

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 730 697**

51 Int. Cl.:

G10L 25/69 (2013.01)

G10L 25/18 (2013.01)

H03G 5/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.10.2016 E 16192913 (8)**

97 Fecha y número de publicación de la concesión europea: **13.03.2019 EP 3306608**

54 Título: **Estimación de audibilidad de muestras de audio**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
12.11.2019

73 Titular/es:
BMAT LICENSING, S.L.U. (100.0%)
C. Bruniquer, 49
08024 Barcelona, ES

72 Inventor/es:
MOLINA, EMILIO;
CANO, PEDRO;
CIURANA, ALEX y
LYDA, JOHANNES

74 Agente/Representante:
CONTRERAS PÉREZ, Yahel

ES 2 730 697 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

Estimación de audibilidad de muestras de audio

5

La presente divulgación se refiere a procedimientos para estimar la audibilidad de muestras de audio en programas de medios de radio difusión.

ANTECEDENTES

- 10 El uso de muestras de audio, especialmente muestras de música, en programas de medios de radio difusión es bastante común y puede tener propósitos muy diferentes. En algunos casos, los tipos de audio, por ejemplo, la música, juegan un papel importante en un programa y se percibe claramente en primer plano. Sin embargo, en muchas otras situaciones, dicho audio puede ser apenas perceptible porque tiene un volumen muy bajo o puede estar enmascarado por otros sonidos (por ejemplo, voz). En dichas situaciones, es posible que la música no sea lo
15 suficientemente audible como para tenerla en cuenta con fines de derechos de autor o copyright.

Dado que las muestras de audio (por ejemplo, música, según se ha mencionado) en programas de medios de radio difusión normalmente se mezclan con otros sonidos (por ejemplo, voz), medir su nivel de audibilidad es un reto. Por un lado, el nivel de potencia relativa de la muestra de audio dentro de la mezcla de audio es desconocido a priori. La
20 solución trivial para determinar la potencia relativa de la muestra de audio es medirla en la etapa de producción, en la que las distintas pistas aún son independientes en la mesa de mezclas. Sin embargo, esta medición generalmente no se realiza en la producción de la radio difusión en el mundo real. Un ejemplo de medición de audibilidad en una mezcla de sonidos lo proporciona el documento EP 2 579 247 A1.

- 25 Sería deseable proporcionar un procedimiento alternativo para estimar la audibilidad de muestras de audio en mezclas de audio en programas de medios de radio difusión.

RESUMEN

- Se proponen unos procedimientos para estimar la audibilidad percibida de muestras de audio en programas de
30 medios de radio difusión. De acuerdo con el procedimiento propuesto, se puede procesar una representación aproximada en tiempo-frecuencia (por ejemplo, un output de bancos de filtros) de dos audios involucrados: un programa de medio de radio difusión (que puede comprender una o más mezclas de audio) y una muestra de audio ajustada en ganancia y alineada en el tiempo. La muestra de audio requiere estar alineada en el tiempo con la mezcla de audio, y la potencia de la muestra de audio alineada requiere ser adaptada a la potencia de la música
35 dentro de la mezcla de audio.

Una forma de conseguir esta alineación puede ser anotar información de potencia y de alineación en el tiempo en una etapa de producción, en la que las pistas aún no están mezcladas; Sin embargo, esto no es habitual en los escenarios del mundo real. Existen otras formas, tales como las técnicas de huellas de audio, para conseguir una
40 alineación óptima en el tiempo y un ajuste de ganancia.

Una vez que estas dos representaciones en tiempo-frecuencia (en forma de matriz) están disponibles, pueden ser restadas término a término para generar una nueva matriz, en lo sucesivo denominada matriz de diferencias. Esta matriz de diferencias generalmente tiene una magnitud baja en las regiones de tiempo-frecuencia que no tienen
45 sonidos superpuestos sobre la muestra de audio.

La matriz de diferencias se puede procesar luego utilizando la potencia absoluta de la muestra de audio con el fin de identificar regiones en las que no es audible debido a su baja potencia (incluso sin estar enmascarada). El resultado de este último proceso es una matriz de audibilidad. Esta matriz de audibilidad se puede procesar en intervalos de
50 tiempo de, por ejemplo, 0,5 s para generar una curva promedio que indique la audibilidad de la música a lo largo del tiempo.

En un primer aspecto, se propone un procedimiento de estimar la audibilidad de una muestra de audio en una mezcla de audio de un programa de medio de radio difusión según la reivindicación independiente 1. El
55 procedimiento comprende proporcionar una matriz de representación de una mezcla de audio, comprendiendo la mezcla de audio la muestra de audio y un audio adicional; proporcionar una matriz de representación de la muestra de audio; restar la matriz de representación de la muestra de audio a la matriz de representación de la mezcla de audio para generar una matriz de diferencias; aplicar un modelo de audibilidad a la matriz de diferencias para generar una matriz de audibilidad; determinar un nivel de audibilidad para cada elemento de la matriz de audibilidad;
60 y promediar los niveles de audibilidad determinados de la matriz para estimar la audibilidad de la muestra de audio.

Esto permite determinar si una muestra de audio es audible en la mezcla de audio. A partir de entonces, es posible decidir si la muestra de audio se puede tener en cuenta con fines de atribución de derechos de autor.

