

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 984 840**

51 Int. Cl.:

H04R 29/00 (2006.01)

H04S 7/00 (2006.01)

H04R 3/12 (2006.01)

H04R 27/00 (2006.01)

H04S 3/00 (2006.01)

H04S 5/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **27.06.2012** **E 21169907 (9)**

97 Fecha y número de publicación de la concesión europea: **19.06.2024** **EP 3893521**

54 Título: **Sistema y método para la generación, codificación e interpretación informática (o renderización) de señales de audio adaptativo**

30 Prioridad:

01.07.2011 US 201161504005 P
20.04.2012 US 201261636429 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
31.10.2024

73 Titular/es:

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

72 Inventor/es:

ROBINSON, CHARLES Q.;
TSINGOS, NICOLAS R. y
CHABANNE, CHRISTOPHE

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 984 840 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

Sistema y método para la generación, codificación e interpretación informática (o renderización) de señales de audio adaptativo

Referencia cruzada a solicitudes relacionadas

Esta solicitud es una solicitud divisional europea de la solicitud de patente Euro-PCT EP12743261.5, presentada el 27 de junio de 2012.

Esta solicitud reivindica prioridad de la solicitud provisional de EE.UU. nº 61/504005 presentada el 1 de julio de 2011 y la solicitud provisional de EE.UU. nº 61/636429 presentada el 20 de abril de 2012.

Campo técnico

Una o más implantaciones se refieren en general al procesamiento de señales de audio, y, más específicamente, al procesamiento de audio basado en canales y objetos híbridos para su uso en el cine, el hogar y otros entornos.

Antecedentes

No debe asumirse que el tema que se analice en la sección de antecedentes sea de técnica anterior simplemente por su mención en la sección de antecedentes. De manera similar, no se debe suponer que un problema mencionado en la sección de antecedentes o asociado con el tema de la sección de antecedentes haya sido anteriormente revisado en la técnica anterior. Un tema en la sección de antecedentes representa simplemente diferentes enfoques, que en sí mismos pueden ser también invenciones.

Desde la introducción del sonido en las películas ha habido una evolución constante de la tecnología utilizada para capturar la intención artística del creador para la banda sonora de la película y reproducirla con precisión en un entorno cinematográfico. Un papel fundamental del sonido cinematográfico es apoyar el argumento que se muestra en la pantalla. Las pistas de sonido de cine típicas comprenden muchos elementos de sonido diferentes correspondientes a elementos e imágenes en la pantalla, diálogos, ruidos y efectos de sonido que emanan de diferentes elementos en pantalla y se combinan con música de fondo y efectos ambientales para crear en el público una experiencia integral. La intención artística de los creadores y productores representa su deseo de que estos sonidos se reproduzcan de una manera que se corresponda lo más certeramente con lo que se muestra en la pantalla con respecto a la posición, intensidad, movimiento y otros parámetros similares a tener en cuenta en la fuente de sonido.

La autoría, distribución y reproducción cinematográficas actuales adolecen de limitaciones que restringen la creación de audios que de verdad proporcionen una experiencia de inmersión y realista. Los sistemas de audio tradicionales basados en canales envían contenido de audio en forma de alimentación de altavoces a altavoces individuales en un entorno de reproducción, tales como los sistemas estéreo y 5.1. La introducción del cine digital ha creado nuevos estándares para el sonido en la película, tales como la incorporación de hasta 16 canales de audio para permitir una mayor creatividad a los creadores de contenido y una experiencia auditiva más envolvente y realista al público. La introducción de los sistemas de sonido envolvente 7.1 ha proporcionado un nuevo formato que aumenta el número de canales de sonido envolvente al dividir los canales de sonido envolvente derecho e izquierdo existentes en cuatro zonas, aumentando de este modo el alcance, para los diseñadores de sonido y mezcladores, de controlar la posición de los elementos de audio en la sala de cine.

Para mejorar adicionalmente la experiencia del oyente, la reproducción de sonido en entornos virtuales tridimensionales se ha convertido en un área de mayor investigación y desarrollo. La presentación espacial del sonido utiliza objetos de audio, que son señales de audio con descripciones paramétricas asociadas de fuente, de posición de fuente aparente (por ejemplo, coordenadas 3D), ancho de fuente aparente y otros parámetros. El audio basado en objetos se utiliza cada vez más para muchas aplicaciones multimedia actuales, tales como películas digitales, videojuegos, simuladores y vídeo 3D.

Es fundamental ir más allá de la alimentación de altavoces tradicional y del audio basado en canales como medio para distribuir audio espacial, y ha provocado un considerable interés la descripción del audio basada en modelos que promete al oyente/expositor la libertad de seleccionar la configuración de reproducción que se adapte a sus necesidades individuales o a su presupuesto, con el audio renderizado específicamente para la configuración elegida. Con alto nivel, existen cuatro formatos principales de descripción de audio espacial en la actualidad: alimentación de altavoz, en el que el audio se describe como señales destinadas a los altavoces en posiciones nominales de altavoz; alimentación de micrófono, en el que el audio se describe como señales capturadas por micrófonos virtuales o reales en una matriz predefinida; descripción basada en modelos, en el que el audio se describe en términos de una secuencia de eventos de audio en posiciones definidas; y binaural, en el que el audio es descrito por las señales que llegan a los oídos de los oyentes. Estos cuatro formatos de descripción a menudo se asocian con una o más tecnologías de renderización que convierten las señales de audio en alimentación de

altavoz. Las tecnologías de renderización actuales incluyen la panoramización, en la que el torrente de audio se convierte en alimentación de altavoz usando un conjunto de leyes de panoramización y posiciones de altavoz conocidas o asumidas (típicamente renderizadas antes de la distribución); la ambisónica, en la que las señales del micrófono se convierten en alimentación para una matriz escalable de altavoces (típicamente renderizados después de la distribución); la WFS (síntesis de campo de ondas) en la que los eventos de sonido se convierten en las señales de altavoz apropiadas para sintetizar el campo de sonido (siendo típicamente renderizadas después de la distribución); y la binaural, en la que las señales binaurales L/R (izquierda/derecha) se suministran al oído L/R, típicamente utilizando auriculares, pero también mediante el uso de altavoces y la anulación de diafonía (renderizada antes o después de la distribución). De estos formatos, el formato de alimentación por altavoz es el más común, porque es simple y efectivo. Los mejores resultados sonoros (más precisos, más fiables) se logran mezclando/monitorizando y distribuyendo a los altavoces la alimentación directamente, ya que no hay procesamiento entre el creador de contenido y el oyente. Si el sistema de reproducción se conoce de antemano, una descripción de la alimentación del altavoz proporciona generalmente la mayor fidelidad. Sin embargo, en muchas aplicaciones prácticas, se desconoce el sistema de reproducción. La descripción basada en modelos se considera la más adaptable porque no asume conceptos sobre la tecnología de renderización, y, por lo tanto, se aplica más fácilmente a cualquier tecnología de renderización. Si bien la descripción basada en modelos captura de manera eficiente la información espacial, se vuelve muy ineficaz a medida que aumenta el número de fuentes de audio.

Durante muchos años, los sistemas de cine han presentado canales de pantalla discretos en forma de canales izquierdo, central, derecho y, ocasionalmente, "interior izquierdo" e "interior derecho". Estas fuentes discretas tienen generalmente suficiente respuesta de frecuencia y manejo de potencia para permitir que los sonidos se coloquen con precisión en diferentes áreas de la pantalla y para permitir la coincidencia de timbres cuando los sonidos se mueven o se panoramizan entre ubicaciones. Los desarrollos recientes para mejorar la experiencia del oyente intentan reproducir con precisión la ubicación de los sonidos con relación al oyente. En una configuración 5.1, las 'zonas' envolventes comprenden una serie de altavoces, todos los cuales llevan la misma información de audio dentro de cada zona envolvente izquierda o envolvente derecha. Tales matrices pueden ser efectivas con efectos de envolventes difusos o "ambientales"; sin embargo, en la vida cotidiana, muchos efectos de sonido se originan en fuentes puntuales colocadas al azar. Por ejemplo, en un restaurante, la música ambiental puede reproducirse aparentemente desde todos lados, pero los sonidos sutiles pero discretos se originan en puntos específicos: una persona charlando desde un lugar, el ruido de un cuchillo en un plato desde otro. Ser capaz de colocar tales sonidos de manera discreta alrededor del auditorio puede añadir una sensación más pronunciada de realismo sin caer en lo obvio. Los sonidos que vienen de arriba también son un componente importante de la definición de sonido envolvente. En el mundo real, los sonidos se propagan en todas direcciones y no siempre en un único plano horizontal. Se puede conseguir una mayor sensación de realismo si se puede escuchar sonido procedente de arriba, en otras palabras, desde el "hemisferio superior". Sin embargo, los sistemas actuales no ofrecen una reproducción de sonido verdaderamente precisa para diferentes tipos de audio en una variedad de entornos de reproducción diferentes. Se requiere un gran esfuerzo de procesamiento, conocimiento y configuración de los entornos de reproducción reales utilizando los sistemas existentes para intentar una representación precisa de los sonidos específicos de la ubicación, lo que hace que los sistemas actuales no sean prácticos para la mayoría de las aplicaciones.

Lo que se necesita es un sistema que soporte múltiples canales de pantalla, dando como resultado una mayor definición y una coherencia audiovisual mejorada para los sonidos o diálogos en pantalla, y la capacidad de posicionar con precisión las fuentes en cualquier lugar de las zonas envolventes para mejorar la transición audiovisual de la pantalla a la sala. Por ejemplo, si un personaje en la pantalla mira dentro de la sala hacia una fuente de sonido, el ingeniero de sonido ("el mezclador") debe tener la capacidad de posicionar con precisión el sonido para que coincida con la línea de visión del personaje y el efecto sea consistente para el público en su totalidad. Sin embargo, en una mezcla de sonido envolvente tradicional 5.1 o 7.1, el efecto depende en gran medida de la posición del oyente en el asiento, lo que es una desventaja para la mayoría de los entornos de audición a gran escala. El aumento de la resolución envolvente crea nuevas oportunidades para usar el sonido de una manera centrada en la sala en contraste con el enfoque tradicional, donde el contenido se crea asumiendo un sólo oyente en el "punto óptimo".

