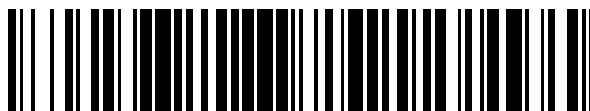


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 377 056**

51 Int. Cl.:

H04R 3/00 (2006.01)

H04R 1/40 (2006.01)

G10L 21/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **10166119 .7**

96 Fecha de presentación: **16.06.2010**

97 Número de publicación de la solicitud: **2293594**

97 Fecha de publicación de la solicitud: **09.03.2011**

54 Título: **Procedimiento de filtrado de los ruidos laterales no estacionarios para un dispositivo de audio multimicrofono, en particular un dispositivo telefónico "manos libres" para vehículo automóvil**

30 Prioridad:
23.07.2009 FR 0955133

45 Fecha de publicación de la mención BOP!:
22.03.2012

45 Fecha de la publicación del folleto de la patente:
22.03.2012

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

72 Inventor/es:
Vitte, Guillaume;
Seris, Julie y
Pinto, Guillaume

74 Agente/Representante:
Fàbrega Sabaté, Xavier

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).

ES 2 377 056 T3

DESCRIPCIÓN

Procedimiento de filtrado de los ruidos laterales no estacionarios para un dispositivo de audio multimicrófono, en particular un dispositivo telefónico “manos libres” para vehículo automóvil.

5 La invención se refiere al procesamiento de la voz en un medio ruidoso.

Se refiere en particular, pero de manera no limitativa, al procesamiento de las señales de voz captadas por dispositivos de telefonía para vehículos automóviles.

10 Estos aparatos constan de un micrófono (“micro”) sensible que capta no sólo la voz de usuario, sino también el ruido ambiental, ruido que constituye un elemento perturbador que puede ir, en algunos casos, hasta hacer incomprensibles las palabras del hablante. Ocurre lo mismo si se quieren aplicar técnicas de reconocimiento de voz, ya que es muy difícil llevar a cabo un reconocimiento de forma en palabras sumergidas en un nivel de ruido elevado.

15 Esta dificultad ligada a los ruidos ambientales es particularmente molesta en el caso de los dispositivos “manos libres”. En particular, la distancia importante entre el micro y el hablante implica un nivel relativo de ruido elevado que dificulta la extracción de la señal útil sumergida en el ruido. Además, el medio muy ruidoso típico del entorno automóvil presenta características espectrales no estacionarias, es decir que evolucionan de manera imprevisible en función de las condiciones de conducción: paso sobre calzadas deformadas o con adoquinadas, autorradio en funcionamiento, etc.

20 Algunos de estos dispositivos prevén el uso de varios micros, generalmente dos micros, y utilizan la media de las señales captadas, u otras operaciones más complejas, para obtener una señal con un nivel menor de perturbaciones. En particular, una técnica denominada conformación de haz permite crear por medios de software una directividad que mejora la relación señal/ruido, pero los rendimientos de esta técnica son muy limitados cuando se usan sólo dos micrófonos.

25 Por otra parte, las técnicas clásicas se adaptan sobre todo al filtrado de los ruidos difusos, estacionarios, que proceden de los alrededores del dispositivo y que se encuentran a niveles comparables en las señales captadas por los dos micros.

30 Por el contrario, un ruido no estacionario, es decir que evoluciona de manera imprevisible en función del tiempo, no se discriminará de la voz y por lo tanto no se atenuará.

Ahora bien, en el entorno automóvil, estos ruidos no estacionarios y directivos son muy frecuentes: toque de bocina, paso de una motocicleta, adelantamiento de un coche, etc.

35 Una de las dificultades del filtrado de estos ruidos no estacionarios se debe a que sus características temporales y espaciales están muy próximas de las de la voz, de ahí la dificultad, por una parte, de estimar la presencia de una voz (ya que el hablante no está hablando todo el tiempo) y, por otra parte, de extraer la señal útil de voz en un entorno muy ruidoso tal como un habitáculo de vehículo automóvil.

40 Uno de los objetivos de la invención es aprovechar la estructura multimicrófono del dispositivo para llevar a cabo una detección espacial de estos ruidos no estacionarios y, a continuación, discriminar entre todas las componentes no estacionarias (en lo sucesivo, “transitorios”), las que son componentes de ruido no estacionario de las que son componentes de voz y, finalmente, tratar la señal captada para eliminar el ruido de la misma de manera eficaz minimizando a la vez las distorsiones introducidas por este procesamiento.

45 50 En lo sucesivo se denominará “ruido lateral” un ruido no estacionario directivo cuya dirección de llegada está alejada de la de la señal útil, y se denominará “cono privilegiado” la dirección o sector angular del espacio donde se encuentra la fuente de señal útil (la voz del hablante) respecto de la red de micros. Cuando una fuente sonora se manifieste fuera del cono privilegiado, se tratará por lo tanto de un ruido lateral, que se intentará atenuar.

55 El punto de partida de la invención consiste en asociar las propiedades de no estacionalidad temporal y frecuencial, por una parte, y de directividad espacial, por otra parte, para detectar un tipo de ruido que es normalmente difícil de discriminar de la voz y, a continuación, para deducir una probabilidad de presencia de la voz que servirá para atenuar este ruido.

60 Más concretamente, la invención tiene por objeto un procedimiento de eliminación de ruido de una señal acústica ruidosa captada por una pluralidad de micrófonos de un dispositivo de audio multimicrófono que opera en un medio ruidoso. La señal acústica ruidosa comprende una componente útil de voz procedente de una fuente de voz directiva y una componente parásita de ruido, incluyendo esta componente de ruido una componente de ruido lateral no estacionaria directiva.

