



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 341 500**

51 Int. Cl.:
H04M 9/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07823413 .5**

96 Fecha de presentación : **10.08.2007**

97 Número de publicación de la solicitud: **2057835**

97 Fecha de publicación de la solicitud: **13.05.2009**

54 Título: **Procedimiento de reducción del eco acústico residual tras la supresión de eco en un dispositivo "manos libres".**

30 Prioridad: **26.10.2006 FR 06 09395**

73 Titular/es: **Parrot**
174 quai de Jemmapes
75010 Paris, FR

45 Fecha de publicación de la mención BOPI:
21.06.2010

72 Inventor/es: **Seydoux, Henri**

45 Fecha de la publicación del folleto de la patente:
21.06.2010

74 Agente: **Forteo Laguna, Juan José**

ES 2 341 500 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 de reducción del eco acústico residual tras la supresión de eco en un dispositivo “manos libres”.

5 La invención se refiere de manera general a la reducción de eco en los dispositivos que utilizan señales de audio captadas por un micrófono, por ejemplo los teléfonos, adaptadores de tipo “manos libres” y dispositivos análogos.

En particular, la invención puede aplicarse ventajosamente a un dispositivo “manos libres” utilizable en asociación con un teléfono móvil, especialmente un dispositivo “manos libres” autónomo adaptable de manera amovible en un
10 vehículo automóvil.

Más concretamente, el eco puede resultar de dos fenómenos de distinta naturaleza. El primero es el eco denominado “eco de línea”, limitado a la vía de transmisión y para el que se conocen diversos métodos de filtrado; el segundo se denomina “eco acústico”, eco realmente captado por el micrófono y debido al fenómeno de reverberación del entorno del hablante, típicamente la estancia en la que se encuentra o el habitáculo de un vehículo.
15

Es a la eliminación de este último tipo de eco (eco acústico que se refiere más particularmente la presente invención.

Este eco acústico constituye un elemento perturbador relevante del dispositivo que puede llegar, a menudo, hasta
20 hacer incomprensible para el hablante remoto (el que se encuentra al otro lado de la vía de transmisión de la señal telefónica) las palabras del hablante próximo (aquel cuyas palabras quedan ahogadas en el eco acústico).

Estos aparatos constan de un micrófono sensible destinado a captar la voz del hablante próximo, y un altavoz relativamente potente que reproduce la voz del hablante remoto en el transcurso de una conversación telefónica.
25 Sin embargo, mediante acoplamiento acústico entre estos dos transductores, el micrófono capta no sólo la voz del hablante próximo, sino también el ruido circundante y, sobre todo, el eco acústico, es decir la reverberación del sonido reproducido por el altavoz - y esto a un nivel tanto más elevado cuanto más próximos se encuentran el micrófono y el altavoz y que la potencia acústica restituída por el altavoz es elevada. Tal es típicamente el caso de los sistemas embarcados en un vehículo automóvil, en el que el nivel sonoro del altavoz es relativamente elevado para cubrir los
30 ruidos que circundan.

Además, un número de estos dispositivos están realizados en forma de aparatos autónomos, amovibles, que constan simultáneamente en una misma carcasa del micrófono y del altavoz así como los botones de mando: la proximidad (algunos centímetros) entre el altavoz y el micrófono implica entonces un nivel de eco acústico considerable, típicamente
35 unas veinte veces más elevado que la señal de voz producida por el hablante próximo.

Además, aparte del eco acústico, la importante distancia entre el micrófono y el hablante implica en estos dispositivos un nivel relativo de ruido elevado en la señal captada, que hace difícil la extracción de la señal útil ahogada en el eco y en el ruido. Este ruido presenta además características espectrales no estacionarias, es decir que evolucionan de manera imprevisible en función de las condiciones de conducción: paso por calzadas deformadas o adoquinadas, autorradio en funcionamiento, etc., que hacen aún más difícil la puesta a punto de algoritmos adecuados de tratamiento de la señal.
40

Por estos motivos, la eliminación del eco acústico es particularmente difícil, en especial en los entornos muy ruidosos, típicos de los vehículos automóviles, donde el ruido ambiental se suma a las señales de voz y de eco captadas por el micrófono.
45

También es importante que el tratamiento de eliminación del eco no introduzca distorsión sensible en la señal útil y no genere, o genere lo menos posible, “ruido musical”, que es un tipo de ruido estridente especialmente perceptible y desagradable de oír.
50

Los dispositivos conocidos más eficientes emplean en tales casos: (i) un módulo de anulación de eco, (ii) un módulo de supresión de eco y (iii) un módulo de reducción de ruido.