Aparte de las cuestiones espaciales, los actuales sistemas multicanal del estado de la técnica sufren con respecto al timbre. Por ejemplo, la calidad tímbrica de algunos sonidos, tales como el estallido de vapor de una tubería rota, puede verse afectada por ser reproducida por una matriz de altavoces. La capacidad de dirigir sonidos específicos a un sólo altavoz le da al mezclador la oportunidad de eliminar los artefactos de la reproducción de matriz y ofrecer una experiencia más realista al público. Tradicionalmente, los altavoces de sonido envolvente no soportan el mismo intervalo completo de frecuencia y nivel de audio que soportan los canales de pantalla grande. Históricamente, este hecho ha dado problemas a los mezcladores, al reducir su capacidad para mover libremente sonidos de intervalo completo de la pantalla a la sala. Como resultado, los propietarios de las salas de cine no se han sentido obligados a actualizar la configuración de sus canales de sonido envolvente, lo que impide la adopción generalizada de instalaciones de mayor calidad.

El documento US2006/206221 A1 divulga de manera general un sistema y método para proporcionar control individual sobre objetos de sonido que se reciben discretamente en un dispositivo de reproducción. Los objetos de

sonido pueden ser representativos de fuentes de sonido individuales, y pueden incluir tanto contenido de sonido producido por los objetos de sonido como otras características de los objetos de sonido. Las otras características de los objetos de sonido pueden comprender uno o más de un patrón de directividad, información de posición, un algoritmo de movimiento de objeto y/u otras características.

El documento US2010/135510 A1 divulga de manera general un aparato de generación/reproducción de contenidos de audio basado en objetos. El aparato de generación/reproducción de contenidos de audio basado en objetos puede incluir una unidad de obtención de señales de audio de objetos para obtener una pluralidad de señales de audio de objetos mediante la grabación de una pluralidad de señales de fuentes de sonido, una unidad de obtención de información de espacio de grabación para obtener información de espacio de grabación con respecto a un espacio de grabación de la pluralidad de señales de fuente de sonido, una unidad de obtención de información de ubicación de fuente de sonido para obtener información de ubicación de sonido de la pluralidad de señales de fuente de sonido, y una unidad de codificación para generar contenidos de audio basado en objetos mediante codificación de al menos uno de la pluralidad de señales de audio de objeto, la información de espacio de grabación y la información de ubicación de fuente de sonido.

Breve resumen de las realizaciones

De acuerdo con la presente divulgación, se proporcionan sistemas para procesar señales de audio, un método de autoría de señales de audio para renderización, un método de renderización de señales de audio, un programa de ordenador, así como un medio de almacenamiento legible por ordenador, que tienen las características de respectivas reclamaciones independientes. Las reivindicaciones dependientes se refieren a realizaciones preferidas.

Se describen sistemas y métodos para un formato de sonido de cine y un sistema de procesamiento que incluyen un nuevo trazado de altavoces (una configuración de canal) y un formato de descripción espacial asociado. Se definen un sistema y un formato de audio adaptativo que soporta múltiples tecnologías de renderización. Los torrents de audio se transmiten junto con metadatos que describen la "intención del mezclador", incluyendo la posición deseada del torrente de audio. La posición se puede expresar como un canal con nombre (desde dentro de la configuración de canal predefinida) o como información de posición tridimensional. Este formato de canales más objetos combina métodos óptimos de descripción de escenas de audio basados en canales y en modelos. Los datos de audio para el sistema de audio adaptativo comprenden una serie de torrents monofónicos de audio independientes. Cada torrente tiene asociados metadatos que especifican si el torrente es un torrente basado en canales o en objetos. Los torrents basados en canales tienen información de renderización codificada por medio del nombre del canal; y los torrents basados en objetos tienen información de ubicación codificada mediante expresiones matemáticas codificadas en metadatos asociados adicionales. Los torrents de audio independientes originales se empaquetan como un único torrente de bits en serie que contiene todos los datos de audio. Esta configuración permite que el sonido se renderice de acuerdo con un marco alocéntrico de referencia, en el que la ubicación de renderización de un sonido se basa en las características del entorno de reproducción (por ejemplo, en el tamaño de la sala, la forma, etc.) para corresponder a la intención del mezclador. Los metadatos de posición del objeto contienen el marco alocéntrico apropiado de la información de referencia necesaria para reproducir el sonido correctamente utilizando las posiciones de altavoz disponibles en una sala que está configurada para reproducir el contenido de audio adaptativo. Esto permite que el sonido se mezcle de manera óptima para un entorno de reproducción particular que puede ser diferente del entorno de mezcla experimentado por el ingeniero de sonido.

El sistema de audio adaptativo mejora la calidad de audio en diferentes habitaciones a través de beneficios tales como una mejor ecualización de la sala y una gestión envolvente de graves, de modo que el mezclador pueda acceder libremente a los altavoces (ya sea en pantalla o fuera de pantalla) sin tener que pensar en la coincidencia tímbrica. El sistema de audio adaptativo agrega la flexibilidad y la potencia de los objetos de audio dinámicos a los flujos de trabajo tradicionales basados en canales. Estos objetos de audio permiten a los creadores controlar elementos de sonido discretos independientemente de las configuraciones específicas de los altavoces de reproducción, incluidos los altavoces superiores. El sistema introduce también nuevas eficiencias en el proceso de postproducción, permitiendo a los ingenieros de sonido reflejar de manera eficiente toda su intención, y monitorizar después en tiempo real, o generar automáticamente, versiones de sonido envolvente 7.1 y 5.1.

El sistema de audio adaptativo simplifica la distribución encapsulando la esencia del audio y la intención artística en un único archivo de pista dentro de un procesador digital de cine, que se puede volver a reproducir fielmente en una amplia gama de configuraciones de sala. El sistema proporciona una reproducción óptima de la intención artística cuando la mezcla y la renderización utilizan la misma configuración de canal y un único inventario con adaptación descendente para la configuración de renderización, es decir, una mezcla descendente.

Estas y otras ventajas se proporcionan a través de unas realizaciones que están dirigidas a una plataforma de sonido de cine, que abordan las limitaciones del sistema actual y que ofrecen una experiencia de audio más allá de la que ofrecen los sistemas disponibles actualmente.

Breve descripción de los dibujos

En los siguientes dibujos se utilizan números de referencia similares para hacer referencia a elementos similares. Aunque las siguientes figuras representen diversos ejemplos, la una o más implantaciones no se limitan a los ejemplos representados en las figuras.

- 5 La figura 1 es una descripción general de alto nivel de un entorno de reproducción y creación de audio que utiliza un sistema de audio adaptativo.
- La figura 2 ilustra la combinación de canales y datos basados en objetos para producir una mezcla de de audio adaptativo.
- 10 La figura 3 es un diagrama de bloques que ilustra el flujo de trabajo de creación, empaquetado y renderización de contenido de audio adaptativo.
- La figura 4 es un diagrama de bloques de una etapa de reproducción de un sistema de audio adaptativo.
- 15 La figura 5 es una tabla que enumera los tipos de metadatos y los elementos de metadatos asociados para el sistema de audio adaptativo.
- La figura 6 es un diagrama que ilustra una postproducción y masterización para un sistema de audio adaptativo.
- 20 La figura 7 es un diagrama de un flujo de trabajo de ejemplo para un proceso de empaquetado de cine digital que utiliza archivos de audio adaptativos.
- La figura 8 es una vista desde arriba de un trazado de ejemplo de ubicaciones de altavoces sugeridas para su uso con un sistema de audio adaptativo en un auditorio típico.
- 25 La figura 9 es una vista frontal de una ubicación de ejemplo de ubicaciones de altavoces sugeridas en la pantalla para su uso en un auditorio típico.
- La figura 10 es una vista lateral de un trazado de ejemplo de ubicaciones de altavoces sugeridas para su uso con un sistema de audio adaptativo en el auditorio típico.
- La figura 11 es un ejemplo de posicionamiento de los altavoces envolventes superiores y de los altavoces envolventes laterales con relación al punto de referencia.
- 35

Descripción detallada

- Se describen sistemas y métodos para un sistema de audio adaptativo y la señal de audio a él asociada y un formato de datos que soporta múltiples tecnologías de renderización. Los aspectos de una o más realizaciones descritas en el presente documento pueden implantarse en un sistema de audio o audiovisual que procese la información de fuente de audio en un sistema de mezcla, renderización y reproducción que incluya uno o más ordenadores o dispositivos de procesamiento que ejecuten instrucciones de equipo lógico informático (software). Cualquiera de las realizaciones descritas puede usarse sola o junto con otras en cualquier combinación. Aunque diversas realizaciones pueden haber estado motivadas por diversas deficiencias de la técnica anterior, que pueden analizarse o aludirse en uno o más lugares de la memoria descriptiva, las realizaciones no abordan necesariamente ninguna de estas deficiencias. En otras palabras, diferentes realizaciones pueden abordar diferentes deficiencias que pueden analizarse en la especificación. Algunas realizaciones pueden abordar sólo parcialmente algunas deficiencias o sólo una deficiencia, la cual o las cuales se puede analizar en la especificación, y algunas realizaciones pueden no abordar ninguna de estas deficiencias.
- 40
- 45
- 50

Para los fines de la presente descripción, los siguientes términos tienen los siguientes significados asociados:

- Canal o canal de audio: una señal de audio monofónica o un torrente de audio más los metadatos en los que la posición se codifica como un ID de canal, por ejemplo, de sonido envolvente frontal izquierdo o superior derecho. Un objeto de canal puede accionar varios altavoces, por ejemplo, los canales envolventes izquierdos (Ls) alimentarán a todos los altavoces de la matriz Ls.
- 55

- Configuración de canal: un conjunto predefinido de zonas de altavoces con ubicaciones nominales asociadas, por ejemplo, 5.1, 7.1, etc.; 5.1 se refiere a un sistema de audio de sonido envolvente de seis canales que tiene canales frontales izquierdo y derecho, canal central, dos canales envolventes y un canal de subwoofer; 7.1 se refiere a un sistema de sonido envolvente de ocho canales que añade dos canales de sonido envolvente adicionales al sistema 5.1. Entre los ejemplos de configuraciones 5.1 y 7.1 se incluyen los sistemas de sonido envolvente Dolby®.
- 60

- Altavoz: un transductor de audio o conjunto de transductores que generan una señal de audio.
- 65

Zona de altavoces: una matriz de uno o más altavoces puede tener una referencia única y que recibe una única señal de audio, por ejemplo de sonido envolvente izquierdo como se encuentra típicamente en el cine, y en particular para la exclusión o inclusión en la renderización de objetos.