65 Tal procedimiento se divulga, por ejemplo, por 1. Cohen, *Analysis of Two-Channel Generalized Sidelobe Canceller*

(GSC) with Post-Filtering", *IEEE Transactions on Speech and Audio Processing*, Vol. 11, nº 6, noviembre de 2003, páginas 684-699.

Esencialmente, y de manera característica de la invención, el procedimiento consta de las siguientes etapas de procesamiento, ejecutadas en el dominio de frecuencia:

- a) combinación de la pluralidad de señales captadas por la pluralidad correspondiente de micrófonos en una señal combinada ruidosa;
- b) a partir de la señal combinada ruidosa, estimación de una componente de ruido pseudoestacionaria contenida en esta señal combinada ruidosa;
- c) a partir de la componente de ruido pseudoestacionaria estimada en la etapa b) y de la señal combinada ruidosa, cálculo de una probabilidad de presencia de transitorios en la señal combinada ruidosa;
- d) a partir de la pluralidad de señales captadas por la pluralidad correspondiente de micrófonos y de la probabilidad de presencia de transitorios calculada en la etapa c), estimación de una dirección principal de llegada de los transitorios;
- e) a partir de la dirección principal de llegada de los transitorios estimada en la etapa d), cálculo de una probabilidad de presencia de voz con respecto a un criterio espacial, propio para distinguir entre voz útil y ruido lateral entre los transitorios;
- f) a partir de la probabilidad de presencia de voz calculada en la etapa e) y de la señal combinada ruidosa, reducción selectiva del ruido mediante aplicación de una ganancia variable propia de cada banda de frecuencias y de cada trama temporal.

Según diversas formas de puesta en práctica subsidiarias ventajosas:

- el procesamiento de la etapa a) es un procesamiento de prefiltrado de tipo conformación fija de haz;
- el procesamiento de la etapa d) comprende las siguientes subetapas sucesivas: d1) partición del espacio en una pluralidad de sectores angulares; d2) para cada sector, evaluación de un estimador de dirección de llegada a partir de la pluralidad de señales captadas por la pluralidad correspondiente de micrófonos; d3) ponderación de cada estimador por la probabilidad de presencia de transitorios calculada en la etapa c); d4) a partir de los valores de estimadores ponderados calculados en la etapa d3), estimación de una dirección principal de llegada de los transitorios; y d5) validación o invalidación de la estimación de la dirección principal de llegada de los transitorios llevada a cabo en la etapa d4).
- en la etapa d5) la estimación sólo se valida si el valor del estimador ponderado que corresponde a la dirección estimada es superior a un umbral predeterminado, y/o en ausencia de máximo local del estimador ponderado en el sector angular de origen de la señal de voz útil, y/o si el valor del estimador es creciente de manera monótona en una pluralidad de tramas temporales sucesivas;
- el procedimiento comprende además una etapa de mantenimiento de la estimación de la dirección principal de llegada durante un lapso de tiempo mínimo predeterminado;
- la probabilidad de presencia de voz calculada en la etapa e) es o bien una probabilidad binaria, que toma un valor de 1 ó 0 dependiendo de si la dirección principal de llegada de los transitorios estimada en la etapa d) está situada o no en el sector angular de origen de la señal de voz útil, o bien una probabilidad de valores múltiples, función de la desviación angular entre la dirección principal de llegada de los transitorios estimada en la etapa d) y la dirección de origen de la señal de voz útil;
- el procesamiento de la etapa f) es un procesamiento de reducción selectiva del ruido mediante la aplicación de una ganancia de amplitud espectral logarítmica modificada optimizada OM-LSA.

A continuación se describirá un ejemplo de aplicación del procedimiento de la invención con referencia a la figura adjunta.

La Figura 1 es un esquema de bloques que muestra los diferentes módulos y funciones aplicadas por el procedimiento de la invención así como sus interacciones.

El procedimiento de la invención se aplica por medios de software, que se pueden descomponer y esquematizar mediante un cierto número de módulos 10 a 24 ilustrados en la Figura 1.

Estos procesamientos se aplican en forma de algoritmos apropiados ejecutados por un microcontrolador o un procesador digital de señal. Aunque, por motivos de claridad, estos diversos procesamientos se presentan en forma de distintos módulos, aplican elementos comunes y corresponden en la práctica a una pluralidad de funciones globalmente ejecutadas por un mismo software.

La señal de la que se desea eliminar el ruido procede de una pluralidad de señales captadas por una red de micros (que, en la configuración mínima, puede ser simplemente una red de dos micros) dispuestos según una configuración predeterminada.

La red de micros capta la señal emitida por la fuente de señal útil (señal de voz) y la diferencia de posición entre los micros induce un conjunto de desfases y variaciones de amplitud en el registro de las señales emitidas por la fuente

de señal útil.

Más concretamente, el micro de índice n emite una señal:

$$x_n(t) = a_n \times s(t - \tau_n) + v_n(t)$$

- 5 donde a_n es la atenuación de amplitud debida a la pérdida de energía entre la posición de la fuente sonora s y el micro, τ_n es el desfase entre la señal emitida y recibida por el micro y v_n representa el valor del campo de ruido difuso en la posición del micro.

- 10 En la medida en que la fuente se encuentra alejada al menos algunos centímetros de los micros, se podrá llevar a cabo la aproximación que la fuente sonora emite una onda plana. Los retardos τ_n se podrán entonces calcular a partir del ángulo θ_s , definido como el ángulo entre las mediatrices de las parejas de micros (n, m) y la dirección de referencia correspondiente a la fuente s de señal útil. Cuando el sistema considerado consta de dos micros cuya mediatriz corta la fuente, el ángulo θ_s es nulo.

