



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 294 817**

51 Int. Cl.:
H04L 25/03 (2006.01)
G10L 21/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **98939720 .3**
86 Fecha de presentación : **20.07.1998**
87 Número de publicación de la solicitud: **0932964**
87 Fecha de publicación de la solicitud: **04.08.1999**

54 Título: **Procedimiento y dispositivo de ecualización ciega de los efectos de un canal de transmisión sobre una señal de voz digital.**

30 Prioridad: **22.07.1997 FR 97 09273**

45 Fecha de publicación de la mención BOPI:
01.04.2008

45 Fecha de la publicación del folleto de la patente:
01.04.2008

73 Titular/es: **FRANCE TELECOM**
6, place d'Alleray
75015 Paris, FR

72 Inventor/es: **Mauuary, Laurent y**
Monne, Jean

74 Agente: **Justo Vázquez, Jorge Miguel de**

ES 2 294 817 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 de ecualización ciega de los efectos de un canal de transmisión sobre una señal de voz digital.

La invención se refiere a un procedimiento y a un dispositivo de ecualización ciega, de los efectos de un canal de transmisión sobre una señal de voz digital.

La transmisión fiel de una señal de voz es, en el momento actual, un objetivo importante con vistas, especialmente, a la mejora de los rendimientos de los sistemas de reconocimiento automático de voz, sistemas RAP, que funcionan por medio de una red telefónica conmutada o de una red de radiotelefonía móvil, GSM.

La dificultad principal que se opone a la obtención de una tasa de reconocimiento satisfactoria y sensiblemente constante se debe a las variabilidades que existen en la señal acústica que transmite la voz.

Las fuentes de variabilidad son numerosas y se distinguen, habitualmente, dos clases de variabilidad, la variabilidad intrínseca y la variabilidad extrínseca al locutor. En efecto, las creaciones acústicas de una misma palabra difieren, por ejemplo, según el estado del locutor, del contexto de la palabra dentro de la frase. Esta referencia resulta más importante si se procede además a una comparación de las creaciones acústicas provenientes de diversos locutores. Las creaciones acústicas de las palabras son, de hecho, ondas acústicas, las cuales deben ser captadas para ser sometidas a un proceso de reconocimiento. En el momento de la adquisición de una onda acústica, onda de voz, diversas perturbaciones vienen a añadirse a aquélla, lo que provoca el aumento de la variabilidad de la señal captada. El ruido ambiente es, así mismo, constitutivo de una onda acústica captada por el micrófono, y mezclada así aditivamente a la onda de voz. La señal electrónica suministrada por el micrófono es una señal que constituye la suma de las señales correspondientes a la voz y al ruido ambiente.

Por otro lado, en particular respecto de aplicaciones que establecen la red electrónica, el módulo de adquisición formado por el micrófono del teléfono celular y por las líneas telefónicas que unen el usuario al sistema de reconocimiento de voz, actúan como un filtro de convolución lineal, que varía lentamente con el tiempo.

Para un teléfono celular clásico, que no funciona en el modo manos libres, los efectos del ruido ambiente son en general desdenables y son esencialmente los efectos convolutivos de las líneas de transmisión las que deben ser tomadas en consideración. Así, cada señal observada en la entrada del sistema de reconocimiento automático de la voz contiene un componente convolutivo casi constante para una llamada dada pero que varía de una llamada a otra.

Este componente convolutivo perjudica a la eficacia del reconocimiento de la voz. Con el fin de caracterizar adecuadamente su efecto, es indispensable proyectar este componente en el espacio de representación, en el cual el reconocimiento se efectúa, a saber el espacio cepstral en la mayoría de los sistemas de reconocimiento.

A título de ejemplo ilustrativo, puede recordarse, en relación con la figura 1a respecto de la representación cepstral que, corriente arriba de un sistema RAP, un módulo de parametrización transforma la señal de voz digital en una serie de vectores de parámetros, calculados sobre unas ventanas de longitud fija, de 10 a 40 metros, y que se solapan. En principio, el solapamiento es del orden del 50%. Estos vectores de parámetros son escogidos para representar la información más pertinente posible dentro de la ventana de la señal. Tal como se ilustra en la figura 1a, con respecto al principio general de cálculo del cepstro, una transformada, FFT, calcula el espectro de la ventana de la señal. El logaritmo de la energía espectral es calculado a continuación. A partir de este logaritmo, el cepstro $\{C_n(i)\}$ es obtenido por la FFT inversa. Generalmente, no se toma en consideración más que la primera decena de los coeficientes cepstrales. Estos coeficientes retenidos se supone que modelizan la respuesta impulsora del conducto vocal y, por tanto, se supone que incorporan la información pertinente para el proceso de reconocimiento. Además, estos coeficientes son insensibles a la energía de la señal de entrada, característica principal en el cuadro de reconocimiento automático de la voz.

También se utilizan otras representaciones del mismo tipo, en particular con el objeto específico del reconocimiento de la voz. Este es el caso del sistema RAP desarrollado en Francia por el Centro Nacional de Estudios de Telecomunicaciones de FRANCE TELECOM, sistema PHIL90, el cual establece como vectores de parámetros unos MFCC, o "*Mel Frequency based Cepstral Coefficients*". Estos últimos emplean un alisado espectral $\{S_n(f)\}$ para la estimación de la envolvente espectral y de los conocimientos psicoacústicos. El alisado espectral se efectúa por medio de un banco de filtros. El sistema auditivo humano que analiza las bajas frecuencias con una resolución más elevada que las demás frecuencias, mientras que además, en una señal de voz, las bajas frecuencias son más ricas en información que las altas frecuencias, las bandas críticas del banco de filtros se reparten sobre una escala perceptiva no lineal llamada MEL o BARK. El principio de cálculo de los MFCC, con referencia a la figura 1b consiste, después de la transformación frecuencial de la ventana de la señal FTT, después del filtrado MEL, en calcular para cada ventana de la señal un vector constituido por las energías existentes dentro de cada una de las bandas de frecuencias. La transformación secuencial inversa, FTT inversa, realizada por una transformación inversa, suministra los coeficientes MFCC.

En el espacio correspondiente a estos tipos de representación, un filtro convolutivo, que representa el canal de transmisión es transformado en una polarización aditiva, casi constante, la cual está contenida en los vectores cepstra-

les. Para un estudio más detallado de estas representaciones, se podrán consultar con provecho los artículos reseñados a continuación, publicados por:

- H. HERMANSTKY, N. MORGAN, A. BAYYA, P. KOHN "Compensation for the Effect of the Communication Channel in Auditory-like Analysis of Speech", (RASTA-PLP), *Eurospeech*, págs. 1367-1370, Génova 1991;

- C. MOKBEL, D. JOUVET, J. MONNE "Deconvolution of Telephone Line Effects for Speech Recognition", *Speech Communication*, Vol. 19, nº 3, septiembre 1996, págs. 185-196.

