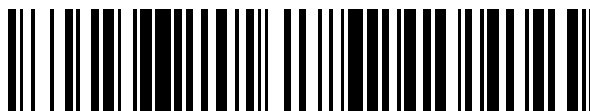


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 664 113**

51 Int. Cl.:

H04R 3/00 (2006.01)

H04S 7/00 (2006.01)

G10K 11/16 (2006.01)

H04M 3/00 (2006.01)

H04M 9/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.11.2014** **PCT/US2014/067803**

87 Fecha y número de publicación internacional: **11.06.2015** **WO15084683**

96 Fecha de presentación y número de la solicitud europea: **28.11.2014** **E 14819181 (0)**

97 Fecha y número de publicación de la concesión europea: **17.01.2018** **EP 3078210**

54 Título: **Cálculo de una respuesta de impulso de sala para cancelación de eco acústico**

30 Prioridad:

05.12.2013 US 201314098283

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
18.04.2018

73 Titular/es:

MICROSOFT TECHNOLOGY LICENSING, LLC
(100.0%)
One Microsoft Way
Redmond, WA 98052, US

72 Inventor/es:

FLORENCIO, DINEI y
YELLEPEDDI, ATULYA

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 664 113 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

Cálculo de una respuesta de impulso de sala para cancelación de eco acústico

5 ANTECEDENTES

La cancelación de eco acústico (AEC) es el problema de eliminar de la señal capturada por un micrófono, las señales radiadas por los altavoces que están situados en las proximidades del micrófono. La AEC ha sido un importante aspecto de la comunicación bidireccional, que incluye sistemas de teleconferencia, teléfonos con altavoz y similares. El escenario del problema estándar incluye un altavoz y un micrófono, y la solución tradicional es intentar un filtro adaptativo para eliminar la indicación de la retroalimentación del altavoz. Debido a la importancia del problema, aumentar la velocidad de convergencia, detectar el estado de convergencia, reducir la complejidad computacional, etc. son áreas en donde la investigación continúa avanzando.

Un segundo escenario ha ganado recientemente importancia, a saber, el control de voz para sistemas de entretenimiento de hogares y de automóviles. Estos sistemas normalmente tienen reproducción de audio multicanal, por ejemplo, estéreo o sistemas envolventes ("surround") 5.1, que funcionan en niveles de reproducción bastante elevados. El usuario (es decir, la fuente de sonido deseada) puede estar alejado varios metros del micrófono, y los niveles de ruido pueden ser significativos.

Demba Ba et al, "L1 regularized room modeling with compact microphone, arrays" (International Conference on Acoustics, Speech and Processing, 2010) describe un método para identificar distancias de pared, posiciones y coeficientes de reflexión con una configuración de micrófono pequeño y altavoz, utilizando un modelo de sala restringida.

25 SIUMARIO

En lo que sigue se presenta un resumen simplificado con el fin de proporcionar un entendimiento básico de algunos aspectos descritos en la presente. Este resumen no es una visión general extensiva de la materia objeto reivindicada. Este resumen no está destinado a identificar los elementos claves o críticos de la materia objeto reivindicada ni a definir el campo de la materia objeto reivindicada. La única finalidad de este resumen es presentar algunos conceptos de la materia objeto reivindicada de una forma simplificada como un avance de la descripción más detallada que se presenta más adelante.

La invención proporciona un método para calcular una respuesta de impulso de sala como este definido en la reivindicación 1; y un sistema para implementar la cancelación de eco acústico como está definido en la reivindicación 7.

Un ejemplo proporciona un método para calcular la respuesta de impulso de sala entre una fuente de audio y una disposición de micrófonos. El ejemplo incluye recibir señales de audio en un micrófono de una disposición de micrófonos, correspondiendo las señales de audio a cada uno de los micrófonos. El ejemplo incluye también determinar una respuesta de impulso de sala en un subespacio que es compatible con una geometría de la disposición de micrófonos en base a las señales de audio recibidas.

Otro ejemplo proporciona un método para calcular una respuesta de impulso de sala entre una fuente de audio y la disposición de micrófonos. El ejemplo incluye recibir señales de audio en un micrófono de una disposición de micrófonos, correspondiendo las señales de audio a cada uno de los micrófonos. El ejemplo incluye también definir un conjunto de señales de fuente única ya que las señales de fuente única afectan a todos los micrófonos de la disposición de micrófonos al mismo tiempo, correspondiendo cada una de las señales de fuente única a una ubicación diferente de una única fuente real o virtual. Además, el ejemplo incluye encontrar una aproximación de la señal de audio recibida como una combinación ponderada del conjunto de señales de fuente única. El ejemplo incluye calcular una respuesta de impulso de sala a partir de una fuente de audio para cada uno de los micrófonos de la disposición de micrófonos, correspondiendo la respuesta de impulso a la combinación ponderada. El ejemplo finaliza cancelando el eco acústico de la señal de audio recibida utilizando la respuesta de impulso.

Otro ejemplo proporciona un sistema para implementar la cancelación de eco acústico de retroalimentación experimentada en una disposición de una fuente de audio y una disposición de micrófonos. El ejemplo incluye un procesador para ejecutar el código ejecutable por procesador, una disposición de micrófonos, y un dispositivo de almacenamiento que almacena el código ejecutable por procesador. Cuando el código ejecutable por procesador es ejecutado por el procesador, del sistema y hace que el procesador reciba señales de audio en un micrófono de la disposición de micrófonos, correspondiendo la señales de audio a cada uno de los micrófonos. El procesador está configurado para definir un conjunto de señales de fuente única ya que las señales de fuente única afectarían a todos los micrófonos de la disposición de micrófonos al mismo tiempo, correspondiendo cada una de las señales de fuente única a una ubicación diferente de una única fuente real o virtual. El procesador encuentra una aproximación de la señal de audio recibida como una combinación del conjunto de señales de fuente única. El procesador está configurado para calcular una respuesta de impulso de sala desde una fuente de audio a cada uno de los micrófonos de la disposición de micrófonos, correspondiendo la respuesta de impulso a la combinación ponderada. El procesador finalmente cancela el eco acústico de la señal de audio recibida utilizando la respuesta de impulso.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La siguiente descripción detallada se puede entender mejor haciendo referencia a los dibujos adjuntos, que contienen ejemplos específicos de numerosas características de la materia objeto descrita.

La Figura 1 es una vista esquemática de una sala en donde las reflexiones de señal están modelizadas como fuentes de imagen virtuales.

La Figura 2 es un diagrama de bloques de un ejemplo de un dispositivo de cálculo que puede ser utilizado para la AEC utilizando procesamiento de señal de disposición dispersa.

La Figura 3 representa un modelo de sistema de la disposición de micrófonos y la configuración de altavoces.

La Figura 4 es una vista esquemática de las señales audibles que llegan a una disposición 1-d plana.

La Figura 5 es un diagrama de flujo de un método a modo de ejemplo para recibir señales de audio y para calcular una respuesta de impulso.

La Figura 6 es un diagrama de flujo de proceso de un método a modo de ejemplo para cancelar el eco acústico utilizando procesamiento de señal de disposición dispersa.

La Figura 7 es un diagrama de bloques que muestra un medio de almacenamiento tangible, leíble por un ordenador, que puede almacenar instrucciones para cancelar el eco acústico utilizando procesamiento de señal de disposición.

DESCRIPCIÓN DETALLADA

Se considera el problema de la cancelación de eco acústico en un ambiente reverberante con una disposición de micrófonos y uno o más altavoces. Cancelar el eco requiere aprender la respuesta de impulso procedente de un cierto número de altavoces para cada micrófono en una disposición de micrófonos. Esto se ha realizado de manera conveniente separadamente en cada micrófono para cada altavoz. Sin embargo, las señales que llegan a la disposición comparten una estructura común, que puede ser aprovechada para mejorar los cálculos de respuesta de impulso.