- 15 *Transformada de Fourier de las señales captadas por los micros (bloques 10)*

La señal en el dominio de tiempo $x_n(t)$ procedente de cada uno de los N micros se digitaliza, se divide en tramas de T puntos temporales, aventanada temporalmente por una ventana de tipo Hanning, y luego la transformada de Fourier rápida FFT (transformada a corto plazo) $X_n(k, l)$ se calcula para cada una de estas señales:

20

$$X_n(k, l) = a_n \cdot d_n(k) \times S(k, l) + V_n(k, l)$$

con :

$$d_n(k) = e^{-i2\pi f_k \tau_n}$$

- 25 1 siendo el índice de la trama temporal,
 k siendo el índice de la banda de frecuencias, y
 f_k siendo la frecuencia central de la banda de frecuencia de índice k .

Constitución de una señal combinada con ruido parcialmente eliminado (bloque 12)

30

Las señales $X_n(k, l)$ pueden combinarse entre sí mediante una técnica sencilla de prefiltrado por conformación de haz del tipo *Delay and Sum*, que se aplica para obtener una señal combinada $X(k, l)$ con ruido parcialmente eliminado:

$$X(k, l) = \frac{1}{N} \sum_{n \in [1, N]} \overline{d_n(k)} \cdot X_n(k, l)$$

35

Cabe resaltar que, concretamente, al estar limitado el número de micros, este procesamiento sólo proporciona una ligera mejora de la relación señal/ruido, del orden de 1 dB solamente.

- 40 Cuando el sistema considerado consta de dos micros cuya mediatriz corta la fuente, el ángulo θ_s es nulo y se trata de una simple media que se realiza en los dos micrófonos.

Estimación del ruido pseudoestacionario (bloque 14)

- 45 Esta etapa tiene por objeto calcular una estimación de la componente de ruido pseudoestacionario $\hat{v}(k, l)$ presente en la señal $X(k, l)$.

- 50 Existe un gran número de publicaciones acerca de este tema, siendo en efecto la estimación y la reducción del ruido pseudoestacionario un problema clásico bastante bien resuelto. Diferentes procedimientos son eficaces y pueden utilizarse para obtener $\hat{v}(k, l)$, en particular un algoritmo de estimación de la energía de la componente de ruido pseudoestacionario con promediación recursiva mediante control de los mínimos (MCRA) como el descrito por I. Cohen y B. Berdugo, *Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement*, *IEEE Signal Processing Letters*, Vol. 9, nº 1, páginas 12-15, enero de 2002.

- 55 *Cálculo de la probabilidad de presencia de los transitorios (bloque 16)*

Los "transitorios" designan todas las señales no estacionarias, incluyendo tanto la voz útil como los ruidos no estacionarios esporádicos, que pueden tener una energía equivalente o a veces superior a la voz útil (paso de un vehículo, sirena, claxon, voz de otras personas, etc.).

5 Es posible detectar estos transitorios con la ayuda de la estimación establecida anteriormente de la componente de ruido pseudoestacionario $\hat{V}(k, l)$, sustrayendo esta última de la señal global $X(k, l)$.

10 Más adelante se verá (descripción detallada de los bloques 18 y 20) de qué manera es posible discriminar de entre estos transitorios los que corresponden a la voz útil y los que corresponden a ruidos no estacionarios y que tienen características similares a la voz útil. El procesamiento llevado a cabo por el bloque 16 consiste solamente en calcular una probabilidad $p_{\text{Transitorio}}(k, l)$ de presencia de señales transitorias, sin distinción entre voz útil y ruidos parásitos no estacionarios. El algoritmo es el siguiente:

15 Para cada trama l y para cada banda de frecuencia k ,

(i) calcular la Relación Transitorios-Estacionarios: $TSR(k, l) = \frac{X(k, l) - \hat{V}(k, l)}{\hat{V}(k, l)}$

20 (ii) Si $TSR(k, l) \leq TSR_{\min}$:

$$p_{\text{Transitorio}}(k, l) = 0$$

25 (iii) Si $TSR(k, l) \geq TSR_{\max}$:

$$p_{\text{Transitorio}}(k, l) = 1$$

30 (iv) Si $TSR_{\min} < TSR(k, l) < TSR_{\max}$: $p_{\text{Transitorio}}(k, l) = \frac{TSR(k, l) - TSR_{\min}}{TSR_{\max} - TSR_{\min}}$

Las constantes $TSR_{\min}$ y $TSR_{\max}$ se eligen de manera que se correspondan con situaciones típicas, próximas a la realidad.

35 *Cálculo de la dirección de llegada de los transitorios (bloque 18)*

Este cálculo aprovecha el hecho de que, a diferencia de la componente pseudoestacionaria del ruido que es difusa, los transitorios son a menudo directivos, es decir, procedentes de una fuente sonora puntual (como la boca del hablante para la voz útil, o el motor de una motocicleta para un ruido lateral). Por lo tanto es conveniente calcular la dirección de llegada de estas señales, que en general se definirá correctamente, y comparar esta dirección de llegada con el ángulo θ_s correspondiente a la dirección de origen de voz útil, para determinar si la señal no estacionaria considerada es útil o parásita, y efectuar de este modo la discriminación entre voz útil y ruido no estacionario.

45 La primera etapa consiste en estimar la dirección de llegada del transitorio. El procedimiento utilizado aquí se basa en el uso de la probabilidad de presencia de los transitorios $p_{\text{Transitorio}}(k, l)$ determinada por el bloque 18 de la manera expuesta más arriba.

