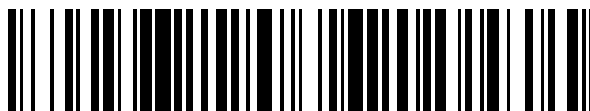


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 788 682**

51 Int. Cl.:

G10H 1/40 (2006.01)

G06F 17/16 (2006.01)

G10H 1/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.03.2017 PCT/EP2017/056257**

87 Fecha y número de publicación internacional: **21.09.2017 WO17158102**

96 Fecha de presentación y número de la solicitud europea: **16.03.2017 E 17711161 (4)**

97 Fecha y número de publicación de la concesión europea: **25.03.2020 EP 3430612**

54 Título: **Aparato y procedimiento para separación de sonido residual percusivo armónico usando un tensor de estructura sobre espectrogramas**

30 Prioridad:

18.03.2016 EP 16161251

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

22.10.2020

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**NIEDERMEIER, ANDREAS;
FÜG, RICHARD;
DISCH, SASCHA;
MÜLLER, MEINARD y
DRIEDGER, JONATHAN**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 788 682 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

Aparato y procedimiento para separación de sonido residual percusivo armónico usando un tensor de estructura sobre espectrogramas

5

[0001] La presente invención se refiere al procesamiento de señales de audio y, en particular, a un aparato y procedimiento para la separación de sonido residual percusivo armónico usando el tensor de estructura sobre espectrogramas.

10 **[0002]** Poder separar un sonido en su componente armónico y percusivo es una etapa de preprocesamiento efectiva para muchas aplicaciones.

[0003] Aunque la "Separación (Residual) Percusiva Armónica" es un término común, es engañoso puesto que implica una estructura armónica sinusoidal que tiene una frecuencia de un entero múltiple de la frecuencia fundamental. 15 Aun cuando el término correcto debería ser "Separación (Residual) Percusiva Tonal", el término y "armónica" en lugar de "tonal" es usado en lo sucesivo para facilitar la comprensión.

[0004] El uso del componente percusivo separado de un registro de música por ejemplo puede llevar a mejorar la calidad por seguimiento de ritmo (véase [1]), análisis de ritmo y transcripción de instrumentos de ritmo. El 20 componente armónico separado es adecuado para la transcripción de instrumentos afinados y detección de acordes (véase [3]). Además, la separación percusiva armónica puede ser usada para propósitos de remezclado como la modificación de la relación de nivel entre ambos componentes de la señal (véase [4]), lo cual conduce a una percepción de sonido total "más suave" o "más fuerte".

25 **[0005]** Algunos procedimientos para la separación de sonido percusivo armónico dependen de la suposición de que los sonidos armónicos tienen una estructura horizontal en el espectrograma de magnitud de la señal de entrada (en la dirección del tiempo), mientras que los sonidos percusivos aparecen como estructuras verticales (en la dirección de la frecuencia). Ono y col. presentó un procedimiento que crea primero espectrogramas mejorados armónicamente/percuscivamente por difusión en el tiempo/dirección de la frecuencia (véase [5]). Al comparar esas 30 representaciones mejoradas posteriormente, podría ser derivada una decisión de si un sonido es armónico o percusivo.

[0006] Un procedimiento similar fue publicado por Fitzgerald, donde los espectrogramas mejorados fueron calculados mediante el uso de filtración mediana en direcciones perpendiculares en lugar de la difusión (véase [6]), lo 35 cual condujo a resultados similares reduciendo a la vez la complejidad computacional.

[0007] Inspirado por el modelo de señal senos+transitorios+ruidos (S+T+N) (véase [7], [8], [9]), un marco de referencia que tiene como propósito describir los componentes de señal respectivos por medio de un conjunto pequeño de parámetros. El procedimiento de Fitzgerald se extendió entonces a la separación residual percusiva armónica (HPR) 40 en [10]. Puesto que las señales de audio con frecuencia consisten en sonidos que no son ni claramente armónicos ni percusivos, este procedimiento captura esos sonidos en un tercer, componente residual. Aunque algunas de esas señales residuales claramente tienen una estructura isotrópica, que no es horizontal ni vertical, (como por ejemplo ruidos), existen sonidos que no tienen una estructura horizontal clara pero que no obstante contienen información tonal y pueden ser percibidos como la parte armónica de un sonido. Un ejemplo son tonos modulados por frecuencia como los que pueden ocurrir en registros de ejecución de violín o voz, donde se dicen que tienen "vibrato". Debido a la 45 estrategia de reconocimiento de las estructuras horizontal o vertical, los procedimientos mencionados anteriormente no siempre son capaces de capturar esos sonidos en su componente armónico.

[0008] Un procedimiento de separación percusiva armónica que se basa en una factorización de matriz no negativa que es capaz de capturar sonidos armónicos con estructuras espectrales no horizontales en el componente 50 armónico fue propuesto en [11]. Sin embargo, este no incluyó un tercer componente residual.

[0009] Resumiendo lo anterior, los procedimientos recientes dependen de la observación de que, en una representación de espectrograma, los sonidos armónicos conducen a estructuras horizontales y los sonidos percusivos conducen a estructuras verticales. Además, esos procedimientos asocian estructuras que no son ni horizontales ni 55 verticales (es decir, sonidos no armónicos, no percusivos) con una categoría residual. Sin embargo, esta suposición no se sostiene para señales como tonos modulados por frecuencia que muestran estructuras espectrales fluctuantes, aunque no obstante contienen información tonal.

[0010] El tensor de estructura, una herramienta usada en el procesamiento de imágenes (véase [12], [13]), es 60 aplicado allí para imágenes en escala de gris para detección de bordes y esquinas (véase [14]) o estimar la orientación de un objeto.

[0011] El tensor de estructura ya ha sido usado para el preprocesamiento y la extracción de características en el procesamiento de audio (véase [15], [16]), representando [15] la técnica anterior más cercana.

65

[0012] El objeto de la presente invención es proporcionar conceptos mejorados para el procesamiento de señales de audio. El objeto de la presente invención es resuelto por un aparato según la reivindicación 1, por un procedimiento según la reivindicación 16 y por un programa informático según la reivindicación 17.

5 **[0013]** Se proporciona un aparato para analizar un espectrograma de magnitud de una señal de audio. El aparato comprende un determinador de cambio de frecuencia que está configurado para determinar un cambio de una frecuencia para cada compartimento de tiempo-frecuencia de una pluralidad de compartimentos de tiempo-frecuencia del espectrograma de magnitud de la señal de audio dependiendo del espectrograma de magnitud de la señal de audio. Además, el aparato comprende un clasificador que está configurado para asignar cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componente de señal de dos o más grupos de componente de señal dependiendo del cambio de la frecuencia determinada por compartimento de tiempo-frecuencia.

[0014] Además, se proporciona un procedimiento para analizar un espectrograma de magnitud de una señal de audio. El procedimiento comprende:

20 La determinación de un cambio de una frecuencia para cada compartimento de tiempo-frecuencia de una pluralidad de compartimentos de tiempo-frecuencia del espectrograma de magnitud de la señal de audio dependiendo del espectrograma de magnitud de la señal de audio. Y:

La asignación de cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señal de dos o más grupos de componentes de señal dependiendo del cambio de la frecuencia determinada del compartimento de tiempo-frecuencia.

25 **[0015]** Además, se proporciona un programa informático, en el que el programa informático está configurado para implementar el procedimiento descrito anteriormente cuando se ejecuta en un ordenador o procesador de señales.

[0016] En lo sucesivo, las realizaciones de la presente invención son descritas con más detalle con referencia a las figuras, en las cuales:

30 La figura 1 ilustra un aparato para analizar un espectrograma de magnitud de una señal de audio según una realización,

35 La figura 2 ilustra un espectrograma de una mezcla de una voz cantante, castañuelas y aplausos con ampliación en la región según una realización, en el que la orientación de las flechas indica la dirección y en el que la longitud de las flechas indica una medida de anisotropía,

40 La figura 3 ilustra un intervalo de valores de orientación/anisotropía calculados mediante el uso del tensor de estructura según una realización,

La figura 4 ilustra una comparación entre el procedimiento HPR-M y HPR-ST para un extracto de una señal de entrada sintética,

45 La figura 5 ilustra un aparato según una realización, en la que el aparato comprende un generador de señales, y

La figura 6 ilustra un aparato según una realización, en la que el aparato comprende uno o más micrófonos para registrar la señal de audio.

[0017] La figura 1 ilustra un aparato para analizar un espectrograma de magnitud de una señal de audio según las realizaciones.

[0018] El aparato comprende un determinador de cambio de frecuencia 110. El determinador de cambio de frecuencia 110 está configurado para determinar un cambio de una frecuencia para cada compartimento de tiempo-frecuencia de una pluralidad de compartimentos de tiempo-frecuencia del espectrograma de magnitud de la señal de audio dependiendo del espectrograma de magnitud de la señal de audio.

[0019] Además, el aparato comprende un clasificador 120. El clasificador 120 está configurado para asignar cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componente de señal de dos o más grupos de componentes de señal dependiendo del cambio de la frecuencia determinada para dicho compartimento de tiempo-frecuencia.

[0020] Según una realización, el determinador de cambio de frecuencia 110 puede, por ejemplo, estar configurado para determinar el cambio de la frecuencia para cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia dependiendo de un ángulo $\alpha(b,k)$ para dicho compartimento de tiempo-frecuencia. El ángulo $\alpha(b,k)$ para dicho compartimento de tiempo-frecuencia depende del espectrograma de magnitud

de la señal de audio.

[0021] En una realización, el determinador de cambio de frecuencia 110 puede, por ejemplo, estar configurado para determinar el cambio de la frecuencia para cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia dependiendo además de una frecuencia de muestreo f_s de la señal de audio, y dependiendo de una longitud N de una ventana de análisis y dependiendo de un tamaño de salto H de la ventana de análisis.

[0022] Según una realización, el determinador de cambio de frecuencia del aparato 110 está configurado para determinar el cambio de frecuencia para cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia dependiendo de la fórmula

$$R(b, k) = \frac{f_s^2}{HN} \cdot \tan(\alpha(b, k))$$

(b, k) indica un compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia, en la que $R(b, k)$ indica el cambio de la frecuencia del compartimento de tiempo-frecuencia (b, k) , en la que b indica el tiempo, en la que k indica la frecuencia, en la que f_s indica la frecuencia de muestreo de la señal de audio, en la que N indica la longitud de la ventana de análisis, en la que H indica el tamaño de salto de la ventana de análisis, y en la que $\alpha(b, k)$ indica el ángulo para el compartimento de tiempo-frecuencia (b, k) , en la que el ángulo $\alpha(b, k)$ depende del espectrograma de magnitud.

[0023] En una realización, el determinador de cambio de frecuencia 110 puede, por ejemplo, estar configurado para determinar una derivada parcial S_b del espectrograma de magnitud S de la señal de audio con respecto a un índice de tiempo. En esa realización, el determinador de cambio de frecuencia 110 puede, por ejemplo, estar configurado para determinar una derivada parcial S_k del espectrograma de magnitud S de la señal de audio con respecto al índice de tiempo.

[0024] Además, en tal realización, el determinador de cambio de frecuencia 110 está configurado para determinar un tensor de estructura $T(b, k)$ para cada compartimento de tiempo-frecuencia (b, k) de la pluralidad de compartimentos de tiempo-frecuencia dependiendo de la derivada parcial S_b del espectrograma de magnitud S de la señal de audio con respecto al índice de tiempo y dependiendo de la derivada parcial S_k del espectrograma de magnitud S de la señal de audio con respecto al índice de frecuencia.

