(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 980 796**

(51) Int. Cl.:

G10L 19/16 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **06.10.2015 E 22166776 (9)**

(97) Fecha y número de publicación de la concesión europea: **24.04.2024 EP 4060661**

(54) Título: **Sonoridad de programa basada en la presentación, independiente de la transmisión**

(30) Prioridad:

10.10.2014 US 201462062479 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
03.10.2024

(73) Titular/es:

DOLBY LABORATORIES LICENSING CORPORATION (50.0%)
1275 Market Street
San Francisco, CA 94103, US y
DOLBY INTERNATIONAL AB (50.0%)

(72) Inventor/es:

KOPPENS, JEROEN y
NORCROSS, SCOTT GREGORY

(74) Agente/Representante:

LINAGE GONZÁLEZ, Rafael

DESCRIPCIÓN

Sonoridad de programa basada en la presentación, independiente de la transmisión

Referencia cruzada a solicitudes relacionadas

Esta solicitud reivindica la prioridad de la solicitud de patente provisional de EE. UU. núm. 62/062.479, presentada el 10 de octubre de 2014.

Esta solicitud es una solicitud divisional europea de la solicitud de patente Euro-PCT EP18209378.1 (referencia: D14109EP02), cuyo formulario 1011 de la OEP fue presentado el 30 de noviembre de 2018.

Campo técnico

La invención se refiere al procesamiento de señales de audio, y, más particularmente, a la codificación y decodificación de flujos de bits de datos de audio con el fin de lograr el nivel de sonoridad deseado de una señal de audio de salida.

Antecedentes de la técnica

Dolby AC-4 es un formato de audio para distribuir eficientemente contenido multimedia enriquecido. AC-4 proporciona un marco flexible para que las emisoras y los productores de contenido distribuyan y codifiquen contenido de manera eficiente. El contenido se puede distribuir en varios flujos secundarios, como, por ejemplo, M&E (música y efectos) en un flujo secundario, y diálogo en un segundo flujo secundario. Para algunos contenidos de audio, puede ser ventajoso, por ejemplo, cambiar el idioma del diálogo de un idioma a otro, o poder añadir, por ejemplo, un flujo secundario de comentarios al contenido o un flujo secundario adicional que comprenda una descripción para personas con problemas de visión.

Para garantizar una nivelación adecuada del contenido presentado al consumidor, es necesario conocer la sonoridad del contenido con cierto grado de precisión. Los requisitos de sonoridad actuales tienen tolerancias de 2 dB (ATSC A/85), 0,5 dB (EBU R128), mientras que algunas especificaciones tienen tolerancias tan bajas como 0,1 dB. Esto significa que la sonoridad de una señal de audio de salida con una pista de comentarios y con diálogo en un primer idioma debería ser substancialmente el mismo que la sonoridad de una señal de audio de salida sin la pista de comentarios y con diálogo en un segundo idioma.

El informe de búsqueda europeo emitido en relación con el presente documento hace referencia a “Digital Audio Compression (AC-4) Standard”, 1 de abril de 2014 (2014-04-01) páginas 1-295, en adelante “documento D1”. El documento D1 especifica una presentación codificada de información de audio, y especifica el proceso de decodificación. La presentación codificada especificada en él es adecuada para su uso en aplicaciones de almacenamiento y transmisión de audio digital. La presentación codificada puede transmitir señales de audio de ancho de banda completo, junto con una señal de perfeccionamiento de baja frecuencia, para reproducción multicanal. Se pueden incluir presentaciones adicionales, dirigidas, por ejemplo, a oyentes con discapacidad visual o auditiva. Los decodificadores implantados de acuerdo con el documento D1 soportan una amplia gama de velocidades de bits codificadas, que van desde la compresión del estado de la técnica hasta velocidades perceptualmente sin pérdida.

Breve descripción de los dibujos

Ahora se describirán realizaciones de ejemplo con referencia a los dibujos que se acompañan, en los cuales:

la figura 1 es un diagrama de bloques generalizado que muestra, a modo de ejemplo, un decodificador para procesar un flujo de bits y lograr el nivel de sonoridad deseado de una señal de audio de salida;

la figura 2 es un diagrama de bloques generalizado de una primera realización de un componente de mezcla del decodificador de la figura 1,

la figura 3 es un diagrama de bloques generalizado de una segunda realización de un componente de mezcla del decodificador de la figura 1;

la figura 4 describe una estructura de datos de presentación de acuerdo con las realizaciones,

la figura 5 muestra un diagrama de bloques generalizado de un codificador de audio a modo de ejemplo, y

la figura 6 describe un flujo de bits formado por el codificador de audio de la figura 5.

Todas las figuras son esquemáticas y muestran, en general, sólo las partes que son necesarias para aclarar la

divulgación, mientras que otras partes pueden omitirse o meramente sugerirse. A menos que se indique lo contrario, números de referencia similares se refieren a partes similares en figuras diferentes.

Descripción detallada

En vista de lo anterior, un objetivo es proporcionar codificadores y decodificadores y métodos asociados destinados a proporcionar el nivel de sonoridad deseado para una señal de audio de salida, independientemente de qué flujos secundarios de contenido se mezclen en la señal de audio de salida.

I. Visión de conjunto - Decodificador

De acuerdo con un primer aspecto, los ejemplos de realización proponen métodos de decodificación, decodificadores y productos de programas informáticos para la decodificación. Los métodos, decodificadores y productos de programas informáticos propuestos pueden tener, en general, las mismas características y ventajas.

De acuerdo con realizaciones de ejemplo, se proporciona un método para procesar un flujo de bits que comprende una pluralidad de flujos secundarios de contenido como se indica en la reivindicación 1.

Los datos que indican una estructura de datos de presentación seleccionada y un nivel de sonoridad deseado son típicamente una configuración de usuario disponible en el decodificador. Un usuario puede, por ejemplo, usar un control remoto para seleccionar una estructura de datos de presentación en la que el diálogo está en francés, y/o aumentar o disminuir el nivel de sonoridad de salida deseado. En muchas realizaciones, el nivel de sonoridad de salida está relacionado con las capacidades del dispositivo de reproducción. De acuerdo con algunas realizaciones, el nivel de sonoridad de salida está controlado por el volumen. En consecuencia, los datos que indican una estructura de datos de presentación seleccionada y el valor de sonoridad deseado no se incluyen típicamente en el flujo de bits recibido por el decodificador.

Como se usa en el presente documento, “sonoridad” representa una medida psicoacústica modelada de la intensidad del sonido; en otras palabras, la sonoridad representa una aproximación del volumen de un sonido o sonidos percibidos por el usuario promedio.

Como se usa en el presente documento, el término “datos de sonoridad” se refiere a los datos que resultan de la medición del nivel de sonoridad de una estructura de datos de presentación específica mediante una función que modela la percepción de sonoridad psicoacústica. En otras palabras, es una colección de valores que indica las propiedades de sonoridad de la combinación de uno o más flujos secundarios de contenido a los que se hace referencia. De acuerdo con las realizaciones, se puede medir el nivel de sonoridad promedio de la combinación de los uno o más flujos secundarios de contenido a los que hace referencia la estructura de datos de presentación específica. Por ejemplo, los datos de sonoridad pueden hacer referencia a un valor de normalización de diálogo (de acuerdo con las recomendaciones ITU-R BS.1770) de los uno o más flujos secundarios de contenido a los que hace referencia la estructura de datos de presentación específica. Se pueden usar otros estándares de medición de sonoridad adecuados, tal como el modelo de sonoridad de Glasberg y Moore, que proporciona modificaciones y extensiones al modelo de sonoridad de Zwicker.

Como se usa aquí, “estructura de datos de presentación” se refiere a metadatos relacionados con el contenido de una señal de audio de salida. La señal de audio de salida también se denominará “programa”. La estructura de datos de presentación también se denominará “presentación”.

El contenido de audio se puede distribuir en varios flujos secundarios. Como se usa aquí, “flujo secundario de contenido” se refiere a tales flujos secundarios. Por ejemplo, un flujo secundario de contenido puede comprender la música del contenido de audio, el diálogo del contenido de audio o una pista de comentarios para que se incluyan en la señal de audio de salida. Un flujo secundario de contenido puede estar basado tanto en canales como en objetos. En el último caso, los datos de posición espacial dependientes del tiempo se incluyen en el flujo secundario de contenido. El flujo secundario de contenido puede estar comprendido en un flujo de bits o ser parte de la señal de audio (es decir, como un grupo de canales o un grupo de objetos).

Tal como se usa en el presente documento, “señal de audio de salida” se refiere a la señal de audio realmente emitida que se ofrecerá al usuario.

Los inventores se han dado cuenta de que al proporcionar datos de sonoridad para cada presentación, por ejemplo, un valor de normalización de diálogo, los datos de sonoridad específicos están disponibles para el decodificador que indica exactamente cuál es la sonoridad para el al menos un flujo secundario referido de contenido al decodificar esa presentación específica.

En la técnica anterior, se pueden proporcionar datos de sonoridad para cada flujo secundario de contenido. El problema de proporcionar datos de sonoridad para cada flujo secundario de contenido es que, en ese caso, depende del decodificador combinar los diversos datos de sonoridad en una sonoridad de presentación. Añadir los valores de

datos de sonoridad individuales de los flujos secundarios, que representan las sonoridades promedio de los flujos secundarios, para llegar a un valor de sonoridad para una cierta presentación, puede no ser preciso, y, en muchos casos, no dará como resultado el valor de sonoridad promedio real de los flujos secundarios combinados. Añadir los datos de sonoridad para cada flujo secundario de contenido referido puede ser matemáticamente imposible debido a las propiedades de la señal, al algoritmo de sonoridad y a la naturaleza de la percepción de la sonoridad, que no es típicamente aditiva, y podría dar lugar a imprecisiones potenciales que superen las tolerancias indicadas anteriormente.

Usando la presente realización, la diferencia entre el nivel de sonoridad promedio de la presentación seleccionada, proporcionada por los datos de sonoridad para la presentación seleccionada, y el nivel de sonoridad deseado, puede usarse para controlar de este modo la ganancia de reproducción de la señal de audio de salida.

Al proporcionar y usar datos de sonoridad como se describió anteriormente, se puede conseguir una sonoridad consistente, es decir, una sonoridad que esté cerca del nivel de sonoridad deseado, entre diferentes presentaciones. Además, se puede conseguir una sonoridad consistente entre diferentes programas en un canal de televisión, por ejemplo, entre un programa de televisión y sus cortes comerciales, y también entre canales de televisión.

