(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 570 961**

(51) Int. CI.:

G10L 21/0208 (2013.01)
G10L 25/18 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **14.03.2008 E 08726859 (5)**

(97) Fecha y número de publicación de la concesión europea: **09.03.2016 EP 2137728**

(54) Título: **Estimación de varianza de ruido para mejorar la calidad de voz**

(30) Prioridad:

19.03.2007 US 918964 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
23.05.2016

(73) Titular/es:

DOLBY LABORATORIES LICENSING CORPORATION (100.0%)
100 POTRERO AVENUE
SAN FRANCISCO, CA 94103-4813, US

(72) Inventor/es:

YU, RONGSHAN

(74) Agente/Representante:

LEHMANN NOVO, María Isabel

DESCRIPCIÓN

Estimación de varianza de ruido para mejorar la calidad de voz

Campo técnico

La invención se refiere al procesamiento de señales de audio. Más en particular, se refiere a la clarificación y a la mejora de la calidad de voz en un entorno ruidoso.

Se hace referencia a las siguientes publicaciones:

[1] "*Speech enhancement using a minimum mean square error short time spectral amplitude estimator*", de Y. Ephraim y D. Malah, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, páginas 1109 a 1121, diciembre de 1984.

[2] "*Single channel speech enhancement based on masking properties of the human auditory system*", de N. Virag, *IEEE Tran. Speech and Audio Processing*, vol. 7, páginas 126 a 137, marzo de 1999.

[3] "*Spectral subtraction based on minimum statistics*", de R. Martin, *Proc. EUSIPCO*, 1994, páginas 1182 a 1185.

[4] "*Efficient alternatives to Ephraim and Malah suppression rule for audio signal enhancement*", de P. J. Wolfe y S. J. Godsill, *EURASIP Journal on Applied Signal Processing*, vol. 2003, n.º 10, páginas 1043 a 1051, 2003.

[5] "*A brief survey of Speech Enhancement*", de Y. Ephraim, H. Lev-Ari y W. J. J. Roberts, *The Electronic Handbook, CRC Press*, abril de 2005.

Técnica anterior

Vivimos en un mundo ruidoso. El ruido ambiental, que procede tanto de fuentes naturales como de actividades humanas, nos rodea. Durante una comunicación de voz, los ruidos ambientales se transmiten junto con la señal de voz deseada, lo que afecta negativamente a la calidad de la señal recibida. Este problema se mitiga mediante técnicas de mejora de la calidad de voz que eliminan tales componentes de ruido no deseadas, generándose así una señal más limpia e inteligible.

Las mayoría de sistemas de mejora de la calidad de voz se basan en varios tipos de una operación de filtrado adaptativo. Tales sistemas atenúan las regiones de tiempo/frecuencia (T/F) de la señal de voz ruidosa que presentan bajas relaciones de señal a ruido (SNR), mientras que conservan aquellas con una alta SNR. Por tanto, se conservan las componentes esenciales de voz mientras que se reduce considerablemente la componente de ruido. Normalmente, una operación de filtrado de este tipo se lleva a cabo en el dominio digital mediante un dispositivo de cálculo, tal como un chip de procesamiento de señales digitales (DSP).

El procesamiento en el dominio de subbanda es una de las maneras preferidas en las que se implementa tal operación de filtrado adaptativo. Expresado brevemente, la señal de voz no modificada en el dominio de tiempo se transforma en varias subbandas usando un banco de filtros, tal como la transformada discreta de Fourier (DFT). Después, las señales dentro de cada subbanda se suprimen en una cantidad deseable según propiedades estadísticas conocidas de voz y ruido. Finalmente, las señales sin ruido en el dominio de subbanda se transforman al dominio de tiempo usando un banco de filtros inverso para generar una señal de voz mejorada, cuya calidad depende en gran medida de los detalles del procedimiento de supresión.

Un ejemplo de un dispositivo de mejora de la calidad de voz de la técnica anterior se muestra en la FIG. 1. La entrada de datos se genera digitalizando una señal de voz analógica que contiene voz nítida y ruido. Esta señal de audio no modificada $y(n)$, donde $n = 0,1,...,\infty$ es el índice de tiempo, se envía después a un dispositivo o función de banco de filtros de análisis ("banco de filtro de análisis") 2, generando múltiples señales de subbanda, $Y_k(m)$, $k = 1,...,K$, $m = 0,1,...,\infty$, donde k es el número de subbanda y m es el índice de tiempo de cada señal de subbanda. Las señales de subbanda pueden tener frecuencias de muestro más bajas en comparación con $y(n)$ debido a la operación de muestreo descendente en el banco de filtro de análisis 2. El nivel de ruido de cada subbanda se estima después usando un dispositivo o función de estimación de varianza de ruido ("estimador de varianza de ruido") 4 con la señal de subbanda como entrada. El estimador de varianza de ruido 4 de la presente invención es diferente de los estimadores conocidos de la técnica anterior y se describe posteriormente, en particular con respecto a las FIG. 2a y 2b. En función de nivel de ruido estimado se determinan ganancias de supresión apropiadas g_k en un dispositivo o función de reglas de supresión ("regla de supresión") 6 y se aplican a las señales de subbanda de la siguiente manera:

$$\tilde{Y}_k(m) = g_k Y_k(m),\ k = 1,...,K. \qquad (1)$$

Tal aplicación de la ganancia de supresión a una señal de subbanda se muestra de manera simbólica mediante un símbolo de multiplicación 8. Finalmente, $\tilde{Y}_k(m)$ se envían a un dispositivo o función de banco de filtros de síntesis

("banco de filtro de síntesis") 10 para generar una señal de voz mejorada $\tilde{y}(n)$. Para una mayor claridad en la ilustración, la FIG. 1 muestra los detalles para generar y aplicar una ganancia de supresión a solamente una de múltiples señales de subbanda (k).

La cantidad de supresión apropiada para cada subbanda depende en gran medida de su nivel de ruido. A su vez, esto se determina mediante la varianza de la señal de ruido, definida como el valor cuadrático medio de la señal de ruido con respecto a una distribución de probabilidad gaussiana de media cero. Evidentemente, una estimación precisa de la varianza de ruido es crucial para el rendimiento del sistema.

Normalmente, la varianza de ruido no está disponible *a priori*, sino que debe estimarse a partir de la señal de audio no modificada. Es bien sabido que la varianza de una señal de ruido "limpia" puede estimarse calculando el promedio en el tiempo del valor cuadrático de las amplitudes de ruido durante un gran periodo de tiempo. Sin embargo, puesto que la señal de audio no modificada contiene voz nítida y ruido, este procedimiento no puede aplicarse directamente.