La polarización produce e introduce así un cambio de una llamada a otra. Debido a esta transformación y a esta representación, los vectores cepstrales, correspondientes a un sonido determinado, ocupan así una parte amplia, debido a la presencia de la polarización, dentro del espacio de representación. La existencia de esta polarización tiene así como efecto la disminución de la capacidad de discriminación entre vectores cepstrales correspondientes a sonidos diferentes e implica, por esta razón, la puesta en práctica de modelos más complejos con el fin de proceder a la discriminación entre las diferentes formas del vocabulario de aplicación.

En consecuencia, para reducir al máximo los efectos de las líneas telefónicas sobre una señal observada, a la vista de un proceso de RAP por ejemplo, el problema planteado lo es esencialmente en términos de deconvolución ciega de dos señales, pues un solo cepsor terminal está disponible. Sin embargo, la señal de voz de origen y la función de transferencia de canal ocupan las mismas zonas de frecuencias y tienen, de hecho, una gran parte en común dentro del espacio cepstral. Es por tanto particularmente delicado, definir unos filtros, definiéndose los filtros, convencionalmente, como módulos de atenuación o filtros dentro del dominio cepstral, con el fin de reducir o suprimir la contribución de la función de transferencia del canal y, a partir de ahí, realizar la deconvolución perseguida.

Las técnicas actuales llevadas a cabo tienen por objeto, por un lado, en el dominio del reconocimiento automático de la voz, la robustez del proceso de reconocimiento en las condiciones de adquisición de la señal telefónica y, por otro lado, dentro del dominio del tratamiento de la señal, la reducción de las perturbaciones en una señal telefónica con el fin de mejorar la inteligibilidad.

Cuando el proceso de reconocimiento es aplicado localmente para el comando de voz para los teléfonos de manos libres, de las computadoras, de los terminales de información u otros, los esfuerzos destinados a reducir las perturbaciones de la señal se concentran en la reducción de las perturbaciones introducidas por el ruido aditivo. En el marco precitado, las técnicas usuales cubren la definición de las representaciones robustas, el filtrado del tipo de la sustracción espectral, el filtrado de antena, el filtrado por estado de los modelos de MARKOV, o incluso la ayuda en línea del ruido ambiente sobre la señal o los modelos de referencia.

En el marco del procedimiento de reconocimiento centralizado, los esfuerzos realizados se refieren igualmente a la reducción de los efectos de las comunicaciones telefónicas. La técnica utilizada, de forma general, es la sustracción, a los vectores cepstrales, de su componente continuo estimado sobre un horizonte suficientemente amplio. Se indica que la noción de horizonte designa, para una señal telefónica digital subdividida, en ventanas, un número entero de ventanas sucesivas. Esta sustracción puede ser efectuada, ya sea de manera explícita, estimación de la media y sustracción de ésta, ya sea de manera implícita por filtrado de paso alto. Los trabajos recientes han mostrado que la media de los vectores cepstrales sobre un horizonte suficientemente amplio presenta exactamente los efectos de las líneas telefónicas. Confróntese el artículo de C. MOKBEL, J. MONNE y D. JOUVET anteriormente citado.

En el marco general de la deconvolución de señales, se distinguen dos grandes clases de procesos de convolución.

La primera clase, designada como deconvolución ciega, está basada en las propiedades espectrales, cepstrales o incluso temporales de las señales para definir unos esquemas de deconvolución. En el dominio de las telecomunicaciones, los algoritmos de ecualización adaptativa se emparentan con la deconvolución ciega. Para una descripción más detallada de este tipo de algoritmo puede consultarse el artículo publicado por A. BENVENISTE y M. GOURSAT, "Blind Equalizers", *IEEE Transactions on Communications*, Vol. COM-32, nº 8, agosto 1984, págs. 871-883, 1984.

La segunda clase, de forma parecida al proceso puesto en práctica por los algoritmos de anulación de eco o de desreverberación, utiliza el filtrado adaptativo, o el filtrado espacial en el caso de una antena acústica. En tal caso, se dispone en general de varios sensores, como mínimo dos, uno utilizado para la señal de referencia y el otro para la señal de entrada. Para una descripción más detallada de este tipo de filtrado adaptativo se podrá consultar el artículo publicado por B. WIDROW *et al.*, "Adaptive Noise Cancelling: Principles and Applications", *Proc. of IEEE*, vol. 63, nº 12, págs. 1692-1716, diciembre 1975.

En el marco más concreto de la transmisión de señales digitales, los problemas planteados por el proceso de ecualización son de la misma naturaleza, debido al hecho de que es difícil de obtener una señal de referencia propia con la finalidad de llevar a cabo un esquema clásico de filtrado para anular el efecto del canal de transmisión. En efecto, la única señal de que se dispone es la señal digital observada y ya transmitida. Con el fin de simplificar el proceso de ecualización, unas secuencias digitales conocidas del receptor pueden ser emitidas hacia este último, con el fin de identificar la función de transferencia del canal. Sin embargo, un modo operativo de este tipo tiene como efecto saturar rápidamente la capacidad de transmisión del canal.

Con el fin de poner remedio a este inconveniente, se han realizado diferentes trabajos tendentes a establecer un proceso de ecualización ciego. Este procedimiento de ecualización ciego utiliza una lógica de decisión y unas estadísticas a largo plazo conocidas, sobre la señal digital transmitida, con el fin de calcular el error que sirve para poner en evidencia unos coeficientes de un filtro por descenso de gradiente estocástico. El esquema general de un proceso de este tipo se representa en la figura 1c. Para una descripción más detallada de este tipo de proceso, se podrá, por ejemplo, consultar el artículo publicado por J.J. SHYNK, "Frequency-Domain and Multirate Adaptive Filtering", IEEE Signal Processing Magazine, págs. 15-37, enero 1992.

Además, un procedimiento y un sistema de filtrado adaptativo por ecualización ciega de una señal telefónica digital constituye el objeto de la solicitud de patente francesa nº 94 08741 depositada el 13 de julio de 1994, en nombre de France TELECOM.

En el procedimiento y sistema precisados, la señal digital es sometida a una transformación secuencial, FFT, y a un filtrado en sub-bandas. Cada señal en sub-bandas es sometida a un filtrado adaptativo a partir de una señal de referencia basada en unas estadísticas a largo plazo sobre la señal telefónica. Una ecualización por deconvolución ciega de los efectos de la línea telefónica digital es así efectuada.

Un modo operativo de este tipo, basado esencialmente en un proceso de ecualización ciega dentro del dominio frecuencial, y por tanto, espectral, justificado sin embargo por las propiedades estadísticas a largo plazo de la señal que se desplaza sobre la línea telefónica, resulta satisfactorio.

Por último, la solicitud de patente EP 0720149 describe un proceso de reconocimiento de voz en el cual son calculados unos coeficientes cepstrales y un vector de ecualización correspondiente a una estimación aproximada de la deformación de las muestras de voz más recientes. Este vector de ecualización debe pues ser reactualizado para asegurar una sustracción eficaz de la distorsión.

La presente invención tiene sin embargo por objeto el establecimiento de un proceso de ecualización ciego por filtrado adaptativo directamente aplicado al dominio cepstral.