55 El módulo denominado de “anulación de eco” (*echo cancellation*) o AEC (*Acoustic Echo Canceller*) ejecuta un algoritmo que modeliza el acoplamiento acústico micrófono/altavoz mediante una transformación lineal entre la señal del hablante remoto (es decir la señal destinada a ser reproducida por el altavoz) y el eco captado por el micrófono. Esta transformación define dinámicamente un filtro adaptativo, que se aplica entonces a la señal incidente procedente del hablante remoto. El resultado de este filtrado se sustrae entonces de la señal recogida por el micrófono, lo que tiene
60 por efecto anular la mayor parte del eco acústico.

El módulo denominado de “supresión de eco” (*echo suppression*) tiene por función suprimir el eco residual presente después del tratamiento mediante el módulo de anulación de eco, atenuando ese eco residual hasta el nivel del ruido de fondo. A diferencia de la anulación de eco que se aplica esencialmente mediante sustracción de señales, la supresión de eco se lleva a cabo mediante control de ganancia. El inconveniente es que la reducción de la ganancia se
65 aplica pues igualmente al componente útil de la señal captada por el micrófono (voz del hablante próximo), ya que esa ganancia se aplica indistintamente a la señal incidente, considerada en su globalidad.

Finalmente, el módulo de reducción de ruido (*noise reduction*) pretende reducir el ruido de fondo captado por el micrófono, preservando a la vez la voz del hablante próximo. Esta reducción de ruido funciona ventajosamente de manera dinámica y adaptativa, discriminando los períodos de silencio de la conversación para identificar en éstos el ruido y caracterizar el espectro del mismo, llevando a cabo a continuación una eliminación de ruido selectiva con una atenuación adaptada.

El documento WO-A-00/25441 describe un dispositivo que emplea los diversos módulos descritos más arriba.

Sin embargo, los dispositivos propuestos hasta ahora no ofrecen total satisfacción en las configuraciones de doble voz, especialmente en el caso de los aparatos en los que la distancia entre el altavoz y el micrófono es muy pequeña en comparación con la distancia entre el hablante próximo y el micrófono de manera que, como se ha indicado más arriba, cuando el hablante remoto habla, el nivel del eco producido es considerablemente superior al nivel medio de la voz del hablante próximo.

Por otro lado, el módulo de anulación de eco, que se basa en un filtro lineal adaptativo, no modeliza perfectamente el acoplamiento y no tiene en cuenta las no-linealidades introducidas especialmente a nivel del amplificador y del altavoz, ni del ruido de fondo eléctrico a nivel de los circuitos de conversión analógica/digital. Estos diversos fenómenos generan inestabilidad en el algoritmo de anulación de eco, que se debe readaptar en un tiempo muy corto.

El eco residual después del tratamiento por el módulo de anulación de eco puede alcanzar así un nivel relativamente elevado, de ahí la importancia de la tarea ejecutada por el módulo de supresión de ese eco residual.

La dificultad de idear un módulo de anulación de eco eficiente, combinada con la necesidad de un tratamiento de reducción del eco residual mediante un "post-procesador", lo mencionan en especial C. Faller y J. Chen, *Suppressing Acoustic Echo in a Spectral Envelope Space*, IEEE Transactions on Speech and Audio Processing, Vol. 13, No. 5, pp. 1048-1062, Sept. 2005.

Para evitar las dificultades antes mencionadas, y para reducir la complejidad del algoritmo de reducción del eco acústico, estos autores preconizan una alternativa a la técnica de anulación de eco con filtro adaptativo lineal: en lugar de intentar modelizar el acoplamiento acústico, el algoritmo analiza la envuelta espectral de la señal de eco y aplica directamente un tratamiento de modificación espectral que tiene en cuenta especialmente la selectividad frecuencial del sistema auditivo humano.

La técnica propuesta por este documento presenta sin embargo serias limitaciones en el caso de señales cuyo nivel de eco es muy elevado y en los entornos fuertemente ruidosos, típicos de los adaptadores manos libres utilizados en un habitáculo de automóvil, para los que las técnicas que emplean una modelización del acoplamiento acústico con filtro lineal adaptativo se revelan netamente más eficientes, con una mejor eliminación del eco y una menor distorsión de la señal de voz.

Uno de los objetivos de la invención es proponer un procedimiento mejorado de supresión de eco del tipo antes citado, es decir utilizable en combinación con un módulo de anulación de eco con filtro adaptativo, que permita preservar, en la medida de lo posible, los componentes útiles de la señal, reduciendo al mismo tiempo de manera más acentuada el componente de eco residual. Y esto sin introducir distorsión sensible en la señal útil y sin generar, o lo menos posible, ruido musical parásito.

Otro objetivo de la invención es proponer un procedimiento mejorado de supresión de eco que pueda igualmente asegurar la reducción del ruido de fondo, discriminando el componente de ruido y el componente de eco residual en la señal captada para determinar la que, teniendo en cuenta su espectro, presenta el carácter más perjudicial con relación al componente útil de voz, y aplicarle prioritariamente el tratamiento de supresión.