En algunos ejemplos, el procedimiento puede comprender además una alineación en el tiempo y un ajuste de ganancia entre la matriz de representación de la muestra de audio y la matriz de representación de la mezcla de audio. Alinear en el tiempo puede comprender alinear en el tiempo usando picos espectrales. El uso de picos espectrales permite una correspondencia de audio simple, resistente al ruido y eficiente, ya que no necesita procesar toda la información de audio, sino solo unos pocos cientos de picos.

El ajuste de ganancia también se puede calcular utilizando picos espectrales, ya que proporcionan información suficiente para relacionar la potencia de los picos en la muestra de audio y la potencia de los picos en la mezcla.

En algunos ejemplos, la matriz de representación puede ser una matriz de representación en tiempo-frecuencia de baja resolución (por ejemplo, un output de bancos de filtros). Esto permite reducir la cantidad de información a procesar y a evitar detalles de alta resolución imperceptibles con afectación al resultado final.

En algunos ejemplos, el modelo de audibilidad puede ser un modelo de contorno de igual sonoridad, tal como un modelo de Fletcher-Munson. Esto permite determinar unos umbrales absolutos de audibilidad con mayor precisión, ya que tiene en cuenta los factores psico acústicos.

En algunos ejemplos, el procedimiento puede comprender además aplicar un modelo de enmascaramiento psico acústico a la mezcla de audio para generar la matriz de representación de la mezcla de audio. Esto permite producir un valor de audibilidad más preciso, ya que tiene en cuenta efectos de enmascaramiento que podrían reducir la audibilidad real de ciertas frecuencias.

En algunos ejemplos, el promedio se puede realizar en intervalos de tiempo para estimar la audibilidad de la muestra de audio en la mezcla de audio en cada intervalo de tiempo. Esto permite determinar partes de una muestra de audio que pueden ser más audibles y partes de la muestra de audio que pueden ser menos audibles. En base a la determinación, puede tener lugar un porcentaje de atribución de derechos de autor para cada intervalo de tiempo.

En algunos ejemplos, el procedimiento puede comprender además cuantificar la audibilidad estimada en un número predeterminado de niveles de audibilidad. Por ejemplo, se pueden establecer cuatro niveles de audibilidad y se puede atribuir la muestra de audio o partes de la muestra de audio a uno de los cuatro niveles.

En algunos ejemplos, la muestra de audio puede ser música y el audio adicional puede ser voz. Este es un escenario típico en programas de medios de radio difusión, tales como programas de entrevistas o programas de noticias, en los que la música se reproduce en el fondo de una discusión o una presentación.

En algunos ejemplos, el procedimiento puede comprender además proporcionar un programa de medio de radio difusión y proporcionar una hoja cue. El programa de medio de radio difusión puede comprender la mezcla de audio y la hoja cue puede comprender una indicación de la hora de inicio y la hora de finalización de la muestra de audio en el programa de medio de radio difusión. Esto puede ser proporcionado en una etapa de post producción. El uso de hojas cue permite automatizar el recorte de audio para estimar la audibilidad de cada muestra de audio individual.

En algunos ejemplos, la mezcla de audio puede comprender una pluralidad de muestras de audio y un audio adicional. El procedimiento puede comprender además recortar la mezcla de audio para aislar la parte de la mezcla de audio correspondiente a una sola muestra de audio. Esto puede permitir procesar y estimar la audibilidad de cada muestra de audio de la mezcla de audio.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

A continuación, se describirán ejemplos no limitativos de la presente divulgación, con referencia a los dibujos adjuntos, en los que:

La Figura 1 ilustra esquemáticamente un diagrama de flujo de un procedimiento de estimar la audibilidad de una muestra de audio en una mezcla de audio de un programa de medio de radio difusión según un ejemplo;

La Figura 2A ilustra esquemáticamente una señal de audio con partes de señal de audibilidad variable;

La Figura 2B ilustra esquemáticamente una curva de audibilidad;

La Figura 3 ilustra esquemáticamente un diagrama de flujo de un procedimiento de estimar la audibilidad de una muestra de audio en una mezcla de audio de un programa de medio de radio difusión según otro ejemplo;

La Figura 4 ilustra esquemáticamente un diagrama de flujo de un procedimiento de estimar la audibilidad de una muestra de audio en una mezcla de audio de un programa de medio de radio difusión según otro ejemplo;

La Figura 5 ilustra esquemáticamente un proceso automático de hojas cue de radio difusión para estimar la audibilidad de varias muestras de audio de acuerdo con un ejemplo.

5 DESCRIPCIÓN DETALLADA DE EJEMPLOS

La Figura 1 ilustra esquemáticamente un diagrama de flujo de un procedimiento de estimar la audibilidad de una muestra de audio en una mezcla de audio de un programa de medio de radio difusión según un ejemplo. Se puede proporcionar una matriz de representación en tiempo-frecuencia de una mezcla de audio en el bloque 105. Se puede proporcionar una matriz de representación en tiempo-frecuencia alineada de una muestra de audio en el bloque 110.