Se han propuesto muchas estrategias de estimación de la varianza de ruido para resolver este problema. La solución más sencilla es estimar la varianza de ruido en la fase de inicialización del sistema de mejora de la calidad de voz, cuando la señal de voz no está presente (referencia [1]). Sin embargo, este procedimiento solo funciona correctamente cuando la señal de ruido y la varianza de ruido son relativamente estacionarias.

Se han propuesto procedimientos más sofisticados para tratar de manera precisa el ruido no estacionario. Por ejemplo, los estimadores de detección de actividad de voz (VAD) usan un detector autónomo para determinar la presencia de una señal de voz. La varianza de ruido solo se actualiza cuando no está presente (referencia [2]). Este procedimiento tiene dos inconvenientes. En primer lugar, es muy difícil obtener resultados VAD fiables cuando la señal de audio es ruidosa, lo que a su vez afecta a la fiabilidad del resultado de la estimación de la varianza de ruido. En segundo lugar, este procedimiento excluye la posibilidad de actualizar la estimación de la varianza de ruido cuando la señal de voz está presente. Esto último hace que el sistema sea ineficaz, ya que la estimación de la varianza de ruido puede actualizarse de manera fiable aun cuando el nivel de voz es débil.

Otra solución a este problema citada con mucha frecuencia es el procedimiento de estadísticas mínimas (referencia [3]). En principio, el procedimiento registra el nivel de señal de muestras históricas para cada subbanda y estima la varianza de ruido basándose en el valor mínimo registrado. La base de este enfoque es que la señal de voz es generalmente un proceso de todo o nada que, naturalmente, tiene pausas. Además, el nivel de señal es normalmente mucho mayor cuando la señal de voz está presente. Por lo tanto, el nivel de señal mínimo del algoritmo se obtiene probablemente a partir de una sección de pausa de voz si el registro tiene una duración suficientemente larga, dando como resultado un nivel de ruido estimado fiable. Sin embargo, el procedimiento de estadísticas mínimas consume mucha memoria y no puede aplicarse en dispositivos cuya memoria disponible es limitada.

El artículo "*Speech enhancement for non-stationary noise environments*" de Israel Cohen y Baruch Berdugo, *Signal Processing*, vol. 81, 2001, páginas 2403 a 2418, da a conocer un estimador de voz y un enfoque de estimación de ruido para una mejora robusta de la calidad de voz. El espectro variable en el tiempo del ruido se estima usando una técnica que aplica un suavizado recursivo temporal a la medición ruidosa durante periodos de ausencia de voz. El espectro de ruido se estima calculando el promedio de valores de potencia espectral anteriores, usando un parámetro de suavizado que se ajusta por la probabilidad de presencia de voz.

Resumen de la invención

La presente invención está definida por las reivindicaciones independientes. Las reivindicaciones dependientes se refieren a características opcionales de algunas formas de realización de la invención.

Según un primer aspecto de la invención, se mejoran las componentes de voz de una señal de audio formada por componentes de voz y de ruido. Una señal de audio se transforma desde el dominio de tiempo a una pluralidad de subbandas en el dominio de frecuencia. Después, las subbandas de la señal de audio se procesan. El procesamiento incluye reducir de manera adaptativa la ganancia de algunas de la subbandas en respuesta a un control. El control se obtiene, al menos en parte, a partir de una estimación de la varianza en componentes de ruido de la señal de audio. A su vez, la estimación se obtiene a partir del promedio de estimaciones previas de la amplitud de las componentes de ruido de la señal de audio. Las estimaciones de la amplitud de las componentes de ruido de la señal de audio que tienen un sesgo de estimación mayor que una cantidad máxima predeterminada de sesgo de estimación se excluyen de o se ponderan con un valor bajo en el promedio de las estimaciones previas de la amplitud de las componentes de ruido de la señal de audio. Finalmente, la señal de audio procesada se transforma del dominio de frecuencia al dominio de tiempo para proporcionar una señal de audio en la que se han mejorado las componentes de voz. Este aspecto de la invención puede incluir además una estimación de la amplitud de las componentes de ruido de la señal de audio en función de una estimación de la varianza en componentes de ruido de la señal de audio, una estimación de la varianza en componentes de voz de la señal de audio y la amplitud de la señal de audio.

Según un aspecto adicional de la invención, se obtiene una estimación de la varianza en componentes de ruido de una señal de audio formada por componentes de voz y de ruido. La estimación de la varianza en componentes de ruido de una señal de audio se obtiene a partir del promedio de estimaciones previas de la amplitud de las componentes de ruido de la señal de audio. Las estimaciones de la amplitud de las componentes de ruido de la señal de audio que tienen un sesgo de estimación mayor que una cantidad máxima predeterminada de sesgo de estimación se excluyen de o se ponderan con un valor bajo en el promedio de las estimaciones previas de la amplitud de las componentes de ruido de la señal de audio. Este aspecto de la invención puede incluir además una estimación de la amplitud de las componentes de ruido de la señal de audio en función de una estimación de la varianza en componentes de ruido de la señal de audio, una estimación de la varianza en componentes de voz de la señal de audio y la amplitud de la señal de audio.

Según cualquiera de los aspectos anteriores de la invención, las estimaciones de la amplitud de las componentes de ruido en la señal de audio que tienen valores superiores a un umbral en el promedio de estimaciones previas de la amplitud de las componentes de ruido de la señal de audio pueden excluirse o ponderarse con un valor bajo.

El umbral mencionado anteriormente puede ser una función de $\psi(1+\hat{\xi}(m))\hat{\lambda}_d(m)$, donde $\hat{\xi}$ es la relación de señal a ruido estimada *a priori*, $\hat{\lambda}_d$ es la varianza estimada en componentes de ruido de la señal de audio y ψ es una constante determinada por la cantidad máxima predeterminada de sesgo de estimación.

Los aspectos de la invención descritos anteriormente pueden implementarse como procedimientos o como aparatos adaptados para llevar a cabo tales procedimientos. Un programa informático, almacenado en un medio legible por ordenador, puede hacer que un ordenador lleve a cabo cualquiera de tales procedimientos.

Un objeto de la presente invención es proporcionar una mejora de la calidad de voz que pueda estimar las intensidades relativas de las componentes de voz y de ruido que están operativas durante la presencia y la ausencia de voz.

