(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 433 100**

(51) Int. CI.:

H04M 9/08 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **07.05.2010 E 10727770 (9)**

(97) Fecha y número de publicación de la concesión europea: **31.07.2013 EP 2430825**

(54) Título: **Procedimiento de selección de un micrófono entre dos o más micrófonos, para un sistema de tratamiento de la palabra, como un dispositivo telefónico "manos libres" que opera en un entorno ruidoso**

(30) Prioridad:

14.05.2009 FR 0953176

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
09.12.2013

(73) Titular/es:

PARROT (100.0%)
174 quai de Jemmapes
75010 Paris, FR

(72) Inventor/es:

VITTE, GUILLAUME;
BRIOT, ALEXANDRE 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).

DESCRIPCIÓN

Procedimiento de selección de un micrófono entre dos o más micrófonos, para un sistema de tratamiento de la palabra, como un dispositivo telefónico “manos libres” que opera en un entorno ruidoso

La invención se refiere al tratamiento de la palabra en un medio ruidoso.

Se refiere, en particular, pero de forma no limitativa, al tratamiento de las señales de palabra captadas por dispositivos de tipo “manos libres” para vehículos automóviles.

Estos aparatos incorporan un micrófono (“micro”) sensible que capta no solamente la voz del usuario, sino igualmente el ruido circundante, ruido que constituye un elemento perturbador hasta el punto de que, en ciertos casos, puede hacer incomprensibles las palabras del locutor. Lo mismo sucede si se quieren emplear técnicas de reconocimiento de voz, pues es muy difícil efectuar un reconocimiento de forma sobre palabras ahogadas en un nivel de ruido intenso.

Esta dificultad conectada con el ruido ambiente es particularmente restringente en el caso de los dispositivos “manos libres” para vehículos automóviles. En particular, la distancia considerable entre el micro y el locutor provoca un nivel relativo de ruido intenso que hace difícil la extracción de la señal útil ahogada en el ruido. Asimismo, el medio muy ruidoso típico del entorno automovilístico presenta características espectrales no constantes, es decir que evolucionan de manera imprevisible en función de las condiciones de la conducción: paso por calzadas deformadas o adoquinadas, radio del coche en funcionamiento, etc.

Algunos de estos dispositivos prevén la utilización 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 de perturbaciones menor. En particular, una técnica llamada *beamforming* emplea una red de micrófonos para crear, por medios software relativamente simples, una directividad que permite mejorar la relación señal/ruido. Esta técnica presupone, sin embargo, que la posición del locutor con relación a los micros sea casi constante, lo que generalmente es el caso respecto de un teléfono de coche utilizado por el conductor (pero no por el pasajero).

Con carácter general, en todas las técnicas referidas de *beamforming*, para obtener una eficacia casi satisfactoria se requiere un número importante (al menos superior a cuatro) de micrófonos, lo que no es el caso en la mayoría de las aplicaciones industriales.

Por esta razón, se utilizan más a menudo técnicas que implementan un solo micro.

En este caso es preferible utilizar un micro unidireccional, que ofrezca una mejor relación señal/ruido, pero en una sola dirección fija y, por tanto, para una sola posición del locutor.

Dicho dispositivo “manos libres” para vehículo automóvil se describe por ejemplo en el documento WO 2008/049982 A1 (Parrot SA), que describe los diversos tratamientos de anulación y de supresión de eco, de supresión de ruido, etc., aplicados a una señal captada por un único micro que recoge la señal acústica (ruidosa) del locutor próximo.

La invención tiene por objetivo dicho sistema de tratamiento de la palabra que no utiliza más que un solo micro (por tanto, diferente de los sistemas de tipo *beamforming* u otros), pero en el que este (único) micro se selecciona automáticamente entre dos micros distintos, o entre varios micros de una red de micrófonos.

El objetivo de la invención consiste en proponer un procedimiento que permita determinar entre estos dos micros o entre todos los micros de la red, aquel que capte la señal de palabra menos ruidosa, para poder activar luego diversos tratamientos (eliminación del ruido, anulación del eco, etc.) sobre la señal captada por el micro así determinado, designado a continuación como “micro de referencia”.