- 10 La muestra de audio puede ser música o una canción. Puede suponerse que la mezcla de audio comprende una copia de la muestra de audio junto con otro audio adicional, por ejemplo, voz. Luego se puede realizar una estimación de audibilidad 100. En el bloque 115, la matriz de representación de la muestra de audio puede ser restada de la matriz de representación de la mezcla de audio para generar una matriz de diferencias. Una vez que se ha generado la matriz de diferencias, se puede usar la información proporcionada por la matriz de diferencias
- 15 para evitar que las regiones de muy baja potencia (incluso si no están enmascaradas) se consideren inaudibles. Para este propósito, se puede aplicar en el bloque 120 un modelo de audibilidad, por ejemplo: curvas de Fletcher-Munson, que se basa en experimentos psico acústicos, para determinar un umbral de audibilidad absoluto. Esto puede proporcionar una matriz de audibilidad como output. Luego, en el bloque 125, se pueden identificar las diversas regiones de tiempo-frecuencia potencialmente audibles, proporcionando así una curva de audibilidad 130.

- 20 La Figura 2A ilustra esquemáticamente una señal de audio con partes de señal de audibilidad variable. Una primera y una cuarta regiones A y D, respectivamente, se pueden caracterizar como regiones de audibilidad baja, una segunda región B como una región de audibilidad media y una tercera y una quinta regiones C y E, respectivamente, como no audibles. La Figura 2B ilustra esquemáticamente una curva de audibilidad. El eje horizontal representa el
- 25 tiempo y el eje vertical el porcentaje de audibilidad desde 0% hasta 100%.

La audibilidad promedio se puede calcular promediando la curva de audibilidad. Este promedio puede entonces cuantificarse en, por ejemplo, cuatro categorías:

- 30 Audibilidad entre 0% y 10%: no audible
 Audibilidad entre 10% y 30%: audibilidad baja
 Audibilidad entre 30% y 70%: audibilidad media
 Audibilidad entre 70% y 100%: audibilidad alta

- 35 La Figura 3 ilustra esquemáticamente un diagrama de flujo de un procedimiento de estimar la audibilidad de una muestra de audio en una mezcla de audio de un programa de medio de radio difusión según otro ejemplo. En el ejemplo de la Figura 3, se puede usar un modelo de enmascaramiento psico acústico para procesar la representación en tiempo-frecuencia de la mezcla de audio antes de realizar la resta. De manera similar al ejemplo comentado con referencia a la Figura 1, se puede proporcionar una matriz de representación de una mezcla de
- 40 audio en el bloque 305. Luego, se puede realizar una estimación de la audibilidad 300. En el bloque 307, se puede usar un modelo de enmascaramiento psico acústico para procesar la representación en tiempo-frecuencia de la mezcla de audio. El modelo de enmascaramiento puede identificar los componentes tonales y no tonales de la señal de audio, y se aplica una función de enmascaramiento alrededor de cada uno de ellos. La suma de todas las funciones de enmascaramiento puede determinar el umbral de enmascaramiento, que puede ser concatenado
- 45 fotograma a fotograma con el fin de crear el umbral de enmascaramiento en tiempo-frecuencia en forma de matriz según se especifica, por ejemplo, en la norma ISO/IEC 11172-3: 1993. El resultado puede ser una representación en tiempo-frecuencia del umbral de enmascaramiento. Cualquier sonido por debajo de este umbral se puede considerar enmascarado y, por lo tanto, se puede considerar inaudible. En el ejemplo de la Figura 3, la matriz de representación de la mezcla de audio es reemplazada por la representación en tiempo-frecuencia del umbral de enmascaramiento
- 50 de la mezcla. De esta manera, se puede mejorar la estimación de la audibilidad para los casos en los que el enmascaramiento auditivo es perceptible. En el bloque 310 se puede proporcionar una representación matricial en tiempo-frecuencia alineada de una muestra de audio. La muestra de audio puede ser música o una canción. Luego, en el bloque 315, la matriz de representación de la muestra de audio puede ser restada de la matriz de umbral de enmascaramiento en tiempo-frecuencia de la mezcla de audio para generar una matriz de diferencias. Luego,
- 55 nuevamente, de manera similar al ejemplo anterior, se puede aplicar un modelo de audibilidad en el bloque 320 para determinar un umbral de audibilidad absoluto. Esto puede proporcionar una matriz de audibilidad como output. Luego, en el bloque 325, las diversas regiones de tiempo-frecuencia potencialmente audibles pueden ser identificadas en la matriz de audibilidad. Entonces se puede generar una curva de audibilidad 330.