De acuerdo con realizaciones de ejemplo, en las que la estructura de datos de presentación seleccionada hace referencia a dos o más flujos secundarios de contenido, y hace adicionalmente referencia a al menos dos coeficientes de mezcla que se aplicarán a estos, formando dicha señal de audio de salida que comprende adicionalmente mezclar de manera aditiva los uno o más flujos secundarios decodificados de contenido aplicando el/los coeficiente/s de mezcla.

Al proporcionar al menos dos coeficientes de mezcla, se logra una mayor flexibilidad del contenido de la señal de audio de salida.

Por ejemplo, la estructura de datos de presentación seleccionada puede hacer referencia, para cada flujo secundario de los dos o más flujos secundarios de contenido, a un coeficiente de mezcla que se aplicará a los respectivos flujos secundarios. De acuerdo con esta realización, se pueden cambiar los niveles de sonoridad relativa entre los flujos secundarios de contenido. Por ejemplo, las preferencias culturales pueden requerir diferentes equilibrios entre los diferentes flujos secundarios de contenido. Considérese una situación en la que las regiones españolas quieran prestar menos atención a la música. Por ello, el flujo secundario de música se atenúa en 3 dB. De acuerdo con otras realizaciones, se puede aplicar un único coeficiente de mezcla a un subconjunto de los dos o más flujos secundarios de contenido.

De acuerdo con realizaciones de ejemplo, el flujo de bits comprende una pluralidad de tramas de tiempo, y los coeficientes de mezcla a los que hace referencia la estructura de datos de presentación seleccionada son asignables independientemente para cada trama de tiempo. Un efecto de proporcionar coeficientes de mezcla que varían con el tiempo es que se puede conseguir la atenuación. Por ejemplo, el nivel de sonoridad para un segmento de tiempo de un flujo secundario de contenido puede reducirse mediante un aumento de sonoridad en el mismo segmento de tiempo de otro flujo secundario de contenido.

De acuerdo con realizaciones de ejemplo, los datos de sonoridad representan valores de una función de sonoridad relacionada con la aplicación de activación de puerta a su señal de entrada de audio.

La señal de entrada de audio es la señal a un lado del codificador a la que se a la que se le ha aplicado la función de sonoridad (es decir, la función de normalización de diálogo). Los datos de sonoridad resultantes se transmiten luego al decodificador en el flujo de bits. La puerta de ruido (también conocida como puerta de silencio) es un dispositivo electrónico o equipo lógico informático (software) que se utiliza para controlar el volumen de una señal de audio. La activación de puerta es el uso de tal puerta. Las puertas de ruido atenúan las señales que se registran por debajo de un umbral. Las puertas de ruido pueden atenuar las señales en una cantidad fija, conocida como intervalo. En su forma más simple, una puerta de ruido permite que una señal pase sólo cuando está por encima de un umbral establecido.

La activación de puerta también puede basarse en la presencia de diálogo en la señal de entrada de audio. En consecuencia, de acuerdo con realizaciones de ejemplo, los datos de sonoridad representan valores de una función de sonoridad relacionada con tales segmentos de tiempo, de su señal de entrada de audio, que representan diálogo. De acuerdo con otras realizaciones, la activación de puerta se basa en un nivel de sonoridad mínimo. Dicho nivel de sonoridad mínimo puede ser un umbral absoluto o un umbral relativo. El umbral relativo puede basarse en el nivel de sonoridad medido con un umbral absoluto.

De acuerdo con realizaciones de ejemplo, la estructura de datos de presentación comprende adicionalmente una referencia a datos de compresión de intervalo dinámico, de DRC, para uno o más flujos secundarios de contenido a los que se hace referencia, incluyendo el método, adicionalmente, el procesamiento de uno o más flujos secundarios de contenido decodificados o la señal de audio de salida sobre la base de los datos de DRC, donde el procesamiento comprende aplicar una o más ganancias de DRC a uno o más flujos secundarios de contenido

decodificados o la señal de audio de salida.

La compresión del intervalo dinámico reduce el volumen de los sonidos fuertes o amplifica los sonidos bajos, por lo tanto, estrecha o "comprime" el intervalo dinámico de una señal de audio. Al proporcionar datos de DRC únicos para cada presentación, se puede conseguir una experiencia de usuario mejorada de la señal de audio de salida, sin importar qué presentación se elija. Lo que es más, al proporcionar datos de DRC para cada presentación, se puede conseguir una experiencia de usuario consistente de la señal de salida de audio en cada presentación de la pluralidad de presentaciones, y también entre programas y a través de canales de TV, como se describió anteriormente.

Las ganancias de DRC varían siempre en el tiempo. En cada segmento de tiempo, las ganancias de DRC pueden ser una sola ganancia para la señal de salida de audio o diferir para cada flujo secundario. Las ganancias de DRC pueden aplicarse a grupos de canales y/o depender de la frecuencia. Además, las ganancias de DRC comprendidas en los datos de DRC pueden representar ganancias de DRC para dos o más segmentos de tiempo de DRC. Por ejemplo para las subtramas de una trama de tiempo definida por el codificador.

De acuerdo con realizaciones de ejemplo, los datos de DRC comprenden al menos un conjunto de las una o más ganancias de DRC. De este modo, los datos de DRC pueden comprender múltiples perfiles de DRC correspondientes a los modos de DRC, cada uno de los cuales proporciona una experiencia de usuario diferente de la señal de salida de audio. Al incluir las ganancias de DRC directamente en los datos de DRC, se puede conseguir una reducción en la complejidad computacional del decodificador.

De acuerdo con realizaciones de ejemplo, los datos de DRC comprenden al menos una curva de compresión, obteniéndose las una o más ganancias de DRC: al calcular uno o más valores de sonoridad de los uno o más flujos secundarios de contenido o la señal de salida de audio usando una función predefinida de sonoridad, y al mapear los uno o más valores de sonoridad para las ganancias de DRC utilizando la curva de compresión. Al proporcionarse curvas de compresión en los datos de DRC y calcular las ganancias de DRC en base a esas curvas, se puede reducir el régimen de bits requerido para transmitir los datos de DRC al codificador. La función de sonoridad predefinida puede tomarse, por ejemplo, de los documentos de recomendación ITU-R BS.1770, pero puede usarse cualquier función de sonoridad adecuada.

De acuerdo con realizaciones de ejemplo, el mapeo de los valores de sonoridad comprende una función de suavizado de las ganancias de DRC. El efecto de esto puede ser una señal de audio de salida mejor percibida. Las constantes de tiempo para suavizar las ganancias de DRC pueden transmitirse como parte de los datos de DRC. Tales constantes de tiempo pueden ser diferentes dependiendo de las propiedades de la señal. Por ejemplo, en algunas realizaciones, la constante de tiempo puede ser menor cuando dicho valor de sonoridad sea mayor que el correspondiente valor anterior de sonoridad, en comparación con cuando dicho valor de sonoridad sea menor que el correspondiente valor anterior de sonoridad.

De acuerdo con realizaciones de ejemplo, dichos datos de DRC a los que se hace referencia están comprendidos en dicho flujo secundario de metadatos. Esto puede reducir la complejidad de decodificación del flujo de bits.

De acuerdo con las realizaciones de ejemplo, cada uno de los uno o más flujos secundarios de contenido decodificados comprende datos de sonoridad de nivel de flujo secundario descriptivos del nivel de sonoridad del flujo secundario de contenido, incluyendo adicionalmente, dicho procesamiento adicional del uno o más flujos secundarios de contenido decodificado o de la señal de audio de salida, garantizar la consistencia de la sonoridad en base al nivel de sonoridad del flujo secundario de contenido.

Como se usa en el presente documento, "consistencia de sonoridad" se refiere a que la sonoridad es consistente entre diferentes presentaciones, es decir, consistente sobre las señales de audio de salida formadas sobre la base de diferentes flujos secundarios de contenido. Lo que es más, el término se refiere a que la sonoridad es consistente entre diferentes programas, es decir, entre señales de audio de salida completamente diferentes, tales como una señal de audio de un programa de televisión y una señal de audio de un anuncio. Además, el término se refiere a que la sonoridad es consistente en diferentes canales de TV.

Proporcionar datos de sonoridad descriptivos de un nivel de sonoridad del flujo secundario de contenido puede, en algunos casos, ayudar al decodificador a proporcionar consistencia de sonoridad. Por ejemplo, en los casos en los que dicha formación de una señal de audio de salida incluye la combinación de dos o más flujos secundarios de contenido decodificado utilizando coeficientes de mezcla alternativos, y cuando los datos de sonoridad a nivel de flujo secundario se usan para compensar los datos de sonoridad para proporcionar consistencia de sonoridad. Estos coeficientes de mezcla alternativos pueden derivarse de la entrada de usuario, en el caso, por ejemplo, de que el usuario decida desviarse de esa presentación por defecto (por ejemplo, con perfeccionamiento de diálogo, atenuación de diálogo, personalización de escena, etc.). Esto puede poner en peligro el cumplimiento de la sonoridad, ya que la influencia del usuario puede hacer que la sonoridad de la señal de salida de audio quede fuera de las normas de cumplimiento. Para ayudar a la consistencia de la sonoridad en esos casos, la presente realización proporciona la opción de transmitir datos de sonoridad a nivel de flujo secundario.

De acuerdo con algunas realizaciones, la referencia a al menos uno de dichos flujos secundarios de contenido es una referencia a al menos un grupo de flujos secundarios de contenido compuesto por uno o más de los flujos secundarios de contenido. Esto puede reducir la complejidad del decodificador, ya que las presentaciones de entre una pluralidad de presentaciones pueden compartir un grupo de flujo secundario de contenido (por ejemplo, un grupo de flujo secundario compuesto por el flujo secundario de contenido relacionado con la música y el flujo secundario de contenido relacionado con los efectos). Esto también puede disminuir el régimen de bits requerido para transmitir el flujo de bits.

De acuerdo con algunas realizaciones, la estructura de datos de presentación seleccionada hace referencia, para un grupo de flujos secundarios de contenido, a un único coeficiente de mezcla que se aplicará a cada uno de dichos uno o más flujos secundarios de contenido que componen el grupo de flujos secundarios.

