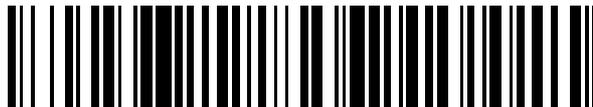


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 709 523**

51 Int. Cl.:

**H04S 7/00** (2006.01)

**G10L 21/0364** (2013.01)

**G10L 21/0208** (2013.01)

**G10L 21/0232** (2013.01)

**G10L 21/034** (2013.01)

**H04S 3/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.02.2011 PCT/US2011/026505**

87 Fecha y número de publicación internacional: **15.09.2011 WO11112382**

96 Fecha de presentación y número de la solicitud europea: **28.02.2011 E 11707537 (4)**

97 Fecha y número de publicación de la concesión europea: **12.12.2018 EP 2545552**

54 Título: **Procedimiento y sistema de escalado de atenuación de canales relevantes de voz en audio multicanal**

30 Prioridad:

**08.03.2010 US 311437 P**

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

**16.04.2019**

73 Titular/es:

**DOLBY LABORATORIES LICENSING CORPORATION (100.0%)  
1275 Market Street  
San Francisco, CA 94103, US**

72 Inventor/es:

**MUESCH, HANNES**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 709 523 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

**DESCRIPCIÓN**

Procedimiento y sistema de escalado de atenuación de canales relevantes de voz en audio multicanal

**Referencia cruzada a aplicaciones relacionadas**

5 La presente solicitud reivindica prioridad a la solicitud provisional de patente de los Estados Unidos N° 61/311.437, presentada el 8 de Marzo de 2010.

**Antecedentes de la invención****1. Campo de la invención**

10 La invención se refiere a sistemas y a procedimientos para mejorar la inteligibilidad de la voz humana (por ejemplo, un diálogo) determinada por una señal de audio multicanal. En algunas realizaciones, la invención es un procedimiento y un sistema para filtrar una señal de audio que tiene un canal de voz ("speech channel") y un canal sin voz ("non-speech channel") para mejorar la inteligibilidad de la voz determinada por la señal, mediante la determinación de al menos un valor de control de atenuación indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por el canal sin voz, y la atenuación del canal sin voz en respuesta al valor de control de atenuación.

**15 2. Antecedentes de la invención**

A lo largo de la presente descripción, incluyendo las reivindicaciones, el término "voz" se usa en un sentido amplio para indicar la voz humana. Por lo tanto, la "voz" determinada por una señal de audio es un contenido de audio de la señal que es percibida como voz humana (por ejemplo, diálogo, monólogo, canto u otra voz humana) durante la reproducción de la señal por un altavoz (u otro transductor emisor de sonidos). Según las realizaciones típicas de la invención, la audibilidad de la voz determinada por una señal de audio es mejorada con relación a otro contenido de audio (por ejemplo, música instrumental o efectos de sonido sin voz) determinado por la señal, mejorando de esta manera la inteligibilidad (por ejemplo, claridad o facilidad de comprensión) de la voz.

20 A lo largo de la presente descripción, incluyendo las reivindicaciones, la expresión "contenido mejorador de voz" de un canal de una señal de audio multicanal es un contenido (determinado por el canal) que mejora la inteligibilidad u otra cualidad percibida del contenido de voz determinado por otro canal (por ejemplo, un canal de voz) de la señal.

Las realizaciones típicas de la invención suponen que la mayor parte de la voz determinada por una señal de audio de entrada multicanal está determinada por el canal central de la señal. Esta suposición es consistente con la convención en la producción de sonido envolvente según la cual la mayor parte de la voz es colocada normalmente en un solo canal (el canal central), y la mayor parte de la música, el sonido ambiental y los efectos de sonido se mezclan normalmente en todos los canales (por ejemplo, los canales Izquierdo, Derecho, Envolvente Izquierdo y Envolvente Derecho, así como el canal central).

30 De esta manera, en la presente memoria, a veces se hará referencia al canal central de una señal de audio multicanal como el canal de "voz" y, en la presente memoria, a veces se hará referencia al resto de canales (por ejemplo, Izquierdo, Derecho, Envolvente Izquierdo y Envolvente Derecho) de la señal como canales "sin voz". De manera similar, en la presente memoria, a veces se hará referencia a un canal "central" generado sumando los canales izquierdo y derecho de una señal estéreo cuya voz está paneada centralmente como canal de "voz", y, en la presente memoria, a veces se hará referencia a un canal "lateral" generado restando dicho canal central del canal izquierdo (o derecho) de la señal estéreo como un canal "sin voz".

40 A lo largo de la presente descripción, incluyendo las reivindicaciones, la expresión realizar una operación "sobre" señales o datos (por ejemplo, filtrar, escalar o transformar las señales o los datos) se usa en un sentido amplio para indicar la realización de la operación directamente sobre las señales o los datos, o sobre versiones procesadas de las señales o los datos (por ejemplo, sobre versiones de las señales que se han sido sometidas a un filtrado preliminar antes de la realización de la operación sobre las mismas).

45 A lo largo de la presente descripción, incluyendo las reivindicaciones, la expresión "sistema" se usa en un sentido amplio para indicar un dispositivo, sistema o subsistema. Por ejemplo, un subsistema que implementa un decodificador puede denominarse un sistema decodificador, y un sistema que incluye dicho subsistema (por ejemplo, un sistema que genera X señales de salida en respuesta a múltiples entradas, en el que el subsistema genera M de las entradas y las otras X-M entradas son recibidas desde una fuente externa) puede denominarse también sistema decodificador.

50 A lo largo de la descripción, incluyendo las reivindicaciones, la expresión "relación" de un primer valor ("A") a un segundo valor ("B") se usa en un sentido amplio para indicar A/B, o B/A, o relación de una versión escalada o desplazada de uno

de entre A y B a una versión escalada o desplazada del otro de entre A y B (por ejemplo,  $(A + x)/(B + y)$ , donde x e y son valores de desplazamiento).

5 A lo largo de la descripción, incluyendo las reivindicaciones, la expresión "reproducción" de señales por transductores emisores de sonido (por ejemplo, altavoces) hace referencia a causar que los transductores produzcan sonido en respuesta a las señales, incluyendo la realización de cualquier amplificación requerida y/u otro procesamiento de las señales.

10 Cuando se escucha una voz en presencia de sonidos competitivos (tal como cuando se escucha a un amigo sobre el ruido de la gente en un restaurante), una parte de las características acústicas que señalan el contenido fonológico de la voz (características o señales de voz) están enmascaradas por los sonidos competitivos y ya no están disponibles para el oyente para decodificar el mensaje. A medida que el nivel del sonido competitivo aumenta con relación al nivel de la voz, el número de características de voz que se reciben correctamente disminuye y la percepción de la voz se hace cada vez más incómoda hasta que, en algún nivel de sonido competitivo, el proceso de percepción de la voz se deteriora. Aunque esta relación es válida para todos los oyentes, el nivel de sonido competitivo tolerable para cualquier nivel de voz no es el mismo para todos los oyentes. Algunos oyentes, por ejemplo, aquellos con pérdida de audición debido a la edad (presbiacusis) o aquellos que escuchan un lenguaje que adquirieron después de la pubertad, toleran menos los sonidos competitivos que los oyentes con buena audición o aquellos que escuchan su idioma nativo.

15 El hecho de que los oyentes tengan diferentes capacidades para comprender la voz en presencia de sonidos competitivos tiene implicaciones para el nivel en el que los sonidos ambientales y la música de fondo en las noticias o en un audio de entretenimiento se mezclan con la voz. Los oyentes con pérdida auditiva o aquellos que escuchan un idioma extranjero frecuentemente prefieren un nivel relativo más bajo de audio sin voz que el proporcionado por el creador del contenido.

20 Para satisfacer estas necesidades especiales, es conocida la aplicación de una atenuación ("ducking") a los canales sin voz de una señal de audio multicanal, pero menos (o ninguna) atenuación al canal de voz de la señal, para mejorar la inteligibilidad de la voz determinada por la señal.

25 Por ejemplo, la publicación de solicitud internacional PCT número WO 2010/011377, que nombra a Hannes Muesch como inventor y asignada a Dolby Laboratories Licensing Corporation (publicada el 28 de Enero de 2010), describe que los canales sin voz (por ejemplo, los canales izquierdo y derecho) de una señal de audio multicanal pueden enmascarar la voz en el canal de voz de la señal (por ejemplo, el canal central) hasta el punto de que se satisface un nivel deseado de inteligibilidad de la voz. El documento WO 2010/011377 describe cómo determinar una función de atenuación a ser aplicada por una circuitería de atenuación a los canales sin voz en un intento de desenmascarar la voz en el canal de voz mientras se conserva la mayor parte posible de la intención del creador del contenido. La técnica descrita en el documento WO 2010/011377 se basa en la suposición de que el contenido en un canal sin voz nunca mejora la inteligibilidad (u otra cualidad percibida) del contenido de voz determinado por el canal de voz.

30 La presente invención se basa, en parte, en el reconocimiento de que, aunque esta suposición es correcta para la gran mayoría del contenido de audio multicanal, no siempre es válida. El presente inventor ha reconocido que cuando al menos un canal sin voz de una señal de audio multicanal incluye contenido que mejora la inteligibilidad (u otra cualidad percibida) del contenido de voz determinado por el canal de voz de la señal, el filtrado de la señal según el procedimiento del documento WO 2010/011377 puede afectar negativamente a la experiencia de entretenimiento de una persona que escucha la señal filtrada reproducida. Según las realizaciones típicas de la presente invención, la aplicación del procedimiento descrito en el documento WO 2010/011377 es suspendida o modificada durante los momentos en los que el contenido no se ajusta a las suposiciones subyacentes al procedimiento del documento WO 2010/011377.

35 Existe una necesidad de un procedimiento y de un sistema para filtrar una señal de audio multicanal para mejorar la inteligibilidad de la voz en el caso común en el que al menos un canal sin voz de la señal de audio incluye contenido que mejora la inteligibilidad del contenido de voz en el canal de voz de la señal de audio.

#### **Breve descripción de la invención**

45 En una primera clase de realizaciones, la invención es un procedimiento para filtrar una señal de audio multicanal que tiene un canal de voz y al menos un canal sin voz, para mejorar la inteligibilidad de la voz determinada por la señal. El procedimiento incluye las etapas de: (a) determinar al menos un valor de control de atenuación indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por al menos un canal sin voz de la señal de audio multicanal; y (b) atenuar al menos un canal sin voz de la señal de audio multicanal en respuesta al por lo menos un valor de control de atenuación. Típicamente, la etapa de atenuación comprende escalar una señal de control de atenuación no procesada (por ejemplo, una señal de control de ganancia de atenuación) para el canal sin voz en respuesta al por lo menos un valor de control de atenuación. Preferiblemente, el canal sin voz es atenuado para mejorar la inteligibilidad de la voz determinada por el canal de voz sin atenuar de manera no deseada el contenido mejorador de la voz determinado por el canal sin voz. En algunas

realizaciones, cada valor de control de atenuación determinado en la etapa (a) es indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por un canal sin voz de la señal de audio, y la etapa (b) incluye la etapa de atenuar este canal sin voz en respuesta a dicho valor de control de atenuación. En algunas otras realizaciones, la etapa (a) incluye una etapa de derivar un canal sin voz derivado a partir de al menos un canal sin voz de la señal de audio, y el al menos un valor de control de atenuación es indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por el canal sin voz derivado. Por ejemplo, el canal sin voz derivado puede ser generado sumando o sino mezclando o combinando al menos dos canales sin voz de la señal de audio. La determinación de cada valor de control de atenuación a partir de un solo canal sin voz derivado puede reducir el coste y la complejidad de implementar algunas realizaciones de la invención, con relación al coste y a la complejidad de determinar diferentes subconjuntos de un conjunto de valores de atenuación a partir de canales sin voz diferentes. En las realizaciones en las que la señal de audio de entrada tiene al menos dos canales sin voz, la etapa (b) puede incluir la etapa de atenuar un subconjunto de los canales sin voz (por ejemplo, cada canal sin voz a partir del cual se ha derivado un canal sin voz) o todos los canales sin voz, en respuesta al por lo menos un valor de control de atenuación (por ejemplo, en respuesta a una secuencia única de valores de control de atenuación).

En algunas realizaciones en la primera clase, la etapa (a) incluye una etapa de generar una señal de control de atenuación indicativa de una secuencia de valores de control de atenuación, en la que cada uno de los valores de control de atenuación es indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por el al menos un canal sin voz en un tiempo diferente (por ejemplo, en un intervalo de tiempo diferente), y la etapa (b) incluye las etapas de: escalar una señal de control de ganancia de atenuación en respuesta a la señal de control de atenuación para generar una señal de control de ganancia escalada, y aplicar la señal de control de ganancia escalada para atenuar el al menos un canal sin voz (por ejemplo, activar, habilitar o emitir ("assert") la señal de control de ganancia escalada a la circuitería de atenuación para controlar la atenuación del al menos un canal sin voz por la circuitería de atenuación). Por ejemplo, en algunas de dichas realizaciones, la etapa (a) incluye una etapa de comparar una primera secuencia de características relacionadas con la voz (indicativas del contenido relacionado con la voz determinado por el canal de voz) con una segunda secuencia de características relacionadas con la voz (indicativas del contenido relacionado con la voz determinado por el al menos un canal sin voz) para generar la señal de control de atenuación, y cada uno de los valores de control de atenuación indicados por la señal de control de atenuación es indicativo de una medida de similitud entre la primera secuencia de características relacionadas con la voz y la segunda secuencia de características relacionadas con la voz en un tiempo diferente (por ejemplo, en un intervalo de tiempo diferente). En algunas realizaciones, cada valor de control de atenuación es un valor de control de ganancia.

En algunas realizaciones en la primera clase, cada valor de control de atenuación está relacionado monotónicamente con la probabilidad de que al menos un canal sin voz de la señal de audio sea indicativo de contenido mejorador de voz que mejora la inteligibilidad (u otra cualidad percibida) del contenido de voz determinado por el canal de voz.

En una segunda clase de realizaciones, la invención es un procedimiento para filtrar una señal de audio multicanal que tiene un canal de voz y al menos un canal sin voz, para mejorar la inteligibilidad de la voz determinada por la señal. El procedimiento incluye las etapas de: (a) comparar una característica del canal de voz y una característica del canal sin voz para generar al menos un valor de atenuación para controlar la atenuación del canal sin voz con relación al canal de voz; y (b) ajustar el al menos un valor de atenuación en respuesta al por lo menos un valor de probabilidad de mejora de voz para generar al menos un valor de atenuación ajustado para controlar la atenuación del canal sin voz con relación al canal de voz. Típicamente, la etapa de ajuste es (o incluye) escalar cada valor de atenuación en respuesta a uno de dichos valores de probabilidad de mejora de voz para generar uno de dichos valores de atenuación ajustada. Típicamente, cada valor de probabilidad de mejora de voz es indicativo (por ejemplo, está relacionado monotónicamente con) la probabilidad de que el canal sin voz (o un canal sin voz derivado a partir del canal sin voz o a partir de un conjunto de canales sin voz de la señal de audio de entrada) sea indicativo de contenido mejorador de voz (contenido que mejora la inteligibilidad u otra cualidad percibida del contenido de voz determinado por el canal de voz).

