



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 292 919**

51 Int. Cl.:
G10L 21/00 (2006.01)
G10L 19/00 (2006.01)
H04M 1/725 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **03291085 .3**
86 Fecha de presentación : **06.05.2003**
87 Número de publicación de la solicitud: **1363272**
87 Fecha de publicación de la solicitud: **19.11.2003**

54 Título: **Terminal de telecomunicaciones que permite modificar la voz transmitida durante una comunicación telefónica.**

30 Prioridad: **16.05.2002 FR 02 06035**

45 Fecha de publicación de la mención BOPI:
16.03.2008

45 Fecha de la publicación del folleto de la patente:
16.03.2008

73 Titular/es:
TCL & ALCATEL MOBILE PHONES LIMITED
Room 1502, 15/F., Tower 6 China Hong Kong City
33 Canto Road, Tsim Sha Tsui
Kowloon, Hong Kong, CN

72 Inventor/es: **Bonnard, Pierre;**
Bourmeyster, Ivan;
Fourquin, Xavier y
Ladouce, Pierre

74 Agente: **Elzaburu Márquez, Alberto**

ES 2 292 919 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

Terminal de telecomunicaciones que permite modificar la voz transmitida durante una comunicación telefónica.

5 La presente invención se refiere a un terminal de telecomunicación que permite modificar la voz transmitida durante una comunicación telefónica, más particularmente adaptado a un sistema de telefonía móvil.

10 Incluso si bien la transmisión de la palabra sigue siendo el elemento esencial de la telefonía móvil, sin embargo los fabricantes buscan diferenciar sus productos ofreciendo nuevos servicios atractivos y de distracción para el consumidor. Los juegos, los servicios ligados al reconocimiento vocal o la multiplicación de los sonidos de llamada son ejemplos.

15 Estos nuevos servicios implican a menudo un coste suplementario en el teléfono ligado a la adición de elementos lógicos o materiales.

La presente invención se dirige a proporcionar un terminal de telecomunicación que ofrece un servicio de modificación de la voz transmitida por el usuario del terminal durante una comunicación telefónica, teniendo este servicio una característica atractiva y de distracción que se ponen en práctica de manera simple y económica.

20 La presente invención propone a este efecto un terminal de telecomunicación *tal como el definido en la reivindicación 1*.

25 El documento WO 98/35340 describe un sistema de conversión de voz en el cual una señal de fuente (voz de un primer usuario) debe ser transformada en una señal de destino (voz de un segundo usuario). Este sistema aplica el principio de los "codebook", es decir de bibliotecas de fonemas registrados previamente para cada usuario posible (fase de aprendizaje), con una correspondencia única entre la biblioteca de fonemas ligada al primer usuario y la biblioteca de fonemas correspondiente al segundo usuario.

30 El documento US 5.956.685 describe un sistema que permite convertir principalmente una señal de palabra en otra señal representativa de una voz modificada. Este sistema descansa sobre el tratamiento de datos de palabras correspondiente a divisiones de la señal de palabra de fuente numerada en función de la puntuación.

35 Gracias a la invención, la colocación de un servicio de modificación de la voz transmitida por un teléfono móvil se hace de manera simple y económica utilizando un codificador de voz ya presente en el teléfono para la codificación de la voz. Los parámetros de reconstrucción son representativos de la voz que se va a transmitir y su modificación permite actuar directamente sobre la sonoridad de la voz transmitida.

40 Ventajosamente, los citados medios para modificar los parámetros de reconstrucción comprenden medios para modificar la frecuencia fundamental de la voz.

De acuerdo con un modo de realización, la citada codificación de fuente de la señal de palabra es una codificación de RPE-LTP (Regular Pulse Excitation Long Term Prediction).

45 De manera ventajosa, los citados medios para modificar los parámetros de reconstrucción comprenden medios para modificar los parámetros de filtro de corto plazo del citado codificador de voz.

Los parámetros de filtro de corto plazo son por ejemplo los coeficientes de reflexión, llamados coeficientes de LAR (Long Area Ratio) determinados durante una codificación del tipo de RPE-LTP.

50 De acuerdo con una variante, los citados medios para modificar los parámetros de reconstrucción comprenden medios para modificar los parámetros de filtro de largo plazo del citado codificador de voz.