El procedimiento de la invención es un procedimiento de un tipo conocido según el documento WO-A-00/25441 antes mencionado, es decir que consta en el ámbito temporal de un tratamiento de anulación de eco que funciona mediante sustracción en la señal captada de una señal de referencia entregada por un circuito adaptativo de modelización de dicho acoplamiento acústico. El procedimiento comprende además, en el ámbito frecuencial y en sucesivas tramas de la señal, un tratamiento de supresión del eco residual que subsiste tras la ejecución del tratamiento de anulación de eco, y un tratamiento de reducción del ruido presente en la señal tras la ejecución del tratamiento de anulación de eco.

El tratamiento en el ámbito frecuencial de supresión del eco residual comprende una estimación de los respectivos valores de la potencia del eco residual en una pluralidad de bandas de frecuencias definidas para sucesivas tramas de la señal, así como el cálculo y la aplicación de una ganancia variable propia a cada banda de frecuencias y a cada trama, siendo la respectiva ganancia más débil en la medida en que, para la banda de frecuencia considerada, es elevada la potencia relativa estimada del eco residual, y a la inversa.

De manera característica, según un primer aspecto de la invención, el tratamiento incluye además, para cada banda de frecuencia y para cada trama, una estimación *a priori* de los respectivos valores de la potencia del ruido ambiental presente en la señal captada, y una comparación de los niveles relativos de la potencia del ruido y de la potencia del eco residual en la señal captada. El cálculo de dicha ganancia variable funciona entonces en función de la potencia

ES 2 341 500 T3

relativa estimada del eco residual cuando el nivel de este último es superior al del ruido, y en función de la potencia relativa estimada del ruido en el caso contrario.

Preferiblemente, según un segundo aspecto de la invención, el procedimiento comprende una estimación de una probabilidad de presencia/ausencia de voz, y una ponderación, mediante la probabilidad estimada de presencia/ausencia de voz, de la potencia relativa estimada del eco residual, siendo el valor de la potencia más reducido por dicha ponderación en la medida en que la probabilidad de voz es escasa, y a la inversa.

En un modo de aplicación ventajoso, la potencia relativa estimada del eco residual viene dada, para la banda de frecuencias k definida para la trama l , por la expresión:

$$Potencia_Eco_res(k,l) = \frac{|E\{E(k,l).\bar{X}(k,l)\}|^2}{E\{|X(k,l)|^2\}}.(1 - q_{ref}(k,l))$$

donde:

- | | |
|----------------|---|
| $E\{\bullet\}$ | representa la esperanza matemática, |
| $E(k,l)$ | es la transformada de Fourier discreta de la señal captada tras la sustracción de la señal de referencia mediante el tratamiento de anulación de eco, |
| $X(k,l)$ | es la transformada de Fourier discreta de la señal de referencia utilizada para el tratamiento de anulación de eco, y |
| $q_{ref}(k,l)$ | representa la probabilidad de presencia de voz. |

El cálculo de la ganancia variable puede efectuarse igualmente en función de una estimación *a posteriori* de los valores respectivos de la potencia del eco residual en la señal captada, y esto tanto cuando el nivel del eco residual es superior al del ruido como en el caso contrario.

Se va a describir ahora un ejemplo de aplicación de la invención, con referencia a los dibujos adjuntos.

La figura 1 es un esquema general que muestra las distintas etapas implicadas en el tratamiento de la señal.

La figura 2 es un esquema más detallado que muestra los distintos bloques funcionales implicados al mismo tiempo en el módulo de anulación de eco y en el módulo de reducción de ruido y de supresión de eco residual combinados.

Con referencia a la figura 1, se va a describir en primer lugar la estructura general de un dispositivo “manos libres” que, de manera conocida en sí misma, comprende medios de anulación de eco, de reducción de eco y de reducción del ruido.

Un dispositivo tal está destinado a soportar una conversación telefónica entre dos personas, es decir un hablante próximo 10 y un hablante remoto 12. El hablante próximo 10 es el que se encuentra a proximidad del dispositivo “manos libres” 14. El hablante remoto 12 es aquel con quien el hablante próximo mantiene la conversación telefónica.

El dispositivo 14 es típicamente un dispositivo instalado en un vehículo automóvil, de manera fija o amovible. Posee un altavoz 16 destinado a reproducir la voz del hablante remoto, y un micrófono 18 destinado a captar la voz del hablante próximo 10. El dispositivo consta igualmente de medios de interfaz, esquematizados en 20, para la transmisión bidireccional de las señales en una red telefónica, por ejemplo (pero de manera no limitativa) una interfaz inalámbrica del tipo *Bluetooth* (marca registrada de Bluetooth SIG, Inc.) u otro que permita acoplar el dispositivo 14 al teléfono celular portátil del hablante próximo, mediante el cual este último está en comunicación radiotelefónica con el hablante remoto.