[0025] Además, en tal realización, el determinador de cambio de frecuencia 110 puede, por ejemplo, estar configurado para determinar el ángulo $\alpha(b, k)$ para cada compartimento de tiempo-frecuencia (b, k) de la pluralidad de compartimentos de tiempo-frecuencia dependiendo del tensor de estructura $T(b, k)$ del compartimento de tiempo-frecuencia (b, k) .

[0026] Según una realización, el determinador de cambio de frecuencia 110 puede, por ejemplo, estar configurado para determinar el ángulo $\alpha(b, k)$ para cada compartimento de tiempo-frecuencia (b, k) de la pluralidad de compartimentos de tiempo-frecuencia determinando dos componentes de $v_1(b, k)$ y $v_2(b, k)$ un vector propio $v(b, k)$ del tensor de estructura $(T(b, k))$ del compartimento de tiempo-frecuencia (b, k) , y determinando el ángulo $(\alpha(b, k))$ del compartimento de tiempo-frecuencia (b, k) según

$$\alpha(b, k) = \text{atan} \left(\frac{v_2(b, k)}{v_1(b, k)} \right) \in [-\pi/2; \pi/2]$$

$\alpha(b, k)$ indica el ángulo del compartimento de tiempo-frecuencia (b, k) , en la que b indica el tiempo, en la que k indica la frecuencia, y en la que $\text{atan}()$ indica una función tangente inversa.

[0027] En una realización, el clasificador 120 puede, por ejemplo, estar configurado para determinar una medida de anisotropía para cada compartimento de tiempo-frecuencia (b, k) de la pluralidad de compartimentos de

tiempo-frecuencia dependiendo de al menos una de las fórmulas:

$$\left(\frac{\mu(b,k) - \lambda(b,k)}{\mu(b,k) + \lambda(b,k)} \right)^2$$

5 y

$$\mu(b,k) + \lambda(b,k) \geq e,$$

$\mu(b,k)$ es un primer valor propio $\lambda(b,k)$ es un segundo valor propio del tensor de estructura ($\mathbf{T}(b,k)$) de dicho
10 compartimento de tiempo-frecuencia (b,k) , y $e \in \mathbb{R}^{>0}$.

[0028] En esa realización, el clasificador 120 puede, por ejemplo, estar configurado para asignar cada
compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de
componentes de señal de dos o más grupos de componentes de señal dependiendo además del cambio de la medida
15 de anisotropía.

[0029] Según una realización, el clasificador 120 puede, por ejemplo, estar configurado para determinar la
medida de anisotropía para dicho compartimento de tiempo-frecuencia (b,k) dependiendo de la fórmula:

$$C(b,k) = \begin{cases} \left(\frac{\mu(b,k) - \lambda(b,k)}{\mu(b,k) + \lambda(b,k)} \right)^2 & , \mu(b,k) + \lambda(b,k) \geq e \\ 0 & , \text{además} \end{cases}$$

20

$C(b,k)$ es la medida de anisotropía dependiendo del compartimento de tiempo-frecuencia (b,k) y en el que el
clasificador 120 está configurado para asignar el compartimento de tiempo-frecuencia (b,k) a un grupo de
componentes residuales de dos o más grupos de componentes de señal, si la medida de anisotropía $C(b,k)$ es menor
25 que un primer valor umbral c , o si el clasificador 120 está configurado para asignar el compartimento de tiempo-
frecuencia (b,k) al grupo de componentes residuales de dos o más grupos de componentes de señal, si la medida de
anisotropía $C(b,k)$ es menor que o igual al primer valor umbral c , en el que $c \in \mathbb{R}^{>0}$.

[0030] En una realización, el clasificador 120 puede, por ejemplo, estar configurado para asignar cada
30 compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de
componentes de señal de dos o más grupos de componentes de señal dependiendo del cambio $R(b,k)$ de la frecuencia
determinada para dicho compartimento de tiempo-frecuencia (b,k) , de modo que el clasificador 120 asigne un
compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de
componentes de señales armónicas de dos o más grupos de componentes de señales dependiendo de si un valor
35 absoluto $|R(b,k)|$ del cambio $R(b,k)$ de la frecuencia determinada para dicho compartimento de tiempo-frecuencia
 (b,k) es menor que un segundo valor umbral r_h , o dependiendo de si el valor absoluto $|R(b,k)|$ del cambio $R(b,k)$ de
la frecuencia determinada para dicho compartimento de tiempo-frecuencia (b,k) es menor que o igual al segundo valor
umbral r_h , en el que $r_h \in \mathbb{R}^{>0}$.

[0031] Según una realización, el clasificador 120 puede, por ejemplo, estar configurado para asignar cada
compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de
componentes de señales de dos o más grupos de componentes de señales dependiendo del cambio $R(b,k)$ de la
frecuencia determinada para dicho compartimento de tiempo-frecuencia (b,k) , de modo que el clasificador 120 asigne
un compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de
45 componentes de señales percusivas de los dos o más grupos de componentes de señales dependiendo de si el valor
absoluto $|R(b,k)|$ del cambio $R(b,k)$ de la frecuencia determinada para dicho compartimento de tiempo-frecuencia
 (b,k) es mayor que un tercer valor umbral r_p , o dependiendo de si el valor absoluto $|R(b,k)|$ del cambio $R(b,k)$ de
la frecuencia determinada para dicho compartimento de tiempo-frecuencia (b,k) es mayor que o igual al tercer valor
umbral r_p , en el que $r_p \in \mathbb{R}^{>0}$.

[0032] En lo sucesivo, se proporciona una descripción detallada de las realizaciones.

[0033] Las realizaciones proporcionan conceptos mejorados para la separación del sonido residual percusivo armónico (HPR) sobre la base del tensor de estructura. Algunas realizaciones capturan sonidos modulados por frecuencia que contienen información tonal en el componente armónico explotando la información acerca de la orientación de estructuras espectrales proporcionadas por el tensor de estructura.

[0034] Algunas realizaciones se basan en el descubrimiento de que una clasificación estricta en horizontal y vertical es inapropiada para esas señales y puede conducir a la fuga de información tonal hacia el componente residual. Las realizaciones se relacionan con un procedimiento novedoso que en su lugar usa el tensor de estructura, una herramienta matemática, para calcular los ángulos de orientación predominantes en el espectrograma de magnitud. Las realizaciones emplean esta información de orientación para distinguir entre componentes de señal armónicos, percusivos y residuales, incluso en el caso de señales moduladas por frecuencia. Finalmente, la efectividad del concepto de las realizaciones es verificada por medio de mediciones de evaluación objetivas, así como ejemplos de audio.

[0035] Además, algunas realizaciones se basan en el descubrimiento de que el tensor de estructura puede ser considerado una caja negra, donde la entrada es una imagen en escala de gris y las salidas son ángulos n por cada píxel correspondiente a la dirección del cambio más baja y una medida de certeza y anisotropía para esta dirección por cada píxel. El tensor de estructura ofrece, adicionalmente la posibilidad de ser más suave, lo cual reduce la influencia del ruido por robustez mejorada. Además, la medida de certeza puede ser usada para determinar la calidad de los ángulos estimados. Un valor bajo de esta medida de certeza indica que el píxel se encuentra en una región de brillantez constante sin ninguna dirección clara.

[0036] Un cambio de frecuencia local puede, por ejemplo, ser extraído de los ángulos obtenidos por el tensor de estructura. De esos ángulos, puede determinarse, si un compartimento de tiempo-frecuencia en el espectrograma pertenece al componente armónico (= cambio de frecuencia local baja) o percusivo (= frecuencia local alta o infinita).

[0037] Se proporcionan realizaciones mejoradas para la clasificación y separación residual percusiva armónica.

[0038] La separación de sonido residual percusivo armónico es una herramienta de preprocesamiento útil para aplicaciones como transcripción de instrumento afinado o extracción de ritmo. En lugar de buscar únicamente estructuras estrictamente horizontales y verticales, algunas realizaciones determinan ángulos de orientación predominantes, así como anisotropía local en el espectrograma usando el tensor de estructura conocido del procesamiento de imágenes.

[0039] En realizaciones, la información proporcionada acerca de la orientación de las estructuras espectrales puede utilizarse entonces para distinguir entre componentes de señal armónicos, percusivos y residuales ajustando los umbrales apropiados, véase la figura 2.

[0040] La figura 2 ilustra un espectrograma de una mezcla de una voz cantante, castañuelas y aplausos con amplificación en la región mostrando adicionalmente la dirección (orientación de las flechas) y medida de anisotropía (longitud de las flechas) obtenidas por el tensor de estructura. El color de las flechas indica si el compartimento de tiempo-frecuencia respectivamente es asignado al componente armónico (áreas 210), al componente percusivo (áreas 230), o al componente residual (áreas 220) sobre la base de la información de orientación y anisotropía.

[0041] Todos los compartimentos que no tienen una velocidad de cambio de frecuencia local ni alta ni baja o una cierta medida que indique una región constante fueron asignados para pertenecer al componente residual. Un ejemplo de esta separación de un espectrograma puede ser observado en la figura 2. Las realizaciones funcionan mejor para señales de audio que contienen sonidos modulados por frecuencia que procedimientos similares que trabajan sobre el espectrograma de magnitud.

[0042] Primero, se describe un concepto de tensor de estructura y este concepto general se extiende para ser aplicable en el contexto del procesamiento de audio.

[0043] En lo sucesivo, las matrices y vectores se escriben con negritas por conveniencia notacional. Además, el operador $(\cdot)$ es usado para indexar un elemento específico. En este caso la matriz o vector se escribe sin negritas para mostrar su uso escalar.

[0044] Primero, se describe el cálculo de un espectrograma según realizaciones. La señal de audio puede, por ejemplo, ser una señal de audio de entrada (discreta).

[0045] El tensor de estructura puede ser aplicado a la representación del espectrograma de una señal de audio de entrada discreta $\mathbf{X} \in \mathbb{R}^M$ con una frecuencia de muestreo de f_s . Para el análisis espectral de $\mathbf{X}$, se utiliza la

transformada de Fourier de tiempo reducido (STFT)

$$X(b, k) := \sum_{n=0}^{N-1} w(n)x(n + Hb) \exp(-i2\pi nk/N)$$

5 donde $X(b, k) \in \mathbb{C}$, b denota el índice de trama, k el índice de frecuencia y $w \in \mathbb{R}^N$ es una función de ventana de longitud N (en otras palabras: N es una longitud de una ventana de análisis). $H \in \mathbb{N}$, $H \leq N$ representa el tamaño del salto análisis de la ventana. Cabe observar que puesto que el espectro STFT tiene una cierta simetría alrededor del punto Nyquist en $\frac{N}{2}$, el procesamiento puede, por ejemplo, ser restringido a $0 \leq k \leq \frac{N}{2}$, puesto que la simetría puede ser reconstruida durante la STFT inversa.

10

[0046] Usando la formula (1) anterior, se puede obtener un espectrograma. Un espectrograma comprende una pluralidad de espectro en el que la pluralidad de espectros sucede cada una a la otra en el tiempo. Un segundo espectro de la pluralidad de espectros sucede a un primer espectro en el tiempo, si existen al menos algunas segundas muestras en el dominio del tiempo que se utilicen para generar el segundo espectro y que no se utilicen para generar el primer espectro, y que sean muestras en el dominio del tiempo que se refieran a un punto en el tiempo posterior que las primeras muestras en el dominio del tiempo que se utilizan para generar el primer espectro. Las ventanas de muestras en el dominio del tiempo usadas para generar espectros muy cercanos en el tiempo pueden, por ejemplo, superponerse.