- 60 La Figura 4 ilustra esquemáticamente un diagrama de flujo de un procedimiento de estimar la audibilidad de una muestra de audio en una mezcla de audio de un programa de medio de radio difusión según otro ejemplo. En el ejemplo de la Figura 4, se usan los picos espectrales (también llamados puntos de referencia) para encontrar el ajuste de ganancia y el desplazamiento en tiempo de la muestra de audio en la mezcla de audio. Estas técnicas se

aplican normalmente en la determinación de huellas de audio, y permiten encontrar la correspondencia o coincidencia óptima de un conjunto de puntos entre una consulta (es decir, la mezcla de audio) y una referencia (es decir, la muestra de audio). Se puede usar cualquier conjunto de picos espectrales, o puntos de referencia, capaces de resumir el contenido de audio original. En su definición más simple, los picos espectrales pueden ser máximos locales en la representación en tiempo-frecuencia (por ejemplo, transformada de Fourier de tiempo corto, transformada de Q constante, etc.). Estos picos espectrales se pueden usar para encontrar la alineación óptima entre dos audios que contienen la misma información (correspondencia o coincidencia). La alineación se puede realizar buscando picos con la misma frecuencia en la consulta y la referencia, y el mismo desplazamiento de tiempo relativo. Esta información coincidente es almacenada como una lista de picos comunes. Esta información se puede usar para calcular la alineación en el tiempo y la relación de amplitud entre la referencia y la consulta, en la que:

Alineación tiempo = ubicación tiempo en consulta – ubicación tiempo en referencia

y

Relación amplitud = amplitud en consulta/amplitud en referencia

Esta información se puede usar para alinear en el tiempo y ajustar la ganancia de la muestra con el fin de que suene como suena en la mezcla.

En la Figura 4, se proporciona la mezcla de audio en el bloque 405 y la muestra de audio en el bloque 410. Luego, se puede realizar la estimación de la alineación en el tiempo y la potencia relativa de la muestra de audio en la mezcla de audio en el bloque 400. En el bloque 415, se alinean las dos señales en el tiempo usando picos espectrales. El bloque 415 puede proporcionar la posición exacta en el tiempo de la muestra de audio en la mezcla de audio y la potencia de la muestra de audio en la mezcla. Luego, se proporcionan el vector de la señal de muestra de audio, la posición exacta en el tiempo de la muestra de audio en la mezcla de audio y la potencia de la muestra de audio en la mezcla al bloque 420, en el que se genera un tiempo desplazado y una señal de audio vectorial ajustada. En el bloque 425, se genera una representación matricial en tiempo-frecuencia 430 del vector de señal de audio. Ésta ya viene alineada en el tiempo del bloque 415. En paralelo, en el bloque 435 se genera una representación matricial en tiempo-frecuencia 440 del vector de señal de mezcla de audio. Luego, se proporcionan las dos matrices al bloque 450 en el que puede tener lugar la estimación de la audibilidad de acuerdo con los ejemplos proporcionados en este documento y luego se puede proporcionar una curva de audibilidad 460.

La Figura 5 ilustra esquemáticamente un proceso automático de hojas cue de radio difusión para estimar la audibilidad de varias muestras de audio de acuerdo con un ejemplo. En el bloque 505, se puede proporcionar el archivo de audio de un programa de medio de radio difusión como un vector de señal de audio largo. En el bloque 510, se pueden proporcionar los archivos de audio de las muestras de audio que pueden estar presentes en el programa de medio de radio difusión como un vector de señal de audio. En el bloque 515 se puede proporcionar la hoja cue. Los vectores se pueden recortar con el uso de la hoja cue en el bloque 520. El recorte puede generar un primer conjunto de vectores de señal de mezcla de audio 522n y un segundo conjunto de vectores de muestra de audio 524n. Los elementos de los conjuntos se pueden corresponder entre sí. Para cada par de vector de señal de mezcla de audio y vector de señal de muestra de audio, el proceso puede continuar en el bloque 525 con la estimación de la alineación en el tiempo y la potencia relativa de la muestra de audio en la mezcla de audio. Luego, en los bloques 530 y 535, se puede generar una matriz de representación en tiempo-frecuencia de la mezcla de audio y de la muestra de audio, respectivamente. Finalmente, en el bloque 540, puede tener lugar la estimación de la audibilidad utilizando procedimientos como los descritos en este documento para generar la curva de audibilidad 550.

Aunque solo se han descrito varios ejemplos en este documento, son posibles otras alternativas, modificaciones, usos y/o equivalentes de los mismos. Además, también se cubren todas las combinaciones posibles de los ejemplos descritos. Por lo tanto, el alcance de la presente divulgación no debe ser limitado por ejemplos particulares, sino que se debe determinar solo por medio de una lectura imparcial de las siguientes reivindicaciones. Los signos de referencia relacionados con los dibujos que se encuentran entre paréntesis en una reivindicación son únicamente para intentar aumentar la inteligibilidad de la reivindicación y no se deben interpretar como limitantes del alcance de la reivindicación.

Además, aunque los ejemplos descritos con referencia a los dibujos comprenden aparatos/sistemas informáticos y procesos realizados en aparatos/sistemas informáticos, la invención también se extiende a programas informáticos, particularmente programas informáticos en un portador, adaptados para poner en práctica el sistema.