La señal incidente recibida en la entrada E del dispositivo 14, que comprende el componente de voz 22 del hablante remoto 12, se aplica a un convertidor digital/analógico 28, y a la entrada de un amplificador de audio 30 que reproduce la señal en el altavoz 16.

El micrófono 18, por su parte, capta una señal acústica que será una combinación (i) de la voz 32 del hablante próximo 10, (ii) del ruido ambiente 34 reinante en el entorno de este último (por ejemplo en el habitáculo del vehículo automóvil) y (iii) de la señal 36 reproducida por el altavoz 16.

La señal 36 es una señal de eco indeseable desde el punto de vista del micrófono 18. Como se ha indicado más arriba, teniendo en cuenta los niveles de funcionamiento del amplificador 30, y del acoplamiento importante debido

ES 2 341 500 T3

a la proximidad entre altavoz 16 y micrófono 18, cuando el hablante remoto 12 habla, el nivel de la señal 36 puede alcanzar veinte veces el de la señal de voz 32 del hablante próximo 10.

La reducción del eco acústico funciona, tras la digitalización de la señal del micrófono 18 por la etapa de conversión analógica/digital 38, mediante tres tratamientos sucesivos: (i) anulación de eco 40, (ii) supresión de eco 42 y (iii) reducción de ruido 44.

Estos tratamientos se aplican en forma de algoritmos adecuados ejecutados por un microcontrolador o un procesador de tratamiento de las señales incorporado al dispositivo 14. Aunque, para claridad de la exposición, estos tres tratamientos se presenten en forma de módulos distintos, éstos aplican elementos comunes, y corresponden de hecho a una pluralidad de funciones globalmente ejecutadas por un mismo software.

Los distintos bloques funcionales de este software, que realizan un tratamiento de la señal al mismo tiempo en el ámbito temporal (anulación de eco 40) y en el ámbito frecuencial (supresión de eco 42 y reducción del ruido 44) se expondrán en detalle más abajo, con referencia a la figura 2.

El primer módulo es el módulo de anulación de eco 40. No tiene esencialmente efecto sobre la voz del hablante próximo 10: la voz de hablante próximo queda por lo tanto intacta tras la anulación de eco 40. La anulación de eco consiste en definir de manera dinámica un filtro de compensación que reciba en la entrada una señal de referencia 46 procedente del hablante remoto (señal en la entrada E del dispositivo) para sustraer de la señal 48 captada por el micrófono 18 una señal adaptada representativa de la señal recibida, para proporcionar en la salida una señal liberada del eco parásito. El filtro lineal puede estar caracterizado por medio de un algoritmo de supresión de eco tal como un algoritmo del tipo APA (*Affin Projection Algorithm*), o del tipo LMS (*Least Mean Squares*) o NLMS (*Normalized LMS*).

Un algoritmo adaptativo perfeccionado del tipo APA tal se describe por ejemplo en el documento FR-A-2 792 146 (Parrot SA).

Sin embargo, por varios motivos (identificación del filtro imperfecta, modelización lineal aproximada, presencia de no linealidades,...) es frecuente que quede eco audible tras tratamiento por el módulo de anulación de eco 40.

El módulo de supresión de eco 42 tiene por objeto suprimir este eco residual, atenuando la señal de eco hasta el nivel del ruido de fondo.

La invención trata en particular de esta reducción de eco residual, cuya manera de funcionar se explicará más abajo.

Finalmente, el módulo de reducción del ruido 44 tiene por función reducir el ruido de fondo 34 captado por el micrófono, preservando al mismo tiempo la voz 32 del hablante próximo.

El tratamiento de eliminación de ruido propiamente dicho se realiza en el ámbito frecuencial, realizándose el paso entre ámbitos temporal y frecuencial mediante transformadas de Fourier rápidas con ventanas y solapamiento. La eliminación de ruido aplica una estimación de ruido y un cálculo de ganancia para cada banda de frecuencia: durante los períodos identificados como silencios, se analiza el ruido captado en sus distintos componentes frecuenciales de manera a determinar la energía de los componentes frecuenciales más ruidosos y aplicar a continuación a la señal ruidosa una ganancia escasa, dejando al mismo tiempo intactas los componentes poco o nada ruidosos. Los valores de ganancia se basan en modelos estadísticos de la voz y del ruido y en la estimación de parámetros de estos modelos.

La señal así tratada, liberada de su eco y sin ruido, se entrega finalmente a la salida S del dispositivo para transmisión al hablante remoto por la interfaz 20.

La invención propone una nueva técnica de supresión del eco residual, más eficiente que las técnicas empleadas hasta ahora. Estas técnicas presentaban en efecto cierto número de limitaciones, especialmente cuando el eco es difícilmente predecible y puede alcanzar niveles máximos muy elevados debido especialmente a la proximidad material entre micrófono y altavoz.