Esto puede ser ventajoso en el caso de que las proporciones mutuas del nivel de sonoridad de los flujos secundarios de contenido de un grupo de flujos secundarios de contenido estén bien, pero el nivel de sonoridad general de los flujos secundarios de contenido del grupo de flujos secundarios de contenido debe aumentar o disminuir en comparación con otro/s flujo secundario/s de contenido o grupo/s de flujos secundarios de contenido a los que la estructura de datos de presentación seleccionada hace referencia.

De acuerdo con algunas realizaciones, el flujo de bits comprende una pluralidad de tramas de tiempo, y allá donde los datos que indican la estructura de datos de presentación seleccionada entre las una o más estructuras de datos de presentación sean asignables de manera independiente para cada trama de tiempo. En consecuencia, en el caso de que se reciba una pluralidad de estructuras de datos de presentación para un programa, la estructura de datos de presentación seleccionada puede ser cambiada, por ejemplo, por el usuario, mientras el programa está en curso. En consecuencia, la presente realización proporciona una forma más flexible de seleccionar el contenido del audio de salida a la vez que proporciona, al mismo tiempo, consistencia de sonoridad de la señal de audio de salida.

De acuerdo con algunas realizaciones, el método comprende adicionalmente: del flujo de bits, y para la primera de dicha pluralidad de tramas de tiempo, extraer una o más estructuras de datos de presentación, y del flujo de bits, y para la segunda de dicha pluralidad de tramas de tiempo, extraer una o más estructuras de datos de presentación diferentes a dichas una o más estructuras de datos de presentación extraídas de la primera de dicha pluralidad de tramas de tiempo, donde los datos que indican la estructura de datos de presentación seleccionada indican una estructura de datos de presentación seleccionada para el trama de tiempo que le ha sido asignado. En consecuencia, se puede recibir una pluralidad de estructuras de datos de presentación en el flujo de bits, donde algunas de las estructuras de datos de presentación se relacionan con un primer conjunto de tramas de tiempo y algunas de las estructuras de datos de presentación se relacionan con un segundo conjunto de tramas de tiempo. Por ejemplo una pista de comentarios sólo puede estar disponible durante un cierto segmento de tiempo del programa. Lo que es más, las estructuras de datos de presentación actualmente aplicables en un momento específico pueden usarse para seleccionar una estructura de datos de presentación seleccionada mientras el programa está en curso. En consecuencia, la presente realización proporciona una forma más flexible de seleccionar el contenido del audio de salida a la vez que proporciona, al mismo tiempo, consistencia de sonoridad de la señal de audio de salida.

De acuerdo con algunas realizaciones, de la pluralidad de flujos secundarios de contenido comprendidos en el flujo de bits, sólo se decodifican los uno o más flujos secundarios de contenido a los que hace referencia la estructura de datos de presentación seleccionada. Esta realización puede proporcionar un decodificador eficiente, con una complejidad computacional reducida.

De acuerdo con algunas realizaciones, el flujo de bits comprende dos o más flujos de bits independientes, cada uno de los cuales comprende al menos uno de dicha pluralidad de flujos secundarios de contenido, donde el paso de decodificar los uno o más flujos secundarios de contenido a los que hace referencia la estructura de datos de presentación seleccionada comprende: decodificar por separado, para cada flujo de bits específico de los dos o más flujos de bits independientes, el/los flujo secundario/s de contenido de los flujos secundarios de contenido a los que se hace referencia comprendidos en el flujo de bits específico. De acuerdo con esta realización, cada flujo de bits independiente puede ser recibido por un decodificador independiente que decodifica el/los flujo/s secundario/s de contenido proporcionados en el flujo de bits independiente que se necesita de acuerdo con la estructura de presentación seleccionada.
Esto puede mejorar la velocidad de decodificación, ya que los decodificadores independientes pueden funcionar en paralelo. En consecuencia, las decodificaciones realizadas por los decodificadores independientes pueden superponerse al menos parcialmente. Sin embargo, cabe señalar que las decodificaciones realizadas por los decodificadores independientes no tienen que superponerse necesariamente.

Lo que es más, al dividir los flujos secundarios de contenido en varios flujos de bits, la presente realización permite recibir los al menos dos flujos de bits independientes a través de diferentes infraestructuras, como se describe más adelante. En consecuencia, la presente realización proporciona un método más flexible para recibir la pluralidad de flujos secundarios de contenido en el decodificador.

Cada decodificador puede procesar el/los flujo/s secundario/s decodificado/s sobre la base de los datos de sonoridad a los que hace referencia la estructura de datos de presentación seleccionada, y/o aplicar ganancias de DRC, y/o aplicar coeficientes de mezcla al/a los flujo secundario/s decodificado/s. Los flujos secundarios de contenido procesados o no procesados pueden ser luego proporcionados desde todos los al menos dos decodificadores a un componente de mezcla para formar la señal de audio de salida. Alternativamente, el componente de mezcla realiza el procesamiento de sonoridad y/o aplica las ganancias de DRC y/o aplica los coeficientes de mezcla. En algunas realizaciones, un primer decodificador puede recibir un primer flujo de bits de los dos o más flujos de bits independientes a través de una primera infraestructura (por ejemplo, con la difusión de televisión por cable) mientras que un segundo decodificador recibe un segundo flujo de bits de los dos o más flujos de bits independientes a través de una segunda infraestructura (por ejemplo, con Internet). De acuerdo con algunas realizaciones, dichas una o más estructuras de datos de presentación están presentes en todos los dos o más flujos de bits independientes. En este caso, la definición de presentación y los datos de sonoridad están presentes en todos los decodificadores independientes. Esto permite el funcionamiento por separado de los decodificadores hasta el componente de mezcla. Las referencias a flujos secundarios que no estén presentes en el flujo de bits correspondiente pueden indicarse como proporcionados externamente.

De acuerdo con realizaciones de ejemplo, se proporciona un decodificador para procesar un flujo de bits que comprende una pluralidad de flujos secundarios de contenido, cada uno de los cuales representa una señal de audio, comprendiendo el decodificador: un componente receptor configurado para recibir el flujo de bits; un demultiplexor configurado para extraer, del flujo de bits, una o más estructuras de datos de presentación, cada una de las cuales comprende una referencia a al menos uno de dichos flujos secundarios de contenido, y comprende adicionalmente una referencia a un flujo secundario de metadatos que representa datos de sonoridad descriptivos de la combinación del uno o más flujos secundarios de contenido a los que se hace referencia; un componente de estado de reproducción configurado para recibir datos que indican una estructura de datos de presentación seleccionada de entre la una o más estructuras de datos de presentación, y el nivel de sonoridad deseado; y un componente de mezcla configurado para decodificar los uno o más flujos secundarios de contenido a los que hace referencia la estructura de datos de presentación seleccionada, y para formar una señal de audio de salida sobre la base de los flujos secundarios de contenido decodificados, donde el componente de mezcla está configurado adicionalmente para procesar los uno o más flujos secundarios de contenido decodificados o la señal de audio de salida para alcanzar dicho nivel de sonoridad deseado sobre la base de la referencia de datos de sonoridad por la estructura de datos de presentación seleccionada.

II. Vista de conjunto - Codificador

Los ejemplos útiles para comprender la divulgación describen métodos de codificación, codificadores y productos de programas informáticos para la codificación. Los métodos, codificadores y productos de programas informáticos pueden tener, en general, las mismas características y ventajas.

De acuerdo con ejemplos útiles para comprender la divulgación, se proporciona un método de codificación de audio, que incluye: recibir una pluralidad de flujos secundarios de contenido que representan señales de audio respectivas; definir una o más estructuras de datos de presentación, cada una de las cuales se refiere a al menos un flujo secundario de dicha pluralidad de flujos secundarios de contenido; para cada una de las una o más estructuras de datos de presentación, aplicar una función de sonoridad predefinida para obtener datos de sonoridad descriptivos de la combinación de uno o más flujos secundarios de contenido a los que se hace referencia, e incluir una referencia a los datos de sonoridad de la estructura de datos de presentación; y formar un flujo de bits que comprenda dicha pluralidad de flujos secundarios de contenido, dichas una o más estructuras de datos de presentación y los datos de sonoridad a los que hace referencia las estructuras de datos de presentación.

Como se describió anteriormente, el término “flujo secundario de contenido” abarca flujos secundarios tanto dentro de un flujo de bits como dentro de una señal de audio. Un codificador de audio recibe típicamente señales de audio que luego se codifican en flujos de bits. Las señales de audio pueden agruparse, pudiendo cada grupo caracterizarse como señales de audio de entrada de codificador individuales. Luego, cada grupo puede codificarse en un flujo secundario.

De acuerdo con algunos ejemplos útiles para comprender la descripción, el método comprende adicionalmente los pasos de: para cada una de las una o más estructuras de datos de presentación, determinar los datos de compresión de intervalo dinámico, de DRC, para los uno o más flujos secundarios de contenido a los que se hace referencia, cuantificando los datos de DRC al menos una curva de compresión deseada o al menos un conjunto de ganancias de DRC, e incluyendo dichos datos de DRC en el flujo de bits.

De acuerdo con algunos ejemplos útiles para comprender la divulgación, el método comprende adicionalmente los pasos de: para cada flujo secundario de la pluralidad de flujos secundarios de contenido, aplicar la función de sonoridad predefinida para obtener datos de sonoridad a nivel de flujo secundario del flujo secundario de contenido; e incluir dichos datos de sonoridad a nivel de flujo secundario en el flujo de bits.

De acuerdo con algunos ejemplos útiles para comprender la divulgación, la función de sonoridad predefinida se relaciona con la aplicación de activación de puerta de la señal de audio.

De acuerdo con algunos ejemplos útiles para comprender la divulgación, la función de sonoridad predefinida se relaciona sólo con tales segmentos de tiempo de la señal de audio que representan el diálogo.

De acuerdo con algunos ejemplos útiles para comprender la divulgación, la función de sonoridad predefinida incluye al menos un elemento de entre: la ponderación dependiente de la frecuencia de la señal de audio, la ponderación dependiente del canal de la señal de audio, sin tener en cuenta los segmentos de la señal de audio con una potencia de la señal por debajo de un valor umbral, calculando la medida de energía de la señal de audio.

De acuerdo con ejemplos útiles para comprender la divulgación, se proporciona un codificador de audio que comprende: un componente de sonoridad configurado para aplicar una función de sonoridad predefinida para obtener datos de sonoridad descriptivos de una combinación de uno o más flujos secundarios de contenido que representan señales de audio respectivas; un componente de datos de presentación configurado para definir una o más estructuras de datos de presentación, comprendiendo, cada estructura, la referencia a uno o más flujos secundarios de contenido de entre una pluralidad de flujos secundarios de contenido, y una referencia a datos de sonoridad descriptivos de la combinación de los flujos secundarios de contenido a los que se hace referencia; y un componente de multiplexación configurado para formar un flujo de bits que comprenda dicha pluralidad de flujos secundarios de contenido, dichas una o más estructuras de datos de presentación y los datos de sonoridad a los que hace referencia las estructuras de datos de presentación.

