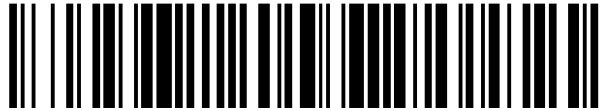


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 025 975**

51 Int. Cl.:

G10L 19/00 (2013.01)

G06N 3/045 (2013.01)

G06N 3/088 (2013.01)

G10L 25/30 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.10.2021 PCT/US2021/055090**

87 Fecha y número de publicación internacional: **21.04.2022 WO22081915**

96 Fecha de presentación y número de la solicitud europea: **14.10.2021 E 21798908 (6)**

97 Fecha y número de publicación de la concesión europea: **09.04.2025 EP 4229627**

54 Título: **Procedimiento y aparato para el procesamiento de audio utilizando una red neuronal**

30 Prioridad:

15.10.2020 US 202063092118 P
01.12.2020 EP 20210968

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
10.06.2025

73 Titular/es:

DOLBY LABORATORIES LICENSING CORPORATION (100.00%)
1275 Market Street
San Francisco, CA 94103, US

72 Inventor/es:

VINTON, MARK S.;
ZHOU, CONG;
FEJGIN, ROY M. y
DAVIDSON, GRANT A.

74 Agente/Representante:

ELZABURU, S.L.P

ES 3 025 975 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 aparato para el procesamiento de audio utilizando una red neuronal

5 Referencia cruzada a solicitudes relacionadas

Esta solicitud reivindica la prioridad de las siguientes solicitudes prioritarias: la solicitud provisional estadounidense 63/092,118, presentada el 15 de octubre de 2020, y la solicitud de patente europea 20210968.2, presentada el 1 de diciembre de 2020.

10

Tecnología

La presente divulgación se refiere en general a un procedimiento de procesamiento de una señal de audio utilizando una red neuronal o utilizando una primera y una segunda red neuronal y, en particular, a un procedimiento de procesamiento de una señal de audio en un dominio perceptual utilizando una red neuronal o utilizando una primera y una segunda red neuronal. La presente divulgación se refiere además a un procedimiento de entrenamiento de dicha red neuronal o de entrenamiento conjunto de un conjunto de dicha primera y dicha segunda red neuronal. La presente divulgación se refiere además a un procedimiento de obtención y transmisión de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal y a un procedimiento de obtención de una señal de audio a partir de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal. La presente divulgación también se refiere a los respectivos aparatos y productos de programas informáticos.

15

20

25

Si bien en la presente memoria se describirán algunas realizaciones con referencia particular a dicha divulgación, se apreciará que la presente divulgación no está limitada a dicho campo de uso y es aplicable en contextos más amplios.

Estado de la técnica

30

Cualquier análisis sobre la técnica anterior a lo largo de la divulgación no debe considerarse de ninguna manera como una admisión de que dicha técnica es ampliamente conocida o forma parte del conocimiento general común en el campo.

35

Los codificadores y descodificadores de audio de alto rendimiento aprovechan las limitaciones del sistema auditivo humano para eliminar información irrelevante que los humanos no pueden oír. Típicamente, los sistemas de codificación utilizan modelos psicoacústicos o perceptuales para calcular el umbral de enmascaramiento respectivo. Luego se utiliza el umbral de enmascaramiento para controlar el proceso de codificación de modo que el ruido introducido tenga un impacto audible mínimo.

40

Hasta ahora, las redes neuronales han demostrado ser prometedoras en muchas aplicaciones, incluida la codificación y/o descodificación de imágenes, vídeos e incluso habla. Sin embargo, todavía existe una necesidad de aplicación de redes neuronales en aplicaciones generales de codificación y/o descodificación de audio utilizando técnicas de entrenamiento típicas y, en particular, en aplicaciones de codificación y/o descodificación que involucran señales de audio en el dominio perceptual.

45

El documento US 2019/164052 A1 se refiere a un procedimiento de entrenamiento de una red neuronal que se aplica a un procedimiento de codificación de señales de audio utilizando un aparato de codificación de señales de audio. El procedimiento de entrenamiento incluye generar un umbral de enmascaramiento de una primera señal de audio antes de realizar el entrenamiento, calcular una matriz de ponderación que se aplicará a un componente de frecuencia de la primera señal de audio en base al umbral de enmascaramiento, generar una función de error ponderada obtenida al corregir una función de error preestablecida que utiliza la matriz de ponderación y generar una segunda señal de audio mediante la aplicación de un parámetro aprendido con la función de error ponderada a la primera señal de audio.

50

55

Compendio

La presente invención se define en las reivindicaciones adjuntas. A continuación, se entiende que las partes de la descripción y los dibujos que se refieren a realizaciones anteriores que no comprenden necesariamente todas las características para implementar realizaciones de la invención reivindicada no representan realizaciones de la invención, sino que se refieren a ejemplos útiles para comprender las realizaciones de la invención.

60

En un ejemplo, se proporciona un procedimiento de procesamiento de una señal de audio utilizando una red neuronal. El procedimiento puede incluir la etapa de (a) obtener una señal de audio en el dominio perceptual. El procedimiento puede incluir además la etapa de (b) introducir la señal de audio en el dominio perceptual

65

dentro de la red neuronal para procesar la señal de audio en el dominio perceptual. El procedimiento puede incluir además la etapa de (c) obtener, como salida de la red neuronal, una señal de audio procesada en el dominio perceptual. Y el procedimiento puede incluir la etapa de (d) convertir la señal de audio procesada en el dominio perceptual en un dominio de la señal original en base a una máscara que indica umbrales de enmascaramiento derivados a partir de un modelo psicoacústico.

5

En algunos ejemplos, el procesamiento de la señal de audio en el dominio perceptual mediante la red neuronal puede realizarse en un dominio del tiempo.

10

En algunos ejemplos, el procedimiento puede incluir, además, antes de la etapa (d), convertir la señal de audio a un dominio de la frecuencia.

En algunos ejemplos, la red neuronal puede estar condicionada a información indicativa de la máscara.

15

En algunos ejemplos, la red neuronal puede estar condicionada a la señal de audio en el dominio perceptual.

En algunos ejemplos, el procesamiento de la señal de audio en el dominio perceptual mediante la red neuronal puede incluir la predicción de la señal de audio procesada en el dominio perceptual a lo largo del tiempo.

20

En algunos ejemplos, el procesamiento de la señal de audio en el dominio perceptual mediante la red neuronal puede incluir la predicción de la señal de audio procesada en el dominio perceptual a lo largo de la frecuencia.

25

En algunos ejemplos, el procesamiento de la señal de audio en el dominio perceptual mediante la red neuronal puede incluir la predicción de la señal de audio procesada en el dominio perceptual a lo largo del tiempo y la frecuencia.

30

En algunos ejemplos, la señal de audio en el dominio perceptual se puede obtener: (a) al convertir una señal de audio desde el dominio de la señal original al dominio perceptual mediante la aplicación de la máscara; (b) al codificar la señal de audio en el dominio perceptual; y (c) al descodificar la señal de audio en el dominio perceptual. En algunos ejemplos, se puede aplicar cuantificación a la señal de audio en el dominio perceptual antes de la codificación y se puede aplicar cuantificación inversa a la señal de audio en el dominio perceptual después de la descodificación.

35

Según un aspecto de la presente divulgación, se proporciona un procedimiento de procesamiento de una señal de audio utilizando una primera y una segunda red neuronal. El procedimiento puede incluir la etapa de (a) obtener, mediante un primer aparato, una señal de audio en el dominio perceptual al aplicar una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico a una señal de audio en un dominio de la señal original. El procedimiento puede incluir además la etapa de (b) introducir la señal de audio en el dominio perceptual dentro de la primera red neuronal para proporcionar la correspondencia entre la señal de audio en el dominio perceptual y una representación del espacio de características latentes. El procedimiento puede incluir además la etapa de (c) obtener, como salida de la primera red neuronal, la representación del espacio de características latentes. El procedimiento puede incluir además la etapa de (d) transmitir la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara a un segundo aparato. El procedimiento puede incluir además la etapa de (e) recibir, mediante el segundo aparato, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara. El procedimiento puede incluir además la etapa de (f) introducir la representación del espacio de características latentes dentro de la segunda red neuronal para generar una señal de audio aproximada en el dominio perceptual. El procedimiento puede incluir además la etapa de (g) obtener, como salida de la segunda red neuronal, la señal de audio aproximada en el dominio perceptual. Y el procedimiento puede incluir la etapa de (h) convertir la señal de audio aproximada en el dominio perceptual al dominio de la señal original en base a la máscara.

40

45

50

55

En algunas realizaciones, el procedimiento puede incluir además codificar la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara en una corriente de bits y transmitir la corriente de bits al segundo aparato, en el que el procedimiento puede incluir además recibir la corriente de bits mediante el segundo aparato y descodificar la corriente de bits para obtener la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara.

60

En algunas realizaciones, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara se pueden cuantificar antes de la codificación en la corriente de bits y descuantificar antes del procesamiento mediante la segunda red neuronal.

65

En algunas realizaciones, la segunda red neuronal puede estar condicionada a la representación del espacio de características latentes de la señal de audio en el dominio perceptual y/o la máscara.

En algunas realizaciones, la correspondencia entre la señal de audio en el dominio perceptual y la representación del espacio de características latentes mediante la primera red neuronal y la generación de la señal de audio aproximada en el dominio perceptual mediante la segunda red neuronal se puede realizar en un dominio del tiempo.

5

En algunas realizaciones, la obtención de la señal en el dominio perceptual en la etapa (a) y la conversión de la señal en el dominio perceptual aproximada en la etapa (h) pueden realizarse en un dominio de la frecuencia.

10

En otro ejemplo, se proporciona un procedimiento para entrenar conjuntamente un conjunto de una primera y una segunda red neuronal. El procedimiento puede incluir la etapa de (a) introducir una señal de entrenamiento de audio en el dominio perceptual dentro de la primera red neuronal para proporcionar la correspondencia entre la señal de entrenamiento de audio en el dominio perceptual y una representación del espacio de características latentes. El procedimiento puede incluir además la etapa de (b) obtener, como salida de la primera red neuronal, la representación del espacio de características latentes de la señal de entrenamiento de audio en el dominio perceptual. El procedimiento puede incluir además la etapa de (c) introducir la representación del espacio de características latentes de la señal de entrenamiento de audio en el dominio perceptual dentro de la segunda red neuronal para generar una señal de entrenamiento de audio aproximada en el dominio perceptual. El procedimiento puede incluir además la etapa de (d) obtener, como salida de la segunda red neuronal, la señal de entrenamiento de audio aproximada en el dominio perceptual. Y el procedimiento puede incluir la etapa de (e) ajustar de forma iterativa los parámetros de la primera y la segunda red neuronal basándose en una diferencia entre la señal de entrenamiento de audio aproximada en el dominio perceptual y una señal de audio original en el dominio perceptual.

15

20

25

En algunos ejemplos, la primera y la segunda red neuronal pueden entrenarse en el dominio perceptual basándose en una o más funciones de pérdida.

En algunos ejemplos, la primera y la segunda red neuronal pueden entrenarse en el dominio perceptual basándose en una condición de probabilidad logarítmica negativa.

30

En otro ejemplo, se proporciona un procedimiento de entrenamiento de una red neuronal. El procedimiento puede incluir la etapa de (a) introducir una señal de entrenamiento de audio en el dominio perceptual dentro de la red neuronal para procesar la señal de entrenamiento de audio en el dominio perceptual. El procedimiento puede incluir además la etapa de (b) obtener, como salida de la red neuronal, una señal de entrenamiento de audio procesada en el dominio perceptual. Y el procedimiento puede incluir la etapa de (c) ajustar de forma iterativa los parámetros de la red neuronal basándose en una diferencia entre la señal de entrenamiento de audio procesada en el dominio perceptual y una señal de audio original en el dominio perceptual.

35

40

En algunos ejemplos, la red neuronal puede entrenarse en el dominio perceptual basándose en una o más funciones de pérdida.

En algunos ejemplos, la red neuronal puede entrenarse en el dominio perceptual basándose en una condición de probabilidad logarítmica negativa.