Se conocen diversas técnicas de selección de un micrófono entre varios, por ejemplo en un dispositivo de teleconferencia como el que se describe en el documento EP 1 564 980 A1, que comprende varios micros dirigidos hacia los participantes instalados alrededor del dispositivo. Este documento describe medios de selección dinámica susceptibles de discriminar entre palabra y ruido para no activar en un instante dado más que el micro situado frente al locutor activo.

En el caso de un dispositivo de tipo “manos libres” para vehículo automóvil indicado con anterioridad, la situación más frecuente (aunque no limitativa) es la de un sistema de dos micros. Se contemplan entonces diversas configuraciones.

Una primera configuración consiste en prever dos micros direccionales situados bastante próximos uno de otro, por ejemplo en los dos extremos del frontal de una radio de automóvil situada en la parte central del habitáculo del vehículo, teniendo uno de los micros el lóbulo principal de su diagrama de directividad orientado hacia el conductor y el otro micro hacia el pasajero.

Otra configuración consiste en prever dos micros alejados, por ejemplo situados en la parte alta, en los ángulos extremos de la luz cenital del vehículo, o en los dos extremos del salpicadero del vehículo.

En uno u otro caso, el micro más eficaz para ser seleccionado es la mayor parte de las veces el que está dirigido hacia el locutor activo (ya sea el conductor o el pasajero). Pero a veces puede ser necesario tener en cuenta situaciones particulares: así, en el caso de un motorista que adelantara al vehículo por la izquierda, podría ser preferible seleccionar el micro derecho, incluso si el locutor es el conductor.

El objetivo de la invención es proponer un sistema robusto de selección automática de micrófono para un sistema de dos micros (pudiendo evidentemente la invención aplicarse a la conmutación de un micro entre más de dos micros para la aplicación de los mismos algoritmos a cada par de micros posibles).

La idea subyacente es que, cualquiera que sea la configuración de los micros (próximos o alejados), es posible diferenciar las tomas de sonido efectuadas por cada uno de los micros partiendo de la hipótesis de que la señal captada por uno de los dos micros será mejor que la otra, es decir que uno de los micros estará en mejores condiciones de captar la señal deseada con menos perturbaciones o menos pérdidas que el otro micro.

Uno de los principios de base de la invención consiste en derivar a este efecto un índice de confianza de presencia de una palabra para cada uno de los dos canales correspondientes a los dos micros, y en condicionar la selección de uno o del otro de estos micros a una regla de decisión en función, entre otros factores, de este índice de confianza de presencia de una palabra.

Este principio permite acrecentar de forma sustancial la robustez de la selección del micro de referencia en el curso del tiempo, en función de los resultados calculados sin interrupción, adaptando esta selección a la evolución de las diversas desviaciones y a las variaciones del entorno acústico.

Los objetivos mencionados se obtienen mediante un procedimiento de tipo general divulgado en el documento EP 1 564 980 A1 precitado, es decir, mediante un procedimiento de selección de un micrófono entre dos o más micrófonos, para un sistema multimicrófono de tratamiento de la palabra que opera en un entorno ruidoso, estando cada uno de los micrófonos asociado a un canal y siendo cada uno de los micrófonos apto para captar una señal acústica ruidosa que comprende un componente útil de palabra emitida por una fuente principal de señal acústica mezclada con un componente de ruido difuso. Este procedimiento comprende las etapas de:

- digitalización de las señales acústicas captadas simultáneamente por los dos micrófonos:

- ejecución de una transformación de las señales captadas sobre los dos canales, para producir una sucesión de tramas en una serie de bandas de frecuencias:

- aplicación de un algoritmo de cálculo de un índice de presencia de palabra sobre cada canal;

- selección de uno de los dos micrófonos por aplicación a las tramas sucesivas de cada uno de los canales de una regla de decisión en función tanto de un criterio de selección de canal como de dicho índice de presencia de palabra; e

- implementación del tratamiento de la palabra a partir de la señal acústica captada por el único micrófono seleccionado.

De manera característica de la invención:

- la transformación de las señales captadas por los dos canales es una transformación de Fourier a corto plazo;

- el índice de presencia de palabra es un índice de confianza calculado para cada banda de frecuencias de cada trama; y

- el criterio de selección se calcula por bandas de frecuencias sobre las únicas bandas de frecuencias cuyo índice de confianza es superior a un primer umbral dado.