- 5 Canal de altavoz o canal de alimentación de altavoz: un canal de audio que está asociado con un altavoz o zona de altavoces con nombre dentro de una configuración de altavoz definida. Un canal de altavoz se renderiza nominalmente utilizando la zona de altavoz asociada.

- 10 Grupo de canales de altavoces: un conjunto de uno o más canales de altavoces correspondientes a una configuración de canal (por ejemplo, una pista estéreo, una pista mono, etc.)

- Objeto o canal de objeto: uno o más canales de audio con una descripción paramétrica de fuente, como la posición de fuente aparente (por ejemplo, con coordenadas 3D), el ancho de fuente aparente, etc. Un torrente de audio más metadatos en los que la posición se codifica como una posición 3D en el espacio.

- 15 Programa de audio: el conjunto completo de los canales de altavoz y/o de los canales de objeto y los metadatos asociados que describen la presentación de audio espacial deseada.

- 20 Referencia alocéntrica: una referencia espacial en la que los objetos de audio se definen con relación a las características dentro del entorno de representación, tal como las paredes y esquinas de la sala, las ubicaciones estándar de altavoz y la ubicación de pantalla (por ejemplo, la esquina frontal izquierda de una sala).

- 25 Referencia egocéntrica: una referencia espacial en la que los objetos de audio se definen con relación a la perspectiva del oyente (del público), y que a menudo se especifica con respecto a ángulos con relación a un oyente (por ejemplo, 30 grados a la derecha del oyente).

Trama: las tramas son segmentos cortos independientemente decodificables en los que se divide un programa de audio total. El límite y la tasa de trama de audio están típicamente alineados con los fotogramas de vídeo.

- 30 Audio adaptativo: las señales de audio basadas en canal y/o basadas en objeto, más los metadatos que generan las señales de audio, en base al entorno de reproducción.

- El sistema de procesamiento y de formato de sonido de cine descrito en el presente documento, también denominado "sistema de audio adaptativo", utiliza una nueva descripción de audio espacial y de tecnología de reproducción para permitir una mejor inmersión del público, más control artístico, flexibilidad y escalabilidad del sistema, y facilidad de instalación y mantenimiento. Una plataforma de audio de cine incluye varios componentes discretos que incluyen herramientas de mezcla, empaquetador/codificador, desempaquetado/decodificador, componentes de renderización y mezcla final en la sala de cine, nuevos diseños de altavoz y amplificadores en red. El sistema incluye recomendaciones para una nueva configuración de canal que vaya a ser utilizada por creadores de contenido y expositores. El sistema utiliza una descripción basada en modelos que soporta varias características tales como: inventario único con adaptación ascendente y descendente a la configuración de renderización, es decir, renderización de retardo y habilitación de un uso óptimo de los altavoces disponibles; envolvente de sonido mejorado, incluyendo la mezcla optimizada descendente para evitar la correlación entre canales; resolución espacial aumentada a través de matrices direccionales (por ejemplo, un objeto de audio asignado dinámicamente a uno o más altavoces dentro de una matriz envolvente); y soporte para métodos de renderización alternativos.

- La figura 1 es una descripción general de alto nivel de un entorno de reproducción y creación de audio que utiliza un sistema de audio adaptativo. Como se muestra en la figura 1, un entorno 100 comprehensivo y de extremo a extremo incluye componentes de creación, empaquetado, distribución y reproducción/renderización de contenido en un amplio número de dispositivos de punto final y casos de uso. El sistema general 100 se origina con contenido capturado de y para varios casos de uso diferentes que comprenden diferentes experiencias de usuario 112. El elemento 102 de captura de contenido incluye, por ejemplo, cine, TV, difusión en vivo, contenido generado por el usuario, contenido grabado, juegos, música y similares, y puede incluir contenido audio/visual o de audio puro. El contenido, a medida que avanza a través del sistema 100 desde la etapa 102 de captura hasta la experiencia 112 del usuario final, atraviesa varios pasos clave de procesamiento a través de componentes discretos del sistema. Estos pasos del proceso incluyen el procesamiento previo del audio 104, las herramientas de autoría y los procesos 106, la codificación mediante un códec 108 de audio que captura, por ejemplo, datos de audio, metadatos adicionales e información de reproducción, y canales de objetos. Se pueden aplicar diversos efectos de procesamiento, tales como compresión (con o sin pérdida), encriptación y similares a los canales de objetos para una distribución eficiente y segura a través de diversos medios. Luego, se aplican procesos 110 de decodificación y renderización específicos del punto final apropiados para reproducir y trasladar una experiencia particular 112 de usuario de audio adaptativo. La experiencia 112 de audio representa la reproducción del contenido de audio o audio/visual a través de altavoces y dispositivos de reproducción apropiados, y puede representar cualquier entorno en el que un oyente esté experimentando la reproducción del contenido capturado, tal como un cine, una sala de conciertos, un sala de cine al aire libre, una casa o una habitación, una cabina de escucha, un automóvil, una

consola de juegos, un sistema de auriculares o audífonos, un sistema de megafonía (PA) o cualquier otro entorno de reproducción.

El sistema 100 incluye un códec 108 de audio que es capaz de distribuir y almacenar eficazmente programas de audio multicanal, y que, por lo tanto, puede denominarse códec "híbrido". El códec 108 combina datos de audio tradicionales basados en canal con metadatos asociados para producir objetos de audio que facilitan la creación y suministro de audio que están adaptados y optimizados para la renderización y reproducción en entornos que pueden ser diferentes del entorno de mezcla. Esto permite al ingeniero de sonido codificar su intención con respecto a cómo el oyente debe escuchar el audio final, en base al entorno de escucha real del oyente.

Los códecs de audio convencionales basados en canal funcionan asumiendo que el programa de audio será reproducido por una matriz de altavoces en posiciones predeterminadas con respecto al oyente. Para crear un programa de audio multicanal completo, los ingenieros de sonido suelen típicamente mezclar un gran número de torrentes de audio separados (por ejemplo, diálogo, música, efectos) para crear la impresión general deseada. Las decisiones de mezcla de audio se toman típicamente escuchando el programa de audio reproducido por una matriz de altavoces en las posiciones predeterminadas, por ejemplo, un sistema 5.1 o 7.1 particular en una sala de cine específica. La señal mezclada final sirve como entrada al códec de audio. Para la reproducción, los campos de sonido espacialmente precisos se consiguen sólo cuando los altavoces están colocados en las posiciones predeterminadas.

Una nueva forma de codificación de audio denominada codificación de objetos de audio proporciona distintas fuentes de sonido (objetos de audio) como entrada al codificador en forma de torrentes de audio separados. Los ejemplos de objetos de audio incluyen pistas de diálogo, instrumentos individuales, efectos de sonido individuales y otras fuentes puntuales. Cada objeto de audio está asociado con parámetros espaciales, que pueden incluir, pero sin limitarse a, información sobre la posición del sonido, el ancho del sonido y la velocidad. Los objetos de audio y los parámetros asociados se codifican luego para su distribución y almacenamiento. La renderización y la mezcla finales del objeto de audio se realizan en el extremo de recepción de la cadena de distribución de audio, como parte de la reproducción del programa de audio. Este paso puede basarse en el conocimiento de las posiciones reales de los altavoces de modo que el resultado sea un sistema de distribución de audio que se puede personalizar para las condiciones de escucha específicas del usuario. Las dos formas de codificación, basadas en canal y basadas en objeto, se comportan de manera óptima para diferentes condiciones de señal de entrada. Los codificadores de audio basados en canal son generalmente más eficientes para codificar señales de entrada que contienen mezclas densas de diferentes fuentes de audio y para sonidos difusos. Por el contrario, los codificadores de objetos de audio son más eficientes para codificar un pequeño número de fuentes de sonido altamente direccionales.

Los métodos y componentes del sistema 100 comprenden un sistema de codificación, distribución y decodificación de audio configurado para generar uno o más torrentes de bits que contienen tanto elementos de audio convencionales basados en canales como elementos de codificación de objetos de audio. Tal enfoque combinado proporciona eficiencia de codificación y flexibilidad de renderización mayores en comparación con los enfoques basados ya sea en canales o en objetos tomados por separado.