50 Más concretamente, se lleva a cabo una partición del espacio en sectores angulares correspondiendo cada uno a una dirección definida por un ángulo θ_i , $i \in [1, M]$, (por ejemplo, $M = 19$, con el conjunto de ángulos $\{-90^\circ, -80^\circ, \dots, 0^\circ, \dots, +80^\circ, +90^\circ\}$). Se observará que no hay ninguna relación entre el número N de micros y el número M de ángulos probados. Por ejemplo, es totalmente posible probar una decena de ángulos ($M = 10$) con una sola pareja de micros ($N=2$).

55 Cada ángulo θ_i se prueba para determinar el que está más próximo a la dirección de llegada de la señal no estacionaria estudiada. Para esto, se considera cada pareja de micros (n, m) y se calcula un estimador de dirección de llegada $P_{n,m}(\theta_i, k, l)$ correspondiente, cuyo módulo será máximo cuando el ángulo θ_i probado sea el más próximo a la dirección de llegada del transitorio.

60 Este estimador puede basarse por ejemplo en un cálculo de intercorrelación y tomar la forma:

$$P_{n,m}(\theta_i, k, l) = E\left(X_m(k, l) \cdot \bar{X}_n(k, l) \cdot e^{-i2\pi f_k \tau_i}\right)$$

con

$$\tau_i = \frac{l_{n,m}}{c} \sin \theta_i$$

$l_{n,m}$ siendo la distancia entre los micros de índices n y m , y siendo c la celeridad del sonido.

Un primer procedimiento, clásico, consiste en tomar para la estimación de la dirección de llegada el ángulo que maximiza el módulo de este estimador, es decir:

$$\hat{\theta}_{std}(k, l) = \arg \max_{\theta_i, i \in [1, M]} \|P_{n,m}(\theta_i, k, l)\|$$

Otro procedimiento, utilizado aquí de manera preferida, consiste en ponderar el estimador $P_{n,m}(\theta_i, k, l)$ por la probabilidad de presencia de transitorios $p_{Transitorio}(k, l)$ y definir una nueva estrategia de decisión. El estimador de dirección de llegada correspondiente será:

$$P_{Nuevo_{n,m}}(\theta_j, k, l) = P_{n,m}(\theta_j, k, l) \times p_{Transitorio}(k, l)$$

El estimador puede promediarse en las parejas de micros (n, m) :

$$P_{Nuevo}(\theta_i, k, l) = \frac{1}{N(N-1)} \sum_{n \neq m} P_{Nuevo_{n,m}}(\theta_i, k, l)$$

La integración de la probabilidad de presencia de transitorios en el estimador de dirección de llegada presenta tres ventajas importantes:

- la estimación de dirección tiene como objetivo las partes no estacionarias de la señal (donde la probabilidad $p_{Transitorio}(k, l)$ es próxima a 1), cuya dirección de llegada está bien definida, lo cual hace que la estimación sea coherente;
- la estimación de dirección es robusta con respecto al ruido difuso (donde la probabilidad $p_{Transitorio}(k, l)$ es próxima a cero), lo que normalmente perturba las estimaciones de dirección de llegada;
- la fiabilidad del estimador $P_{Nuevo_{n,m}}(\theta_i, k, l)$ permite distinguir varias señales no estacionarias que corresponden a diferentes direcciones y simultáneamente presentes (se verá más abajo que esta distinción puede realizarse por banda de frecuencias o por análisis de los máximos angulares locales en una misma banda de frecuencias). De este modo, si se tiene al mismo tiempo una señal de voz útil y un ruido lateral potente, se detectarán los dos tipos de señal, evitando que la señal de voz útil concomitante se elimine por error en el resto del proceso, aunque su energía sea baja.

Ahora se explicitarán las reglas de decisión que permiten a partir de P_{Nuevo} :

- o bien proporcionar una estimación $\hat{\theta}(k, l)$ de la dirección de llegada del transitorio,
- o bien indicar que no puede proporcionarse ninguna estimación de dirección de llegada, si estas reglas no se cumplen.

1º) Significado de $P_{Nuevo}(\theta_{max}, k, l)$, (siendo θ_{max} el ángulo que maximiza el valor $\|P_{Nuevo}(\theta_i, k, l)\|$)

Regla 1:

Una estimación de dirección sólo puede proporcionarse si $\|P_{Nuevo}(\theta_{max}, k, l)\|$ sobrepasa un umbral dado P_{MIN} .

Esta primera regla permite asegurar que en la parte (k, l) de la señal considerada, la probabilidad de presencia de un transitorio y el nivel de intercorrelación son lo bastante elevados para que la estimación sea coherente.

2º) Monotonía de P_{Nuevo} en el intervalo $[\theta_s - \theta_{max}; \theta_{max}]$ (para aligerar las notaciones, en lo sucesivo se retirarán las barras de módulo de P_{Nuevo})

Regla 2:

Si θ_{max} está fuera del cono privilegiado, una estimación de ángulo sólo será validada si P_{Nuevo} aumenta de manera monótona en el intervalo $[\theta_s - \theta_{max}; \theta_{max}]$.

Esta segunda regla analiza el contenido del “cono privilegiado”, correspondiente al sector angular en el cual está centrada la fuente s y que presenta una extensión angular de θ_0 . Este cono privilegiado se define mediante los ángulos $\hat{\theta}$ tales que $|\theta - \theta_s| \leq \theta_0$.

El "ruido lateral" corresponderá a una señal cuya dirección de llegada es exterior al cono privilegiado, y se considerará por lo tanto que un ruido lateral está presente si $|\theta_{\max} - \theta_s|$ sobrepasa el umbral θ_0 .