Las realizaciones presentadas aquí están destinadas al cálculo de la función de transferencia de sala (RTF) inicial, en base a una señal de calibración. Las técnicas descritas aquí se pueden extender al problema más complejo de filtros adaptativos, sin embargo las disposiciones de micrófonos más comúnmente desplegadas utilizan una señal de calibración para iniciar los filtros, y más tarde adaptar un subconjunto de coeficientes de filtro. Las realizaciones presentes están dirigidas a, pero no limitadas, al problema más relevante de la estimación de RTF inicial. La dispersión de RTF es bien conocida, y ha sido utilizada en cancelación de eco. Véase J Benesty, et al. "Adaptive algorithms for the identification of sparse impulse responses," Selected methods for acoustic echo and noise control, vol. 5, pp. 125-153, 2006. La interacción entre la formación de haces (o filtrado espacial) y la cancelación de eco también ha sido considerada. Véase W. Herbordt, et al. "Joint optimization of acoustic echo cancellation and adaptive beamforming," Topics in acoustic echo and noise control, pp. 19-50, 2006. Sin embargo, en aquellos casos el cálculo de filtro de cancelación de eco por sí mismo no obtiene ventaja de la información de disposición. Las técnicas anteriores calculan cada filtro AEC separadamente y no aprovechan la estructura de disposición como se detalla en la presente memoria.

Como materia preliminar, algunas de las figuras describen conceptos en el contexto de uno o más componentes estructurales, referidos como funcionalidades, módulos, características, elementos, etc. Los distintos componentes mostrados en las figuras pueden ser implementados de cualquier forma, por ejemplo, mediante software, (por ejemplo componentes lógicos discretos, etc.), firmware, etcétera, o cualquier combinación de estas implementaciones. En algunas realizaciones, los diversos componentes pueden reflejar el uso de componentes correspondientes en una implementación real. En otras realizaciones, cualquier componente único ilustrado en las figuras puede ser implementado por un cierto número de componentes reales. La descripción de cualesquiera dos o más componentes separados en las figuras puede reflejar diferentes funciones realizadas por un componente real único.

Otras figuras describen los conceptos en forma de diagrama de flujo. De esta forma, ciertas operaciones están descritas como bloques distintos constituyentes realizados en cierto orden. Tales implementaciones son a modo de ejemplo y no limitativas. Ciertos bloques descritos aquí pueden ser agrupados juntos y realizados en una única operación, ciertos bloques pueden ser separados en varios bloques de componentes y ciertos bloques pueden ser realizados en un orden que difiera del que se ilustra en la presente, incluyendo una manera paralela de realizar los bloques. Los bloques mostrados en los diagramas de flujo pueden ser implementados por software, hardware, firmware, procesamiento manual, y similares, o cualquier combinación de estas implementaciones. Como se ha utilizado aquí, el hardware puede incluir sistemas de ordenador, componentes lógicos discretos, tales como circuitos integrados de aplicación específica (ASICs), y similares, así como cualquier combinación de los mismos.

Por terminología, la expresión "configurado para" engloba cualquier manera en la que cualquier tipo de componente estructural puede ser construido para realizar una operación identificada. El componente estructural puede estar configurado para realizar una operación utilizando software hardware, firmware y similares, o cualquier combinación de los mismos.

El término "lógico" engloba cualquier funcionalidad para realizar una tarea. Por ejemplo, cada operación ilustrada en los diagrama de flujo corresponde a la lógica para realizar esa operación. Una operación puede ser realizada utilizando software, hardware, firmware, etc., o cualquier combinación de los mismos.

Como se ha utilizado aquí, los términos "componente", "sistema", "cliente" y similares están destinados a referirse a una entidad relacionada con ordenadores, o bien hardware, software (por ejemplo, en ejecución) y/o firmware, o una combinación de los mismos. Por ejemplo, un componente puede ser un proceso que se ejecute en un procesador, un objeto, un ejecutable, un programa, una función, una librería, una subrutina, y/o un ordenador o una combinación de software y hardware. A modo de ilustración, tanto una aplicación que se ejecuta en un servidor como el servidor pueden ser un componente. Uno o más componentes pueden residir dentro de un proceso y un componente puede estar localizado en un ordenador y/o distribuido entre dos o más ordenadores.

Además, la materia objeto reivindicada puede ser implementada como un método, aparato, o artículo de fabricación, utilizando técnicas de programación y/o ingeniería para producir software, firmware hardware o cualquier combinación de los mismos, para controlar un ordenador para implementar la materia objeto descrita. La expresión "artículo de fabricación" como se ha utilizado aquí, está destinada a englobar un programa de ordenador accesible desde cualquier dispositivo o medio tangible, leíble por un ordenador.

El medio de almacenamiento leíble con un ordenador puede incluir, pero no estar limitado a, dispositivos de almacenamiento magnéticos, (por ejemplo, discos duros, discos flexibles, o cintas magnéticas entre otros), discos ópticos, (por ejemplo discos compactos (CD), y discos versátiles digitales (DVD), entre otros), tarjetas inteligentes, dispositivos de memoria temporal (por ejemplo tarjetas, pinchos y lápices de memoria, entre otros). Por el contrario, los medios leíbles con un ordenador generalmente (es decir medios no de almacenamiento) pueden adicionalmente incluir medios de comunicación, tales como medios de transmisión o señales inalámbricas y similares.

Las realizaciones de las técnicas descritas en la presente están dirigidas a, pero no se limitan a, el cálculo de la función de transferencia de sala. Las presentes realizaciones consideran la incorporación de información de disposición en el cálculo de los filtros de cancelación de eco para conseguir el resultado deseado de cancelación de eco acústico.

La Figura 1 es un esquema de una sala 100 en donde las reflexiones de señal audibles 110 están modelizadas como fuentes de imagen virtuales 108. El altavoz 102 emite ondas de sonido 104 en todas direcciones que reverberan dentro de la sala 100, se reflejan en las paredes y otras superficies antes de ser recibidas en la disposición de micrófonos 106. Cada reflexión 110 corresponde a una fuente de imagen 108 que es utilizada en los ejemplos descritos en la presente para simular acústicos dentro de la sala 100. La señal recibida en la disposición de micrófonos 106 es la superposición de la señal recibida a lo largo de un cierto número de trayectorias entre el altavoz 102 que transmite la señal y el receptor en la disposición de micrófonos 106.

Una realización de la técnica actual incorpora un algoritmo que obtiene ventaja de la estructura de disposición de micrófonos 106, así como la dispersión de las reflexiones 110 que llegan a la disposición con el fin de formar cálculos de la respuesta de impulso entre cada altavoz 102 y el micrófono 106. El algoritmo se muestra para mejorar el rendimiento sobre un algoritmo de filtro coincidente tanto en los datos sintéticos como reales. Una realización utiliza una geometría de disposición de micrófono conocida para mejorar el cálculo de la función de transferencia de sala (RTF) en cada micrófono. Esto se consigue explotando las regularidades que aparecen en las RTFs, que pertenecen al hecho de que los micrófonos están en la misma sala, recibiendo señales de los mismos altavoces y rebotando en los mismos obstáculos. Aunque suponemos que la geometría de la disposición de micrófonos es conocida, se ha de observar que la geometría de la disposición de micrófonos 106 no necesita ser conocida antes de recoger las señales descritas en la presente. En otras palabras, no es necesario conocer la configuración de la disposición de antemano, ya que ésta se puede calcular utilizando las señales recibidas. Una persona familiarizada con la técnica sabrá cómo utilizar la correlación y las múltiples señales recibidas para calcular la geometría de disposición.

La Figura 2 es un diagrama de bloques de un ejemplo de un dispositivo de ordenador 200 que puede ser utilizado para AEC utilizando procesamiento de señal de disposición dispersa. El sistema de ordenador 200 puede ser, por ejemplo el Xbox Kinect® utilizado en combinación con la plataforma Xbox 360® o Xbox One® de Microsoft®, o dispositivos de captura de audio similares. En algunas realizaciones, el sistema de ordenador 200 puede ser un ordenador portátil, por ejemplo, utilizado por un revisor para proporcionar retroalimentación. En otra realización, el sistema de ordenador 200 puede estar incorporado dentro de la presentación de cuadro de instrumentos y la interfaz de usuario en muchos automóviles más nuevos. El sistema de ordenador 200 puede incluir un procesador 202 que está adaptado para ejecutar las instrucciones almacenadas, así como un dispositivo de memoria 204 que almacene las instrucciones que son ejecutadas por el procesador 202. El procesador 202 puede ser un procesador de núcleo único, un procesador de múltiples núcleos, un grupo de ordenadores, o cualquier número de otras configuraciones. El dispositivo de memoria 204 puede incluir memoria de acceso aleatoria (por ejemplo SRAM, DRAM, RAM de condensador cero, SONOS, eDRAM, EDO RAM, DDR RAM, RRAM, PRAM, etc.), memoria de solo lectura (por ejemplo ROM de Máscara, PROM, EPROM, EEPROM, etc.) memoria temporal, o cualesquiera otros sistemas de memoria adecuados. Las instrucciones que son ejecutadas por el procesador 202 pueden ser utilizadas para

cancelación de eco acústico utilizando procesamiento de señales basado en disposiciones representadas dispersamente.