55 Un parámetro de filtro de largo plazo es por ejemplo la frecuencia fundamental asociada a un paso (pitch) determinado durante una codificación del tipo de RPE-LTP.

Ventajosamente, los citados medios para modificar los parámetros de reconstrucción comprenden medios para modificar los parámetros asociados a amplitudes de excitación del citado codificador de voz.

60 Un parámetro asociado a una amplitud de excitación puede ser por ejemplo un parámetro que define la señal de impulso de excitación determinada durante una codificación de RPE-LTP.

65 De acuerdo con un modo de realización, los citados medios para modificar los parámetros de reconstrucción comprenden medios para modificar los citados parámetros en función de datos codificados representativos de un conjunto de notas definidas por sus frecuencias y sus duraciones respectivas.

Ventajosamente, el terminal comprende medios de selección de un tipo de modificación de la voz que se va a transmitir.

ES 2 292 919 T3

De manera ventajosa, el terminal comprende medios para configurar manualmente las modificaciones de los citados parámetros de reconstrucción de la señal de palabra.

5 De acuerdo con un modo de realización, el terminal comprende medios para ejecutar las citadas modificaciones en un modo de prueba.

De manera ventajosa, el terminal es un teléfono móvil.

10 Otras características y ventajas de la presente invención se desprenderán de la descripción siguiente de un modo de realización de la invención dado a título ilustrativo y en absoluto limitativo.

En las figuras siguientes:

- 15 • La figura 1 representa esquemáticamente un terminal de telecomunicación de acuerdo con la invención,
- La figura 2 representa esquemáticamente los medios de codificación y de modificación de la voz utilizados en un terminal de telecomunicación de acuerdo con la invención, que utilizan una codificación de RPE-LTP de acuerdo con la norma GSM 06.10.

20 En todas las figuras, los elementos comunes llevan los mismos números de referencia.

La figura 1 representa un terminal de telecomunicación 8 de acuerdo con la invención, tal como un teléfono móvil.

25 El terminal 8 comprende un micrófono 9, un convertidor analógico numérico 10, un codificador de voz 11, un formateador de datos 12, medios 13 para modificar parámetros de reconstrucción de la voz, medios 14 de selección de un tipo de modificación de la voz que se va a transmitir, medios 15 para configurar manualmente las modificaciones de los parámetros de reconstrucción de la señal de palabra y medios 16 para ejecutar modificaciones en un modo de prueba.

30 El codificador de voz 11, el formateador de datos 12 y los medios 13 para modificar los parámetros de reconstrucción de la voz se describirán a continuación en relación con la figura 2.

35 La figura 2 representa un codificador de voz 11 para ejecutar una codificación de RPE-LTP de acuerdo con la norma GSM 06.10, un formateador 12 de flujo de datos y medios 13 de modificaciones de parámetros de reconstrucción de la voz. El codificador de voz 11, el formateador 12 y los medios 13 son medios lógicos ejecutados por un procesador programable no representado.

40 El codificador de voz 11 incluye medios 2 de determinación de coeficientes de reflexión de LAR, medios 3 de determinación de una frecuencia fundamental de la voz y medios 4 de determinación de una señal de impulso de excitación, a partir de una señal S muestreada.

45 El principio de la codificación de RPE-LTP consiste en modelizar la señal de palabra como un filtro lineal cuyos parámetros evolucionan en el tiempo, siendo el citado filtro lineal excitado por una señal de impulso. Por lo tanto, se comprende que el objetivo de esta codificación es determinar la señal de impulso de excitación, definiendo los diferentes parámetros el filtro lineal y la frecuencia fundamental de la voz que se va a transmitir. Este tipo de codificación de la palabra opera en tramas de 20 ms en la banda [300-3400 Hz]. El filtro lineal se descompone en dos partes: una primera parte que tiene en cuenta las correlaciones de corto plazo de la señal de palabra y una segunda parte que tiene en cuenta las correlaciones de largo plazo de la señal de palabra. Se hablará en lo que sigue del filtro de corto plazo y del filtro de largo plazo.

50 La señal S constituye una trama de palabra de 20 ms muestreada a 8 kHz.

55 Los medios 2 permiten determinar los coeficientes de reflexión de LAR del filtro de corto plazo. Se utiliza para esto el algoritmo de Durbin que se aplica a la señal filtrada con el fin de obtener una predicción lineal. El algoritmo de Durbin genera diez coeficientes de reflexión, llamados coeficientes de LAR, comprendidos entre -1 y 1, expresados en la escala logarítmica de LAR (Long Area Ratio).