Para validar esta detección de un ruido lateral, hay que verificar que una señal de voz útil no se encuentra simultáneamente en la entrada del sistema.

Para esto, $P_{\text{Nuevo}}(\theta_{\max}, k, l)$ se confronta con los valores de $P_{\text{Nuevo}}(\theta_i, k, l)$ obtenidos para otros ángulos, en particular los que pertenecen al cono privilegiado. La regla permite de este modo garantizar que no hay ningún máximo local en el cono privilegiado.

3º) Fiabilización de la detección de un ruido lateral

Regla 3:

Si $\theta_{\max}$ se encuentra fuera del cono privilegiado por primera vez en la trama l considerada, una estimación de ángulo solo se validará si:

$$P_{\text{Nuevo}}(\theta_{\max}, k, l) \geq \alpha_1 \times P_{\text{Nuevo}}(\theta_{\max}, k, l-1),$$

y si

$$P_{\text{Nuevo}}(\theta_{\max}, k, l) \geq \alpha_2 \times \frac{1}{5} \sum_{i \in [l-5:l-1]} P_{\text{Nuevo}}(\theta_{\max}, k, i).$$

Si se detecta un ruido lateral, esta tercera regla tiene en cuenta las tramas anteriores para evitar las activaciones falsas. Se aplica sólo a la primera trama de un presunto ruido lateral, y verifica que $P_{\text{Nuevo}}(\theta_{\max}, k, l)$ aumente de manera significativa con respecto a los datos correspondientes obtenidos en las cinco tramas anteriores.

Los parámetros α_1 y α_2 se eligen para que correspondan a situaciones típicas, próximas a la realidad. Si se cumplen las tres reglas 1 a 3 anteriores, la estimación $\hat{\theta}(k, l)$ de la dirección de llegada vendrá dada por: $\hat{\theta}(k, l) = \theta_{\max}$.

4º) Estabilización de la detección de un ruido lateral:

Las dos últimas reglas están destinadas a impedir los cortes en la detección de un ruido lateral. Después de un periodo de detección, siguen manteniendo este estado durante un lapso de tiempo denominado de persistencia, aunque ya no se verifiquen las reglas de decisión anteriores. Esto permite detectar los eventuales periodos de baja energía de un ruido no estacionario.

Regla 4:

Si $\hat{\theta}(k, l-1)$ está fuera del cono privilegiado (trama anterior), si $cpt_1 \leq \text{TiempoPersistencia}_1$, (es decir, el periodo de persistencia no ha terminado), y si $P_{\text{Nuevo}}(\hat{\theta}(k, l-1), k, l)$ es superior a un umbral dado P_1 entonces la estimación de ángulo se mantiene y cpt_1 se incrementa.

Regla 5:

Si $\hat{\theta}(k, l-1)$ está fuera del cono privilegiado (trama anterior), si $cpt_2 \leq \text{TiempoPersistencia}_2$ y si $\frac{1}{5} \sum_{i \in [l-5:l-1]} P_{\text{Nuevo}}(\hat{\theta}(k, l-1), k, i)$ es superior a un umbral dado P_2 , entonces la estimación de ángulo se mantiene y cpt_2 se incrementa.

Si una de estas dos últimas reglas (Regla nº 4 o nº 5) se cumple, es prioritaria, y de ella resulta que: $\hat{\theta}(k, l) = \hat{\theta}(k, l-1)$, por consiguiente con corrección eventual del valor de $\hat{\theta}(k, l)$, que no será igual a $\hat{\theta}_{\max}$ pero que se mantendrá en su valor anterior.

Resumiendo, el cálculo de $\hat{\theta}(k, l)$ sigue tres casos posibles:

- (i) si se cumple la regla nº 4 o nº 5, entonces $\hat{\theta}(k, l) = \hat{\theta}(k, l-1)$;
- (ii) en caso contrario (no se cumplen ni la regla nº 4, ni la regla nº 5), si se cumplen las reglas nº 1, nº 2 y nº 3, entonces $\hat{\theta}(k, l) = \theta_{\max}$;
- (iii) si no (no se cumplen ni la regla nº 4, ni la regla nº 5, y al menos no se cumple una de las reglas nº 1, nº 2 y nº 3), entonces no se define $\hat{\theta}(k, l)$.

En una variante, el estimador P_{Nuevo} es media en paquetes de bandas de frecuencias $K_1, K_2, \dots, K_p$:

$$P_{Nuevo}(\theta_i, K_j, l) = \frac{1}{N(N-1)} \frac{1}{C_j} \sum_{n \neq m} \left[\sum_{k \in K_j} P_{Nuevo, n, m}(\theta_i, k, l) \right]$$

C_j designando el cardinal de K_j .

En este caso, la estimación de ángulo θ_{max} no se realiza en cada banda de frecuencias, sino en cada paquete K_j de bandas de frecuencias.

Se observará asimismo que un enfoque de "banda completa" es posible ($p=1$, estimándose un solo ángulo por trama).

Finalmente se observará que el procedimiento propuesto es compatible con el uso de micros unidireccionales. En este caso, será habitual utilizar una red lineal (micros alineados y cuyas direcciones privilegiadas son idénticas) y orientada hacia el hablante. En este caso, el valor de θ_s es por lo tanto naturalmente conocido e igual a cero.

Cálculo de una probabilidad de presencia de voz con criterio espacial (bloque 20)

La siguiente etapa, característica del procedimiento de la invención, consiste en calcular una probabilidad de presencia de voz en función de la estimación de dirección de llegada $\hat{\theta}(k, l)$ obtenida de la manera indicada más arriba.