45

Según otro aspecto de la presente divulgación, se proporciona un procedimiento de obtención de y transmisión de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal. El procedimiento puede incluir la etapa de (a) obtener una señal de audio en el dominio perceptual mediante la aplicación de una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico a una señal de audio en un dominio de la señal original. El procedimiento puede incluir además la etapa de (b) introducir la señal de audio en el dominio perceptual dentro de una red neuronal para proporcionar la correspondencia entre la señal de audio en el dominio perceptual y una representación del espacio de características latentes. El procedimiento puede incluir además la etapa de (c) obtener, como salida de la red neuronal, la representación del espacio de características latentes de la señal de audio en el dominio perceptual. Y el procedimiento puede incluir la etapa de (d) emitir la representación del espacio de características latentes de la señal de audio en el dominio perceptual como una corriente de bits.

50

55

En algunos ejemplos, se puede emitir información adicional indicativa de la máscara como la corriente de bits en la etapa (d).

60

En algunos ejemplos, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y/o la información indicativa de la máscara se pueden cuantificar antes de la salida como la corriente de bits.

65

En algunos ejemplos, la correspondencia entre la señal de audio en el dominio perceptual y la representación del espacio de características latentes mediante la red neuronal se puede realizar en un dominio del tiempo.

En algunos ejemplos, la obtención de la señal de audio en el dominio perceptual puede realizarse en un dominio de la frecuencia.

5 Según otro aspecto de la presente divulgación, se proporciona un procedimiento de obtención de una señal de audio a partir de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal. El procedimiento puede incluir la etapa de (a) recibir una representación del espacio de características latentes de una señal de audio en el dominio perceptual como una corriente de bits. El procedimiento puede incluir además la etapa de (b) introducir la representación del espacio de características latentes dentro de una red neuronal para generar la señal de audio en el dominio perceptual. El procedimiento puede incluir además la etapa de (c) obtener, como salida de la red neuronal, la señal de audio en el dominio perceptual. Y el procedimiento puede incluir la etapa de (d) convertir la señal de audio en el dominio perceptual a un dominio de la señal original en base a una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico.

15 En algunos ejemplos, la red neuronal puede estar condicionada a la representación del espacio de características latentes de la señal de audio en el dominio perceptual.

En algunos ejemplos, en la etapa (a) se puede recibir información adicional indicativa de la máscara como la corriente de bits y la red neuronal puede estar condicionada a dicha información.

20 En algunos ejemplos, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y/o la información indicativa de la máscara pueden recibirse cuantificadas y puede realizarse una cuantificación inversa antes de la etapa (b).

25 En algunos ejemplos, la generación de la señal de audio en el dominio perceptual mediante la red neuronal puede realizarse en un dominio del tiempo.

En algunos ejemplos, la conversión de la señal de audio en el dominio perceptual al dominio de la señal original puede realizarse en un dominio de la frecuencia.

30 En otro ejemplo, se proporciona un aparato para procesar una señal de audio utilizando una red neuronal. El aparato puede incluir una red neuronal y uno o más procesadores configurados para realizar un procedimiento que incluye las etapas de (a) obtener una señal de audio en el dominio perceptual; (b) introducir la señal de audio en el dominio perceptual dentro de la red neuronal para procesar la señal de audio en el dominio perceptual; (c) obtener, como salida de la red neuronal, una señal de audio procesada en el dominio perceptual; y (d) convertir la señal de audio procesada en el dominio perceptual en un dominio de la señal original en base a una máscara que indica umbrales de enmascaramiento derivados a partir de un modelo psicoacústico.

40 En otro ejemplo, se proporciona un aparato para obtener y transmitir una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal. El aparato puede incluir una red neuronal y uno o más procesadores configurados para realizar un procedimiento que incluye las etapas de: (a) obtener una señal de audio en el dominio perceptual mediante la aplicación de una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico a una señal de audio en un dominio de la señal original; (b) introducir la señal de audio en el dominio perceptual dentro de una red neuronal para proporcionar la correspondencia entre la señal de audio en el dominio perceptual y una representación del espacio de características latentes; (c) obtener, como salida de la red neuronal, la representación del espacio de características latentes de la señal de audio en el dominio perceptual; y (d) emitir la representación del espacio de características latentes de la señal de audio en el dominio perceptual como una corriente de bits.

50 En otro ejemplo, se proporciona un aparato para obtener una señal de audio a partir de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal. El aparato puede incluir una red neuronal y uno o más procesadores configurados para realizar un procedimiento que incluye las etapas de: (a) recibir una representación del espacio de características latentes de una señal de audio en el dominio perceptual como una corriente de bits; (b) introducir la representación del espacio de características latentes dentro de una red neuronal para generar la señal de audio en el dominio perceptual; (c) obtener, como salida de la segunda red neuronal, la señal de audio en el dominio perceptual; y (d) convertir la señal de audio en el dominio perceptual en un dominio de la señal original en base a una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico.

60 Según aspectos adicionales de la presente divulgación, se proporcionan productos de programas informáticos que comprenden medios de almacenamiento legibles por ordenador con instrucciones adaptadas para hacer que los dispositivos lleven a cabo los procedimientos descritos en la presente memoria cuando son ejecutados por un dispositivo que tiene capacidad de procesamiento.

65

Breve descripción de los dibujos

A continuación, se describirán realizaciones de ejemplo de la divulgación únicamente a modo de ejemplo, en referencia a los dibujos adjuntos, en los que:

5 La FIG. 1 ilustra un ejemplo de un procedimiento de procesamiento de una señal de audio utilizando una red neuronal.

10 La FIG. 2 ilustra un ejemplo adicional de un procedimiento de procesamiento de una señal de audio utilizando una red neuronal.

La FIG. 3 ilustra un ejemplo de un sistema que incluye un aparato para procesar una señal de audio utilizando una red neuronal.

15 Las FIGS. 4a y 4b ilustran un ejemplo de un procedimiento de procesamiento de una señal de audio utilizando una primera y una segunda red neuronal.

20 La FIG. 5 ilustra un ejemplo de un sistema de un aparato para obtener y transmitir una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal y un aparato para obtener una señal de audio a partir de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal.

La FIG. 6 ilustra un ejemplo de un procedimiento de entrenamiento de una red neuronal.

25 La FIG. 7 ilustra un ejemplo de un procedimiento para entrenar conjuntamente un conjunto de una primera y una segunda red neuronal.

La FIG. 8 ilustra un ejemplo de una señal de audio original y una máscara en función del nivel y la frecuencia.

30 La FIG. 9 ilustra un ejemplo de una señal de audio en el dominio perceptual en función del nivel y la frecuencia obtenida a partir de la aplicación de la máscara a la señal de audio original.

35 La FIG. 10 ilustra un ejemplo de conversión de una señal de audio a un dominio perceptual y el procesamiento de dicha señal de audio utilizando una red neuronal.

40 La FIG. 11 ilustra un ejemplo de un codificador y descodificador de audio que funciona en el dominio perceptual con una red neuronal tanto en el codificador como en el descodificador. La figura también ilustra un ejemplo del uso de funciones de pérdida sencillas para el entrenamiento de las redes neuronales mientras las redes funcionan en el dominio perceptual.

La FIG. 12 ilustra un ejemplo de un codificador y descodificador de audio que funciona en el dominio perceptual con una red neuronal en el descodificador. La figura también ilustra un ejemplo del uso de funciones de pérdida sencillas para el entrenamiento de la red neuronal mientras la red funciona en el dominio perceptual.

45 Descripción de las realizaciones de ejemplo

Descripción general

50 Si bien las redes neuronales han demostrado ser prometedoras para codificar y/o descodificar imágenes, vídeos e incluso habla, codificar y/o descodificar audio general es complejo mediante el uso de redes neuronales. Existen dos factores que hacen que la compresión de audio general sea complicada con redes neuronales: en primer lugar, los codificadores y descodificadores de audio necesitan aprovechar las limitaciones del sistema auditivo humano para lograr un alto rendimiento. Para aprovechar la limitación perceptual del sistema auditivo humano, las redes neuronales no se pueden entrenar directamente con
55 funciones de pérdida no perceptual tales como L1 o L2:

$$L1 = \frac{1}{N} \sum_{n=1}^N |x_n - \bar{x}_n|$$

$$L2 = \frac{1}{N} \sum_{n=1}^N |x_n - \bar{x}_n|^2$$

donde, x_n es el objetivo (verdad fundamental) y \bar{x}_n es la predicción (salida de la red).

5 En segundo lugar, las señales de audio generales tienen un rango dinámico muy alto y son de naturaleza muy diversa, lo que complica el entrenamiento de redes neuronales.

10 La presente divulgación describe procedimientos y aparatos para transformar la señal de audio en un dominio perceptual antes de la aplicación de una red neuronal en los respectivos codificadores y/o descodificadores de audio. La conversión en el dominio perceptual de la señal de audio no solo reduce significativamente el rango dinámico, sino que también permite utilizar funciones de pérdida no perceptual, tales como L1 y L2, para entrenar la red.

Un procedimiento de procesamiento de una señal de audio utilizando una red neuronal

15 En referencia al ejemplo de la Figura 1, se ilustra un procedimiento de procesamiento de una señal de audio utilizando una red neuronal. En la etapa S101, se obtiene una señal de audio en el dominio perceptual. El término dominio perceptual tal como se usa en la presente memoria se refiere a una señal en la que la diferencia relativa de nivel entre los componentes de frecuencia es (aproximadamente) proporcional a su importancia subjetiva relativa. En general, una señal de audio convertida a un dominio perceptual minimiza el impacto audíble de añadir ruido blanco (ruido espectralmente plano) a la señal en el dominio perceptual, puesto que el ruido se moldeará para minimizar la audibilidad cuando la señal se convierta nuevamente en el dominio de la señal original.

25 En referencia al ejemplo de la Figura 2, la señal de audio en el dominio perceptual se puede obtener a partir de las etapas S101a, S101b y S101c, en el que en la etapa S101a una señal de audio se puede convertir desde el dominio de la señal original al dominio perceptual mediante la aplicación de una máscara.

30 Una forma de convertir una señal de audio al dominio perceptual puede ser, por ejemplo, utilizar un modelo psicoacústico para estimar una máscara o curva de enmascaramiento. Una curva de enmascaramiento generalmente define el nivel de distorsión apenas perceptible (JND) que el sistema auditivo humano puede detectar para una señal de estímulo determinada. Una vez se ha derivado la curva de enmascaramiento a partir del modelo psicoacústico, el espectro de la señal de audio se puede dividir por la curva de enmascaramiento para producir una señal de audio en el dominio perceptual. La señal de audio en el dominio perceptual derivada a partir de la multiplicación por la estimación de la máscara inversa se puede convertir nuevamente a la señal original multiplicándola por la máscara después de la codificación y/o descodificación de la red neuronal. La multiplicación por la máscara después de la descodificación asegurará que el error introducido por el proceso de codificación y descodificación siga la curva de enmascaramiento. Si bien esta es una forma de convertir una señal de audio original en el dominio perceptual, obsérvese que también son concebibles varias otras formas, por ejemplo, el filtrado en el dominio del tiempo mediante un filtro variable en el tiempo diseñado apropiadamente. En referencia al ejemplo de las Figuras 8 y 9, se ilustra una conversión de un espectro de una señal de audio original al dominio perceptual. La gráfica de la Figura 8 ilustra el espectro de una señal de audio original (línea continua) y la máscara estimada o curva de enmascaramiento (línea de puntos y guiones) calculada con un modelo psicoacústico. La señal en el dominio perceptual resultante de la multiplicación por la estimación de la máscara inversa se ilustra en el gráfico de la Figura 9. La señal en el dominio perceptual no solo permite utilizar términos de pérdida sencillas durante el entrenamiento de una red neuronal, sino que, como se ilustra en la Figura 8, presenta un rango dinámico mucho menor que el espectro de la señal de audio original.

50 En referencia nuevamente al ejemplo de la Figura 2, en la etapa S101b, la señal de audio en el dominio perceptual puede luego codificarse y posteriormente descodificarse en la etapa S101c para obtener la señal de audio en el dominio perceptual. En algunas realizaciones, se puede aplicar cuantificación a la señal de audio en el dominio perceptual antes de que se pueda aplicar codificación y cuantificación inversa a la señal de audio en el dominio perceptual después de la descodificación.

55

Volviendo al ejemplo de la Figura 1, en la etapa S102, la señal de audio en el dominio perceptual se introduce dentro de la red neuronal para procesar la señal de audio en el dominio perceptual. La red neuronal utilizada no está limitada y puede elegirse según los requisitos de procesamiento. Si bien la red neuronal puede funcionar en un dominio de la frecuencia, así como también en un dominio del tiempo, en algunas realizaciones, el procesamiento de la señal de audio en el dominio perceptual mediante la red neuronal puede realizarse en el dominio del tiempo. Además, en algunas realizaciones, la red neuronal puede estar condicionada a información indicativa de la máscara. De forma adicional o alternativa, en algunas realizaciones, la red neuronal puede estar condicionada a la señal de audio en el dominio perceptual.