Otro objeto de la presente invención, en razón misma de esta aplicación directa al dominio cepstral, es igualmente una reducción del coste global del cálculo.

Otro objeto de la presente invención es igualmente una reducción del número de "salidas" del filtro llevadas a cabo.

Otro objeto de la presente invención en razón misma de esta aplicación directa del dominio cepstral, es la puesta en práctica de una mejor adecuación del proceso de tratamiento y reconocimiento de voz, especialmente en el sistema PHIL90.

Otro objeto de la presente invención es, por último, la puesta en práctica de un proceso del tratamiento que permita, en ciertas situaciones específicas, una mejora de las tasas de reconocimiento comparadas con las obtenidas gracias a la ecualización ciega en el dominio espectral de la técnica anterior.

El procedimiento y el dispositivo, objetos de la presente invención, son dignos de destacar en el sentido de que la señal de voz $S_n(t)$ es transformada en vectores cepstrales. Cada vector cepstral es sometido a un filtrado adaptativo a partir de un cepstro de referencia, representativo del cepstro a largo plazo de la señal de voz, para engendrar unos vectores cepstrales ecualizados representativos de la señal de voz ecualizada.

El procedimiento y el dispositivo de ecualización ciega de los efectos de un canal de transmisión sobre una señal digital, objetos de la presente invención, se comprenderán mejor con la lectura de la descripción y con la observación de los dibujos adjuntos, en los cuales, además de las figuras 1a a 1c relativas a la técnica anterior:

la figura 2a representa, en forma de esquemas de bloques, un organigrama general del procedimiento objeto de la presente invención;

la figura 2b representa, en forma de esquemas de bloques, un detalle de la puesta en práctica del procedimiento objeto de la invención, tal como se representa en la figura 2a;

la figura 3a representa, en forma de esquema funcional, un dispositivo de ecualización ciega de los efectos de un canal de transmisión de una señal de voz digital, objeto de la presente invención;

las figuras 3b y 3c representan un modo de realización concreto del dispositivo objeto de la invención, representado en la figura 3a.

A continuación se expondrá, en relación con la figura 2a y con la figura 2b, una descripción más detallada del procedimiento de ecualización ciega de los efectos de un canal de transmisión sobre una señal de voz digital.

En la figura 2a precitada, la señal de voz numérica es anotada como $\{S_n(t)\}$, siendo esta señal considerada como capaz de transitar por el canal de transmisión para el cual la ecualización ciega, conforme al procedimiento objeto de la presente invención, debe ser llevada a cabo.

5 En un sentido general, la señal de voz digital está, típicamente, dividida en bloques sucesivos que se solapan eventualmente, designando n el rango del bloque actual y por extensión el rango de todo el conjunto de datos, o trama, obtenidas gracias a la puesta en práctica del procedimiento objeto de la presente invención, a partir de este bloque actual.

10 Conforme a la figura precitada, el procedimiento consiste al menos en someter la señal de voz digital $\{S_n(t)\}$ de una transformación en un conjunto de vectores cepstrales, designándose el vector asociado a la trama de rango n $C_n(i)$, designando la referencia $\{C_n(i)\}$ el componente, o coeficiente cepstral, del rango i de cada vector cepstral, efectuándose esta transformación en una etapa con la referencia 1000 en la figura precitada. El conjunto de los vectores cepstrales es representativo de la señal de voz digital $\{S_n(t)\}$ sobre un horizonte determinado, definiéndose la noción
15 de horizonte conforme a la definición dada anteriormente en la descripción.

La etapa 1000 precitada, después de la cual los vectores cepstrales $\{C_n(i)\}$ son disponibles, es seguida de una etapa con la referencia 1001, la cual consiste en calcular un cepstro de referencia, con la referencia $\{R_n(i)\}$ siendo este cepstro representativo para cada uno de los vectores cepstrales $\{C_n(i)\}$ del conjunto de vectores cepstrales del cepstro a largo plazo de esta señal de voz.

Se indica en efecto que está establecido, con referencia a la publicación de C. MOKBEL, D. JOUVET y J. MONNE anteriormente mencionada en la descripción, que el cepstro a largo plazo de la señal de voz es casi constante en todas las frecuencias, siendo la noción de frecuencia, dentro del dominio cepstral, el análogo de la noción de
25 frecuencia en el dominio espectral. Además, y con referencia a la publicación anteriormente mencionada, la media de los logaritmos de las densidades espectrales y, por tanto, también la media de los coeficientes espectrales, sobre un horizonte suficientemente amplio, representa una componente convolutiva constante en la señal observada, la cual puede ser válidamente asimilada al efecto del canal de transmisión.

30 En consecuencia, la etapa 1001 precitada va seguida de una etapa 1002 consistente en someter cada uno de los vectores cepstrales $\{C_n(i)\}$ a un filtrado adaptativo a partir del cepstro de referencia $\{R_n(i)\}$ para engendrar un conjunto de vectores cepstrales ecualizados, con la referencia $\{C_n(i)\}$, en los cuales el efecto del canal de transmisión es esencialmente suprimido. Este conjunto de vectores cepstrales ecualizados es representativo de una señal de voz digital ecualizada.

35 Así, conforme al procedimiento objeto de la presente invención, el filtrado adaptativo realizado en la etapa 1002 es conducido a partir de la entrada de referencia constituida por el cepstro de referencia $\{R_n(i)\}$, siendo el filtrado adaptativo por supuesto implantado en el dominio cepstral y aplicado sobre los coeficientes cepstrales $C_n(i)$.

40 Por lo que respecta al establecimiento del filtrado adaptativo precitado, se indica que este filtrado puede ser un filtrado adaptativo de tipo LMS, o *Least Mean Square*, habiendo sido descrito este tipo de filtrado en la publicación anteriormente citada, publicada por J.J. SCHINK.

45 Sin embargo, conforme a un aspecto particularmente ventajoso del procedimiento objeto de la presente invención, el filtrado adaptativo precitado es aplicado en el dominio cepstral y no en el dominio espectral, tal como se realiza en la técnica anterior.

En consecuencia, y conforme a un aspecto particularmente ventajoso del procedimiento objeto de la presente invención:

50 - La ecualización se basa así en el hecho de que la ecualización cepstral a largo plazo de esta señal puede ser aproximada por una constante.

55 Para una descripción más detallada de los elementos relativos a la representación cepstral a largo plazo de la señal, en particular sobre las estadísticas relativas a esta representación sobre la posibilidad de identificar el canal de transmisión utilizando el cepstro a largo plazo, se podrá con provecho consultar el artículo publicado por C. MOKBEL, D. JOUVET y J. MONNE anteriormente mencionado en la descripción.

60 - El filtrado adaptativo es aplicado directamente en el dominio cepstral, lo que por supuesto indica la relación de un tratamiento por bloques de la señal digital considerada, tal como se ha mencionado anteriormente en la descripción.

A continuación se expone un justificante del proceso de convolución ciega por filtrado adaptativo, conforme al objeto de la presente invención.