III. Realizaciones de ejemplo

La figura 1 muestra a modo de ejemplo un diagrama de bloques generalizado de un decodificador 100 para procesar un flujo P de bits y alcanzar el nivel de sonoridad deseado de la señal 114 de audio de salida.

El decodificador 100 comprende un componente de recepción (no mostrado) configurado para recibir el flujo P de bits que comprende una pluralidad de flujos secundarios de contenido, cada uno de los cuales representa una señal de audio.

El decodificador 100 comprende adicionalmente un demultiplexor 102 configurado para extraer, del flujo P de bits, una o más estructuras 104 de datos de presentación. Cada estructura de datos de presentación comprende una referencia a al menos uno de dichos flujos secundarios de contenido. En otras palabras, una estructura de datos de presentación, o presentación, es una descripción de qué flujos secundarios de contenido se van a combinar. Como se indicó anteriormente, los flujos secundarios de contenido codificados en dos o más flujos secundarios independientes pueden combinarse en una sola presentación.

Cada estructura de datos de presentación comprende adicionalmente una referencia a un flujo secundario de metadatos que representa datos de sonoridad descriptivos de la combinación de uno o más flujos secundarios de contenido a los que se hace referencia.

El contenido de una estructura de datos de presentación y sus diferentes referencias se describirán ahora junto con la figura 4.

En la figura 4, se muestran los diferentes flujos secundarios 412, 205 a los que se puede hacer referencia mediante una o más estructuras 104 de datos de presentación extraídas. De las tres estructuras 104 de datos de presentación, se elige una estructura 110 de datos de presentación seleccionada. Como se desprende claramente de la figura 4, el flujo P de bits comprende los flujos secundarios 412 de contenido, el flujo secundario 205 de metadatos y las una o más estructuras 104 de datos de presentación. Los flujos secundarios 412 de contenido pueden comprender, por ejemplo, un flujo secundario para la música, un flujo secundario para los efectos, un flujo secundario para el ambiente, flujo secundario para diálogo en inglés, un flujo secundario para diálogo en español, un flujo secundario para audio asociado (AA) en inglés, por ejemplo, una pista de comentarios en inglés, y un flujo secundario para AA en español, por ejemplo, una pista de comentarios en español.

En la figura 4, todos los flujos secundarios 412 de contenido están codificados en el mismo flujo P de bits, pero, como se indicó anteriormente, no sucede siempre así. Los organismos de radiodifusión del contenido de audio pueden utilizar una única configuración de flujo de bits, por ejemplo, una configuración de identificador de paquete único (PID) en el estándar MPEG, o una configuración de flujo de bits múltiple, por ejemplo, una configuración dual-PID, para transmitir el contenido de audio a sus clientes, es decir, a un decodificador.

La presente divulgación introduce un nivel intermedio en forma de grupos de flujos secundarios, que residen entre la capa de presentación y la capa de flujos secundarios. Los grupos de flujos secundarios de contenido pueden agrupar o hacer referencia a uno o más flujos secundarios de contenido. Las presentaciones pueden entonces hacer referencia a grupos de flujos secundarios de contenido. En la figura 4, los flujos secundarios de música, efectos y ambiente de contenido se agrupan para formar un grupo 410 de flujos secundarios de contenido, al que se refiere

404 la estructura 110 de datos de presentación seleccionada.

Los grupos de flujos secundarios de contenido ofrecen más flexibilidad en la combinación de flujos secundarios de contenido. En particular, el nivel de grupo de flujos secundarios proporciona un medio para recopilar o agrupar varios flujos secundarios de contenido en un grupo único, por ejemplo, en el grupo 410 de flujos secundarios de contenido, que comprende música, efectos y ambiente.

Esto puede resultar ventajoso, ya que un grupo de flujos secundarios de contenido (por ejemplo, para música y efectos, o para música, efectos y ambiente) se puede usar para más de una presentación, por ejemplo, junto con un diálogo en inglés o español. Del mismo modo, un flujo secundario de contenido se puede también utilizar en más de un grupo de flujos secundarios de contenido.

Lo que es más, dependiendo de la sintaxis de la estructura de datos de presentación, el uso de grupos de flujos secundarios de contenido puede proporcionar posibilidades para mezclar un mayor número de flujos secundarios de contenido para una presentación.

De acuerdo con algunas realizaciones, una presentación 104, 110 constará siempre de uno o más grupos de flujos secundarios.

La estructura 110 de datos de presentación seleccionada de la figura 4 comprende una referencia 404 al grupo 410 de flujos secundarios de contenido compuesto por uno o más de los flujos secundarios de contenido. La estructura 110 de datos de presentación seleccionada comprende adicionalmente una referencia a un flujo secundario de contenido para diálogo en español y una referencia a un flujo secundario de contenido para AA en español. Lo que es más, la estructura 110 de datos de presentación seleccionada comprende una referencia 406 a un flujo secundario 205 de metadatos que representa datos 408 de sonoridad descriptivos de la combinación de uno o más flujos secundarios de contenido a los que se hace referencia. Obviamente, las otras dos estructuras de datos de presentación de la pluralidad de estructuras 104 de datos de presentación pueden comprender datos similares a los de la estructura 110 de datos de presentación seleccionada. De acuerdo con otras realizaciones, el flujo P de bits puede comprender flujos secundarios de metadatos adicionales similares al flujo secundario 205 de metadatos, haciendo referencia, las otras estructuras de datos de presentación, a estos flujos secundarios adicionales de metadatos. En otras palabras, cada estructura de datos de presentación de la pluralidad de estructuras 104 de datos de presentación puede hacer referencia a datos de sonoridad dedicados.

La estructura de datos de presentación seleccionada puede cambiar con el tiempo, es decir, si el usuario decide desactivar la pista de comentarios en español, AA (ES). En otras palabras, el flujo P de bits comprende una pluralidad de tramas de tiempo, siendo asignables independientemente, para cada trama de tiempo, los datos (con la referencia 108 de la figura 1) que indican la estructura de datos de presentación seleccionada entre las una o más estructuras 104 de datos de presentación.

Como se ha descrito anteriormente, el flujo P de bits comprende una pluralidad de tramas de tiempo. De acuerdo con algunas realizaciones, las una o más estructuras 104 de datos de presentación pueden relacionarse con diferentes segmentos de tiempo del flujo P de bits. En otras palabras, el demultiplexor (con la referencia 102 de la figura 1) puede configurarse para extraer, del flujo P de bits, y para la primera de dicha pluralidad de tramas de tiempo, una o más estructuras de datos de presentación, y configurarse adicionalmente para extraer, del flujo P de bits, y para la segunda de dicha pluralidad de tramas de tiempo, una o más estructuras de datos de presentación diferentes de las dichas una o más estructuras de datos de presentación extraídas de la primera de dicha pluralidad de tramas de tiempo. En este caso, los datos (con la referencia 108 de la figura 1), que indican la estructura de datos de presentación seleccionada, indican una estructura de datos de presentación seleccionada para la trama de tiempo a la que se han asignado.

Ahora, volviendo a la figura 1, el decodificador 100 comprende adicionalmente un componente 106 de estado de reproducción. El componente 106 de estado de reproducción está configurado para recibir datos 108 que indican una estructura 110 de datos de presentación seleccionada de entre una o más estructuras 104 de datos de presentación. Los datos 108 comprenden también el nivel de sonoridad deseado. Como se describió anteriormente, los datos 108 pueden ser proporcionados por un consumidor del contenido de audio que será decodificado por el decodificador 100. El valor de sonoridad deseado puede también ser una configuración específica del decodificador, que dependa del equipo de reproducción que se utilizará para la reproducción de la señal de audio de salida. El consumidor puede, por ejemplo, elegir que el contenido de audio comprenda un diálogo en español del modo en que se expuso anteriormente.

El decodificador 100 comprende adicionalmente un componente de mezcla que recibe la estructura 110 de datos de presentación seleccionada del componente 106 de estado de reproducción, y decodifica uno o más flujos secundarios de contenido a los que hace referencia la estructura 110 seleccionada de datos de presentación del flujo P de bits. De acuerdo con algunas realizaciones, el componente de mezcla sólo decodifica los uno o más flujos secundarios de contenido a los que hace referencia la estructura 110 de datos de presentación seleccionada. En consecuencia, en caso de que el consumidor haya elegido una presentación con, por ejemplo, diálogo en español,

no se decodificará ningún flujo secundario de contenido que represente diálogo en inglés, lo que reduce la complejidad computacional del decodificador 100.

El componente 112 de mezcla está configurado para formar una señal 114 de audio de salida sobre la base de los flujos secundarios de contenido decodificados.

Lo que es más, el componente 112 de mezcla está configurado para procesar los uno o más flujos secundarios de contenido decodificados o la señal de audio de salida para alcanzar dicho nivel de sonoridad deseado sobre la base de los datos de sonoridad a los que hace referencia la estructura 110 de datos de presentación seleccionada.

Las figuras 2 y 3 describen diferentes realizaciones del componente 112 de mezcla.

En la figura 2, el flujo P de bits es recibido por un componente 202 de decodificación de flujo secundario que, en base a la estructura 110 de datos de presentación seleccionada, decodifica los uno o más flujos secundarios 204 de contenido a los que hace referencia la estructura 110 seleccionada de datos de presentación del flujo P de bits. Los uno o más flujos secundarios 204 de contenido decodificado se transmiten luego a un componente 206 para formar una señal 114 de audio de salida sobre la base de los flujos secundarios 204 de contenido decodificados y de un flujo secundario 205 de metadatos. El componente 206 puede, por ejemplo, tener en cuenta cualesquiera datos de posición espacial dependientes del tiempo incluidos en el/los flujo/s secundario/s 204 de contenido cuando se forme la señal de salida de audio. El componente 206 puede tener en cuenta adicionalmente los datos de DRC comprendidos en el flujo secundario 205 de metadatos. Alternativamente, el componente 210 de sonoridad (descrito más adelante) procesa la señal 114 de audio de salida sobre la base de los datos de DRC. En algunas realizaciones, el componente 206 recibe coeficientes de mezcla (descritos más adelante) de la estructura 110 de datos de presentación (no mostrada en la figura 2) y los aplica a los flujos secundarios 204 de contenido correspondientes. La señal 114* de audio de salida se transmite luego a un componente 210 de sonoridad, el cual, sobre la base de los datos de sonoridad (incluidos en el flujo secundario 205 de metadatos) a los que hacen referencia la estructura seleccionada 110 de datos de presentación y el nivel de sonoridad deseado comprendido en los datos 108, procesa la señal 114* de audio de salida para alcanzar dicho nivel de sonoridad deseado, y emite, de este modo, una señal 114 de audio de salida procesada por sonoridad.