Se trata de una probabilidad denotada como $p_{spa}(k, l)$, que tiene por lo tanto como novedad ser calculada con un criterio espacial (a partir de $\hat{\theta}(k, l)$), y que permitirá distinguir entre las señales no estacionarias la voz útil de los ruidos parásitos. Esta probabilidad se utilizará entonces en una estructura clásica de eliminación de ruido (bloque 22, descrito a continuación).

La probabilidad $p_{spa}(k, l)$ puede calcularse de diferentes maneras, dando un valor binario o bien múltiples valores. A continuación se darán dos ejemplos de cálculo de $p_{spa}(k, l)$, sabiendo que pueden utilizarse otras leyes para expresar $p_{spa}(k, l)$ a partir de $\hat{\theta}(k, l)$.

1º) *Cálculo de una probabilidad $P_{spa}(k, l)$ binaria:*

La probabilidad de presencia de voz tomará los valores '0' o '1':

- valdrá '0' cuando se detecte un ruido lateral, es decir un transitorio procedente de una dirección exterior al cono privilegiado;
- valdrá '1' cuando la dirección de llegada del transitorio se encuentre en el interior del cono privilegiado, o cuando no se haya podido realizar ninguna estimación fiable en esta dirección.

El algoritmo correspondiente es el siguiente:

Si $\hat{\theta}(k, l)$ está en el interior del cono privilegiado ($|\hat{\theta}(k, l) - \theta_s| \leq \theta_0$) entonces $p_{spa}(k, l) = 1$
 Si $\hat{\theta}(k, l)$ está en el exterior del cono privilegiado ($|\hat{\theta}(k, l) - \theta_s| > \theta_0$) entonces $p_{spa}(k, l) = 0$
 Si $\hat{\theta}(k, l)$ no se define, entonces $p_{spa}(k, l) = 1$

2º) *Cálculo de una probabilidad $p_{spa}(k, l)$ con valores continuos en [0;1]:*

Es posible utilizar para $p_{spa}(k, l)$ un cálculo progresivo, por ejemplo según el siguiente algoritmo:

Si $\hat{\theta}(k, l)$ está en el interior del cono privilegiado ($|\hat{\theta}(k, l) - \theta_s| \leq \theta_0$) entonces $p_{spa}(k, l) = 1$
 Si $\hat{\theta}(k, l)$ está en el exterior del cono privilegiado ($|\hat{\theta}(k, l) - \theta_s| > \theta_0$) entonces $p_{spa}(k, l) = 1 - \frac{|\hat{\theta}(k, l) - \theta_0|}{\frac{\pi}{2} - \theta_0}$
 Si $\hat{\theta}(k, l)$ no se define, entonces $p_{spa}(k, l) = 1$

Reducción de ruido lateral (bloque 22)

La probabilidad $p_{spa}(k, l)$ de presencia de voz calculada en el bloque 20, que depende de la probabilidad $p_{Transitorio}(k, l)$ de presencia de transitorios calculada en el bloque 16, se va a utilizar como parámetro de entrada en una técnica clásica de eliminación de ruido.

Se sabe que la probabilidad de presencia de voz es un estimador crucial para el buen funcionamiento de un algoritmo de eliminación de ruido, ya que sirve de base a la buena estimación del ruido y el cálculo de una ganancia óptima eficaz. Puede utilizarse de manera ventajosa un procedimiento de eliminación de ruido de tipo OM-LSA (*Optimally Modified - Log Spectral Amplitude*) como el descrito por I. Cohen *Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator*, IEEE Signal Processing Letters, Vol. 9, nº 4, abril de 2002.

Esencialmente, 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 más adecuada al comportamiento del oído humano y da por lo tanto cualitativamente mejores resultados. En todos los casos, la idea esencial es disminuir la energía de las componentes de frecuencia muy parasitadas aplicándoles una ganancia baja, dejando a la vez intactas (por aplicación de una ganancia igual a 1) las que lo están poco o nada.

El algoritmo "OM-LSA" (*Optimally-Modified Log-Spectral Amplitude*) mejora el cálculo de la ganancia LSA que ha de aplicarse ponderándola por la probabilidad condicional de presencia de voz.

En este procedimiento, la probabilidad de presencia de voz interviene en dos momentos importantes, para la estimación de la energía del ruido y para el cálculo de la ganancia final, y la probabilidad $p_{spa}(k,l)$ se utilizará a estos dos niveles.

Si se denota como $\hat{\lambda}_{Ruido}(k,l)$ la estimación de la densidad espectral de potencia del ruido, esta estimación viene dada por:

$$\hat{\lambda}_{Ruido}(k,l) = \alpha_{Ruido}(k,l) \cdot \hat{\lambda}_{Ruido}(k,l-1) + [1 - \alpha_{Ruido}(k,l)] |X(k,l)|^2$$

con:

$$\alpha_{Ruido}(k,l) = \alpha_B + (1 - \alpha_B) \cdot p_{spa}(k,l)$$

Puede observarse aquí que la probabilidad $p_{spa}(k,l)$ modula el factor de olvido en la estimación del ruido, que se actualiza más rápidamente en la señal ruidosa $X(k,l)$ cuando la probabilidad de voz es baja, condicionando este mecanismo totalmente la calidad de $\hat{\lambda}_{Ruido}(k,l)$.

La ganancia de eliminación de ruido $G_{OM-LSA}(k,l)$ viene dada por:

$$G_{OM-LSA}(k,l) = \{G_{H1}(k,l)\}^{p_{spa}(k,l)} \cdot G_{min}^{1-p_{spa}(k,l)}$$

$G_{H1}(k,l)$ siendo una ganancia de eliminación de ruido (cuyo cálculo depende de la estimación del ruido $\hat{\lambda}_{Ruido}$) descrita en el artículo anteriormente mencionado de Cohen, y G_{min} siendo una constante correspondiente a la eliminación de ruido aplicada cuando se considera la voz como ausente.