El procesamiento de la señal de audio en el dominio perceptual mediante la red neuronal, en algunas realizaciones, puede incluir la predicción de la señal de audio procesada en el dominio perceptual a lo largo del tiempo. De forma alternativa, en algunas realizaciones, el procesamiento de la señal de audio en el dominio perceptual por la red neuronal puede incluir la predicción de la señal de audio procesada en el dominio perceptual a lo largo de la frecuencia. Además, de forma alternativa, en algunas realizaciones, el procesamiento de la señal de audio en el dominio perceptual por la red neuronal puede incluir la predicción de la señal de audio procesada en el dominio perceptual a lo largo del tiempo y la frecuencia.

En la etapa S103, se obtiene una señal de audio procesada en el dominio perceptual como salida de la red neuronal. En algunas realizaciones, la señal de audio procesada en el dominio perceptual se puede convertir al dominio de la frecuencia antes de la etapa siguiente S104.

En la etapa S104, la señal de audio procesada en el dominio perceptual se convierte en un dominio de la señal original en base a una máscara que indica umbrales de enmascaramiento derivados a partir de un modelo psicoacústico. Por ejemplo, para calcular la máscara, el modelo psicoacústico puede utilizar coeficientes de frecuencia de la transformación de tiempo a frecuencia aplicada para convertir la señal de audio en el dominio perceptual procesada en el dominio de la frecuencia. De forma adicional o alternativa, la máscara utilizada en la etapa S104 puede basarse en la máscara que se había utilizado para convertir la señal de audio original en el dominio perceptual. En este caso, la máscara puede obtenerse como información secundaria; opcionalmente, la máscara puede cuantificarse.

El término "señal de audio original" tal como se usa en la presente memoria se refiere al dominio de la señal respectivo de la señal de audio antes de la conversión de la señal de audio al dominio perceptual.

El procedimiento como se describe anteriormente puede implementarse de diversas maneras. Por ejemplo, dicho procedimiento puede implementarse mediante un aparato para procesar una señal de audio utilizando una red neuronal, en el que el aparato incluye una red neuronal y uno o más procesadores configurados para realizar dicho procedimiento.

En referencia al ejemplo de la Figura 3, se ilustra un sistema que incluye un aparato para procesar una señal de audio utilizando una red neuronal. El aparato puede ser un descodificador. En este caso, la red neuronal solo se utiliza en el descodificador.

Como se ilustra en el ejemplo de la Figura 3, una señal de audio en el dominio perceptual puede ser sometida a una cuantificación en un cuantificador, 101, y puede ser codificada (por entropía) por un codificador heredado respectivo, 102, por ejemplo. La señal de audio perceptual codificada cuantificada puede luego transmitirse, por ejemplo, como una corriente de bits, al descodificador, 103, para obtener la señal de audio en el dominio perceptual cuantificada, por ejemplo, mediante descodificación (por entropía) de la corriente de bits recibida. La señal de audio en el dominio perceptual cuantificado puede luego ser sometida a una cuantificación inversa en un cuantificador inverso respectivo, 104. La señal de audio en el dominio perceptual obtenida puede luego introducirse dentro de una red neuronal (red neuronal descodificadora), 105, para obtener una señal de audio procesada en el dominio perceptual como salida de la red neuronal, 105.

De forma adicional o alternativa, el procedimiento descrito anteriormente puede implementarse mediante un producto de programa informático que comprende un medio de almacenamiento legible por ordenador con instrucciones adaptadas para hacer que un dispositivo lleve a cabo dicho procedimiento cuando es ejecutado por un dispositivo que tiene capacidad de procesamiento.

Un procedimiento de procesamiento de una señal de audio utilizando una primera y una segunda red neuronal

En referencia al ejemplo de las Figuras 4a y 4b, se ilustra un procedimiento de procesamiento de una señal de audio utilizando una primera y una segunda red neuronal. La primera red neuronal puede, por ejemplo, implementarse en un sitio codificador, mientras que la segunda red neuronal puede implementarse en un sitio descodificador.

Como se ilustra en el ejemplo de la Figura 4a, en la etapa S201, un primer aparato obtiene una señal de audio en el dominio perceptual al aplicar una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico a una señal de audio en un dominio de la señal original. El primer aparato puede ser, por ejemplo, un codificador. En algunas realizaciones, la obtención de la señal de audio en el dominio perceptual puede realizarse en un dominio de la frecuencia.

En la etapa S202, la señal de audio en el dominio perceptual obtenida se introduce luego dentro de la primera red neuronal para proporcionar la correspondencia entre la señal de audio en el dominio perceptual y una representación del espacio de características latentes.

En algunas realizaciones, la correspondencia entre la señal de audio en el dominio perceptual y la representación del espacio de características latentes mediante la primera red neuronal se puede realizar en un dominio del tiempo.

Como salida de la primera red neuronal, en la etapa S203, se obtiene la representación del espacio de características latentes.

En la etapa S204, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara se transmiten luego a un segundo aparato. En algunas realizaciones, el procedimiento descrito anteriormente puede incluir además la codificación de la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara en una corriente de bits y la transmisión de la corriente de bits al segundo aparato. En algunas realizaciones, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara pueden cuantificarse adicionalmente antes de la codificación en la corriente de bits.

En referencia a continuación al ejemplo de la Figura 4b, en la etapa S205, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara se reciben mediante el segundo aparato. El segundo aparato puede ser, por ejemplo, un descodificador. En algunas realizaciones, el procedimiento puede incluir además recibir la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara como una corriente de bits mediante el segundo aparato y descodificar la corriente de bits para obtener la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara. En algunas realizaciones, en el caso de que la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara estén cuantificadas, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara pueden descuantificarse antes del procesamiento mediante la segunda red neuronal.

En la etapa S206, la representación del espacio de características latentes se introduce dentro de la segunda red neuronal para generar una señal de audio aproximada en el dominio perceptual. En algunas realizaciones, la segunda red neuronal puede estar condicionada a la representación del espacio de características latentes de la señal de audio en el dominio perceptual y/o la máscara. En algunas realizaciones, la generación de la señal de audio aproximada en el dominio perceptual mediante la segunda red neuronal puede realizarse en el dominio del tiempo.

En la etapa S207, como salida de la segunda red neuronal, se obtiene la señal de audio aproximada en el dominio perceptual.

La señal de audio aproximada en el dominio perceptual se convierte en la etapa S208 al dominio de la señal original en base a la máscara. En algunas realizaciones, la conversión de la señal en el dominio perceptual aproximado se puede realizar en el dominio de la frecuencia.

El procedimiento descrito anteriormente puede implementarse mediante un sistema de un primer y un segundo aparato respectivos. De forma adicional o alternativa, el procedimiento descrito anteriormente a continuación también puede implementarse mediante un producto de programa informático respectivo que comprende un medio de almacenamiento legible por ordenador con instrucciones adaptadas para hacer que un dispositivo lleve a cabo dichos procedimientos cuando son ejecutados por un dispositivo que tiene capacidad de procesamiento.

De forma alternativa, el procedimiento descrito anteriormente puede implementarse en parte mediante un aparato para obtener y transmitir una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal y en parte mediante un aparato para obtener una señal de audio a partir de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal. Los aparatos pueden entonces implementarse como aparatos autónomos o como un sistema.

El procedimiento de obtención y transmisión de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal incluye las etapas siguientes. En la etapa (a), se obtiene una señal de audio en el dominio perceptual mediante la aplicación de una máscara
 5 indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico a una señal de audio en un dominio de la señal original. En algunas realizaciones, la obtención de la señal de audio en el dominio perceptual puede realizarse en un dominio de la frecuencia.

En la etapa (b), la señal de audio en el dominio perceptual se introduce dentro de una red neuronal para proporcionar la correspondencia entre la señal de audio en el dominio perceptual y una representación del
 10 espacio de características latentes. En algunas realizaciones, la correspondencia entre la señal de audio en el dominio perceptual y la representación del espacio de características latentes mediante la red neuronal se puede realizar en un dominio del tiempo.

Como salida de la red neuronal, en la etapa (c), se obtiene la representación del espacio de características latentes de la señal de audio en el dominio perceptual. Y en la etapa (d), la representación del espacio de características latentes de la señal de audio en el dominio perceptual se emite luego como una corriente de bits.
 15

En algunas realizaciones, se puede emitir información adicional indicativa de la máscara como la corriente de bits en la etapa (d). En algunas realizaciones, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y/o la información indicativa de la máscara se pueden cuantificar antes de la salida como la corriente de bits.
 20

Un procedimiento de obtención de una señal de audio a partir de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal incluye las etapas siguientes. En la etapa (a), se recibe una representación del espacio de características latentes de una señal de audio en el dominio perceptual como una corriente de bits. En la etapa (b), la representación del espacio de características latentes se introduce dentro de una red neuronal para generar la señal de audio en el dominio perceptual. En la etapa (c), como salida de la red neuronal, se obtiene la señal de audio en el dominio perceptual. Y en la etapa (d), la señal de audio en el dominio perceptual se convierte en un dominio de la señal original en base a una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico.
 25
 30

En algunas realizaciones, la red neuronal puede estar condicionada a la representación del espacio de características latentes de la señal de audio en el dominio perceptual. En algunas realizaciones, además, en la etapa (a), la información indicativa de la máscara puede recibirse como la corriente de bits y la red neuronal puede estar condicionada a dicha información. En algunas realizaciones, la representación del espacio de características latentes de la señal de audio en el dominio perceptual y/o la información indicativa de la máscara pueden recibirse cuantificadas y puede realizarse una cuantificación inversa antes de la etapa (b). En algunas realizaciones, la generación de la señal de audio en el dominio perceptual mediante la red neuronal puede realizarse en un dominio del tiempo. En algunas realizaciones, la conversión de la señal de audio en el dominio perceptual al dominio de la señal original puede realizarse en un dominio de la frecuencia.
 35
 40

En referencia al ejemplo de la Figura 5, se ilustra un sistema de un aparato para obtener y transmitir una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal (también primer aparato) y un aparato para obtener una señal de audio a partir de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red neuronal (también segundo aparato).
 45

En el ejemplo de la Figura 5, en el (primer) aparato, 201, una señal de audio en el dominio perceptual puede introducirse en una (primera) red neuronal, 202, para su procesamiento como se describe anteriormente. La primera red neuronal, 202, puede ser una red neuronal codificadora. La representación del espacio de características latentes resultante de la (primera) red neuronal se puede cuantificar en un cuantificador, 203, y transmitir al (segundo) aparato, 204. La representación del espacio de características latentes cuantificado se puede codificar y transmitir al (segundo) aparato, 204, como una corriente de bits. En el (segundo) aparato, 204, la representación del espacio de características latentes recibida primero puede cuantificarse inversamente en un cuantificador inverso, 205, y opcionalmente descodificarse antes de introducirla dentro de la (segunda) red neuronal, 206, para generar una señal de audio aproximada en el dominio perceptual basada en la representación del espacio de características latentes. Como salida de la (segunda) red neuronal, 206, se puede obtener la señal de audio aproximada en el dominio perceptual.
 50
 55
 60

Un procedimiento de entrenamiento de una red neuronal

En referencia al ejemplo de la Figura 6, se ilustra un procedimiento de entrenamiento de una red neuronal. En la etapa S301, se introduce una señal de entrenamiento de audio en el dominio perceptual dentro de la red neuronal para procesar la señal de entrenamiento de audio en el dominio perceptual. La señal de entrenamiento
 65

de audio en el dominio perceptual es procesada por la red neuronal y en la etapa S302, se obtiene una señal de entrenamiento de audio procesada en el dominio perceptual como salida de dicha red neuronal. Basándose en una diferencia entre la señal de entrenamiento de audio procesada en el dominio perceptual y una señal de audio original en el dominio perceptual de la cual puede obtenerse la señal de entrenamiento de audio en el dominio perceptual, los parámetros de la red neuronal se ajustan de forma iterativa en la etapa S303. Basándose en este ajuste iterativo, la red neuronal se entrena para generar señales de entrenamiento de audio en el dominio perceptual cada vez mejor procesadas. El objetivo de este ajuste iterativo es hacer que la red neuronal genere una señal de entrenamiento de audio procesada en el dominio perceptual que sea indistinguible de la respectiva señal de audio original en el dominio perceptual.