60 Los medios 3 de determinación de una frecuencia fundamental de la voz utilizan un procedimiento de predicción asociado al filtro de largo plazo en tramas de 5 ms y permiten reproducir la frecuencia fundamental asociada a un paso P comprendido entre 56 y 444 Hz y característico del locutor, así como la ganancia asociada a esta frecuencia.

Los medios 4 permiten determinar la señal de impulso I de excitación.

65 El conjunto de los datos proporcionados por los medios 2, 3 y 4 para una trama de 20 ms es formateado en el formateador 12 de flujo de datos que va a producir una señal de palabra C codificada.

Cuando estos parámetros están definidos, se pueden producir diferentes efectos en la voz modificando los citados parámetros de LAR, P e I. Los medios de modificación 13 permiten modificar estos diferentes parámetros.

ES 2 292 919 T3

Así, los medios 13 de modificación pueden modificar únicamente el valor del paso asociado a la frecuencia fundamental de manera que transformen una voz de hombre en una voz de mujer. Para ello, el valor de paso P de cada trama es modificado en un valor $\text{Min}(P+P_{\text{effet}}, \text{Max}(P))$, siendo P_{effet} un valor predefinido, por ejemplo a 90 Hz, y siendo $\text{Max}(P)$ el valor de paso máximo aceptado por el sistema estándar. En efecto, una voz masculina tiene un valor de paso aproximadamente de 120 Hz mientras que una voz femenina tiene un valor de paso de aproximadamente 210 Hz.

Los medios 13 de modificación pueden proceder de manera similar para transformar una voz de mujer en una voz de hombre. Para ello, el valor de paso P de cada trama es modificado en un valor $\text{Max}(P-P_{\text{effet}}, \text{Min}(P))$, siendo P_{effet} un valor predefinido de 90 Hz y siendo $\text{Min}(P)$ el valor de paso mínimo aceptado por el sistema estándar.

Se puede igualmente obtener un efecto de susurro cuando los medios 13 modifican el paso y la ganancia. Para ello, los medios de modificación 13 aumentan el paso y disminuyen la ganancia de manera que la salida de la señal del filtro de largo plazo sea substancialmente modificada disminuyendo el efecto de las vibraciones de las cuerdas vocales. Los parámetros de LAR permanecen sin cambios.

Se puede obtener también un efecto de voz robotizada cuando los medios 13 modifican los parámetros del filtro de largo plazo. Para ello, el paso se fija en un valor relativamente elevado. Como para el efecto de susurro, la prosodia, es decir el conjunto de los elementos fónicos que caracterizan el lenguaje hablado, es completamente modificada y transformada en un sonido monótono que tiene una frecuencia fundamental única y parámetros de LAR no cambiados.

Otra propiedad notable de los parámetros de los codificadores de voz se refiere al hecho de que los parámetros de LAR representan el trayecto vocal y el trayecto nasal ligados a la producción de palabra, pudiendo estos dos trayectos ser modelizados por filtros de resonancia del tipo de filtro “todo polo”, representando cada polo una frecuencia de resonancia. Por lo tanto, haciendo la producción de la palabra intervenir el trayecto vocal y el trayecto nasal, los medios 13 pueden modificar estos parámetros de manera que las vocales resuenan como las de una persona que tiene la nariz taponada.

El trayecto vocal y el trayecto nasal pueden ser cada uno modelizados por un filtro todo polo pero la suma de estos dos filtros no es todo polo. Para modelizar la combinación de estos dos trayectos en forma de un filtro todo polo, es preciso construir un cero situando dos polos uno al lado de otro. Este cero representa la contribución de la cavidad vocal.

Aproximando estos dos polos uno al lado de otro, los medios de modificación 13 disminuyen la contribución de la cavidad nasal y dan así un efecto “resfriado” a la voz.