Según diversas características subsidiarias de implementación:

- el procedimiento comprende además una etapa de eliminación, en las tramas sucesivas, de las bandas de frecuencias situadas por debajo de un segundo umbral dado;

- el índice de confianza de presencia de palabra es una probabilidad de presencia de palabra;

- el índice de confianza de presencia de palabra se calcula para cada banda de frecuencias y la selección se aplica si el índice es superior a un tercer umbral dado, en un número de bandas de frecuencias superior a un cuarto umbral dado;

- el sistema multimicrófono de tratamiento de la palabra es un sistema de micrófonos directivos próximos y el criterio de selección de canal es un criterio angular basado en una comparación de las fases respectivas de las señales captadas de forma simultánea sobre los dos canales, siendo el micrófono que debe seleccionarse aquel

para el que la señal se capta antes que el otro;

- el sistema multimicrófono de tratamiento de la palabra es un sistema de micrófonos directivos o no directivos alejados y el criterio de selección de canal es un criterio energético basado en una comparación de los valores de la relación señal/ruido respectivos de las señales captadas de forma simultánea sobre los dos canales, siendo el micrófono que debe seleccionarse aquel para el que la relación señal/ruido es superior al otro;

- la selección de uno de los dos micrófonos se lleva a cabo de manera condicionada solamente si la regla de decisión se verifica sobre un número de tramas sucesivas superior a un quinto umbral dado;

- si se ha decidido una selección de uno de los micrófonos, esta se activa de forma progresiva sobre un periodo de tiempo de transición dado, por aplicación de una ganancia creciente en el canal del micrófono que debe seleccionarse y de una ganancia decreciente en el canal del micrófono que no debe seleccionarse.

A continuación se describirá un ejemplo de implementación del dispositivo de la invención, con referencia a los dibujos adjuntos, en los que las mismas referencias designan elementos funcionalmente similares.

La Figura 1 ilustra de forma esquemática una configuración con dos micrófonos próximos.

La Figura 2 ilustra de forma esquemática una configuración con dos micrófonos alejados.

A continuación se describirá el principio de base de la invención y su aplicación a los dos supuestos de las figuras indicadas anteriormente (micros próximos y micros alejados).

La hipótesis de base es que una de las señales captadas por los dos micrófonos es *a priori* mejor que la otra, es decir que contiene menos perturbaciones o pérdidas que la otra con relación a la señal de palabra útil.

El objetivo de la invención es determinar entre las dos señales captadas cuál es la mejor, con el fin de aplicar ulteriormente un tratamiento de la palabra (eliminación del ruido, anulación del eco, etc.) a esa única señal seleccionada, denominada a continuación “señal de referencia” o “señal procedente del micro de referencia”.

Dos micros captan una señal emitida por una fuente de señal principal s y la diferencia de posición entre los micros va a producir un conjunto de desfases y de variaciones de amplitud con relación a la señal emitida por s. Más exactamente, se tendrá en función del instante t:

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

donde a_n es la atenuación de amplitud debida a la pérdida de energía sobre la trayectoria acústica comprendida entre la posición de la fuente sonora s y el micro indexado n, τ_n es el desfase entre la señal emitida y recibida por el micro *n* representando el término suplementario v_n el valor del campo de ruido difuso en el lugar donde se encuentra el micro *n*.

Si se anotan en mayúsculas las transformaciones de Fourier a corto plazo de las señales anteriores, tendremos:

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

donde *k* y *l* designan, respectivamente, la banda de frecuencias y la trama considerada, con:

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

donde f_k es la frecuencia central de la banda de frecuencias indexada por *k*.

Se observará que los coeficientes $d_n(k)$ dependen únicamente de los datos geométricos del sistema y no de la naturaleza de la señal emitida s. Estos coeficientes son por otro lado característicos de la configuración geométrica considerada.