20 **[0047]**

En realizaciones, la longitud de ventana de análisis N puede, por ejemplo, ser definida como:

$$256 \text{ muestras} \leq N \leq 2048 \text{ muestras}$$

[0048]

En algunas realizaciones la longitud de ventana de análisis puede ser, por ejemplo, 2048. En otras realizaciones, la longitud de ventana de análisis puede ser, por ejemplo, 1024 muestras. En realizaciones adicionales, la longitud de ventana de análisis puede ser, por ejemplo, 768 muestras. En realizaciones aun adicionales, la longitud de ventana de análisis puede ser, por ejemplo, 256 muestras.

[0049]

En realizaciones, el tamaño del salto de análisis H puede estar, por ejemplo, en un intervalo entre el 25% y el 75% de la ventana de análisis. En tales realizaciones:

30

$$0,25 N \leq H \leq 0,75 N.$$

[0050]

De este modo, en tales realizaciones, si la ventana de análisis tiene, por ejemplo, 2048 muestras ($N = 2048$), el tamaño del salto de análisis puede estar, por ejemplo, en el intervalo:

35

$$512 \text{ muestras} \leq H \leq 1.536 \text{ muestras.}$$

[0051]

Si la ventana de análisis tiene, por ejemplo, 256 muestras ($N = 256$), el tamaño del salto de análisis puede estar, por ejemplo, en el intervalo:

40

$$64 \text{ muestras} \leq H \leq 192 \text{ muestras.}$$

[0052]

En realizaciones preferidas, el tamaño del salto de análisis puede ser, por ejemplo, el 50% de la ventana de análisis. Esto corresponde a una superposición de ventana de dos ventanas de análisis subsecuentes del 50%.

45

[0053]

En algunas realizaciones, el tamaño del salto de análisis puede ser, por ejemplo, el 25% de la ventana de análisis. Esto corresponde a una superposición de ventana de dos ventanas de análisis subsecuentes del 75%.

50 **[0054]**

En otras realizaciones, el tamaño del salto de análisis puede ser, por ejemplo, el 75% de la ventana de análisis. Esto corresponde a una superposición de ventana de dos ventanas de análisis subsecuentes del 25%.

[0055]

Cabe observar que los conceptos de la presente invención son aplicables para cualquier tipo de transformación en el dominio del tiempo al dominio espectral, como para la MDCT (Transformada de Coseno Discreta Modificada), MDST (Transformada Sinusoidal Discreta Modificada), DSTFT (Transformada de Fourier de tiempo reducido), etc.

55

[0056]

El espectrograma algorítmico de valor real puede ser calculado, por ejemplo, como:

$$S(b, k) = 20 \log_{10} |X(b, k)|$$

[0057] El espectrograma de magnitud de la señal de audio puede ser referido como $\mathbf{S}$ y un valor del espectrograma de magnitud para una bandeja de tiempo-frecuencia (b, k) puede ser referido como $S(b, k)$.

[0058] En lo sucesivo, se describe el cálculo del tensor de estructura según las realizaciones.

[0059] Para el cálculo del tensor de estructura, son necesarias las derivadas parciales de $\mathbf{S}$. La derivada parcial con respecto al índice de tiempo b es dada por

$$\mathbf{S}_b = \mathbf{S} * \mathbf{d}$$

mientras que la derivada parcial, con respecto al índice de frecuencia k es definida como

$$\mathbf{S}_k = \mathbf{S} * \mathbf{d}^T$$

donde $\mathbf{d}$ es un operador de diferenciación discreta (por ejemplo, para diferencias centrales podría elegirse $\mathbf{d} = [-1, 0, 1] / 2$) y $*$ denota la convolución bidimensional.

[0060] Además, puede definirse:

$$\mathbf{T}_{11} = (\mathbf{S}_b \odot \mathbf{S}_b) * \mathbf{G} \quad (5)$$

$$\mathbf{T}_{21} = \mathbf{T}_{12} = (\mathbf{S}_k \odot \mathbf{S}_b) * \mathbf{G} \quad (6)$$

$$\mathbf{T}_{22} = (\mathbf{S}_k \odot \mathbf{S}_k) * \mathbf{G} \quad (7)$$

donde $\odot$ es la multiplicación matricial de punto, también conocida como el producto Hadamard y $\mathbf{G}$ es un filtro de suavización Gaussiano 2-D que tiene la desviación estándar σ_b en la dirección del índice del tiempo y σ_k en la dirección del índice de frecuencia. El tensor de estructura $\mathbf{T}(b, k)$ es entonces dado por una matriz semidefinida simétrica y positiva de 2×2

$$\mathbf{T}(b, k) = \begin{bmatrix} T_{11}(b, k) & T_{12}(b, k) \\ T_{21}(b, k) & T_{22}(b, k) \end{bmatrix} \quad (8)$$

[0061] El tensor de estructura contiene información sobre la orientación dominante del espectrograma en la posición (b, k) . Cabe observar que en el caso especial donde $\mathbf{G}$ es un escalador, $\mathbf{T}(b, k)$ no contiene más información que el gradiente en esta posición en el espectrograma. Sin embargo, en contraste con el gradiente, el tensor de estructura puede ser suavizado por $\mathbf{G}$ sin efectos de cancelación, lo cual hace esto más robusto contra el ruido.

[0062] Cabe observar, que se define un tensor de estructura $\mathbf{T}(b, k)$ para cada compartimento de tiempo-frecuencia (b, k) de la pluralidad de compartimentos de tiempo-frecuencia. De modo que cuando se considera una pluralidad de compartimentos de tiempo y frecuencia, por ejemplo, los compartimentos de tiempo-frecuencia $(0, 0); (0, 1); (0, 2); \dots (1, 0); (1, 1); (1, 2); \dots$ entonces existe una pluralidad de tensores de estructura $\mathbf{T}(0, 0); \mathbf{T}(0, 1); \mathbf{T}(0, 2); \dots \mathbf{T}(1, 0); \mathbf{T}(1, 1); \mathbf{T}(1, 2); \dots$ Por ejemplo, para cada compartimento de tiempo-frecuencia (b, k) de la pluralidad de compartimentos de tiempo-frecuencia, se determina un tensor de estructura

$\mathbf{T}(b, k)$

[0063] En lo sucesivo, se describe el cálculo de ángulos y una medida de anisotropía según las realizaciones.

5 **[0064]** La información acerca de la orientación para cada compartimento en el espectrograma se obtiene calculando los valores propios $\lambda(b, k)$, $\mu(b, k)$ con $\lambda(b, k) \leq \mu(b, k)$ y los vectores propios correspondientes $\mathbf{v}(b, k) = [v_1(b, k), v_2(b, k)]^T$ y $\mathbf{w}(b, k) = [w_1(b, k), w_2(b, k)]^T$ del tensor de estructura $\mathbf{T}(b, k)$. Cabe observar que $\mathbf{v}(b, k)$, el vector propio correspondiente al valor propio más pequeño $\lambda(b, k)$, está apuntando hacia la dirección del cambio más baja en el espectrograma en el índice (b, k) , mientras que $\mathbf{w}(b, k)$ está apuntando hacia la dirección del cambio más alta. De este modo, el ángulo de orientación de un compartimento específico se puede obtener por

$$\alpha(b, k) = \text{atan} \left(\frac{v_2(b, k)}{v_1(b, k)} \right) \in [-\pi/2; \pi/2] \quad (9)$$

15 $v_1(b, k)$ y $v_2(b, k)$ son los componentes del vector propio $\mathbf{v}(b, k)$.

$\text{atan}()$ indica una función tangente inversa.

20 **[0065]** Además, se puede determinar una medida de anisotropía

$$C(b, k) = \begin{cases} \left(\frac{\mu(b, k) - \lambda(b, k)}{\mu(b, k) + \lambda(b, k)} \right)^2 & , \mu(b, k) + \lambda(b, k) \geq e \\ 0 & , \text{además} \end{cases} \quad (10)$$

con $e \in \mathbb{R}^{>0}$ para cada compartimento. Cabe observar que $C(b, k) \in [0; 1]$. Los valores de $C(b, k)$ cercanos a 1 indican una anisotropía alta del espectrograma en el índice (b, k) , mientras que una constante cercana conduce a valores cercanos a 0. El umbral e , que define un límite sobre lo que será considerado anisotrópico, puede ser elegido para incrementar aún más la robustez contra el ruido.

[0066] El significado físico del ángulo $\alpha(b, k)$ puede ser comprendido considerando una señal continua con un cambio con la frecuencia instantánea Δf durante un intervalo de tiempo Δt . De este modo la velocidad de cambio de frecuencia instantánea R es denotada por

$$R = \frac{\Delta f}{\Delta t} \quad (11)$$

35 **[0067]** Por ejemplo, según las realizaciones, los ángulos (indicados por la dirección de las flechas en la figura 2) obtenidos por el tensor de estructura pueden, por ejemplo, ser traducidos en una velocidad de cambio de frecuencia local

$$R = \frac{\Delta f}{\Delta t} [Hz/s] \quad 11a)$$

40 para cada compartimento de tiempo-frecuencia del espectrograma.

[0068] El cambio de la frecuencia para cada compartimento de tiempo-frecuencia puede ser referido, por ejemplo, como velocidad de cambio de frecuencia instantánea.

[0069] Considerando la velocidad de muestreo, la longitud y el tamaño de salto del análisis de STFT aplicado, puede derivarse una relación entre los ángulos en el espectrograma y la velocidad de cambio de frecuencia instantánea $R(b, k)$ para cada compartimento por

5

$$R(b, k) = \frac{f_s^2}{HN} \cdot \tan(\alpha(b, k)) \quad (12)$$

[0070] También las desviaciones estándar del filtro de suavización $\mathbf{G}$ en el dominio discreto σ_b y σ_k pueden ser convertidas a parámetros físicos continuos σ_t y σ_f por

10

$$\sigma_t = \frac{H}{f_s} \sigma_b, \quad \sigma_f = \frac{f_s}{N} \sigma_k \quad (13)$$

[0071] En lo sucesivo se describe la separación residual percusiva armónica usando el tensor estructural.

15 [0072] La información obtenida vía el tensor estructural puede aplicarse al problema de separación HPR, por ejemplo, para clasificar cada compartimento en el espectrograma como parte de cualquiera del componente armónico, percusivo o residual de la señal de entrada.

20 [0073] Las realizaciones se basan en el descubrimiento de que los compartimentos asignados a los componentes armónicos deberán pertenecer a estructuras horizontales, mientras que los compartimentos que pertenecen a estructuras bastante verticales serán asignados al componente percusivo. Además, los compartimentos que no pertenezcan a ningún tipo de estructura orientada serán asignados al componente residual.

25 [0074] Según las realizaciones, un compartimento (b, k) puede ser asignado, por ejemplo, al componente armónico, si satisface ambas de las dos restricciones siguientes.