Un objeto adicional de la presente invención es proporcionar una mejora de la calidad de voz que pueda estimar las intensidades relativas de las componentes de voz y ruido a pesar de la presencia de una componente de ruido significativa.

Un objeto adicional de la presente invención es proporcionar una mejora de la calidad de voz que pueda aplicarse en dispositivos cuya memoria disponible sea limitada.

Estas y otras características y ventajas de la presente invención se darán a conocer o resultarán más evidentes en la siguiente descripción y en las reivindicaciones adjuntas. Las características y ventajas pueden implementarse y obtenerse mediante los instrumentos y combinaciones descritos de manera particular en las reivindicaciones adjuntas. Además, las características y ventajas de la invención pueden obtenerse llevando la invención a la práctica o resultarán evidentes a partir de la siguiente descripción.

Descripción de los dibujos

La FIG. 1 es un diagrama de bloques funcional que muestra un dispositivo de mejora de la calidad de voz de la técnica anterior.
La FIG. 2a es un diagrama de bloques funcional de un estimador de varianza de ruido a modo de ejemplo según aspectos de la presente invención. Tales estimadores de varianza de ruido pueden usarse para mejorar los dispositivos de mejora de la calidad de voz de la técnica anterior, como el del ejemplo de la FIG. 1, o pueden usarse con otros fines.
La FIG. 2b es un diagrama de flujo útil para entender el funcionamiento del estimador de varianza de ruido de la FIG. 2a.
La FIG. 3 muestra gráficos idealizados de la estimación de sesgo de la amplitud de ruido en función de la SNR estimada *a priori* para cuatro valores de una SNR real.

Mejor modo de llevar a cabo la invención

En el apéndice A se ofrece un glosario de acrónimos y términos usados en el presente documento. En el apéndice B se ofrece una lista de símbolos junto con sus respectivas definiciones. El apéndice A y el apéndice B son una parte integrante y forman parte de la presente solicitud.

En la FIG. 2a se muestra un diagrama de bloques de una forma de realización a modo de ejemplo de un estimador de varianza de ruido según los aspectos de la invención. Puede estar integrado en un dispositivo de mejora de la calidad de voz, como el de la FIG. 1, con el fin de estimar el nivel de ruido de cada subbanda. Por ejemplo, el estimador de varianza de ruido según los aspectos de la invención puede utilizarse como el estimador de varianza de ruido 4 de la FIG. 1, proporcionando así un dispositivo de mejora de la calidad de voz mejorado. La entrada al estimador de varianza de ruido es la señal de subbanda no modificada $Y(m)$ y su salida es un valor actualizado de la estimación de la varianza de ruido.

Con fines explicativos, el estimador de varianza de ruido puede caracterizarse por presentar tres componentes principales: un dispositivo o función de estimación de amplitud de ruido (“estimación de amplitud de ruido”) 12, un dispositivo o función de estimación de varianza de ruido que funciona en respuesta a una estimación de amplitud de ruido (“estimación de varianza de ruido”) 14 y un dispositivo o función de estimación de varianza de voz (“estimación de varianza de voz”) 16. El ejemplo de estimador de varianza de ruido de la FIG. 2a incluye además un retardo 18, mostrado usando la notación del dominio z (“Z^{-1}”).

El funcionamiento del ejemplo de estimador de varianza de ruido de la FIG. 2a puede entenderse mejor haciendo referencia también al diagrama de flujo de la FIG. 2b. Debe apreciarse que varios dispositivos, funciones y procesos mostrados y descritos en varios ejemplos del presente documento pueden mostrarse de manera combinada o por separado, de distinta manera a la mostrada en las figuras del presente documento. Por ejemplo, cuando se implementan mediante secuencias de instrucciones de software informático, todas las funciones de las FIG. 2a y 2b pueden implementarse mediante secuencias de instrucciones de software de múltiples hilos que se ejecutan en hardware adecuado de procesamiento de señales digitales, en cuyo caso los diversos dispositivos y funciones de los ejemplos mostrados en las figuras pueden corresponder a partes de las instrucciones de software.

Se estima la amplitud de la componente de ruido (estimación de amplitud de ruido 12, FIG. 2a; estimación N(m) 24, FIG. 2b). Puesto que la señal de entrada de audio contiene voz y ruido, tal estimación solo puede realizarse aprovechando las diferencias estadísticas que distinguen una componente de otra. Además, la amplitud de la componente de ruido puede estimarse modificando de manera apropiada los modelos estadísticos existentes usados en la actualidad para estimar la amplitud de la componente de voz (referencias [4] y [5]).

Tales modelos de voz y ruido asumen normalmente que las componentes de voz y de ruido son distribuciones gaussianas de media cero no correlacionadas. Los parámetros de modelo principales, más específicamente la varianza de la componente de voz y la varianza de la componente de ruido, deben estimarse a partir de la señal de audio de entrada no modificada. Como se ha indicado anteriormente, las propiedades estadísticas de las componentes de voz y de ruido son muy diferentes. En la mayoría de los casos, la varianza de la componente de ruido es relativamente estable. Por el contrario, la componente de voz es un proceso de "todo o nada" y su varianza puede cambiar drásticamente incluso en pocos milisegundos. Por consiguiente, la estimación de la varianza de la componente de ruido implica una ventana de tiempo relativamente larga, mientras que la operativa análoga de la componente de voz puede implicar solamente muestras de entrada actuales y anteriores. Un ejemplo de esto último es el "procedimiento basado en decisiones" propuesto en la referencia [1].

Una vez que se hayan determinado los modelos estadísticos y sus parámetros de distribución para las componentes de voz y de ruido, es factible estimar las amplitudes de ambas componentes a partir de la señal de audio. En la forma de realización a modo de ejemplo, el estimador de potencia del mínimo error cuadrático medio (MMSE), presentado anteriormente en la referencia [4] para estimar la amplitud de la componente de voz, está adaptado para estimar la amplitud de la componente de ruido. La elección de un modelo de estimación no es crítica para la invención.

Expresado brevemente, el estimador de potencia MMSE determina primero, de manera respectiva, la distribución de probabilidad de las componentes de voz y de ruido basándose en modelos estadísticos y en la señal de audio no modificada. Después, la amplitud de ruido se determina como el valor que minimiza el valor cuadrático medio del error de estimación.