El procesador 202 puede estar conectado a través de un bus de sistema 206 (por ejemplo un bus propietario, PCI, ISA, ECI-Express, HyperTransport®, NuBus, etc.) a una interfaz de dispositivo (I/O) de entrada/salida 208 adaptada para conectar el sistema de ordenador 200 a uno o más dispositivos I/O 210. Los dispositivos I/O 210 pueden incluir, por ejemplo, una cámara, un dispositivo de entrada de reconocimiento de gesto, un teclado, un dispositivo de señalamiento, un dispositivo de reconocimiento de voz, y una interfaz de red, entre otros. El dispositivo señalamiento puede incluir un panel táctil o una pantalla táctil, entre otros. En la descripción actual, un dispositivo I/O importante es la disposición de micrófono 106 que está conectada al sistema de ordenador 200. Los dispositivos I/O 210 pueden ser componentes integrados del sistema de ordenador 200, o pueden ser dispositivos que están conectados externamente al sistema de ordenador 200.

El procesador 202 también puede estar enlazado a través del bus de sistema 206 a una interfaz de dispositivo de presentación 212 adaptado para conectar el sistema de ordenador 200 a un dispositivo de presentación 214. El dispositivo de presentación 214 puede incluir una pantalla de presentación que es un componente integrado del sistema de ordenador 200. El dispositivo de presentación 214 también puede incluir un monitor de ordenador, televisión, o proyector, entre otros, que esté conectado externamente al sistema de ordenador 200.

El almacenamiento 216 puede estar conectado al procesador 202 a través del bus 206. El almacenamiento 216 puede incluir un disco duro, un disco duro sólido, o un disco óptico, un disco de memoria temporal USB, o una disposición de discos, o cualquier combinación de los mismos. El almacenamiento 216 puede incluir un cierto número de módulos configurados para implementar la cancelación de eco acústico utilizando procesamiento de señal de disposición dispersa como se ha descrito en la presente. Por ejemplo, el almacenamiento 216 puede incluir un módulo de disposición de señal 218 configurado para disponer las señales recibidas 110 en todos los micrófonos de la configuración de micrófonos 106 en un único vector grande.

El almacenamiento 216 puede incluir además un módulo de vector de base 220 para calcular un vector de base que incorpora una distancia de fuente desde y un ángulo de una señal recibida a la disposición de micrófono 106. Para cada posible ángulo y cada posible tiempo de retraso, que puede ser calculado como una distancia, el vector base puede ser definido como la señal de altavoz si ha estado llegando desde un ángulo y retraso específico al centro de la disposición, apilada para cada micrófono de la disposición de micrófonos 106. En una realización, todos los ángulos y retrasos posibles están discretizados y sometidos a un solucionador disperso antes de que se pueda generar una respuesta de impulso que cancele el eco acústico experimentado en la disposición de micrófonos 106. De este modo, la señal recibida se convierte en una combinación ponderada de unos pocos de los vectores base (unos pocos debido a que el posible número de ángulos y retrasos es mucho mayor que los realmente recibidos en la disposición de micrófonos 106).

También incluido en el almacenamiento 106 puede haber un módulo solucionador disperso 222. Debido a que hay millones de ángulos y retrasos, y dado que cada uno de ellos necesita ser representado por un vector base, el tamaño total de la matriz base es enorme. Cada posible ubicación va a ser representada, si aquellas señales representadas por el vector base son reales o virtuales. A través de un proceso de optimización convexo, el módulo solucionador disperso 222 utiliza vectores base ponderados y determina los pesos de un valor mínimo que satisface un error de reconstrucción particular. El proceso de optimización puede ser, por ejemplo Eliminación de Ruido de Búsqueda de Base (BPDN), un método bien conocido para resolver el problema de la representación aproximada con restricciones dispersas. Un solucionador disperso también incluye tales solucionadores como el solucionador SPGL, un solucionador Matlab™ para mínimos cuadrados regularizados de una normal de gran escala, que confía en las operaciones de vector de matriz.

Un módulo de respuesta de impulso 224 también puede estar incluido en el almacenamiento 216. El módulo de respuesta de impulso 224 está configurado para utilizar los pesos base determinados a partir del módulo solucionador disperso 222 y estimar la respuesta de impulso del altavoz 102 a cada uno de los micrófonos de la disposición de micrófonos 106. Una respuesta de impulso se puede estimar utilizando un vector de base que se corresponde con el vector de base que es calculado por el módulo de vector base 220. Una matriz base correspondiente también puede ser configurada por el módulo de respuesta de impulso 224, y un vector procedente de esta matriz base puede contener los cálculos de respuesta de impulso apilados en orden de cada micrófono en la disposición de micrófonos 106. El almacenamiento 216 puede además incluir un módulo de cancelación de eco 226 que utiliza los cálculos de respuesta de impulso para cancelar el eco acústico recibido en la disposición de micrófonos 106. El módulo de cancelación de eco 226 filtra las señales recibidas en cada micrófono y resuelve el problema de AEC.

Se ha de entender que el diagrama de bloques de la Figura 2 no está destinado a indicar que el sistema de ordenador 200 es para incluir todos los componentes mostrados en la Figura 2. En su lugar, el sistema de ordenador 200 puede incluir menos o adicionales componentes no ilustrados en la Figura 2, por ejemplo, aplicaciones adicionales, módulos adicionales, dispositivos de memoria adicionales, interfaces de red adicionales (no mostradas) y similares. Además, el sistema de ordenador 200 no está limitado a los módulos mostrados, ya que cualesquiera

combinaciones del código utilizados para implementar estas funciones pueden ser implementadas.

La Figura 3 representa un modelo de sistema de la disposición de micrófono y la configuración de altavoz. Durante la fase de calibración, una señal de prueba $l(t)$ es reproducida en el altavoz 302. Utilizamos una notación de tiempo discreto, y utilizamos n como nuestro índice de tiempo. De este modo designamos la versión de tiempo discreto de $l(t)$ como $l[n]$. Una realización a modo de ejemplo utiliza múltiples altavoces, y este proceso puede tener que ser repetido para cada altavoz separadamente. La RTF desde el altavoz 302 al $k^{\text{ésimo}}$ micrófono 304 se designa como $h_k[n]$ 306, y se supone que es invariable de tiempo, pero la adaptación se puede realizar para rastrear características que varían en el tiempo. De este modo, la señal recibida en el micrófono k 304 esta designada por la Ecuación (1):

$$x_k[n] = \sum_{m=0}^{N-1} h_k[m]l[n-m] + d_k[n] + v_k[n] \quad (1)$$

En la Ecuación 1, m es una variable ficticia utilizada para computar el suma $d_k[n]$ es la señal deseada, n es la notación tradicional para índice de tiempo, y $v_k[n]$ es el ruido de interferencia. Nótese que se supone que $d_k[n]$ es cero durante la fase de calibración. Si las RTFs $h_k[n]$ 306 fueran conocidas, el eco simplemente sería restado de la señal recibida. En su lugar, los filtros de cancelación $\hat{h}_k[n]$, son aproximaciones que son calculadas y que son utilizadas para eliminar (parcialmente) el eco de $x_k[n]$ calculando una señal de salida $y_k[n]$ dada por:

$$y_k[n] = x_k[n] - \sum_{m=0}^{N-1} \hat{h}_k[m]l[n-m]. \quad (2)$$

En la ecuación (2), las variables iguales son las mismas que las definidas en la Ecuación (1).