En algunas realizaciones en la segunda clase, el al menos un valor de probabilidad de mejora de voz es una secuencia de valores de comparación (por ejemplo, valores de diferencia) determinados por un procedimiento que incluye una etapa de comparar una primera secuencia de características relacionadas con la voz indicativas del contenido relacionado con la voz determinado por el canal de voz con una segunda secuencia de características relacionadas con la voz indicativas del contenido relacionado con la voz determinado por el canal sin voz, y cada uno de los valores de comparación es una medida de similitud entre la primera secuencia de características relacionadas con la voz y la segunda secuencia de características relacionadas con la voz en un tiempo diferente (por ejemplo, en un intervalo de tiempo diferente). En realizaciones típicas en la tercera clase, el procedimiento incluye también la etapa de atenuar el canal sin voz en respuesta al por lo menos un valor de atenuación ajustado. La etapa (b) puede comprender escalar el al menos un valor de atenuación (que típicamente es, o está determinado por, una señal de control de ganancia de atenuación u otra señal de control de atenuación no procesada) en respuesta al por lo menos un valor de probabilidad de mejora de voz.

En algunas realizaciones en la segunda clase, cada valor de atenuación generado en la etapa (a) es un primer factor indicativo de una cantidad de atenuación del canal sin voz necesario para limitar la relación de potencia de señal en el canal sin voz a la potencia de señal en el canal de voz de manera que no exceda un umbral predeterminado, escalado por un segundo factor relacionado monotónicamente con la probabilidad de que el canal de voz sea indicativo de voz.

- 5 En algunas realizaciones en la segunda clase, la etapa (a) incluye las etapas de generar cada uno de dichos valores de atenuación, incluyendo mediante la determinación de un espectro de potencia (indicativo de la potencia como una función de la frecuencia) de cada uno de los canales de voz y el canal sin voz, y realizar una determinación en el dominio de la frecuencia del valor de atenuación en respuesta a dicho espectro de potencia. Preferiblemente, los valores de atenuación generados de esta manera determinan la atenuación como una función de la frecuencia a aplicar a los componentes de frecuencia del canal sin voz.

Los aspectos de la invención incluyen un sistema configurado (por ejemplo, programado) para realizar cualquier realización del procedimiento de la invención, y un medio legible por ordenador (por ejemplo, un disco) que almacena código para implementar cualquier realización del procedimiento de la invención.

### Breve descripción de los dibujos

- 15 La Fig. 1 es un diagrama de bloques de una realización del sistema de la invención.  
La Fig. 1A es un diagrama de bloques de otra realización del sistema de la invención.  
La Fig. 2 es un diagrama de bloques de otra realización del sistema de la invención. La Fig. 2A es un diagrama de bloques de otra realización del sistema de la invención. La Fig. 3 es un diagrama de bloques de otra realización del sistema de la invención.
- 20 La Fig. 4 es un diagrama de bloques de un procesador de señal digital (DSP) de audio que es una realización del sistema de la invención.  
La Fig. 5 es un diagrama de bloques de un sistema de ordenador, que incluye un medio 504 de almacenamiento legible por ordenador que almacena código informático para programar el sistema para realizar una realización del procedimiento de la invención.

### 25 Descripción detallada de las realizaciones preferidas

Muchas realizaciones de la presente invención son tecnológicamente posibles. A partir de la presente descripción, la implementación de las realizaciones será evidente para las personas con conocimientos ordinarios en la materia. Las realizaciones del sistema, del procedimiento y del medio de la invención se describirán con referencia a las Figs. 1, 1A, 2, 2A y 3-5, y se definen en las reivindicaciones adjuntas.

- 30 El presente inventor ha observado que algunos contenidos de audio multicanal tienen contenido de voz diferente, pero aún así relacionado, en el canal de voz y al menos en un canal sin voz. Por ejemplo, las grabaciones de audio multicanal de algunos espectáculos teatrales se mezclan de manera que la voz "seca" (es decir, la voz sin reverberación notable) se coloca en el canal de voz (típicamente, el canal central, C, de la señal) y la misma voz, pero con un componente de reverberación significativo (voz "húmeda") se coloca en los canales sin voz de la señal. En un escenario típico, la voz seca es la señal desde el micrófono que el artista sostiene cerca de su boca y la voz húmeda es la señal desde los micrófonos colocados en la audiencia. La voz húmeda está relacionada con la voz seca, ya que es la interpretación tal como es escuchada por el público en el sitio. Sin embargo, es diferente de la voz seca. Típicamente, la voz húmeda está retrasada con relación a la voz seca, y tiene un espectro diferente y componentes aditivos diferentes (por ejemplo, ruidos del público y reverberación).
- 40 Dependiendo de los niveles relativos de la voz seca y la voz húmeda, es posible que la componente de voz húmeda enmascare la componente de voz seca en un grado tal que la atenuación de los canales sin voz en la circuitería de atenuación (por ejemplo, como en el procedimiento descrito en el documento WO 2010/011377 indicado anteriormente) atenúa de manera no deseable la señal de voz húmeda. Aunque las componentes de voz seca y húmeda pueden describirse como entidades separadas, un oyente fusiona perceptualmente las dos y las escucha como un solo flujo de voz. La atenuación de la componente de voz húmeda (por ejemplo, en la circuitería de atenuación) puede tener el efecto de reducir la intensidad percibida del flujo de voz fusionado junto con el colapso de su anchura de imagen. El presente inventor ha reconocido que para las señales de audio multicanal que tienen componentes de voz húmeda y seca del tipo indicado, frecuentemente sería perceptualmente más agradable, así como más favorable para la inteligibilidad de la voz, si el nivel de las componentes de voz húmeda no se alterara durante el procesamiento de mejora de voz de las señales.
- 45 La invención se basa, en parte, en el reconocimiento de que, cuando al menos un canal sin voz de una señal de audio multicanal incluye contenido que mejora la inteligibilidad (u otra cualidad percibida) del contenido de voz determinado por
- 50

el canal de voz de la señal, el filtrado de los canales sin voz de la señal usando circuitería de atenuación (por ejemplo, según el procedimiento del documento WO 2010/011377) puede afectar negativamente a la experiencia de entretenimiento de una persona que escucha la señal filtrada reproducida. Según las realizaciones típicas de la invención, la atenuación (en circuitería de atenuación) de al menos un canal sin voz de una señal de audio multicanal es suspendida o modificada durante los tiempos en los que el canal sin voz incluye contenido mejorador de voz (contenido que mejora la inteligibilidad u otra cualidad percibida del contenido de voz determinada por el canal de voz de la señal). En los tiempos en los que el canal sin voz no incluye contenido mejorador de voz (o no incluye contenido mejorador de voz que cumple con un criterio predeterminado), el canal sin voz es atenuado normalmente (la atenuación no se suspende ni se modifica).

Una señal multicanal típica (que tiene un canal de voz) para la cual un filtrado convencional en circuitería de atenuación es inapropiado es aquella que incluye al menos un canal sin voz que transporta características de voz que son sustancialmente idénticas a las características de voz en el canal de voz. Según las realizaciones típicas de la presente invención, una secuencia de características relacionadas con la voz en el canal de voz es comparada con una secuencia de características relacionadas con la voz en el canal sin voz. Una similitud sustancial de las dos secuencias de características indica que el canal sin voz (es decir, la señal en el canal sin voz) contribuye a la información útil para comprender la voz en el canal de voz y que debería evitarse la atenuación del canal sin voz.

Para apreciar la importancia del examen de la similitud entre dichas secuencias de características relacionadas con la voz en lugar de las propias señales, es importante reconocer que el contenido de voz "seca" y "húmeda" (determinado por los canales de voz y sin voz) no es idéntico; las señales indicativas de los dos tipos de contenido de voz están típicamente desfasadas temporalmente, y han sido sometidas a procesos de filtrado diferentes y se les han añadido diferentes componentes extraños. Por lo tanto, una comparación directa entre las dos señales producirá una baja similitud, independientemente de si el canal sin voz contribuye o no a características de voz que son las mismas que en el canal de voz (como en el caso de la voz seca y húmeda), características de voz no relacionadas (como en el caso de dos voces no relacionadas en el canal de voz y sin voz [por ejemplo, una conversación objetivo en el canal de voz y murmullo de fondo en el canal sin voz]), o ninguna señal de voz en absoluto (por ejemplo, el canal sin voz transporta música y efectos). Al basar la comparación en las características de la voz (como en las realizaciones preferidas de la presente invención), se consigue un nivel de abstracción que disminuye el impacto de los aspectos de señales irrelevantes, tales como pequeñas cantidades de retardo, diferencias espectrales y señales extrañas agregadas. De esta manera, las implementaciones preferidas de la invención generan típicamente al menos dos flujos de características de voz: una que representa la señal en el canal de voz; y al menos uno que representa la señal de un canal sin voz.

Un primer ejemplo (125) de un sistema que implementa los procedimientos reivindicados se describirá con referencia a la Fig. 1. En respuesta a una señal de audio multicanal que comprende un canal 101 de voz (canal C central) y dos canales 102 y 103 sin voz (canales L y R izquierdo y derecho), el sistema de la Fig. 1 filtra los canales sin voz para generar una señal de audio de salida multicanal filtrada que comprende el canal 101 de voz y los canales 118 y 119 sin voz filtrados (canales L' y R' izquierdo y derecho filtrados). De manera alternativa, uno o ambos canales 102 y 103 sin voz pueden ser otro tipo de canal sin voz de una señal de audio multicanal (por ejemplo, canales trasero izquierdo y/o trasero derecho de una señal de audio de 5.1 canales) o puede ser un canal in voz derivado, derivado a partir de (por ejemplo, es una combinación de) cualquiera de los muchos subconjuntos diferentes de canales sin voz de una señal de audio multicanal. De manera alternativa, el sistema puede ser implementado para filtrar solo un canal sin voz, o más de dos canales sin voz, de una señal de audio multicanal.

Con referencia una vez más a la Fig. 1, los canales 102 y 103 sin voz son emitidos o activados ("asserted") a los amplificadores 117 y 116 de atenuación, respectivamente. Durante el funcionamiento, el amplificador 116 de atenuación es accionado por una señal S3 de control (que es indicativa de una secuencia de valores de control y, de esta manera, se denomina también secuencia S3 de valores de control) emitida desde el elemento 114 de multiplicación, y el amplificador 117 de atenuación es accionado por la señal S4 de control (que es indicativa de una secuencia de valores de control, y de esta manera se denomina también secuencia S4 de valores de control) emitida desde el elemento 115 de multiplicación.

La potencia de cada canal de la señal de entrada multicanal es medida con un banco de estimadores de potencia (104, 105 y 106) y es expresada en una escala logarítmica [dB]. Estos estimadores de potencia pueden implementar un mecanismo de suavizado, como un integrador con fugas, de manera que el nivel de potencia medido refleje el nivel de potencia promediado durante la duración de una oración o un pasaje completo. El nivel de potencia de la señal en el canal de voz se resta del nivel de potencia en cada uno de los canales sin voz (restando los elementos 107 y 108) para dar una medida de la relación de potencia entre los dos tipos de señal. La salida del elemento 107 es una medida de la relación de potencia en el canal 103 sin voz a la potencia en el canal 101 de voz. La salida del elemento 108 es una medida de la relación de potencia en el canal 102 sin voz a la potencia en el canal 101 de voz.

El circuito 109 de comparación determina, para cada canal sin voz, el número de decibelios (dB) en los que el canal sin voz debe ser atenuado para que su nivel de potencia permanezca al menos  $\vartheta$  dB por debajo del nivel de potencia de la señal en el canal de voz (donde el símbolo  $\vartheta$  conocido también como "script theta", indica un valor de umbral

predeterminado). En una implementación del circuito 109, el elemento 120 de adición suma el valor  $\vartheta$  de umbral (almacenado en el elemento 110, que puede ser un registro) a la diferencia de nivel de potencia (o "margen") entre el canal 103 sin voz y el canal 101 de voz, y el elemento 121 de adición añade el valor  $\vartheta$  de umbral a la diferencia de nivel de potencia entre el canal 102 sin voz y el canal 101 de voz. Los elementos 111-1 y 112-1 cambian el signo de la salida de los elementos 120 y 121 de adición, respectivamente. Esta operación de cambio de signo convierte los valores de atenuación en valores de ganancia. Los elementos 111 y 112 limitan cada resultado de manera que sea igual o menor que cero (la salida del elemento 111-1 es emitida al limitador 111 y la salida del elemento 112-1 es emitida al limitador 112). El valor C1 actual emitido desde el limitador 111 determina la ganancia (atenuación negada) en dB que debe ser aplicada al canal 103 sin voz para mantener su nivel de potencia  $\vartheta$  dB por debajo del nivel de potencia del canal 101 de voz (en el tiempo relevante o en la ventana temporal relevante, de la señal de entrada multicanal). El valor C2 actual emitido desde el limitador 112 determina la ganancia (atenuación negada) en dB que debe ser aplicada al canal 102 sin voz para mantener su nivel de potencia  $\vartheta$  dB por debajo del nivel de potencia del canal 101 de voz (en el tiempo relevante, o en la ventana temporal relevante, de la señal de entrada multicanal). Un valor adecuado típico para  $\vartheta$  es 15 dB.

Debido a que hay una relación única entre una medida expresada en una escala logarítmica (dB) y esa misma medida expresada en una escala lineal, puede construirse un circuito (o procesador programado o sino configurado) que es equivalente a los elementos 104, 105, 106, 107, 108 y 109 de la Fig. 1, en el que la potencia, la ganancia y el umbral se expresan en una escala lineal. En dicha implementación, todas las diferencias de nivel se reemplazan por relaciones de las medidas lineales. Implementaciones alternativas pueden reemplazar la medida de potencia con medidas que están relacionadas con la intensidad de la señal, tales como el valor absoluto de la señal.

La señal C1 emitida desde el limitador 111 es una señal de control de atenuación no procesada para el canal 103 sin voz (una señal de control de ganancia para el amplificador 116 de atenuación) que podría ser emitida directamente al amplificador 116 para controlar la atenuación del canal 103 sin voz. La señal C2 emitida desde el limitador 112 es una señal de control de atenuación no procesada para el canal 102 sin voz (una señal de control de ganancia para el amplificador 117 de atenuación) que podría ser emitida directamente al amplificador 117 para controlar la atenuación del canal 102 sin voz.

Las señales C1 y C2 de control de atenuación no procesadas son escaladas en los elementos 114 y 115 de multiplicación para generar las señales S3 y S4 de control de ganancia para controlar la atenuación de los canales sin voz por parte de los amplificadores 116 y 117. La señal C1 es escalada en respuesta a una secuencia de valores S1 de control de atenuación, y la señal C2 es escalada en respuesta a una secuencia de valores S2 de control de atenuación. Cada valor S1 de control es emitido a partir de la salida del elemento 134 de procesamiento (que se describirá más adelante) a una entrada del elemento 114 de multiplicación, y la señal C1 (y, de esta manera, cada valor C1 de control de ganancia "no procesado" determinado de esta manera) es emitida desde el limitador 111 a la otra entrada del elemento 114. El elemento 114 escala el valor C1 actual en respuesta al valor S1 actual multiplicando estos valores entre sí para generar el valor S3 actual, que es emitido al amplificador 116. Cada valor S2 de control es emitido desde la salida del elemento 135 de procesamiento (que se describirá más adelante) a una entrada del elemento 115 de multiplicación, y la señal C2 (y, de esta manera, cada valor C2 de control de ganancia "no procesado" determinado) es emitida desde el limitador 112 a la otra entrada del elemento 115. El elemento 115 escala el valor C2 actual en respuesta al valor S2 actual multiplicando estos valores entre sí para generar el valor S4 actual, que es emitido al amplificador 117.