En algunas realizaciones, la red neuronal puede entrenarse en el dominio perceptual basándose en una o más funciones de pérdida. Una red neuronal diseñada para codificar señales de audio en el dominio perceptual se puede entrenar con funciones de pérdida sencillas tales como L1 y L2, ya que pueden introducir un error espectral blanco. En el caso de L1 y L2, la red neuronal puede predecir la media de la señal de entrenamiento de audio procesada en el dominio perceptual.

De forma alternativa, en algunas realizaciones, la red neuronal puede entrenarse en el dominio perceptual basándose en una condición de probabilidad logarítmica negativa (NLL). En el caso de NLL, la red neuronal puede predecir la media y la escala como parametrización de una distribución preseleccionada. Típicamente se puede utilizar una operación logarítmica del parámetro de escala para evitar la inestabilidad numérica. La distribución preseleccionada puede ser laplaciana. De forma alternativa, la distribución preseleccionada puede ser una distribución logística o gaussiana. En el caso de una distribución gaussiana, el parámetro de escala puede reemplazarse por un parámetro de varianza. En el caso NLL, se puede utilizar una operación de muestreo para convertir los parámetros de distribución a la señal de entrenamiento de audio procesada en el dominio perceptual. La operación de muestreo se puede escribir como:

$$\tilde{x} = \text{mean} + F(u, \text{escala})$$

donde \tilde{x} es la señal de entrenamiento de audio procesada en el dominio perceptual prevista, la media y la escala son los parámetros previstos de la red neuronal, FQ es la función de muestreo determinada por la distribución preseleccionada y u se muestrea a partir de una distribución uniforme.

Por ejemplo, en un caso laplaciano,

$$F = -\text{escala} * \text{signo}(u) * \log(1 - 2 * |u|), u \sim (-0.5, 0.5)$$

se puede aplicar una función de ponderación derivada a partir de una máscara cuantificada al parámetro de escala en la función de muestreo F(). Además, en el caso del muestreo de una mezcla para cada coeficiente de salida (p. ej., mezcla gaussiana), puede haber un vector de parámetros.

Un procedimiento para entrenar conjuntamente un conjunto de una primera y una segunda red neuronal

En referencia al ejemplo de la Figura 7, se ilustra un procedimiento para entrenar conjuntamente un conjunto de una primera y una segunda red neuronal.

En la etapa S401, se introduce una señal de entrenamiento de audio en el dominio perceptual dentro de la primera red neuronal para proporcionar la correspondencia entre la señal de entrenamiento de audio en el dominio perceptual y una representación del espacio de características latentes. En la etapa S402, como salida de la primera red neuronal, se obtiene la representación del espacio de características latentes de la señal de entrenamiento de audio en el dominio perceptual. En la etapa S403, la representación del espacio de características latentes de la señal de entrenamiento de audio en el dominio perceptual se introduce dentro de la segunda red neuronal para generar una señal de entrenamiento de audio aproximada en el dominio perceptual. Como salida de la segunda red neuronal, en la etapa S404, se obtiene la señal de entrenamiento de audio aproximada en el dominio perceptual. Y en la etapa S405, los parámetros de la primera y la segunda red neuronal se ajustan de forma iterativa basándose en una diferencia entre la señal de entrenamiento de audio aproximada en el dominio perceptual y una señal de audio original en el dominio perceptual en base a la cual se ha derivado la señal de entrenamiento de audio en el dominio perceptual.

En algunas realizaciones, la primera y la segunda red neuronal pueden entrenarse en el dominio perceptual basándose en una o más funciones de pérdida. En algunas realizaciones, la primera y la segunda red neuronal pueden entrenarse en el dominio perceptual basándose en una condición de probabilidad logarítmica negativa (NLL). El objetivo del ajuste iterativo es hacer que la primera y la segunda red neuronal generen una señal de entrenamiento de audio aproximada en el dominio perceptual que sea indistinguible de la respectiva señal de audio original en el dominio perceptual.

Otras realizaciones ejemplares

En referencia a los ejemplos de las Figuras 10 a 12, se ilustran realizaciones ejemplares adicionales de los procedimientos y aparatos descritos en la presente memoria. En el ejemplo de la Figura 10, se ilustra un esquema que muestra la conversión de una señal de audio a un dominio perceptual para la reducción de datos utilizando una red neuronal. En el ejemplo de la Figura 10, se utilizan datos de audio PCM como entrada.

En el ejemplo de la Figura 11, se ilustra un esquema de un codificador y descodificador de audio que funciona en el dominio perceptual con una red neuronal tanto en el codificador como en el descodificador. La Figura 11 también muestra el uso de funciones de pérdida sencillas para el entrenamiento de las redes neuronales mientras las redes funcionan en el dominio perceptual. En el ejemplo de la Figura 11, la señal de la verdad fundamental se refiere a una señal de audio original en el dominio perceptual en base a la cual se puede derivar la respectiva señal de entrenamiento de audio en el dominio perceptual y la cual se puede comparar con la señal de audio aproximada en el dominio perceptual para ajustar de forma iterativa las redes neuronales.

En el ejemplo de la Figura 12, se ilustra un esquema de un codificador y descodificador de audio que funciona en el dominio perceptual con una red neuronal en el descodificador. La figura 12 también muestra el uso de funciones de pérdida sencillas para el entrenamiento de la red neuronal mientras la red funciona en el dominio perceptual. Además, en este caso, la señal de la verdad fundamental se refiere a una señal de audio original en el dominio perceptual en base a la cual se puede derivar la respectiva señal de entrenamiento de audio en el dominio perceptual y la cual se puede comparar con la señal de audio procesada en el dominio perceptual para ajustar de forma iterativa la red neuronal.

Interpretación

A menos que se indique específicamente lo contrario, como se desprende de los análisis siguientes, se entiende que, en toda la divulgación, los análisis que utilizan términos como "procesamiento", "cálculo", "determinación", "análisis" o similares, se refieren a la acción y/o procesos de un ordenador o sistema informático, o dispositivos informáticos electrónicos similares, que manipulan y/o transforman datos representados como cantidades físicas, tales como electrónicas, en otros datos representados de forma similar tales como cantidades físicas.

De forma similar, el término "procesador" puede referirse a cualquier dispositivo o porción de un dispositivo que procese datos electrónicos, p. ej., de registros y/o memoria para transformar esos datos electrónicos en otros datos electrónicos que, p. ej., puedan almacenarse en registros y/o memoria. Una "ordenador", una "máquina de cálculo" o una "plataforma informática" pueden incluir uno o más procesadores.

Las metodologías descritas en la presente memoria son, en una realización de ejemplo, ejecutables por uno o más procesadores que aceptan código legible por ordenador (también llamado legible por máquina) que contiene un conjunto de instrucciones que cuando son ejecutadas por uno o más de los procesadores llevan a cabo por lo menos uno de los procedimientos descritos en la presente memoria. Se incluye cualquier procesador capaz de ejecutar un conjunto de instrucciones (secuenciales o de otro tipo) que especifiquen acciones a realizar. Así, un ejemplo es un sistema de procesamiento típico que incluye uno o más procesadores. Cada procesador puede incluir uno o más de los siguientes: una CPU, una unidad de procesamiento de gráficos y una unidad DSP programable. El sistema de procesamiento puede incluir además un subsistema de memoria que incluye RAM principal y/o una RAM estática y/o ROM. Se puede incluir un subsistema de bus para la comunicación entre los componentes. El sistema de procesamiento también puede ser un sistema de procesamiento distribuido con procesadores acoplados por una red. Si el sistema de procesamiento requiere un visualizador, se puede incluir dicho visualizador, p. ej., un visualizador de cristal líquido (LCD) o un visualizador de tubo de rayos catódicos (CRT). Si se requiere la entrada de datos manual, el sistema de procesamiento también incluye un dispositivo de entrada, tal como una o más unidades de entrada alfanumérica tal como un teclado, un dispositivo de control apuntador tal como un ratón, etc. El sistema de procesamiento también puede englobar un sistema de almacenamiento, tal como una unidad de disco. El sistema de procesamiento en algunas configuraciones puede incluir un dispositivo de salida de sonido y un dispositivo de interfaz de red. El subsistema de memoria incluye así un medio portador legible por ordenador que transporta un código legible por ordenador (p. ej., software) que incluye un conjunto de instrucciones para hacer que se realicen, cuando es ejecutado por uno o más procesadores, uno o más de los procedimientos descritos en la presente memoria. Obsérvese que cuando el procedimiento incluye varios elementos, p. ej., varias etapas, no hay ningún orden de dichos elementos implicado, a menos que se indique específicamente. El software puede residir en el disco duro, o también puede residir, total o por lo menos parcialmente, en la memoria RAM y/o en el procesador durante la ejecución del mismo por el sistema informático. Así, la memoria y el procesador también constituyen un medio portador legible por ordenador que transporta código legible por ordenador. Es más, un medio portador legible por ordenador puede formar o estar incluido en un producto de programa informático.

- En realizaciones de ejemplo alternativas, el uno o más procesadores funcionan como un dispositivo autónomo o pueden estar conectados, p. ej., en red con otro(s) procesador(es), en una implementación en red, el uno o más procesadores pueden funcionar en la capacidad de un servidor o una máquina de usuario en un entorno de red servidor-usuario, o como una máquina del mismo nivel en un entorno de red entre pares o distribuido.
- 5 El uno o más procesadores pueden formar un ordenador personal (PC), una tableta electrónica, un asistente personal digital (PDA), un teléfono celular, un electrodoméstico conectado a Internet, un enrutador, conmutador o puente de red, o cualquier máquina capaz de ejecutar un conjunto de instrucciones (secuenciales o de otro tipo) que especifiquen acciones que debe realizar esa máquina.
- 10 Obsérvese que el término "máquina" también debe entenderse como cualquier conjunto de máquinas que, individual o conjuntamente, ejecuten un conjunto (o múltiples conjuntos) de instrucciones para realizar una o más de las metodologías analizadas en la presente memoria.
- 15 Así, una realización de ejemplo de cada uno de los procedimientos descritos en la presente memoria es en forma de un medio portador legible por ordenador que transporta un conjunto de instrucciones, p. ej., un programa informático que se va a ejecutar en uno o más procesadores, p. ej., uno o más procesadores que forman parte de una disposición de servidor web. Así, como apreciarán los expertos en la materia, las realizaciones de ejemplo de la presente divulgación pueden materializarse como un procedimiento, un aparato tal como un aparato de propósito especial, un aparato tal como un sistema de procesamiento de datos o un
- 20 medio portador legible por ordenador, p. ej., un producto de programa informático. El medio portador legible por ordenador transporta un código legible por ordenador que incluye un conjunto de instrucciones que, cuando se ejecutan en uno o más procesadores, hacen que el procesador o los procesadores implementen un procedimiento. En consecuencia, los aspectos de la presente divulgación pueden adoptar la forma de un procedimiento, una realización de ejemplo completamente de hardware, una realización de ejemplo
- 25 completamente de software o una realización de ejemplo que combina aspectos de software y hardware. Es más, la presente divulgación puede adoptar la forma de un medio portador (p. ej., un producto de programa informático en un medio de almacenamiento legible por ordenador) que transporta un código de programa legible por ordenador incorporado en el medio.
- 30 El software también puede transmitirse o recibirse a través de una red por medio de un dispositivo de interfaz de red. Si bien en una realización de ejemplo el medio portador es un medio único, el término "medio portador" debe entenderse como que incluye un medio único o múltiples medios (p. ej., una base de datos centralizada o distribuida, y/o memorias cachés y servidores asociados) que almacenan el uno o más conjuntos de
- 35 instrucciones. El término "medio portador" también se entenderá que incluye cualquier medio que sea capaz de almacenar, codificar o transportar un conjunto de instrucciones para su ejecución por uno o más de los procesadores y que haga que el uno o más procesadores realicen una o más de las metodologías de la presente divulgación. Un medio de este tipo puede adoptar muchas formas, incluyendo pero sin limitarse a, medios no volátiles, medios volátiles y medios de transmisión. Un medio portador puede adoptar muchas formas que incluyen, pero no se limitan a, medios no volátiles, medios volátiles y medios de transmisión. Los
- 40 medios no volátiles incluyen, por ejemplo, discos ópticos, discos magnéticos y discos magnetoópticos. Los medios volátiles incluyen la memoria dinámica, tal como la memoria principal. Los medios de transmisión incluyen cables coaxiales, cables de cobre y fibra óptica, incluidos los cables que comprenden un subsistema de bus. Los medios de transmisión también pueden adoptar la forma de ondas acústicas o luminosas, tales como las que se generan durante las comunicaciones de datos por ondas radioeléctricas o infrarrojos. Por
- 45 ejemplo, el término "medio portador" se entenderá que incluye, pero no se limita a, memorias de estado sólido, un producto informático incorporado en medios ópticos y magnéticos; un medio que tiene una señal propagada detectable por al menos un procesador o uno o más procesadores y que representa un conjunto de instrucciones que, cuando se ejecutan, implementan un procedimiento; y un medio de transmisión en una red que tiene una señal propagada detectable por al menos un procesador del uno o más procesadores y que
- 50 representa el conjunto de instrucciones. Se entenderá que las etapas de los procedimientos analizados se realizan en una realización de ejemplo mediante un procesador (o procesadores) apropiado de un sistema de procesamiento (p. ej., un ordenador) que ejecuta instrucciones (código legible por ordenador) almacenadas en el almacenamiento. También se entenderá que la divulgación no está limitada a ninguna implementación o técnica de programación en particular y que la divulgación puede implementarse utilizando cualquier técnica apropiada para implementar la funcionalidad descrita en la presente memoria. La divulgación no se limita a ningún lenguaje de programación o sistema operativo en particular.
- 55 La referencia a lo largo de la presente divulgación a "una realización", "algunas realizaciones" o "una realización de ejemplo" significa que una función, estructura o característica en particular descrito en relación con la realización está incluido en por lo menos una realización de la presente divulgación. Así, cuando aparecen expresiones como "en una realización", "en algunas realizaciones" o "en una realización de ejemplo" en diversos lugares de la presente divulgación no necesariamente se refieren todas a la misma realización de ejemplo. Es más, las funciones, estructuras o características en particular pueden combinarse de cualquier
- 60 manera adecuada, como sería evidente para una persona con experiencia ordinaria en la técnica a partir de la presente divulgación, en una o más realizaciones de ejemplo.
- 65