[0075] Según realizaciones preferidas, un compartimento (b, k) puede ser asignado, por ejemplo, al componente armónico, si satisface ambas de las dos restricciones siguientes:

30 - La primera restricción puede ser, por ejemplo, que el valor absoluto del ángulo $\alpha(b, k)$ sea menor que (o igual a) un umbral α_h . El umbral α_h puede estar, por ejemplo, en el intervalo $\alpha_h \in [0; \pi/2]$. Esto significa que el compartimento deberá ser parte de alguna estructura espectral que no tenga una pendiente mayor o menor que α_h . De esta manera también pueden ser considerados sonidos modulados por frecuencia que sean parte del componente armónico, dependiendo del parámetro α_h .

35

- La segunda restricción puede ser, por ejemplo, que la medida de anisotropía $C(b, k)$ soporta que la estructura (b, k) sea parte de alguna estructura anisotrópica directa, y por lo tanto excede un segundo umbral c . Cabe observar que para un compartimento dado (b, k) , el ángulo $\alpha(b, k)$ y la medida de anisotropía $C(b, k)$ juntas definen un punto $\mathbb{R}^2$ dado en coordenadas polares.

40

[0076] De manera similar, en las realizaciones, se asigna otro umbral de ángulo α_p para definir cuándo un compartimento deberá ser asignado al componente percusivo (áreas 330 con líneas verticales en la figura 3).

45 [0077] De este modo, según las realizaciones, un compartimento (b, k) puede ser asignado, por ejemplo, al componente percusivo, si satisface la primera de las dos restricciones siguientes.

[0078] Según las realizaciones preferidas, un compartimento (b, k) puede ser asignado, por ejemplo, al componente percusivo, si satisface ambas de las dos restricciones siguientes:

5 - La primera restricción puede ser, por ejemplo, que el valor absoluto del ángulo $\alpha(b, k)$ sea mayor que (o igual a) un umbral α_p . El umbral α_p puede estar, por ejemplo, en el intervalo $\alpha_p \in [0; \pi/2]$. Esto significa, que el compartimento será parte de alguna estructura espectral que no tiene una pendiente mayor o menor que α_p . De este modo también pueden ser considerados sonidos modulados por frecuencia como parte del componente armónico, dependiendo del parámetro α_p .

10 - La segunda restricción puede ser, por ejemplo, que la medida de anisotropía $C(b, k)$ soporte que el compartimento (b, k) sea parte de alguna estructura anisotrópica dirigida, y por lo tanto excede un segundo umbral c . Cabe observar que para un compartimento dado (b, k) , el ángulo $\alpha(b, k)$ y la medida de anisotropía $C(b, k)$ juntas definen un punto $\mathbb{R}^2$ dado en coordenadas polares.

15 **[0079]** Finalmente, en las realizaciones, todos los compartimentos que no sean asignados al componente armónico ni al percusivo pueden ser asignados, por ejemplo, al componente residual.

20 **[0080]** El proceso de asignación descrito anteriormente puede ser expresado definiendo una máscara para el componente armónico M_h , una máscara para el componente percusivo M_p , y una máscara para el componente residual M_r .

[0081] Cabe observar, que en lugar de usar el umbral α_h y el umbral α_p los umbrales se pueden definirse en las realizaciones, por ejemplo, sobre la velocidad de cambio de frecuencia absoluta máxima $r_h, r_p \in \mathbb{R}^{>0}$ con $r_p \geq r_h$ para dar a la elección de los parámetros una mejor interpretación física. Las máscaras son entonces dadas por:

$$M_h(b, k) = \begin{cases} 1 & , |R(b, k)| \leq r_h \wedge C(b, k) > c \\ 0 & , \text{ademas} \end{cases} \quad (14)$$

$$M_p(b, k) = \begin{cases} 1 & , |R(b, k)| > r_p \wedge C(b, k) > c \\ 0 & , \text{ademas} \end{cases} \quad (15)$$

$$M_r(b, k) = 1 - M_h(b, k) - M_p(b, k) \quad (16)$$

[0082] Finalmente, la STFT del componente armónico X_h , el componente percusivo X_p y el componente residual X_r son obtenidos por

$$\mathbf{X}_h = \mathbf{M}_h \odot \mathbf{X} \quad (17)$$

$$\mathbf{X}_p = \mathbf{M}_p \odot \mathbf{X} \quad (18)$$

$$\mathbf{X}_r = \mathbf{M}_r \odot \mathbf{X} \quad (19)$$

[0083] Las señales de tiempo correspondientes pueden ser calculadas a continuación vía STFT inversa.

[0084] La figura 3 ilustra un intervalo de valores de orientación/anisotropía calculados por el tensor de

estructura.

[0085] En particular, la figura 3 describe el subconjunto de todos los puntos que conducen a una asignación al componente armónico. En particular, los valores en las áreas 310 con líneas onduladas conducen a una asignación al

[0086] Los valores en las áreas 330 con líneas verticales conducen a una asignación al componente percusivo.

[0087] Los valores en las áreas 320 que están punteadas conducen a una asignación al componente residual.

[0088] El umbral α_h define la línea 301 en la figura 3, y el umbral α_p define la línea 302 en la figura 3.

[0089] La figura 5 ilustra un aparato según una realización, en la que el aparato comprende un generador de señales 130 que es configurado para generar una señal de salida de audio dependiendo de la asignación de la pluralidad de compartimentos tiempo-frecuencia a los dos o más grupos de componentes de señal.

[0090] Por ejemplo, el generador de señales puede filtrar los diferentes componentes de la señal de audio aplicando diferentes factores de peso o ponderación sobre los valores de magnitud de los compartimentos de tiempo-frecuencia de los grupos de componentes de señal diferentes. Por ejemplo, el grupo de componentes de señal armónica pueden tener un primer factor de peso o ponderación w_h , el grupo de componentes de señal percusiva pueden tener un segundo factor de peso w_p , y el grupo de componente de señal residual puede tener un primer factor de peso w_r , y el valor de la magnitud de cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia puede ser ponderado, por ejemplo, con el factor de peso del grupo de componentes de señal, al que el compartimento de tiempo-frecuencia está asignado.

[0091] Por ejemplo, para enfatizar los componentes de señal armónica, en una realización, donde los factores de peso son multiplicados con valores de magnitud lineal, por ejemplo, $w_h = 1,3$, $w_p = 0,7$, y $w_r = 0,2$

[0092] Por ejemplo, para enfatizar los componentes de señal armónica, en una realización, donde los factores de peso son agregados a los valores de magnitud logarítmica, por ejemplo, $w_h = +0,26$, $w_p = -0,35$ y $w_r = -1,61$

[0093] Por ejemplo, para enfatizar los componentes de señal percusiva, en una realización, donde los factores de peso son multiplicados con los valores de magnitud lineal, por ejemplo, $w_h = 0,7$, $w_p = 1,3$, y $w_r = 0,2$

[0094] Por ejemplo, para enfatizar los componentes de señal percusiva, en una realización, donde los factores de peso son agregados a los valores de magnitud logarítmica, por ejemplo, $w_h = -0,35$, $w_p = +0,26$ y $w_r = -1,61$

[0095] De este modo, el generador de señal 130 está configurado para aplicar un factor de peso sobre el valor de magnitud de cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia para obtener la señal de salida de audio, en la que el factor de peso que es aplicado sobre el compartimento de tiempo-frecuencia depende del grupo de componentes de señal al cual el compartimento de tiempo-frecuencia es asignado.

[0096] En una realización particular de la figura 5, el procesador de señales 130 puede ser, por ejemplo, un mezclador ascendente que está configurado para mezclar de forma ascendente la señal de audio para obtener la señal de salida de audio que comprende dos o más canales de salida de audio. El mezclador ascendente puede estar configurado, por ejemplo, para generar los dos o más canales de salida de audio dependiendo de la asignación de la pluralidad de compartimentos de tiempo-frecuencia a los dos o más grupos de componentes de señal.

[0097] Por ejemplo, los dos o más canales de audio de salida pueden ser generados a partir del filtro de señal de audio de los diferentes componentes de la señal de audio aplicando diferentes factores de peso sobre los compartimentos de tiempo-frecuencia de valores de magnitud de los diferentes grupos de componentes de señal como se describió anteriormente.

[0098] Sin embargo, para generar los canales de audio diferentes pueden utilizarse diferentes pesos para los grupos de componentes de señal que pueden ser, por ejemplo, específicos para cada uno de los canales de salida de audio diferentes.

[0099] Por ejemplo, para un primer canal de salida de audio, los pesos que serán agregados a los valores de magnitud logarítmica pueden ser, por ejemplo, $w_{1h} = +0,26$, $w_{1p} = -0,35$, y $w_{1r} = -1,61$.

[0100] Y para un segundo canal de salida de audio, los pesos que serán agregados a los valores de magnitud logarítmica pueden ser, por ejemplo, $w_{2h} = +0,35$, $w_{2p} = -0,26$ y $w_{2r} = -1,61$.

5

[0101] Por ejemplo, cuando se mezcla de forma ascendente la señal de audio para obtener cinco canales de salida de audio izquierdo frontal, central, derecho, envolvente izquierdo y envolvente derecho:

10 - El factor de peso armónico w_{1h} puede ser mayor para generar los canales de audio de salida izquierdo, central y derecho en comparación con el factor de peso armónico w_{2h} para generar los canales de salida de audio envolvente izquierdo y envolvente derecho.

15 - El factor de peso percusivo w_{1p} puede ser más pequeño para generar los canales de salida de audio izquierdo, central y derecho en comparación con el factor de peso percusivo w_{2p} para generar los canales de salida de audio envolvente izquierdo y envolvente derecho.

[0102] Se pueden utilizar factores de peso individuales para cada canal de salida de audio que se genere.

20 **[0103]** La figura 6 ilustra un aparato según una realización, en la que el aparato comprende uno o más micrófonos 171, 172 para registrar la señal de audio.

[0104] En la figura 6, el primer micrófono 171 registra un primer canal de audio de la señal de audio. El segundo micrófono opcional 172 registra un segundo canal de audio opcional de la señal de audio.

25 **[0105]** Además, el aparato de la figura 6 comprende además un generador de espectrograma de magnitud 180 para generar el espectrograma de magnitud de la señal de audio a partir de la señal de audio que comprende el primer canal de audio y comprende opcionalmente el segundo canal de audio opcional. La generación de un espectrograma de magnitud a partir de una señal de audio es un concepto bien conocido por el experto en la materia.

30 **[0106]** En lo sucesivo, se considera la evaluación de las realizaciones.

[0107] Para mostrar la efectividad de las realizaciones en la captura de sonidos modulados por frecuencia en el componente armónico, el procedimiento HPR basado en el tensor de estructura (HPR-ST) según las realizaciones es comparado con el procedimiento no iterativo sobre la base de la filtración mediana presentada en [10] (HPR-M).
35 Adicionalmente, se calculan también las métricas para los resultados de separación con máscaras binarias ideales (IBM) que sirvieron como referencia para la calidad de separación máxima alcanzable.

[0108] Considerando un parámetro de prueba del sistema, para, ambas de HPR-ST así como HPR-M, se eligió que los parámetros de STFT fueran $f_s = 22050\text{Hz}$, $N = 1024$ y $H = 256$, usando una ventana sinusoidal w . Los parámetros de separación para HPR-M fueron elegidos como en los experimentos efectuados en [10]. Según las realizaciones, el tensor de estructura se calcula usando un operador diferencial, por ejemplo, el operador Scharr [17] como operador de diferenciación discreto d . La suavización se efectuó usando un filtro Gaussiano isotrópico 9×9 con

las desviaciones estándar $\sigma_b = \sigma_k = 1.4$ lo cual conduce a $\sigma_t \approx 16\text{ms}$ y $\sigma_f \approx 30\text{Hz}$. Finalmente, los umbrales para la separación se fijaron como $e = 20$, $c = 0.2$ y $r_h = r_p = 10000\text{Hz} / s$.