Los parámetros pueden igualmente ser modificados en función de datos codificados representativos de un conjunto de notas definidas por sus frecuencias y sus duraciones respectivas. Así la voz modificada puede seguir una partitura definida en un formato de codificación que define un conjunto de pares (frecuencia, duración) y tener así un efecto cantado. Una partitura es proporcionada por los medios 13 de modificación de los parámetros de reconstrucción. Esta partitura puede ser en un formato de codificación de música MIDI, SMAF de Yamaha®, EMR R5 polifónico, IrDA iMelody del IrMC (Infrared Mobile Communications) u otro formato de descripción vectorial de la música. En la etapa de tratamiento previo, la partitura es transcrita automáticamente si es necesario, para que las frecuencias de las notas estén en el intervalo [56 Hz, 444 Hz].

Las notas son interpretadas en duración y en frecuencia, correspondiendo a cada nota una fecha de inicio, una fecha de finalización y una frecuencia. Las tramas de palabra de 5 ms son modificadas por los medios 13 para que el paso de la palabra sea igual a la frecuencia de la nota correspondiente en el mismo instante.

Las modificaciones de los parámetros se hacen generalmente por medio de intercambios entre el formateador de flujo de datos y los medios 13 con el fin de modificar los parámetros. Sin embargo, se pueden igualmente considerar modificaciones de los parámetros hechas directamente al nivel del codificador de voz 11.

En la figura 1, el usuario tiene la posibilidad de seleccionar el tipo de efectos deseado sobre la voz que se va a transmitir por medio de los medios 14 de selección de un tipo de modificación. Puede igualmente configurar manualmente las modificaciones de los parámetros de reconstrucción de la señal de palabra por medio de los medios 15. Así, puede por ejemplo influir sobre el valor del paso modificado a $\text{Min}(P+P_{\text{effet}}, \text{Max}(P))$ en el caso de una transformación de voz masculina en voz femenina fijando a su vez el valor de P_{effet} . Puede igualmente probar el efecto obtenido por medio de los medios 16 que le permiten ejecutar el tipo de modificación seleccionado con los parámetros tal como los ha fijado.

Durante una comunicación móvil, la voz del usuario es primero muestreada y numerada por los medios 9. A continuación es codificada por el codificador de voz 11, el cual realiza la etapa 7 tal como se describe en relación con la figura 1.

La voz codificada es a continuación formateada por el formateador 12 en forma de parámetros de reconstrucción de la voz obtenidos gracias al codificador de voz 11.

ES 2 292 919 T3

Los medios 13, gobernados por los medios 14, 15 y 16 van entonces a modificar los parámetros formateados de reconstrucción de la voz para obtener los efectos buscados por el usuario. Los parámetros modificados son a continuación retransmitidos a los medios 12 y después son enviados al interlocutor por medio de la red de telecomunicación o ejecutados en modo de prueba por el usuario.

5

Se puede igualmente considerar una modificación de los parámetros de reconstrucción directamente al nivel de los medios 11 por los medios 13.

10

Se ha de entender que la invención no está limitada al modo de realización que acaba de describirse.

Particularmente, el modo de realización porta en una codificación de fuente del tipo de RPE-LTP, pero se puede considerar cualquier otra codificación de fuente tal como, por ejemplo, la ETS 300 726 GSM EFR (Enhanced Full Rate) y la 3GPP TS 26.071 AM (Adaptive Multirate).

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

5 1. Terminal (8) de telecomunicación que comprende medios (9) de entrada por el usuario del citado terminal de una señal de palabra analógica, medios (10) para convertir la citada señal de palabra analógica en una señal de palabra numérica (S), un codificador de voz (11) para ejecutar una codificación de fuente de la citada señal de palabra numérica (S), incluyendo el citado codificador de voz (11) medios (2, 3, 4) para extraer parámetros (LAR, P, I) de reconstrucción de la citada señal de palabra numérica y un formateador de datos (12) que entrega una señal de palabra codificada (C) función de los parámetros de reconstrucción extraídos, **caracterizado** porque el citado terminal (8) comprende medios 10 (13) para modificar, durante una comunicación telefónica, al menos uno de los citados parámetros de reconstrucción de manera que la voz transmitida asociada a la citada señal de palabra codificada (C) sea modificada, estando los citados medios gobernados por medios (14) de selección por el usuario de un tipo de modificación de la voz que se va a transmitir y/o por medios (15) para configurar manualmente las modificaciones de los citados parámetros de reconstrucción de palabra.