Tal como se usa en la presente memoria, a menos que se especifique lo contrario, el uso de los adjetivos ordinales "primero", "segundo", "tercero", etc., para describir un objeto común, meramente indica que se hace referencia a diferentes instancias de objetos similares y no pretende implicar que los objetos así descritos deben estar en una secuencia dada, ya sea temporalmente, espacialmente, en clasificación o de cualquier otra manera.

5

En las reivindicaciones siguientes y la descripción en la presente memoria, uno cualquiera de los términos comprendiendo, compuesto de o que comprende es un término abierto que significa incluir por lo menos los elementos/características que siguen, pero sin excluir otros. Así, el término que comprende, cuando se utiliza en las reivindicaciones, no debe interpretarse como limitativo de los medios, elementos o etapas enumerados después. Por ejemplo, el alcance de la expresión un dispositivo que comprende A y B no debe limitarse a dispositivos que consisten únicamente en los elementos A y B. Uno cualquiera de los términos incluyendo o el cual incluye o que incluye como se usa en la presente memoria también es un término abierto que también significa incluir por lo menos los elementos/características que siguen al término, pero sin excluir otros. Así, incluir es sinónimo de comprender y significa comprender.

10

15

Se debe tener en cuenta que, en la descripción anterior de realizaciones de ejemplo de la divulgación, diversas características de la divulgación a veces se agrupan juntas en una única realización de ejemplo, Fig., o descripción de la misma con el propósito de simplificar la divulgación y ayudar en la comprensión de uno o más de los diversos aspectos de la invención. Sin embargo, este procedimiento de divulgación no debe interpretarse como un reflejo de una intención de que las reivindicaciones requieran más características que las que se mencionan expresamente en cada reivindicación. Más bien, como reflejan las reivindicaciones siguientes, los aspectos de la invención residen en menos de todas las características de una única realización de ejemplo divulgada anteriormente. Así, las reivindicaciones que siguen a la descripción se incorporan por este medio expresamente a esta descripción, y cada reivindicación constituye por sí sola una realización de ejemplo independiente de esta divulgación.

20

25

Es más, aunque algunas realizaciones de ejemplo descritas en la presente memoria incluyen algunas, pero no otras características incluidas en otras realizaciones de ejemplo, se pretende que las combinaciones de las características de diferentes realizaciones de ejemplo estén dentro del alcance de la divulgación, y formen realizaciones de ejemplo distintas, como entenderán las personas con experiencia en la materia. Por ejemplo, en las siguientes reivindicaciones, cualquiera de las realizaciones de ejemplo reivindicadas puede utilizarse en cualquier combinación.

30

En la descripción proporcionada en la presente memoria, se exponen numerosos detalles concretos. Sin embargo, se entiende que las realizaciones de ejemplo de la divulgación pueden ponerse en práctica sin estos detalles concretos. En otros casos, los procedimientos, estructuras y técnicas bien conocidos no se han mostrado en detalle para no oscurecer la comprensión de esta descripción.

35

Así, si bien se han descrito los que se cree que son los mejores modos de la divulgación, los expertos en la materia reconocerán que pueden realizarse otras modificaciones adicionales. Por ejemplo, las fórmulas dadas anteriormente son meramente representativas de procedimientos que pueden utilizarse. Se pueden añadir o eliminar funcionalidades de los diagramas de bloques y se pueden intercambiar operaciones entre bloques funcionales. Se pueden añadir o eliminar etapas a los procedimientos descritos dentro del alcance de la presente divulgación.

40

45

REIVINDICACIONES

1. Un procedimiento implementado por ordenador para codificar una señal de audio utilizando una red (202) neuronal, incluyendo el procedimiento las etapas de:
- 5 (a) obtener una señal de audio en el dominio perceptual mediante la aplicación de una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico a la señal de audio en un dominio de la señal original;
- 10 (b) introducir la señal de audio en el dominio perceptual dentro de una red (202) neuronal para proporcionar la correspondencia entre la señal de audio en el dominio perceptual y una representación del espacio de características latentes;
- (c) obtener, como salida de la red (202) neuronal, la representación del espacio de características latentes de la señal de audio en el dominio perceptual; y
- 15 (d) emitir la representación del espacio de características latentes de la señal de audio en el dominio perceptual en una corriente de bits.
2. Un procedimiento implementado por ordenador para descodificar una señal de audio utilizando una red (206) neuronal, en el que el procedimiento incluye las etapas de:
- 20 (a) obtener una representación de una señal de audio en el dominio perceptual mediante la descodificación de una corriente de bits recibida;
- (b) introducir la representación de la señal de audio en el dominio perceptual dentro de la red (206) neuronal para procesar la representación de la señal de audio en el dominio perceptual;
- 25 (c) obtener, como salida de la red (206) neuronal, una señal de audio procesada en el dominio perceptual; y
- (d) convertir la señal de audio procesada en el dominio perceptual en un dominio de la señal original en base a una máscara que indica umbrales de enmascaramiento derivados a partir de un modelo psicoacústico.
- 30 3. El procedimiento según la reivindicación 2, en el que la red (206) neuronal está condicionada a información indicativa de la máscara; y/o en el que la red (206) neuronal está condicionada a la señal de audio en el dominio perceptual.
- 35 4. El procedimiento según la reivindicación 3, en el que el procesamiento de la señal de audio en el dominio perceptual mediante la red (206) neuronal incluye por lo menos uno de:
- 40 predecir la señal de audio procesada en el dominio perceptual a lo largo del tiempo;
- predecir la señal de audio procesada en el dominio perceptual a lo largo de la frecuencia; y
- predecir la señal de audio procesada en el dominio perceptual a lo largo del tiempo y la frecuencia.
- 45 5. El procedimiento según una cualquiera de las reivindicaciones 2 a 4, en el que la representación de la señal de audio en el dominio perceptual comprende la señal de audio en el dominio perceptual, y/o en el que la representación de la señal de audio en el dominio perceptual se obtiene a partir de:
- convertir una señal de audio desde el dominio de la señal original al dominio perceptual mediante la aplicación de la máscara;
- codificar la señal de audio en el dominio perceptual; y
- descodificar la señal de audio en el dominio perceptual; y opcionalmente
- 50 en el que la cuantificación se aplica a la señal de audio en el dominio perceptual antes de la codificación y la cuantificación inversa se aplica a la señal de audio en el dominio perceptual después de la descodificación.
6. El procedimiento según la reivindicación 2,
- 55 en el que la etapa (a) implica recibir una representación del espacio de características latentes de la señal de audio en el dominio perceptual en una corriente de bits; y en el que la etapa (b) implica introducir la representación del espacio de características latentes dentro de la red (206) neuronal para generar la señal de audio procesada en el dominio perceptual, y/o en el que la red (206) neuronal está condicionada a la representación del espacio de características latentes de la señal de audio en el dominio perceptual; y/o
- 60 en el que el procedimiento comprende además recibir información adicional indicativa de la máscara como la corriente de bits, y en el que la red (206) neuronal está condicionada a dicha información adicional.

7. Un procedimiento implementado por ordenador para procesar una señal de audio utilizando una primera (202) y una segunda red (206) neuronal, en el que el procedimiento incluye las etapas de:

- 5 (a) obtener, mediante un primer aparato (201), una señal de audio en el dominio perceptual mediante la aplicación a una señal de audio en un dominio de la señal original de una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo (203) psicoacústico;
- (b) introducir la señal de audio en el dominio perceptual dentro de la primera red (202) neuronal para proporcionar la correspondencia entre la señal de audio en el dominio perceptual y una representación del espacio de características latentes;
- 10 (c) obtener, como salida de la primera red (202) neuronal, la representación del espacio de características latentes;
- (d) transmitir la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara a un segundo aparato (204);
- 15 (e) recibir, mediante el segundo aparato (204), la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara;
- (f) introducir la representación del espacio de características latentes dentro de la segunda red (206) neuronal para generar una señal de audio aproximada en el dominio perceptual;
- (g) obtener, como salida de la segunda red (206) neuronal, la señal de audio aproximada en el dominio perceptual; y
- 20 (h) convertir la señal de audio aproximada en el dominio perceptual al dominio de la señal original en base a la máscara.

8. El procedimiento según la reivindicación 7, en el que el procedimiento incluye además codificar la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara en una corriente de bits y transmitir la corriente de bits al segundo aparato (204), y en el que el procedimiento incluye además recibir la corriente de bits mediante el segundo aparato (204) y descodificar la corriente de bits para obtener la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara.

9. El procedimiento según la reivindicación 8, en el que la representación del espacio de características latentes de la señal de audio en el dominio perceptual y la máscara se cuantifican antes de la codificación en la corriente de bits y se descuantifican antes del procesamiento mediante la segunda red (206) neuronal.

10. El procedimiento según una cualquiera de las reivindicaciones 7 a 9, en el que la segunda red (206) neuronal está condicionada a la representación del espacio de características latentes de la señal de audio en el dominio perceptual y/o la máscara; y/o

en el que la correspondencia entre la señal de audio en el dominio perceptual y la representación del espacio de características latentes mediante la primera red (202) neuronal y la generación de la señal de audio aproximada en el dominio perceptual mediante la segunda red (206) neuronal se realiza en un dominio del tiempo; y/o

en el que la obtención de la señal en el dominio perceptual en la etapa (a) y la conversión de la señal en el dominio perceptual aproximada en la etapa (h) se realizan en un dominio de la frecuencia.

11. Un procedimiento implementado por ordenador para obtener y transmitir una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red (202) neuronal, incluyendo el procedimiento las etapas de:

- 50 (a) obtener una señal de audio en el dominio perceptual mediante la aplicación de una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico a una señal de audio en un dominio de la señal original;
- (b) introducir la señal de audio en el dominio perceptual dentro de una red (202) neuronal para proporcionar la correspondencia entre la señal de audio en el dominio perceptual y una representación del espacio de características latentes;
- 55 (c) obtener, como salida de la red (202) neuronal, la representación del espacio de características latentes de la señal de audio en el dominio perceptual; y
- (d) emitir la representación del espacio de características latentes de la señal de audio en el dominio perceptual como una corriente de bits.