Los valores S1 y S2 de control son generados de la siguiente manera.

En los elementos 130, 131 y 132 de procesamiento de probabilidad de voz, se genera una señal de probabilidad de voz (cada una de las señales P, Q y T de la Fig. 1) para cada canal de la señal de entrada multicanal. La señal P de probabilidad de voz es indicativa de una secuencia de valores de probabilidad de voz para el canal 102 sin voz; la señal Q de probabilidad de voz es indicativa de una secuencia de valores de probabilidad de voz para el canal 101 de voz, y la señal T de probabilidad de voz es indicativa de una secuencia de valores de probabilidad de voz para el canal 103 sin voz.

La señal Q de probabilidad de voz es un valor relacionado monotónicamente con la probabilidad de que la señal en el canal de voz sea, de hecho, indicativa de voz. La señal P de probabilidad de voz es un valor relacionado monotónicamente con la probabilidad de que la señal en el canal 102 sin voz sea una voz, y la señal T de probabilidad de voz es un valor relacionado monotónicamente con la probabilidad de que la señal en el canal 103 sin voz sea una voz. Los procesadores 130, 131 y 132 (que son típicamente idénticos entre sí, pero no son idénticos entre sí en algunas realizaciones) pueden implementar cualquiera de entre diversos procedimientos para determinar automáticamente la probabilidad de que las señales de entrada emitidas a los mismos sean indicativas de voz. En un ejemplo, los procesadores 130, 131 y 132 de probabilidad son idénticos entre sí, el procesador 130 genera la señal P (a partir de la información en el canal 102 sin voz), de manera que la señal P sea indicativa de una secuencia de valores de probabilidad de la voz, cada uno relacionado monotónicamente con la probabilidad de que la señal en el canal 102 en un tiempo (o ventana temporal) diferente sea una voz, el procesador 131 genera la señal Q (a partir de la información en el canal 101), de manera que la señal Q sea indicativa de una secuencia de valores de probabilidad de voz, cada uno relacionado

monotónicamente con la probabilidad de que la señal en el canal 101 en un tiempo (o ventana temporal) diferente sea una voz, el procesador 132 genera la señal T (a partir de la información en el canal 103 sin voz) de manera que la señal T sea indicativa de una secuencia de valores de probabilidad de voz, cada una relacionada monotónicamente con la probabilidad de que la señal en el canal 102 en un tiempo (o ventana temporal) diferente sea una voz, y cada uno de los procesadores 130, 131 y 132 lo hace implementando (en el canal relevante de entre los canales 102, 101 y 103) el mecanismo descrito por Robinson y Vinton en "Automated Speech/Other Discrimination for Loudness Monitoring" (Audio Engineering Society, número de prepublicación 6437 de la Convención 118, Mayo de 2005). De manera alternativa, la señal P puede ser creada manualmente, por ejemplo, por el creador de contenido, y puede ser transmitida junto con la señal de audio en el canal 102 al usuario final, y el procesador 130 puede simplemente extraer dicha señal P creada previamente a partir del canal 102 (o el procesador 130 puede ser eliminado y la señal P creada previamente puede ser emitida directamente al procesador 134). De manera similar, la señal Q puede ser creada manualmente y puede ser transmitida junto con la señal de audio en el canal 101, el procesador 131 puede simplemente extraer dicha señal Q creada previamente a partir del canal 101 (o el procesador 131 puede ser eliminado y la señal Q creada previamente puede ser emitida directamente a los procesadores 134 y 135), la señal T puede ser creada manualmente y puede ser transmitida junto con la señal de audio en el canal 103, y el procesador 132 puede simplemente extraer dicha señal T creada previamente a partir del canal 103 (o el procesador 132 puede ser eliminado y la señal T creada previamente puede ser emitida directamente al procesador 135).

En una implementación típica del procesador 134, los valores de probabilidad de voz determinados por las señales P y Q son comparados por pares para determinar la diferencia entre los valores actuales de las señales P y Q para cada uno de entre una secuencia de valores actuales de la señal P. En una implementación típica del procesador 135, los valores de probabilidad de voz determinados por las señales T y Q son comparados por pares para determinar la diferencia entre los valores actuales de las señales T y Q para cada uno de entre una secuencia de valores actuales de la señal Q. Como resultado, cada uno de los procesadores 134 y 135 genera una secuencia temporal de valores de diferencia para un par de señales de probabilidad de voz.

Los procesadores 134 y 135 se implementan preferiblemente para suavizar cada una de dichas secuencias de valores de diferencia mediante promedio de tiempo, y opcionalmente para escalar cada secuencia resultante de valores de diferencia promediados. El escalado de las secuencias de valores de diferencia promediados puede ser necesario para que los valores promediados escalados emitidos desde los procesadores 134 y 135 estén en un rango tal que las salidas de los elementos 114 y 115 de multiplicación sean útiles para accionar los amplificadores 116 y 117 de atenuación.

En una implementación típica, la señal S1 emitida desde el procesador 134 es una secuencia de valores de diferencia promediados escalados (en el que cada uno de estos valores de diferencia promediados escalados es un promedio escalado de la diferencia entre valores actuales de los valores de diferencia de las señales P y Q en una ventana temporal diferente). La señal S1 es una señal de control de ganancia de atenuación para el canal 102 sin voz, y se emplea para escalar la señal C1 de control de ganancia de atenuación no procesada generada de manera independiente para el canal 102 sin voz. De manera similar, en una implementación típica, la señal S2 emitida desde el procesador 135 es una secuencia de valores de diferencia promediados escalados (en el que cada uno de estos valores de diferencia promediados escalados es un promedio escalado de la diferencia entre los valores actuales de las señales T y Q en una ventana temporal diferente). La señal S2 es una señal de control de ganancia de atenuación para el canal 103 sin voz, y se emplea para escalar la señal C2 de control de ganancia de atenuación no procesada generada de manera independiente para el canal 103 sin voz.

El escalado de la señal C1 de control de ganancia de atenuación no procesada en respuesta a la señal S1 de control de ganancia de atenuación puede ser realizado multiplicando (en el elemento 114) cada valor de control de ganancia no procesado de la señal C1 por un valor correspondiente de entre los valores de diferencia promediados escalados de la señal S1, para generar la señal S3. El escalado de la señal C2 de control de ganancia de atenuación no procesada en respuesta a la señal S2 de control de ganancia de atenuación puede ser realizado multiplicando (en el elemento 115) cada valor de control de ganancia no procesado de la señal C2 por un valor correspondiente de entre los valores de diferencia promediados escalados de la señal S2, para generar la señal S4.

Otro ejemplo (125') del sistema se describirá con referencia a la Fig. 1A. En respuesta a una señal de audio multicanal que comprende un canal 101 de voz (canal C central) y dos canales 102 y 103 sin voz (canales L y R izquierdo y derecho), el sistema de la Fig. 1A filtra los canales sin voz para generar una señal de audio de salida multicanal filtrada que comprende el canal 101 de voz y los canales 118 y 119 sin voz (canales L' y R' izquierdo y derecho, filtrados).

En el sistema de la Fig. 1A (al igual que en el sistema de la Fig. 1), los canales 102 y 103 sin voz son emitidos a los amplificadores 117 y 116, respectivamente. Durante el funcionamiento, el amplificador 117 de atenuación es accionado por una señal S4 de control (que es indicativa de una secuencia de valores de control y, de esta manera, se denomina también secuencia S4 de valores de control) emitida desde el elemento 115 de multiplicación, y el amplificador 116 de atenuación es accionado por una señal S3 de control (que es indicativa de una secuencia de valores de control y, de esta

manera, se denomina también secuencia S3 de valores de control) emitida desde el elemento 114 de multiplicación. Los elementos 104, 105, 106, 107, 108, 109 (incluyendo los elementos 110, 120, 121, 111-1, 112-1, 111 y 112), 114, 115, 130, 131, 132, 134 y 135 de la Fig. 1A son idénticos a (y funcionan de manera idéntica a) los elementos de la Fig. 1, y no se repetirá la descripción anterior de los mismos.

5 El sistema de la Fig. 1A difiere del de la Fig. 1 en que una señal V1 de control (emitida en la salida del multiplicador 214) es usada para escalar la señal C1 de control (emitida en la salida del elemento 111 limitador) en lugar de la señal S1 de control (emitida en la salida del procesador 134), y una señal V2 de control (emitida en la salida del multiplicador 215) es usada para escalar la señal C2 de control (emitida en la salida del elemento 112 limitador) en lugar de la señal S2 de control (emitida en la salida del procesador 135). En la Fig. 1A, el escalado de la señal C1 de control de ganancia de  
10 atenuación no procesada en respuesta a la secuencia de valores V1 de control de atenuación según la invención se realiza multiplicando (en el elemento 114) cada valor de control de ganancia no procesado de la señal C1 por un valor correspondiente de entre los valores V1 de control de atenuación, para generar la señal S3, y el escalado de la señal C2 de control de ganancia de atenuación no procesada en respuesta a la secuencia de valores V2 de control de atenuación según la invención se realiza multiplicando (en el elemento 115) cada valor de control de ganancia no procesado de la  
15 señal C2 por un valor correspondiente de entre los valores V2 de control de atenuación, para generar la señal S4.

Para generar la secuencia de los valores V1 de control de atenuación, la señal Q (emitida en la salida del procesador 131) es emitida a una entrada del multiplicador 214, y la señal S1 de control (emitida en la salida del procesador 134) es emitida a la otra entrada del multiplicador 214. La salida del multiplicador 214 es la secuencia de los valores V1 de control de atenuación. Cada uno de los valores V1 de control de atenuación es uno de los valores de probabilidad de voz  
20 determinados por la señal Q, escalados por un valor correspondiente de entre los valores S1 de control de atenuación.

De manera similar, para generar la secuencia de valores V2 de control de atenuación, la señal Q (emitida en la salida del procesador 131) es emitida a una entrada del multiplicador 215, y la señal S2 de control (emitida en la salida del procesador 135) es emitida a la otra entrada del multiplicador 215. La salida del multiplicador 215 es la secuencia de los  
25 valores V2 de control de atenuación. Cada uno de los valores V2 de control de atenuación es uno de los valores de probabilidad de voz determinados por la señal Q, escalado por un valor correspondiente de entre los valores S2 de control de atenuación.

El sistema de la Fig. 1 (o el de la Fig. 1A) puede ser implementado en un software por un procesador (por ejemplo, el procesador 501 de la Fig. 5) que ha sido programado para implementar las operaciones descritas en el sistema de la Fig. 1 (o 1A). De manera alternativa, puede ser implementado en hardware con elementos de circuito conectados tal como se  
30 muestra en la Fig. 1 (o 1A).

En las variaciones en el ejemplo de la Fig. 1 (o en el de la Fig. 1A), el escalado de la señal C1 de control de ganancia de atenuación no procesada en respuesta a la señal S1 (o V1) de control de ganancia de atenuación según la invención (para generar una señal de control de ganancia de atenuación para accionar el amplificador 116) puede ser realizado de manera  
35 no lineal. Por ejemplo, dicho escalado no lineal puede generar una señal de control de ganancia de atenuación (que reemplaza a la señal S3) que no causa una atenuación por parte del amplificador 116 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 116 y, de esta manera, una atenuación nula del canal 103) cuando el valor actual de la señal S1 (o V1) está por debajo de un umbral, y causa que el valor actual de la señal de control de ganancia de atenuación (que reemplaza a la señal S3) sea igual al valor actual de la señal C1 (de manera que la señal S1 (o V1) no modifique el valor actual de C1) cuando el valor actual de la señal S1 supera el umbral. De manera alternativa, puede  
40 realizarse otro escalado lineal o no lineal de la señal C1 (en respuesta a la señal S1 o V1 de control de ganancia de atenuación de la invención) para generar una señal de control de ganancia de atenuación para accionar el amplificador 116. Por ejemplo, dicho escalado de la señal C1 puede generar una señal de control de ganancia de atenuación (que reemplaza a la señal S3) que no causa una atenuación por parte del amplificador 116 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 116) cuando el valor actual de la señal S1 (o V1) está por debajo de un  
45 umbral, y causa que el valor actual de la señal de control de ganancia de atenuación (que reemplaza a la señal S3) sea igual al valor actual de la señal C1 multiplicado por el valor actual de la señal S1 o V1 (o algún otro valor determinado a partir de este producto) cuando el valor actual de la señal S1 (o V1) supera el umbral.

De manera similar, en variaciones del ejemplo de la Fig. 1 (o de la Fig. 1A), el escalado de la señal C2 de control de ganancia de atenuación no procesada en respuesta a la señal S2 (o V2) de control de ganancia de atenuación según la  
50 invención (para generar una señal de control de ganancia de atenuación para accionar el amplificador 117) puede ser realizado de manera no lineal. Por ejemplo, dicho escalado no lineal puede generar una señal de control de ganancia de atenuación (que reemplaza a la señal S4) que no causa una atenuación por parte del amplificador 117 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 117 y, de esta manera, una atenuación nula del canal 102) cuando el valor actual de la señal S2 (o V2) está por debajo de un umbral, y causa que el valor actual de la señal de control de ganancia de atenuación (que reemplaza a la señal S4) sea igual al valor actual de la señal C2 (de manera que  
55 la señal S2 o V2 no modifique el valor actual de C2) cuando el valor actual de la señal S2 (o V2) supera el umbral. De

manera alternativa, puede realizarse otro escalado lineal o no lineal de la señal C2 (en respuesta a la señal S2 o V2 de control de ganancia de atenuación) para generar una señal de control de ganancia de atenuación para accionar el amplificador 117. Por ejemplo, dicho escalado de la señal C2 puede generar una señal de control de ganancia de atenuación (que reemplaza a la señal S4) que no causa una atenuación por parte del amplificador 117 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 117) cuando el valor actual de la señal S2 (o V2) está por debajo de un umbral, y causa que el valor actual de la señal de control de ganancia de atenuación (que reemplaza a la señal S4) sea igual al valor actual de la señal C2 multiplicado por el valor actual de la señal S2 o V2 (o algún otro valor determinado a partir de este producto) cuando el valor actual de la señal S2 (o V2) supera el umbral.

Otro ejemplo (225) de un sistema de la invención se describirá con referencia a la Fig. 2. En respuesta a una señal de audio multicanal que comprende un canal 101 de voz (canal C central) y dos canales 102 y 103 sin voz (canales L y R izquierdo y derecho), el sistema de la Fig. 2 filtra los canales sin voz para generar una señal de audio de salida multicanal filtrada que comprende el canal 101 de voz y los canales 118 y 119 sin voz (canales L' y R' izquierdo y derecho filtrados).