65 Suponiendo que la señal observada $s(k)$ es decir la señal digital transmitida $\{S_n(t)\}$ es el producto de la convolución de una señal limpia, es decir el producto de la señal de la palabra emitida $x(k)$ por el filtro que identifica el canal telefónico del canal de transferencia $w(k)$, la señal observada $s(k)$ verifica la relación (1):

$$s(k) = x(k) * w(k)$$

En esta relación, el operador * representa el operador producto de la convolución.

5 *Deconvolución en el dominio espectral*

Con el fin de introducir la deconvolución en el dominio cepstral, se introduce primero un recordatorio teórico de la deconvolución en el dominio espectral.

10 Con referencia a la relación (1) precitada, la expresión de la densidad espectral de potencia de los dos términos de la relación (1) precitada, se escribe, para cada ventana o bloque de la señal de voz, según la relación (2):

$$15 \quad S_s(f) = S_x(f)W^2(f)$$

En esta relación, $S_s(f)$, $S_x(f)$ representan respectivamente las densidades espectrales de potencia de la señal observada $s(k)$ y de la señal de voz emitida $x(k)$, mientras que $W(f)$ representa la función de transferencia del canal telefónico. Se recuerda que $W(f)$ designa de hecho la transformada de Fourier del filtro que identifica el canal telefónico $w(k)$.

20 Teniendo en cuenta las relaciones precedentes, un filtro adaptativo de función de transferencia $H(f)$ puede ser directamente aplicado sobre la densidad espectral de potencia de la señal observada $S_s(f)$ con el fin de obtener el espectro ecualizado $S_n(f)$ en el cual el efecto de la línea telefónica o del canal de transmisión ha sido suprimido.

25 En esas condiciones, el espectro ecualizado $S_n(f)$ verifica la relación (3):

$$S_n(f) = S_s(f)H(f) = S_x(f)W^2(f)H(f)$$

30 A partir de un espectro plano constante constitutivo de una señal de referencia $R(f)$, el error $E(f)$ para cada ventana de la señal observada especifica la relación (4):

$$35 \quad E(f) = R(f) - S_x(f)W^2(f)H(f) .$$

El filtro óptimo hacia el cual la función de transferencia $H(f)$ converge es el que permite reducir al mínimo el error cuadrático medio, consignado como EQM, en cada una de las bandas de frecuencia f dentro de la cual la descomposición en bandas de frecuencia, descomposición espectral, ha sido realizada. El error cuadrático medio, $EQM(f)$, verifica la relación (5):

$$45 \quad EQM(f) = E[E^2(f)] .$$

Teniendo en cuenta algunas hipótesis ampliamente verificadas, el espectro a largo plazo de la señal de voz constante y la función de transferencia del canal telefónico $W(f)$ constante sobre un amplio horizonte, el filtro óptimo es el que permite reducir al mínimo la expresión dada por la relación (6):

$$50 \quad EQM(f) = R^2(f) + \overline{S_x^2(f)}W^4(f)H^2(f) - 2R(f)\overline{S_x(f)}W^2(f)H(f)$$

55 cualquiera que sea el valor de f , es decir en el conjunto de las ondas de frecuencia en el cual la señal observada ha sido descompuesta.

La reducción al mínimo del error cuadrático medio, $EQM(f)$ dado por la relación (6) precedente permite obtener la función de transferencia del filtro óptimo $H_{opt}(f)$ la cual verifica la relación (7):

$$60 \quad H_{opt}(f) = \frac{R(f)\overline{S_x(f)}}{\overline{S_x^2(f)}} \frac{1}{W^2(f)} = Cte \frac{1}{W^2(f)} .$$

65 El filtro óptimo así obtenido permite compensar el efecto del canal de transmisión, es decir de la comunicación telefónica.

Por otro lado, si se considera una señal de referencia $R(f)$ específica, es decir una señal de referencia que tenga la misma potencia que la señal, la expresión $E[R(f)S_x(f)]$ tiende entonces a ser igual a $E[S_x^2(f)]$ y, en estas condiciones, el filtro óptimo se aproxima al inverso del canal de transmisión.

5 Deconvolución en el dominio cepstral

Por analogía, en el dominio cepstral, la relación (3) precedente se escribe según la relación (8):

$$10 \quad C_n(i) = C_s(i) + C_H(i) = C_x(i) + C_w(i) + C_H(i) .$$

En esta relación, $C_n(i)$, $C_s(i)$, $C_x(i)$, $C_w(i)$ y $C_H(i)$ representan respectivamente el cepstro ecualizado, el cepstro de la señal observada, el cepstro de la señal de voz limpia, es decir, antes de la transmisión por la línea telefónica o por el canal de transmisión, la retransmisión cepstral del efecto de la línea y del cepstro ecualizador.

15 Teniendo en cuenta la puesta en práctica de un cepstro constante $R(i)$ como referencia, el error $E(i)$ para cada ventana de la señal observada verifica la relación (9):

$$20 \quad E(i) = R(i) - (C_x(i) + C_w(i) + C_H(i))$$

El filtro óptimo hacia el cual converge la función de transferencia $H(i)$ del filtro es el que permite reducir al mínimo el error cuadrático medio, $EQM(i)$ en cada una de las referencias según la relación (10):

$$25 \quad EQM(i) = E[E^2(i)]$$

30 Teniendo en cuenta ciertas hipótesis similares a las hipótesis tomadas en cuenta en el dominio frecuencial, siendo expresadas estas hipótesis en el dominio frecuencial, es decir el cepstro a largo plazo de voz constante, la representación cepstral del efecto de la línea de transmisión constante sobre un amplio horizonte, el filtro óptimo es el que permite reducir al mínimo el error cuadrático medio, y el que, en consecuencia, verifica la relación (11):

$$35 \quad C_{Hopt}(i) = R(i) \quad \overline{C_x(i)} \quad C_w(i) = Cte \quad C_w(i)$$

El filtrado óptimo aplicado en el dominio cepstral permite así compensar el efecto del canal de transmisión.

40 Cuando se escoge el cepstro de referencia $R(i)$ igual al valor medio del cepstro a largo plazo consignado como $\overline{C_x(i)}$, el filtrado óptimo se aproxima entonces inverso del canal de transmisión.

45 La comparación de aproximación correspondiente al proceso de ecualización ciega conforme al procedimiento de la presente invención, con la aproximación clásica del filtrado paso alto o de la sustracción cepstral muestra de hecho que desde el punto de vista de los rendimientos de reconocimiento, este modo operativo iguala sustancialmente los rendimientos de la técnica anterior y a veces los sobrepasa de manera significativa en ciertas bases de datos, como se describirá de forma más detallada más adelante en la descripción.

50 Por lo que respecta a la etapa 1000 de transformación de la señal digital observada $\{s_n(t)\}$ en un conjunto de vectores cepstrales, a continuación se ofrecerá, en relación con la figura 2b, una descripción detallada del proceso de transformación propiamente dicho.