12. Un procedimiento implementado por ordenador para obtener una señal de audio a partir de una representación del espacio de características latentes de una señal de audio en el dominio perceptual utilizando una red (206) neuronal, incluyendo el procedimiento las etapas de:

- 65 (a) recibir una representación del espacio de características latentes de una señal de audio en el dominio perceptual como una corriente de bits;

(b) introducir la representación del espacio de características latentes dentro de una red (206) neuronal para generar la señal de audio en el dominio perceptual;

(c) obtener, como salida de la red (206) neuronal, la señal de audio en el dominio perceptual; y

5 (d) convertir la señal de audio en el dominio perceptual a un dominio de la señal original en base a una máscara indicativa de los umbrales de enmascaramiento derivados a partir de un modelo psicoacústico.

13. Un dispositivo configurado para llevar a cabo el procedimiento según una cualquiera de las reivindicaciones 1 a 12.

10 14. Un programa informático que comprende instrucciones adaptadas para hacer que un dispositivo que tiene capacidad de procesamiento lleve a cabo el procedimiento según una cualquiera de las reivindicaciones 1 a 12 cuando es ejecutado por el dispositivo.

15 15. Un medio de almacenamiento legible por ordenador con instrucciones adaptadas para hacer que un dispositivo que tiene capacidad de procesamiento lleve a cabo el procedimiento según una cualquiera de las reivindicaciones 1 a 12 cuando son ejecutadas por el dispositivo.