En la figura 3, se muestra un componente 112 de mezcla similar. La diferencia con el componente 112 de mezcla descrito en la figura 2 es que el componente 206 que forma la señal de audio de salida y el componente 210 de sonoridad han cambiado de posición entre sí. En consecuencia, el componente 210 de sonoridad procesa uno o más flujos secundarios 204 de contenido decodificados para alcanzar dicho nivel de sonoridad deseado (sobre la base de los datos de sonoridad incluidos en el flujo secundario 205 de metadatos) y emite uno o más flujos secundarios 204* de contenido de sonoridad procesados. Luego, se transmiten éstos al componente 206 para formar una señal de audio de salida que emite la señal 114 de Como se describe junto con la figura 2, se pueden aplicar datos de DRC (incluidos en el flujo secundario 205 de metadatos), ya sea en el componente 206 o en el componente 210 de sonoridad. Lo que es más, en algunas realizaciones, el componente 206 recibe coeficientes de mezcla (descritos más adelante) de la estructura 110 de datos de presentación (no mostrada en la figura 3) y los aplica a los flujos secundarios 204* de contenido correspondientes.

Cada una de las una o más estructuras 104 de datos de presentación comprende datos de sonoridad dedicados que indican exactamente cuál será la sonoridad de los flujos secundarios de contenido a los que hace referencia la estructura de datos de presentación cuando se decodifiquen. Los datos de sonoridad pueden representar, por ejemplo, el valor de normalización de diálogo. De acuerdo con algunas realizaciones, los datos de sonoridad representan valores de una función de sonoridad que aplica activación de puerta a su señal de entrada de audio. Esto puede mejorar la precisión de los datos de sonoridad. Por ejemplo, si los datos de sonoridad se basan en una función de sonoridad de limitación de banda, el ruido de fondo de la señal de entrada de audio no se tomará en consideración al calcular los datos de sonoridad, ya que las bandas de frecuencia que sólo contienen estática pueden descartarse.

Lo que es más, los datos de sonoridad pueden representar valores de una función de sonoridad relacionada con tales segmentos de tiempo de una señal de entrada de audio que representan un diálogo. Esto está en línea con el estándar ATSC A/85, en el que la normalización de diálogo se define explícitamente con respecto a la sonoridad del diálogo (elemento de ancla): "*The value of the dialnorm parameter indicates the loudness of the Anchor Element of the consent*".

El procesamiento de uno o más flujos secundarios de contenido decodificados o la señal de audio de salida para alcanzar dicho nivel de sonoridad deseado, ORL, sobre la base de los datos de sonoridad a los que hace referencia la estructura de datos de presentación seleccionada, o la nivelación, g_L, de la señal de audio de salida puede, de este modo, realizarse utilizando la normalización de diálogo de la presentación, *DN(pres)*, calculada de acuerdo con lo anterior:

$$g_L = ORL - DN(pres),$$

donde DN(pres) y ORL son ambos valores que típicamente se expresan en dB_{FS} (dB con referencia a una onda sinusoidal (o cuadrada) de 1 kHz a escala completa).

De acuerdo con algunas realizaciones, en las que la estructura de datos de presentación seleccionada hace referencia a dos o más flujos secundarios de contenido, la estructura de datos de presentación seleccionada hace referencia adicionalmente a al menos un coeficiente de mezcla que se aplicará a los dos o más flujos secundarios de contenido. El o los coeficientes de mezcla pueden utilizarse para proporcionar un nivel de sonoridad relativo modificado entre los flujos secundarios de contenido a los que hace referencia la presentación seleccionada. Estos coeficientes de mezcla pueden aplicarse como ganancias de banda ancha a un canal/objeto en un flujo secundario de contenido antes de mezclarlo con el canal/objeto en el/los otro/s flujo secundario/s de contenido.

Al menos un coeficiente de mezcla es típicamente estático, pero puede asignarse independientemente para cada trama de tiempo de un flujo de bits, por ejemplo, para conseguir una atenuación.

Por consiguiente, no es necesario transmitir los coeficientes de mezcla en el flujo de bits para cada trama de tiempo; pueden permanecer válidos hasta que se sobrescriban.

El coeficiente de mezcla se puede definir por flujo secundario de contenido. En otras palabras, la estructura de datos de presentación seleccionada puede hacer referencia, para cada flujo secundario de los dos o más flujos secundarios, a un coeficiente de mezcla que se aplicará a los respectivos flujos secundarios.

De acuerdo con otras realizaciones, el coeficiente de mezcla puede definirse por grupo de flujos secundarios de contenido y aplicarse a todos los flujos secundarios de contenido pertenecientes al grupo de flujos secundarios de contenido. En otras palabras, la estructura de datos de presentación seleccionada puede hacer referencia, para un grupo de flujos secundarios de contenido, a un único coeficiente de mezcla que se aplicará a cada uno de dichos uno o más flujos secundarios de contenido que componen el grupo de flujos secundarios.

De acuerdo con otra realización más, la estructura de datos de presentación seleccionada puede hacer referencia a un único coeficiente de mezcla para aplicar a cada uno de los dos o más flujos secundarios de contenido.

La tabla 1 de más abajo indica un ejemplo de transmisión de objetos. Los objetos se agrupan en categorías que se distribuyen en varios flujos secundarios. Todas las estructuras de datos de presentación combinan la música y los efectos que contienen la parte principal del contenido de audio sin el diálogo. Esta combinación es, de este modo, un grupo de flujos secundarios de contenido. Dependiendo de la estructura de datos de presentación seleccionada, se elige un determinado idioma, por ejemplo, inglés (D#1) o español D#2. Lo que es más, el flujo secundario de contenido comprende un flujo secundario de audio asociado en inglés (Desc#1) y un flujo secundario de audio asociado en español (Desc#2). El audio asociado puede comprender un audio perfeccionado, tal como una descripción de audio, un narrador para personas con problemas de audición, un narrador para personas con problemas de visión, una pista de comentarios, etc.

Tabla 1: Ejemplos de coeficientes de mezcla

Flujos secundarios / grupos	M&E		D#1	D#2	Desc#1	Desc#2
Flujos secundarios / Presentación	Música	Efectos	D#1	D#2	Desc#1	Decs#2
1	(0 dB)	(0 dB)	(0 dB)	-	-	-
2	(-3 dB)	(0 dB)	-	(0 dB)	-	
3	(-3 dB)	(0 dB)	-	(0 dB)	-	(-6 dB)
4	(-3 dB)	(-3 dB)	(-3 dB)	-	(0 dB)	-

En la presentación 1, no se debe aplicar ninguna ganancia de mezcla mediante coeficientes de mezcla; la presentación 1, de este modo, no hace en absoluto referencia a los coeficientes de mezcla.

Las preferencias culturales pueden requerir diferentes equilibrios entre las categorías. Esto se ejemplifica en la presentación 2. Considérese la situación en la que las regiones españolas quieran prestar menos atención a la música. El flujo secundario de música se atenuará, por lo tanto, en 3 dB. En este ejemplo, la presentación 2 hace referencia, para cada flujo secundario de los dos o más flujos secundarios, a un coeficiente de mezcla que se aplicará a los respectivos flujos secundarios.

La presentación 3 incluye un flujo de descripción en español para personas con problemas de visión. Este flujo fue grabado en una cabina, y resulta demasiado alto para mezclarlo directamente en la presentación, así que se atenúa, por lo tanto, en 6 dB. En este ejemplo, la presentación 3 hace referencia, para cada flujo secundario de los dos o más flujos secundarios, a aplicar un coeficiente de mezcla a los respectivos flujos secundarios.

En la presentación 4, tanto el flujo secundario de música como el flujo secundario de efectos se atenúan en 3 dB. En este caso, la presentación 4 hace referencia, para el grupo de flujos secundarios de M&E, a aplicar un único coeficiente de mezcla a cada uno de dichos uno o más flujos secundarios de contenido que componen el grupo de flujos secundarios de M&E.

De acuerdo con algunas realizaciones, el usuario o consumidor del contenido de audio puede proporcionar una entrada de usuario, de tal manera que la señal de audio de salida se desvíe de la estructura de datos de presentación seleccionada. Por ejemplo, el usuario puede solicitar el perfeccionamiento o la atenuación del diálogo, o el usuario puede querer realizar algún tipo de personalización de la escena, así, por ejemplo, aumentar el volumen de los efectos. En otras palabras, se pueden proporcionar coeficientes de mezcla alternativos que se usen cuando se combinen dos o más flujos secundarios de contenido decodificados para formar la señal de audio de salida. Esto puede influir en el nivel de sonoridad de la señal de salida de audio. Con el fin de proporcionar consistencia de sonoridad en este caso, cada uno de los uno o más flujos secundarios de contenido decodificados pueden comprender datos de sonoridad a nivel de flujo secundario descriptivos del nivel de sonoridad del flujo secundario de contenido. Los datos de sonoridad a nivel de flujo secundario pueden usarse entonces para compensar los datos de sonoridad y proporcionar consistencia de sonoridad.

Los datos de sonoridad a nivel de flujo secundario pueden ser similares a los datos de sonoridad a los que hace referencia la estructura de datos de presentación, y pueden representar ventajosamente valores de una función de sonoridad, opcionalmente con un intervalo mayor para cubrir las señales en general más silenciosas de un flujo secundario de contenido.

Hay muchas maneras de usar estos datos para conseguir consistencia de sonoridad. Los siguientes algoritmos se muestran a modo de ejemplo.

Sea DN(P) la normalización de diálogo de presentación, y DN(Si) la sonoridad de flujo secundario del flujo secundario i.