En el sistema de la Fig. 2 (como en el sistema de la Fig. 1), los canales 102 y 103 sin voz son emitidos a los amplificadores 117 y 116 de atenuación, respectivamente. Durante el funcionamiento, el amplificador 117 de atenuación es accionado por una señal S6 de control (que es indicativa de una secuencia de valores de control y, de esta manera, se denomina también secuencia S6 de valores de control) emitida desde el elemento 115 de multiplicación, y el amplificador 116 de atenuación es accionado por la señal S5 de control (que es indicativa de una secuencia de valores de control y, de esta manera, se denomina también secuencia S5 de valores de control), emitida desde el elemento 114 de multiplicación. Los elementos 114, 115, 130, 131, 132, 134 y 135 de la Fig. 2 son idénticos a (y funcionan de manera idéntica a) los elementos de la Fig. 1, y no se repetirá la descripción de los mismos.

El sistema de la Fig. 2 mide la potencia de las señales en cada uno de los canales 101, 102 y 103 con un banco de estimadores 201, 202 y 203 de potencia. A diferencia de sus homólogos en la Fig. 1, cada uno de los estimadores 201, 201 y 203 de potencia mide la distribución de la potencia de la señal a lo largo de la frecuencia (es decir, la potencia en cada banda de frecuencias diferente de entre un conjunto de bandas de frecuencia del canal relevante), resultando en un espectro de potencia en lugar de un único número para cada canal. La resolución espectral de cada espectro de potencia coincide idealmente con la resolución espectral de los modelos de predicción de inteligibilidad implementados por los elementos 205 y 206 (descritos más adelante).

Los espectros de potencia son suministrados al circuito 204 de comparación. El propósito del circuito 204 es determinar la atenuación a aplicar a cada canal sin voz para garantizar que la señal en el canal sin voz no reduzca la inteligibilidad de la señal en el canal de voz de manera que sea menor que un criterio predeterminado. Esta funcionalidad se consigue empleando un circuito de predicción de inteligibilidad (205 y 206) que predice la inteligibilidad de la voz a partir de los espectros de potencia de la señal (201) del canal de voz y señales (202 y 203) de los canales sin voz. Los circuitos 205 y 206 de predicción de inteligibilidad pueden implementar un modelo de predicción de inteligibilidad adecuado según las elecciones y los compromisos de diseño. Los ejemplos son el Índice de inteligibilidad de voz según se especifica en ANSI S3.5- 1997 ("Methods for Calculation of the Speech Intelligibility Index") y el modelo de sensibilidad de reconocimiento de voz de Muesch y Buus ("Using statistical decision theory to predict speech intelligibility. I. Model structure" Journal of the Acoustical Society of America, 2001, Vol. 109, p. 2896-2909). Es evidente que la salida del modelo de predicción de inteligibilidad no tiene ningún significado cuando la señal en el canal de voz es algo diferente de una voz. A pesar de esto, en adelante, se hará referencia a la salida del modelo de predicción de inteligibilidad como inteligibilidad de voz predicha. El error percibido se contabiliza en el procesamiento subsiguiente escalando los valores de ganancia emitidos desde el circuito 204 de comparación con los parámetros S1 y S2, cada uno de los cuales está relacionado con la probabilidad de que la señal en el canal de voz sea indicativa de una voz.

Los modelos de predicción de inteligibilidad tienen en común que predicen una inteligibilidad de voz incrementada o inalterada como resultado de la reducción del nivel de la señal sin voz. Continuando en el flujo del procedimiento de la Fig. 2, los circuitos 207 y 208 de comparación comparan la inteligibilidad predicha con un valor de criterio predeterminado. Si el elemento 205 determina que el nivel del canal 103 sin voz es tan bajo que la inteligibilidad predicha excede el criterio, un parámetro de ganancia, que se inicializa a 0 dB, es recuperado desde el circuito 209 es proporcionado al circuito 211 como la salida C3 del circuito 204 de comparación. Si el elemento 206 determina que el nivel del canal 102 sin voz es tan bajo que la inteligibilidad predicha excede el criterio, un parámetro de ganancia, que se inicializa a 0 dB, es recuperado desde el circuito 210 y es proporcionado al circuito 212 como la salida C4 del circuito 204 de comparación. Si el elemento 205 o 206 determina que no se cumple el criterio, el parámetro de ganancia (en el elemento relevante de entre los elementos 209 y 210) se reduce en una cantidad fija y se repite la predicción de inteligibilidad. Un tamaño de paso adecuado para reducir la ganancia es de 1 dB. La iteración que se acaba de describir continúa hasta que la inteligibilidad predicha cumple o supera el valor del criterio.

Por supuesto, es posible que la señal en el canal de voz sea tal que no pueda alcanzarse el criterio de inteligibilidad incluso en ausencia de una señal en el canal sin voz. Un ejemplo de dicha situación es una señal de voz de nivel muy bajo

o con un ancho de banda severamente restringido. En este caso, se llegará a un punto en el que cualquier reducción adicional de la ganancia aplicada al canal sin voz no afectará a la inteligibilidad de voz predicha y el criterio nunca se cumplirá. En dicha condición, el bucle formado por los elementos 205, 207 y 209 (o los elementos 206, 208 y 210) continúa indefinidamente, y puede aplicarse una lógica adicional (no mostrada) para interrumpir el bucle. Un ejemplo particularmente simple de dicha lógica es contar el número de iteraciones y salir del bucle una vez que se haya excedido un número predeterminado de iteraciones.

El escalado de la señal C3 de control de ganancia de atenuación no procesada en respuesta a la señal S1 de control de ganancia de atenuación según la invención puede ser realizado multiplicando (en el elemento 114) cada valor de control de ganancia no procesado de la señal C3 por un valor correspondiente de entre los valores de diferencia promediados escalados de la señal S1, para generar la señal S5. El escalado de la señal C4 de control de ganancia de atenuación no procesada en respuesta a la señal S2 de control de ganancia de atenuación según la invención puede ser realizado multiplicando (en el elemento 115) cada valor de control de ganancia no procesado de la señal C4 por un valor correspondiente de entre los valores de diferencia promediados escalados de la señal S2, para generar la señal S6.

El sistema de la Fig. 2 puede ser implementado en software por un procesador (por ejemplo, el procesador 501 de la Fig. 5) que ha sido programado para implementar las operaciones descritas en el sistema de la Fig. 2. De manera alternativa, puede ser implementado en hardware con elementos de circuito conectados tal como se muestra en la Fig. 2.

En variaciones del ejemplo de la Fig. 2, el escalado de la señal C3 de control de ganancia de atenuación no procesada en respuesta a la señal S1 de control de ganancia de atenuación (para generar una señal de control de ganancia de atenuación para accionar el amplificador 116) puede ser realizado de una manera no lineal. Por ejemplo, dicho escalado no lineal puede generar una señal de control de ganancia de atenuación (que reemplaza a la señal S5) que no causa una atenuación por parte del amplificador 116 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 116 y, de esta manera, una atenuación nula del canal 103) cuando el valor actual de la señal S1 está por debajo de un umbral, y causa que el valor actual de la señal de control de ganancia de atenuación (que reemplaza a la señal S5) sea igual al valor actual de la señal C3 (de manera que la señal S1 no modifique el valor actual de C3) cuando el valor actual de la señal S1 supera el umbral. De manera alternativa, puede realizarse otro escalado lineal o no lineal de la señal C3 (en respuesta a la señal S1 de control de ganancia de atenuación de la invención) para generar una señal de control de ganancia de atenuación para accionar el amplificador 116. Por ejemplo, dicho escalado de la señal C3 puede generar una señal de control de ganancia de atenuación (que reemplaza a la señal S5) que no causa una atenuación por parte del amplificador 116 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 116) cuando el valor actual de la señal S1 está por debajo de un umbral, y causa que el valor actual de la señal de control de ganancia de atenuación (que reemplaza a la señal S5) sea igual al valor actual de la señal C3 multiplicado por el valor actual de la señal S1 (o algún otro valor determinado a partir de este producto) cuando el valor actual de la señal S1 supera el umbral.

De manera similar, en variaciones del ejemplo de la Fig. 2, el escalado de la señal C4 de control de ganancia de atenuación no procesada en respuesta a la señal S2 de control de ganancia de atenuación (para generar una señal de control de ganancia de atenuación para accionar el amplificador 117) puede ser realizado de una manera no lineal. Por ejemplo, dicho escalado no lineal puede generar una señal de control de ganancia de atenuación (que reemplaza a la señal S6) que no causa una atenuación del amplificador 117 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 117 y, de esta manera, una atenuación nula del canal 102) cuando el valor actual de la señal S2 está por debajo de un umbral y causa que el valor actual de la señal de control de ganancia de atenuación (que reemplaza a la señal S6) sea igual al valor actual de la señal C4 (de manera que la señal S2 no modifique el valor actual de C4) cuando el valor actual de la señal S2 supera el umbral. De manera alternativa, puede realizarse otro escalado lineal o no lineal de la señal C4 (en respuesta a la señal de control S2 de ganancia de atenuación de la invención) para generar una señal de control de ganancia de atenuación para accionar el amplificador 117. Por ejemplo, dicho escalado de la señal C4 puede generar una señal de control de ganancia de atenuación (que reemplaza la señal S6) que no causa una atenuación del amplificador 117 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 117) cuando el valor actual de la señal S2 está por debajo de un umbral, y causa que el valor actual de la señal de control de ganancia de atenuación (que reemplaza a la señal S6) sea igual al valor actual de la señal C4 multiplicado por el valor actual de la señal S2 (o algún otro valor determinado a partir de este producto) cuando el valor actual de la señal S2 supera el umbral.

Otro ejemplo (225') del sistema se describirá con referencia a la Fig. 2A. En respuesta a una señal de audio multicanal que comprende un canal 101 de voz (canal C central) y dos canales 102 y 103 sin voz (canales L y R izquierdo y derecho), el sistema de la Fig. 2A filtra los canales sin voz para generar una señal de audio de salida multicanal filtrada que comprende el canal 101 de voz y los canales 118 y 119 sin voz filtrados (canales L' y R' izquierdo y derecho filtrados).

En el sistema de la Fig. 2A (como en el sistema de la Fig. 2), los canales 102 y 103 sin voz son emitidos a los amplificadores 117 y 116 de atenuación, respectivamente. Durante el funcionamiento, el amplificador 117 de atenuación es accionado por una señal S6 de control (que es indicativa de una secuencia de valores de control y, de esta manera, se denomina también secuencia S6 de valores de control) emitida desde el elemento 115 de multiplicación, y el amplificador

116 de atenuación es accionado por la señal S5 de control (que es indicativa de una secuencia de valores de control y, de esta manera, se denomina también secuencia S5 de valores de control), emitida desde el elemento 114 de multiplicación. Los elementos 201, 202, 203, 204, 114, 115, 130 y 134 de la Fig. 2A son idénticos a (y funcionan de manera idéntica a) los elementos con numeración idéntica de la Fig. 2, y no se repetirá la descripción de los mismos.

5 El sistema de la Fig. 2A difiere del de la Fig. 2 en dos aspectos principales. En primer lugar, el sistema está configurado para generar (es decir, derivar) un canal sin voz "derivado" (L + R) a partir de dos canales sin voz individuales (102 y 103) de la señal de audio de entrada, y para determinar los valores de control de atenuación (V3) en respuesta a este canal sin voz derivado. Por el contrario, el sistema de la Fig. 2 determina los valores S1 de control de atenuación en respuesta a un canal sin voz (canal 102) de la señal de audio de entrada y determina los valores S2 de control de atenuación en  
10 respuesta a otro canal sin voz (canal 103) de la señal de audio de entrada. Durante el funcionamiento, el sistema de la Fig. 2A atenúa cada canal sin voz de la señal de audio de entrada (cada uno de los canales 102 y 103) en respuesta al mismo conjunto de valores V3 de control de atenuación. Durante el funcionamiento, el sistema de la Fig. 2 atenúa el canal 102 sin voz de la señal de audio de entrada en respuesta a los valores S2 de control de atenuación, y atenúa el canal 103 sin voz de la señal de audio de entrada en respuesta a un conjunto diferente de valores de control de atenuación (valores S1).  
15

El sistema de la Fig. 2A incluye un elemento 129 de adición cuyas entradas están acopladas para recibir los canales 102 y 103 sin voz de la señal de audio de entrada. El canal sin voz derivado (L + R) es emitido en la salida del elemento 129. El elemento 130 de procesamiento de probabilidad de voz emite la señal P de probabilidad de voz en respuesta al canal L + R sin voz derivado desde el elemento 129. En la Fig. 2A, la señal P es indicativa de una secuencia de valores de  
20 probabilidad de voz para el canal sin voz derivado. Típicamente, la señal P de probabilidad de voz de la Fig. 2A es un valor relacionado monótonicamente con la probabilidad de que la señal en el canal sin voz derivado sea una voz. La señal Q de probabilidad de voz (generada por el procesador 131) de la Fig. 2A es idéntica a la señal Q de probabilidad de voz de la Fig. 2 descrita anteriormente.

Un segundo aspecto principal en el que el sistema de la Fig. 2A difiere del de la Fig. 2 es el siguiente. En la Fig. 2A, se usa la señal V3 de control (emitida en la salida del multiplicador 214) (en lugar de la señal S1 de control emitida en la salida del procesador 134) para escalar la señal C3 de control de ganancia de atenuación no procesada (emitida en la salida del elemento 211), y la señal V3 de control se usa también (en lugar de la señal S2 de control emitida en la salida del procesador 135 de la Fig. 2) para escalar la señal C4 de control de ganancia de atenuación no procesada (emitida en la salida del elemento 212). En la Fig. 2A, el escalado de la señal C3 de control de ganancia de atenuación no procesada en  
25 respuesta a la secuencia de valores de control de atenuación indicados por la señal V3 (a los que se hará referencia como valores V3 de control de atenuación) es realizado multiplicando (en el elemento 114) cada valor de control de ganancia no procesado de la señal C3 por un valor correspondiente de entre los valores V3 de control de atenuación, para generar la señal S5, y el escalado de la señal C4 de control de ganancia de atenuación no procesada en respuesta a la secuencia de valores V3 de control de atenuación es realizado multiplicando (en el elemento 115) cada valor de control de ganancia no procesado de la señal C4 por un valor correspondiente de entre los valores V3 de control de atenuación, para generar la señal S6.  
30  
35

Durante el funcionamiento, el sistema de la Fig. 2A genera la secuencia de valores V3 de control de atenuación de la siguiente manera. La señal Q de probabilidad de voz (emitida en la salida del procesador 131 de la Fig. 2A) es emitida a una entrada del multiplicador 214, y la señal S1 de control de atenuación (emitida en la salida del procesador 134) es emitida a la otra entrada del multiplicador 214. La salida del multiplicador 214 es la secuencia de valores V3 de control de atenuación. Cada uno de los valores V3 de control de atenuación es uno de los valores de probabilidad de voz determinados por la señal Q, escalado por un valor correspondiente de entre los valores S1 de control de atenuación.  
40

Otro ejemplo (325) de un sistema se describirá con referencia a la Fig. 3. En respuesta a una señal de audio multicanal que comprende un canal 101 de voz (canal C central) y dos canales 102 y 103 sin voz (canales L y R izquierdo y derecho), el sistema de la Fig. 3 filtra los canales sin voz para generar una señal de audio de salida multicanal filtrada que comprende el canal 101 de voz y los canales 118 y 119 sin voz (canales L' y R' izquierdo y derecho filtrados).  
45

En el sistema de la Fig. 3, cada una de las señales en los tres canales de entrada es dividida en sus componentes espectrales por el banco 301 de filtros (para el canal 101), el banco 302 de filtros (para el canal 102) y el banco 303 de filtros (para el canal 103). El análisis espectral puede conseguirse con bancos de filtros de canal N en el dominio del tiempo. Según un ejemplo, cada banco de filtros divide el rango de frecuencias en bandas de 1/3 de octava o imita el filtrado que se supone que ocurre en el oído interno humano. El hecho de que la señal emitida desde cada banco de filtros consiste en N sub-señales se ilustra mediante el uso de líneas gruesas.  
50

En el sistema de la Fig. 3, las componentes de frecuencia de las señales en los canales 102 y 103 sin voz son emitidas a los amplificadores 117 y 116 de atenuación, respectivamente. Durante el funcionamiento, el amplificador 117 de atenuación es accionado por una señal S8 de control (que es indicativa de una secuencia de valores de control y, de esta manera, se denomina también secuencia S8 de valores de control) emitida desde el elemento 115' de multiplicación, y el  
55

amplificador 116 de atenuación es accionado por una señal S7 de control (que es indicativa de una secuencia de valores de control y, de esta manera, se denomina también secuencia S7 de valores de control) emitida desde el elemento 114' de multiplicación. Los elementos 130, 131, 132, 134 y 135 de la Fig. 3 son idénticos a (y funcionan de manera idéntica a) los elementos con numeración idéntica de la Fig. 1, y no se repetirá la descripción de los mismos.