REIVINDICACIONES

1. Un procedimiento para estimar la audibilidad de una muestra de audio en una mezcla de audio de un programa de medio de radio difusión, que comprende:
 - 5 proporcionar una matriz de representación de una mezcla de audio, comprendiendo la mezcla de audio la muestra de audio y un audio adicional;
proporcionar una matriz de representación de la muestra de audio;
restar la matriz de representación de la muestra de audio a la matriz de representación de la mezcla de audio para generar una matriz de diferencias;
 - 10 aplicar un modelo de audibilidad a la matriz de diferencias para generar una matriz de audibilidad;
determinar un nivel de audibilidad para cada elemento de la matriz de audibilidad;
promediar los niveles de audibilidad determinados de la matriz para estimar la audibilidad de la muestra de audio.
2. El procedimiento según la reivindicación 1, que comprende además una alineación en el tiempo entre la matriz de representación de la muestra de audio y la matriz de representación de la mezcla de audio.
15
3. El procedimiento según la reivindicación 2, en el que alinear en el tiempo comprende alinear en el tiempo usando picos espectrales.
- 20 4. El procedimiento según cualquiera de las reivindicaciones 1 a 3, que comprende además un ajuste de ganancia entre la matriz de representación de la muestra de audio y la matriz de representación de la mezcla de audio.
5. El procedimiento según cualquiera de las reivindicaciones 1 a 4, en el que la matriz de representación es una matriz de representación en tiempo-frecuencia de baja resolución.
25
6. El procedimiento según cualquiera de las reivindicaciones 1 a 5, en el que el modelo de audibilidad es un modelo de contorno de igual sonoridad.
7. El procedimiento según cualquiera de las reivindicaciones 1 a 6, en el que el modelo de audibilidad es un modelo de Fletcher-Munson.
30
8. El procedimiento según cualquiera de las reivindicaciones 1 a 7, que comprende además aplicar un modelo de enmascaramiento psico acústico a la mezcla de audio para generar la matriz de representación de la mezcla de audio.
35
9. El procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que el promediado comprende promediar en intervalos de tiempo para estimar la audibilidad de la muestra de audio en la mezcla de audio en cada intervalo de tiempo.
- 40 10. El procedimiento según cualquiera de las reivindicaciones 1 a 9, que comprende además cuantificar la audibilidad estimada en un número predeterminado de niveles de audibilidad.
11. El procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que la muestra de audio es música y el audio adicional es voz.
45
12. El procedimiento según cualquiera de las reivindicaciones 1 a 11, que comprende además
proporcionar un programa de medio de radio difusión, comprendiendo el programa de medio de radio difusión la mezcla de audio,
proporcionar una hoja cue, comprendiendo la hoja cue una indicación del momento de inicio y del momento de finalización de la muestra de audio en el programa de medio de radio difusión.
50
13. El procedimiento según la reivindicación 12, en el que la mezcla de audio comprende una pluralidad de muestras de audio y un audio adicional.
- 55 14. El procedimiento según la reivindicación 13, que comprende además recortar la mezcla de audio para aislar cada muestra de audio.
15. El procedimiento según la reivindicación 14, que comprende además estimar la audibilidad de cada muestra de audio del programa de radio difusión.
60