Finalmente, como preparación a cálculos posteriores, la varianza de la componente de ruido se actualiza incluyendo el valor absoluto actual elevado al cuadrado de la amplitud de ruido estimada en la varianza de ruido global. Este valor adicional forma parte de una operación acumulativa en una memoria intermedia razonablemente grande que contiene la amplitud de componente de ruido actual y las anteriores. Con el fin de mejorar más la precisión de la estimación de la varianza de ruido, puede añadirse un procedimiento de evitación de estimación sesgada.

Estimación de la amplitud de ruido
(Estimación de amplitud de ruido 12, FIG. 2a; estimación N(m) 24, FIG. 2b)

Como se ilustra en las FIG. 1, 2a y 2b (20), la entrada al estimador de varianza de ruido (en este contexto, el "estimador de varianza de ruido" es el bloque 4 de la FIG. 1 y la combinación de los elementos 12, 14, 16 y 18 de la FIG. 2a) es la subbanda:

$$Y(m) = X(m) + D(m) \qquad (2)$$

donde $X(m)$ es la componente de voz y $D(m)$ es la componente de ruido. Aquí, m es el índice de tiempo, omitiéndose el índice de número de subbanda k debido a que se usa el mismo estimador de varianza de ruido para cada subbanda. Puede suponerse que el banco de filtros de análisis genera cantidades complejas, tal como hace una DFT. Por tanto, la componente de subbanda es también compleja y puede representarse además como

$$Y(m) = R(m)\exp\left(j\vartheta(m)\right) \qquad (3)$$

$$X(m) = A(m)\exp\left(j\alpha(m)\right) \qquad (4)$$

y

$$D(m) = N(m)\exp\left(j\phi(m)\right) \qquad (5)$$

donde $R(m)$, $A(m)$ y $N(m)$ son, respectivamente, las amplitudes de la señal de audio no modificada y de las componentes de voz y de ruido, y $\vartheta(m)$, $\alpha(m)$ y $\phi(m)$ son sus respectivas fases.

Suponiendo que las componentes de voz y de ruido son distribuciones gaussianas de media cero no correlacionadas, la amplitud de $X(m)$ puede estimarse usando el estimador de potencia MMSE, obtenido en la referencia [4], de la siguiente manera:

$$\hat{A}(m) = G_{\mathrm{SP}}\left(\xi(m), \gamma(m)\right) \cdot R(m) \qquad (6)$$

donde la función de ganancia viene dada por

$$G_{\mathrm{SP}}\left(\xi(m), \gamma(m)\right) = \sqrt{\frac{\xi(m)}{1+\xi(m)}\left(\frac{1+\upsilon(m)}{\gamma(m)}\right)} \qquad (7)$$

donde

$$\upsilon(m) = \frac{\xi(m)}{1+\xi(m)}\gamma(m) \qquad (8)$$

$$\xi(m) = \frac{\lambda_x(m)}{\lambda_d(m)} \qquad (9)$$

y

$$\gamma(m) = \frac{R^2(m)}{\lambda_d(m)} \qquad (10)$$

Aquí, $\lambda_x(m)$ y $\lambda_d(m)$ son, respectivamente, las varianzas de la componente de voz y de la componente de ruido. $\xi(m)$ y $\gamma(m)$ se consideran normalmente como las relaciones de componente a ruido *a priori* y *a posteriori*, utilizándose dicha notación en el presente documento. Dicho de otro modo, la SNR "a priori" es la relación de la varianza de voz supuesta (aunque no conocida en la práctica, de ahí el término "a priori") con respecto a la varianza de ruido. La SNR "a posteriori" es la relación del cuadrado de la amplitud de la señal observada (de ahí el término "a posteriori") con respecto a la varianza de ruido.

En el modelo de estimación de potencia MMSE, las varianzas respectivas de las componentes de voz y de ruido pueden intercambiarse para estimar la amplitud de la componente de ruido:

$$\hat{N}(m) = G_{\mathrm{SP}}\left(\xi'(m), \gamma'(m)\right) \cdot R(m) \qquad (11)$$

donde

$$\xi'(m) = \frac{\lambda_d(m)}{\lambda_x(m)} \qquad (12)$$

y

$$\gamma'(m) = \frac{R^2(m)}{\lambda_x(m)} \qquad (13)$$

Estimación de la varianza de voz
(Estimación de varianza de voz 16, FIG. 2a; estimación $\hat{\lambda}_x(m)$ 22, FIG. 2b)

La estimación de la varianza de componente de voz $\hat{\lambda}_x(m)$ puede calcularse usando el procedimiento basado en decisiones propuesto en la referencia [1]:

$$\hat{\lambda}_x(m) = \mu\hat{A}^2(m-1) + (1-\mu)\max\left(R^2(m) - \hat{\lambda}_d(m), 0\right) \qquad (14)$$

Aquí

$$0 << \mu < 1 \qquad (15)$$

es una constante preseleccionada y $\hat{A}(m)$ es la estimación de la amplitud de la componente de voz. A continuación se describe la estimación del cálculo de la varianza de componente de ruido $\hat{\lambda}_d(m)$.

Estimación de la amplitud de ruido (continuación de lo anterior)

La estimación de la amplitud de la componente de ruido se obtiene finalmente como

$$\hat{N}(m) = G_{SP}\left(\hat{\xi}'(m), \hat{\gamma}'(m)\right) \cdot R(m) \qquad (16)$$

donde

$$\hat{\xi}'(m) = \frac{\hat{\lambda}_d(m)}{\hat{\lambda}_x(m)} \qquad (17)$$

y

$$\hat{\gamma}'(m) = \frac{R^2(m)}{\hat{\lambda}_x(m)} \qquad (18)$$

Aunque en este ejemplo se utiliza un banco de filtros complejo, resulta sencillo modificar las ecuaciones para un banco de filtros que solo tenga valores reales.

El procedimiento descrito anteriormente solo se proporciona a modo de ejemplo. Modelos más sofisticados o más sencillos pueden utilizarse dependiendo de la aplicación. También pueden usarse múltiples entradas de micrófono para obtener una mejor estimación de las amplitudes de ruido.

Estimación de la varianza de ruido
(Estimación de varianza de ruido 14, FIG. 2a; estimación $\lambda_d(m)$ 26, FIG. 2b)

La componente de ruido en la entrada de subbanda en un índice de tiempo dado *m* se determina, en parte, mediante su varianza $\lambda_d(m)$. Para una distribución gaussiana de media cero, esto se define como el valor medio del cuadrado de la amplitud de la componente de ruido:

$$\lambda_d(m) = E\left\{N^2(m)\right\} \qquad (19)$$

En este caso, la expectación $E\{N^2(m)\}$ se toma con respecto a la distribución de probabilidad de la componente de ruido en el índice de tiempo *m*.