Si un decodificador está formando una señal de salida de audio en base a una presentación que hace referencia a un flujo secundario de contenido de música, S_M, y a un flujo secundario de contenido de efectos, S_E, como a un grupo de flujo secundario de contenido, $S_{M\&E}$, y un flujo secundario de contenido de diálogo, S_D, intenta mantener una sonoridad consistente mientras aplica 9 dB de perfeccionamiento de diálogo, DE, el decodificador podría predecir la nueva sonoridad de presentación, $DN(P_{DE})$, con DE al sumar los valores de sonoridad de flujo secundario de contenido:

$$DN(P_{DE}) = log_{10}(10^{DN(S_{M\&E})} + 10^{DN(S_D)+9})$$

Como se describió anteriormente, realizar tal suma de sonoridades de flujo secundario cuando se aproxima a la sonoridad de presentación puede dar como resultado una sonoridad muy diferente a la sonoridad real. Por consiguiente, como alternativa, se puede calcular la aproximación sin DE, para encontrar una compensación de la sonoridad real:

$$desplazamiento = DN(P) - log_{10}(10^{DN(S_{M\&E})} + 10^{DN(S_D)})$$

Dado que la ganancia en el DE no significa una gran modificación del programa, en la manera en que las diferentes señales de flujo secundario interactúan entre sí, es probable que la aproximación de $DN(P_{DE})$ sea más precisa cuando se use un desplazamiento para corregirla:

$$DN(P_{DE}) = log_{10}(10^{DN(S_{M\&E})}) + 10^{DN(S_D)+9})) + desplazamiento$$

De acuerdo con algunas realizaciones, la estructura de datos de presentación comprende adicionalmente una referencia a datos de compresión de intervalo dinámico, DRC, para uno o más flujos secundarios 204 de contenido a los que se hace referencia. Estos datos de DRC se pueden usar para procesar uno o más flujos secundarios 204 de contenido decodificados aplicando una o más ganancias de DRC a uno o más flujos secundarios 204 de contenido decodificados o a la señal 114 de audio de salida. Las una o más ganancias de DRC pueden incluirse en los datos de DRC, o pueden calcularse en base a una o más curvas de compresión comprendidas en los datos de DRC. En ese caso, el decodificador 100 calcula un valor de sonoridad para cada uno o más flujos secundarios 204 de

contenido a los que se hace referencia o para la señal 114 de audio de salida usando una función de sonoridad predefinida, y luego usa el/los valor/es de sonoridad para mapear las ganancias de DRC usando la/s curva/s de compresión. El mapeo de los valores de sonoridad puede comprender una función de suavizado de las ganancias de DRC.

De acuerdo con algunas realizaciones, los datos de DRC a los que hace referencia la estructura de datos de presentación corresponden a múltiples perfiles de DRC. Estos perfiles de DRC se hacen a medida conforme a la señal de audio particular a la que se pueden aplicar. Los perfiles pueden variar desde la compresión nula ("Ninguna"), pasando por una compresión bastante ligera (por ejemplo, "Música ligera") hasta una compresión extremadamente agresiva (por ejemplo, "Discurso"). En consecuencia, los datos de DRC pueden comprender múltiples conjuntos de ganancias de DRC, o múltiples curvas de compresión a partir de las cuales se pueden obtener múltiples conjuntos de ganancias de DRC.

Los datos de DRC a los que se hace referencia pueden, de acuerdo con las realizaciones, estar incluidos en el flujo secundario 205 de metadatos de la figura 4.

Debe señalarse que el flujo P de bits puede, de acuerdo con algunas realizaciones, comprender dos o más flujos de bits independientes, y que los flujos secundarios de contenido pueden codificarse en este caso en diferentes flujos de bits. En este caso, una o más estructuras de datos de presentación se incluyen ventajosamente en todos los flujos de bits independientes, lo que significa que varios decodificadores, uno para cada flujo de bits independiente, pueden trabajar por separado y de manera totalmente independiente para decodificar los flujos secundarios de contenido a los que hace referencia la estructura de datos de presentación seleccionada (que también se proporciona a cada decodificador por separado). De acuerdo con algunas realizaciones, los decodificadores pueden trabajar en paralelo. Cada decodificador independiente decodifica los flujos secundarios que existen en el flujo de bits independiente que recibe. De acuerdo con las realizaciones, cada decodificador independiente realiza el procesamiento de los flujos secundarios de contenido decodificados por él, para alcanzar el nivel de sonoridad deseado. Los flujos secundarios de contenido procesados se envían luego a un componente de mezcla adicional que forma la señal de audio de salida, con el nivel de sonoridad deseado.

De acuerdo con otras realizaciones, cada decodificador independiente proporciona sus flujos secundarios decodificados y sin procesar al componente de mezcla adicional que realiza el procesamiento de sonoridad, y luego forma la señal de audio de salida de todos los uno o más flujos secundarios de contenido a los que hace referencia la estructura de datos de presentación seleccionada, o primero mezcla los uno o más flujos secundarios de contenido y realiza el procesamiento de sonoridad en la señal mezclada. De acuerdo con otras realizaciones, cada decodificador independiente realiza una función de mezcla en dos o más de sus flujos secundarios decodificados. A continuación, otro componente de mezcla mezcla las contribuciones premezcladas de los decodificadores independientes.

La figura 5, junto con la figura 6, muestra a modo de ejemplo un codificador 500 de audio. El codificador 500 comprende un componente 504 de datos de presentación configurado para definir una o más estructuras 506 de datos de presentación, cada una de las cuales comprende una referencia 604, 605 a uno o más flujos secundarios 612 de contenido pertenecientes a una pluralidad de flujos secundarios 502 de contenido, y una referencia 608 a los datos 510 de sonoridad descriptivos de una combinación de los flujos secundarios 612 de contenido a los que se hace referencia. El codificador 500 comprende adicionalmente un componente 508 de sonoridad configurado para aplicar una función 514 de sonoridad predefinida para obtener datos 510 de sonoridad descriptivos de una combinación de uno o más flujos secundarios de contenido que representan señales de audio respectivas. El codificador comprende adicionalmente un componente 512 de multiplexación configurado para formar un flujo P de bits que comprende dicha pluralidad de flujos secundarios de contenido, dichas una o más estructuras 506 de datos de presentación y los datos 510 de sonoridad a los que hacen referencia dichas una o más estructuras 506 de datos de presentación. Cabe señalar que los datos 510 de sonoridad comprenden típicamente varias instancias de datos de sonoridad, una para cada una de dichas una o más estructuras 506 de datos de presentación.

El codificador 500 puede adaptarse adicionalmente para cada una de las una o más estructuras 506 de datos de presentación, determinando datos de compresión de intervalo dinámico, DRC, para los uno o más flujos secundarios de contenido a los que se hace referencia. Los datos de DRC cuantifican al menos una curva de compresión deseada o al menos un conjunto de ganancias de DRC. Los datos de DRC se incluyen en el flujo P de bits. Los datos de DRC y los datos 510 de sonoridad pueden, de acuerdo con las realizaciones, incluirse en el flujo secundario 614 de metadatos. Como se analizó anteriormente, los datos de sonoridad dependen típicamente de la presentación. Lo que es más, los datos de DRC pueden también depender de la presentación. En estos casos, los datos de sonoridad y, si corresponde, los datos de DRC para una estructura de datos de presentación específica se incluyen en un flujo secundario 614 de metadatos dedicado a esa estructura de datos de presentación específica.

El codificador puede adaptarse adicionalmente para, para cada flujo secundario de la pluralidad de flujos secundarios 502 de contenido, aplicar la función de sonoridad predefinida para obtener datos de sonoridad a nivel de flujo secundario del flujo secundario de contenido; e incluir dichos datos de sonoridad a nivel de flujo secundario en el flujo de bits. La función de sonoridad predefinida puede estar relacionada con la activación de puerta de la

señal de audio. De acuerdo con otras realizaciones, la función de sonoridad predefinida se refiere únicamente a los segmentos de tiempo de la señal de audio que representen diálogo. La función de sonoridad predefinida puede, de acuerdo con algunas realizaciones, incluir al menos un elemento de entre:

- la ponderación dependiente de la frecuencia de la señal de audio,

- la ponderación dependiente del canal de la señal de audio,

- no tener en cuenta los segmentos de la señal de audio con una potencia de señal por debajo de un valor de umbral,

- no tener en cuenta los segmentos de la señal de audio que no se detecten como discurso,

- calcular una medida de energía/potencia/raíz cuadrática media de la señal de audio.

Como se desprende de lo anterior, la función de sonoridad no es lineal. Esto significa que, en caso de que los datos de sonoridad sólo se calcularan a partir de los diferentes flujos secundarios de contenido, la sonoridad de una cierta presentación no podría calcularse sumando los datos de sonoridad de los flujos secundarios de contenido a los que se hace referencia. Lo que es más, cuando se combinan diferentes pistas de audio, es decir, flujos secundarios de contenido, conjuntamente, para la reproducción simultánea, puede aparecer un efecto combinado entre partes coherentes/incoherentes o en diferentes regiones de frecuencia de las diferentes pistas de audio, lo que hace que la suma de los datos de sonoridad para la pista de audio sea matemáticamente imposible.

IV. Equivalentes, extensiones, alternativas y miscelánea

Otras realizaciones de la presente divulgación resultarán evidentes para el experto en la técnica después de estudiar la descripción anterior. Aunque la presente descripción y los dibujos divulguen realizaciones y ejemplos, la divulgación no se limita a estos ejemplos específicos. Se pueden realizar numerosas modificaciones y variaciones sin apartarse del alcance de la presente divulgación, que se define en las reivindicaciones que se acompañan. Cualesquiera signos de referencia que aparezcan en las reivindicaciones no deben entenderse como limitantes de su alcance.

Además, las variaciones a las realizaciones divulgadas pueden ser comprendidas y efectuadas por el experto en la técnica al practicar la divulgación, a partir de un estudio de los dibujos, la divulgación y las reivindicaciones adjuntas. En las reivindicaciones, la palabra “que comprende” no excluye otros elementos o pasos, y el artículo indefinido “un” o “una” no excluye la pluralidad. El mero hecho de que ciertas medidas se citen en reivindicaciones dependientes diferentes mutuamente no indica que una combinación de estas medidas no pueda utilizarse ventajosamente.