DIBUJOS

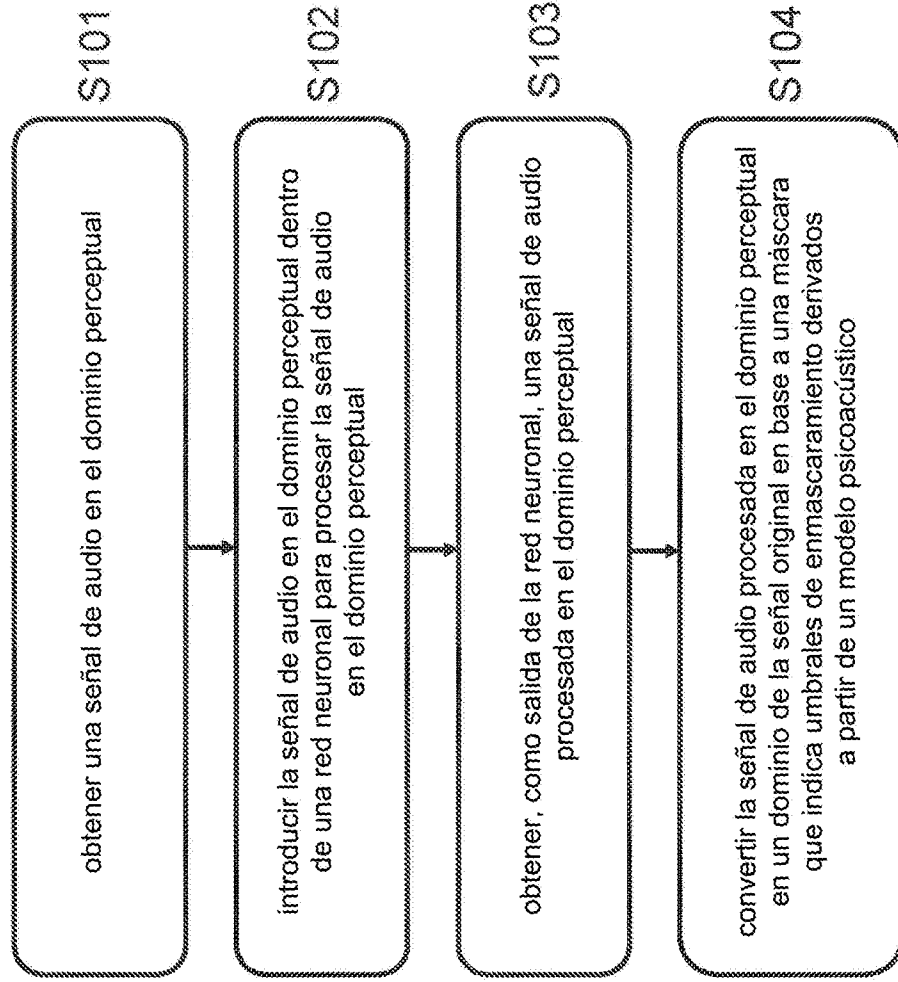


Fig. 1

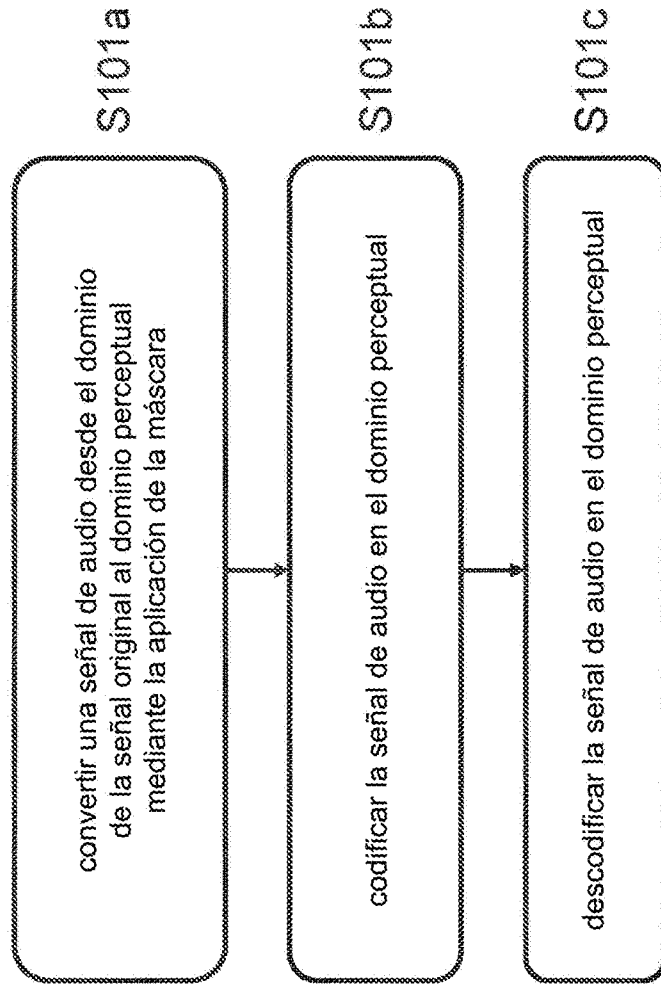


Fig. 2

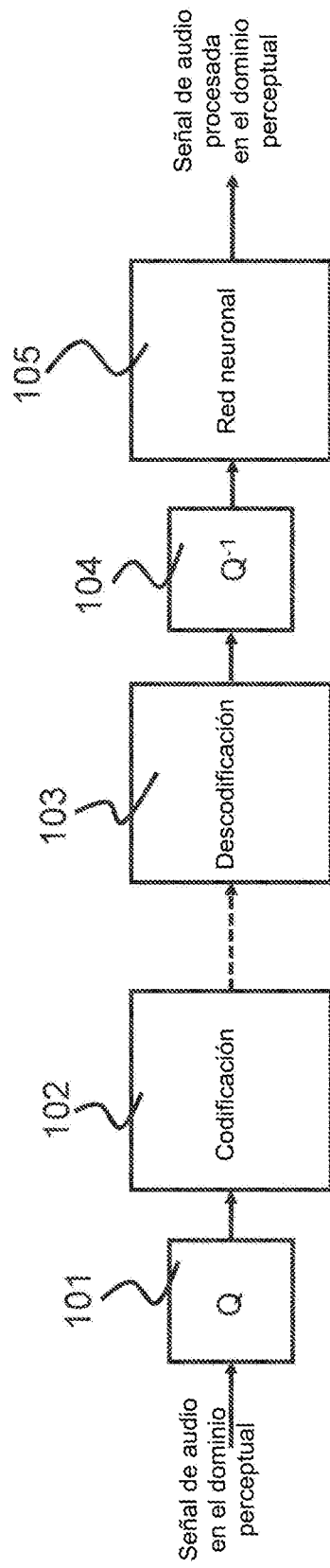


Fig. 3

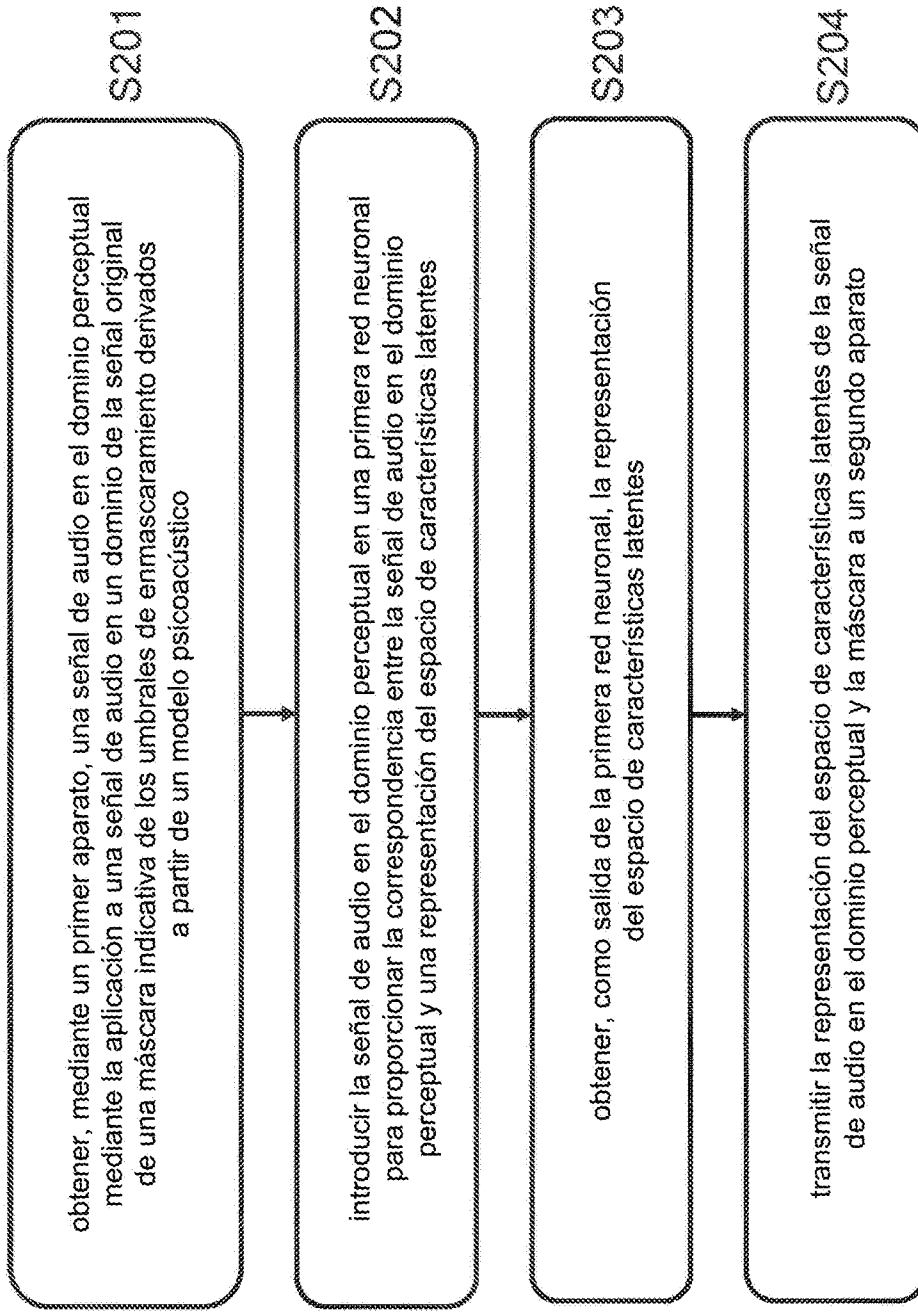


Fig. 4a

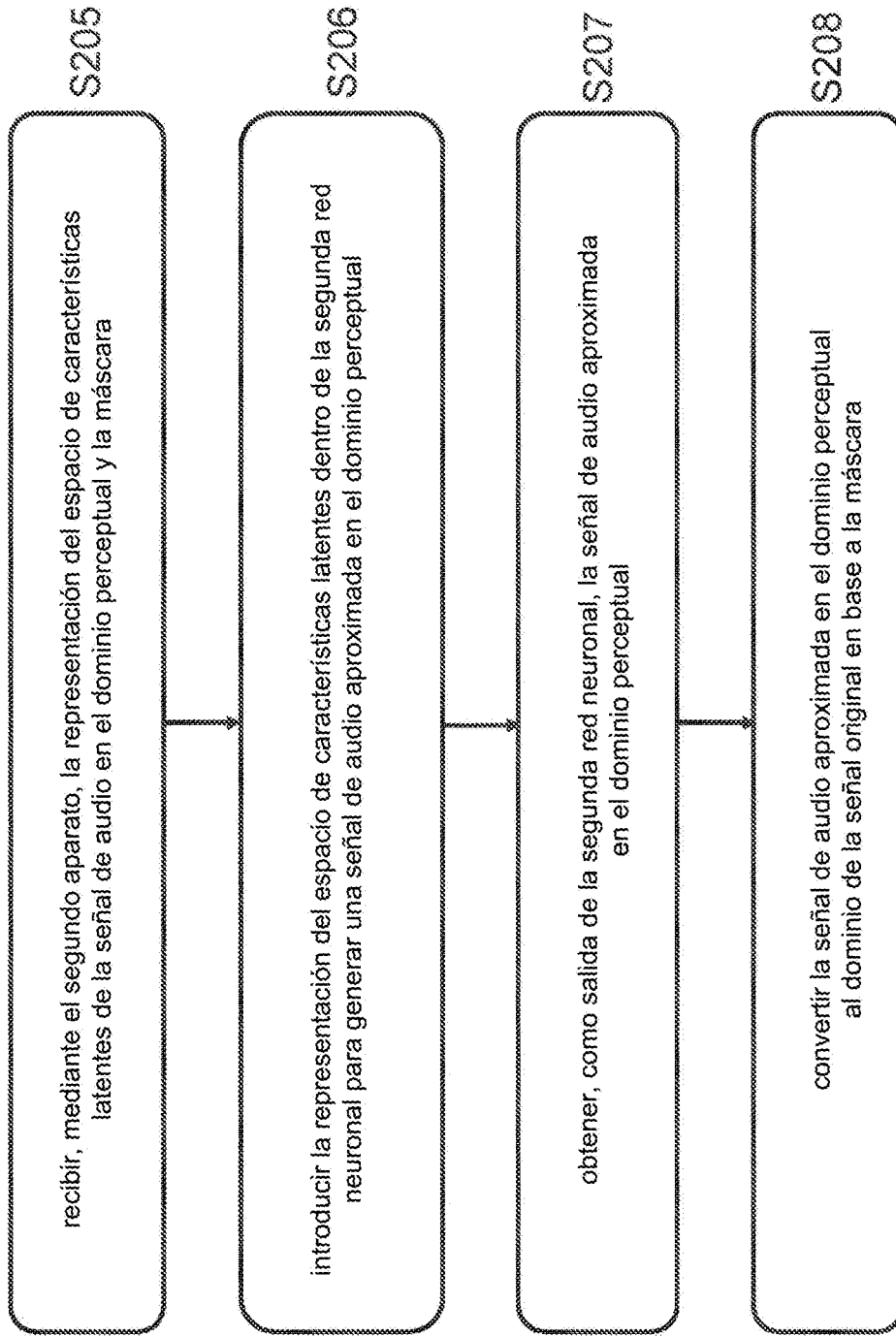


Fig. 4b

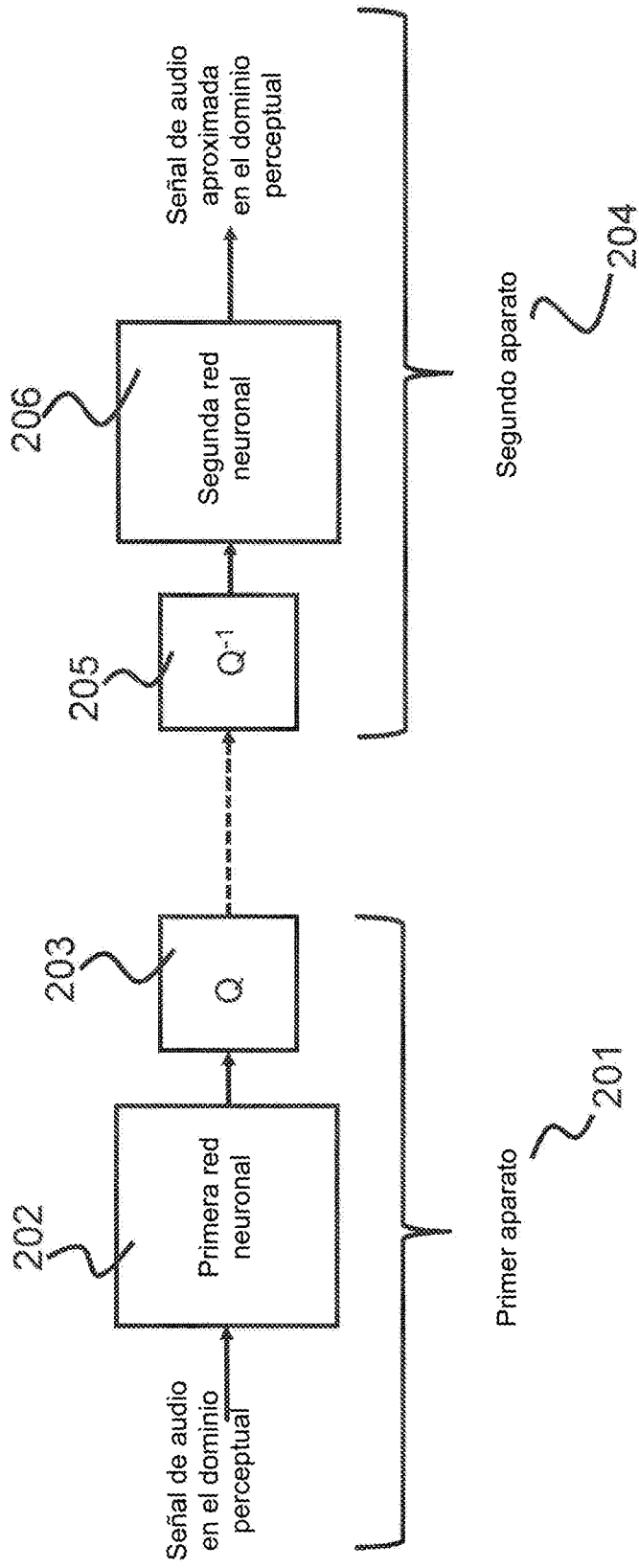


Fig. 5