A continuación se analizarán los dos supuestos de las figuras, a saber:

- configuración con micros próximos, donde se considerará que la distancia entre los dos micros es pequeña ante la distancia entre la fuente y los micros (una definición matemática más precisa de esta condición se ofrecerá más adelante); y

- configuración con micros alejados, donde se considerará que la distancia entre la fuente y uno de los dos micros es relativamente mayor que entre la fuente y el otro micro.

Configuración con micrófonos próximos

Esta configuración, ilustrada de forma esquemática en la Figura 1, se encuentra por ejemplo en el caso de que los dos micros estén dispuestos al nivel del frontal de una radio de automóvil, estando dicha radio colocada en posición central sobre el salpicadero del vehículo. La distancia entre los micros es típicamente del orden de 5 cm, como máximo 10 cm.

Se utilizarán de forma muy ventajosa en esta configuración micros directivos, orientados de manera diferente. Si se designan los dos micros como *N* y *M*, el micro *M* tendrá, idealmente, un diagrama de directividad cuyo lóbulo principal cubra el medio plano P_m, mientras que el micro *N* tendrá un diagrama de directividad que cubra el otro medio plano P_n.

En el caso considerado aquí de micros próximos, se podrá hacer la aproximación que la fuente sonora emite una onda plana (basta con que la fuente esté alejada en al menos algunos centímetros de los micros) y se buscará simplemente evaluar la dirección de propagación de la onda en cuestión (en el espacio o en proyección sobre un plano particular). Entonces sólo será útil la información de desfase entre los micros y se utilizará entonces la fórmula:

$$\tau_n - \tau_m = \frac{1}{2\pi f_k} \arg\left(E\left(\overline{X}_m(k,l) X_n(k,l)\right)\right)$$

donde *E* designa la esperanza matemática, que podrá evaluarse por el cálculo del valor medio del producto considerado sobre varias tramas.

La hipótesis con “micros próximos” puede expresarse de manera más precisa por una condición sobre la distancia $l_{n,m}$ en los dos micros, debiendo el desfase Δø entre $x_n(t)$ y $x_m(t)$ pertenecer al intervalo $[-\pi;\pi]$ para evitar las confusiones entre retrasos y avances, es decir:

$$|\Delta\phi| = \left|\frac{2\pi f_k . l_{n,m}}{c} . \text{sen}\ \theta_{n,m}\right| < \frac{2\pi f_k . l_{n,m}}{c} < \pi$$

Si, no obstante, la distancia *D* es mayor, es posible utilizar el mismo algoritmo, pero retirando del conjunto analizado las frecuencias más altas. La frecuencia máxima se expresará entonces:

$$f_{\max} < \frac{c}{2.l_{n,m}}$$

Basta con determinar en cuál de los dos medios planos P_n o P_m se sitúa la fuente. El problema se reduce pues al estudio del signo de la expresión arg($E(\overline{X}_m(k, l) X_n(k. l))$):

$$sign\left[\theta_{n,m}(k,l)\right] = sign\left[\tau_n - \tau_m\right] = sign\left[\arg\left(E\left(\overline{X}_m(k,l) X_n(k,l)\right)\right)\right]$$

Teóricamente, para una trama dada, los ángulos $\theta_{n,m}(k, l)$ deberían ser los mismos para todas las frecuencias. En la práctica, la componente de ruido introduce ligeras variaciones. El signo de:

$$\tau_n - \tau_m = \frac{1}{2\pi f_k} \arg\left(E\left(\overline{X}_m(k,l) X_n(k,l)\right)\right)$$

por tanto se va a calcular para diferentes frecuencias y se adoptará una decisión global en función de la repartición de los resultados por frecuencias, calculando una suma de decisiones elementales no ponderadas.

En la práctica, diversos ruidos de fase entran en juego y para valores débiles de $\theta_{n,m}\ (k, l)$ la determinación del signo de ($\theta_{n,m}\ (k, l)$ puede desviarse. No obstante, la fórmula:

$$\frac{2\pi f_k l_{n,m}}{c}(\tau_n - \tau_m) + \phi_{\text{ruido}} = \arg\left(E\left(\overline{X}_m(k,l)X_n(k,l)\right)\right)$$

pone de manifiesto el hecho de que el término de ruido tendrá menos impacto en las altas frecuencias. En consecuencia, los cálculos realizados para valores débiles de *k* serán menos fiables, especialmente si el ángulo de incidencia de la señal es pequeño. Será, por tanto, deseable utilizar una ponderación con un índice de confianza más elevado en las altas frecuencias.