15 2. Terminal (8) de acuerdo con la reivindicación precedente, **caracterizado** porque los citados medios (13) para modificar los citados parámetros de reconstrucción comprenden medios para modificar un parámetro representativo de la frecuencia fundamental de la voz.

20 3. Terminal (8) de acuerdo con una de las reivindicaciones precedentes, **caracterizado** porque la citada codificación de fuente de la señal de palabra es una codificación de RPE-LTP.

25 4. Terminal (8) de acuerdo con una de las reivindicaciones precedentes, **caracterizado** porque los citados medios (13) para modificar los citados parámetros de reconstrucción comprenden medios para modificar los parámetros de filtro de corto plazo del citado codificador de voz.

5. Terminal (8) de acuerdo con una de las reivindicaciones precedentes, **caracterizado** porque los citados medios (13) para modificar los citados parámetros de reconstrucción comprenden medios para modificar los citados parámetros asociados a amplitudes de excitación.

30 6. Terminal (8) de acuerdo con una de las reivindicaciones precedentes, **caracterizado** porque los citados medios (13) para modificar los citados parámetros de reconstrucción comprenden medios para modificar los parámetros de filtro de largo plazo del citado codificador de voz.

35 7. Terminal (8) de acuerdo con una de las reivindicaciones precedentes, **caracterizado** porque los citados medios (13) para modificar los citados parámetros de reconstrucción comprenden medios para modificar los citados parámetros en función de datos codificados representativos de un conjunto de notas definidas por sus frecuencias y sus duraciones respectivas.

40 8. Terminal (8) de acuerdo con una de las reivindicaciones precedentes, que comprende medios (16) para ejecutar las citadas modificaciones en un modo de prueba.

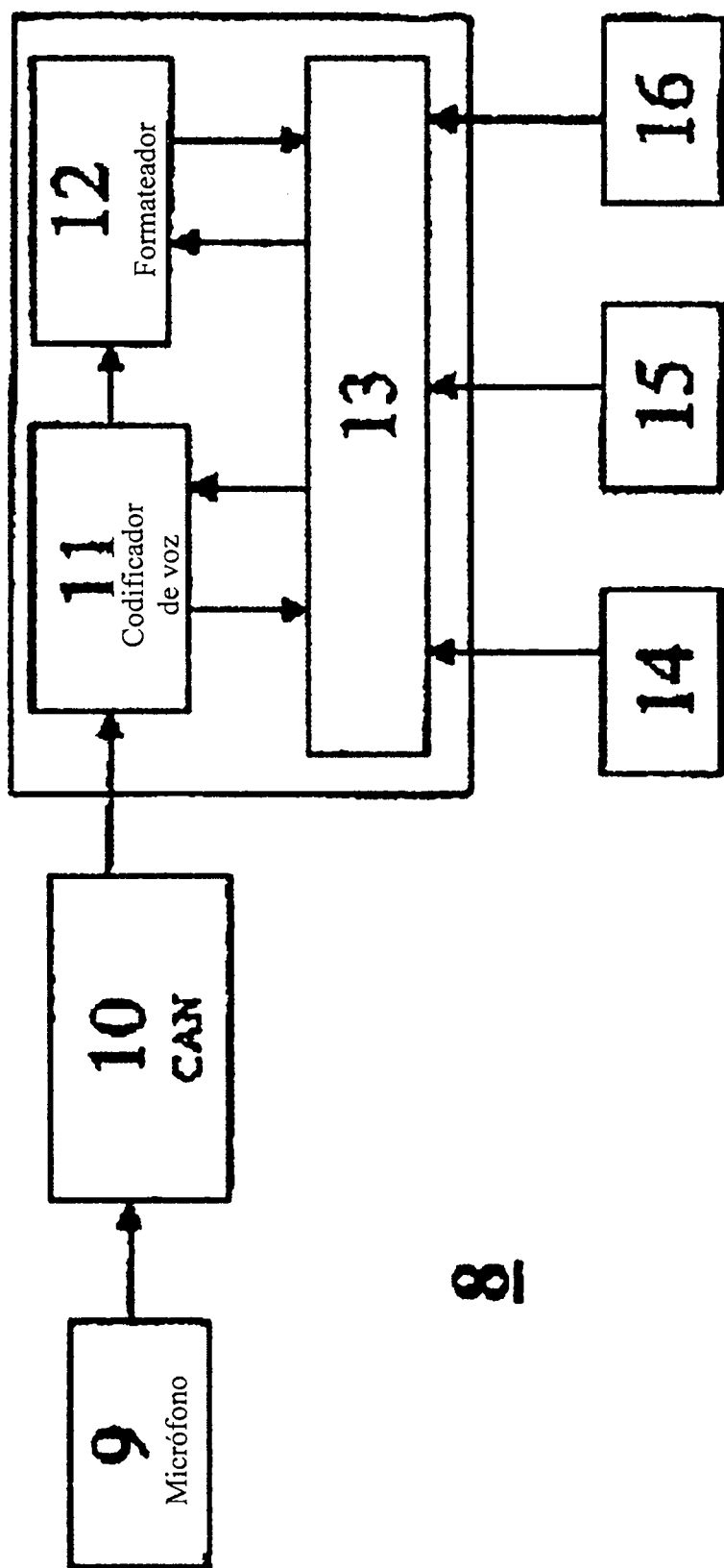
45 9. Terminal de acuerdo con una de las reivindicaciones precedentes, **caracterizado** porque el citado terminal es un teléfono móvil.