5 El procedimiento de la Fig. 3 puede ser reconocido como un proceso de rama lateral. Siguiendo la ruta de señal mostrada en la Fig. 3, las N sub-señales generadas en el banco 302 para el canal 102 sin voz son escaladas, cada una, por un miembro de un conjunto de N valores de ganancia por el amplificador 117 de atenuación y las N sub-señales generadas en el banco 303 para el canal 103 sin voz son escaladas por un miembro de un conjunto de N valores de ganancia por el amplificador 116 de atenuación. La derivación de estos valores de ganancia se describirá más adelante. A continuación,  
10 las sub-señales escaladas son recombinadas en una sola señal de audio. Esto puede realizarse mediante una suma simple (por el circuito 313 de suma para el canal 102 y por el circuito 314 de suma para el canal 103). De manera alternativa, puede usarse un banco de filtros de síntesis que coincide con el banco de filtros de análisis. Este procedimiento resulta en la señal R' sin voz modificada (118) y la señal L' sin voz modificada (119).

Ahora, con relación a la descripción de la ruta de rama lateral del procedimiento de la Fig. 3, cada salida de banco de filtro se pone a disposición de un banco correspondiente de N estimadores (304, 305 y 306) de potencia. Los espectros de potencia resultantes para los canales 101 y 102 sirven como entradas a un circuito 307 de optimización que tiene como salida un vector C6 de ganancias N-dimensional. Los espectros de potencia resultantes para los canales 101 y 103 sirven como entradas a un circuito 308 de optimización que tiene como salida un vector C5 de ganancias N-dimensional. La optimización emplea tanto un circuito (309 y 310) de predicción de inteligibilidad como un circuito (311 y 312) de cálculo de intensidad sonora para encontrar el vector de ganancias que maximiza la intensidad sonora de cada canal sin voz mientras mantiene un nivel predeterminado de inteligibilidad predicha de la señal de voz en el canal 101. Los modelos adecuados para predecir la inteligibilidad se han descrito con referencia a la Fig. 2. Los circuitos 311 y 312 de cálculo de intensidad sonora pueden implementar un modelo de predicción de intensidad sonora adecuado según las opciones y los compromisos de diseño. Los ejemplos de modelos adecuados son el American National Standard ANSI S3.4-2007 "Procedure for the Computation of Loudness of Steady Sounds" y el estándar alemán DIN 45631 "Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum".  
15  
20  
25

Dependiendo de los recursos computacionales disponibles y de las restricciones impuestas, la forma y la complejidad de los circuitos (307, 308) de optimización pueden variar ampliamente. Según un ejemplo, se utiliza una optimización restringida, multidimensional, iterativa, de N parámetros libres. Cada parámetro representa la ganancia aplicada a una de las bandas de frecuencias del canal sin voz. Pueden aplicarse técnicas estándar, tales como seguir el gradiente más pronunciado ("steepest gradient") en el espacio de búsqueda N dimensional para encontrar el máximo. En otra realización, un enfoque computacionalmente menos exigente impone una restricción a las funciones ganancia vs. frecuencia de manera que sean miembros de un pequeño conjunto de posibles funciones ganancia vs. frecuencia, tales como un conjunto de diferentes gradientes espectrales o filtros atenuadores o limitadores ("shelf"). Con esta restricción adicional, el problema de optimización puede reducirse a un pequeño número de optimizaciones unidimensionales. En todavía otro ejemplo, se realiza una búsqueda exhaustiva sobre un conjunto muy pequeño de posibles funciones de ganancia. Este último enfoque podría ser particularmente deseable en aplicaciones en tiempo real en las que se desea una carga computacional y una velocidad de búsqueda constantes.  
30  
35

Las personas con conocimientos ordinarios en la materia reconocerán fácilmente las restricciones adicionales que podrían imponerse en la optimización y están reflejadas también por las realizaciones respectivas de las reivindicaciones dependientes.  
40

Un ejemplo es restringir la intensidad sonora del canal sin voz modificado de manera que no sea mayor que la intensidad sonora antes de la modificación. Otro ejemplo es imponer un límite a las diferencias de ganancia entre bandas de frecuencia adyacentes con el fin de limitar la posibilidad de solapamiento temporal en el banco de filtros de reconstrucción (313, 314) o con el fin de reducir la posibilidad de modificaciones de timbre objetables. Las restricciones deseables dependen tanto de la implementación técnica del banco de filtros como del compromiso elegido entre la mejora de inteligibilidad y la modificación del timbre. En aras de una mayor claridad de la ilustración, estas restricciones se omiten en la Fig. 3.  
45

El escalado del vector C6 de control de ganancia de atenuación no procesado N-dimensional en respuesta a la señal S2 de control de ganancia de atenuación según la invención puede ser realizado multiplicando (en el elemento 115') cada valor de control de ganancia no procesado del vector C6 por un valor correspondiente de entre los valores de diferencia promediados escalados de la señal S2, para generar el vector S8 de control de ganancia de atenuación N-dimensional. El escalado del vector C5 de control de ganancia de atenuación no procesado N-dimensional en respuesta a la señal S1 de control de ganancia de atenuación según la invención puede ser realizado multiplicando (en el elemento 114') cada valor de control de ganancia no procesado del vector C5 por un valor correspondiente de entre los valores de diferencia promediados escalados de la señal S1, para generar el vector S7 de control de ganancia de atenuación N-dimensional.  
50  
55

El sistema de la Fig. 3 puede ser implementado en software por un procesador (por ejemplo, el procesador 501 de la Fig. 5) que ha sido programado para implementar las operaciones descritas del sistema de la Fig. 3. De manera alternativa, puede ser implementado en hardware con elementos de circuito conectados tal como se muestra en la Fig. 3.

En variaciones en el ejemplo de la Fig. 3, el escalado del vector C5 de control de ganancia de atenuación no procesado en respuesta a la señal S1 de control de ganancia de atenuación según la invención (para generar un vector de control de ganancia de atenuación para accionar el amplificador 116) puede ser realizado de una manera no lineal. Por ejemplo, dicho escalado no lineal puede generar un vector de control de ganancia de atenuación (que reemplaza al vector S7) que no causa una atenuación del amplificador 116 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 116 y, de esta manera, una atenuación nula del canal 103) cuando el valor actual de la señal S1 está por debajo de un umbral, y causa que los valores actuales del vector de control de ganancia de atenuación (que reemplaza al vector S7) sean iguales a los valores actuales del vector C5 (de manera que la señal S1 no modifique los valores actuales de C5) cuando el valor actual de la señal S1 supera el umbral. De manera alternativa, puede realizarse otro escalado lineal o no lineal del vector C5 (en respuesta a la señal S1 de control de ganancia de atenuación de la invención) para generar un vector de control de ganancia de atenuación para accionar el amplificador 116. Por ejemplo, dicho escalado del vector C5 puede generar un vector de control de ganancia de atenuación (que reemplaza al vector S7) que no causa una atenuación por parte del amplificador 116 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 116) cuando el valor actual de la señal S1 está por debajo de un umbral, y causa que el valor actual del vector de control de ganancia de atenuación (que reemplaza al vector S7) sea igual al valor actual del vector C5 multiplicado por el valor actual de la señal S1 (o algún otro valor determinado a partir de este producto) cuando el valor actual de la señal S1 supera el umbral.

De manera similar, en variaciones del ejemplo de la Fig. 3, el escalado del vector C6 de control de ganancia de atenuación no procesado en respuesta a la señal S2 de control de ganancia de atenuación (para generar un vector de control de ganancia de atenuación para accionar el amplificador 117) puede ser realizado de una manera no lineal. Por ejemplo, dicho escalado no lineal puede generar un vector de control de ganancia de atenuación (que reemplaza al vector S8) que no causa una atenuación por parte del amplificador 117 (es decir, la aplicación de una ganancia unitaria por parte del amplificador 117 y, de esta manera, una atenuación nula del canal 102) cuando el valor actual de la señal S2 está por debajo de un umbral y causa que los valores actuales del vector de control de ganancia de atenuación (que reemplaza al vector S8) sean iguales a los valores actuales del vector C6 (de manera que la señal S2 no modifique los valores actuales de C6) cuando el valor actual de la señal S2 supera el umbral. De manera alternativa, puede realizarse otro escalado lineal o no lineal del vector C6 (en respuesta a la señal S2 de control de ganancia de atenuación de la invención) para generar un vector de control de ganancia de atenuación para accionar el amplificador 117. Por ejemplo, dicho escalado del vector C6 puede generar un vector de control de ganancia de atenuación (que reemplaza al vector S8) que no causa una atenuación por parte del amplificador 117 (es decir, la aplicación de una ganancia unitaria por el amplificador 117) cuando el valor actual de la señal S2 está por debajo de un umbral, y causa que el valor actual del vector de control de ganancia de atenuación (que reemplaza al vector S8) sea igual al valor actual del vector C6 multiplicado por el valor actual de la señal S2 (o algún otro valor determinado a partir de este producto) cuando el valor actual de la señal S2 supera el umbral.

A partir de esta descripción, será evidente para las personas con conocimientos ordinarios en la materia la manera en la que puede modificarse el sistema de la Fig. 1, 1A, 2, 2A o 3 (y las variaciones en cualquiera de los mismos) para filtrar una señal de entrada de audio multicanal que tiene un canal de voz y cualquier número de canales sin voz. Se proporcionaría un amplificador de atenuación (o un equivalente de software del mismo) para cada canal sin voz, y se generaría una señal de control de ganancia de atenuación (por ejemplo, escalando una señal de control de ganancia de atenuación) para accionar cada amplificador de atenuación (o software equivalente al mismo).

Tal como se ha descrito, el sistema de la Fig. 1, 1A, 2, 2A o 3 (y cada una de las muchas variaciones en el mismo) son operables para llevar a la práctica realizaciones del procedimiento de la invención para filtrar una señal de audio multicanal que tiene un canal de voz y al menos un canal sin voz para mejorar la inteligibilidad de la voz determinada por la señal. En una primera clase de dichas realizaciones, el procedimiento incluye las etapas de:

(a) determinar al menos un valor de control de atenuación (por ejemplo, la señal S1 o S2 de la Fig. 1, 2 o 3, o la señal V1, V2 o V3 de la Fig. 1A o 2A) indicativa de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por al menos un canal sin voz de la señal de audio;

y

(b) atenuar al menos un canal sin voz de la señal de audio en respuesta al por lo menos un valor de control de atenuación (por ejemplo, en el elemento 114 y el amplificador 116, o en el elemento 115 y el amplificador 117, de la Fig. 1, 1A, 2, 2A, o 3).

Típicamente, la etapa de atenuación comprende escalar una señal de control de atenuación no procesada (por ejemplo, la señal C1 o C2 de control de ganancia de atenuación de la Fig. 1 o 1A, o la señal C3 o C4 de la Fig. 2 o 2A) para el canal sin voz en respuesta al por lo menos un valor de control de atenuación. Preferiblemente, el canal sin voz es atenuado para

mejorar la inteligibilidad de la voz determinada por el canal de voz sin atenuar de manera no deseada el contenido mejorador de la voz determinado por el canal sin voz. En algunos ejemplos de realizaciones en la primera clase, la etapa (a) incluye una etapa de generar una señal de control de atenuación (por ejemplo, la señal S1 o S2 de la Fig. 1, 2 o 3, o la señal V1, V2 o V3 de la Fig. 1A o 2A) indicativa de una secuencia de valores de control de atenuación, en el que cada uno de los valores de control de atenuación es indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por al menos un canal sin voz de la señal de audio en un tiempo diferente (por ejemplo, en un intervalo de tiempo diferente), y la etapa (b) incluye las etapas de: escalar una señal de control de ganancia de atenuación (por ejemplo, la señal C1 o C2 de la Fig. 1 o 1A, o la señal C3 o C4) de la Fig. 2 o 2A) en respuesta a la señal de control de atenuación para generar una señal de control de ganancia escalada (por ejemplo, la señal S3 o S4 de la Fig. 1 o 1A, o la señal S5 o S6 de la Fig. 2 o 2A), y aplicar la señal de control de ganancia escalada para atenuar el canal sin voz (por ejemplo, emitir la señal de control de ganancia escalada a la circuitería 116 o 117 de atenuación, de la Fig. 1, 1A, 2 o 2A, para controlar la atenuación de al menos un canal sin voz por la circuitería de atenuación). Por ejemplo, la etapa (a) incluye una etapa de comparar una primera secuencia de características relacionadas con la voz (por ejemplo, la señal Q de la Fig. 1 o 2) indicativas del contenido relacionado con la voz determinado por el canal de voz con una segunda secuencia de características relacionadas con la voz (por ejemplo, la señal P de la Fig. 1 o 2) indicativas del contenido relacionado con la voz determinado por el canal sin voz para generar la señal de control de atenuación, y cada uno de los valores de control de atenuación indicados por la señal de control de atenuación es indicativo de una medida de similitud entre la primera secuencia de características relacionadas con la voz y la segunda secuencia de características relacionadas con la voz en un tiempo diferente (por ejemplo, en un intervalo de tiempo diferente). En algunos ejemplos de las realizaciones, cada valor de control de atenuación es un valor de control de ganancia.