Como no se trata aquí de una media, sino solamente de un descuento de los signos sobre un conjunto de frecuencias, se pueden excluir de este cálculo las frecuencias más bajas (aquellas para las que $k \leq k_{min}$). En definitiva, la regla de selección será la siguiente (siendo *NbPlusThreshold* un umbral predeterminado):

para la trama *l:*

- *M* es el mejor micro si $Card(E_+) - Card(E_-) > NbPlusThreshold$

- *N* es el mejor micro si $Card(E_-) - Card(E_+) > NbPlusThreshold$

- Si no, ninguna decisión.

con:

$$\begin{cases} E_+ = \left\{k\ t.q.\ k > k_{\min}\ \text{y}\ \arg\left(E\left(\overline{X}_m(k,l)X_n(k,l)\right)\right) > 0\right\} \\ E_- = \left\{k\ t.q.\ k > k_{\min}\ \text{y}\ \arg\left(E\left(\overline{X}_m(k,l)X_n(k,l)\right)\right) < 0\right\} \end{cases}$$

En otras palabras, en lugar de medir el ángulo de incidencia de la señal procedente de la fuente a partir de los desfases relativos de las señales captadas por los micros, se determina el signo de la expresión anterior para cada banda de un subconjunto de bandas de frecuencias y se descuentan estos signos. Si el descuento muestra una mayoría significativa (determinada por el umbral *NbPlusThreshold*) a favor de uno de los micros *N* o *M*, entonces este será considerado como el micro de referencia.

Configuración con dos micros alejados

Esta configuración se ilustra de forma esquemática en la Figura 2.

La distancia entre los micros es aquí del orden de 1 a 1,5 m, correspondiente por ejemplo a una instalación en la que los dos micros están situados a cada lado de la luz cenital del habitáculo del vehículo.

En esta configuración, la información de desfase ya no es explotable, pues la distancia entre los dos micrófonos es demasiado elevada (correlación insuficiente y confusión entre retardo y avance).

Por contra, los términos a_n y a_m de atenuación de la señal son mucho más pertinentes. En efecto, en la señal captada globalmente por uno de los micros, la contribución de la fuente, es decir la señal útil, será tanto más débil cuanto más alejado esté este micro de dicha fuente.

Aquí ya no es necesario utilizar micros directivos para diferenciar las dos tomas de sonido. En efecto, esta diferenciación será intrínseca a la dirección de los micros con relación a la fuente: por ejemplo, en el caso de un automóvil, uno de los micros estará mucho más cerca del conductor y el otro del pasajero. El ruido será, por tanto, mucho más atenuado en la señal captada por uno de los micros que en la señal captada por el otro.

El principio de base consiste, en estas condiciones, en utilizar como criterio de selección el valor de la relación señal/ruido. El criterio de selección ya no es, por tanto, un criterio angular, como en el caso de los micros próximos, sino un criterio energético.

La decisión se tomará con relación a un umbral *ΔSNRThreshold* sobre el mismo principio que para el criterio angular en la configuración de micros próximos. La regla de decisión será la siguiente:

para la trama *l* :

- *M* es mejor micro si $SNR_m(l) - SNR_n(l) > \Delta SNRThreshold$

- *N* es mejor micro si $SNR_n(l) - SNR_m(l) > \Delta SNRThreshold$

- Si no, ninguna decisión

con:

$$SNR_m(l) = \frac{\sum_k X_m(k,l)}{\sum_k \hat{V}_m(k,l)} \quad \text{y} \quad SNR_n(l) = \frac{\sum_k X_n(k,l)}{\sum_k \hat{V}_n(k,l)}$$

designando $\hat{V}_m(k,l)$ y $\hat{V}_n(k,l)$ las estimaciones de la componente de ruido respectivamente captada por los micros *M* y *N.*

Refuerzo de la robustez de la selección del micro de referencia

Las técnicas de selección automática del micro de referencia están sometidas a diversas desviaciones ligadas al ruido y a las aproximaciones de cálculo.