55 Según la figura precitada, la etapa de transformación 1000 consiste sucesivamente en someter a una etapa 1000a la señal de voz digital $\{S_n(t)\}$ en una transformada frecuencial que suministre un espectro de frecuencia $\{S_n(f)\}$ de la señal de voz digital $\{s_n(t)\}$ sobre el horizonte considerado. Se recuerda en efecto que la señal numérica observada $\{s_n(t)\}$ está subdividida en bloques sucesivos de muestras con el fin de efectuar un tratamiento por bloques. La transformada frecuencial utilizada puede, por ejemplo, estar constituida por una transformada de Fourier rápida.

60 La propia etapa 1000a es seguida de una etapa 1000b consistente en someter el espectro de frecuencia $\{S_n(f)\}$ obtenido gracias a la transformada frecuencial puesta en práctica en la etapa 1000a en una descomposición en sub-bandas de frecuencias, para engendrar una pluralidad de señales en sub-bandas de frecuencias consignadas como $\{V_n(j)\}$. Se indica que el índice j designa aquí el rango de cada sub-banda de frecuencias considerada. De una manera práctica, se indica que la descomposición en sub-bandas de frecuencias realizada en la etapa 1000b puede ser puesta en práctica por medio de un banco de filtros de frecuencias adaptado, pudiendo el espectro de la señal $\{S_n(f)\}$ por ejemplo ser dividido en 24 bandas de frecuencias adyacentes.

La propia etapa 1000b anteriormente citada es seguida de una etapa 1000c consistente en someter cada señal en sub-bandas de frecuencias, es decir cada señal $V_n(j)$ en una atenuación logarítmica para engendrar una pluralidad de señales en sub-bandas de frecuencias atenuadas, consignadas para esta razón $\{LV_n(j)\}$.

El conjunto atenuado de las señales de las sub-bandas de frecuencias $\{LV_n(j)\}$ es a continuación sometido a una transformación frecuencial inversa para engendrar el conjunto de vectores espectrales consignados como $\{C_n(i)\}$. La transformada frecuencial inversa es por ejemplo realizada por medio de una transformada de Fourier rápida inversa cuando la transformada frecuencial realizada en la etapa 1000a es realizada bajo la forma de una transformada de Fourier rápida directa.

Por lo que respecta al cálculo del cepstro de referencia $\{R_n(i)\}$ se indica que, de manera ventajosa, este puede estar constituido por una señal de cepstro de valor constante para cada uno de los vectores cepstrales $\{C_n(i)\}$. Esta señal de cepstro constante es representativa, dentro de las señales de voz digital y del conjunto de los vectores cepstrales, del cepstro a largo plazo de la señal de voz. De forma general, se indica que el cepstro de referencia puede ser obtenido a partir de una base de datos como se describirá ulteriormente en la descripción.

A continuación se ofrecerá, en relación con las figuras 3a, 3b, y 3c, una descripción más detallada de un dispositivo de ecualización ciega de los efectos de un canal de transmisión sobre una señal de voz digital que permite la puesta en práctica del procedimiento.

Tal como se representa en la figura 3a, y para una señal de voz digital $\{s_n(t)\}$ que se desplaza en el marco de una comunicación telefónica, el dispositivo conforme al objeto de la presente invención incluye al menos un módulo 1 de transformación de la señal de voz digital $\{s_n(t)\}$ en un conjunto de vectores cepstrales $\{C_n(i)\}$ representativo de la señal de voz digital sobre el horizonte determinado y en particular para la ventana de rango n de la señal digital de voz observada. El módulo 1 precitado va seguido de un módulo 2 generador de un cepstro de referencia representativo para cada uno de los valores cepstrales $\{C_n(i)\}$ del cepstro a largo plazo de la señal de voz. El cepstro de referencia permite asociar a cada vector cepstral $\{C_n(i)\}$ un cepstro de referencia $\{R_n(i)\}$ representativo del cepstro a largo plazo de la señal de voz para cada uno de los vectores, en unas condiciones que serán explicitadas ulteriormente en la descripción. Se comprende así que el valor del cepstro de referencia podrá ser ventajosamente aproximado por una constante en el tiempo. Sin embargo, la distribución $R_n(i)$ de este valor para cada componente $C_n(i)$ de rango i de los vectores cepstrales puede ser diferente en función del rango i considerado. Sin embargo, en un modo de realización particular no limitativo, el cepstro de referencia $\{R_n(i)\}$ tenía un valor constante idéntico para cada componente, o coeficiente cepstral $C_n(i)$ de rango i correspondiente, componente constitutivo de cada vector cepstral $\{C_n(i)\}$.

Un módulo 3 de filtrado adaptativo está previsto a partir del cepstro de referencia de cada uno de los vectores cepstrales $\{C_n(i)\}$, permitiendo este módulo 3 engendrar unos vectores cepstrales ecualizados en los cuales el efecto de la comunicación telefónica es sustancialmente suprimido.

Se comprende en particular que el modo de filtrado 3 adaptativo permite, a partir del cepstro de referencia $\{R_n(i)\}$ representativo del cepstro a largo plazo para esta señal de voz para cada uno de los vectores cepstrales, efectuar un filtrado por ejemplo del tipo LMS que permite, a partir del cálculo de error entre el cepstro de referencia $\{R_n(i)\}$ y el conjunto de los vectores cepstrales ecualizados, engendrar unos vectores cepstrales ecualizados $\{\hat{C}_n(i)\}$. El conjunto de los vectores cepstrales ecualizados es representativo de una señal de voz cepstral ecualizada.

Como se ha representado en la figura 3a, el módulo 1 de transformación de la señal de voz digital en un conjunto de vectores cepstrales incluye al menos, recibiendo la señal de voz observada $\{S_n(t)\}$, un módulo 10 de transformada frecuencial de la señal de voz digital que suministra un espectro de frecuencias de la señal de voz digital consignada $\{S_n(f)\}$ sobre el horizonte considerado. Un banco de filtros 11 de descomposición en N sub-bandas de frecuencias del espectro de frecuencias de la señal de voz digital, suministrando este banco de filtros 11 N señales en sub-bandas de frecuencias consignadas como $\{V_n(j)\}$.

El propio banco de filtros 11 va seguido de un módulo 12 de atenuación logarítmica de cada señal en sub-bandas de frecuencias, suministrando este módulo 12 una pluralidad de señales en sub-bandas de frecuencias atenuadas consignadas como $\{LV_n(j)\}$.

Por último, el módulo 1 incluye igualmente un módulo 13 de transformada frecuencial inversa que recibe las señales en sub-bandas de frecuencias atenuadas $\{LV_n(j)\}$ que permiten engendrar, a partir de las señales en sub-bandas de frecuencias atenuadas, el conjunto de vectores cepstrales $\{C_n(i)\}$.

Por lo que respecta al módulo 2 generador del cepstro de referencia, se indica que, para un conjunto de vectores cepstrales $\{C_n(i)\}$ con $i \in [1, M]$, el cepstro de referencia $\{R_n(i)\}$ es representativo para cada uno de los vectores cepstrales del cepstro a largo plazo de la señal de voz. Se comprende así que la señal de datos constitutiva del generador del cepstro de referencia 2 puede estar organizada de forma que suministre el cepstro de referencia representativo del cepstro a largo plazo de la señal de voz en función del índice i que designa la componente del vector cepstral $\{C_n(i)\}$.