Se observa aquí que la probabilidad $p_{spa}(k,l)$ desempeña una función importante en la determinación de la ganancia $G_{OM-LSA}(k,l)$. En particular, cuando esta probabilidad es nula, la ganancia es igual a G_{min} y se aplica una reducción de ruido máxima: si por ejemplo se elige un valor de 20 dB para G_{min} , los ruidos no estacionarios detectados anteriormente se atenúan en 20 dB.

La señal sin ruidos $\hat{S}(k,l)$ a la salida del bloque 22 viene dada por:

$$\hat{S}(k,l) = G_{OM-LSA}(k,l) \cdot X(k,l)$$

Se observará que comúnmente tal estructura de eliminación de ruido produce un resultado poco natural y agresivo sobre los ruidos no estacionarios, que se confunden con la voz útil. Uno de los principales intereses de la presente invención, es eliminar eficazmente estos ruidos no estacionarios.

Por otra parte, es posible utilizar en las expresiones anteriores una probabilidad de presencia de voz híbrida $p_{hibrida}(k,l)$, es decir, calculada con la ayuda de $p_{spa}(k,l)$ combinada con otra probabilidad de presencia de voz $p(k,l)$, por ejemplo calculada según el procedimiento descrito en el documento WO 2007/099222 A1 (Parrot SA). Se obtiene que:

$$p_{\text{híbrida}}(k, l) = \min(p(k, l), p_{\text{spa}}(k, l))$$

Esta probabilidad híbrida permite beneficiarse de la localización de los ruidos no estacionarios asociada a los pequeños valores de $p_{\text{spa}}(k, l)$, y completar la estimación de la probabilidad $p_{\text{híbrida}}(k, l)$ en las partes (k, l) donde la estimación de dirección de llegada $\hat{\theta}(k, l)$ no se ha definido (produciendo una probabilidad $p_{\text{spa}}(k, l)$ forzada en el valor 1 por seguridad).

La probabilidad híbrida $p_{\text{híbrida}}(k, l)$ integra de este modo a la vez los ruidos no estacionarios detectados por $p_{\text{spa}}(k, l)$ y los otros ruidos (por ejemplo pseudoestacionarios) detectados por $p(k, l)$.

Reconstitución temporal de la señal (bloque 24)

La última etapa consiste en aplicar a la señal $\hat{S}(k, l)$ una transformada de Fourier rápida inversa iFFT para obtener en el dominio de tiempo la señal de voz con ruido eliminado $\hat{s}(t)$.

REIVINDICACIONES

1. Procedimiento de eliminación de ruido de una señal acústica ruidosa captada por una pluralidad de micrófonos de un dispositivo de audio multimicrófono que opera en un medio ruidoso, en particular un dispositivo telefónico “manos libres” para vehículo automóvil, comprendiendo la señal acústica ruidosa una componente útil de voz procedente de una fuente de voz direccional y una componente parásita de ruido, incluyendo esta componente de ruido en sí misma una componente de ruido lateral no estacionaria direccional, procedimiento **caracterizado por que** consta, en el dominio de frecuencia para una pluralidad de bandas de frecuencias definidas para tramas temporales sucesivas de señal, de las siguientes etapas de procesamiento de la señal:
 - a) combinación (12) de la pluralidad de señales captadas por la pluralidad correspondiente de micrófonos en una señal combinada ruidosa ($X(k,l)$);
 - b) a partir de la señal combinada ruidosa, estimación (14) de una componente de ruido pseudoestacionaria ($\hat{V}(k,l)$) contenida en esta señal combinada ruidosa;
 - c) a partir de la componente de ruido pseudoestacionaria estimada en la etapa b) y de la señal combinada ruidosa, cálculo (16) de una probabilidad de presencia de transitorios ($p_{Transitorio}(k,l)$) en la señal combinada ruidosa;
 - d) a partir de la pluralidad de señales captadas por la pluralidad correspondiente de micrófonos y de la probabilidad de presencia de transitorios calculada en la etapa c), estimación (18) de una dirección principal de llegada de los transitorios ($\hat{\theta}(k,l)$);
 - e) a partir de la dirección principal de llegada de los transitorios estimados en la etapa d), cálculo (20) de una probabilidad de presencia de voz con respecto a un criterio espacial ($p_{ssp}(k,l)$), adecuado para distinguir entre voz útil y ruido lateral entre los transitorios;
 - f) a partir de la probabilidad de presencia de voz calculada en la etapa e) y de la señal combinada ruidosa, reducción selectiva del ruido (22) mediante la aplicación de una ganancia variable propia de cada banda de frecuencias y de cada trama temporal.
2. El procedimiento de la reivindicación 1, en el que el procesamiento de la etapa a) es un procesamiento de prefiltrado de tipo conformación fija de haz.
3. El procedimiento de la reivindicación 1, en el que el procesamiento de la etapa d) comprende las siguientes subetapas sucesivas:
 - d1) partición del espacio en una pluralidad de sectores angulares;
 - d2) para cada sector, evaluación de un estimador de dirección de llegada a partir de la pluralidad de señales captadas por la pluralidad correspondiente de micrófonos;
 - d3) ponderación de cada estimador por la probabilidad de presencia de transitorios calculada en la etapa c);
 - d4) a partir de los valores de estimadores ponderados calculados en la etapa d3), estimación de una dirección principal de llegada de los transitorios;
 - d5) validación o invalidación de la estimación de la dirección principal de llegada de los transitorios llevada a cabo en la etapa d4).
4. El procedimiento de la reivindicación 3, en el que, en la etapa d5), la estimación sólo se valida si el valor del estimador ponderado que corresponde a la dirección estimada es superior a un umbral predeterminado.
5. El procedimiento de la reivindicación 3, en el que, en la etapa d5), la estimación sólo se valida en ausencia de un máximo local del estimador ponderado en el sector angular de origen de la señal de voz útil.
6. El procedimiento de la reivindicación 3, en el que, en la etapa d5), la estimación sólo se valida si el valor del estimador es creciente de manera monótona en una pluralidad de tramas temporales sucesivas.
7. El procedimiento de la reivindicación 3, que comprende además una etapa de mantenimiento de la estimación de la dirección principal de llegada durante un periodo de tiempo mínimo predeterminado.
8. El procedimiento de la reivindicación 1, en el que la probabilidad de presencia de voz calculada en la etapa e) es una probabilidad binaria, que toma un valor de 1 ó 0 dependiendo de si la dirección principal de llegada de los transitorios estimada en la etapa d) está situada o no en el sector angular de origen de la señal de voz útil.
9. El procedimiento de la reivindicación 1, en el que la probabilidad de presencia de voz calculada en la etapa e) es una probabilidad de múltiples valores, función de la desviación angular entre la dirección principal de llegada de los transitorios estimada en la etapa d) y la dirección de origen de la señal de voz útil;