Otros aspectos incluyen ampliar un códec predefinido de audio basado en canales de una manera retrocompatible para incluir elementos de codificación de objetos de audio. Se define una nueva 'capa de ampliación' que contiene los elementos de codificación del objeto de audio y se añade a la capa 'base' o 'retrocompatible' del torrente de bits del códec de audio basado en canales. Este enfoque permite que uno o más torrentes de bits, que incluyen la capa de ampliación, sean procesados por decodificadores heredados, al tiempo que proporciona una experiencia de escucha perfeccionada para usuarios con nuevos decodificadores. Un ejemplo de una experiencia de usuario perfeccionada incluye el control de la renderización de objetos de audio. Una ventaja adicional de este enfoque es que los objetos de audio se pueden añadir o modificar en cualquier sitio a lo largo de la cadena de distribución sin decodificar/mezclar/volver a codificar el audio multicanal codificado con el códec de audio basado en canales.

Con respecto al marco de referencia, los efectos espaciales de las señales de audio son críticos para proporcionar una experiencia de inmersión para el oyente. Los sonidos que están destinados a emanar de una región específica de una pantalla o sala de visualización deben reproducirse a través de los altavoces ubicados en esa misma ubicación relativa. De este modo, el metadato de audio principal de un evento de sonido en una descripción basada en modelos es la posición, aunque también se pueden describir otros parámetros como el tamaño, la orientación, la velocidad y la dispersión acústica. Para transferir la posición, una descripción espacial de audio 3D basada en modelos requiere un sistema de coordenadas 3D. El sistema de coordenadas usado para la transmisión (euclidiana, esférica, etc.) se elige generalmente por conveniencia o compacidad, sin embargo, se pueden usar otros sistemas de coordenadas para el procesamiento de renderización. Además de un sistema de coordenadas, se requiere un marco de referencia para representar las ubicaciones de los objetos en el espacio. Para que los sistemas reproduzcan con precisión, en una variedad de entornos diferentes, el sonido basado en posición, la selección del marco de referencia adecuado puede ser un factor crítico. Con un marco de referencia alocéntrico, la posición de una fuente de audio se define con relación a las características dentro del entorno de renderización, tales como las paredes y esquinas de la habitación, las ubicaciones estándar de los altavoces y la ubicación de la pantalla. En un marco de referencia egocéntrico, las ubicaciones se representan con respecto a la perspectiva del oyente, tal como

“frente a mí, ligeramente a la izquierda”, etc. Los estudios científicos de la percepción espacial (audio y otros), han demostrado que la perspectiva egocéntrica se utiliza casi universalmente. Sin embargo, para el cine, la alocéntrica resulta generalmente más apropiada, por varias razones. Por ejemplo, la ubicación precisa de un objeto de audio es más importante cuando hay un objeto asociado en la pantalla. Usando una referencia alocéntrica, para cada posición de escucha y para cualquier tamaño de pantalla, el sonido se localizará en la misma posición relativa en la pantalla, por ejemplo, un tercio a la izquierda del centro de la pantalla. Otra razón es que los mezcladores tienden a pensar y a mezclar en términos alocéntricos, y que las herramientas de panoramización se encuentran en un marco alocéntrico (las paredes de la sala), y los mezcladores esperan que se rendericen de esa manera, por ejemplo, este sonido tiene que estar en la pantalla, este sonido tiene que estar fuera de la pantalla o desde la pared izquierda, etc.

A pesar del uso del marco de referencia alocéntrico en el entorno cinematográfico, hay algunos casos en los que un marco de referencia egocéntrico puede ser útil y más apropiado. Estos incluyen sonidos no diégéticos, es decir, aquellos que no están presentes en el “espacio de la historia”, por ejemplo música ambiental, para la que puede ser deseable una presentación egocéntrica uniforme. Otro caso son los efectos de campo cercano (por ejemplo, un zumbido de mosquito en el oído izquierdo del oyente) que requieren una representación egocéntrica. Actualmente, no existen medios para renderizar un campo tal de sonido sin utilizar auriculares o altavoces de campo muy cercano. Además, las fuentes de sonido infinitamente lejanas (y las ondas planas resultantes) parecen provenir de una posición egocéntrica constante (por ejemplo, 30 grados a la izquierda), y tales sonidos son más fáciles de describir en términos egocéntricos que en términos alocéntricos.

En algunos casos, es posible utilizar un marco de referencia alocéntrico siempre que se defina una posición de escucha nominal, mientras que otros ejemplos requieren una representación egocéntrica que aún no es posible renderizar. Aunque una referencia alocéntrica puede ser más útil y apropiada, la representación de audio debe ser ampliable, ya que muchas características nuevas, incluida la representación egocéntrica, pueden ser más deseables en determinadas aplicaciones y entornos de escucha. El sistema de audio adaptativo incluye un enfoque de descripción espacial híbrida que incluye una configuración de canal recomendada para una fidelidad óptima y para la renderización de fuentes multipunto difusas o complejas (por ejemplo, para la multitud de un estadio, para ambiente) utilizando una referencia egocéntrica, más una descripción de sonido basada en modelo alocéntrica para permitir de manera eficiente aumentar la resolución espacial y la escalabilidad.

Componentes del sistema

Con referencia a la figura 1, los datos 102 de contenido de sonido original se procesan primero en un bloque 104 de procesamiento previo. El bloque 104 de procesamiento previo del sistema 100 incluye un componente de filtrado de canal de objeto. En muchos casos, los objetos de audio contienen fuentes de sonido individuales para permitir una panorámica independiente de los sonidos. En algunos casos, tales como cuando se crean programas de audio con sonido natural o de “producción”, puede ser necesario extraer objetos de sonido individuales de una grabación que contiene múltiples fuentes de sonido. Las realizaciones incluyen un método para aislar señales de fuente independientes de una señal más compleja. Los elementos indeseables van a separarse de las señales de fuente independientes pueden incluir, pero sin limitarse a, otras fuentes de sonido independientes y ruido de fondo. Además, se puede eliminar la reverberación para recuperar fuentes de sonido “secas”.

El preprocesador 104 también incluye la funcionalidad de separación de fuentes y de detección de tipo de contenido. El sistema proporciona la generación automatizada de metadatos a través del análisis del audio de entrada. Los metadatos posicionales se derivan de una grabación multicanal a través de un análisis de los niveles relativos de entrada correlacionada entre pares de canales. La detección del tipo de contenido, tal como de “voz” o de “música”, puede lograrse, por ejemplo, mediante extracción y clasificación de características.

Herramientas de autoría

El bloque 106 de herramientas de autoría incluye características para mejorar la autoría de programas de audio optimizando la entrada y la codificación de la intención creativa del ingeniero de sonido permitiéndole crear la mezcla de audio final una vez optimizada para su reproducción en prácticamente cualquier entorno de reproducción. Esto se logra a través del uso de objetos de audio y de datos de posición asociados y codificados con el contenido de audio original. Con el fin de colocar con precisión los sonidos alrededor en un auditorio, el ingeniero de sonido necesita controlar cómo se renderizará el sonido en última instancia en función de las limitaciones y características reales del entorno de reproducción. El sistema de audio adaptativo proporciona este control al permitir que el ingeniero de sonido cambie la manera en que se diseña y se mezcla el contenido de audio a través del uso de objetos de audio y datos de posición.

Los objetos de audio se pueden considerar como grupos de elementos de sonido que pueden percibirse como emanados de una ubicación o de varias ubicaciones física/s particular/es en el auditorio. Tales objetos pueden ser estáticos o pueden moverse. En el sistema 100 de audio adaptativo, los objetos de audio están controlados por metadatos que, entre otras cosas, detallan la posición del sonido en un punto dado en el tiempo. Cuando los objetos se monitorizan o reproducen en una sala de cine, se renderizan de acuerdo con los metadatos posicionales utilizando los altavoces que están presentes, en lugar de ser necesariamente emitirse a un canal físico. Una pista de

una sesión puede ser un objeto de audio, y los datos de panorámica estándar son análogos a los metadatos posicionales. De esta manera, el contenido colocado en la pantalla puede desplazarse de manera efectiva de la misma manera que con el contenido basado en canales, pero el contenido envolvente colocado alrededor se puede renderizar a un altavoz individual si se desea. Si bien el uso de objetos de audio proporciona el control deseado para efectos discretos, otros aspectos de la banda sonora de una película funcionan de manera efectiva en un entorno basado en canales. Por ejemplo, muchos efectos ambientales o la reverberación se aprovechan más al ser alimentados a matrices de altavoces. Aunque éstos podrían tratarse como objetos con suficiente ancho para llenar una matriz, es provechoso conservar alguna funcionalidad basada en canal.

El sistema de audio adaptativo soporta 'lechos' además de objetos de audio, donde los lechos son efectivamente submezclas o vástagos basados en canal. Estos se pueden suministrar para la reproducción final (renderización) ya sea individualmente o combinados en un solo lecho, dependiendo de la intención del creador de contenido. Estos lechos se pueden crear en diferentes configuraciones basadas en canal, como 5.1, 7.1, y son ampliables a formatos más amplios, tales como 9.1, y a matrices que incluyen altavoces aéreos.

La figura 2 ilustra la combinación de canales y datos basados en objetos para producir una mezcla de audio adaptativo. Como se muestra en el proceso 200, los datos 202 basados en canales, que, por ejemplo, pueden ser datos de sonido envolvente 5.1 o 7.1 proporcionados en forma de datos modulados por código de pulso (PCM), se combinan con datos 204 de objetos de audio para producir una mezcla 208 de audio. Los datos 204 de objeto de audio se producen combinando los elementos de los datos originales basados en el canal con metadatos asociados que especifican ciertos parámetros pertenecientes a la ubicación de los objetos de audio.