Así mismo, según se representa en la figura 3a, el módulo 3 de filtrado adaptativo incluye al menos un módulo 30 de cálculo de señal de voz entre cada vector cepstral ecualizado $\{\hat{C}_n(i)\}$ correspondiente y el cepstro de referencia

correspondiente $\{R_n(i)\}$. La señal de error $E(i) = R_n(i) - \hat{C}_n(i)$ es calculada entre cada componente de rango i del cepstro de referencia y del vector cepstral ecualizado. Así mismo, se prevé un módulo de ecualización 31 de cada vector cepstral $\{C_n(i)\}$, permitiendo este módulo de ecualización engendrar de hecho, a partir de cada componente $C_n(i)$ y de esta señal de error, un vector cepstral ecualizado $\{\hat{C}_n(i)\}$ correspondiente, en las condiciones que serán explicitadas más adelante en relación con las figuras 3b y 3c. De hecho, la señal de error puede ser ponderada o adaptada.

En la figura 3b, se ha representado el dispositivo objeto de la presente invención, en un modo de realización particular orientado a una aplicación en el sistema de reconocimiento automático de voz PHIL90 anteriormente mencionado en la descripción.

Por supuesto, las mismas referencias designan los mismos elementos estando, sin embargo, el módulo 1 de la figura 3a, organizado para realizar el cálculo de los MFCC anteriormente designado en la descripción y suministrando, en consecuencia, el vector cepstral correspondiente $\{C_n(i)\}$ de componentes $C_n(1)$, $C_n(i)$ a $C_n(M)$ para cada ventana sucesiva de la señal digital de voz observada. Los módulos 2 y 3 de la figura 3b representan módulos parecidos a los de los módulos 2 y 3 de la figura 3a. No obstante, el módulo de cálculo de error entre el cepstro de referencia $\{R_n(i)\}$ y cada vector cepstral ecualizado $\{\hat{C}_n(i)\}$ es explicitado y representado como 3i como relativo a cada componente $\hat{C}_n(i)$ del vector cepstral ecualizado $\{\hat{C}_n(i)\}$, siendo cada módulo 3i idéntico y asegurando de hecho, para cada componente $C_n(i)$ el vector cepstral $\{C_n(i)\}$ la misma función de cálculo del error y de ecualización de esta componente para suministrar una componente del vector cepstral ecualizado correspondiente $\{\hat{C}_n(i)\}$.

Como se representa en la figura 3c, cada módulo 3i incluye ventajosamente un circuito sustractor 30i que recibe, por un lado, la componente $R_n(i)$ del cepstro de referencia $\{R_n(i)\}$ correspondiente sobre su límite positivo, así como, por otro lado, el coeficiente cepstral ecualizado $\hat{C}_n(i)$ sobre un límite negativo para asegurar el cálculo de error según una señal de error $E(i)$.

Cada módulo 3i recibe además el coeficiente cepstral correspondiente $C_n(i)$ sobre un circuito sumador 31i, el cual permite de hecho realizar la ecualización de este coeficiente cepstral $C_n(i)$ para suministrar un coeficiente cepstral ecualizado $\hat{C}_n(i)$.

Así mismo, como se representa en la figura 3c, cada módulo 3i, y en particular el modo de ecualización de cada coeficiente cepstral $C_n(i)$, comprende ventajosamente un circuito multiplicador 300i para un coeficiente multiplicador μ , recibiendo este circuito multiplicador la señal de error $E(i)$ suministrada por un circuito sustractor 30i y suministrando una señal de error ponderada $E^*(i)$. El circuito multiplicador 300i va seguido por un circuito sumador 301i de dos entradas y una salida, recibiendo una primera entrada del primer circuito sumador 301i la señal de error operada $E^*(i)$ y recibiendo una segunda entrada de este circuito sumador la señal suministrada por este último, por medio de un circuito retardador 302i. El circuito retardador 302i introduce de hecho un retraso igual a la duración de un bloque de muestras de señal digital. La salida del circuito retardador 302i suministra una señal de adaptación $H_n(i)$, la cual es enviada sobre el circuito sumador de ecualización 31i. El circuito de adaptación así constituido por el circuito multiplicador 300i, el circuito sumador 301i y el circuito retardador 302i, permite así adaptar o ponderar la señal de error $E(i)$ para suministrar la señal de adaptación $H_n(i)$. Sobre la adaptación de la señal de error realizada por el circuito de adaptación, el circuito sumador de ecualización 31i suministra entonces el coeficiente cepstral ecualizado $\hat{C}_n(i)$ correspondiente. Como se observará en la figura 3b, lo mismo ocurre para todas las componentes de los vectores cepstrales de rango i comprendidos entre 1 y M del modo de realización considerado.

El dispositivo objeto de la presente invención, tal como se representa en las figuras 3b y 3c, ha sido puesto en práctica y ensayos comparativos han permitido verificar la técnica de ecualización ciega propuesta sobre diversas bases de datos específicas.

Los ensayos han sido realizados tomando como entrada un fichero de diez coeficientes cepstrales, la aplicación del filtrado adaptativo y la salvaguardia de los vectores MFCC como salida del filtro así obtenidos. Los ficheros MFCC son directamente utilizados por el sistema PHIL90.

El cuadro ofrecido más abajo presenta las mejoras obtenidas respecto de una técnica clásica, designada como sistema de base, un proceso de sustracción cepstral tal como se define, por ejemplo, en la técnica anterior por C. MOKBEL, J. MONNE y D. JOUVET anteriormente citada en la descripción, un filtrado adaptativo en el dominio espectral tal como se realiza, por ejemplo, conforme al procedimiento y sistema descritos en la solicitud de patente francesa No. 94 08741 anteriormente citada en la descripción y, por último, un dominio adaptativo en el dominio cepstral conforme al procedimiento y el dispositivo objetos de la presente invención.

Estas técnicas son aplicadas sobre cuatro bases de datos diferentes, designadas en el cuadro precitado, sucesivamente, como Dígitos, Números de dos dígitos, Trégor y Baladins.

Las tres primeras bases de datos son bases de datos llamadas de laboratorio cuyos registros son los de los locutores prevenidos y cooperantes.

ES 2 294 817 T3

La última base de datos, Baladins, es una base de datos llamada de explotación obtenida registrando las llamadas a un servidor en explotación. Las condiciones de registro de las bases de datos de explotación son más próximas a las condiciones reales de explotación. Los resultados de reconocimiento sobre estas bases de datos son, por tanto, más conformes a los rendimientos obtenidos en explotación. Las mejoras, última columna del cuadro, indicadas como Reducción del porcentaje de error, son ofrecidas con respecto a la versión de referencia del sistema PHIL90. La horquilla indicada entre corchetes al lado del porcentaje de error del sistema de base representa el intervalo de confianza en un 95%.

Por lo que respecta al proceso de sustracción cepstral, este proceso aparece como llevándose a cabo al menos las bases de datos Dígitos y Números de dos dígitos. No obstante, la realización de este proceso plantea un problema en línea en tiempo real, pues esta implantación se basa en la estimación de la media de los vectores cepstrales en el silencio o en un amplio horizonte de voz.