45

[0109] Cabe observar que, mediante la elección de r_h y r_p según las realizaciones, se asignan estructuras aún muy inclinadas en el espectrograma en el componente armónico. Las realizaciones emplean las observaciones acerca de sonidos con vibrato del mundo real como por ejemplo se muestra en la figura 2. Allí, se puede observar que en algunos casos el vibrato en la voz cantante tiene una velocidad de cambio de frecuencia instantánea muy alta.
50 Además, cabe observar que mediante la elección de $r_h = r_p$, una asignación de un compartimento en el espectrograma al componente residual depende puramente de su medida de anisotropía.

[0110] La efectividad de HPR-ST según las realizaciones se evaluó comparando este con el procedimiento HPR-M basado en la filtración mediana del estado de la técnica presentado en [10] por medio de medidas de
55 evaluación objetivas, así como ejemplos de audio.

[0111] Para comparar el comportamiento de HPR-ST según las realizaciones y HPR-M del estado de la técnica cuando son aplicados a señales que contienen sonidos modulados por frecuencia para obtener resultados objetivos, se generaron dos elementos de prueba.

60

[0112] El elemento de prueba 1 consiste en la superposición de sonidos puramente sintéticos. La fuente

armónica elegida fue un tono de vibrato con una frecuencia fundamental de 1.000Hz, una frecuencia de vibrato de 3Hz, una extensión de vibrato de 50Hz y 4 sobretonos. Para la fuente percusiva se utilizan varios impulsos, mientras que el ruido blanco no representa la fuente residual ni armónica ni percusiva.

5 **[0113]** El elemento de prueba 2 se generó superponiendo señales del mundo real de voz cantante con vibrato (armónico), castañuelas (percusivo), y aplausos (ni armónico ni percusivo).

[0114] Interpretando la separación HPR de esos elementos como un problema de separación de fuente u origen, se han calculado métricas de evaluación de separación de fuente estándar (relación de fuente a distorsión SDR, relación de fuente a interferencia SIR, y relaciones de fuentes a artefactos SAR, como se introduce en [18]) para los resultados de separación de ambos procedimientos. Los resultados se muestran en la tabla 1.

[0115] La Tabla 1 describe las medidas de la evaluación objetiva, donde todos los valores se dan en dB:

15

		SDR			SIR			SAR		
		IBM	HPR-M	HPR-ST	IBM	HPR-M	HPR-ST	IBM	HPR-M	HPR-ST
Elemento 1	Vibrato	29,43	11,51	21,25	34,26	27,94	30,01	31,16	11,61	21,88
	Impulsos	8,56	-10,33	-1,47	20,31	-7,96	12,03	8,90	2,02	-1,00
	Ruido	8,49	-13,53	2,58	24,70	-11,99	14,12	8,61	3,97	3,06
Elemento 2	Voz	14,82	6,48	9,18	22,75	20,83	15,61	15,60	6,68	10,42
	Castañuelas	8,48	3,79	2,37	21,59	16,09	19,96	8,73	4,16	2,56
	Aplausos	7,39	-2,03	-0,37	20,31	1,11	6,34	7,66	3,33	1,58

(Tabla 1)

[0116] Para el elemento 1 HPR-ST produce una SDR de 21,25dB para el tono de vibrato, y por lo tanto está más cercano al resultado de separación óptimo de IBM (29,43dB) que del resultado de separación de HPR-M (11,51dB). Esto indica que el HPR-ST mejora tras capturar este sonido modulado por frecuencia en el componente armónico en comparación con HPRM. Esto también se muestra en la figura 4.

[0117] La figura 4 ilustra una comparación entre los procedimientos HPR-M y HPR-ST para un extracto de la señal de entrada sintética (elemento 1). Para una mejor visibilidad los espectrogramas se calcularon con parámetros de STFT diferentes que los usados por los algoritmos de separación.

[0118] La figura 4(a) ilustra la frecuencia de la señal de entrada con respecto al tiempo. En la figura 4, se trazan los espectrogramas de los componentes armónicos y la suma de los componentes percusivo y residual para ambos procedimientos. Puede observarse que para el HPR-M las pendientes pronunciadas de tono de vibrato se fugaron hacia el componente residual (figura 4(b) y (c)), mientras que el HPR-ST (figura 4(d) y 4(e)) produce una buena separación. Esto también explica los valores de SIR muy bajos de HPRM para el componente residual en comparación con HPR-ST (-11,99dB contra 14,12dB).

[0119] Cabe observar que el valor de SIR alto de HPR-M del componente armónico únicamente refleja que únicamente existen pocos sonidos interferentes de los otros componentes, no que el sonido del vibrato es bien capturado como un todo. En general la mayoría de las observaciones para el elemento 1 son menos pronunciadas, pero también válidas para la mezcla de sonidos del mundo real en el elemento 2. Para este elemento, el valor de SIR de HPR-M para la voz excede aún el valor de SIR de HPR-ST (20,83dB contra 15,61dB). De nuevo, el valor de SIR bajo para los aplausos apoya que porciones del vibrato en las voces se fuguen hacia el componente residual por el HPR-M (1,11dB) mientras que el componente residual del HPR-ST contiene menos sonidos interferentes (6,34dB). Esto indica que las realizaciones fueron capaces de capturar las estructuras moduladas por frecuencia de las voces mucho mejor que el HPR-M.

[0120] Resumiendo, los resultados, para señales que contienen tonos modulados por frecuencia, el concepto del HPR-ST de realizaciones proporciona resultados de separación mucho mejores en comparación con el HPR-M.

[0121] Algunas realizaciones emplean el tensor de estructura para la detección de voz cantante. (La detección de voz cantante según el estado de la técnica se describe en [2]).

[0122] Aunque algunos aspectos se han descrito en el contexto de un aparato, está claro que esos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. De manera análoga, los aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un bloque o elemento o característica de un aparato correspondiente. Algunas o todas las etapas de procedimiento pueden ser ejecutadas por (o usando) un aparato de hardware como, por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, una o más de las etapas de procedimiento más importantes pueden ser ejecutadas por ese aparato.

[0123] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software o al menos parcialmente en hardware o al menos parcialmente en software. La implementación puede ser efectuada usando un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tenga señales de control legibles electrónicamente almacenadas en él, que puedan cooperar (o sean capaces de cooperar) con un sistema informático programable de modo que se efectúe el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

[0124] Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control legibles electrónicamente, las cuales son capaces de cooperar con un sistema informático programable, de modo que uno de los procedimientos descritos en esta invención pueda ser efectuado.

[0125] Generalmente, las realizaciones de la presente invención pueden ser implementadas como un producto de programa informático con un código de programa, operando el código de programa para efectuar uno de los procedimientos cuando el producto de programa informático sea ejecutado en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un soporte legible por una máquina.

[0126] Otras realizaciones comprenden el programa informático para efectuar uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por una máquina.

[0127] En otras palabras, una realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para efectuar uno de los procedimientos descritos en esta invención, donde el programa informático se ejecuta en un ordenador.

[0128] Una realización adicional de los procedimientos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, registrado en él, el programa informático para efectuar uno de los procedimientos descritos en esta invención. El soporte de datos, el medio de almacenamiento digital o el medio registrado son típicamente tangibles y/o no transitorios.

[0129] Una realización adicional del procedimiento de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para efectuar uno de los procedimientos descritos en esta invención. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

[0130] Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado para o adaptado para efectuar uno de los procedimientos descritos en esta invención.

[0131] Una realización adicional comprende un ordenador que tiene instalado en él un programa informático para efectuar uno de los procedimientos descritos en esta invención.

[0132] Una realización adicional según la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, electrónicamente u ópticamente) un programa informático para efectuar uno de los procedimientos descritos en esta invención a un receptor. El receptor puede, por ejemplo, ser un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir el programa informático al receptor.

[0133] En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programable por campo) para efectuar algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas realizaciones, una matriz de puertas programable por campo puede cooperar con un

microprocesador para efectuar uno de los procedimientos descritos en esta invención. Generalmente, los procedimientos son efectuados preferiblemente por un aparato de hardware.

[0134] El aparato descrito en esta invención puede ser implementado usando un aparato de hardware, o usando un ordenador, o usando una combinación de un aparato de hardware y un ordenador.

[0135] Los procedimientos descritos en esta invención pueden ser efectuados usando un aparato de hardware, o usando un ordenador, o usando una combinación de un aparato de hardware y un ordenador.

[0136] Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Debe comprenderse que modificaciones y variaciones de las disposiciones y detalles descritos en esta invención serán evidentes a otros expertos en la materia. Se pretende, por lo tanto, que sean limitadas únicamente por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en esta invención.

Bibliografía:

[0137]

[1] Aggelos Gkiokas, Vassilios Katsouros, George Carayannis, and Themis Stafylakis, "Music tempo estimation and beat tracking by applying source separation and metrical relations", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2012, pp. 421–424.

[2] Bernhard Lehner, Gerhard Widmer, and Reinhard Sonnleitner, "On the reduction of false positives in singing voice detection", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 7480–7484.

[3] Yushi Ueda, Yuuki Uchiyama, Takuya Nishimoto, Nobutaka Ono, and Shigeki Sagayama, "HMM-based approach for automatic chord detection using refined acoustic features", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, Texas, USA, 2010, pp. 5518–5521.

[4] Nobutaka Ono, Kenichi Miyamoto, Hirokazu Kameoka, and Shigeki Sagayama, "A real-time equalizer of harmonic and percussive components in music signals", in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Philadelphia, Pennsylvania, USA, 2008, pp. 139–144.

[5] Nobutaka Ono, Kenichi Miyamoto, Jonathan LeRoux, Hirokazu Kameoka, and Shigeki Sagayama, "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram", in European Signal Processing Conference, Lausanne, Switzerland, 2008, pp. 240–244.

[6] Derry Fitzgerald, "Harmonic/percussive separation using median filtering", in Proceedings of the International Conference on Digital Audio Effects (DAFX), Graz, Austria, 2010, pp. 246–253.

[7] Scott N. Levine and Julius O. Smith III, "A sines+transients+noise audio representation for data compression and time/pitch scale modulations", in Proceedings of the AES Convention, 1998.

[8] Tony S. Verma and Teresa H.Y. Meng, "An analysis/synthesis tool for transient signals that allows a flexible sines+transients+noise model for audio", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Seattle, Washington, USA, May 1998, pp. 3573–3576.

[9] Laurent Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit", IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 5, pp. 1808–1816, September 2006.

[10] Jonathan Driedger, Meinard Müller, and Sascha Disch, "Extending harmonic-percussive separation of audio signals", in Proceedings of the International Conference on Music Information Retrieval (ISMIR), Taipei, Taiwan, 2014, pp. 611–616.

[11] Jeongsoo Park and Kyogu Lee, "Harmonic-percussive source separation using harmonicity and sparsity constraints", in Proceedings of the International Conference on Music Information Retrieval (ISMIR), Málaga, Spain, 2015, pp. 148–154.

[12] Josef Bigun and Gösta H. Granlund, "Optimal orientation detection of linear symmetry", in Proceedings of the IEEE First International Conference on Computer Vision, London, UK, 1987, pp. 433–438.