El principio básico del supresor de eco residual de la invención consiste en utilizar una técnica comparable con la aplicada para la reducción del ruido. Dicho de otro modo, se asimila el eco residual a un ruido y se le aplica un tratamiento de sustracción espectral y de cálculo de una ganancia óptima en cada uno de los componentes frecuenciales.

La supresión de eco residual según la invención consiste por lo tanto, para un componente frecuencial k y una trama (temporal) indexada l , en calcular y en aplicar una ganancia $G_{Eco_res}(k,l)$ que será:

- débil cuando los componentes frecuenciales contengan mucho eco residual, y
- próxima a la unidad, o igual a la unidad, cuando los componentes frecuenciales contengan poco o nada de eco residual.

ES 2 341 500 T3

Este tratamiento tiene como efecto disminuir la energía de los componentes frecuenciales en los que el eco residual es importante y dejar intactos aquellos que no lo contengan.

Para poder calcular la ganancia, es necesario tener una estimación de las relaciones señales a ruido *a priori* y *a posteriori*, necesitando éstas conocer la potencia del eco residual.

La invención propone el siguiente método, con las siguientes anotaciones:

$y(n)$	señal captada por el micrófono (señal 48),
$e(n)$	señal captada, de la que se ha sustraído el eco estimado,
$eco_res(n)$	eco residual (tras el tratamiento de anulación de eco),
$s(n)$	señal de voz del hablante próximo,
$n(n)$	señal de ruido del entorno del hablante próximo,
$x(n)$	señal de referencia (señal 46 utilizada para el tratamiento de anulación de eco),
n	muestra corriente,
h	respuesta impulsional real (desconocida) de la parte lineal del acoplamiento entre el altavoz y el micrófono,
$\hat{h}$	respuesta impulsional estimada, actualizada con la ayuda de métodos adaptativos.
$\otimes$	representa el producto de convolución

Con las anotaciones anteriores, la señal captada se expresa en la forma:

$$y(n) = s(n) + eco(n) + n(n). \quad (1)$$

Despreciando la parte no lineal del acoplamiento, se obtiene la relación:

$$y(n) = s(n) + h \otimes x(n) + n(n). \quad (2)$$

Asimismo, es posible expresar la señal captada de la que se ha sustraído el eco estimado:

$$e(n) = s(n) + (h - \hat{h}) \otimes x(n) + n(n). \quad (3)$$

Tomando la transformada de Fourier discreta de la expresión (3), se obtiene:

$$E(k,l) = S(k,l) + \alpha(k,l).X(k,l) + N(k,l), \quad (4)$$

k y l representando respectivamente el componente frecuencial (banda de frecuencia) y el número de la trama.

Esto viene a suponer que, para cada banda de frecuencias k , el eco residual es proporcional a la señal de referencia, correspondiendo el parámetro $\alpha(k,l)$ a una atenuación. Todo ocurre como si el eco residual fuese igual a la referencia, atenuada de $\alpha(k,l)$.

Si se multiplica (4) por $\bar{X}(k,l)$ y si se toma su esperanza, anotada como $E\{\bullet\}$:

$$E\{E(k,l)\bar{X}(k,l)\} = \{S(k,l)\bar{X}(k,l)\} + E\{\alpha(k,l)X(k,l)\bar{X}(k,l)\} + E\{N(k,l)\bar{X}(k,l)\} \quad (5)$$

Supongamos que $\alpha(k,l)$ es independiente de la trama considerada, es decir que tenemos $\alpha(k,l) = \alpha(k)$. Supongamos además que las variables $S(k,l)$ y $\bar{X}(k,l)$ y las variables $N(k,l)$ y $X(k,l)$ no están correlacionadas. Se supondrá asimismo que tienen una media nula, hipótesis no restrictiva para señales de voz y de ruido. Se obtiene la expresión simplificada (6):

$$E\{E(k,l)\bar{X}(k,l)\} = \alpha(k) \cdot E\{|X(k,l)|^2\}. \quad (6)$$

A partir de ahora, es posible definir una estimación de la potencia del eco residual.

Notemos que $Esp_EX(k,l) = E\{E(k,l) \cdot \bar{X}(k,l)\}$. $Esp_EX(k,l)$ se obtiene por alisamiento de la trama precedente:

$$Esp_EX(k,l) = (1 - \lambda) \cdot Esp_EX(k,l-1) + \lambda \cdot \bar{X}(k,l) \cdot E(k,l),$$

siendo λ el factor de alisamiento. De lo que resulta:

$$\alpha(k,l) = \frac{Esp_EX(k,l)}{E\{|X(k,l)|^2\}}.$$

Dado que $Eco_res(k,l) = \alpha(k,l) \cdot X(k,l)$, resulta:

$$Potencia_Eco_res(k,l) = E\{\alpha(k,l)^2 \cdot X(k,l) \cdot \bar{X}(k,l)\}$$

$$Potencia_Eco_res(k,l) = E\{\alpha(k,l)^2 |X(k,l)|^2\}$$

De lo que resulta:

$$Potencia_Eco_res(k,l) = \frac{|E\{E(k,l)\bar{X}(k,l)\}|^2}{E\{|X(k,l)|^2\}} \quad (7)$$