En algunas realizaciones en la primera clase, cada valor de control de atenuación está relacionado monotónicamente con la probabilidad de que el canal sin voz sea indicativo de contenido mejorador de voz que mejora la inteligibilidad (u otra cualidad percibida) del contenido de voz determinado por el canal de voz. En algunos otros ejemplos relacionados con realizaciones en la primera clase, cada valor de control de atenuación está relacionado monotónicamente con un valor mejorador de voz esperado del canal sin voz (por ejemplo, una medida de probabilidad de que el canal sin voz sea indicativo de contenido mejorador de voz, multiplicado por una medida de la mejora de la cualidad percibida que el contenido mejorador de voz determinado por el canal sin voz proporcionaría al contenido de voz determinado por la señal multicanal). Por ejemplo, cuando la etapa (a) incluye una etapa de comparar (por ejemplo, en el elemento 134 o 135 de la Fig. 1 o la Fig. 2) una primera secuencia de características relacionadas con la voz indicativas del contenido relacionado con la voz determinada por el canal de voz con una segunda secuencia de características relacionadas con la voz indicativas del contenido relacionado con la voz determinada por el canal sin voz, la primera secuencia de características relacionadas con la voz puede ser una secuencia de valores de probabilidad de voz, cada una de las cuales indica la probabilidad en un tiempo diferente (por ejemplo, en un intervalo de tiempo diferente) de que el canal de voz sea indicativo de voz (en lugar de contenido de audio distinto de una voz), y la segunda secuencia de características relacionadas con la voz puede ser también una secuencia de valores de probabilidad de voz, cada uno de los cuales indica la probabilidad en un tiempo diferente (por ejemplo, en un intervalo de tiempo diferente) de que el canal sin voz sea indicativo de voz.

Tal como se ha descrito, el sistema de la Fig. 1, 1A, 2, 2A o 3 (y cada una de las muchas variaciones en el mismo) es operable también para realizar una segunda clase de realizaciones del procedimiento de la invención para filtrar una señal de audio multicanal que tiene un canal de voz y al menos un canal sin voz para mejorar la inteligibilidad de la voz determinada por la señal. En la segunda clase de realizaciones, el procedimiento incluye las etapas de:

- (a) comparar una característica del canal de voz y una característica del canal sin voz para generar al menos un valor de atenuación (por ejemplo, valores determinados por la señal C1 o C2 de la Fig. 1, o por la señal C3 o C4 de la Fig. 2, o por la señal C5 o C6 de la Fig. 3) para controlar la atenuación del canal sin voz con relación al canal de voz; y
- (b) ajustar el al menos un valor de atenuación en respuesta a al menos un valor de probabilidad de mejora de voz (por ejemplo, la señal S1 o S2 de la Fig. 1, 2 o 3) para generar al menos un valor de atenuación ajustado (por ejemplo, valores determinados de la señal S3 o S4 de la Fig. 1, o por la señal S5 o S6 de la Fig. 2, o por la señal S7 o S8 de la Fig. 3) para controlar la atenuación del canal sin voz con relación al canal de voz. Típicamente, la etapa de ajuste es o incluye escalar (por ejemplo, en el elemento 114 o 115 de la Fig. 1, 2 o 3) cada uno de dichos valores de atenuación en respuesta a uno de dichos valores de probabilidad de mejora de voz para generar uno de dichos valores de atenuación ajustados. Típicamente, cada valor de probabilidad de mejora de voz es indicativo de (por ejemplo, está relacionado monotónicamente con) la probabilidad de que el canal sin voz sea indicativo de contenido mejorador de voz (contenido que mejora la inteligibilidad u otra cualidad percibida del contenido de voz determinado por el canal de voz). En ejemplos relacionados, el valor de probabilidad de mejora de voz es indicativo de un valor mejorador de voz esperado del canal sin voz (por ejemplo, una medida de la probabilidad de que el canal sin voz sea indicativo de contenido mejorador de voz multiplicado por una medida de mejora de la cualidad percibida que el contenido mejorador de voz determinado por el canal sin voz proporcionaría al contenido de voz determinado por la señal de audio multicanal). En algunas realizaciones en la segunda clase, el valor de probabilidad de mejora de voz es una

5 secuencia de valores de comparación (por ejemplo, valores de diferencia) determinados por un procedimiento que incluye una etapa de comparar una primera secuencia de características relacionadas con la voz, indicativas del contenido relacionado con la voz determinada por el canal de voz, con una segunda secuencia de características relacionadas con la voz, indicativas del contenido relacionado con la voz determinada por el canal sin voz, y cada uno de los valores de comparación es una medida de similitud entre la primera secuencia de características relacionadas con la voz y la segunda secuencia de características relacionadas con la voz en un tiempo diferente (por ejemplo, en un intervalo de tiempo diferente). En realizaciones típicas en la segunda clase, el procedimiento incluye también la etapa de atenuar el canal sin voz (por ejemplo, en el amplificador 116 o 117 de la Fig. 1, 2 o 3) en respuesta al por lo menos un valor de atenuación ajustado. La etapa (b) puede comprender escalar el al menos un valor de atenuación (por ejemplo, cada valor de atenuación determinado por la señal C1 o C2 de la Fig. 1), u otro valor de atenuación determinado por una señal de control de ganancia de atenuación u otra señal de control de atenuación no procesada) en respuesta al por lo menos un valor de probabilidad de mejora de voz (por ejemplo, el valor correspondiente determinado por la señal S1 o S2 de la Fig. 1).

15 Durante el funcionamiento del sistema de la Fig. 1 para llevar a la práctica una realización en la segunda clase, cada valor de atenuación determinado por la señal C1 o C2 es un primer factor indicativo de una cantidad de atenuación del canal sin voz necesario para limitar la relación de potencia de señal en el canal sin voz a la potencia de señal en el canal de voz para que no exceda un umbral predeterminado, escalado por un segundo factor relacionado monotónicamente con la probabilidad de que el canal de voz sea indicativo de voz. Típicamente, la etapa de ajuste en estas realizaciones es (o incluye) escalar cada valor C1 o C2 de atenuación por un valor de probabilidad de mejora de voz (determinado por la señal S1 o S2) para generar un valor de atenuación ajustado (determinado por la señal S3 o S4), donde el valor de probabilidad de mejora de voz es un factor relacionado monotónicamente con uno de entre: una probabilidad de que el canal sin voz sea indicativo de contenido mejorador de voz (contenido que mejora la inteligibilidad u otra cualidad percibida del contenido de voz determinado por la señal multicanal) y un valor mejorador de voz esperado del canal sin voz (por ejemplo, una medida de probabilidad de que el canal sin voz sea indicativo de contenido mejorador de voz multiplicado por una medida de la mejora de la cualidad percibida que el contenido mejorador de voz en el canal sin voz proporcionaría al contenido de voz determinado por la señal multicanal).

20 Durante el funcionamiento del sistema de la Fig. 2 para llevar a la práctica una realización en la segunda clase, cada valor de atenuación determinado por la señal C3 o C4 es un primer factor indicativo de una cantidad (por ejemplo, la cantidad mínima) de atenuación del canal sin voz suficiente para causar la inteligibilidad de voz predicha determinada por el canal de voz en presencia de contenido determinado por el canal sin voz exceda un valor de umbral predeterminado, escalado por un segundo factor relacionado monotónicamente con la probabilidad de que el canal de voz sea indicativo de una voz. Preferiblemente, la inteligibilidad de voz predicha determinada por el canal de voz en presencia del contenido determinado por el canal sin voz es determinada según un modelo de predicción de inteligibilidad basado en psico-acústica. Típicamente, la etapa de ajuste en estas realizaciones es (o incluye) escalar cada uno de dichos valores de atenuación por uno de dichos valores de probabilidad de mejora de voz (determinado por la señal S1 o S2) para generar un valor de atenuación ajustado (determinado por la señal S5 o S6), donde el valor de probabilidad de mejora de voz es un factor relacionado monotónicamente con uno de entre: una probabilidad de que el canal sin voz sea indicativo de contenido de mejora de voz y un valor esperado de mejora de voz del canal sin voz.

30 Durante el funcionamiento del sistema de la Fig. 3 para llevar a la práctica una realización en la segunda clase, cada valor de atenuación determinado por la señal C1 o C2 es determinado mediante etapas que incluyen determinar (en el elemento 301, 302 o 303) un espectro de potencia indicativo de la potencia como una función de la frecuencia, de cada uno de entre el canal 101 de voz y los canales 102 y 103 sin voz, y realizar una determinación en el dominio de la frecuencia del valor de atenuación, determinando de esta manera la atenuación como una función de la frecuencia a aplicar a los componentes de frecuencia del canal sin voz.

45 En una clase de realizaciones, la invención es un procedimiento y un sistema para mejorar la voz determinada por una señal de entrada de audio multicanal. En algunos ejemplos de las mismas, el sistema de la invención incluye un módulo o subsistema de análisis (por ejemplo, elementos 130-135, 104-109, 114 y 115 de la Fig. 1, o elementos 130-135, 201-204, 114 y 115 de la Fig. 2) configurado para analizar la señal multicanal de entrada para generar valores de control de atenuación, y un subsistema de atenuación (por ejemplo, los amplificadores 116 y 117 de la Fig. 1 o la Fig. 2). El subsistema de atenuación incluye circuitería de atenuación (accionada por al menos algunos de los valores de control de atenuación) acoplada y configurada para aplicar atenuación ("ducking") a cada canal sin voz de la señal de entrada para generar una señal de salida de audio filtrada. La circuitería de atenuación es accionada por valores de control en el sentido de que la atenuación que aplica a los canales sin voz viene determinada por los valores actuales de los valores de control.

55 En algunas realizaciones, una relación de la potencia del canal de voz (por ejemplo, el canal central) a la potencia del canal sin voz (por ejemplo, el canal lateral y/o el canal trasero) es usada para determinar cuánta atenuación ("ducking") debería aplicarse a cada canal sin voz. Por ejemplo, en el ejemplo de la Fig. 1, la ganancia aplicada por cada uno de los

amplificadores 116 y 117 de atenuación es reducida en respuesta a una disminución en un valor de control de ganancia (emitido desde el elemento 114 o el elemento 115) que es indicativo de una menor potencia (dentro de los límites) del canal 101 de voz con relación a la potencia de un canal sin voz (canal 102 izquierdo o canal 103 derecho) determinada en el módulo de análisis (es decir, un amplificador de atenuación atenúa más un canal sin voz con relación al canal de voz cuando la potencia del canal de voz disminuye (dentro de los límites) con relación a la potencia del canal sin voz), suponiendo que no haya cambios en la probabilidad (según se determina en el módulo de análisis) de que el canal sin voz incluya contenido mejorador de voz que mejora el contenido de voz determinado por el canal de voz.

En algunos ejemplos alternativos, una versión modificada del módulo de análisis de la Fig. 1 o la Fig. 2 procesa individualmente cada una o más sub-bandas de frecuencia de cada canal de la señal de entrada. Específicamente, la señal en cada canal puede ser pasada a través de un banco de filtros pasa banda, que proporcionan tres conjuntos de  $n$  sub-bandas:  $\{L_1, L_2, \dots, L_n\}$ ,  $\{C_1, C_2, \dots, C_n\}$  y  $\{R_1, R_2, \dots, R_n\}$ . Las sub-bandas coincidentes son pasadas a  $n$  instancias del módulo de análisis de la Fig. 1 (o de la Fig. 2), y las sub-señales filtradas (las salidas de los amplificadores de atenuación para los canales sin voz, y las sub-señales del canal de voz no filtradas) son combinadas mediante circuitos de suma para generar la señal de salida de audio multicanal filtrada. Para llevar a cabo en cada sub-banda las operaciones realizadas por el elemento 109 de la Fig. 1, puede seleccionarse un valor  $\vartheta_n$  de umbral separado (correspondiente al valor  $\vartheta$  de umbral del elemento 109) para cada sub-banda. Una buena opción es un conjunto en el que  $\vartheta_n$  es proporcional al número promedio de características de voz transportadas en la región de frecuencia correspondiente; es decir, a las bandas en los extremos del espectro de frecuencias se les asignan umbrales más bajos que a las bandas correspondientes a las frecuencias de voz dominantes. Esta implementación puede ofrecer un muy buen compromiso entre complejidad computacional y rendimiento.

La Fig. 4 es un diagrama de bloques de un sistema 420 (un DSP de audio configurable) que ha sido configurado para llevar a la práctica una realización del procedimiento de la invención. El sistema 420 incluye circuitería 422 DSP programable (un módulo de mejora de voz activo del sistema 420) acoplado para recibir una señal de entrada de audio multicanal. Por ejemplo, los canales Lin y Rin sin voz de la señal pueden corresponder a los canales 102 y 103 de la señal de entrada descritos con referencia a las Figs. 1, 1A, 2, 2A y 3, la señal puede incluir también canales sin voz adicionales (por ejemplo, los canales trasero izquierdo y trasero derecho), y el canal Cin de voz de la señal puede corresponder al canal 101 de la señal de entrada descrito con referencia a las Figs. 1, 1A, 2, 2A y 3. La circuitería 422 es configurada en respuesta a los datos de control desde una interfaz 421 de control para llevar a la práctica una realización del procedimiento de la invención, para generar una señal de audio de salida multicanal con voz mejorada en respuesta a la señal de entrada de audio. Para programar el sistema 420, el software apropiado es activado desde un procesador externo para controlar la interfaz 421 de control, y la interfaz 421 emite en respuesta los datos de control apropiados a la circuitería 422 para configurar la circuitería 422 para realizar el procedimiento de la invención.

Durante el funcionamiento, un DSP de audio que ha sido configurado para realizar una mejora de voz según la invención (por ejemplo, el sistema 420 de la Fig. 4) está acoplado para recibir una señal de entrada de audio de canal  $N$ , y el DSP realiza típicamente una diversidad de operaciones sobre el audio de entrada (o una versión procesada del mismo) además de (así como) una mejora de la voz. Por ejemplo, el sistema 420 de la Fig. 4 puede ser implementado para realizar otras operaciones (sobre la salida de la circuitería 422) en el subsistema 423 de procesamiento.

Según todavía otro ejemplo, un DSP de audio es operable para llevar a la práctica una realización del procedimiento de la invención después de ser configurado (por ejemplo, programado) para generar una señal de audio de salida en respuesta a una señal de audio de entrada mediante la realización del procedimiento sobre la señal de audio de entrada.

En algunos ejemplos, el sistema de la invención es o incluye un procesador de propósito general acoplado para recibir o para generar datos de entrada indicativos de una señal de audio multicanal. El procesador está programado con software (o firmware) y/o si no configurado (por ejemplo, en respuesta a los datos de control) para realizar cualquiera de entre una diversidad de operaciones sobre los datos de entrada, incluyendo una realización del procedimiento de la invención. El sistema de ordenador de la Fig. 5 es un ejemplo de dicho sistema. El sistema de la Fig. 5 incluye un procesador 501 de propósito general que está programado para realizar cualquiera de entre una diversidad de operaciones sobre los datos de entrada, incluyendo una realización del procedimiento de la invención.