Quando solo están presentes un micrófono y un altavoz, el error cuadrático mínimo estimado de la RTF se conoce como el "filtro coincidente" que puede ser calculado como:

$$h_k^{MF}[n] = \frac{\sum_m \tilde{l}[m]\tilde{x}_k[n+m]}{\sum_m (\tilde{l}[m])^2} \quad (3)$$

En la Ecuación (3) $\tilde{l}[n]$ se refiere a la versión blanqueada de $l[n]$, y $\tilde{x}_k[n]$ se refiere a $x_k[n]$ después de que sea filtrada por el mismo filtro utilizado para blanquear la señal del altavoz $l[n]$. Un procedimiento de blanqueado estadístico transforma beneficiosamente los datos de manera que tengan una matriz de covarianza de identidad, en donde todas las muestras son estadísticamente independientes. El filtro de cancelación de la Ecuación (3) es el cálculo de línea de base con el que se compara la técnica de cancelación de eco actual. Además, un cálculo de cada RTF obtenida en el calculador 308 aprovechando los múltiples micrófonos en la disposición de micrófonos puede ser mejor que el óptimo anterior.

La Figura 4 es una vista esquemática de las señales audibles que llegan a una disposición plana 400. A lo largo de cualquier trayectoria, la señal tiene un retraso de propagación al centro de la disposición, que depende de la trayectoria, y una ganancia (que posiblemente depende de la frecuencia), que depende de la naturaleza de las reflexiones 110 tomadas. Suponiendo que la reflexión 110 llega en la configuración 106 como una onda plana, la estructura de la señal recibida a lo largo de una trayectoria particular está indicada en la Figura 4. Para una mayor simplicidad de notación, se supone una disposición de micrófonos lineal, y que los micrófonos están alineados a lo largo del eje z , como en la Figura 4, y situados en las posiciones $(z_1, z_2, \dots, z_k)$ 106. Para la onda plana que incide en la disposición lineal 106, si la señal de tiempo continua en el origen 402 es $s(t)$, entonces la señal en un micrófono situado en z_k es $s(t-T(z_k, \theta))$, en donde $T(z_k, \theta) = -z_k \cos(\theta)/c$, y c es la velocidad del sonido. En otras palabras, la señal producida en cada micrófono 106 por la onda plana es la misma, excepto para un retraso (T) que depende solo de la geometría de disposición en la dirección de llegada (θ) 404.

Cada reflexión 110 puede corresponder a una fuente de imagen virtual 108 como se indica en la Figura 1. La ecuación (4), más adelante, supone primero que todas las dimensiones de la disposición 106 son lo suficientemente pequeñas como para que los frentes de ondas puedan ser aproximadas como planos, supone además que cada reflector se desvanece de forma uniforme, y finalmente suponen que P número de fuentes (real o imagen) son suficientes para proporcionar una aproximación de la señal de llegada. Después de estas suposiciones, la señal recibida en el micrófono k de la disposición de micrófonos 106 se puede escribir como:

$$x_k[n] = \sum_{p=1}^P \alpha_p l \left(nT_s - T_p - \frac{z_k}{c} \cos \theta_p \right) + \tilde{v}_k[n] \quad (4)$$

En la ecuación (4), T_p y α_p son, respectivamente, el retraso y la atenuación sufrida por el frente de ondas p (incluyendo la pérdida de propagación y las directividades del micrófono y del altavoz). Además, c es la velocidad del sonido, $\tilde{v}_k[n]$ incluye tanto el ruido real $v_k[n]$ como la posibilidad de cualesquiera componente de señal no modelizados, y θ_p es el ángulo hecho por la señal que llega a lo largo de la p -ésima trayectoria con el eje de la disposición. T_s es el intervalo de muestreo y z_k es la posición del k -ésimo micrófono en el eje z como se ha definido anteriormente. Un conjunto de señales de fuente única pueden ser definidas como señales de fuente única que inciden en todos los micrófonos de la disposición de micrófonos al mismo tiempo, cada una de las señales de única fuente correspondiente a una ubicación diferente una única fuente real o virtual.

La Ecuación (4) y la Figura 4 asumen que la disposición de micrófonos 106 es uniforme y está linealmente alineada. Sin embargo, esta configuración solo proporciona un ejemplo y no se requiere implementar la revelación actual. Efectivamente la disposición no necesita ser lineal en general. Además, la Ecuación (4) implícitamente supone que la ganancia es independiente de la frecuencia, que tampoco es requerida, pero se supone en la implementación de las técnicas descritas aquí.

La Figura 5 es un diagrama de flujo de proceso de un método a modo de ejemplo para recibir señales de audio y calcular una respuesta de impulso. El método 500 empieza en el bloque 502 cuando una señal de audio es reproducida en uno o más altavoces, y una correspondiente señal es recibida en un micrófono de una disposición de micrófonos, posiblemente contaminados con el ruido ambiente. El método incluye una etapa en el bloque 504 en donde es calculada la familia de señales que posiblemente podría llegar a los micrófonos como resultado del sonido reproducido en el altavoz. El método incluye también una determinación en el bloque 506 de una respuesta de impulso de sala en un subespacio que es compatible con una geometría de la disposición de micrófonos basada en las señales de audio recibidas.

El diagrama de flujo de proceso de la Figura 5 no está destinado a indicar que las etapas del método 500 van a ser ejecutadas en un orden particular, o que todas las etapas del método 500 van a ser incluidas en cada caso. Además, cualquier número de etapas adicionales puede ser incluido dentro del método 500, dependiendo de la aplicación específica.

La Figura 6 es un diagrama de flujo de proceso de un ejemplo de un método 500 para implementar la cancelación de eco acústico utilizando el procesamiento de señal de disposición dispersa. El método puede ser implementado por un dispositivo de ordenador 200 descrito con respecto a la Figura 2. El método 600 empieza en el bloque 602 en donde las señales de audio son recibidas en la disposición de micrófono. Las señales recibidas son después dispuestas en el bloque 604 y representadas con un único vector matemático grande. En el bloque 606, un vector base está definido como una señal enviada por un altavoz para múltiples ángulos de audición y retrasos de tiempo que corresponden a muchas fuentes virtuales. Las fuentes virtuales o fuentes de imagen pueden ser representaciones del amplio número de posibles reflexiones de señal que pueden reverberar dentro de una sala antes de ser recibidas en la disposición de micrófonos. No todos los vectores posibles serán de este modo analizados, y se puede imponer la construcción ponderada para proporcionar un conjunto de datos más fácilmente calculable.

El método de este modo continúa en el bloque 608 en donde las señales de audio recibidas, si son virtuales o reales, son interpretadas como una función ponderada de una combinación de un número de vectores base. Los vectores base y los correspondientes pesos pueden ser optimizados como resultado de la mejor representación del vector observado en algún espacio de eco deseado. La optimización puede ser modelizada a través de lo siguiente:

$$\hat{w} = \underset{w}{\operatorname{argmin}} \|w\|_1, \quad \text{s.t.} \quad \|Bw - x\|_2 \leq \sigma. \quad (5)$$

En la Ecuación (5) $\hat{w}$ es el vector de peso mejor calculado (es decir, el vector que maximiza la expresión en (5)), B es la matriz grande que describe el espacio de eco deseado, y formada componiendo todos los vectores base individuales; x es el vector observado; y σ es el error de reconstrucción permitido. En una realización a modo de ejemplo, σ es establecido en un valor ligeramente mayor que la desviación estándar de los componentes no modelizados. Más concretamente, mientras que cualquier valor de σ producirá resultados, en este ejemplo, $\sigma = \beta \sigma_v$ para algún $\beta > 1$, en donde σ_v es la desviación estándar de ruido, y β es un número entero positivo. En otra realización, β puede estar establecido en 1,5.