Dígitos	Tasa de error	Reducción de la tasa de error
Sistema de base	0,63% [0,44 a 0,90]	
Sustracción cepstral	0,40%	36,7%
Filtrado adaptativo (dominio espectral)	0,52%	16,7%
Filtrado adaptativo (dominio cepstral)	0,46%	26,7%
Números de dos dígitos	Tasa de error	Reducción de la tasa de error
Sistema de base	3,2% [2,82 a 3,63]	
Sustracción cepstral	2,5%	21,1%
Filtrado adaptativo (dominio espectral)	2,6%	18,5%
Filtrado adaptativo (dominio cepstral)	2,5%	21,1%
Trégor	Tasa de error	Reducción de la tasa de error
Sistema de base	0,40% [0,30 a 0,53]	
Sustracción cepstral	0,34%	13,7%
Filtrado adaptativo (dominio espectral)	0,33%	17,6%
Filtrado adaptativo (dominio cepstral)	0,26%	35,3%
Baladins	Tasa de error	Reducción de la tasa de error
Sistema de base	3,8% [2,44 a 3,68]	
Sustracción cepstral	3,2%	12,7%
Filtrado adaptativo (dominio espectral)	3%	19,1%
Filtrado adaptativo (dominio cepstral)	2,7%	29,1%

En lo que respecta a la realización efectiva del dispositivo objeto de la presente invención, se indica que éste puede, por supuesto, utilizar estructuras ya puestas en práctica en el marco del sistema PHIL90 anteriormente mencionado en la descripción.

El dispositivo de ecualización, conforme al objeto de la presente invención, apreciado en el espacio de los coeficientes cepstrales, es muy poco costoso en tiempo de cálculo. En efecto, exige para cada vector de parámetros, cada 16 ms, M multiplicaciones y 3M adiciones donde M designa por supuesto el número de coeficientes cepstrales. El número de operaciones necesarias resulta, por tanto, igual a una multiplicación y tres adiciones por salida de filtro. El coste de cálculo es sensiblemente desdeñable con respecto al volumen de cálculo llevado a la práctica en el momento de la determinación de los vectores de parámetros.

Teniendo en cuenta estas indicaciones, se indica que el coste de cálculo del filtro puesto en práctica es débil y que la implantación de este último no plantea, por tanto, ningún problema de cálculo en tiempo real.

Con respecto a la solución propuesta por la solicitud de patente francesa No. 94 08741, el coste en volumen de cálculo se reduce, pues no hay más que una multiplicación y tres adiciones por salida de filtro, en lugar de tres mul-

ES 2 294 817 T3

tiplicaciones y dos adiciones en la solución anteriormente mencionada, mientras que, además, el filtro, en el dominio cepstral, incluye menos salidas que en el dominio espectral. Así mismo, siendo la dinámica de los coeficientes cepstrales más débil que la de los coeficientes espectrales y que la de los vectores correspondientes, la precisión en número de elementos de código asignados a las variables precitadas exigida en los cálculos es, por tanto, menos importante. Los resultados obtenidos son equivalentes o ligeramente superiores; es, en particular, el caso para las bases de datos mencionadas en el cuadro precitado respecto de la técnica observada en la solicitud de Patente nº 94 08741 anteriormente citada.

10

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Procedimiento de ecualización ciega de los efectos de un canal de transmisión sobre una señal de voz digital ($\{S_n(t)\}$) desplazándose sobre el canal de transmisión, en el que se somete dicha señal de voz digital a una transformación en un conjunto de vectores cepstrales, siendo el conjunto de vectores cepstrales representativo de dicha señal de voz digital sobre un horizonte determinado, siendo definido un horizonte determinado como un número entero de ventanas sucesivas aplicadas a dicha señal de voz digital, **caracterizado** porque dicho procedimiento consiste además:
- en calcular un cepstro de referencia ($\{R_n(i)\}$) representativo, para cada uno de los vectores cepstrales de dicho conjunto de vectores cepstrales, del cepstro a largo plazo de esa señal de voz;
 - en someter cada uno de los vectores cepstrales a un filtrado adaptativo a partir de dicho cepstro de referencia, para engendrar un conjunto de vectores cepstrales ecualizados ($\{\hat{C}_n(i)\}$) en el cual el efecto del canal de transmisión es sustancialmente suprimido, siendo dicho conjunto de vectores cepstrales representativo de una señal de voz digital ecualizada.
2. Procedimiento según la reivindicación 1, **caracterizado** porque dicha transformación consiste sucesivamente:
- en someter dicha señal de voz digital ($\{S_n(t)\}$) en una transformada frecuencial que suministra un espectro de frecuencias ($\{S_n(f)\}$) de dicha señal de voz digital ($\{S_n(t)\}$) sobre el horizonte considerado;
 - en someter dicho espectro de frecuencias ($\{S_n(f)\}$) en una descomposición de sub-bandas de frecuencias para engendrar una pluralidad de señales en sub-bandas de frecuencias ($\{V_n(j)\}$);
 - en someter cada señal en sub-bandas de frecuencias en una atenuación logarítmica para engendrar una pluralidad de señales en sub-bandas de frecuencias atenuadas ($\{LV_n(j)\}$);
 - en someter el conjunto de las señales en sub-bandas de frecuencias atenuado en una transformación frecuencial inversa para engendrar dicho conjunto de vectores cepstrales ($\{C_n(i)\}$).
3. Procedimiento según las reivindicaciones 1 ó 2, **caracterizado** porque el cepstro de referencia ($\{R_n(i)\}$) está constituido por una señal de cepstro constante para cada uno de los vectores cepstrales, representativo, en la señal de voz digital y en el conjunto de los vectores cepstrales, del cepstro a largo plazo de la señal de voz.
4. Procedimiento según la reivindicación 3, **caracterizado** porque el cepstro de referencia ($\{R_n(i)\}$) tiene un valor constante idéntico para cada componente, o coeficiente cepstral ($C_n(i)$) de rango i correspondiente, constitutivo de cada vector cepstral ($\{C_n(i)\}$).
5. Procedimiento según una de las reivindicaciones 1 a 4, **caracterizado** porque dicho filtrado adaptativo consiste, para cada uno de los vectores cepstrales a partir de dicho cepstro de referencia ($\{R_n(i)\}$) representativo del cepstro a largo plazo de esta señal de voz para cada uno de los vectores cepstrales, en efectuar un filtrado de tipo LMS, permitiendo dicho filtrado de tipo LMS, a partir del cálculo de error entre dicho cepstro de referencia ($\{R_n(i)\}$) y el conjunto de dichos vectores cepstrales ecualizados, engendrar dichos vectores cepstrales ecualizados.
6. Dispositivo de ecualización ciega de los efectos de un canal de transmisión sobre una señal de voz digital ($\{S_n(t)\}$) que se desplaza sobre este canal de transmisión que comprende unos medios de transformación de la señal de voz digital en un conjunto de vectores cepstrales, siendo el conjunto de vectores cepstrales representativo de dicha señal de voz digital sobre un horizonte determinado, siendo definido un horizonte determinado como un número entero de ventanas sucesivas aplicadas a dicha señal de voz digital, **caracterizado** porque comprende al menos:
- unos medios generadores del cepstro de referencia, representativo, para cada uno de los vectores cepstrales de dicho conjunto de vectores cepstrales, del cepstro a largo plazo de esta señal de voz;
 - unos medios de filtrado adaptativo, a partir de dicho cepstro de referencia, de cada uno de los vectores cepstrales que permiten engendrar unos vectores cepstrales ecualizados en los cuales el efecto de dicho canal de transmisión es sustancialmente suprimido, siendo dicho conjunto de vectores cepstrales ecualizados representativo de una señal de voz digital ecualizada.
7. Dispositivo según la reivindicación 6, **caracterizado** porque dichos medios de transformación de dicha señal de voz digital en un conjunto de vectores cepstrales incluyen al menos:
- unos medios de transformada frecuencial de dicha señal de voz digital, que suministran un espectro de frecuencias de dicha señal de voz digital ($\{S_n(t)\}$) sobre el horizonte considerado;
 - un banco de filtros de descomposición en N sub-bandas de frecuencias de dicho espectro de frecuencias de dicha señal de voz digital que suministra N señales en sub-bandas de frecuencias ($\{V_n(j)\}$);