La presencia de un ruido de fondo en la señal de referencia X (debido al ruido ambiental captado del lado del hablante remoto y al ruido introducido por la vía de transmisión) implica la estimación de un eco residual incluso cuando el hablante remoto no habla, ya que la esperanza en el numerador no puede ser realmente nula. Sin embargo, el eco residual efectivo introducido por este ruido de fondo del lado de hablante remoto no es molesto.

Dado que el proceso de supresión de eco residual por el método propuesto implica una ligera distorsión de la voz, es preferible no sobreestimar demasiado el eco residual cuando ello no es necesario. Es por ello que, en la práctica, es deseable ponderar la expresión (7) por la probabilidad de presencia de voz, que vale (1-probabilidad de ausencia). La expresión se convierte en:

$$Potencia_Eco_res(k,l) = \frac{|E\{E(k,l).\bar{X}(k,l)\}|^2}{E\{|X(k,l)|^2\}} \cdot (1 - q_{ref}(k,l)), \quad (8)$$

donde $q_{ref}(k,l)$ representa la probabilidad de presencia de voz en la señal de referencia, para la frecuencia k y la trama l .

Se obtiene así una estimación de la potencia del eco residual, que se utilizará a continuación en el cálculo de las relaciones de señales a eco *a priori* y *a posteriori*.

Según otro aspecto, la invención propone calcular una ganancia global, en lugar de calcular dos ganancias por separado tras el tratamiento por el módulo de anulación de eco (uno para la supresión del eco residual, función aplicada por el módulo 42, y el otro para la reducción del ruido, función aplicada por el módulo 44).

Esta ganancia global, anotada $G(k,l)$, tendrá en cuenta al mismo tiempo, para una banda de frecuencia k y una trama l , la presencia del ruido y la presencia del eco residual.

La figura 2 ilustra las distintas funciones aplicadas para el cálculo de esta ganancia $G(k,l)$.

El principio básico de este tratamiento está fundado en los trabajos descritos por: [1] Y. Ephraim y D. Malah, Speech Enhancement using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No 6, pp. 1109-1121, Dic. 1984, y [2] Y. Ephraim y D. Malah, Speech Enhancement using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-33, No 2, pp. 443-445, Abril 1985.

Como se ha indicado más arriba, la invención propone utilizar estas técnicas, inicialmente ideadas para la eliminación de ruido, para la supresión del eco residual tras la anulación de eco.

Esencialmente, la señal que contiene el eco residual se analiza en componentes frecuenciales independientes por una transformada de Fourier discreta. Se calcula y se aplica una ganancia óptima en cada uno de estos componentes y luego se recombina la señal así tratada en el ámbito temporal. En [1], la ganancia aplicada se denomina "ganancia STSA" (*Short-Time Spectral Amplitude*) y permite minimizar la distancia cuadrática media entre la señal estimada (a la salida del algoritmo) y la señal de voz original (sin ruido). En [2], la aplicación de una ganancia denominada "ganancia LSA" (*Log-Spectral Amplitude*) permite minimizar la distancia cuadrática media entre el logaritmo de la amplitud de la señal estimada y el logaritmo de la amplitud de la señal de voz original. Este segundo criterio se muestra superior al primero ya que la distancia elegida es mejor adecuación con el comportamiento del oído humano y ofrece por lo tanto cualitativamente mejores resultados.

En todos los casos, la idea esencial es disminuir la energía de los componentes frecuenciales muy parasitados, aplicándoles una ganancia débil, dejando al mismo tiempo intactos (por la aplicación de una ganancia igual a 1) aquellos que lo están poco o nada.

La aplicación de este tratamiento permite calcular un valor $G(k,l)$ para la ganancia LSA. Un algoritmo de tipo "OM-LSA" (*Optimally-Modified Log-Spectral Amplitude*) permite mejorar el cálculo de la ganancia LSA a aplicar, ponderándolo por la probabilidad condicional de presencia de voz, según una técnica descrita especialmente por: [3] I. Cohen, Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator, IEEE Signal Processing Letters, Vol. 9, No 4, Abril 2002.

Más concretamente, como se ilustra en la figura 2, el tratamiento funciona a partir de, por una parte, la señal 48 captada por el micrófono tras la anulación de eco por el módulo 40 (*AEC, Acoustic Echo Celler*) y, por otra parte, a partir de la señal de referencia 46 representativa de la señal reproducida por el altavoz.