Debido a la gran cantidad de datos que pueden representar las señales audibles, la matriz B probablemente va a ser enorme, y probablemente va a constituir una base más que completar del espacio de señal. De este modo, se hace necesario encontrar formas eficientes de resolver (5), así como de encontrar formas eficientes de manejar la

condición de sobrecompleto. Esto se puede realizar imponiendo una restricción de dispersión (incluido en la métrica de norma-1 indicada en (5)), y utilizando un solucionador disperso. Esto permite que los datos en B sean almacenados y se acceda a ellos más fácilmente. En el bloque 610, un solucionador disperso es utilizado para determinar los pesos que cumplen un error de reconstrucción particular. Como se ha indicado, este error de reconstrucción puede idealmente ser ligeramente mayor que la desviación estándar de los componentes no modelizados. El error de reconstrucción puede ser, por ejemplo, una figura próxima a un error de modelización más ruido. Un proceso de limpieza puede ser también implementado para mejorar más el rendimiento. La limpieza recalcula todos los coeficientes $w_{\theta, t}$, encontrados en (5) que no son cero (o por encima de cierto umbral), minimizando el valor de $\|x - \sum_{\theta, t} w_{\theta, t} b_{\theta, t}\|_2$. En donde $b_{\theta, t}$ es el vector base correspondiente a una reflexión que llega en un ángulo θ con un retraso t y $w_{\theta, t}$ es el peso asociado con ese vector base (y que ha sido encontrado como distinto de cero en (5)).

En el bloque 612, una respuesta de impulso es calculada a partir del altavoz para cada uno de los micrófonos de la disposición de micrófonos. La respuesta de impulso corresponde con una estimación para el eco acústico generado por el altavoz y es utilizada de manera efectiva para cancelar tal eco cuando se recibe en la disposición de micrófonos. El cálculo de la respuesta de impulso puede ser calculada como, por ejemplo, el siguiente vector:

$$\hat{h} = \sum_{\theta, t} \hat{w}_{\theta, t} h_{\theta, t} \quad (6)$$

En el bloque 614, el vector de estimación de respuesta de impulso de la ecuación (6) puede ser utilizado para cancelar el eco acústico de la representación de señales de audio recibidas en todos los micrófonos de la disposición de micrófonos.

El diagrama de flujo de proceso de la Figura 6 no está destinado a indicar que las etapas del método 600 van a ser ejecutadas en un orden particular, o que todas las etapas del método 600 van a ser incluidas en cada caso. Además, cualquier número de etapas adicionales puede ser incluido dentro del método 600, dependiendo de la aplicación específica.

La Figura 7 es un diagrama de bloques que muestra un medio de almacenamiento tangible 700, leíble por un ordenador, para ejecutar AEC utilizando procesamiento de señal de disposición dispersa. Al medio de almacenamiento tangible 700, leíble con ordenador, se puede acceder mediante un procesador 702 sobre un bus de ordenador 704. Además, el medio de almacenamiento tangible 700, leíble con ordenador, puede incluir el código para dirigir el procesador 702 para realizar las etapas de las técnicas descritas en la presente.

El medio de almacenamiento tangible 700, leíble con ordenador, puede incluir el código 706 configurado para dirigir el procesador 702 para recibir y procesar señales de audio en una disposición de micrófonos. Además, el medio de almacenamiento tangible 700, leíble con ordenador, puede incluir el código 708 configurado para dirigir el procesador 702 para facilitar la disposición de vector. Por ejemplo, el código puede dar instrucciones al procesador 702 para definir un único vector grande de las posibles representaciones de la señal de audio recibida. Otro bloque de código 710 puede dar instrucciones al procesador para definir el vector base que puede ser una combinación de señales de fuente reales o virtuales. Un bloque de código 712 puede entonces dar instrucciones al procesador 702 para crear una función ponderada sobre los valores del vector base, con el fin de reducir la cantidad de datos representativos a un tamaño más manejable. El medio de almacenamiento tangible 700, leíble por un ordenador, también puede incluir el código 714 para dar instrucciones al procesador 702 para implementar una técnicas solucionadoras para imponer una restricción de dispersión sobre la todavía gran cantidad de datos definidos por el vector de base ponderado. El bloque del código 716 puede entonces dirigir el procesador 702 para calcular las estimaciones de respuesta de impulso en cada micrófono de la disposición de micrófonos. Las estimaciones de respuesta de impulso pueden entonces ser utilizadas a través del bloque de código 718 para cancelar el eco acústico que es típicamente generado en sistemas de audio que tienen micrófonos que registran señales de audio reflejadas retrasadas.

Se ha de entender que cualquier número de componentes de software adicionales no mostrado en la Figura 7 puede ser incluido dentro del medio de almacenamiento tangible 700, leíble con ordenador, dependiendo de la aplicación específica. Aunque la materia objeto ha sido descrita en un lenguaje específico de las características y/o métodos estructurales, se ha de entender que la materia objeto definida en las reivindicaciones adjuntas no está necesariamente limitada a las características estructurales específicas o a los métodos descritos anteriormente. En su lugar, las características estructurales específicas y los métodos descritos anteriormente están descritos como formas a modo de ejemplo de implementación de las reivindicaciones.

REIVINDICACIONES

1. Un método para calcular una respuesta de impulso de sala (100) entre una fuente de audio (102, 302) y una disposición de micrófonos (106), que comprende:

recibir señales de audio en cada micrófono de la disposición de micrófonos (106), estando la señales de audio basadas en una señal de calibración;
calcular la geometría de la disposición de micrófonos;
definir un conjunto de señales de fuente única ya que la señales de fuente única incidirían en todos los micrófonos de la disposición de micrófonos (106) al mismo tiempo, correspondiendo cada una de las señales de fuente única a una ubicación diferente de una única fuente real o virtual en donde la fuente virtual representa una reflexión;
crear una aproximación de la señales de audio recibidas como una combinación ponderada del conjunto de señales de fuente única, y
calcular una respuesta de impulso de sala (100) desde cada fuente de audio (102, 302) a cada uno de los micrófonos de la disposición de micrófonos (106), correspondiendo la respuesta de impulso a la combinación ponderada y estando la respuesta de impulso de sala basada al menos en parte en la geometría de la disposición de micrófonos, y un vector base que comprende una señal enviada por un altavoz para múltiples ángulos de audición y retrasos de tiempo que corresponden a muchas fuentes virtuales; y
cancelar el eco acústico de la señal recibida utilizando la respuesta de impulso.

2. El método de la reivindicación 1, en el que el vector base comprende un intervalo de muestreo, un retraso de propagación desde la fuente de audio a un centro de la disposición de micrófonos, y un micrófono individual desde la disposición de micrófonos, y un ángulo de llegada desde las señales de audio recibidas.

3. El método de la reivindicación 1, que comprende utilizar respuesta de impulso de sala (100) calculada para iniciar una cancelación acústico.

4. El método de la reivindicación 1, en el que la etapa de crear una aproximación de la señal de audio recibida como una combinación ponderada del conjunto de señales de fuente única comprende utilizar un solucionador disperso para determinar un peso que satisfaga un error de reconstrucción, y en donde el peso es optimizado utilizando técnicas de optimización convexas, y en donde el error de reconstrucción corresponde al error de modelización más el ruido.

5. El método de la reivindicación 1, en el que la respuesta de impulso es calculada utilizando técnicas que se basan en el algoritmo de transformación de Fourier.

6. El método de la reivindicación 1, en el que las señales de audio recibidas son representadas por los vectores base; en donde los vectores base están configurados para representar señales de imagen reales o virtuales que serían recibidas por la configuración de micrófono (106) desde una fuente no atenuada.

7. Un sistema para implementar la cancelación de eco acústico de retroalimentación audible experimentada en una disposición de una fuente de audio (102, 302) y una disposición de micrófonos (106), que comprende:

un procesador para ejecutar un código ejecutable por procesador;
una disposición de micrófonos (106); y
un dispositivo de almacenamiento que almacena el código ejecutable por procesador, en donde el código ejecutable por procesador, cuando es ejecutado por el procesador, hace que el procesador:

reciba señales de audio en cada micrófono (304) de una disposición de micrófonos (106), estando las señales basadas en una señal de calibración;
calcule la geometría de la disposición de micrófonos;
defina un conjunto de señales de fuente única ya que la señales de fuente única incidirían sobre todos los micrófonos (304) de la disposición de micrófonos (106) al mismo tiempo, correspondiendo cada una de las señales de fuente única a una ubicación diferente de una única fuente real o virtual, en donde la fuente virtual representa una reflexión;
encuentre una aproximación de la señal de audio recibida como una combinación ponderada del conjunto de señales de fuente única;
calcule una respuesta de impulso de sala (100) a partir de la fuente de audio (102, 302) para cada uno de los micrófonos de la disposición de micrófonos (106),

correspondiendo la respuesta de impulso de sala a la combinación ponderada y estando la respuesta de impulso de sala basada al menos en parte en la geometría de disposición de micrófonos, y un vector base que comprende una señal enviada por un altavoz para múltiples ángulos de audición y retrasos de tiempo que corresponden a muchas fuentes virtuales; y
cancele el eco acústico procedente de la señal recibida utilizando la respuesta de impulso.