Los dispositivos y métodos divulgados anteriormente pueden implantarse como equipo lógico informático (software), soporte lógico inalterable (firmware), equipo físico informático (hardware) o una combinación de los mismos. En una implantación de hardware, la división de tareas entre unidades funcionales a que se refiere la descripción anterior no corresponde necesariamente a la división en unidades físicas; por el contrario, un componente físico puede tener múltiples funcionalidades, y una tarea puede ser realizada por varios componentes físicos en cooperación. Ciertos componentes o todos los componentes pueden implantarse como software ejecutado por un procesador o microprocesador de señal digital, o implantarse como hardware o como un circuito integrado específico de aplicación. Tal software puede distribuirse en medios legibles por ordenador, que pueden comprender medios de almacenamiento informáticos (o medios no transitorios) y medios de comunicación (o medios transitorios). Como es bien sabido por el experto en la técnica, el término medios de almacenamiento informáticos incluye medios tanto volátiles como no volátiles, extraíbles y no extraíbles, implantados en cualquier método o tecnología para el almacenamiento de información, tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento informáticos incluyen, entre otros, RAM, ROM, EEPROM, memoria flash u otra tecnología de memoria, CD-ROM, discos versátiles digitales (DVD) u otro almacenamiento en disco óptico, casetes magnéticos, cinta magnética, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda utilizarse para almacenar la información deseada y al que pueda acceder un ordenador. Además, es bien sabido por el experto en la técnica que los medios de comunicación incorporan típicamente instrucciones, estructuras de datos, módulos de programa u otros datos legibles por ordenador en una señal de datos modulada, tal como una onda portadora u otro mecanismo de transporte, e incluyen cualquier medio de entrega de información.

REIVINDICACIONES

1. Un método de procesamiento de un flujo (P) de bits que comprende una pluralidad de flujos secundarios (412) de contenido, cada uno de los cuales representa una señal de audio, incluyendo el método:

a partir del flujo de bits, extraer una o más estructuras (104) de datos de presentación, cada una de las cuales comprende una referencia (404, 405) a una pluralidad de dichos flujos secundarios de contenido, comprendiendo adicionalmente cada estructura de datos de presentación una referencia (406) a datos (408) de sonoridad y a datos de compresión de intervalo dinámico (DRC) incluidos en un flujo secundario (205) de metadatos, en el que dichos datos de sonoridad están dedicados a dicha estructura de datos de presentación e indican cuál será la sonoridad de la combinación de la pluralidad de flujos secundarios de contenido (204) a la que se hace referencia cuando se decodifiquen los flujos secundarios, y en el que los datos de DRC incluyen múltiples conjuntos de una o más ganancias de DRC, y corresponden a múltiples perfiles de DRC;

recibir datos (108) que indican la estructura de datos de presentación seleccionada de entre dichas una o más estructuras (104) de datos de presentación, y el nivel de sonoridad deseado;

decodificar la pluralidad de flujos secundarios (204) de contenido a los que hace referencia la estructura (110) de datos de presentación seleccionada; y

formar una señal (114) de audio de salida sobre la base de los flujos secundarios (204) de contenido decodificados,

incluyendo adicionalmente el método procesar la pluralidad descodificada de flujos secundarios (204) de contenido o la señal (114) de audio de salida sobre la base de los datos de sonoridad a los que hace referencia la estructura de datos de presentación seleccionada y al menos un conjunto de los múltiples conjuntos de una o más ganancias de DRC para alcanzar el nivel de sonoridad deseado.

2. El método de la reivindicación 1, en el que la estructura de datos de presentación seleccionada hace también referencia a al menos dos coeficientes de mezcla que se aplicarán a la pluralidad de flujos secundarios de contenido,

comprendiendo adicionalmente, dicha formación de una señal de audio de salida, mezclar de manera aditiva la pluralidad descodificada de flujos secundarios de contenido aplicando los coeficientes de mezcla.

3. El método de la reivindicación 2, en el que el flujo de bits comprende una pluralidad de tramas de tiempo, y en el que los coeficientes de mezcla a los que hace referencia la estructura de datos de presentación seleccionada son asignables independientemente para cada trama de tiempo.

4. El método de la reivindicación 2 o 3, en el que la estructura de datos de presentación seleccionada hace referencia, para cada flujo secundario de la pluralidad de flujos secundarios, a un coeficiente de mezcla para aplicar a los respectivos flujos secundarios.

5. El método de cualquiera de las reivindicaciones anteriores, en el que el flujo de bits comprende una pluralidad de tramas de tiempo, y en el que los datos que indican la estructura de datos de presentación seleccionada entre las una o más estructuras de datos de presentación son asignables independientemente a cada trama de tiempo.

6. El método de la reivindicación 5, que comprende adicionalmente:

a partir del flujo de bits, y para la primera de dicha pluralidad de tramas de tiempo, extraer una o más estructuras de datos de presentación, y

a partir del flujo de bits, y para la segunda de dicha pluralidad de tramas de tiempo, extraer una o más estructuras de datos de presentación diferentes de las dichas una o más estructuras de datos de presentación extraídas de la primera de dicha pluralidad de tramas de tiempo,

y en el que los datos que indican la estructura de datos de presentación seleccionada indican una estructura de datos de presentación seleccionada para la trama de tiempo para la que se asigna.

7. Un decodificador para procesar un flujo (P) de bits que comprende una pluralidad de flujos secundarios (412) de contenido, cada uno de los cuales representa una señal de audio, comprendiendo, el decodificador, uno o más componentes configurados para realizar el método de cualquiera de las reivindicaciones 1 - 6.

8. Un producto de programa informático que comprende instrucciones, las cuales, cuando son ejecutadas por un dispositivo o sistema informático, realizan el método de cualquiera de las reivindicaciones 1-6.

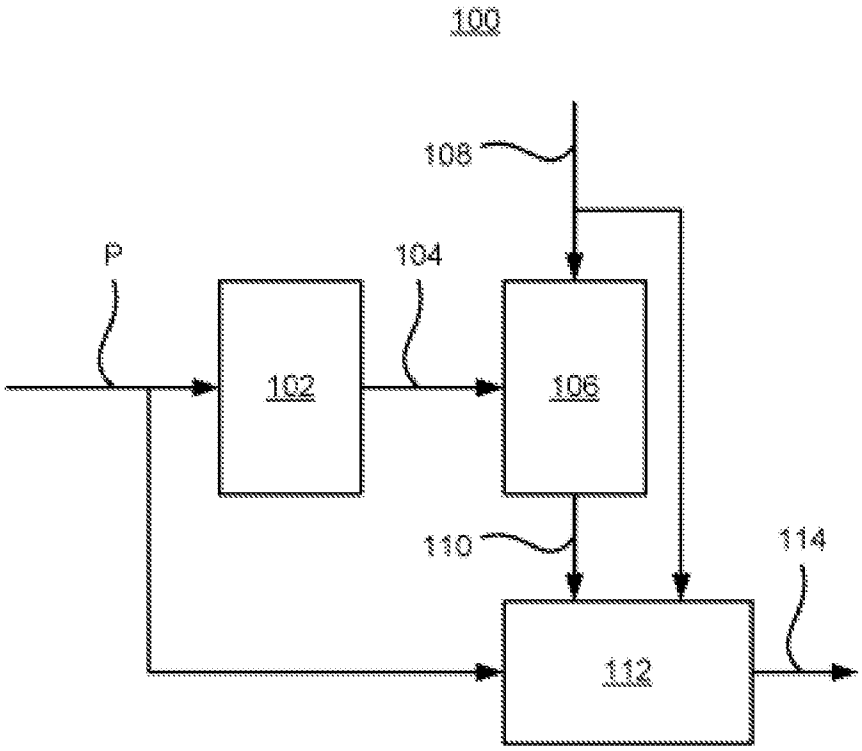
100
108
P
104
102
106
110
114
112

Fig. 1

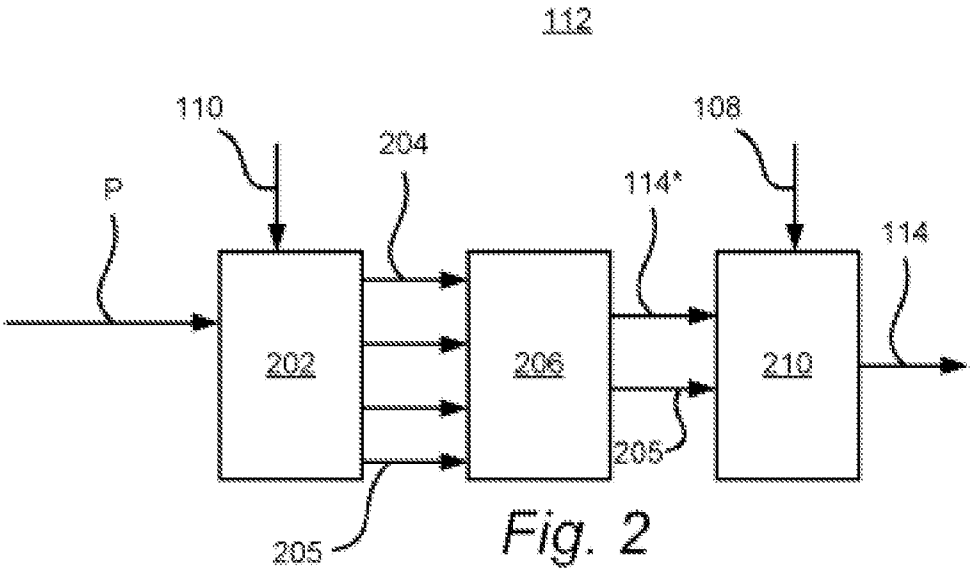
112
110
204
108
P
114*
114
202
206
210
205
205

Fig. 2

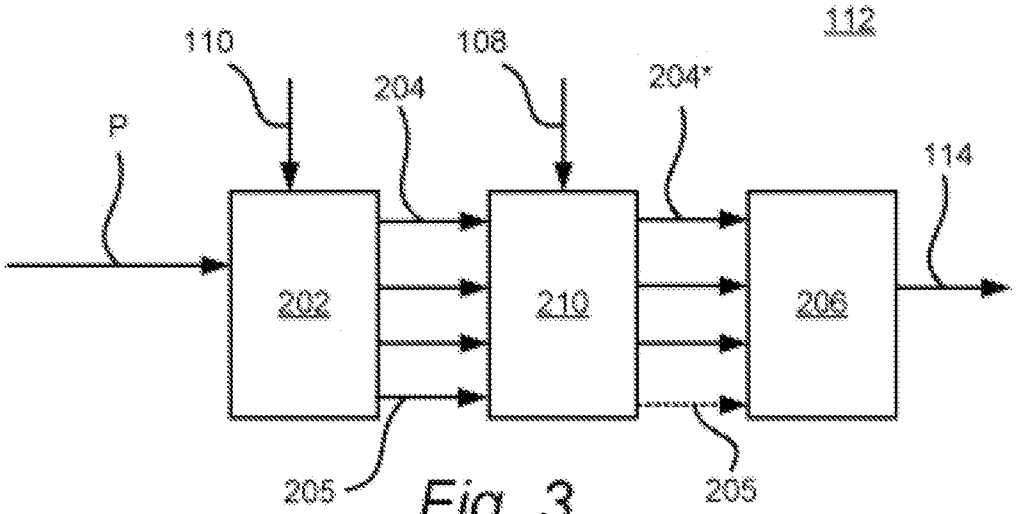
112
110
108
204'
204
P
114
202
210
206
205
205