El sistema de ordenador de la Fig. 5 incluye también un dispositivo 503 de entrada (por ejemplo, un ratón y/o un teclado) acoplado al procesador 501, un medio 504 de almacenamiento acoplado al procesador 501 y un dispositivo 505 de visualización acoplado al procesador 501. El procesador 501 está programado para implementar el procedimiento inventivo en respuesta a las instrucciones y los datos introducidos por la manipulación del usuario del dispositivo 503 de entrada. El medio 504 de almacenamiento legible por ordenador (por ejemplo, un disco óptico u otro objeto tangible) tiene un código de ordenador almacenado en el mismo que es adecuado para programar el procesador 501 para realizar una realización del procedimiento de la invención. Durante el funcionamiento, el procesador 501 ejecuta el código de ordenador para procesar datos indicativos de una señal de entrada de audio multicanal según la invención para generar datos de salida indicativos de una señal de salida de audio multicanal.

5 El sistema de la Fig. 1, 1A, 2, 2A o 3 descrito anteriormente podría ser implementado en el procesador 501 de propósito general, en el que los canales 101, 102 y 103 de señal de entrada son datos indicativos de los canales de entrada de audio central (voz) e izquierdo y derecho (sin voz) (por ejemplo, de una señal de sonido envolvente), y en el que los canales 118 y 119 de señal de salida son datos de salida indicativos de canales de salida de audio izquierdo y derecho enfatizados con voz enfatizada (por ejemplo, de una señal de sonido envolvente con voz mejorada). Un convertidor de digital a analógico (DAC) convencional podría operar sobre los datos de salida para generar versiones analógicas de las señales de canal de audio de salida para su reproducción por altavoces físicos.

10 Los aspectos de la invención son un sistema de ordenador programado para realizar cualquier realización del procedimiento de la invención, y un medio legible por ordenador que almacena código legible por ordenador para implementar cualquier realización del procedimiento de la invención.

Aunque en la presente memoria se han descrito realizaciones específicas de la presente invención y aplicaciones de la invención, será evidente para las personas con conocimientos en la materia que son posibles muchas variaciones en las realizaciones y las aplicaciones descritas en la presente memoria sin apartarse del alcance de la invención descrita. y reivindicada en la presente memoria.

15

**REIVINDICACIONES**

1. Procedimiento para filtrar una señal de audio multicanal que tiene un canal de voz y al menos un canal sin voz, para mejorar la inteligibilidad de la voz determinada por la señal, en el que dicho procedimiento incluye las etapas de:
- 5 (a) determinar al menos un valor de control de atenuación indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por al menos un canal sin voz de la señal de audio multicanal; y
- (b) atenuar al menos un canal sin voz de la señal de audio multicanal en respuesta al por lo menos un valor de control de atenuación.
- 10 2. Procedimiento según la reivindicación 1, en el que cada valor de control de atenuación determinado en la etapa (a) es indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por un canal sin voz de la señal de audio, y la etapa (b) incluye una etapa de atenuar dicho canal sin voz en respuesta a cada uno de dichos valores de control de atenuación.
- 15 3. Procedimiento según la reivindicación 1, en el que la etapa (a) incluye una etapa de derivar un canal sin voz derivado a partir del por lo menos un canal sin voz de la señal de audio, y el al menos un valor de control de atenuación es indicativo de una medida de similitud entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por el canal sin voz derivado.
- 20 4. Procedimiento según la reivindicación 3, en el que el canal sin voz derivado es derivado combinando un primer canal sin voz de la señal de audio multicanal y un segundo canal sin voz de la señal de audio multicanal.
5. Procedimiento según la reivindicación 3, en el que la señal de audio multicanal tiene al menos dos canales sin voz, y la etapa (b) incluye la etapa de atenuar algunos, pero no todos, de los canales sin voz en respuesta al por lo menos un valor de control de atenuación.
- 25 6. Procedimiento según la reivindicación 3, en el que la señal de audio multicanal tiene al menos dos canales sin voz, y la etapa (b) incluye la etapa de atenuar todos los canales sin voz en respuesta al por lo menos un valor de control de atenuación.
7. Procedimiento según la reivindicación 1, en el que la etapa (b) comprende escalar una señal de control de atenuación no procesada para el canal sin voz en respuesta al por lo menos un valor de control de atenuación.
- 30 8. Procedimiento según la reivindicación 1, en el que la etapa (a) incluye la etapa de generar una señal de control de atenuación indicativa de una secuencia de valores de control de atenuación, en el que cada uno de los valores de control de atenuación es indicativo de una medida de similitud en un tiempo diferente entre el contenido relacionado con la voz determinado por el canal de voz y el contenido relacionado con la voz determinado por el al menos un canal sin voz de la señal de audio multicanal, y la etapa (b) incluye las etapas de:
- 35 escalar una señal de control de ganancia de atenuación en respuesta a la señal de control de atenuación para generar una señal de control de ganancia escalada; y
- aplicar la señal de control de ganancia escalada para atenuar al menos un canal sin voz de la señal de audio multicanal.
- 40 9. Procedimiento según la reivindicación 8, en el que la etapa (a) incluye una etapa de comparar una primera secuencia de características relacionadas con la voz, indicativas del contenido relacionado con la voz determinada por el canal de voz, con una segunda secuencia de características relacionadas con la voz indicativas del contenido relacionado con la voz determinada por el al menos un canal sin voz de la señal de audio multicanal para generar la señal de control de atenuación, y cada uno de los valores de control de atenuación indicados por la señal de control de atenuación es indicativo de una medida de similitud en un tiempo diferente entre la primera secuencia de características relacionadas con la voz y la segunda secuencia de características relacionadas con la voz.
- 45 10. Procedimiento según la reivindicación 1, en el que cada uno de dichos valores de control de atenuación está relacionado monotónicamente con la probabilidad de que el al menos un canal sin voz de la señal de audio multicanal sea indicativo de contenido mejorador de voz, en el que dicho contenido mejorador de voz comprende contenido que mejora la inteligibilidad u otra cualidad percibida del contenido de voz determinado por el canal de voz.
- 50

11. Procedimiento para filtrar una señal de audio multicanal que tiene un canal de voz y al menos un canal sin voz, para mejorar la inteligibilidad de la voz determinada por la señal, en el que dicho procedimiento incluye las etapas de:
- 5 (a) comparar una característica del canal de voz y una característica del canal sin voz para generar al menos un valor de atenuación para controlar la atenuación del canal sin voz con relación al canal de voz; y
- (b) ajustar el al menos un valor de atenuación en respuesta al por lo menos un valor de probabilidad de mejora de voz para generar al menos un valor de atenuación ajustado para controlar la atenuación del canal sin voz con relación al canal de voz.
12. Procedimiento según la reivindicación 11, en el que el al menos un valor de probabilidad de mejora de voz es una secuencia de valores de comparación, y el procedimiento incluye una etapa de:
- 10 determinar la secuencia de valores de comparación comparando una primera secuencia de características relacionadas con la voz, indicativa del contenido relacionado con la voz determinada por el canal de voz, con una segunda secuencia de características relacionadas con la voz, indicativa del contenido relacionado con la voz determinada por el canal sin voz, en el que cada uno de los valores de comparación es una medida de similitud en un tiempo diferente entre la primera secuencia de características relacionadas con la voz y la segunda secuencia de características relacionadas con la voz.
- 15 13. Procedimiento según la reivindicación 11, en el que cada uno de dichos valores de atenuación generado en la etapa (a) es un primer factor indicativo de una cantidad de atenuación del canal sin voz necesario para limitar la relación de la potencia de señal en el canal sin voz a la potencia de la señal en el canal de voz de manera que exceda un umbral predeterminado, escalado por un segundo factor relacionado monotónicamente con la probabilidad de que el canal de voz sea indicativo de voz.
- 20 14. Procedimiento según la reivindicación 11, en el que cada uno de dichos valores de atenuación generado en la etapa (a) es un primer factor indicativo de una cantidad de atenuación del canal sin voz suficiente para causar que la inteligibilidad predicha de la voz determinada por el canal de voz en presencia del contenido determinado por el canal sin voz exceda un valor de umbral predeterminado, escalado por un segundo factor relacionado monotónicamente con la probabilidad de que el canal de voz sea indicativo de voz.
- 25 15. Procedimiento según la reivindicación 11, en el que la generación de cada uno de dichos valores de atenuación en la etapa (a) incluye las etapas de:
- 30 determinar un espectro de potencia indicativo de la potencia como una función de la frecuencia del canal de voz y un segundo espectro de potencia indicativo de la potencia como una función de la frecuencia del canal sin voz, y
- realizar una determinación en el dominio de la frecuencia del valor de atenuación en respuesta al espectro de potencia y al segundo espectro de potencia.
- 35 16. Medio de almacenamiento legible por ordenador que comprende instrucciones, que cuando son ejecutadas con uno o más procesadores, controlan el uno o más procesadores para realizar el procedimiento descrito en cualquiera de las reivindicaciones 1-15.
17. Sistema configurado para realizar un procedimiento según cualquiera de las reivindicaciones 1-15.