Suponiendo que la componente de ruido es estacionaria y ergódica, $\lambda_d(m)$ puede obtenerse calculando el promedio en el tiempo de amplitudes de ruido estimadas anteriormente. Más específicamente, la varianza de ruido $\lambda_d(m+1)$

del índice de tiempo *m*+1 puede estimarse calculando un promedio ponderado del cuadrado de las amplitudes de ruido previamente estimadas:

$$\hat{\lambda}_d(m+1) = \frac{\sum_{i=0}^{\infty} w(i)\hat{N}^2(m-i)}{\sum_{i=0}^{\infty} w(i)} \qquad (20)$$

donde $w(i)$,i=0,...,∞ es una función de ponderación. En la práctica, $w(i)$ puede elegirse como una ventana de longitud L: $w(i)$=1, i=0,...,L-1. En el procedimiento de ventana rectangular (RWM), la varianza de ruido estimada viene dada por:

$$\hat{\lambda}_d(m+1) = \frac{1}{L}\sum_{i=0}^{L-1} \hat{N}^2(m-i) \qquad (21)$$

También es posible usar una ventana exponencial:

$$w(i) = \beta^{i+1} \qquad (22)$$

donde

$$0 < \beta < 1. \qquad (23)$$

En el procedimiento de promedio móvil (MAM), la varianza de ruido estimada es el promedio móvil del cuadrado de las amplitudes de ruido:

$$\hat{\lambda}_d(m+1) = (1-\beta)\hat{\lambda}_d(m) + \beta\hat{N}_k^2(m) \qquad (24)$$

donde el valor inicial $\hat{\lambda}_d(0)$ puede fijarse a un valor predeterminado elegido de manera razonable.

Evitación de estimación de sesgo

En ocasiones, el modelo no puede proporcionar una representación precisa de las componentes de voz y de ruido. En estas situaciones, la estimación de varianza de ruido puede resultar imprecisa, produciéndose así un resultado muy sesgado. El procedimiento de evitación de estimación de sesgo (BEA) se ha desarrollado para mitigar este problema.

Esencialmente, el procedimiento BEA asigna un peso reducido a las estimaciones de amplitud de ruido $\hat{N}(m)$, de manera que:

$$sesgo(m) = E\left\{N^2(m) - \hat{N}^2(m)\right\} \Big/ E\left\{N^2(m)\right\} \qquad (25)$$

donde el sesgo, *sesgo*(*m*), es mayor que un máximo predeterminado B_{max}, es decir:

$$\left|sesgo(m)\right| > B_{max} \qquad (26)$$

La precisión de la estimación de amplitud de ruido $\hat{N}(m)$ está sujeta a la precisión del modelo, particularmente a las varianzas de las componentes de voz y de ruido descritas en las secciones anteriores. Puesto que la componente de ruido es relativamente estacionaria, su varianza evoluciona lentamente en el tiempo. Por este motivo, el análisis supone que:

$$\hat{\lambda}_d(m) = \lambda_d(m) \qquad (27)$$

Por el contrario, la componente de voz tiene una naturaleza transitoria y es propensa a grandes errores. Suponiendo que la SNR real es

$$\xi^*(m) = \lambda_x(m)/\lambda_d(m) \qquad (28)$$

y que la SNR estimada *a priori* es

$$\tilde{\xi}(m) = \hat{\lambda}_x(m) / \lambda_d(m) \qquad (29)$$

el sesgo de estimación de $N^2(m)$ viene dado realmente por

$$sesgo(m) = \frac{\tilde{\xi}(m) - \xi^*(m)}{\left(1 + \tilde{\xi}(m)\right)^2} \qquad (30)$$

Evidentemente, si

$$\tilde{\xi}(m) = \xi^*(m) \qquad (31)$$

se obtiene un estimador no sesgado y

$$E\left\{\hat{N}^2(m)\right\} = E\left\{N^2(m)\right\} = \lambda_d(m) \qquad (32)$$

Como se observa en la FIG. 3, el sesgo de estimación es asimétrico con respecto a la línea de puntos de la figura, la línea de sesgo cero. La parte inferior del gráfico indica valores muy variables del sesgo de estimación para valores variables de ξ^*, mientras que la parte superior muestra poca dependencia con respecto a $\tilde{\xi}$ o ξ^*.

Para el intervalo SNR de interés, la subestimación de la amplitud de ruido, es decir:

$$E\left\{\hat{N}^2(m)\right\} < E\left\{N^2(m)\right\} \qquad (33)$$

dará como resultado un sesgo positivo, correspondiente a la parte superior del gráfico. Como puede observarse, el efecto es relativamente pequeño y, por tanto, no es problemático.

Sin embargo, la parte inferior del gráfico corresponde a casos en los que la varianza de la componente de voz está subestimada, dando como resultado un gran sesgo de estimación negativo como se expresa mediante la ecuación (30), es decir:

$$\lambda_x(m) > \hat{\lambda}_x(m) \qquad (34)$$

y

$$\lambda_d(m) > \hat{\lambda}_x(m) \qquad (35)$$

o, como alternativa,

$$\xi^*(m) > \tilde{\xi}(m) \qquad (36)$$

y

$$\tilde{\xi}(m) < 1 \qquad (37)$$

así como una gran dependencia de diferentes valores de ξ^*. Hay situaciones en las que la estimación de la amplitud de ruido es muy elevada. Por consiguiente, a tales amplitudes se les asigna un peso reducido o se evitan completamente.