- unos medios de atenuación logarítmica de cada señal en sub-bandas de frecuencias que suministran una pluralidad de señales en sub-bandas de frecuencias atenuadas ($\{LV_n(j)\}$),

- unos medios de transformada frecuencial inversa que permiten engendrar, a partir de dichas señales en sub-bandas de frecuencias atenuadas ($\{LV_n(j)\}$), dicho conjunto de vectores cepstrales ($\{C_n(i)\}$).

8. Dispositivo según una de las reivindicaciones 6 ó 7, **caracterizado** porque, para cada vector cepstral ($\{C_n(i)\}$), $i \in [1, M]$ y para un cepstro de referencia ($\{R_n(i)\}$) representativo, para cada uno de los vectores cepstrales, del cepstro a largo plazo de esta señal de voz, dichos medios de filtrado adaptativo incluyen al menos:

- unos medios de cálculo de una señal de error, $E(i)$, entre cada componente de rango i , $\hat{C}_n(i)$, de cada vector cepstral ecualizado ($\{\hat{C}_n(i)\}$) y la componente correspondiente de mismo rango, $R_n(i)$, del cepstro de referencia ($\{R_n(i)\}$),

$$E(i) = R_n(i) - \hat{C}_n(i),$$

y

- unos medios de ecualización de dicho vector cepstral ($\{C_n(i)\}$) que suministran a partir de cada componente ($C_n(i)$) de cada vector cepstral ($\{C_n(i)\}$) y de dicha señal de error ($E(i)$) una componente ($\hat{C}_n(i)$) del vector cepstral ecualizado ($\{\hat{C}_n(i)\}$).

9. Dispositivo según la reivindicación 8, **caracterizado** porque dichos medios de cálculo de la señal de error ($E(i)$) y dichos medios de ecualización de dicho vector cepstral ($\{C_n(i)\}$) comprenden, para cada componente ($C_n(i)$) de dicho vector cepstral:

- un circuito sustractor que recibe dicha componente ($R_n(i)$) del cepstro de referencia ($\{R_n(i)\}$) y dicha componente ($C_n(i)$) de dicho vector cepstral ecualizado ($\{\hat{C}_n(i)\}$) y que suministra dicha señal de error ($E(i)$);

- unos medios de adaptación de esta señal de error ($E(i)$) que incluyen:

- un circuito multiplicador para un coeficiente multiplicador μ que suministra una señal de error ponderada ($E^*(i)$);
- un circuito sumador con dos entradas y una salida, recibiendo una primera entrada dicha señal de error ponderada ($E^*(i)$) y recibiendo una segunda entrada la señal suministrada por dicho primer circuito sumador por medio de un circuito retardador con una duración determinada, suministrando la salida de dicho circuito retardador una señal de adaptación ($H_n(i)$);

- un circuito sumador de ecualización que recibe dicho coeficiente cepstral ($C_n(i)$) y dicha señal de adaptación ($H_n(i)$) y que suministra dicho coeficiente cepstral ecualizado ($\hat{C}_n(i)$).

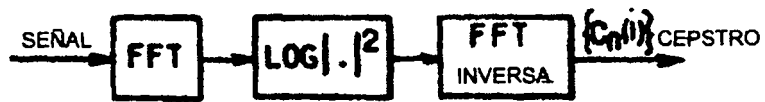


FIG.1a. (TECNICA ANTERIOR)
CALCULO DE CEPSTRADO



FIG.1b. (TECNICA ANTERIOR)
CALCULO DE LOS MFCC

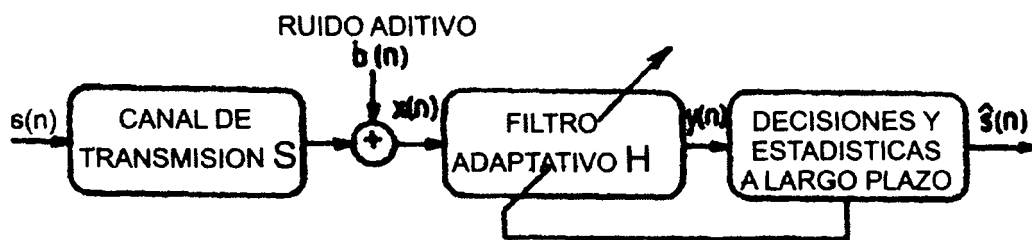


FIG.1c. (TECNICA ANTERIOR)
ECUALIZADOR AUTODIDACTA (CIEGO)
UTILIZADO EN TRANSMISION DIGITAL

FIG.2a.

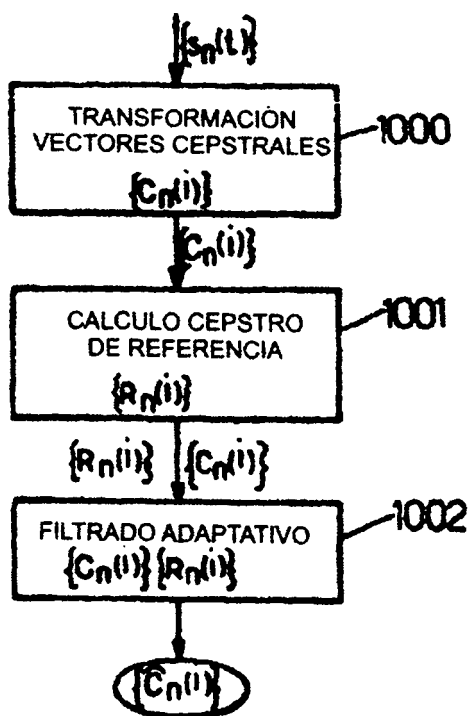


FIG.2b.