8. El sistema de la reivindicación 7, en el que el procesador representa las señales de audio recibidas por los vectores base; en donde los vectores base están configurados para representar señales de imagen reales o virtuales que serían recibidas por la disposición de micrófonos (106) procedentes de una fuente no atenuada.

9. El sistema de la reivindicación 7, en el que la combinación ponderada del conjunto de señales de fuente única es optimizada utilizando técnicas de optimización convexas.

10. El sistema de la reivindicación 7, en el que la respuesta de impulso es calculada utilizando técnicas que se basan en el algoritmo rápido de transformación de Fourier.

11. El sistema de la reivindicación 7, en el que la estructura de la señal recibida es modelizada por el siguiente vector base:

$$b_{T,\theta} = \begin{bmatrix} l(0 - T - \tau(z_1, \theta)) \\ \vdots \\ l((L_x - 1)T_s - T - \tau(z_1, \theta)) \\ \vdots \\ l(0 - T - \tau(z_K, \theta)) \\ \vdots \\ l((L_x - 1)T_s - T - \tau(z_K, \theta)) \end{bmatrix}.$$

en donde $l(t)$ es una función de la versión de tiempo continua de la señal de fuente de audio; L_x designa la longitud de la señal de calibración recibida en cada micrófono; T_s es el intervalo de muestreo; T es el retraso de propagación desde la fuente al centro de la disposición de micrófonos; T es una función de z_k y θ ; en donde z_k es un micrófono individual de la disposición de micrófonos; y θ es el ángulo de llegada de la señal recibida.