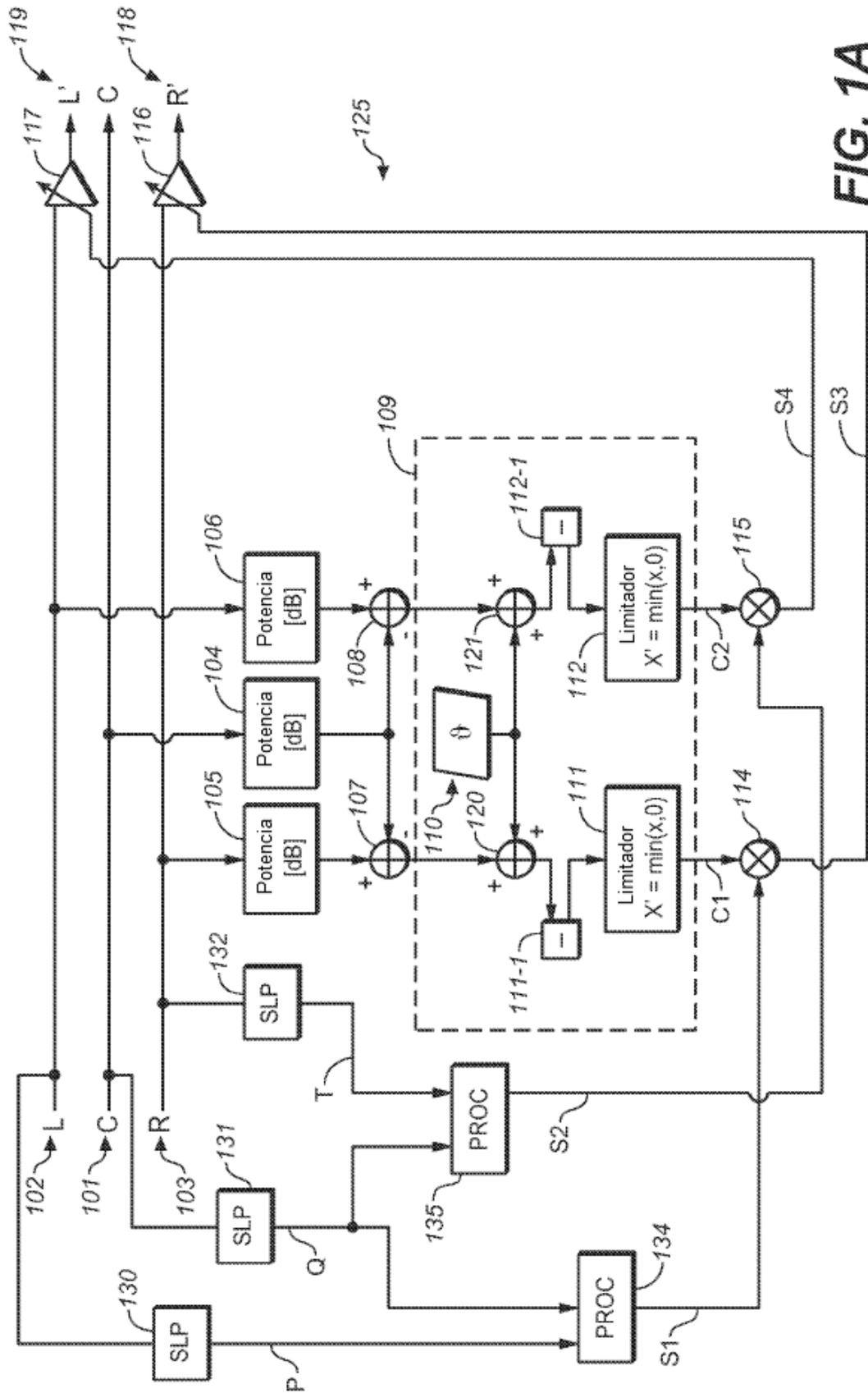


FIG. 1A

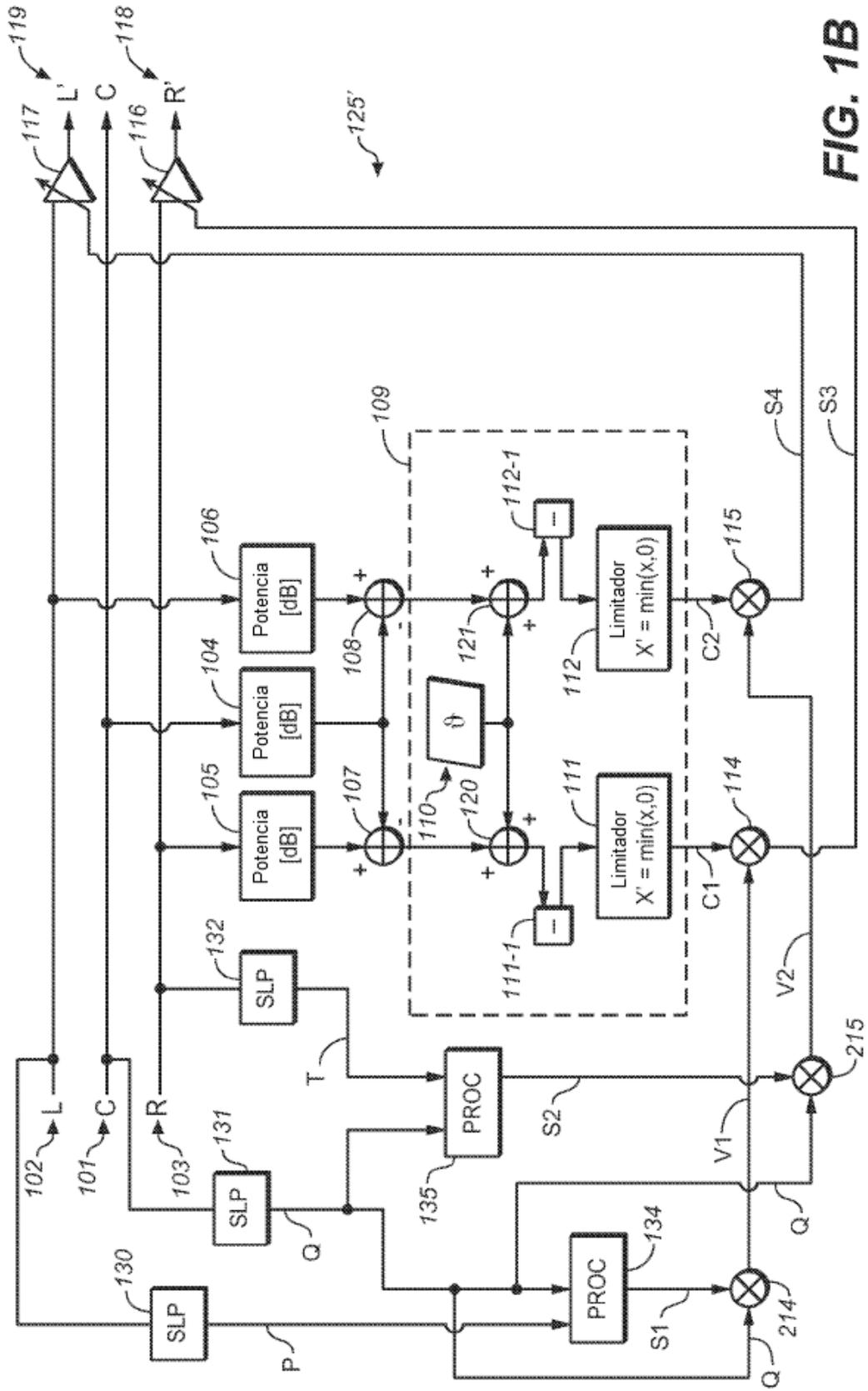


FIG. 1B

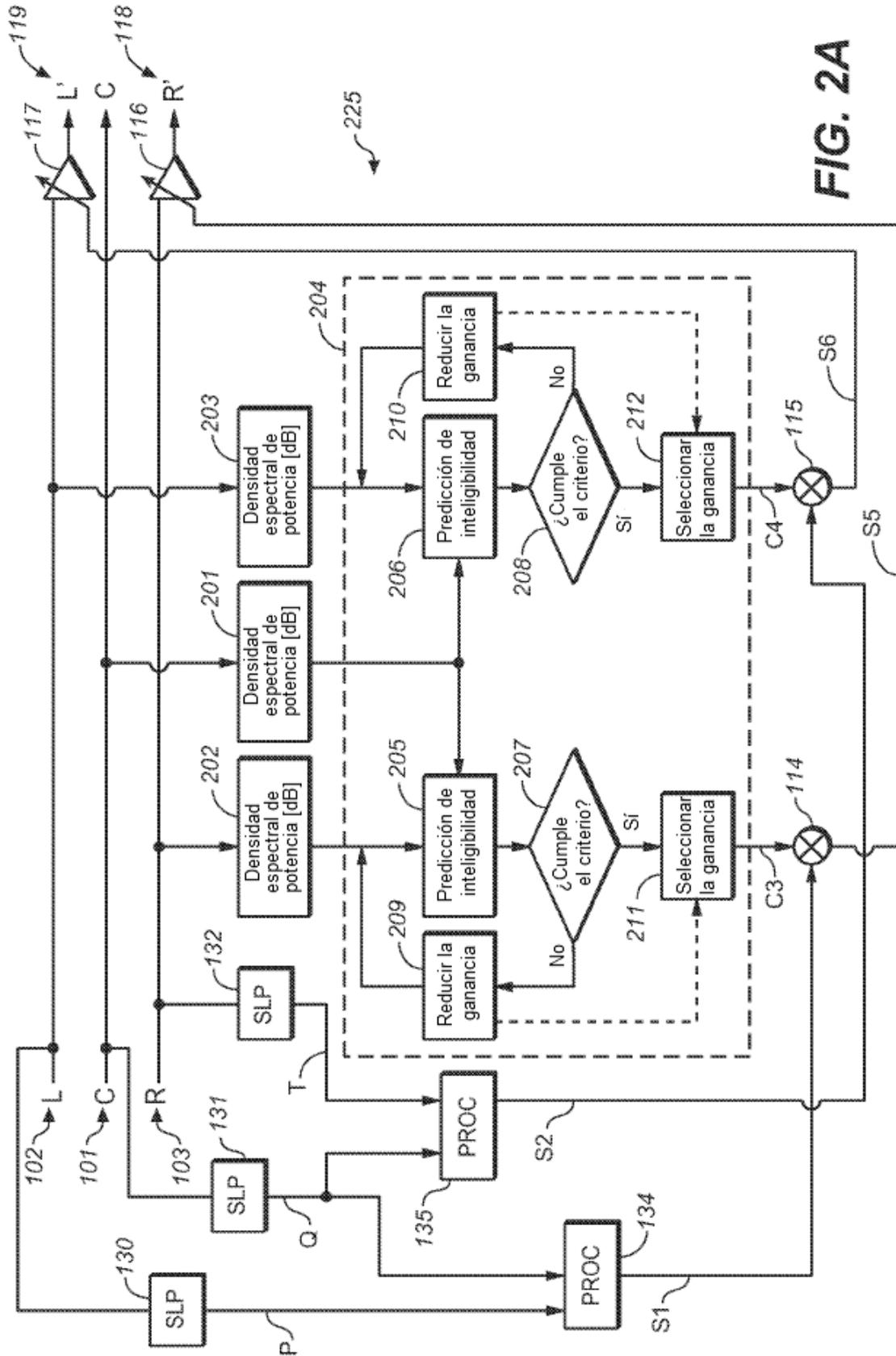


FIG. 2A

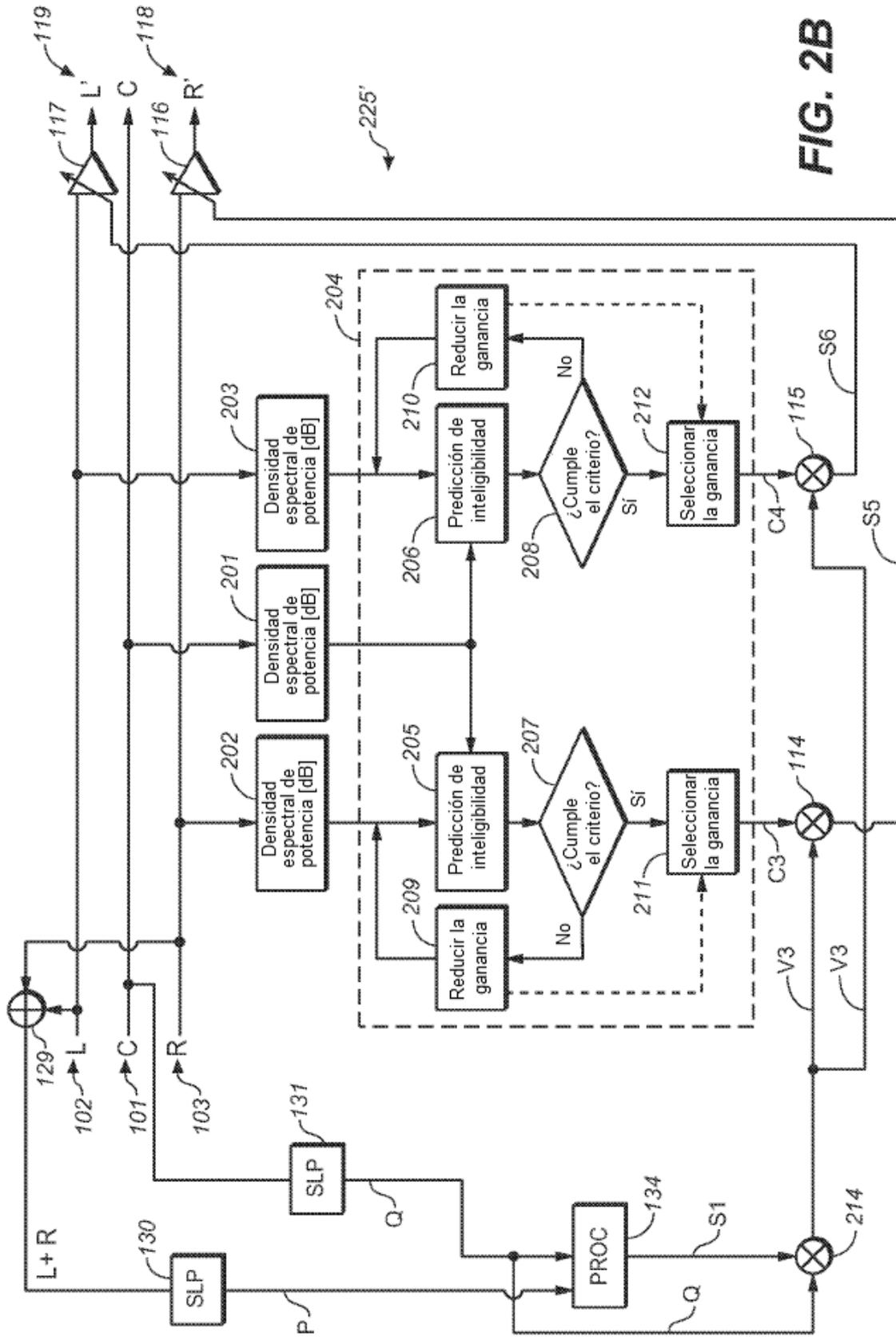


FIG. 2B

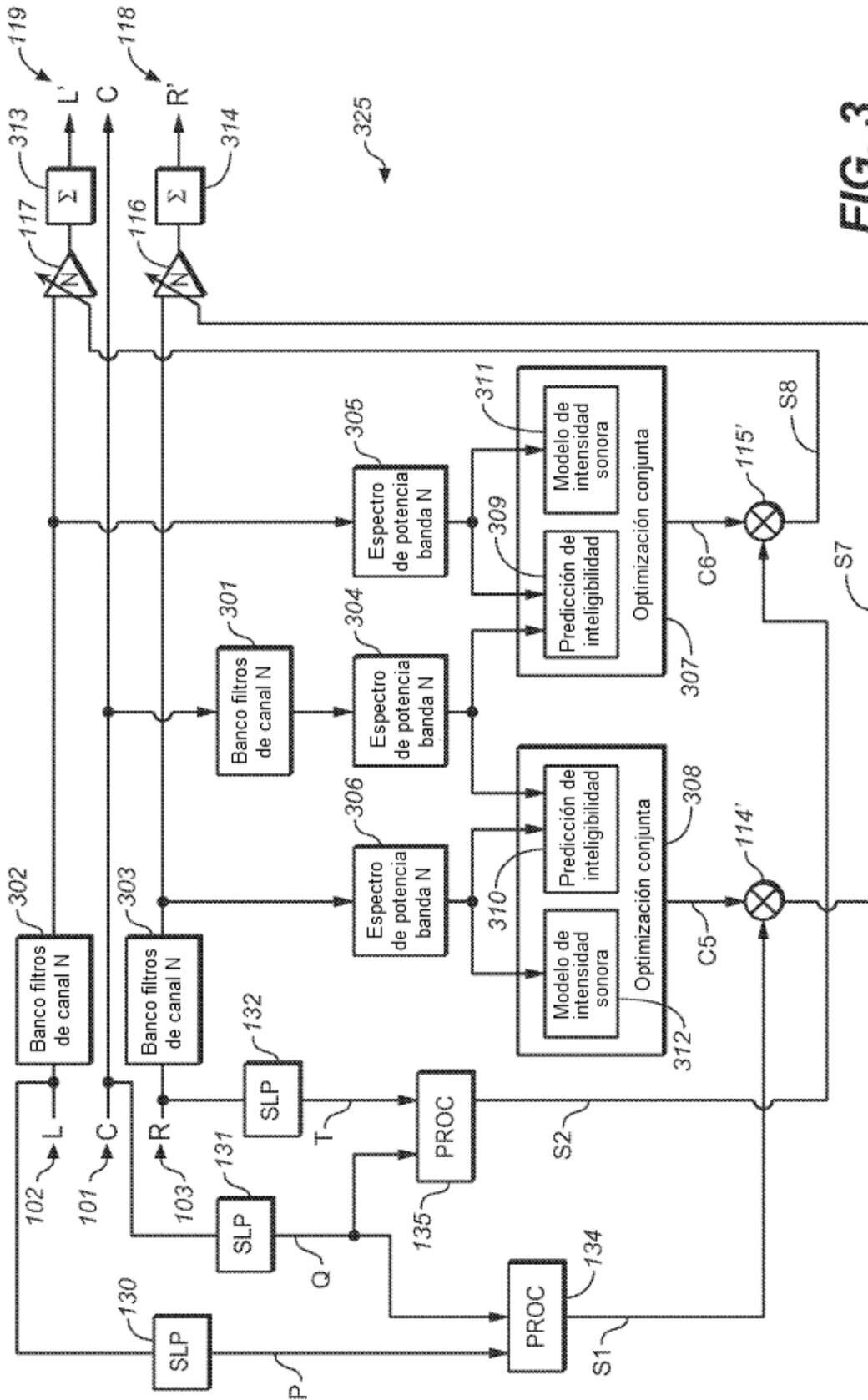
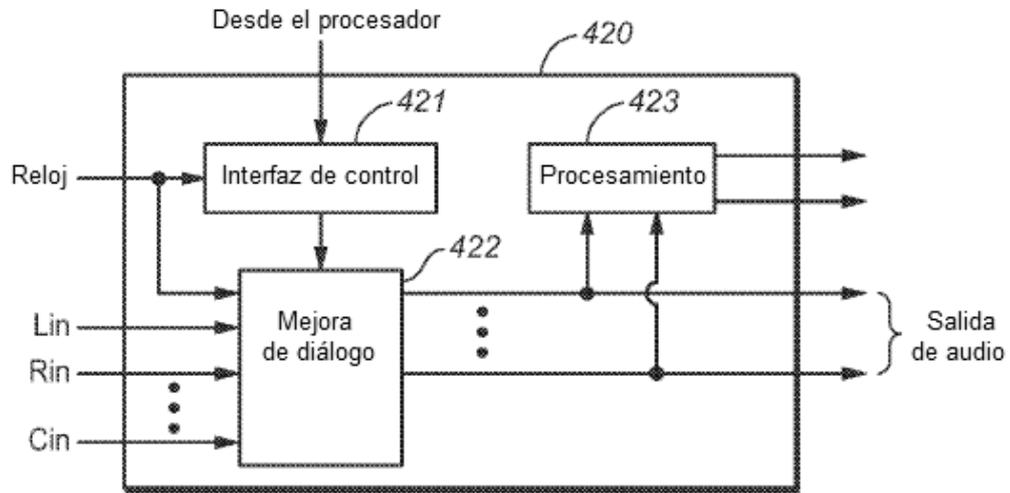
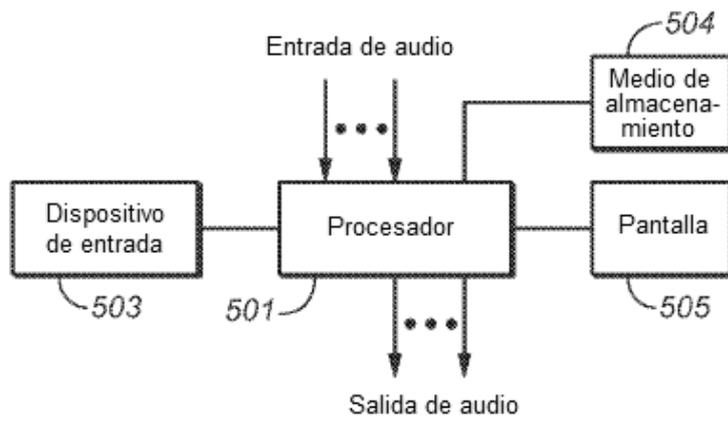


FIG. 3



**FIG. 4**



**FIG. 5**