En la práctica, la experiencia ha demostrado que tales amplitudes dudosas $R(m)$ satisfacen:

$$R^2(m) > \psi\left(1 + \hat{\xi}(m)\right)\lambda_d(m) \qquad (38)$$

donde ψ es una constante positiva predefinida. Esta regla proporciona un límite inferior para el sesgo:

$$sesgo(m) > 1 - \frac{1}{2}\psi \qquad (39)$$

donde

$$\psi = 2\left(B_{\max} + 1\right) \tag{40}$$

En resumen, un sesgo positivo es insignificante. Un sesgo negativo es sostenible si a las amplitudes de ruido estimadas $\hat{N}(m)$ definidas en la ecuación (16) y coherentes con la ecuación (38) se les asigna un peso reducido. En una aplicación práctica, puesto que no se conoce el valor de $\lambda_d(m)$, la regla de la ecuación (38) puede aproximarse mediante:

$$R^2(m) > \psi\left(1 + \hat{\xi}(m)\right)\hat{\lambda}_d(m) \tag{41}$$

donde

$$\hat{\xi}(m) = \frac{\hat{\lambda}_x(m)}{\hat{\lambda}_d(m)} \tag{42}$$

Estos dos ejemplos del procedimiento BEA son el procedimiento de ventana rectangular (RWM) con BEA y el procedimiento de promedio móvil (MAM) con BEA. En la primera implementación, el peso asignado a las muestras que son coherentes con la ecuación (38) es cero:

$$\hat{\lambda}_d(m+1) = \frac{1}{L}\sum_{i\in\Phi_m} \hat{N}^2(i) \tag{43}$$

donde Φ_m es un conjunto que contiene los L valores $\hat{N}^2(i)$ más cercanos con respecto al índice de tiempo m que satisfacen

$$R^2(i) \le \psi\left(1 + \hat{\xi}(i)\right)\hat{\lambda}_d(i) \tag{44}$$

En la segunda implementación, tales muestras pueden incluirse con un peso reducido:

$$\hat{\lambda}_d(m+1) = (1-\beta)\hat{\lambda}_d(m) + \beta\hat{N}_k^2(m) \tag{45}$$

donde

$$\beta = \begin{cases} \beta_0 & R^2(m) \le \psi\left(1 + \hat{\xi}(m)\right)\hat{\lambda}_d(m) \\ \beta_1 & \text{en caso contrario.} \end{cases} \tag{46}$$

y

$$\beta_1 < \beta_0 \tag{47}$$

Para finalizar la descripción del diagrama de flujo de la FIG. 2b, el índice de tiempo m se hace avanzar en uno ("m←m+1" 56) y el proceso de la FIG. 2b se repite.

Implementación

La invención puede implementarse en hardware o software, o en una combinación de ambos (por ejemplo, matrices de lógica programable). A menos que se especifique lo contrario, los procesos incluidos como parte de la invención no están relacionados intrínsecamente con ningún ordenador particular ni con otros aparatos. En particular, varias máquinas de propósito general pueden usarse con programas escritos según las enseñanzas del presente documento, o puede resultar más conveniente fabricar aparatos más especializados (por ejemplo, circuitos integrados) para llevar a cabo las etapas de procedimiento requeridas. Por tanto, la invención puede implementase en uno o más programas informáticos que se ejecutan en uno o más sistemas informáticos programables, donde cada uno comprende al menos un procesador, al menos un sistema de almacenamiento de datos (que incluye memoria y/o elementos de almacenamiento volátiles y no volátiles), al menos un dispositivo o puerto de entrada y al menos un dispositivo o puerto de salida. El código de programa se aplica a datos de entrada para llevar a cabo las funciones descritas en el presente documento y para generar información de salida. La información de salida se aplica, de manera conocida, a uno o más dispositivos de salida.

Cada programa de este tipo puede implementarse en cualquier lenguaje informático deseado (incluyendo lenguaje máquina, lenguaje de ensamblador, lenguaje procedural de alto nivel, lenguaje lógico o lenguaje orientado a objetos)

para comunicarse con un sistema informático. En cualquier caso, el lenguaje puede ser un lenguaje compilado o interpretado.

Preferiblemente, cada programa informático de este tipo se almacena o se descarga en un medio o dispositivo de almacenamiento (por ejemplo, una memoria o un medio de estado sólido o un medio magnético u óptico) legible por un ordenador programable de propósito general o especial para configurar y hacer funcionar el ordenador cuando el medio o dispositivo de almacenamiento es leído por el sistema informático para llevar a cabo los procedimientos descritos en el presente documento. El sistema inventivo también puede implementarse como un medio de almacenamiento legible por ordenador, configurado con un programa informático, donde el medio de almacenamiento configurado de este modo hace que un sistema informático funcione de manera específica y predefinida para llevar a cabo las funciones descritas en el presente documento.

Se han descrito varias formas de realización de la invención. Sin embargo, debe entenderse que pueden realizarse varias modificaciones sin apartarse del alcance de la invención, la cual está definida en las reivindicaciones. Por ejemplo, algunas de las etapas descritas en el presente documento pueden no seguir un orden determinado y, por tanto, pueden llevarse a cabo en un orden diferente al descrito.

Apéndice A
Glosario de acrónimos y términos

BEA	Evitación de estimación sesgada
DFT	Transformada discreta de Fourier
DSP	Procesamiento de señales digitales
MAM	Procedimiento de promedio móvil
RWM	Procedimiento de ventana rectangular
SNR	Relación de señal a ruido
T/F	Tiempo/frecuencia
VAD	Detección de actividad de voz

Apéndice B
Lista de símbolos

$y(n),\ n=0,1,...,\infty$	Señal de tiempo digitalizada
$\tilde{y}(n)$	Señal de voz mejorada
$Y_k(m),\ k=1,...,K,\ m=0,1,...,\infty$	Señal de subbanda *k*
$\tilde{Y}_k(m)$	Señal de subbanda mejorada *k*
$X(m)$	Componente de voz de subbanda *k*
$D(m)$	Componente de ruido de subbanda *k*
g_k	Ganancia de supresión para subbanda *k*
$R(m)$	Amplitud de voz ruidosa
$\vartheta(m)$	Fase de voz ruidosa
$A(m)$	Amplitud de componente de voz
$\hat{A}(m)$	Amplitud de componente de voz estimada
$\alpha(m)$	Fase de componente de voz
$N(m)$	Amplitud de componente de ruido
$\hat{N}(m)$	Amplitud de componente de ruido estimada
$\phi(m)$	Fase de componente de ruido
G_{SP}	Función de ganancia
$\lambda_x(m)$	Varianza de componente de voz
$\hat{\lambda}_x(m)$	Varianza de componente de voz estimada
$\lambda_d(m)$	Varianza de componente de ruido
$\hat{\lambda}_d(m)$	Varianza de componente de ruido estimada
$\xi(m)$	Relación de componente de voz a ruido *a priori*
$\gamma(m)$	Relación de componente de voz a ruido *a posteriori*
$\xi'(m)$	Relación de componente de ruido a voz *a priori*
$\gamma'(m)$	Relación de componente de ruido a voz *a posteriori*
α	Constante preseleccionada
β	Constante preseleccionada para estimación de sesgo

REIVINDICACIONES

1. Un procedimiento para obtener una estimación de varianza en componentes de ruido de una señal de audio formada por componentes de voz y de ruido, que comprende:

obtener dicha estimación de varianza en componentes de ruido de una señal de audio a partir del promedio de estimaciones previas de la amplitud de las componentes de ruido de la señal de audio,
en el que las estimaciones de la amplitud de las componentes de ruido de la señal de audio que tienen valores mayores que un umbral se excluyen de o se ponderan con un valor bajo en el promedio de las estimaciones previas de la amplitud de las componentes de ruido de la señal de audio, y
en el que cada estimación de la amplitud de las componentes de ruido de la señal de audio es una función de una estimación de varianza en las componentes de ruido de la señal de audio, una estimación de varianza en las componentes de voz de la señal de audio y la amplitud de la señal de audio.