12. El método de la reivindicación 1, en el que la estructura de la señal de audio recibida está modelizada por el siguiente vector base:

$$b_{T,\theta} = \begin{bmatrix} l(0 - T - \tau(z_1, \theta)) \\ \vdots \\ l((L_x - 1)T_s - T - \tau(z_1, \theta)) \\ \vdots \\ l(0 - T - \tau(z_K, \theta)) \\ \vdots \\ l((L_x - 1)T_s - T - \tau(z_K, \theta)) \end{bmatrix}.$$

en donde $l(t)$ es una función de la versión de tiempo continua de la señal de fuente de audio; L_x designa la longitud de la señal de calibración recibida en cada micrófono; T_s es el intervalo de muestreo; T es el retraso de propagación desde la fuente al centro de la disposición de micrófonos; T es una función de z_k y θ ; en donde z_k es un micrófono individual de la disposición de micrófonos; y θ es el ángulo de llegada de la señal recibida.

13. Un programa de ordenador que comprende instrucciones que cuando se ejecutan en un ordenador implementan el método de cualquiera de las reivindicaciones 1 a 6 ó 12.

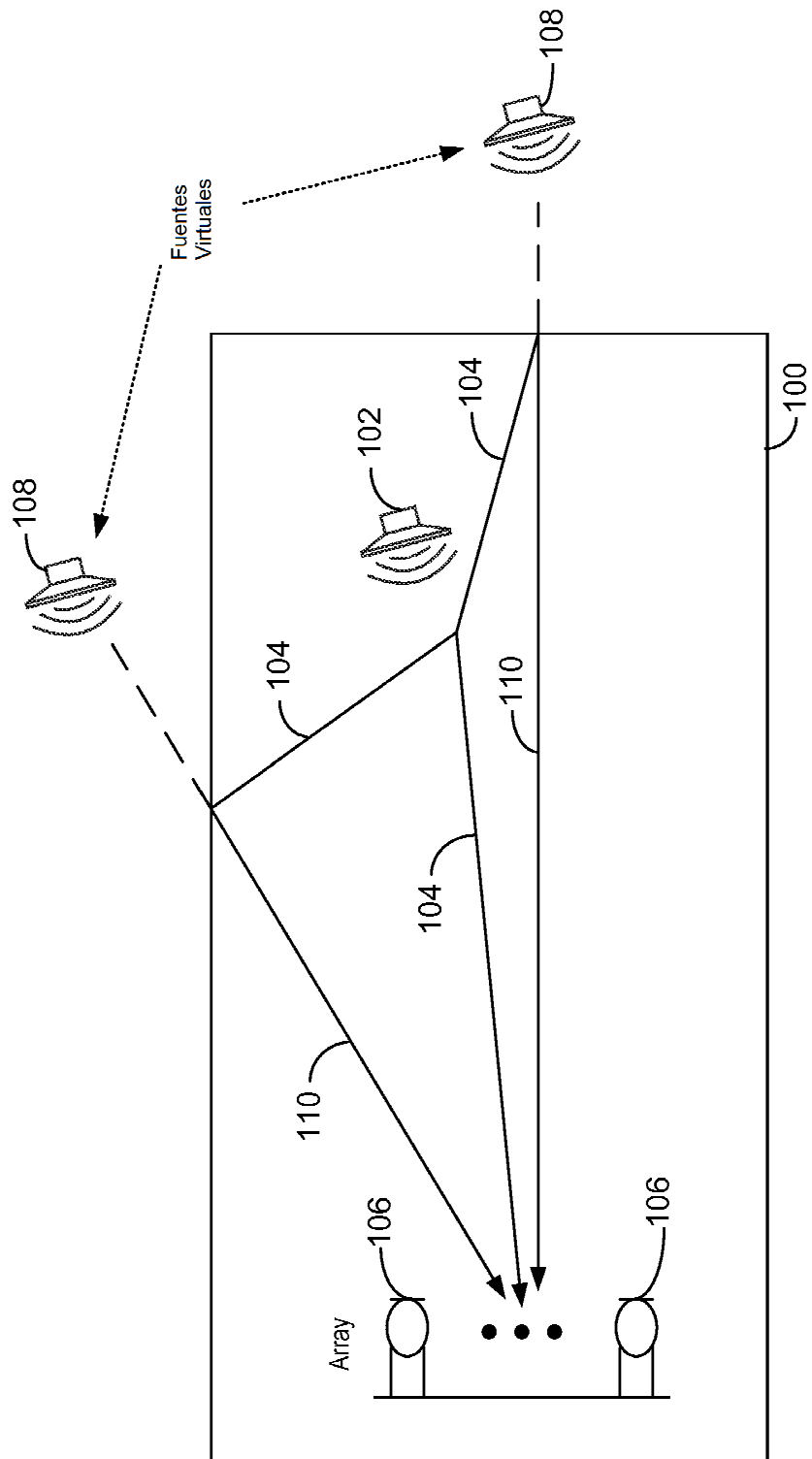
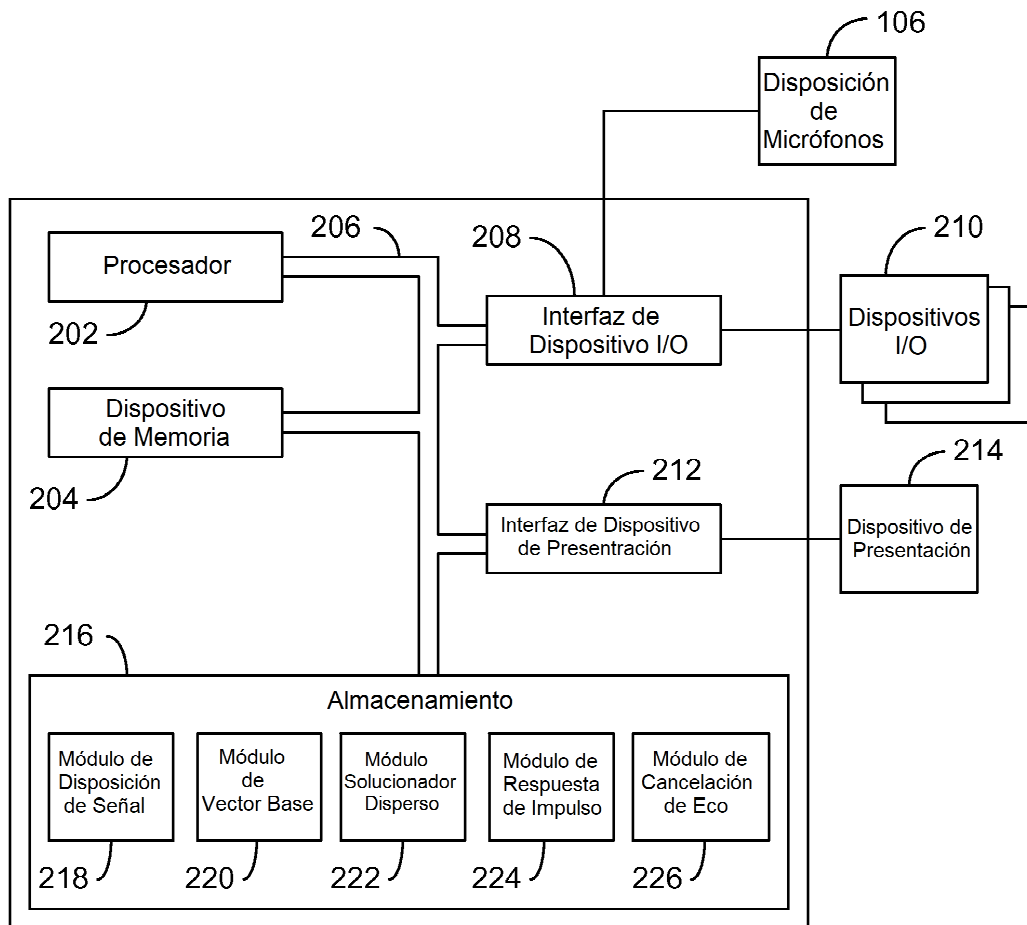
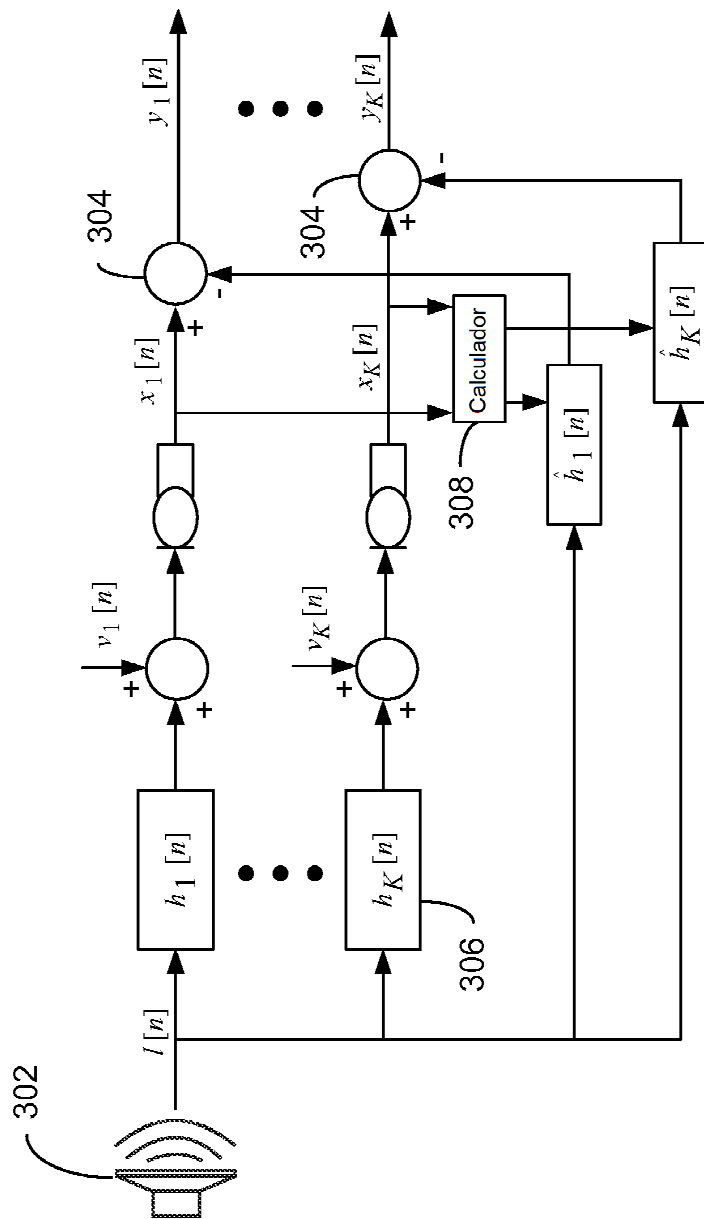