Para evitar un número excesivo de malas detecciones, la presente invención propone completar estas técnicas de selección automática por un mecanismo global que procura una robustez excelente a los ruidos difusos y eficaz cualquiera que sea la técnica de tratamiento aplicada posteriormente a la señal de referencia seleccionada (interrelación, suma, …). Esencialmente, la invención propone derivar un índice de confianza de presencia de palabra en cada una de las señales captadas por los micros y condicionar la selección del micro de referencia a un nivel mínimo de este índice de confianza de presencia de palabra.

En efecto, es inútil pretender seleccionar uno de los dos micros antes que el otro si la señal útil de la fuente (la palabra del locutor) es nula. Convendrá, por tanto, no aplicar los criterios de selección indicados anteriormente más que sobre las tramas para las que la señal de la fuente está -verosímilmente- presente.

El índice de confianza puede ser en particular una probabilidad de presencia de palabra, calculada de la manera descrita por el documento WO 2007/0992221 A1 (Parrot SA) por ejemplo, que describe un procedimiento de eliminación de ruidos que implica el cálculo de dicha probabilidad, procedimiento derivado de la técnica expuesta por I. Cohen y B. Berdugo, Speech Enhancement for Non-Stationary Noise Environments, Signal Processing, Elsevier, vol. 81, páginas 2403 a 2481, 2001.

El cálculo de una probabilidad de presencia de palabra (o cualquier otra técnica conocida comparable), descrito en el marco de un tratamiento de reducción de ruido para el que la señal de la fuente se corresponde con una señal de palabra, puede aplicarse a la presente invención para incrementar la robustez del algoritmo de selección del micro de referencia (antes, por tanto, de cualquier tratamiento de eliminación del ruido).

Como sucede con frecuencia en el ámbito de la eliminación del ruido de una señal de palabra mezclada con una componente de ruido, no se trata de identificar precisamente sobre qué componentes frecuenciales de qué tramas la palabra está presente o ausente, sino de proporcionar un índice de confianza comprendido entre 0 y 1, indicando un valor 1 que la palabra está (según el algoritmo) ausente con seguridad, mientras que un valor 0 indica lo contrario. Por su propia naturaleza, este índice es asimilable a la probabilidad de ausencia de la palabra *a priori*, es decir la probabilidad de que la palabra esté ausente en una componente frecuencial dada de la trama considerada.

La probabilidad de presencia de palabra $P_{PresProba}$ puede, por tanto, utilizarse como un criterio suplementario para condicionar la selección de un micro o del otro, comparándola con un umbral *predeterminado PresProba-Threshold*: no se conservará así, para aplicar la regla de la selección, más que el conjunto *F(l)* de las “frecuencias pertinentes” para tener, para cada trama *l* de la señal:

$$F(l) = \{k \; t.q. \; P_{Pres\ Proba}(k\ l) > PresProbaTreshold\}$$

Si este conjunto formado por las frecuencias contenidas en la señal útil es demasiado pequeño (es decir, si *Card(F(l) < MinSize)*, no se tomará ninguna decisión para la trama actual *l*. Este criterio permite, de la misma forma que la introducción de los umbrales *NbPlusThreshold* o *ΔSNRThreshold*, limitar y controlar el número de decisiones erróneas.

En caso contrario (es decir si *Card(F(l)) > MinSize)*, el cálculo de selección del mejor micro se efectuará a partir de las únicas frecuencias pertinentes, sin tener en cuenta para la toma de decisión las frecuencias para las que el contenido de la fuente es nulo (sin palabra).

Las dos reglas explicitadas más arriba de selección del mejor micro para una trama se reescriben pues como sigue.