Como se muestra conceptualmente en la figura 2, las herramientas de autoría brindan la capacidad de crear programas de audio que contienen una combinación simultánea de grupos de canales de altavoces con canales de objetos. Por ejemplo, un programa de audio podría contener uno o más canales de altavoz organizados opcionalmente en grupos (o pistas, por ejemplo, una pista estéreo o 5.1), metadatos descriptivos para uno o más canales de altavoces, uno o más canales de objetos y metadatos descriptivos para uno o más canales de objetos. Dentro de un programa de audio, cada grupo de canales de altavoz y cada canal de objeto pueden representarse usando una o más tasa de muestreo diferentes. Por ejemplo, las aplicaciones de cine digital (D-Cinema) soportan tasas de muestreo de 48 kHz y 96 kHz, pero también pueden soportarse otras tasas de muestreo. Además, también se puede soportar la asimilación, el almacenamiento y la edición de canales con diferentes tasas de muestreo.

La creación de un programa de audio requiere el paso de diseñar el sonido, que incluye combinar elementos de sonido como una suma de elementos de sonido constituyentes ajustados a nivel para crear un nuevo efecto de sonido deseado. Las herramientas de autoría del sistema de audio adaptativo permiten la creación de efectos de sonido como una colección de objetos de sonido con posiciones relativas utilizando una interfaz gráfica de usuario de diseño de sonido espacial-visual. Por ejemplo, se puede utilizar como plantilla una representación visual del objeto generador de sonido (por ejemplo, un automóvil) para ensamblar elementos de audio (detonación de tubo de escape, zumbido de neumático, ruido de motor) como canales de objetos que contienen el sonido y la posición espacial apropiados (en el tubo de escape, los neumáticos, el capó). A continuación, los canales de objetos individuales se pueden vincular y manipular como un grupo. La herramienta 106 de autoría incluye varios elementos de interfaz de usuario para permitir que el ingeniero de sonido introduzca información de control y vea parámetros de mezcla, y mejore la funcionalidad del sistema. El proceso de autoría y diseño de sonido también se mejora al permitir que los canales de los objetos y los canales de los altavoces se vinculen y manipulen como un grupo. Un ejemplo es la combinación de un canal de objeto con una fuente seca y discreta de sonido con un conjunto de canales de altavoz que contienen una señal de reverberación asociada.

La herramienta 106 de autoría de audio soporta la capacidad de combinar múltiples canales de audio, comúnmente denominada mezcla. Soporta múltiples métodos de mezcla, y puede incluir mezcla tradicional basada en niveles y mezcla basada en intensidad sonora. En la mezcla basada en niveles, se aplica una escala de banda ancha a los canales de audio, y, luego, se suman los canales de audio escalados. Los factores de escala de banda ancha para cada canal se eligen para controlar el nivel absoluto de la señal mixta resultante, y también los niveles relativos de los canales mixtos dentro de la señal mixta. En la mezcla basada en intensidad sonora, se modifican una o más señales de entrada usando escalado de amplitud dependiente de frecuencia, donde la amplitud dependiente de frecuencia se elige para proporcionar la intensidad sonora absoluta y relativa percibida deseada, que conserve el timbre percibido del sonido de entrada.

Las herramientas de autoría facilitan la capacidad de crear canales de altavoces y grupos de canales de altavoces. Esto permite asociar metadatos con cada grupo de canales de altavoces. Cada grupo de canales de altavoces se puede etiquetar de acuerdo con el tipo de contenido. El tipo de contenido se puede ampliar mediante una descripción de texto. Los tipos de contenido pueden incluir, pero sin limitarse a, diálogo, música y efectos. A cada grupo de canales de altavoces se le puede asignar instrucciones únicas sobre cómo realizar una mezcla ascendente de una configuración de canal a otra, donde la mezcla ascendente se define como la creación de M canales de audio a partir de N canales donde $M > N$. Las instrucciones de mezcla ascendente pueden incluir, pero sin limitarse a, los siguientes elementos: un indicador de habilitar/deshabilitar para indicar si se permite la mezcla ascendente; una matriz de mezcla ascendente para controlar el mapeo entre cada canal de entrada y salida; y la configuración

por defecto de habilitación y de matriz se puede asignar en base al tipo de contenido, como, por ejemplo, al habilitar la mezcla ascendente sólo para música. A cada grupo de canales de altavoces también se le pueden asignar instrucciones únicas sobre cómo realizar una mezcla descendente de una configuración de canal a otra, donde la mezcla descendente se define como la creación de Y canales de audio a partir de X canales donde $Y < X$. Las instrucciones de mezcla descendente pueden incluir, pero no están limitadas a, los siguientes elementos: una matriz para controlar el mapeo entre cada canal de entrada y salida; y la configuración de matriz por defecto se puede asignar en base al tipo de contenido, así, por ejemplo, el diálogo se mezclará de manera descendente en la pantalla; los efectos se mezclarán de manera descendente fuera de la pantalla. Cada canal de altavoz se puede también asociar con una bandera de metadatos para deshabilitar la gestión de graves durante la renderización.

Una característica permite la creación de canales de objetos y grupos de canales de objetos. Esta invención permite asociar metadatos con cada grupo de canales de objetos. Cada grupo de canales de objetos se puede etiquetar de acuerdo con el tipo de contenido. El tipo de contenido se puede ampliar mediante una descripción de texto, donde los tipos de contenido pueden incluir, pero sin limitarse a, diálogo, música y efectos. A cada grupo de canales de objetos se le pueden asignar metadatos para describir cómo se deben representar los objetos.

Se proporciona información de posición para indicar la posición de fuente aparente deseada. La posición puede indicarse utilizando un marco de referencia egocéntrico o alocéntrico. La referencia egocéntrica es apropiada cuando la posición de la fuente va a referirse al oyente. En el caso de la posición egocéntrica, son útiles las coordenadas esféricas para la descripción de la posición. Una referencia alocéntrica es el marco de referencia típico para cine u otras presentaciones audiovisuales en las que la posición de la fuente hace referencia a objetos en el entorno de presentación, tales como la pantalla de visualización o los límites de sala. Se proporciona información de la trayectoria tridimensional (3D) para permitir la interpolación de la posición o para el uso de otras decisiones de renderización, decisiones tales como la de habilitar un "ajuste a modo". Se proporciona información de tamaño para indicar el tamaño de fuente de audio percibido aparente deseado.

La cuantificación espacial se proporciona a través de un control de "ajuste al altavoz más cercano", que indica la intención del ingeniero de sonido o del mezclador de hacer que un objeto sea renderizado por exactamente un altavoz (con algún sacrificio potencial a la precisión espacial). Se puede indicar un límite a la distorsión espacial permitida a través de umbrales de tolerancia de elevación y acimut de tal manera que, si se excede el umbral, no se producirá la función "ajuste". Además de los umbrales de distancia, se puede indicar un parámetro de tasa de fundido cruzado para controlar la rapidez con la que un objeto en movimiento transitará o saltará de un altavoz a otro cuando la posición deseada cruce entre los altavoces.

Se utilizan metadatos espaciales dependientes para ciertos metadatos de posición. Por ejemplo, los metadatos se pueden generar automáticamente para un objeto "esclavo" asociándolo con un objeto "maestro" al que el objeto esclavo debe seguir. Se puede asignar un retardo de tiempo o una velocidad relativa al objeto esclavo. También se pueden proporcionar mecanismos para permitir la definición de un centro de gravedad acústico para conjuntos o grupos de objetos, tal que un objeto se pueda renderizar de tal manera que se perciba que se mueve alrededor de otro objeto. En tal caso, uno o más objetos pueden girar alrededor de un objeto o de un área definida, tal como un punto dominante o un área seca de la sala. El centro de gravedad acústico se usaría luego en la etapa de renderización para ayudar a determinar la información de ubicación para cada sonido apropiado basado en objetos, incluso aunque la información de ubicación final se expresara como una ubicación relativa a la sala, en lugar de cómo una ubicación relativa a otro objeto.

Cuando se renderiza un objeto, se asigna a uno o más altavoces de acuerdo con los metadatos de posición y con la ubicación de los altavoces de reproducción. Se pueden asociar metadatos adicionales con el objeto para limitar los altavoces que se utilizarán. El uso de restricciones puede prohibir el uso de los altavoces indicados o simplemente inhibir los altavoces indicados (permitir que entre menos energía en el altavoz o altavoces de la que en otro caso se aplicaría). Los conjuntos de altavoces que se restringirán pueden incluir, pero sin limitarse a, cualquiera de los altavoces nombrados o zonas de altavoces (por ejemplo, L, C, R, etc.) o áreas de altavoces, tales como: pared frontal, pared posterior, pared izquierda, pared derecha, techo, suelo, altavoces dentro de la sala, etc. Asimismo, en el curso de especificar la mezcla deseada de múltiples elementos de sonido, es posible hacer que uno o más elementos de sonido se vuelvan inaudibles o "enmascarados" debido a la presencia de otros elementos de sonido "enmascarados". Por ejemplo, cuando se detectan elementos enmascarados, el usuario podría identificarlos mediante un dispositivo gráfico de visualización.

Como se describe en otra parte, la descripción del programa de audio se puede adaptar para renderizar en una amplia variedad de instalaciones de altavoces y configuraciones de canales. Cuando se realiza la autoría de un programa de audio, es importante monitorizar el efecto de renderizar el programa en las configuraciones de reproducción anticipadas para verificar que se consiguen los resultados deseados. Esta invención incluye la capacidad de seleccionar configuraciones de reproducción de destino y de monitorizar el resultado. Además, el sistema puede monitorizar automáticamente los niveles de señal en el peor de los casos (es decir, los más altos) que se generarían en cada configuración de reproducción anticipada, y proporcionar una indicación si se produjera un recorte o limitación.