Fig. 3

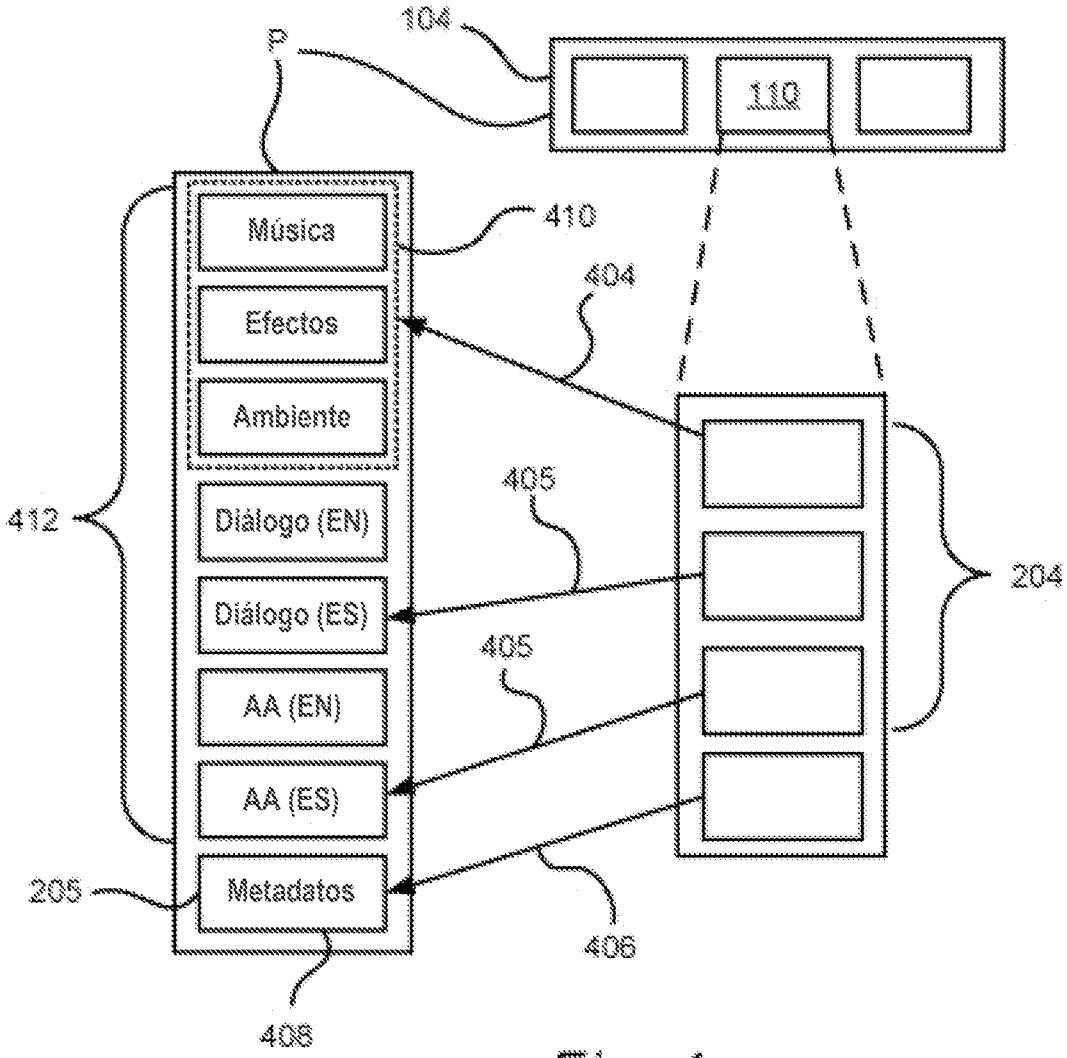
104
P
110
Música
410
404
Efectos
Ambiente
405
412
Diálogo (EN)
Diálogo (ES)
204
405
AA (EN)
AA (ES)
205
Metadatos
406
408

Fig. 4

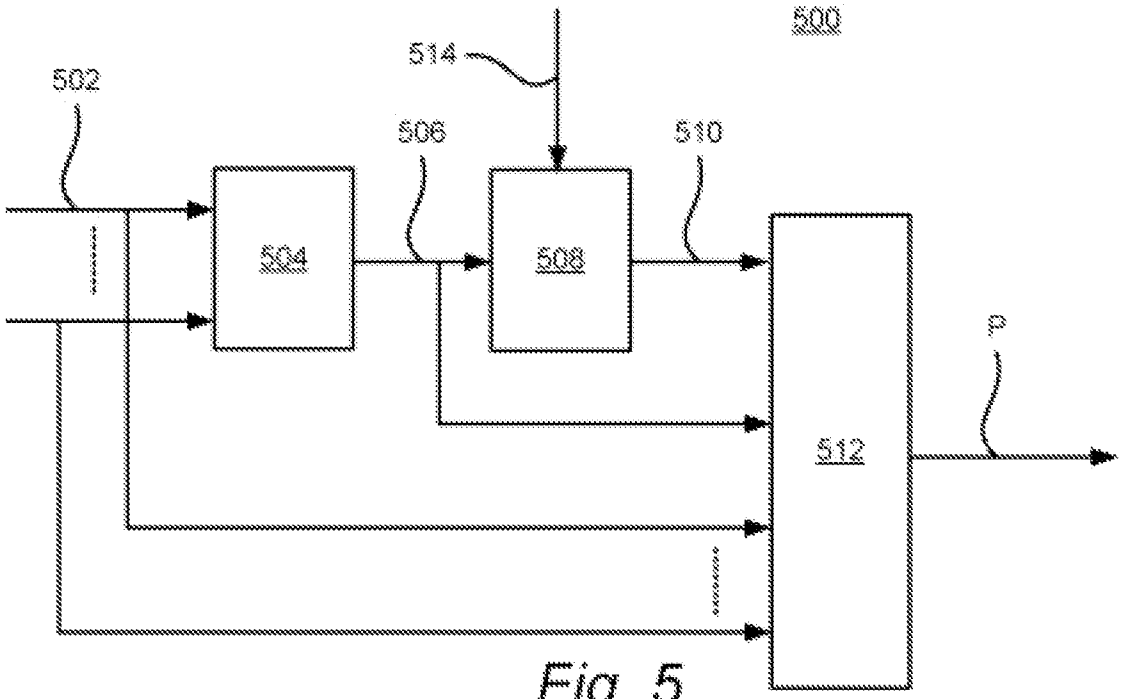
500
514
502
506
510
504
508
512
P

Fig. 5

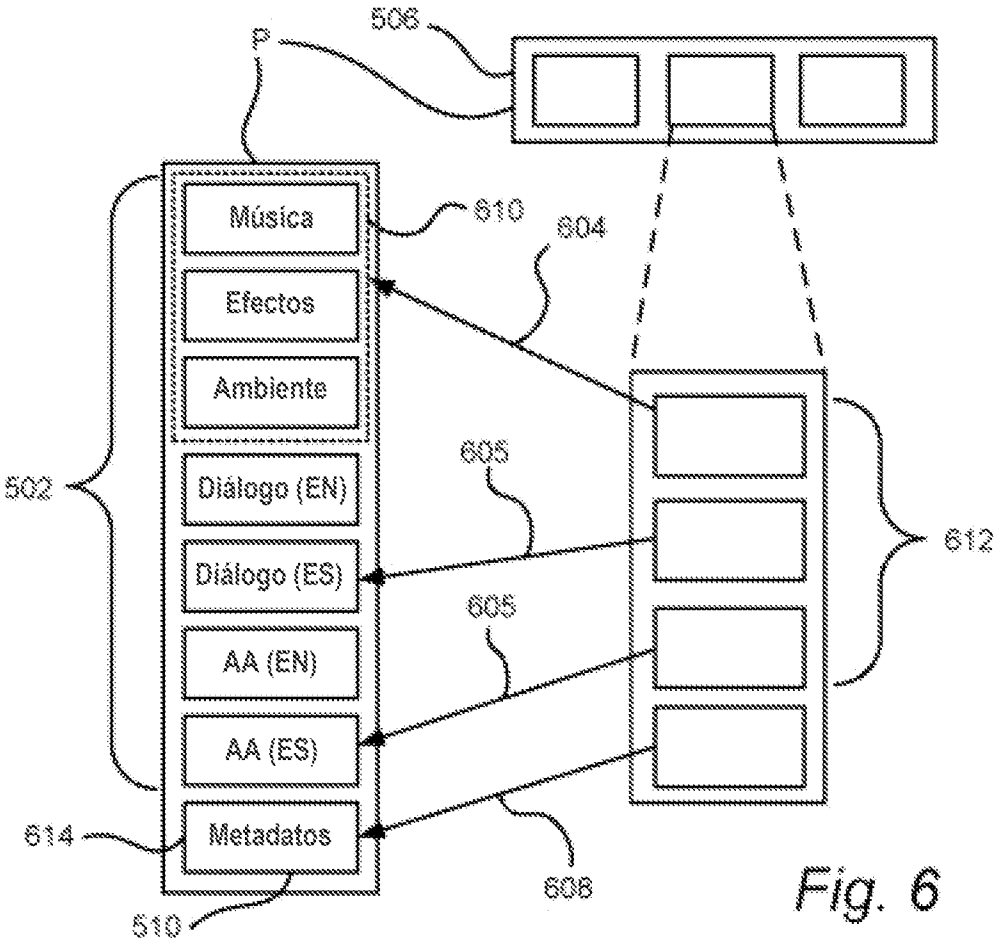
P
506
Música
610
604
Efectos
Ambiente
502
Diálogo (EN)
605
Diálogo (ES)
612
605
AA (EN)
AA (ES)
614
Metadatos
608
510

Fig. 6