En el primer supuesto de las figuras (micros próximos):

para la trama *l*

- *M* es mejor micro si Card(*E*+) - Card(*E*-) > *NbPlusThreshold* y si Card*(F(l)) > MinSize*
- *N* es mejor micro si Card(*E*-) - Card(E_+) > *NbPlusThreshold* y si Card*(F(l)) > MinSize*
- Si no, ninguna decisión

con:

$$\begin{cases} E_+ = \left\{ k \in F(l) \ t.q. \ k > k_{\min} \ \text{y} \ \arg\left(E\left(\overline{X}_m(k,l) X_n(k,l)\right)\right) > 0 \right\} \\ E_- = \left\{ k \in F(l) \ \ t.q. \ k > k_{\min} \ \text{y} \ \arg\left(E\left(\overline{X}_m(k,l) X_n(k,l)\right)\right) < 0 \right\} \end{cases}$$

En el segundo caso de las figuras (micros alejados):

para la trama *l*

- *M* es mejor micro si $SNR_m(l) - SNR_n(l) > \Delta SNRThreshold$ y si Card (*F (l)) > MinSize*
- *N* es mejor micro si $SNR_n(l) - SNR_m(l) > \Delta SNRThreshold$ y si Card (*F (l)) > MinSize*
- Si no, ninguna decisión

con:

$$SNR_m(l) = \frac{\sum_{k \in F(l)} X_m(k,l)}{\sum_{k \in F(l)} \hat{V}_m(k,l)} \quad \text{y} \quad SNR_n(l) = \frac{\sum_{k \in F(l)} X_n(k,l)}{\sum_{k \in F(l)} \hat{V}_n(k,l)}$$

Cambio de micrófono de referencia

Uno de los objetivos del sistema de acuerdo con la invención es poder seguir los desplazamientos de la fuente, por tanto cambiar el micro de referencia con el paso del tiempo cuando la fuente es móvil, o cuando dos fuentes emiten en alternancia (por ejemplo el conductor y el pasajero del vehículo). Para disponer de un sistema estable, es sin embargo importante no cambiar de micro con demasiada brusquedad ni con demasiada frecuencia.

Para conseguir esto, la invención propone igualmente que las fases de cambio del micro de referencia resulten más robustas.

Para evitar las conmutaciones intempestivas, es posible condicionar la conmutación de un micro de referencia al otro (*M* en lugar de *N*, o a la inversa) a un cierto de número de reglas de estabilidad como las siguientes:

- el micro de referencia sólo se modificará si varias repeticiones consecutivas del algoritmo de selección del mejor micro por trama conducen a una decisión de cambio del micro de referencia y/o

- el cambio de micro no será repentino, sino que se producirá de forma progresiva, sobre varias tramas consecutivas aplicando una ganancia variable, creciendo en uno de los micros y decreciendo en el otro, con el fin de obtener un efecto de “fundido” progresivo.

REIVINDICACIONES

1. Un procedimiento de selección de un micrófono entre dos o más micrófonos, para un sistema multimicrófono de tratamiento de la palabra que opera en un entorno ruidoso, estando cada uno de los micrófonos asociado a un canal y siendo apto para captar una señal acústica ruidosa que comprende una componente útil de palabra procedente de una fuente principal *(s(t))* de señal acústica mezclada con una componente de ruido difusa, comprendiendo este procedimiento las etapas siguientes:

- digitalización de las señales acústicas captadas simultáneamente por los dos micrófonos (*N, M*);

- transformación de las señales ($x_n(t)$, $x_m(t)$) captadas sobre los dos canales, de forma que se produzca una sucesión de tramas en una serie de bandas de frecuencias;

- aplicación de un algoritmo de cálculo de un índice de presencia de palabra sobre cada canal;

- selección de uno de los dos micrófonos por aplicación a las tramas sucesivas de cada uno de los canales de una regla de decisión en función tanto de un criterio de selección de canal como de dicho índice de presencia de palabra; e

- implementación del tratamiento de la palabra a partir de la señal acústica captada por el único micrófono seleccionado,

procedimiento **caracterizado por que**:

- dicha transformación de las señales ($x_n(t)$, $x_m(t)$) captadas sobre los dos canales es una transformación de Fourier a corto plazo;

- dicho índice de presencia de palabra es un índice de confianza calculado para cada banda de frecuencias de cada trama; y

- dicho criterio de selección se calcula por bandas de frecuencias sobre las únicas bandas de frecuencias cuyo índice de confianza es superior a un primer canal dado.

2. El procedimiento de la reivindicación 1, que comprende también una etapa de eliminación, en las tramas sucesivas, de las bandas de frecuencias situadas por debajo de un segundo umbral dado.