[13] Hans Knutsson, "Representing local structure using tensors", in 6th Scandinavian Conference on Image Analysis, Oulu, Finland, 1989, pp. 244–251.

[14] Chris Harris and Mike Stephens, “A combined corner and edge detector”, in Proceedings of the 4th Alvey Vision Conference, Manchester, UK, 1988, pp. 147–151.

5 [15] Rolf Bardeli, “Similarity search in animal sound databases”, IEEE Transactions on Multimedia, vol. 11, no. 1, pp. 68–76, January 2009.

[16] Matthias Zeppelzauer, Angela S. Stöger, and Christian Breiteneder, “Acoustic detection of elephant presence in
10 Data, Barcelona, Spain, 2013, pp4. 3–8.

[17] Hanno Schar, “Optimale Operatoren in der digitalen Bildverarbeitung“, Dissertation, IWR, Fakultät für Physik und Astronomie, Universität Heidelberg, Heidelberg, Germany, 2000.

15 [18] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte, “Performance measurement in blind audio source separation”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 4, pp. 1462–1469, 2006.

REIVINDICACIONES

1. Un aparato para analizar un espectrograma de magnitud de una señal de audio, **caracterizado por** un determinador de cambio de frecuencia (110) que está configurado para determinar un cambio de una frecuencia $R(b, k)$ para cada compartimento de tiempo-frecuencia (b, k) de una pluralidad de compartimentos de tiempo-frecuencia del espectrograma de magnitud de la señal de audio dependiendo del espectrograma de magnitud de la señal de audio, y un clasificador (120) que está configurado para asignar cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales de dos o más grupos de componentes de señales dependiendo del cambio de la frecuencia determinada para dicho compartimento de tiempo-frecuencia.

2. Un aparato según la reivindicación 1, en el que el determinador de cambio de frecuencia (110) está configurado para determinar el cambio de la frecuencia para cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia dependiendo de un ángulo $(\alpha(b, k))$ del compartimento de tiempo-frecuencia, en el que el ángulo $(\alpha(b, k))$ de dicho compartimento de tiempo-frecuencia depende del espectrograma de magnitud de la señal de audio.

3. Un aparato según la reivindicación 2, en el que el determinador de cambio de frecuencia (110) está configurado para determinar el cambio de la frecuencia para cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia dependiendo además de una frecuencia de muestreo (f_s) de la señal de audio, y dependiendo de una longitud (N) de una ventana de análisis y dependiendo de un tamaño de salto (H) de la ventana de análisis.

4. Un aparato según la reivindicación 3, en el que el determinador de cambio de frecuencia del aparato (110) está configurado para determinar el cambio de la frecuencia para cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia dependiendo de la fórmula

$$R(b, k) = \frac{f_s^2}{HN} \cdot \tan(\alpha(b, k))$$

en la que (b, k) indica un compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia, en la que $R(b, k)$ indica el cambio de la frecuencia para dicho compartimento de tiempo-frecuencia (b, k) , en la que b indica el tiempo, en la que k indica la frecuencia, en la que f_s indica la frecuencia de muestreo de la señal de audio, en la que N indica la longitud de la ventana de análisis, en la que H indica el tamaño de salto de la ventana de análisis, y en la que $\alpha(b, k)$ indica el ángulo para dicho compartimento de tiempo-frecuencia (b, k) , en la que el ángulo $\alpha(b, k)$ depende del espectrograma de magnitud.

5. Un aparato según cualquiera de las reivindicaciones 2 a 4, en el que el determinador de cambio de frecuencia (110) está configurado para determinar una derivada parcial (S_b) del espectrograma de magnitud (S) de la señal de audio con respecto a un índice de tiempo, en el que el determinador de cambio de frecuencia (110) está configurado para determinar una derivada parcial (S_k) del espectrograma de magnitud (S) de la señal de audio con respecto al índice de tiempo, y en el que el determinador de cambio de frecuencia (110) está configurado para determinar un tensor de estructura $(T(b, k))$ para cada compartimento de tiempo-frecuencia (b, k) de la pluralidad de compartimentos de tiempo-frecuencia dependiendo de la derivada parcial (S_b) del espectrograma de magnitud (S) de la señal de audio con respecto al índice de tiempo y dependiendo de la derivada parcial (S_k) del espectrograma de magnitud (S) de la señal de audio con respecto al índice de frecuencia, y en el que el determinador de cambio de frecuencia (110) está configurado para determinar el ángulo $(\alpha(b, k))$ para cada compartimento de tiempo-frecuencia (b, k) de la pluralidad de compartimentos de tiempo-frecuencia dependiendo del tensor de estructura $(T(b, k))$ del compartimento de tiempo-frecuencia (b, k) .

6. Un aparato según la reivindicación 5,

en el que el determinador de cambio de frecuencia (110) está configurado para determinar el ángulo ($\alpha(b, k)$) para cada compartimento de tiempo-frecuencia ((b, k)) de la pluralidad de compartimentos de tiempo-frecuencia para determinar dos componentes de $v_1(b, k)$ y $v_2(b, k)$ un vector propio $\mathbf{v}(b, k)$ del tensor de estructura ($\mathbf{T}(b, k)$) del compartimento de tiempo-frecuencia ((b, k)), y para determinar el ángulo ($\alpha(b, k)$) para dicho compartimento de tiempo-frecuencia ((b, k)) según

$$\alpha(b, k) = \text{atan} \left(\frac{v_2(b, k)}{v_1(b, k)} \right) \in [-\pi/2; \pi/2]$$

en la que $\alpha(b, k)$ indica el ángulo para dicho compartimento de tiempo-frecuencia ((b, k)),

10 en la que b indica el tiempo,
en la que k indica la frecuencia, y
en la que $\text{atan}()$ indica una función tangente inversa.

7. Un aparato según la reivindicación 5 o 6,

15 en el que el clasificador (120) está configurado para determinar una medida de anisotropía para cada compartimento de tiempo-frecuencia ((b, k)) de la pluralidad de compartimentos de tiempo-frecuencia dependiendo de al menos una de las fórmulas:

$$\left(\frac{\mu(b, k) - \lambda(b, k)}{\mu(b, k) + \lambda(b, k)} \right)^2$$

20 y

$$\mu(b, k) + \lambda(b, k) \geq e$$

25 en la que $\mu(b, k)$ es un primer valor propio $\lambda(b, k)$ es un segundo valor propio del tensor de estructura ($\mathbf{T}(b, k)$) del compartimento de tiempo-frecuencia ((b, k)), y $e \in \mathbb{R}^{>0}$,

en la que el clasificador (120) está configurado para asignar cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales de dos o más grupos de componentes de señales dependiendo además del cambio de la medida de anisotropía.

30

8. Un aparato según la reivindicación 7,

en el que el clasificador está configurado para determinar la medida de anisotropía del compartimento de tiempo-frecuencia ((b, k)) dependiendo de la fórmula:

35

$$C(b, k) = \begin{cases} \left(\frac{\mu(b, k) - \lambda(b, k)}{\mu(b, k) + \lambda(b, k)} \right)^2 & , \mu(b, k) + \lambda(b, k) \geq e \\ 0 & , \text{además} \end{cases}$$

en la que $C(b, k)$ es la medida de anisotropía dependiendo del compartimento de tiempo-frecuencia ((b, k)), y

40 en la que el clasificador (120) está configurado para asignar dicho compartimento de tiempo-frecuencia ((b, k)) a un grupo de componentes residuales de los dos o más grupos de componentes de señal, si la medida de anisotropía $C(b, k)$ es menor que un primer valor umbral c , o en la que el clasificador (120) está configurado para asignar dicho compartimento de tiempo-frecuencia ((b, k)) al grupo de componentes residuales de los dos o más grupos de componentes de señal, si la medida de anisotropía $C(b, k)$ es menor que o igual al primer valor umbral

c ,
 en la que $c \in \mathbb{R}^{>0}$.

9. Un aparato según cualquiera de las reivindicaciones anteriores, en el que el clasificador (120) está configurado para asignar cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales de dos o más grupos de componentes de señales dependiendo del cambio ($R(b, k)$) de la frecuencia determinada por dicho compartimento de tiempo-frecuencia ((b, k)), de modo que el clasificador (120) asigne un compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales armónicas de los dos o más grupos de componentes de señales dependiendo de si un valor absoluto ($|R(b, k)|$) del cambio ($R(b, k)$) de la frecuencia determinada para dicho compartimento de tiempo-frecuencia ((b, k)) es menor que un segundo valor umbral r_h , o dependiendo de si el valor absoluto ($|R(b, k)|$) del cambio ($R(b, k)$) de la frecuencia determinada para dicho compartimento de tiempo-frecuencia ((b, k)) es menor que o igual al segundo valor umbral r_h ,
 en el que $r_h \in \mathbb{R}^{>0}$.

10. Un aparato según cualquiera de las reivindicaciones 1 a 8, en el que el clasificador (120) está configurado para asignar cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales de dos o más grupos de componentes de señales dependiendo del cambio ($R(b, k)$) de la frecuencia determinada para dicho compartimento de tiempo-frecuencia ((b, k)), de modo que el clasificador (120) asigne un compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales percusivas de los dos o más grupos de componentes de señales dependiendo de si el valor absoluto ($|R(b, k)|$) del cambio ($R(b, k)$) de la frecuencia determinada por dicho compartimento de tiempo-frecuencia ((b, k)) es mayor que un tercer valor umbral r_p , o dependiendo de si el valor absoluto ($|R(b, k)|$) del cambio ($R(b, k)$) de la frecuencia determinada para dicho compartimento de tiempo-frecuencia ((b, k)) es mayor que o igual a un tercer valor umbral r_p ,
 en el que $r_p \in \mathbb{R}^{>0}$.

11. Un aparato según la reivindicación 9, en el que el clasificador (120) está configurado para asignar cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales de dos o más grupos de componentes de señales dependiendo del cambio ($R(b, k)$) de la frecuencia determinada para dicho compartimento de tiempo-frecuencia ((b, k)), de modo que el clasificador asigne un compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales percusivas de los dos o más grupos de componentes de señales dependiendo de si el valor absoluto ($|R(b, k)|$) del cambio ($R(b, k)$) de la frecuencia determinada para dicho compartimento de tiempo-frecuencia ((b, k)) es mayor que un tercer valor umbral r_p , o dependiendo de si el valor absoluto ($|R(b, k)|$) del cambio ($R(b, k)$) de la frecuencia determinada para dicho compartimento de tiempo-frecuencia ((b, k)) es mayor que o igual al tercer valor umbral r_p ,
 en el que $r_p \in \mathbb{R}^{>0}$.

12. Un aparato según cualquiera de las reivindicaciones anteriores, en el que el aparato comprende un generador de señales (130) que está configurado para generar una señal de salida de audio dependiendo de la asignación de la pluralidad de compartimentos de tiempo-frecuencia a los dos o más grupos de componentes de señal.

13. Un aparato según la reivindicación 12, en el que el generador de señales (130) está configurado para aplicar un factor de ponderación sobre los valores de magnitud (w_h , w_p , w_r) de cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia para obtener la señal de salida de audio, en el que el factor de ponderación (w_h , w_p , w_r) que es aplicado sobre dicho compartimento de tiempo-frecuencia depende del grupo de componentes de señales para el cual el compartimento de tiempo-frecuencia es asignado.