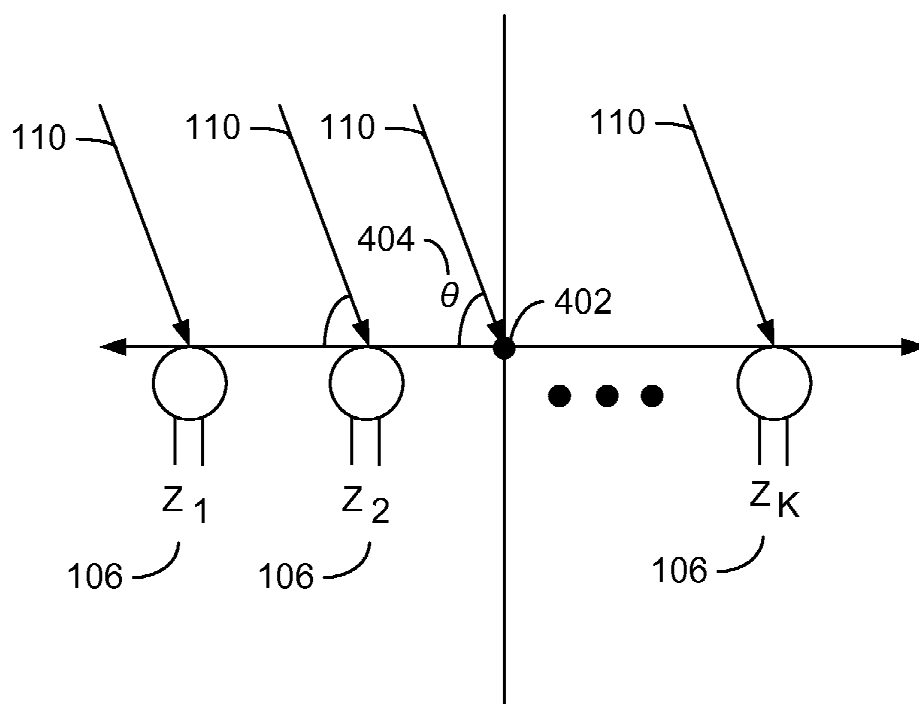
FIG. 1



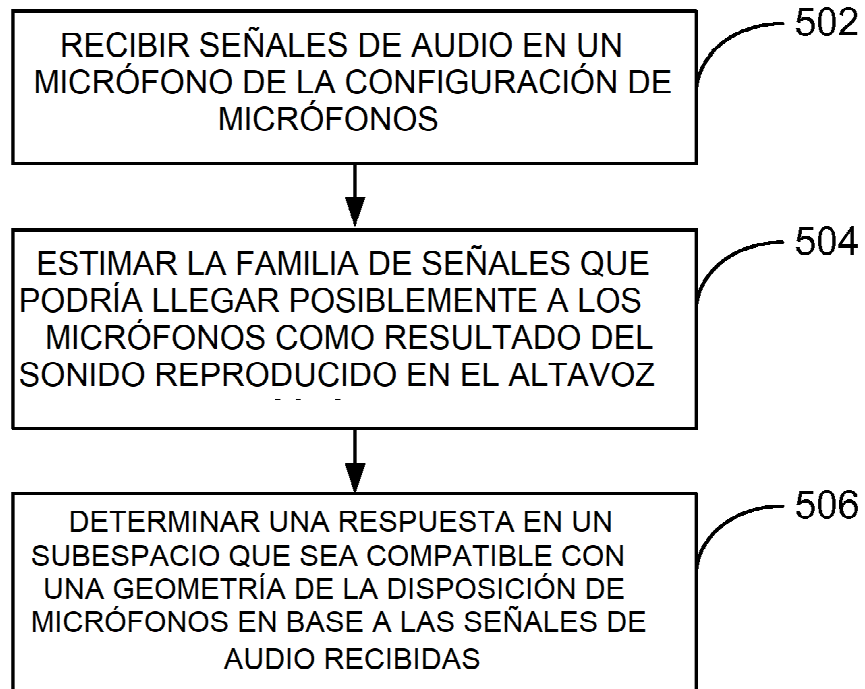
200
FIG. 2



300
FIG. 3

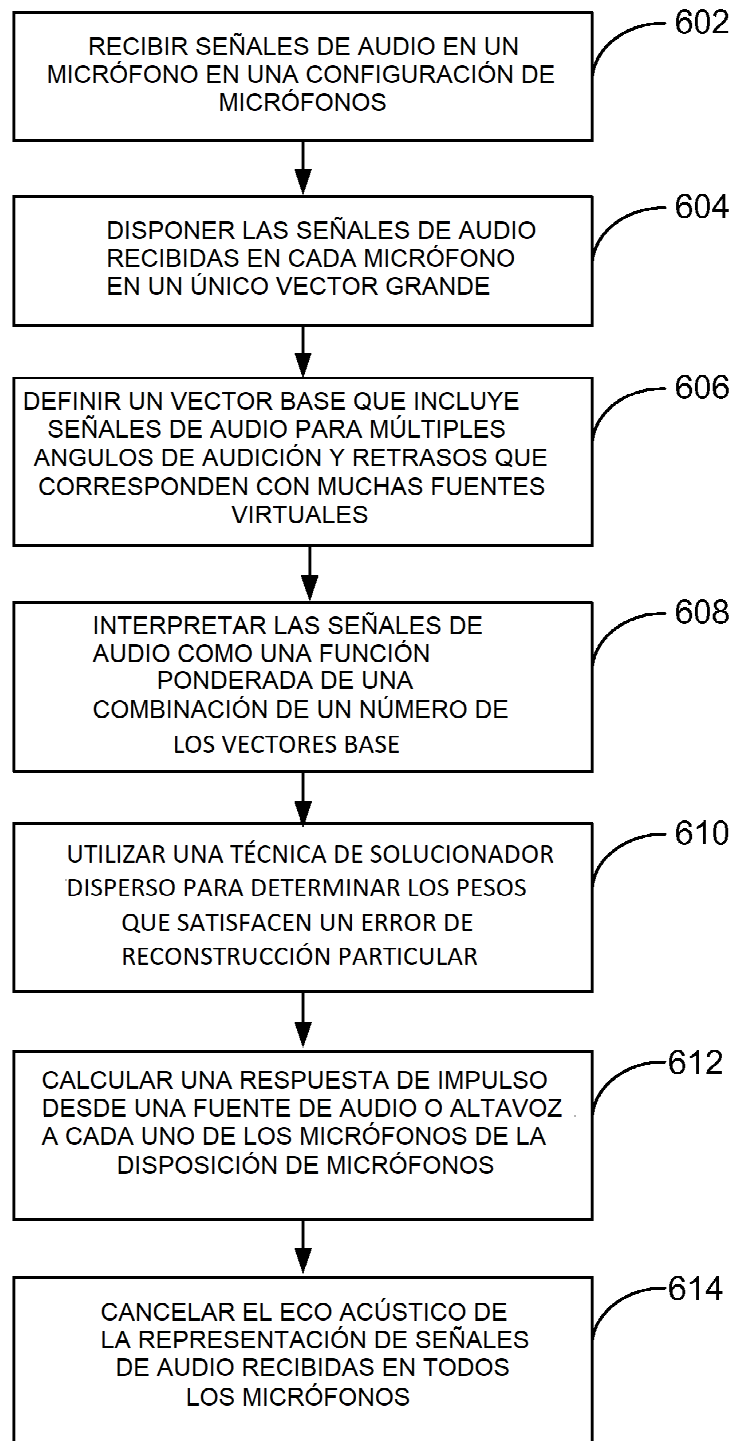


400
FIG. 4



500

FIG. 5



600

FIG. 6

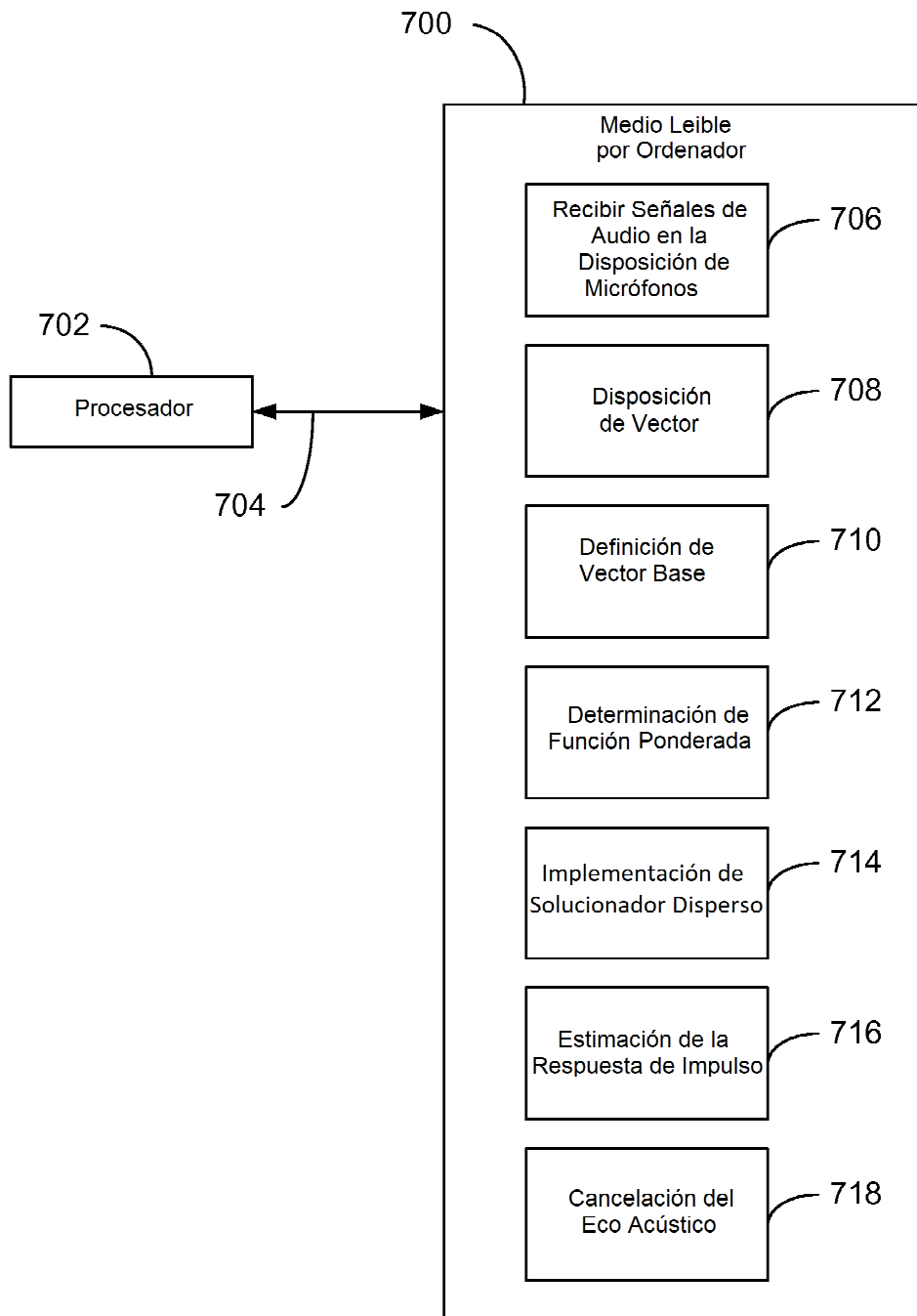


FIG. 7