2. Un procedimiento para mejorar las componentes de voz de una señal de audio formada por componentes de voz y de ruido, que comprende:

transformar la señal de audio desde el dominio de tiempo a una pluralidad de subbandas en el dominio de frecuencia,
procesar subbandas de la señal de audio, incluyendo dicho procesamiento reducir de manera adaptativa la ganancia de algunas de dichas subbandas en respuesta a un control, donde el control se obtiene, al menos en parte, de una estimación de varianza en componentes de ruido de la señal de audio según la reivindicación 1, y
transformar la señal de audio procesada desde el dominio de frecuencia al dominio de tiempo para proporcionar una señal de audio en la que se han mejorado las componentes de voz.

3. Un procedimiento según la reivindicación 1 o la reivindicación 2, en el que $\hat{\lambda}_d(m+1)$, la estimación de varianza en componentes de ruido de la señal de audio de índice de tiempo $m+1$, es un promedio ponderado de $\hat{N}^2(m-i)$, el cuadrado de las amplitudes de ruido previamente estimadas de índice de tiempo $m-i$, en el que:

$$\hat{\lambda}_d(m+1) = \frac{\sum_{i=0}^{\infty} w(i)\hat{N}^2(m-i)}{\sum_{i=0}^{\infty} w(i)},$$

donde $w(i), i = 0,\ldots,\infty$ es una función de ponderación.

4. Un procedimiento según la reivindicación 3, en el que $w(i)$ es una ventana de longitud L, de manera que $w(i)=1$ para $i=0, \ldots, L-1$ y $w(i)=0$ para $i=L,\ldots,\infty$.

5. Un procedimiento según la reivindicación 1 o la reivindicación 2, en el que $\hat{\lambda}_d(m+1)$, la estimación de varianza en componentes de ruido de la señal de audio de índice de tiempo $m+1$, es el promedio móvil de $\hat{N}_k^2$, el cuadrado de las amplitudes de ruido previamente estimadas, en el que:

$$\hat{\lambda}_d(m+1) = (1-\beta)\hat{\lambda}_d(m) + \beta\hat{N}_k^2(m),$$

donde β es una constante preseleccionada y $\hat{\lambda}_d(0)$ es un valor predeterminado.

6. Un procedimiento según la reivindicación 1 o la reivindicación 2, en el que dicho umbral es una función de $\psi(1+\hat{\xi}(m))\hat{\lambda}_d(m)$, donde $\hat{\xi}$ es la relación de señal a ruido estimada *a priori*, $\hat{\lambda}_d$ es la varianza estimada en componentes de ruido de la señal de audio y ψ es una constante determinada por una cantidad máxima predeterminada de un sesgo de estimación.

7. Aparato adaptado para llevar a cabo los procedimientos según una cualquiera de las reivindicaciones 1 a 6.

8. Un programa informático, almacenado en un medio legible por ordenador, para hacer que un ordenador lleve a cabo los procedimientos según una cualquiera de las reivindicaciones 1 a 6.

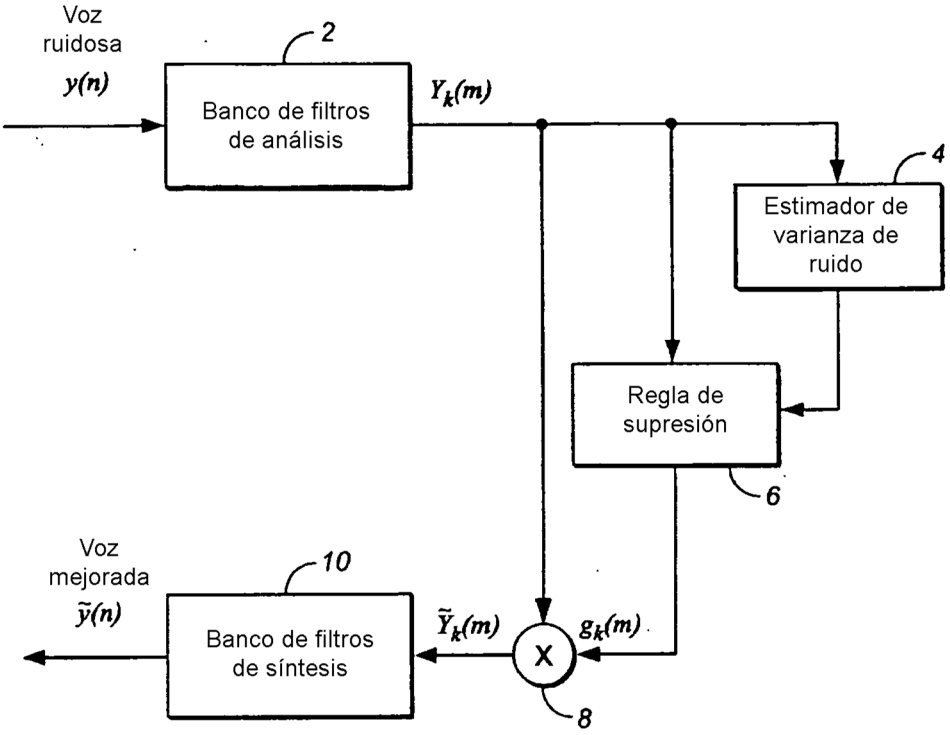
Voz ruidosa
y(n)
2
Banco de filtros de análisis
$Y_k(m)$
4
Estimador de varianza de ruido
Regla de supresión
6
Voz mejorada
$\tilde{y}(n)$
10
Banco de filtros de síntesis
$\tilde{Y}_k(m)$
$g_k(m)$
X
8