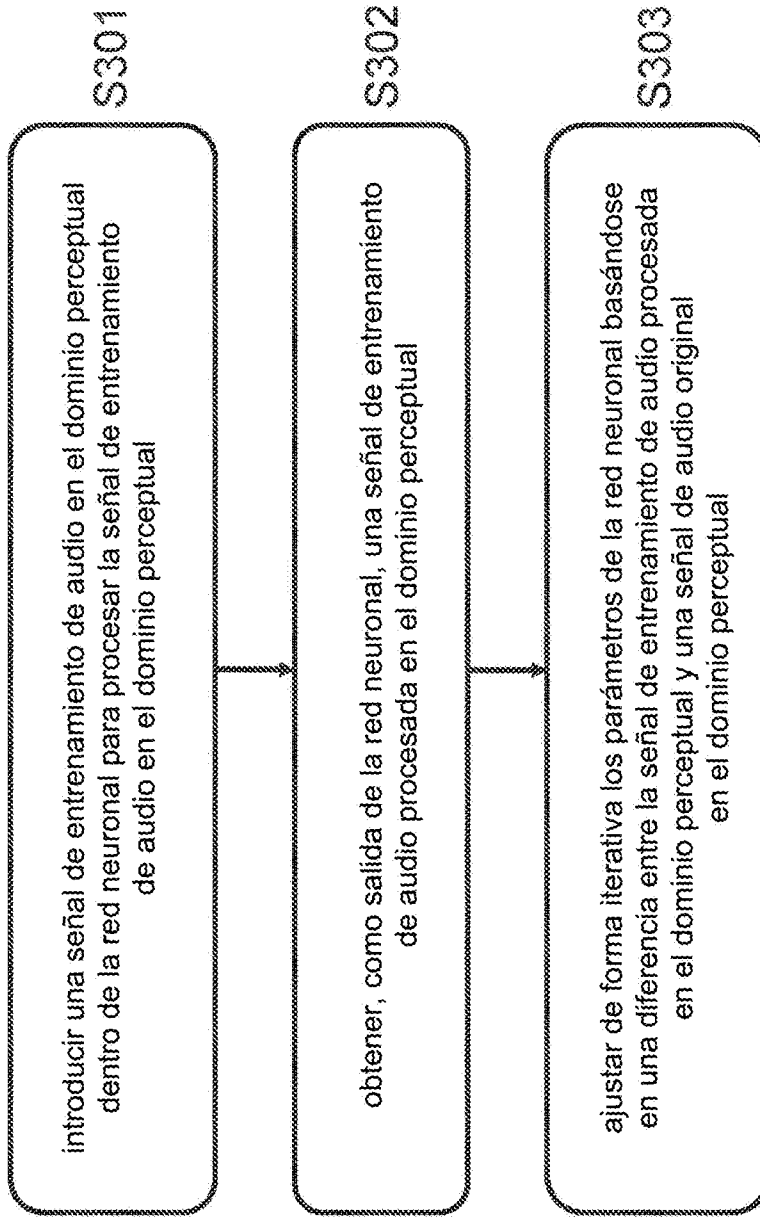


Fig. 6

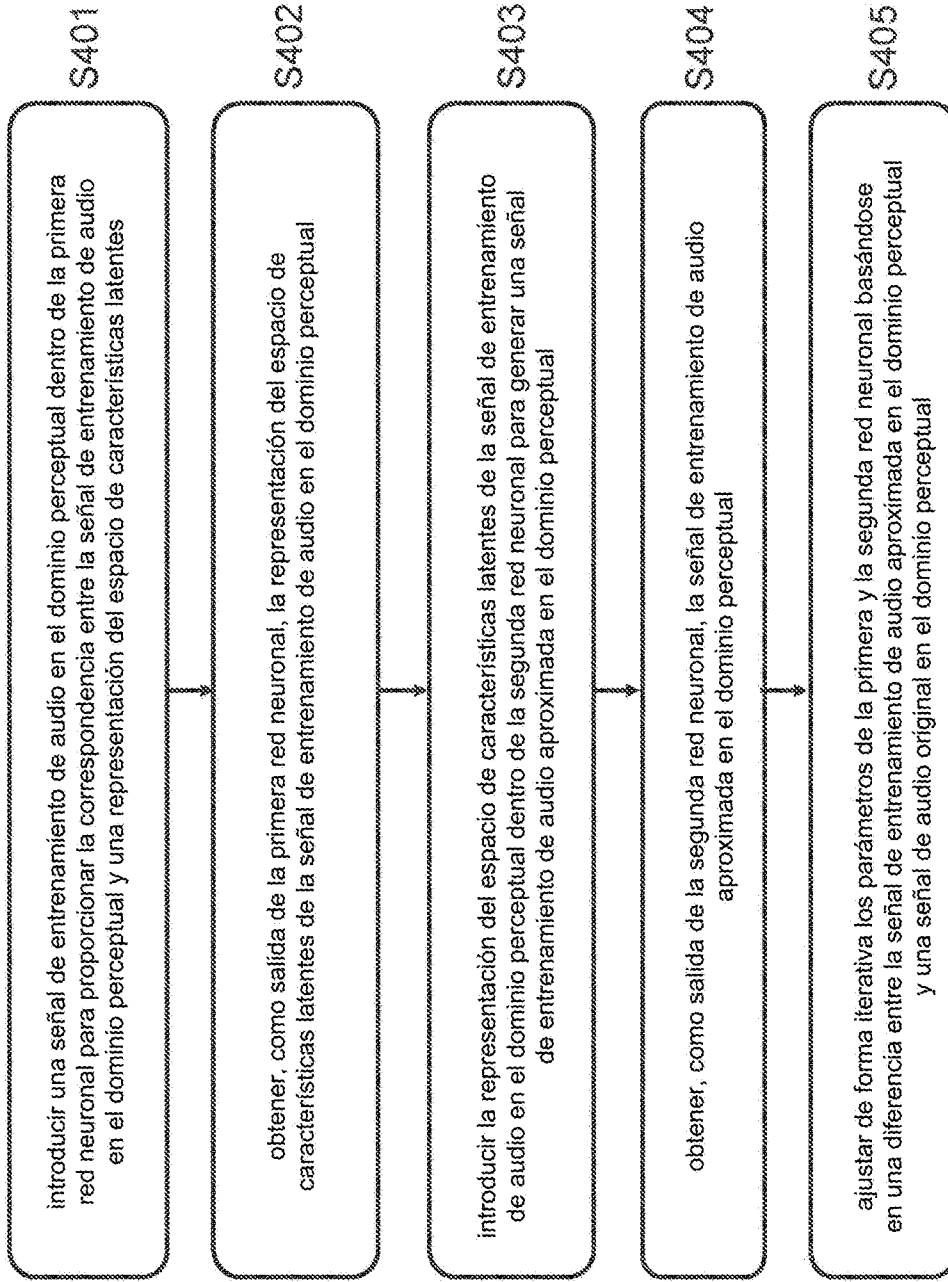


Fig. 7

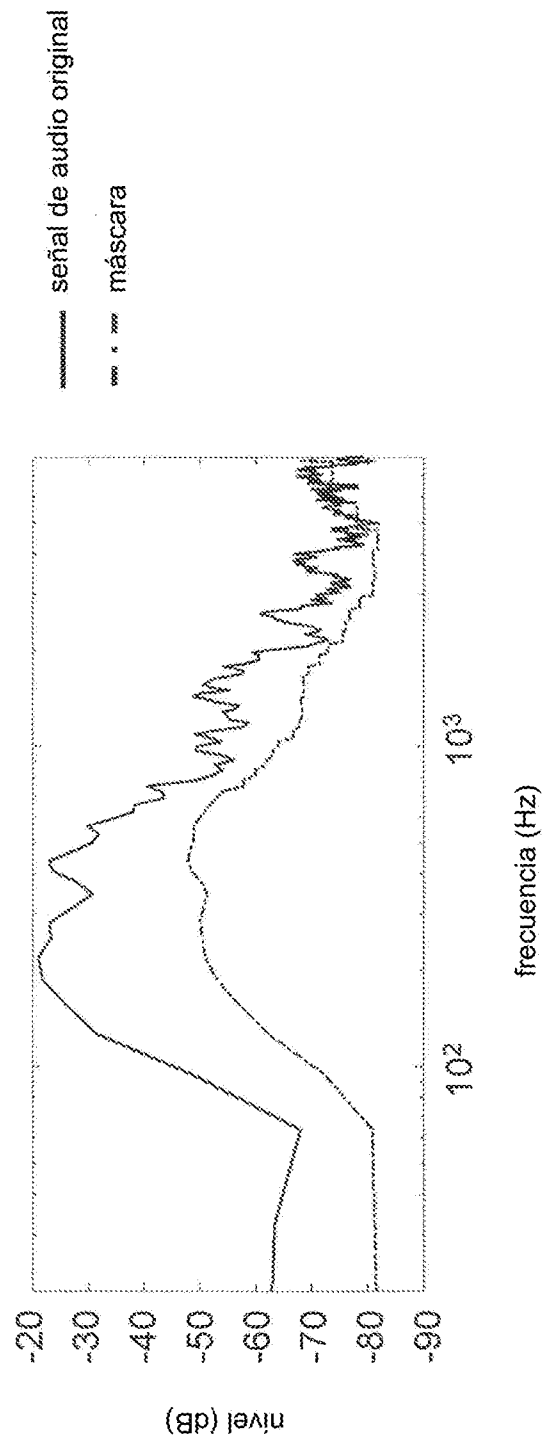


Fig. 8

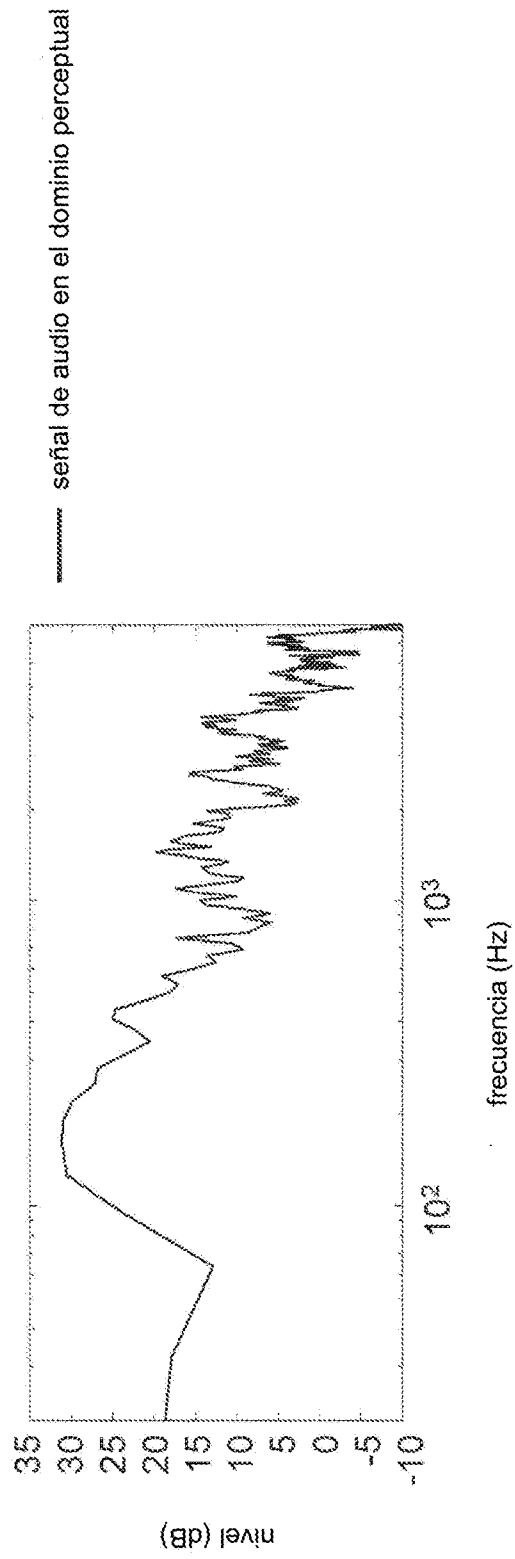


Fig. 9

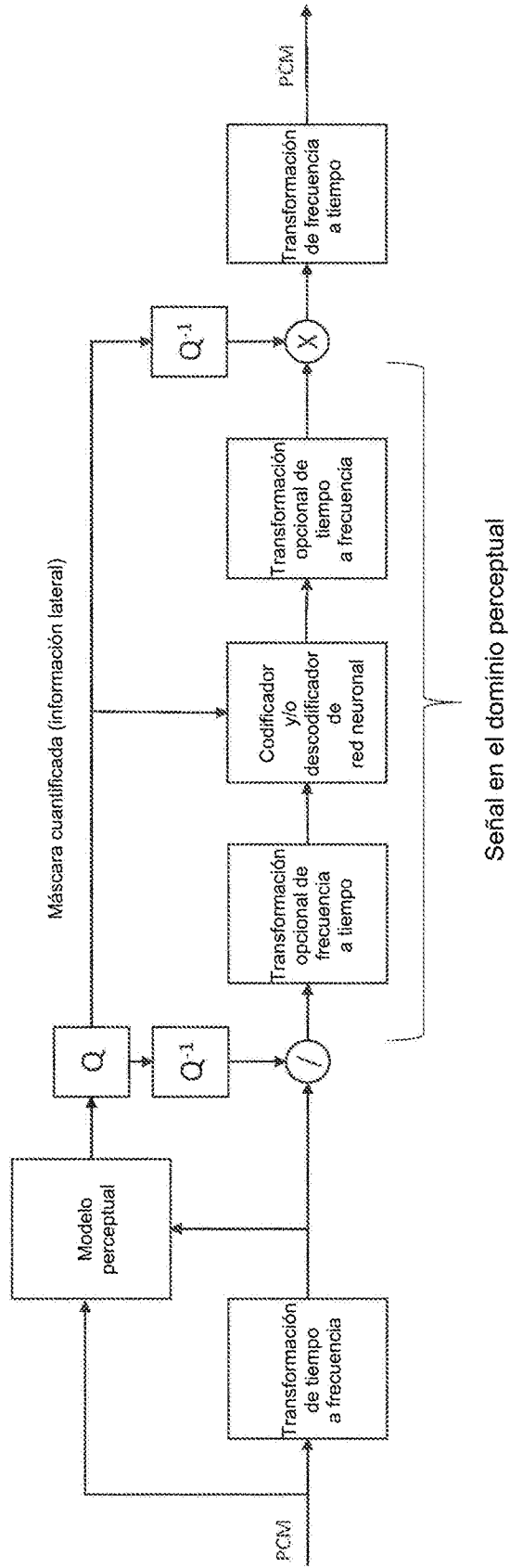


Fig. 10

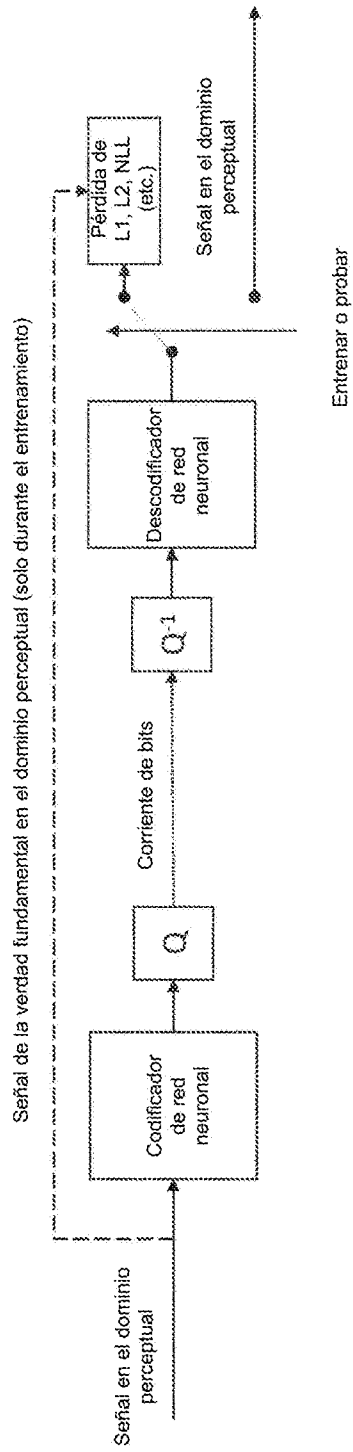


Fig. 11

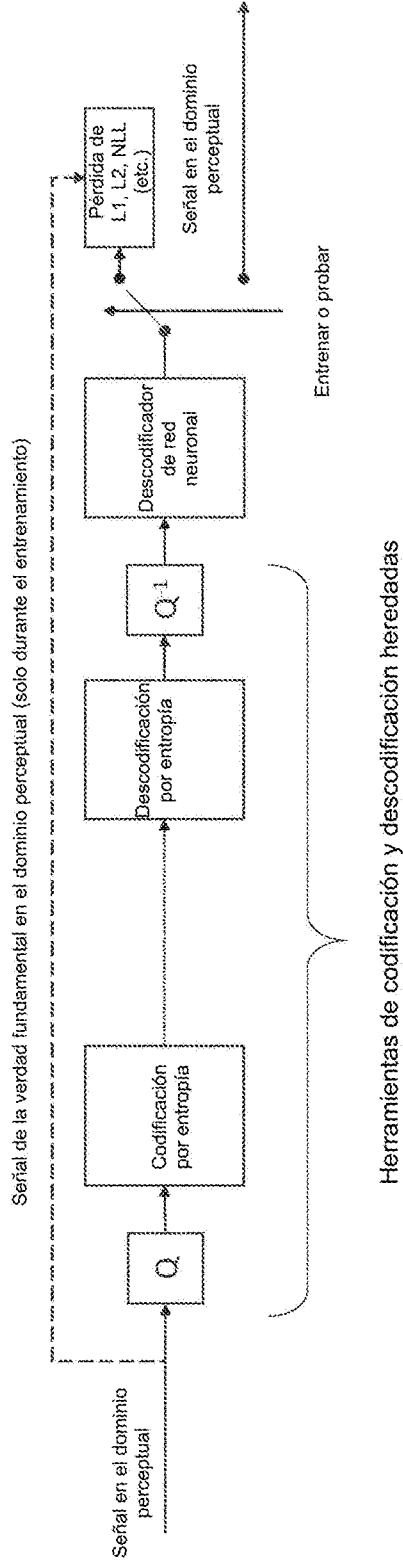


Fig. 12