14. Un aparato según la reivindicación 12 o 13,
en el que el procesador de señales es un mezclador ascendente que está configurado para mezclar de forma ascendente la señal de audio para obtener la señal de salida de audio que comprende dos o más canales de salida de audio,
- 5 en el que el mezclador ascendente está configurado para generar los dos o más canales de salida de audio dependiendo de la asignación de la pluralidad de compartimentos de tiempo-frecuencia a los dos o más grupos de componentes de señal.
15. Un aparato según cualquiera de las reivindicaciones anteriores,
- 10 en el que el aparato comprende uno o más micrófonos (171, 172) para registrar la señal de audio, y
en el que el aparato comprende además un generador de espectrograma de magnitud (180) para generar el espectrograma de magnitud de la señal de audio a partir de la señal de audio.
16. Un procedimiento para analizar un espectrograma de magnitud de una señal de audio, **caracterizado**
- 15 **por:**
la determinación de un cambio de una frecuencia para cada compartimento de tiempo-frecuencia de una pluralidad de compartimentos de tiempo-frecuencia del espectrograma de magnitud de la señal de audio dependiendo del espectrograma de magnitud de la señal de audio, y
- 20 la asignación de cada compartimento de tiempo-frecuencia de la pluralidad de compartimentos de tiempo-frecuencia a un grupo de componentes de señales de dos o más grupos de componentes de señales dependiendo del cambio de la frecuencia determinada para dicho compartimento de tiempo-frecuencia.
17. Un programa informático para la implementación del procedimiento de la reivindicación 16 cuando se
- 25 ejecuta en un ordenador o procesador de señal.