La figura 3 es un diagrama de bloques que ilustra el flujo de trabajo de creación, empaquetado y renderización de contenido de audio adaptativo. El flujo 300 de trabajo de la figura 3 se divide en tres grupos de tareas distintos denominados creación/autoría, empaquetado y exposición. En general, el modelo híbrido de lechos y objetos que se muestra en la figura 2 permite que la mayor parte del diseño, la edición, la mezcla previa y la mezcla final del sonido se realicen de la misma manera que en la actualidad y sin añadir una sobrecarga excesiva a los procesos actuales. La funcionalidad de audio adaptativo se proporciona en forma de software, soporte lógico inalterable (firmware) o circuitería que se utilizan junto con el equipo de procesamiento y producción de sonido, donde tal equipo pueden ser nuevos sistemas de hardware o actualizaciones de sistemas existentes. Por ejemplo, se pueden proporcionar aplicaciones posibles de enchufar para estaciones de trabajo de audio digital que permitan que las técnicas de panoramización existentes dentro del diseño y la edición de sonido permanezcan sin cambios. De esta manera, es posible colocar tanto lechos como objetos dentro de la estación de trabajo en salas de edición 5.1 o similares equipadas con sonido envolvente. El audio y los metadatos del objeto se graban en la sesión como preparación para las etapas de mezcla previa y final en la sala de doblaje.

Como se muestra en la figura 3, las tareas de creación o autoría implican la entrada de controles 302 de mezcla por parte de un usuario, por ejemplo, de un ingeniero de sonido en el siguiente ejemplo, para una mesa 304 de mezclas o estación de trabajo de audio. Los metadatos están integrados en la superficie de la mesa de mezclas, lo que permite los actuadores de fundido de las cintas, la panoramización y el procesamiento de audio para trabajar tanto con lechos o vástagos como con objetos de audio. Los metadatos se pueden editar usando la superficie de la mesa o la interfaz de usuario de la estación de trabajo, y el sonido se monitoriza usando una unidad 306 de renderización y masterización (RMU). Los datos de audio del lecho y del objeto y los metadatos asociados se graban durante la sesión de masterización para crear un 'maestro de impresión', que incluye una mezcla de audio adaptativo 310 y cualesquiera otros entregables renderizados 308 (como una mezcla de cine envolvente 7.1 o 5.1). Se pueden usar herramientas de autoría existentes (por ejemplo, estaciones de trabajo de audio digital tales como Pro Tools) para permitir a los ingenieros de sonido etiquetar pistas de audio individuales dentro de una sesión de mezcla. Este concepto se amplía al permitir a los usuarios etiquetar subsegmentos individuales dentro de una pista para ayudar a encontrar o identificar rápidamente elementos de audio. La interfaz de usuario de la mesa de mezclas que permite la definición y creación de los metadatos puede implantarse a través de elementos de interfaz gráfica de usuario, controles físicos (por ejemplo, controles deslizantes y perillas) o cualquier combinación de los mismos.

En la etapa de empaquetado, el archivo maestro de impresión se embala utilizando procedimientos de embalaje de MXF estándar de la industria, se pica y opcionalmente se encripta con el fin de garantizar la integridad del contenido de audio para su entrega a la instalación de empaquetado de cine digital. Este paso puede ser realizado por un procesador 312 de cine digital (DCP) o por cualquier procesador de audio apropiado dependiendo del entorno último de reproducción, como una sala 318 de cine equipada con sonido envolvente estándar, una sala 320 de cine capacitada con audio adaptativo o cualquier otro entorno de reproducción. Como se muestra en la figura 3, el procesador 312 emite las señales 314 y 316 de audio apropiadas dependiendo del entorno de exposición.

El maestro de impresión de audio adaptativo contiene una mezcla de audio adaptativo, junto con una mezcla estándar de código de pulso modulado (PCM) compatible con DCI. La mezcla PCM puede ser renderizada por la unidad de renderización y masterización en una sala de doblaje, o creada por un pase de mezcla separado si se desea. El audio de PCM forma el archivo de pista de audio principal estándar dentro del procesador 312 de cine digital, y el audio adaptativo forma un archivo de pista adicional. Un archivo tal de seguimiento puede ser compatible con los estándares industriales existentes, y es ignorado por los servidores compatibles con DCI que no pueden utilizarlo.

En un entorno de reproducción de cine de ejemplo, un servidor reconoce el DCP que contiene un archivo de pista de audio adaptativo como un paquete válido, y se asimila en el servidor y luego se retransmite a un procesador de cine de audio adaptativo. Un sistema que tiene tanto PCM lineal como archivos de audio adaptativos disponibles puede alternar entre ellos según sea necesario. Para su distribución al escenario de exposición, el esquema de empaquetado de audio adaptativo permite la entrega para que un solo tipo de paquete sea ofrecido en un cine. El paquete de DCP contiene tanto PCM como archivos de audio adaptativo. Se puede incorporar el uso de claves de seguridad, tal como en un mensaje de entrega de claves (KDM), para facilitar la entrega segura de contenido de películas u otro contenido similar.

Como se muestra en la figura 3, la metodología de audio adaptativo se realiza al facilitar al ingeniero de sonido expresar su intención con respecto a la renderización y reproducción de contenido de audio a través de la estación 304 de trabajo de audio. Al controlar ciertos controles de entrada, el ingeniero puede especificar dónde y cómo se reproducen los objetos de audio y los elementos de sonido dependiendo del entorno de escucha. Los metadatos se generan en la estación 304 de trabajo de audio en respuesta a las entradas 302 de mezcla del ingeniero para proporcionar colas de renderización que controlan los parámetros espaciales (por ejemplo, la posición, la velocidad, la intensidad, el timbre, etc.) y especifican qué altavoz, altavoces o grupos de altavoces en el entorno de escucha reproduce los sonidos respectivos durante la exposición. Los metadatos están asociados con los datos de audio respectivos en la estación 304 de trabajo o en la RMU 306 para su empaquetado y transporte por el DCP 312.

Una interfaz gráfica de usuario y herramientas de software que proporcionan el control de la estación 304 de trabajo por parte del ingeniero comprenden al menos parte de las herramientas 106 de autoría de la figura 1.

Códec de audio híbrido

5 Como se muestra en la figura 1, el sistema 100 incluye un códec 108 de audio híbrido. Este componente comprende un sistema de codificación, distribución y decodificación de audio que está configurado para generar un único torrente de bits que contiene elementos de audio convencionales basados en canal y elementos de codificación de objetos de audio. El sistema de codificación de audio híbrido se basa en un sistema de codificación basado en canales que está configurado para generar un torrente de bits único (unificado) que es simultáneamente compatible con (es decir, decodificable por) un primer decodificador configurado para decodificar datos de audio codificados de acuerdo con un primer protocolo de codificación (basado en canales) y uno o más decodificadores secundarios configurados para decodificar datos de audio codificados de acuerdo con uno o más protocolos de codificación secundarios (basados en objetos). El torrente de bits puede incluir datos codificados (en forma de ráfagas de datos) decodificables por el primer decodificador (e ignorados por cualquier decodificador secundario) y datos codificados (por ejemplo, otras ráfagas de datos) decodificables por uno o más decodificadores secundarios (e ignorados por el primer decodificador). El audio decodificado y la información asociada (metadatos) del primero y uno o más de los decodificadores secundarios se pueden combinar de manera tal que tanto la información basada en canal como la información basada en objeto se rendericen simultáneamente para recrear un facsímil del entorno, canales, información espacial y objetos presentados al sistema de codificación híbrido (es decir, dentro de un espacio 3D o entorno de escucha).

El códec 108 genera un torrente de bits que contiene información de audio codificada e información relacionada con múltiples conjuntos de posiciones de canal (altavoces). Un conjunto de posiciones de canal es fijo y se usa para el protocolo de codificación basado en canales, mientras que otro conjunto de posiciones de canal es adaptativo y se usa para el protocolo de codificación basado en objetos de audio, de tal modo que la configuración de canal para un objeto de audio puede cambiar en función del tiempo (dependiendo de dónde se coloque el objeto en el campo de sonido). De este modo, el sistema de codificación de audio híbrido puede llevar información sobre dos conjuntos de ubicaciones de altavoces para su reproducción, donde un conjunto puede ser fijo y ser un subconjunto del otro. Los dispositivos que soporten información de audio codificada heredada decodificarán y renderizarán la información de audio del subconjunto fijo, mientras que un dispositivo capaz de soportar el conjunto más grande podría decodificar y renderizar la información de audio codificada adicional que se asignaría de manera variable en el tiempo a diferentes altavoces del conjunto más grande. Lo que es más, el sistema no depende de que el primero y uno o más de los decodificadores secundarios estén presentes simultáneamente dentro de un sistema y/o dispositivo. Por consiguiente, un dispositivo/sistema heredado y/o existente que contenga sólo un decodificador que soporte el primer protocolo producirá un campo de sonido totalmente compatible que se renderizará mediante sistemas de reproducción tradicionales basados en canal. En este caso, las partes desconocidas o no soportadas del protocolo de torrente de bits híbrido (es decir, la información de audio representada por un protocolo de codificación secundario) serían ignoradas por el decodificador del sistema o dispositivo que soporta el primer protocolo de codificación híbrido.

Alternativamente, el códec 108 está configurado para funcionar en un modo en el que el primer subsistema de codificación (que soporta el primer protocolo) contiene una representación combinada de toda la información del campo de sonido (canales y objetos) representada tanto en el primero como en uno o más de los subsistemas del codificador secundario presentes dentro del codificador híbrido. Esto asegura que el torrente de bits híbrido incluya retrocompatibilidad con decodificadores que soportan sólo el protocolo del primer subsistema de codificador, al permitir que los objetos de audio (que van normalmente en uno o más protocolos de codificador secundarios) se representen y rendericen dentro de decodificadores que soportan sólo el primer protocolo.

50 Alternativamente, el códec 108 incluye dos o más subsistemas de codificación, donde cada uno de estos subsistemas está configurado para codificar datos de audio de acuerdo con un protocolo diferente, y está configurado para combinar las salidas de los subsistemas para generar un torrente de bits de formato híbrido (unificado).