El paso en el ámbito frecuencial de estas dos señales 46, 48 se obtiene por una transformación de Fourier rápida FFT (bloques 50, 50').

Para la eliminación de ruido, se somete la señal a un tratamiento tal como el descrito en las referencias [1] y [2], con estimación de la potencia del ruido en las distintas bandas de frecuencias y para una trama de señal dada (bloque 52) y determinación de una relación señal a ruido *a priori* (bloque 54). Esta señal permite igualmente calcular una probabilidad de ausencia de voz (bloque 56). Un tratamiento por un algoritmo de búsqueda de mínimos cuadrados LSA (bloque 58) permite, por nueva conexión en bucle, obtener una relación señal a ruido *a posteriori* (bloque 54) tras ponderación por la probabilidad de presencia de voz condicional (bloque 60). Se obtiene así, para cada componente frecuencial (banda de frecuencia) k de cada trama l , un valor de relación señal a ruido *a priori* y un valor de señal a ruido *a posteriori*. Para más detalles sobre estos algoritmos de reducción de eco, podremos referirnos a las referencias [1] y [2] mencionadas anteriormente. La referencia [3] anteriormente citada indica, por su parte, la manera de ponderar la ganancia por la probabilidad condicional de presencia de voz.

Si se designan las relaciones señal a ruido *a priori* y señal con ruido *a posteriori* respectivamente mediante $SNR_prio(k,l)$ y $SNR_post(k,l)$, la referencia [3] ofrece, para la expresión de la ganancia $G_ruido(k,l)$, en función de los parámetros $\xi(k,l) = SNR_prio(k,l)$ y $\gamma(k,l) = SNR_post(k,l)$:

$$G_ruido(k,l) = G_ruido(\xi(k,l), \gamma(k,l)) = \frac{\xi(k,l)}{\xi(k,l) + 1} \exp \left\{ \frac{1}{2} \int_{v(k,l)}^{\infty} \frac{e^{-t}}{t} dt \right\} \quad (9)$$

$$\text{con } v(k,l) = \frac{\xi(k,l)}{\xi(k,l) + 1} \cdot \gamma(k,l)$$

La invención propone llevar a cabo un tratamiento parecido al que se acaba de describir, para calcular una relación ya no señal a ruido sino señal a eco (aquí y a continuación, para simplificar, se entenderá por “eco” por supuesto el eco residual, que se intenta minimizar).

Las relaciones señal a eco *a priori* y señal a eco *a posteriori* se escriben respectivamente $SER_prio(k,l)$ y $SER_post(k,l)$. Estas relaciones se calculan por los bloques funcionales 52', 54', 56' y 58' de la figura 2, que ejecutan algoritmos similares a los de los bloques homólogos 52, 54, 56 y 58 descritos más arriba a propósito del tratamiento del ruido. Este cálculo se efectúa a partir de los datos producidos por la aplicación de una transformada de Fourier rápida FFT 50' a la señal de referencia 46.

Los valores de las relaciones señal a eco *a priori* y *a posteriori*, y de las relaciones señal a ruido *a priori* y *a posteriori* son así conocidos, y se aplican conjuntamente al algoritmo de cálculo de la ganancia LSA (bloque 62).

La ganancia OM-LSA se calcula por el bloque 64. Una transformación de Fourier inversa (bloque 66) restituye en la salida la señal del micrófono, sin ruido y sin eco.

La invención propone, para el cálculo de la ganancia global $G(k,l)$, es decir de la ganancia que tiene en cuenta al mismo tiempo el eco y el ruido, utilizar la expresión (9) indicada más arriba para el cálculo del ruido, pero seleccionando parámetros $\xi(k,l)$ y $\gamma(k,l)$ dependientes al mismo tiempo de las relaciones señal a ruido y de las relaciones señal a eco.

Se pueden utilizar especialmente, en lugar de los parámetros $\xi(k,l)$ y $\gamma(k,l)$, los parámetros $\xi'(k,l)$ y $\gamma'(k,l)$ tales como:

$$\xi'(k,l) = \min(SER_prio(k,l), SNR_prio(k,l)) \quad (10)$$

$$\gamma'(k,l) = SER_post(k,l) \quad (11)$$

En efecto, si el eco residual es superior al ruido, la relación señal a eco será más débil que la relación señal a ruido; se elegirá por lo tanto el parámetro $SER_prio(k,l)$ para aplicar una ganancia que reducirá más bien el eco. En esta situación, de hecho el eco oculta el ruido, y se buscará entonces a devolver el eco al nivel del ruido.

A la inversa, si el ruido es superior al eco residual, la relación señal a ruido es más débil que la relación señal a eco, y se elegirá el parámetro $SNR_prio(k,l)$ para aplicar una ganancia que reducirá más bien el ruido.