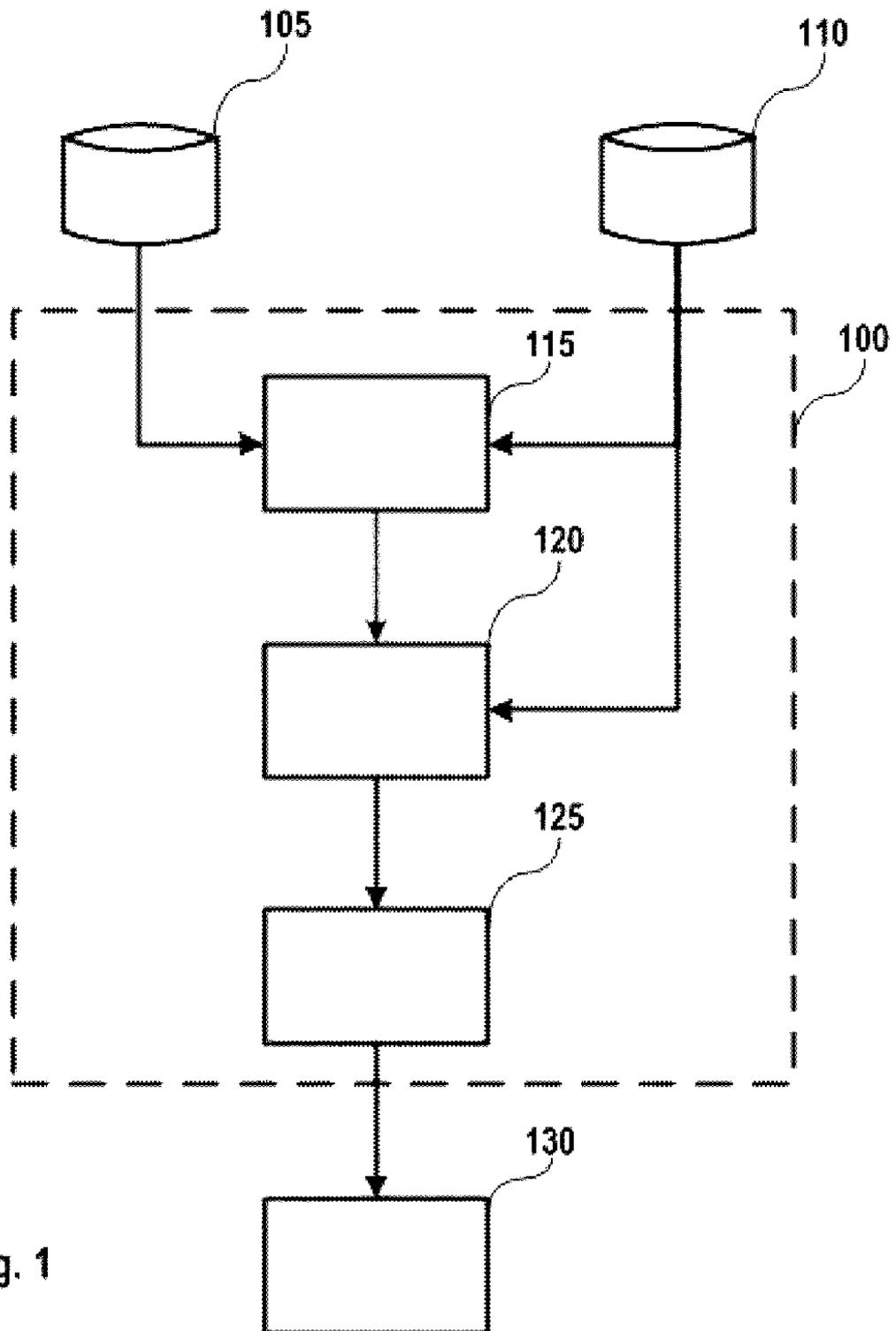


Fig. 1

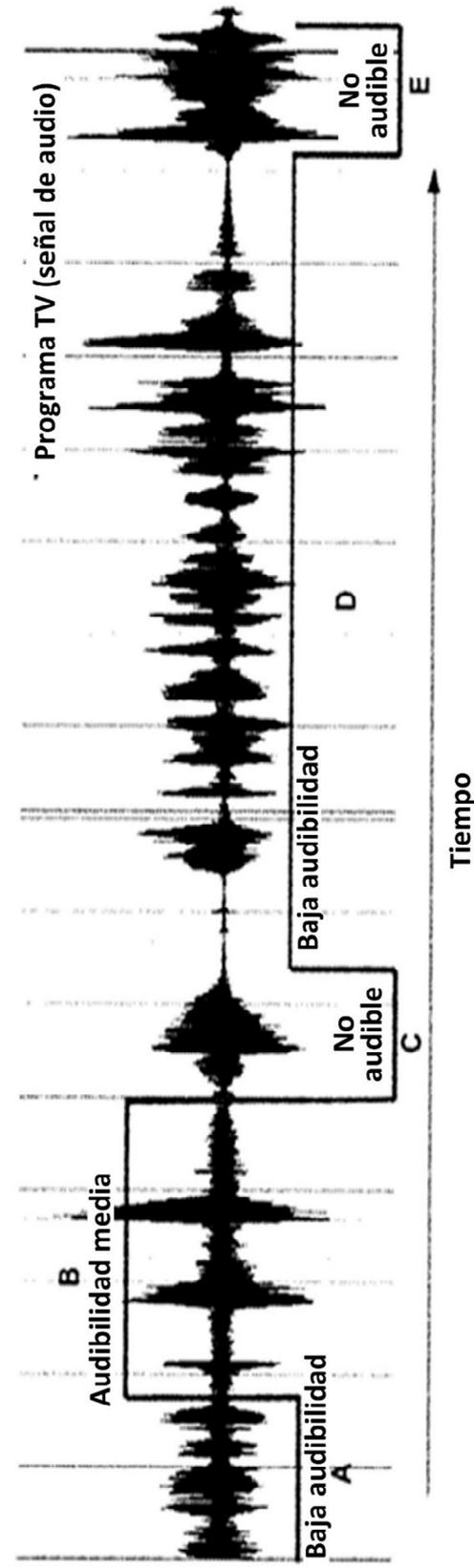