55 Uno de los beneficios es la capacidad de llevar un torrente de bits de audio codificado híbrido a través de una amplia gama de sistemas de distribución de contenido, donde cada uno de los sistemas de distribución soporta convencionalmente sólo datos codificados de acuerdo con el primer protocolo de codificación. Esto elimina la necesidad, para cualquier sistema y/o transporte, de modificaciones/cambios de protocolo de nivel con el fin de soportar específicamente el sistema de codificación híbrido.

60 Los sistemas de codificación de audio utilizan típicamente elementos de torrente de bits estandarizados para permitir el transporte de datos adicionales (arbitrarios) dentro del propio torrente de bits. Estos datos adicionales (arbitrarios) se omiten (es decir, se ignoran) generalmente durante la decodificación del audio codificado incluido en el torrente de bits, pero pueden usarse para un fin diferente a la decodificación. Los diferentes estándares de codificación de audio expresan estos campos de datos adicionales utilizando una nomenclatura única. Los elementos de torrente de bits de este tipo general pueden incluir, pero sin limitarse a, datos auxiliares, campos de omisión, elementos de

torrente de datos, elementos de relleno, datos accesorios y elementos de subtorrente. A menos que se indique lo contrario, el uso de la expresión “datos auxiliares” en el presente documento no implica un tipo o formato específico de datos adicionales, sino que debe interpretarse como una expresión genérica que abarca cualquiera o todos los ejemplos asociados con la presente invención.

5 Un canal de datos habilitado mediante elementos de torrente de bits “auxiliares” de un primer protocolo de codificación dentro de un torrente de bits de sistema de codificación híbrido combinado podría llevar uno o más
10 torrentes secundarios (independientes o dependientes) de bits de audio (codificados de acuerdo con uno o más protocolos de codificación secundarios). El uno o más torrentes de bits de audio secundarios podría/n dividirse en bloques de N muestras y multiplexarse en los campos de “datos auxiliares” de un primer torrente de bits. El primer
15 torrente de bits se puede decodificar mediante un decodificador (complemento) apropiado. Además, los datos auxiliares del primer torrente de bits podrían extraerse, recombinados en uno o más torrentes de bits de audio secundarios, ser decodificados mediante un procesador que soporte la sintaxis de uno o más de los torrentes de bits secundarios y luego combinarse y renderizarse juntos o independientemente. Lo que es más, también es posible
20 invertir los papeles de los torrentes de bits primero y segundo, de modo que los bloques de datos de un primer torrente de bits se multiplexen en los datos auxiliares de un segundo torrente de bits.

25 Los elementos de torrente de bits asociados con un protocolo de codificación secundario también llevan y transfieren características de información (metadatos) del audio subyacente, que pueden incluir, pero no se limitan a, la posición, velocidad y tamaño deseados de la fuente de sonido. Estos metadatos se utilizan durante los procesos de decodificación y renderización para recrear la posición adecuada (es decir, original) para el objeto de audio asociado que va dentro del torrente de bits aplicable. También es posible llevar los metadatos descritos anteriormente, que son aplicables a los objetos de audio contenidos en uno o más torrentes de bits secundarios presentes en el torrente híbrido, dentro de elementos de torrente de bits asociados con el primer protocolo de codificación.

30 Los elementos de torrente de bits asociados con uno o ambos protocolos primero y segundo de codificación del sistema de codificación híbrido llevan/transfieren metadatos contextuales que identifican parámetros espaciales (es decir, la esencia de las propiedades de la señal en sí) y más información adicional que describe el tipo de esencia de audio subyacente en forma de clases de audio específicas que se llevan dentro del torrente de bits de audio
35 codificado híbrido. Tales metadatos podrían indicar, por ejemplo, la presencia de diálogo hablado, música, diálogo sobre música, aplausos, voz cantada, etc., y podrían utilizarse para modificar de manera adaptativa el comportamiento de los módulos de procesamiento previo o posterior interconectados aguas arriba o aguas abajo del sistema híbrido de codificación.

40 El códec 108 está configurado para funcionar con un fondo de bits compartido o común en el que los bits disponibles para codificación se “comparten” entre todos o parte de los subsistemas de codificación que soportan uno o más protocolos. Tal códec puede distribuir los bits disponibles (a partir del fondo común de bits “compartidos”) entre los subsistemas de codificación con el fin de optimizar la calidad de audio general del torrente de bits unificado. Por ejemplo, durante un primer intervalo de tiempo, el códec puede asignar más bits disponibles a un primer subsistema
45 de codificación y menos bits disponibles a los subsistemas restantes, mientras que durante un segundo intervalo de tiempo, el códec puede asignar una cantidad menor de los bits disponibles al primer subsistema de codificación y una cantidad mayor de los bits disponibles a los subsistemas restantes. La decisión de cómo asignar bits entre subsistemas de codificación puede depender, por ejemplo, de los resultados del análisis estadístico del fondo de bits compartidos y/o del análisis del contenido de audio codificado por cada subsistema. El códec puede asignar bits del fondo compartido de tal manera que un torrente de bits unificado construido multiplexando las salidas de los subsistemas de codificación mantenga una tasa de bits/longitud de trama constante durante un intervalo de tiempo específico. También es posible, en algunos casos, que la tasa de bits/longitud de trama del torrente de bits unificado
50 varíe durante un intervalo de tiempo específico.

55 Alternativamente, el códec 108 genera un torrente de bits unificado que incluye datos codificados de acuerdo con el primer protocolo de codificación configurado y transmitido como un subtorrente independiente de un torrente de datos codificados (que un decodificador que soporte el primer protocolo de codificación decodificará) y datos codificados de acuerdo con un segundo protocolo enviado como un subtorrente independiente o dependiente del torrente de datos codificados (que un decodificador que soporte el primer protocolo ignorará). De manera más general, el códec genera un torrente de bits unificado que incluye dos o más subtorrentes independientes o dependientes (donde cada subtorrente incluye datos codificados de acuerdo con un protocolo de codificación diferente o idéntico).

60 Alternativamente, el códec 108 genera un torrente de bits unificado que incluye datos codificados de acuerdo con el primer protocolo de codificación configurado y transmitido con un identificador de torrente de bits único (que un decodificador que soporte un primer protocolo de codificación asociado con el identificador de torrente de bits único decodificará), y datos codificados de acuerdo con un segundo protocolo configurado y transmitido con un identificador de torrente de bits único, que un decodificador que soporte el primer protocolo ignorará. De manera más general, el códec genera un torrente de bits unificado que incluye dos o más subtorrentes (donde cada subtorrente incluye datos codificados de acuerdo con un protocolo de codificación diferente o idéntico y donde cada uno lleva un identificador de torrente de bits único). Los métodos y sistemas para crear un torrente de bits unificado
65

descritos anteriormente brindan la capacidad de señalar inequívocamente (a un decodificador) qué entrelazado y/o protocolo se ha utilizado dentro de un torrente de bits híbrido (por ejemplo, para señalar si es el enfoque descrito de datos AUX, de SKIP, de DSE o de subtorrente el utilizado).

- 5 El sistema híbrido de codificación está configurado para soportar el desintercalado/demultiplexado y el reintercalado/remultiplexado de torrentes de bits que soportan uno o más protocolos secundarios en un primer torrente de bits (que soporta un primer protocolo) en cualquier punto de procesamiento que se encuentre en un sistema de entrega de medios. El códec híbrido está también configurado para ser capaz de codificar torrentes de entrada de audio con diferentes tasas de muestreo en un torrente de bits. Esto proporciona un medio para codificar y
- 10 distribuir de manera eficiente fuentes de audio que contienen señales con anchos de banda inherentemente diferentes. Por ejemplo, las pistas de diálogo tienen típicamente un ancho de banda inherentemente menor que las pistas de música y efectos.

Renderización

- 15 El sistema de audio adaptativo permite empaquetar múltiples pistas (por ejemplo, hasta 128), habitualmente como una combinación de lechos y objetos. El formato básico de los datos de audio para el sistema de audio adaptativo comprende varios torrentes monofónicos de audio independientes. Cada torrente tiene asociados metadatos que especifican si el torrente es un torrente basado en canales o un torrente basado en objetos. Los torrentes basados
- 20 en canal tienen información de renderización codificada por medio del nombre o de la etiqueta del canal; y los torrentes basados en objeto tienen información de ubicación codificada a través de expresiones matemáticas codificadas en metadatos asociados adicionales. Los torrentes de audio independientes originales se empaquetan luego como un único torrente de bits en serie que contiene todos los datos de audio de manera ordenada. Esta configuración de datos adaptativos permite que el sonido se renderice de acuerdo con un marco de referencia
- 25 alocéntrico, en el que la ubicación última de renderización de un sonido se basa en el entorno de reproducción para que se corresponda con la intención del mezclador. De este modo, se puede especificar que un sonido se origine en un marco de referencia de la sala de reproducción (por ejemplo, en el medio de la pared izquierda), en lugar de especificarse en un altavoz o grupo de altavoces etiquetado (por ejemplo, envolvente izquierdo). Los metadatos de posición del objeto contienen el marco alocéntrico apropiado de la información de referencia necesaria para
- 30 reproducir el sonido correctamente utilizando las posiciones de los altavoces disponibles en una sala que está configurada para reproducir el contenido de audio adaptativo.