ES 2 341 500 T3

Dicho de otro modo, tomando el mínimo del $SER_{prio}(k,l)$ y del $SNR_{prio}(k,l)$, nos colocamos siempre en el caso más favorable, ya que se intentará siempre reducir el componente parásito más perjudicial.

Se podría aplicar el mismo razonamiento al parámetro $\gamma(k,l)$, tomando el mínimo de las relaciones señal a ruido y señal a eco *a posteriori*. Sin embargo, en la práctica, se constata que la utilización, en todos los casos, de las relaciones señal a eco *a posteriori* introduce menos ruido musical en la señal restituida.

Se obtiene por lo tanto el valor de la ganancia teniendo en cuenta el eco y el ruido:

$$G(k,l) = G(\xi'(k,l), \gamma'(k,l)) = \frac{\xi'(k,l)}{\xi'(k,l) + 1} \exp \left\{ \frac{1}{2} \int_{v'(k,l)}^{+\infty} \frac{e^{-t}}{t} dt \right\} \quad (11)$$

con
$$v'(k,l) = \frac{\xi'(k,l)}{\xi(k',l) + 1} \cdot \gamma(k',l)$$

REIVINDICACIONES

1. Un procedimiento de tratamiento de una señal de audio captada por el micrófono (18) de un dispositivo “manos libres” (14) que consta igualmente de un altavoz (16) capaz de introducir por acoplamiento acústico con el micrófono un eco acústico parásito,

la señal captada por el micrófono constando de un componente útil de voz (32), un componente parásito de eco acústico (36) y un componente parásito de ruido ambiental (34), este procedimiento comprendiendo en el ámbito temporal:

- un tratamiento de anulación de eco (40) funcionando por sustracción en la señal captada de una señal de referencia entregada por un circuito adaptativo de modelización de dicho acoplamiento acústico,

comprendiendo además este procedimiento, en el ámbito frecuencial y en tramas sucesivas de la señal:

- un tratamiento de supresión del eco residual (42) que subsiste tras la ejecución del tratamiento de anulación de eco, comprendiendo:

- una estimación de los valores respectivos de la potencia del eco residual en una pluralidad de bandas de frecuencias definidas para tramas sucesivas de la señal, y

- el cálculo, y la aplicación, de una ganancia variable propia a cada banda de frecuencia y a cada trama, siendo la ganancia respectiva más débil en la medida en que, para la banda de frecuencia considerada, la potencia relativa estimada del eco residual es elevada, y a la inversa; y

- un tratamiento de reducción del ruido (44) presente en la señal tras la ejecución del tratamiento de anulación de eco, procedimiento **caracterizado**:

- porque comprende además, para cada banda de frecuencias y para cada trama:

- una estimación *a priori* de los valores respectivos de la potencia del ruido ambiental presente en la señal captada, y

- una comparación de los niveles relativos de la potencia del ruido y de la potencia del eco residual en la señal captada,

- y porque el cálculo de dicha ganancia variable se lleva a cabo en función de la potencia relativa estimada del eco residual cuando el nivel de este último es superior al del ruido, y en función de la potencia relativa estimada del ruido en el caso contrario.

2. El procedimiento de la reivindicación 1, que incluye además:

- una estimación de una probabilidad de presencia/ausencia de voz, y

- una ponderación, por la probabilidad estimada de presencia/ausencia de voz, de la potencia relativa estimada del eco residual, siendo el valor de la potencia más reducido por esta ponderación en la medida en que la probabilidad de voz es escasa, y a la inversa.

3. El procedimiento de la reivindicación 2, en el que la potencia relativa estimada del eco residual viene dada, para la banda de frecuencias k definida para la trama l , por la expresión:

$$Potencia_Eco_res(k,l) = \frac{|E\{E(k,l).\bar{X}(k,l)\}|^2}{E\{|X(k,l)|^2\}}.(1 - q_{ref}(k,l))$$

donde:

$E\{\bullet\}$ representa la esperanza matemática,

$E(k,l)$ es la transformada de Fourier discreta de la señal captada tras la sustracción de la señal de referencia por el tratamiento de anulación de eco,

ES 2 341 500 T3

$X(k,l)$ es la transformada de Fourier discreta de la señal de referencia utilizada para el tratamiento de anulación de eco, y

$q_{\text{ref}}(k,l)$ representa la probabilidad de presencia de voz.

5

4. El procedimiento de la reivindicación 1, en el que el cálculo de dicha ganancia variable se efectúa igualmente en función de una estimación *a posteriori* de los valores respectivos de la potencia del eco residual en la señal captada.

10 5. El procedimiento de la reivindicación 4, en el que el cálculo de dicha ganancia variable se efectúa en función de la estimación *a posteriori* de los valores respectivos de la potencia del eco residual en la señal captada, tanto cuando el nivel del eco residual es superior al del ruido, como en el caso contrario.

15

20

25

30

35

40

45

50

55

60

65

FIG_1