Fig. 2A

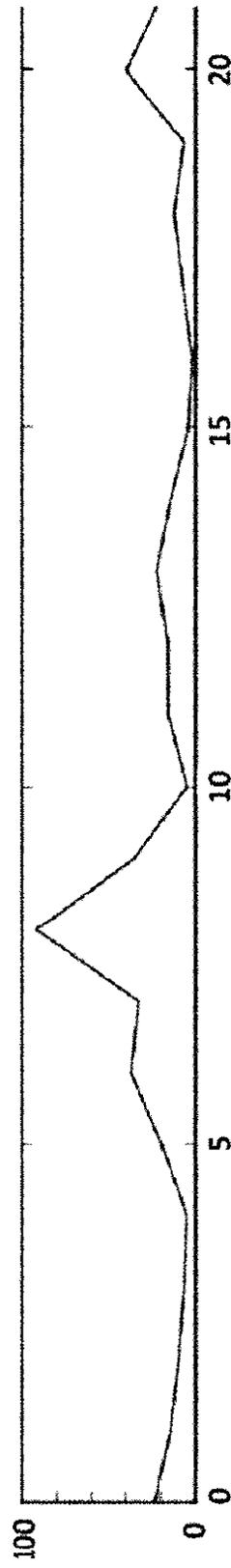
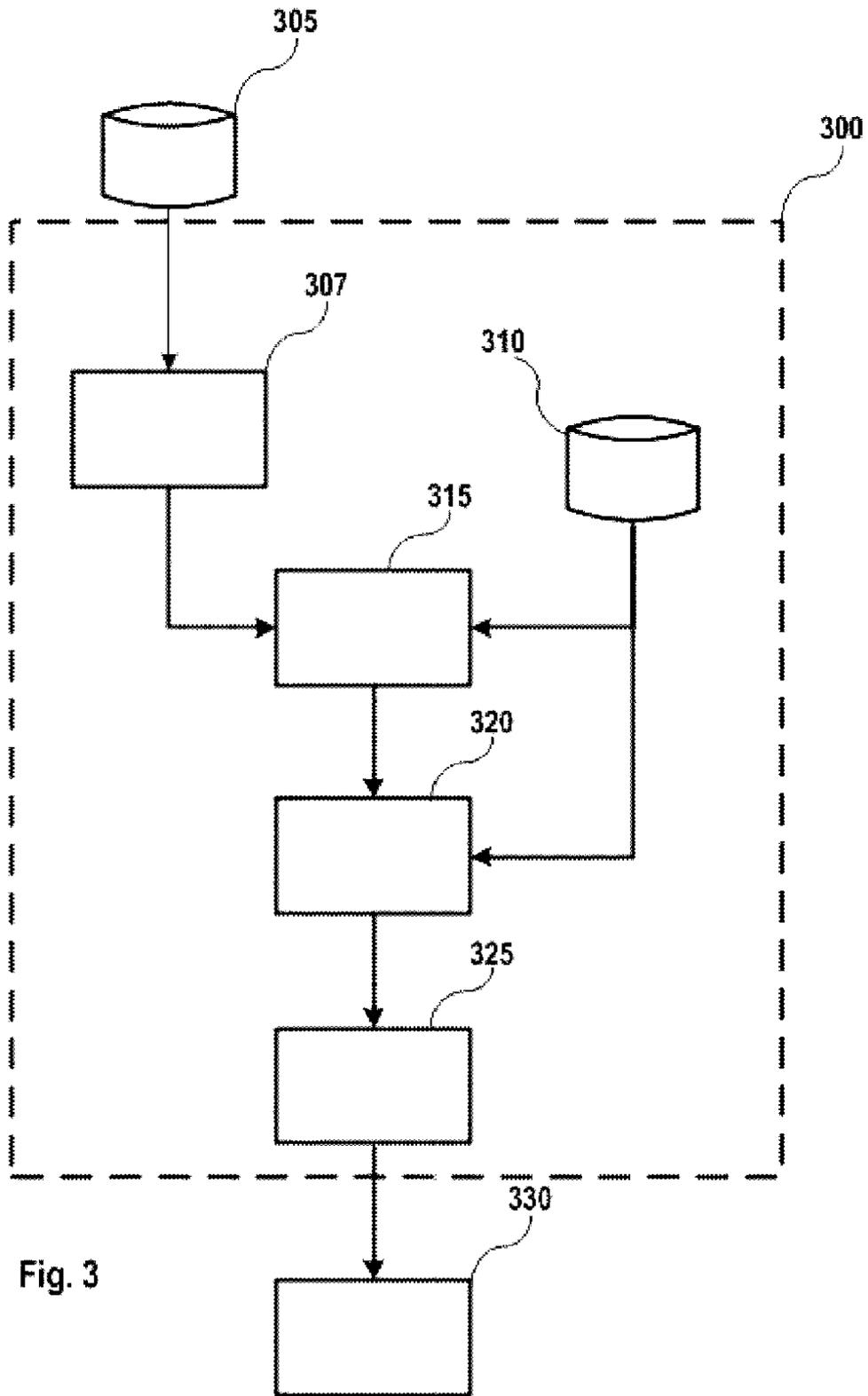