50

55

60

65



8

Fig. 1

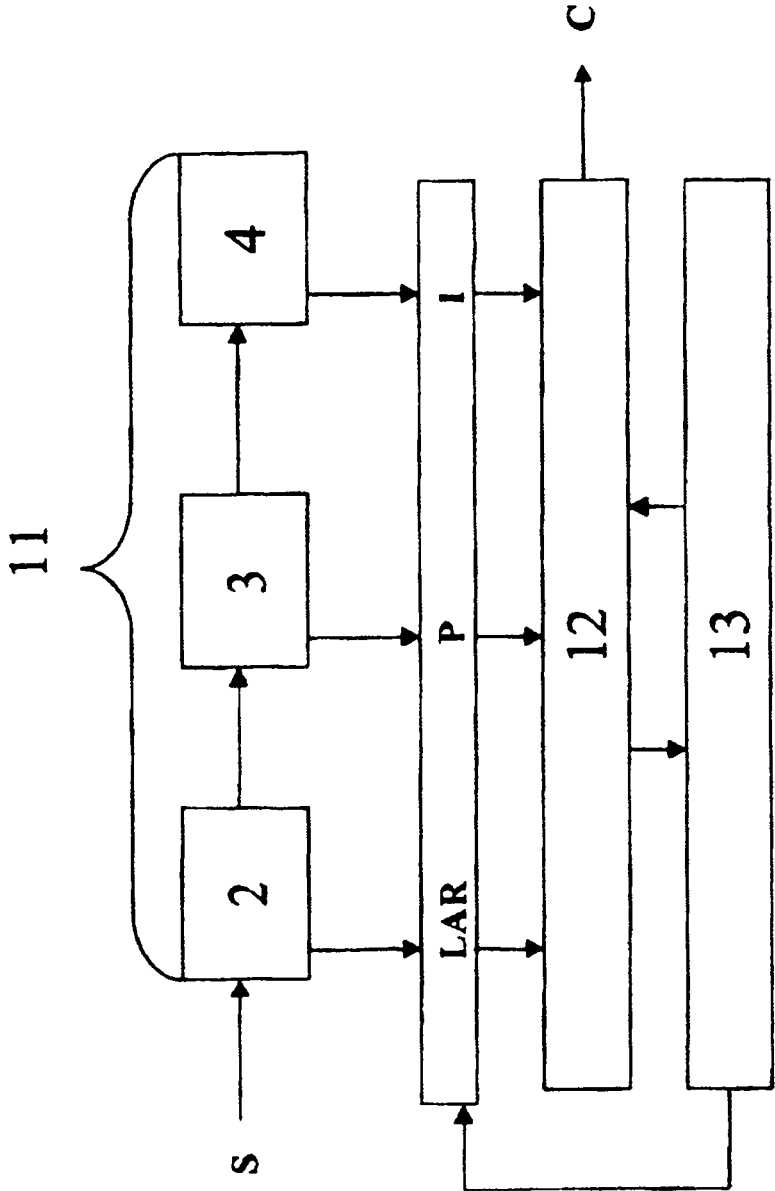


Fig. 2