10. El procedimiento de la reivindicación 1, en el que el procesamiento de la etapa f) es un procesamiento de reducción selectiva del ruido mediante la aplicación de una ganancia de amplitud espectral logarítmica modificada optimizada OM-LSA.

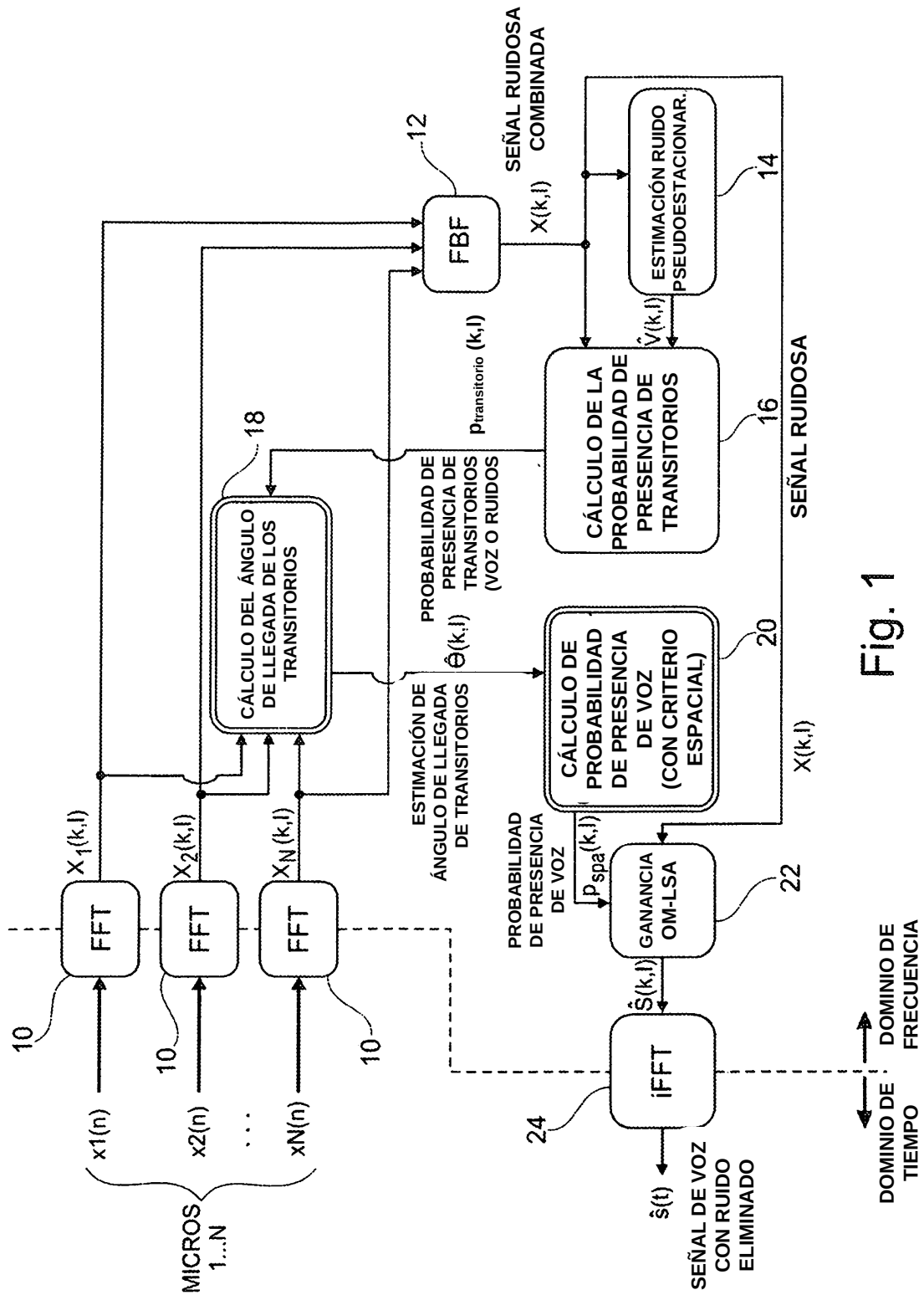


Fig. 1