Fig. 2B



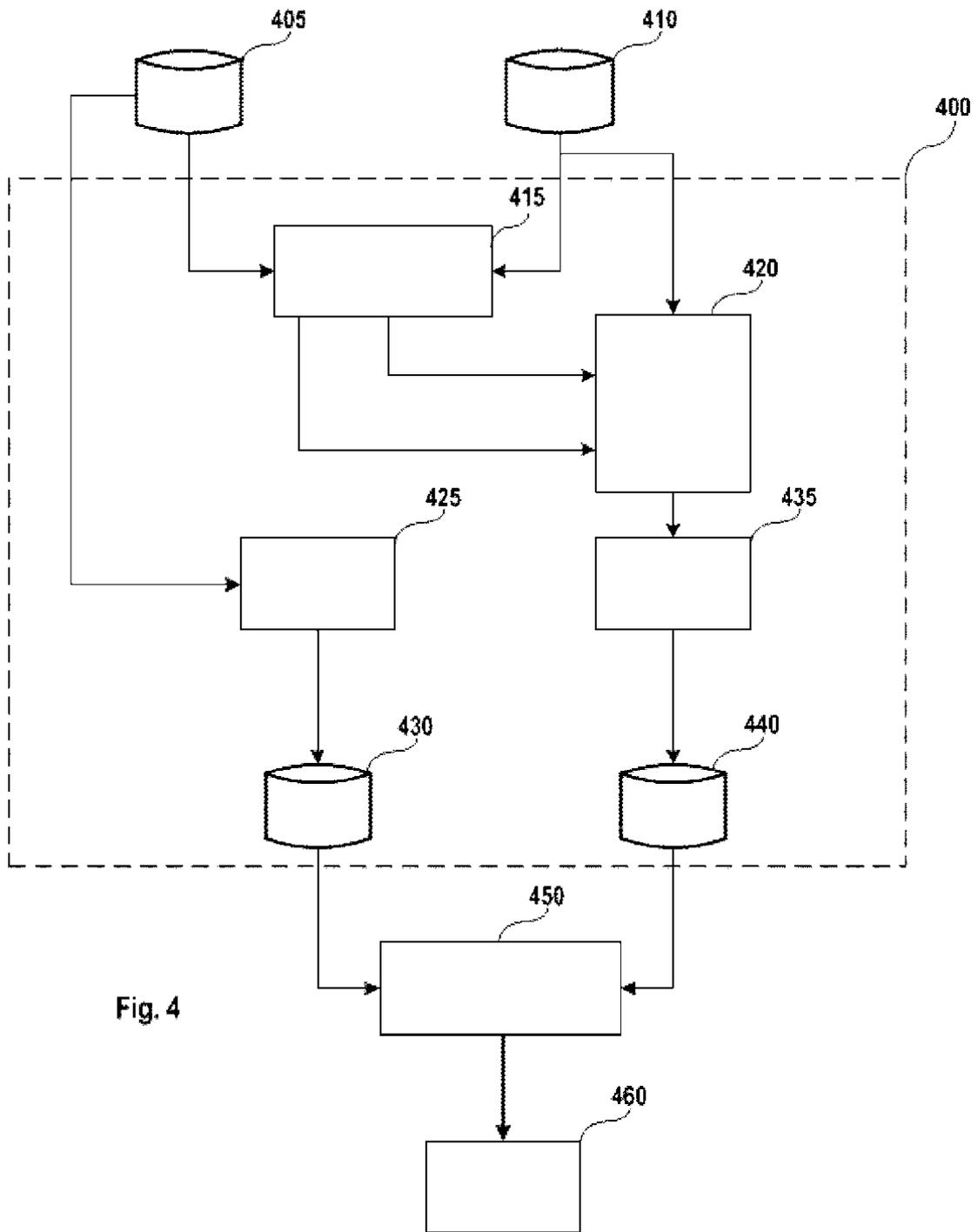


Fig. 4

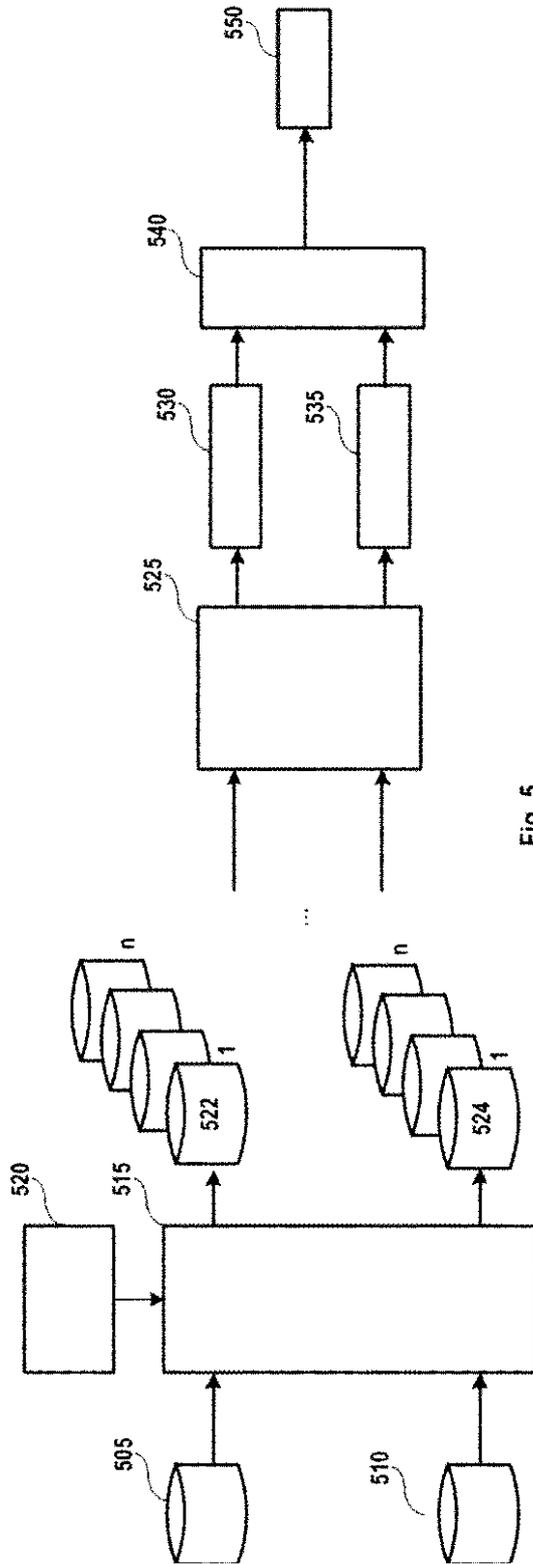


Fig. 5