3. El procedimiento de la reivindicación 1, en el que el índice de confianza de presencia de palabra es una probabilidad de presencia de palabra.

4. El procedimiento de la reivindicación 1, en el que el índice de confianza de presencia de palabra se calcula para cada banda de frecuencias y dicha selección se aplica si el índice es superior a un tercer umbral dado, en un número de bandas de frecuencias superior a un cuarto umbral dado.

5. El procedimiento de la reivindicación 1, en el que:

- dicho sistema multimicrófono de tratamiento de la palabra es un sistema con micrófonos directivos próximos, y

- dicho criterio de selección de canal es un criterio angular basado en una comparación de las fases respectivas de las señales captadas de forma simultánea sobre los dos canales, siendo el micrófono que hay que seleccionar aquel para el que la señal se capta antes que el otro.

6. El procedimiento de la reivindicación 1, en el que:

- dicho sistema multimicrófono de tratamiento de la palabra es un sistema con micrófonos directivos o no directivos alejados, y

- dicho criterio de selección de canal es un criterio energético basado en una comparación de los valores de relación señal/ruido respectivos de las señales captadas de forma simultánea sobre los dos canales, siendo el micrófono que hay que seleccionar aquel para el que la relación señal/ruido es superior al otro.

7. El procedimiento de la reivindicación 1, en el que la selección de uno de los dos micrófonos se aplica condicionalmente sólo si dicha regla de decisión se verifica sobre un número de tramas sucesivas superior a un quinto umbral dado.

8. El procedimiento de la reivindicación 1, en el que si se ha decidido una selección de uno de los micrófonos, ésta se aplica de forma progresiva durante un periodo de tiempo de transición dado, por aplicación de una ganancia creciente en el canal del micrófono que hay que seleccionar y de una ganancia decreciente en el canal del micrófono que no hay que seleccionar.

FIG_1

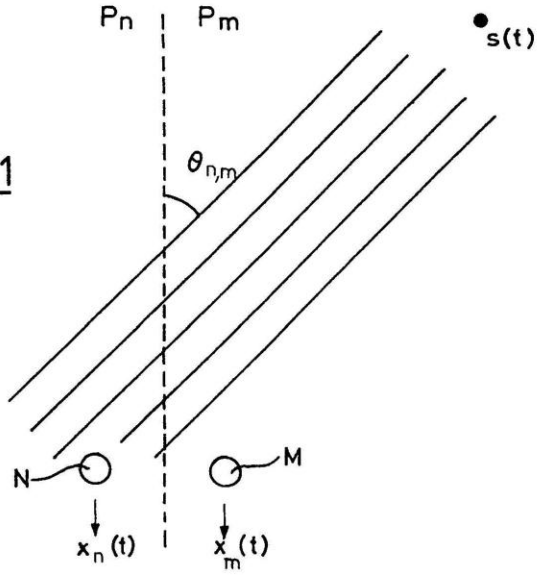
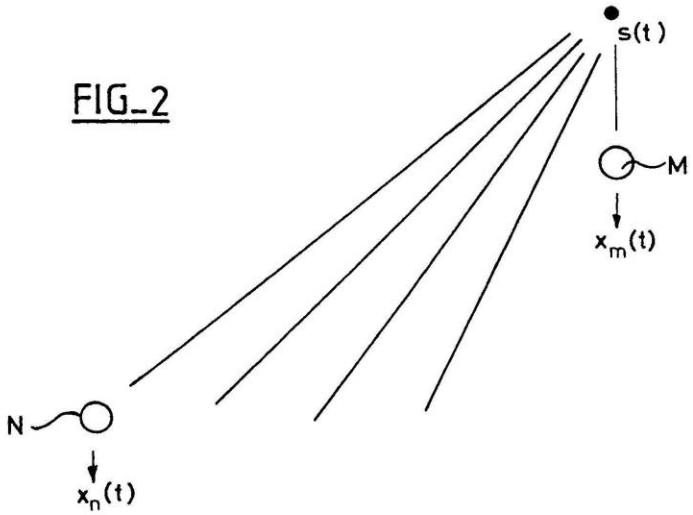
FIG_2