FIG. 1

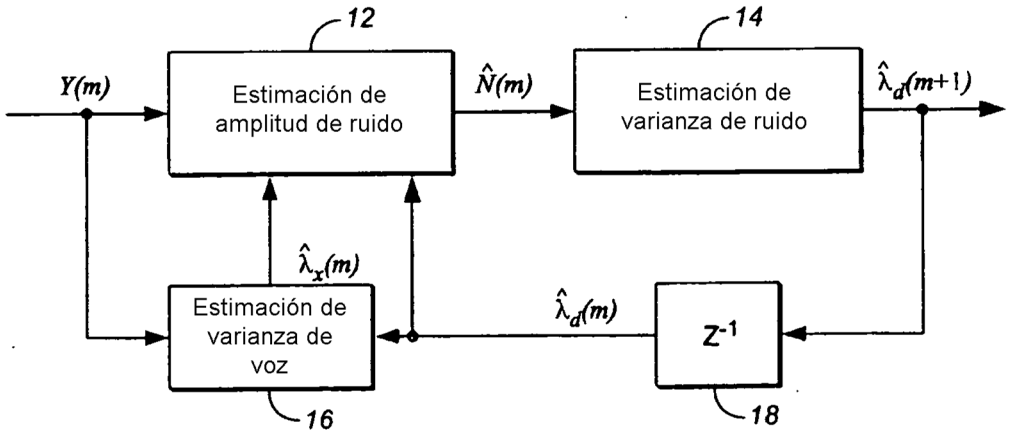
12
Y(m)
Estimación de amplitud de ruido
$\hat{N}(m)$
14
Estimación de varianza de ruido
$\hat{\lambda}_d(m+1)$
$\hat{\lambda}_x(m)$
Estimación de varianza de voz
$\hat{\lambda}_d(m)$
Z^{-1}
16
18

FIG. 2a

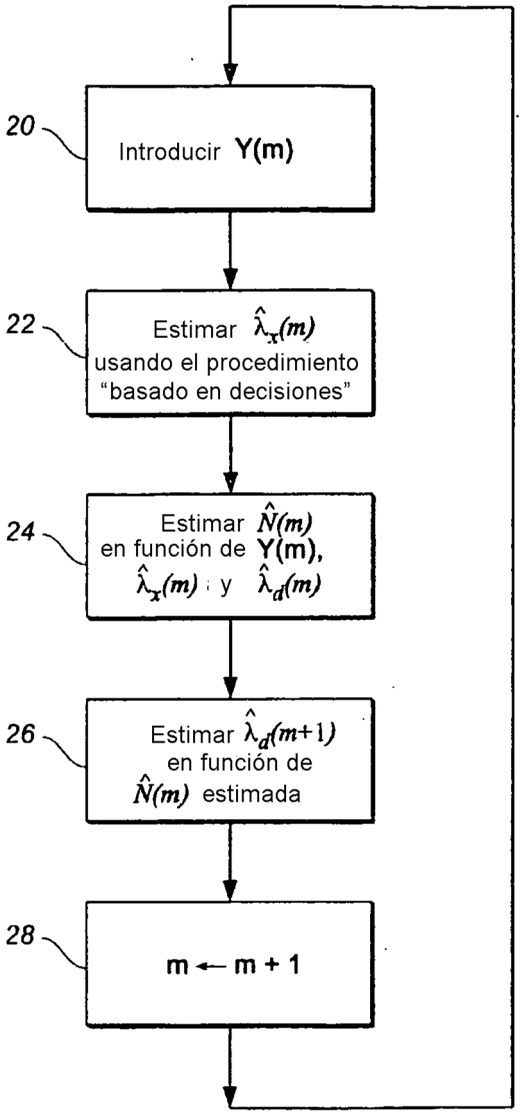
20
Introducir Y(m)
22
Estimar $\hat{\lambda}_x(m)$ usando el procedimiento "basado en decisiones"
24
Estimar $\hat{N}(m)$ en función de Y(m), $\hat{\lambda}_x(m)$ y $\hat{\lambda}_d(m)$
26
Estimar $\hat{\lambda}_d(m+1)$ en función de $\hat{N}(m)$ estimada
28
m ← m + 1

FIG. 2b

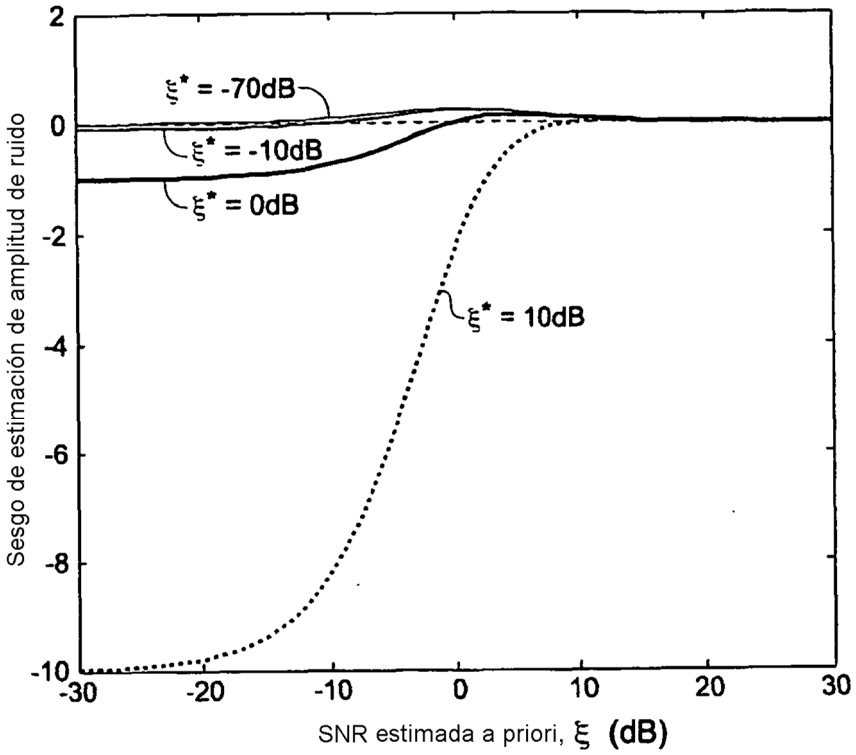
2
0
-2
-4
-6
-8
-10
-30
-20
-10
0
10
20
30
ξ* = -70dB
ξ* = -10dB
ξ* = 0dB
ξ* = 10dB
Sesgo de estimación de amplitud de ruido
SNR estimada a priori, ξ (dB)

FIG. 3