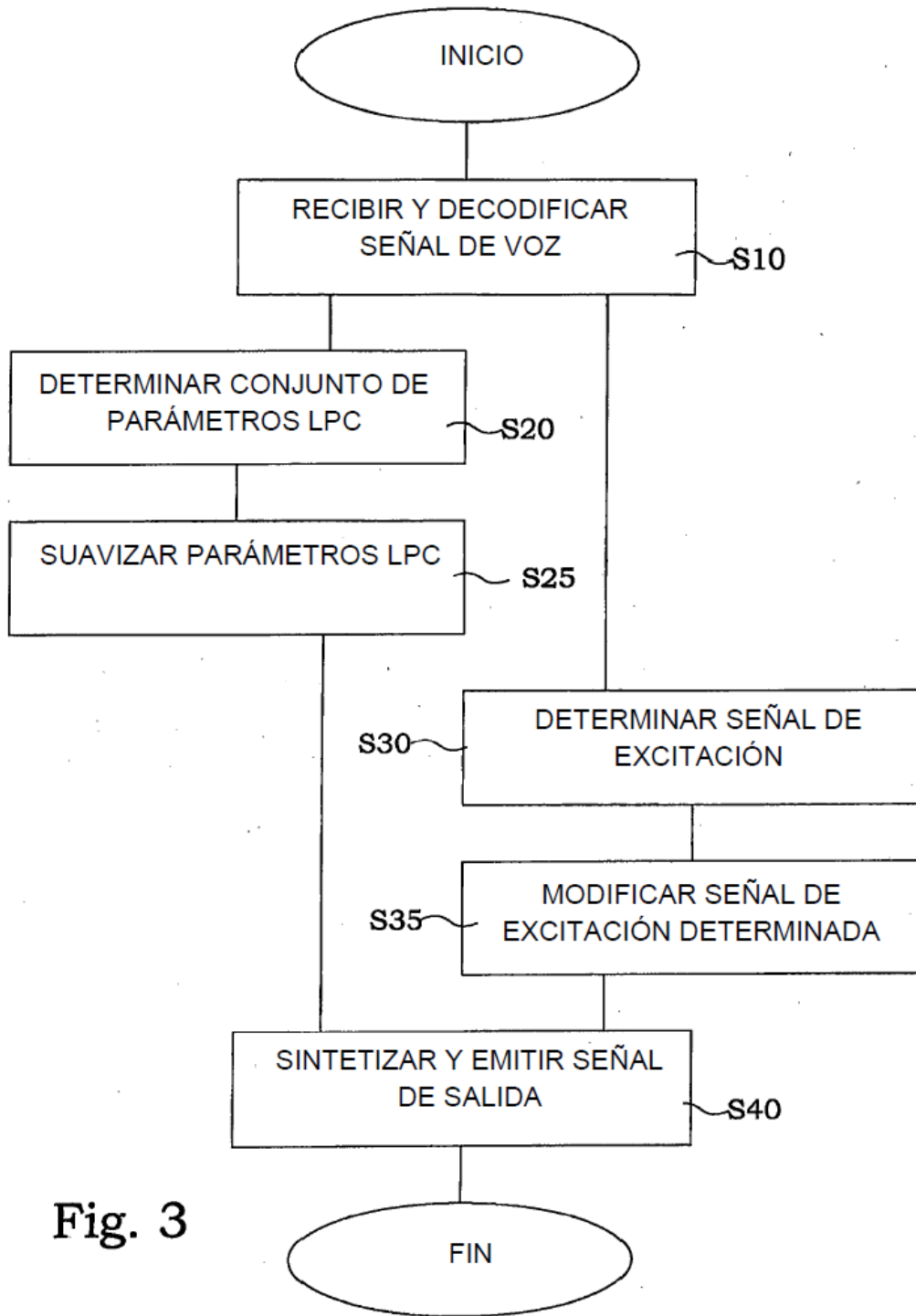


Fig. 3

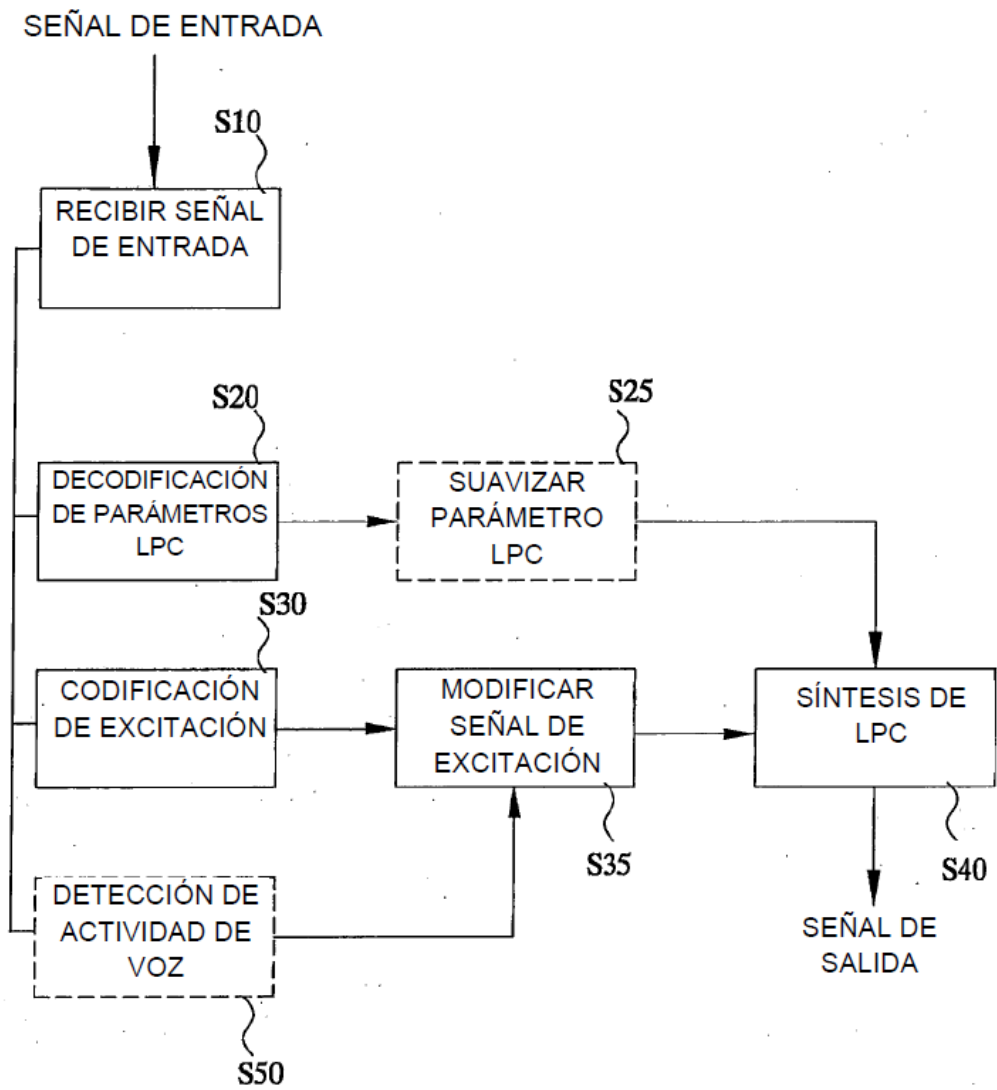


Fig. 4