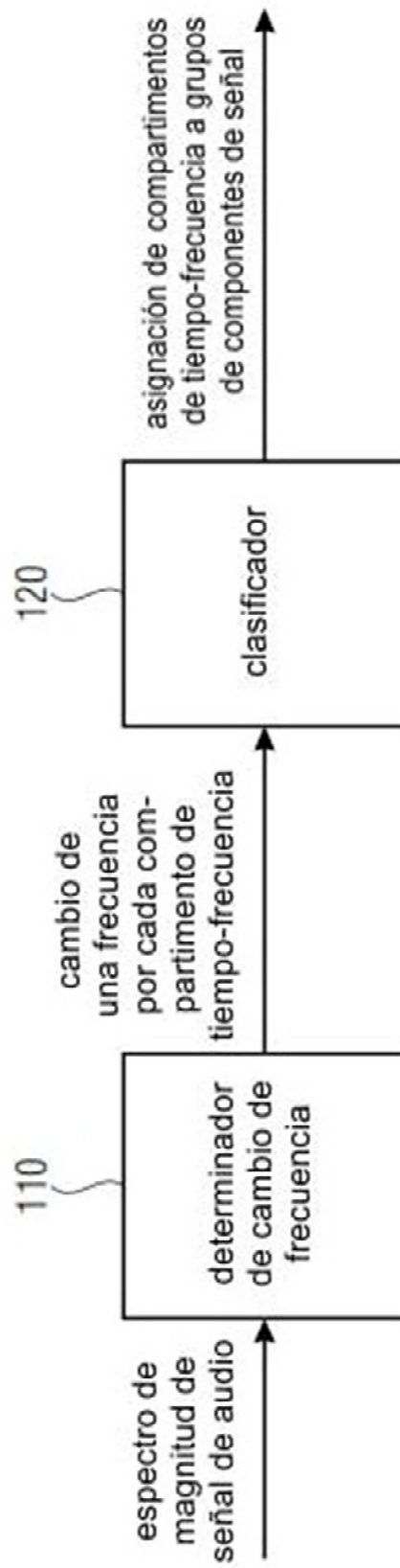


Fig. 1

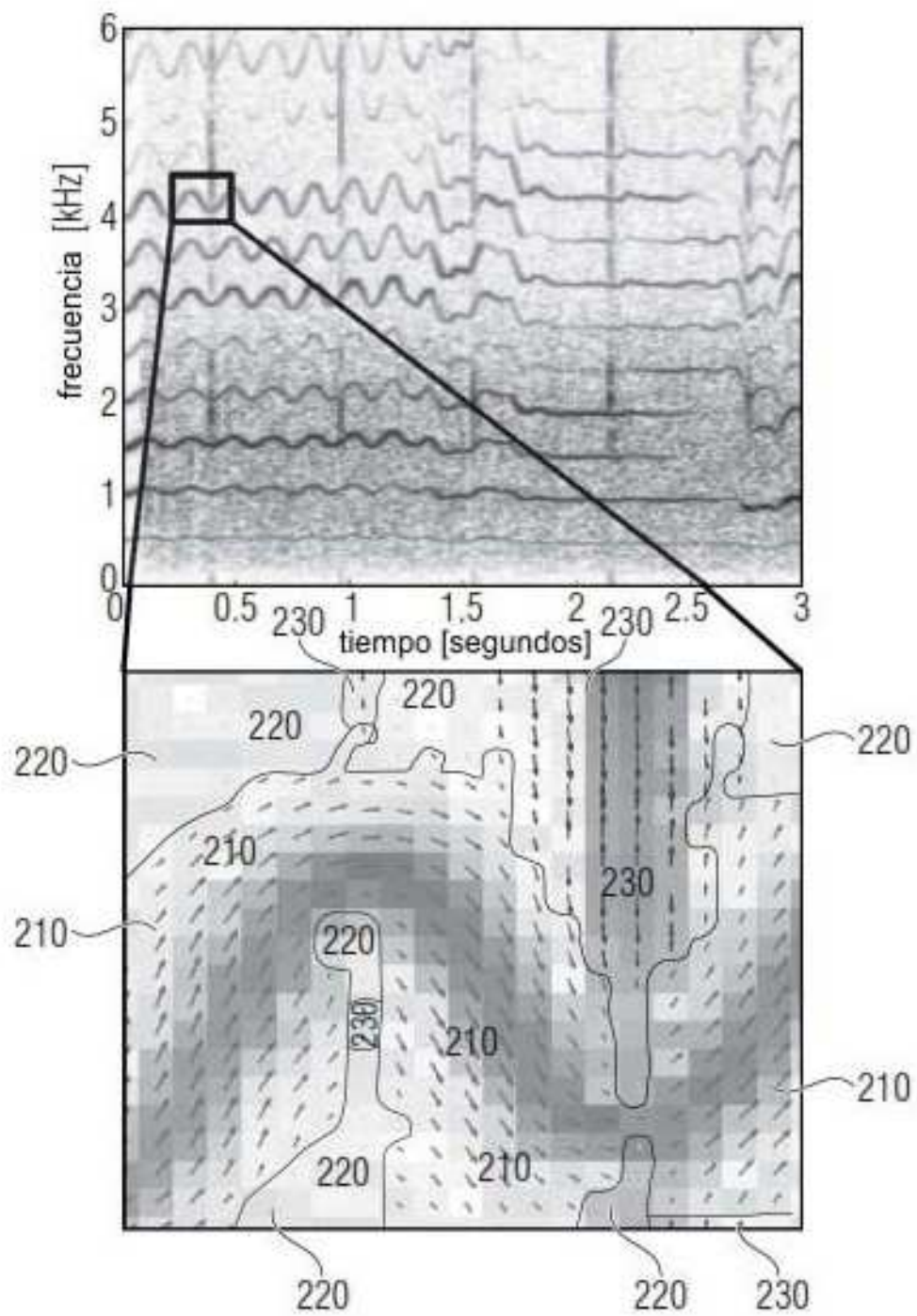


Fig. 2

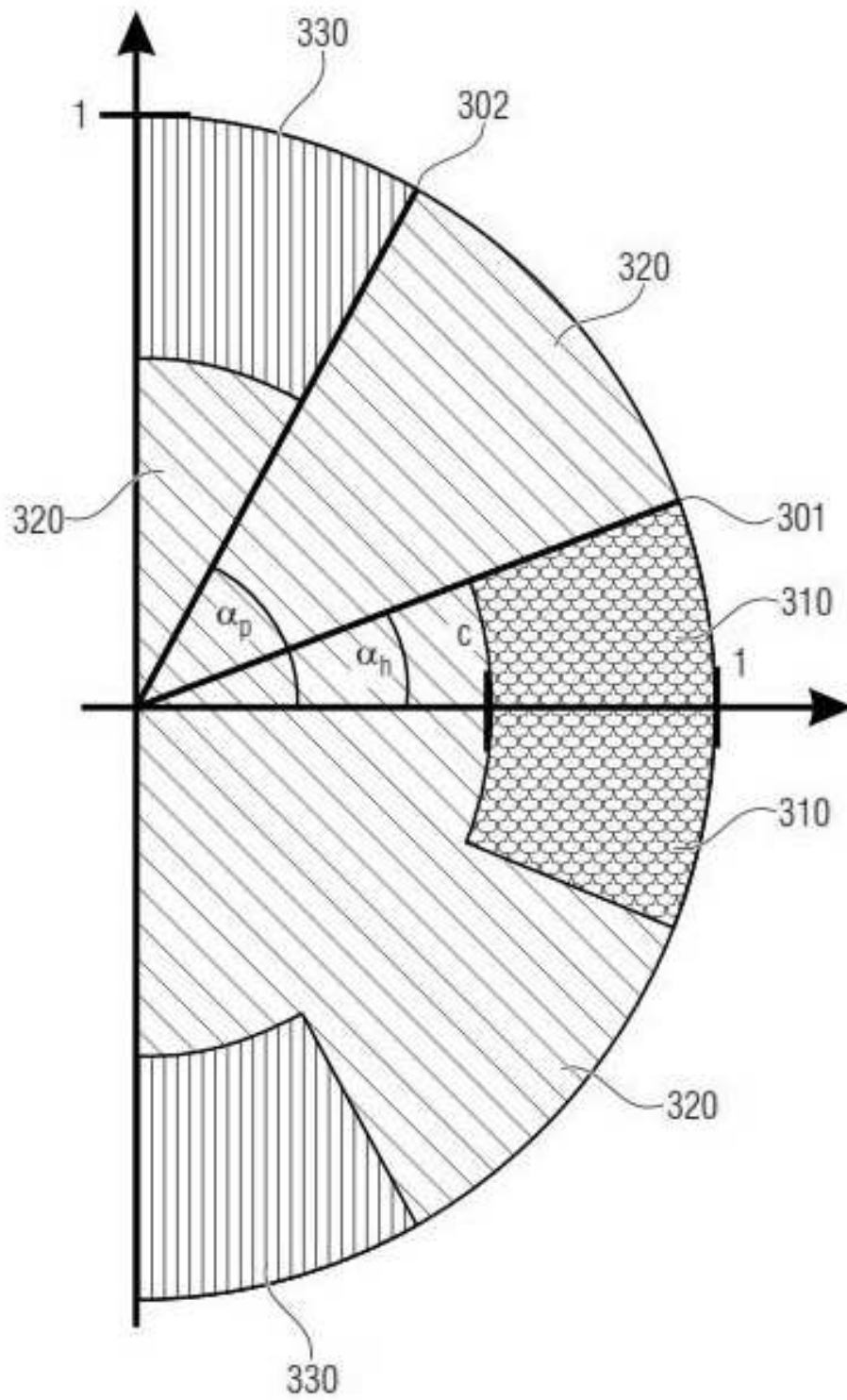


Fig. 3

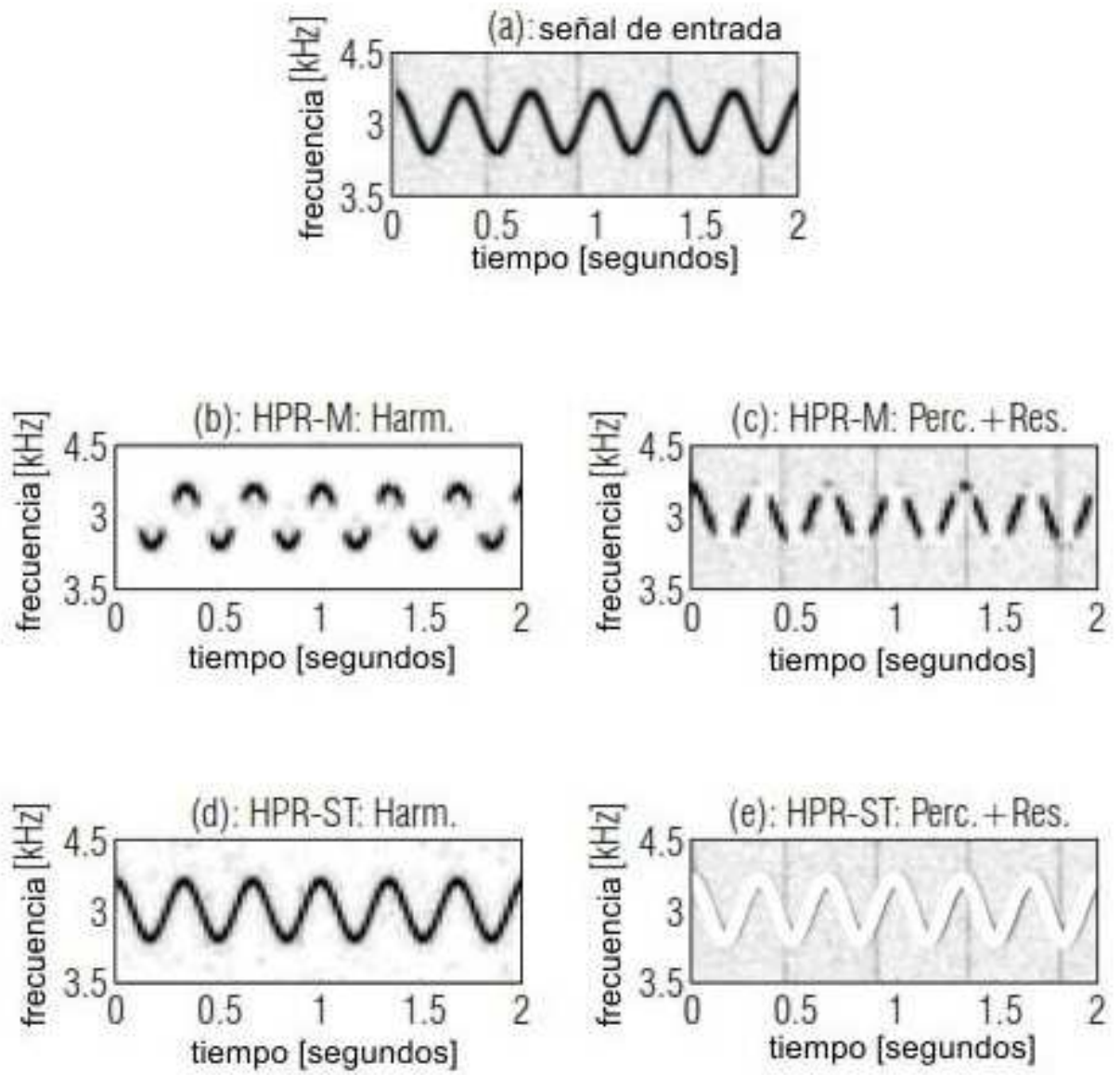


Fig. 4

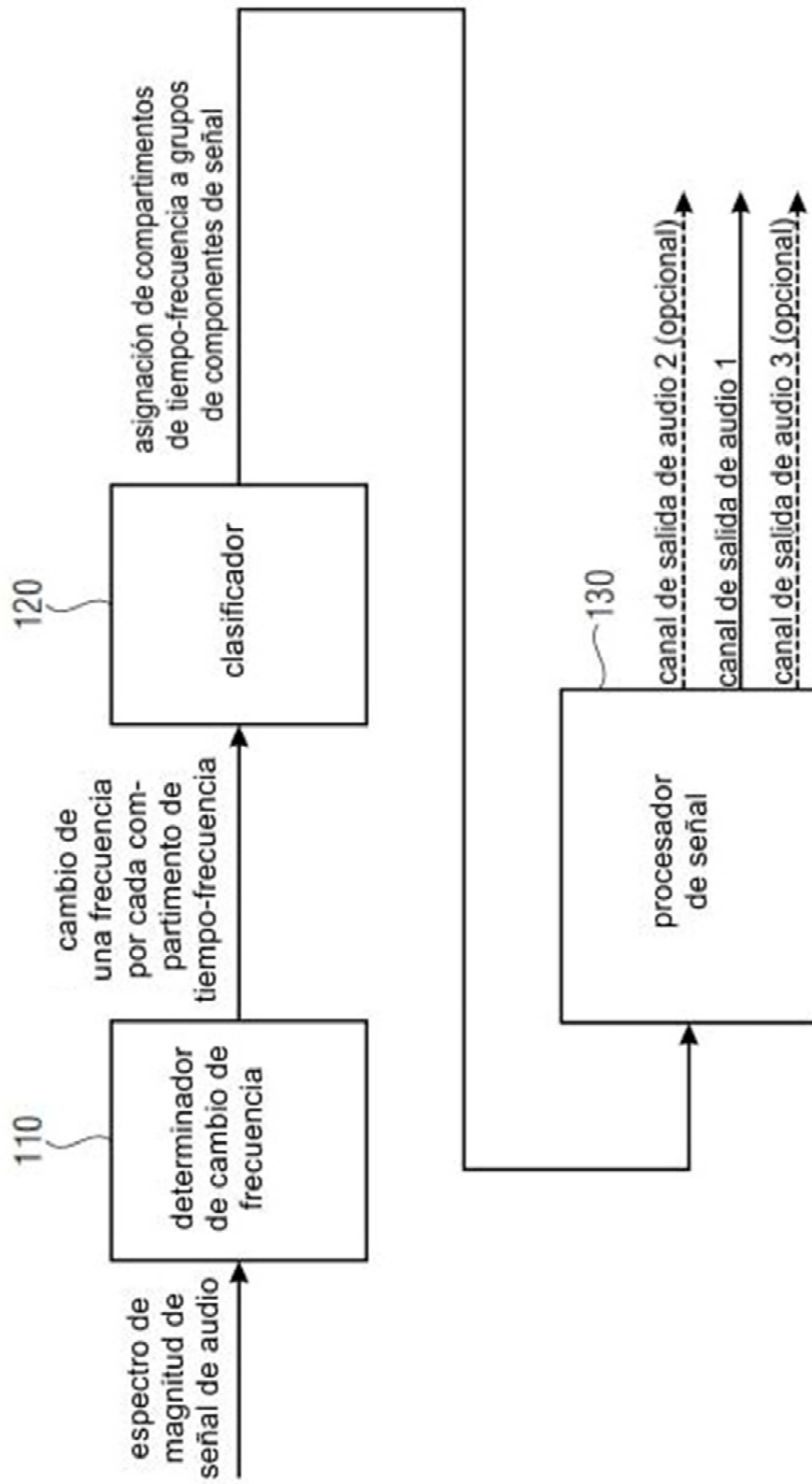


Fig. 5

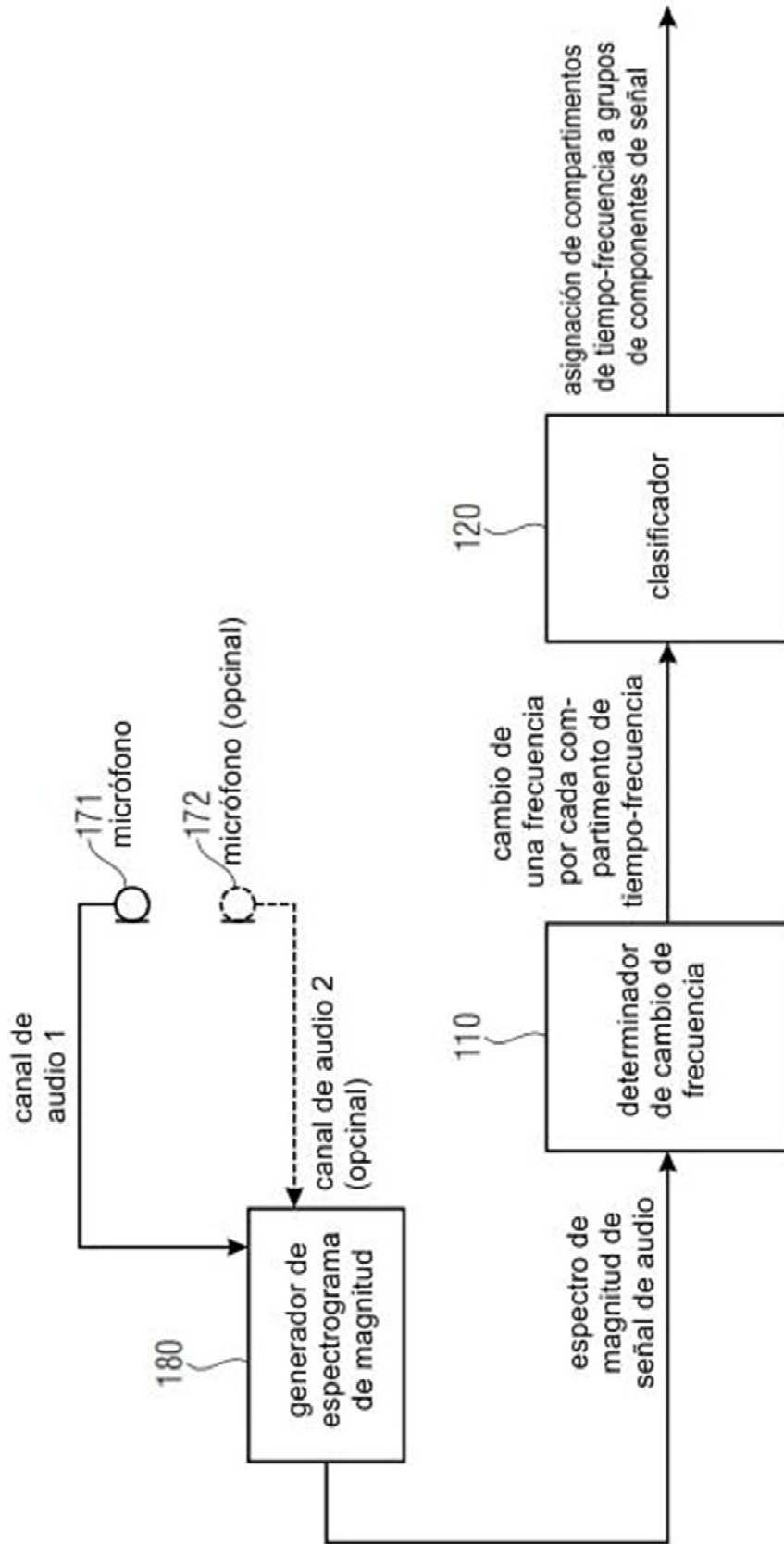


Fig. 6