- El renderizador toma el torrente de bits que codifica las pistas de audio y procesa el contenido de acuerdo con el tipo de señal. Los lechos son alimentados para matrices, que potencialmente requerirán diferentes retardos y
- 35 procesamiento de ecualización que los objetos individuales. El proceso soporta la renderización de estos lechos y objetos en múltiples (hasta 64) salidas de altavoz. La figura 4 es un diagrama de bloques de una etapa de renderización de un sistema de audio adaptativo, según una realización. Como se muestra en el sistema 400 de la figura 4, ciertos componentes de las etapas de creación, autoría y empaquetado del sistema 300, tales como RMU
 - 40 306 y el procesador 312, proporcionan una serie de señales de entrada, tal como hasta 128 pistas de audio, que comprenden las señales 402 de audio adaptativo. Estas señales comprenden los objetos y lechos basados en canal que son utilizados por el renderizador 404. Los objetos y el audio (los lechos) basado en canales se introducen en un administrador 406 de nivel que proporciona control sobre los niveles de salida o amplitudes de los diferentes
 - 45 componentes de audio. Ciertos componentes de audio pueden ser procesados por un componente 408 de corrección de matriz. Las señales de audio adaptativo pasan luego a través de un componente 410 de procesamiento de cadena B, que genera un número (por ejemplo, hasta 64) de señales de salida de alimentación de altavoz. En general, la alimentación de la cadena B se refiere a las señales procesadas por los amplificadores de potencia, los cruces y los altavoces, en contraste con el contenido de la cadena A, que constituye la banda sonora de la película.

- 50 El renderizador 404 ejecuta un algoritmo de renderización que usa inteligentemente los altavoces de sonido envolvente en el cine lo mejor que puede. Al mejorar el manejo de potencia y la respuesta de frecuencia de los altavoces envolventes, y mantener el mismo nivel de referencia de monitorización para cada altavoz o canal de salida en el cine, los objetos que se panoramizan entre la pantalla y los altavoces envolventes pueden mantener su nivel de presión sonora y tener una coincidencia de timbre más cercana sin, lo que es más importante, aumentar el
- 55 nivel general de presión sonora en la sala de cine. Una matriz de altavoces envolventes especificados de forma adecuada tendrá típicamente suficiente margen dinámico para reproducir el intervalo dinámico máximo disponible dentro de una banda sonora envolvente 7.1 o 5.1 (es decir, 20 dB por encima del nivel de referencia), sin embargo, es poco probable que un único altavoz envolvente tenga el mismo margen dinámico que tiene un gran altavoz de pantalla multimodo. Como resultado, es probable que haya casos en los que un objeto colocado en el campo
- 60 envolvente requiera una presión de sonido mayor que la que se puede obtener con un solo altavoz envolvente. En estos casos, el renderizador dispersará el sonido a través de un número apropiado de altavoces para lograr el nivel de presión sonora requerido. El sistema de audio adaptativo mejora la calidad y el manejo de la potencia de los altavoces envolventes para proporcionar una mejora en la fidelidad de la renderización. Proporciona soporte para la gestión de graves de los altavoces de sonido envolvente mediante el uso de subwoofers traseros opcionales que
- 65 permiten que para cada altavoz de sonido envolvente se consiga un mejor manejo de potencia y, al mismo tiempo, se utilicen potencialmente cajas acústicas de altavoz más pequeñas. También permite la adición de altavoces

envolventes laterales más cerca de la pantalla que la práctica actual para garantizar que los objetos puedan pasar sin problemas de la pantalla al sonido envolvente.

Mediante el uso de metadatos para especificar información de ubicación de objetos de audio junto con ciertos procesos de renderización, el sistema 400 proporciona un método comprehensivo y flexible para que los creadores de contenido se muevan más allá de las restricciones de los sistemas existentes. Como se indicó anteriormente, los sistemas actuales crean y distribuyen audio que se fija a ubicaciones particulares de los altavoces con un conocimiento limitado del tipo de contenido transferido en la esencia del audio (la parte del audio que se reproduce). El sistema 100 de audio adaptativo proporciona un nuevo enfoque híbrido que incluye opciones para elementos tanto del audio específico de la ubicación del altavoz (canal izquierdo, canal derecho, etc.) como del audio orientado a objeto, que tengan información espacial generalizada que puede incluir, pero sin limitarse a, posición, tamaño y velocidad. Este enfoque híbrido proporciona un enfoque equilibrado de fidelidad (proporcionada por ubicaciones fijas de altavoces) y flexibilidad en la renderización (objetos de audio generalizados). El sistema también proporciona información adicional útil sobre el contenido de audio que el creador del contenido emparejó con la esencia del audio en el momento de la creación del contenido. Esta información proporciona información potente y detallada sobre los atributos del audio que se puede usar de maneras muy fructíferas durante la renderización. Tales atributos pueden incluir, pero sin limitarse a, el tipo de contenido (diálogo, música, efecto, Foley, trasfondo/ambiente, etc.), atributos espaciales (posición 3D, tamaño 3D, velocidad) e información de renderización (ajuste a la ubicación del altavoz, pesos de canal, ganancia, información de gestión de graves, etc.).

El sistema de audio adaptativo descrito en el presente documento proporciona una información potente que puede ser utilizada para la renderización por un número muy variable de puntos finales. En muchos casos, la técnica óptima de renderización aplicada depende en gran medida del dispositivo de punto final. Por ejemplo, los sistemas de cine en casa y las barras de sonido pueden tener 2, 3, 5, 7 o incluso 9 altavoces separados. Muchos otros tipos de sistemas, tales como televisores, ordenadores y bases de música tienen sólo dos altavoces, y casi todos los dispositivos de uso común tienen una salida de auriculares binaural (PC, ordenador portátil, tableta, teléfono celular, reproductor de música, etc.). Sin embargo, para el audio tradicional que se distribuye en la actualidad (mono, estéreo, canales 5.1, 7.1), los dispositivos de punto final necesitan a menudo tomar decisiones y compromisos simplistas para renderizar y reproducir audio que ahora se distribuye en una forma específica de canal/altavoz. Además, se transfiere poca o ninguna información sobre el contenido real (diálogo, música, ambiente, etc.) que se está distribuyendo y poca o ninguna información sobre la intención del creador del contenido para la reproducción de audio. Sin embargo, el sistema 100 de audio adaptativo proporciona esta información, y proporciona, potencialmente, acceso a los objetos de audio, lo que puede emplearse para crear una experiencia de usuario de próxima generación convincente.

El sistema 100 permite al creador de contenido incrustar la intención espacial de la mezcla dentro del torrente de bits utilizando metadatos tales como posición, tamaño, velocidad, etc., a través de un formato de transmisión de audio adaptativo y metadatos únicos y potentes. Esto permite una gran flexibilidad en la reproducción espacial de audio. Desde el punto de vista de la renderización espacial, el audio adaptativo permite la adaptación de la mezcla a la posición exacta de los altavoces en una sala en particular para evitar la distorsión espacial que se produce cuando la geometría del sistema de reproducción no es idéntica a la del sistema de autoría. En los sistemas de reproducción de audio actuales en los que sólo se envía audio para un canal de altavoz, se desconoce la intención del creador del contenido. El sistema 100 utiliza metadatos transferidos a lo largo del proceso de creación y distribución. Un sistema de reproducción con reconocimiento de audio adaptativo puede usar esta información de metadatos para reproducir el contenido de una manera que coincida con la intención original del creador del contenido. Asimismo, la mezcla se puede adaptar a la configuración del hardware exacta del sistema de reproducción. En la actualidad, existen muchas configuraciones y tipos de altavoces posibles diferentes en equipos de renderización tales como televisores, sistemas de cine en casa, barras de sonido, bases para reproductores portátiles de música, etc. Cuando a estos sistemas se les envía información de audio específica de canal en la actualidad (es decir, audio de canal izquierdo y derecho o audio multicanal), el sistema debe procesar el audio para que coincida de manera apropiada con las capacidades del equipo de renderización. Un ejemplo es el envío de audio estéreo estándar a una barra de sonido con más de dos altavoces. En la reproducción de audio actual, en la que sólo se envía audio para un canal de altavoz, se desconoce la intención del creador del contenido. Mediante el uso de metadatos transferidos a lo largo de la canalización de creación y distribución, un sistema de reproducción de audio adaptativo puede usar esta información para reproducir el contenido de una manera que coincida con la intención original del creador del contenido. Por ejemplo, algunas barras de sonido tienen altavoces laterales para crear una sensación de envolvimiento. Con audio adaptativo, la barra de sonido puede usar la información espacial y el tipo de contenido (tal como efectos ambientales) para enviar sólo el audio apropiado a estos altavoces laterales.

El sistema de audio adaptativo permite la interpolación ilimitada de altavoces en un sistema en todas las dimensiones frontal/posterior, izquierda/derecha, arriba/abajo, cerca/lejos. En los sistemas de reproducción de audio actuales, no existe información sobre cómo manejar el audio donde se puede desear posicionar el audio de tal manera que el oyente lo perciba entre dos altavoces. En la actualidad, con el audio que se asigna sólo a un altavoz específico, se introduce un factor de cuantificación espacial. Con audio adaptativo, el posicionamiento espacial del audio puede conocerse con precisión y reproducirse en consecuencia en el sistema de reproducción de audio.

Con respecto a la renderización de auriculares, la intención del creador se realiza haciendo coincidir las funciones de transferencia relacionadas con la cabeza (HRTF) con la posición espacial. Cuando el audio se reproduce a través de auriculares, se puede conseguir la virtualización espacial mediante la aplicación de una función de transferencia relacionada con la cabeza, que procesa el audio, añadiendo apuntes perceptuales que crean la percepción de que el audio se reproduce en el espacio 3D y no a través de auriculares. La precisión de la reproducción espacial depende de la selección de la HRTF apropiada, que puede variar en base a varios factores, incluida la posición espacial. El uso de la información espacial proporcionada por el sistema de audio adaptativo puede dar como resultado la selección de una HRTF o de un número variable continuo de estas funciones para mejorar en gran medida la experiencia de reproducción.

La información espacial transferida por el sistema de audio adaptativo no sólo puede ser utilizada por un creador de contenido para crear una experiencia de entretenimiento convincente (películas, televisión, música, etc.), sino que la información espacial también puede indicar dónde se posiciona el oyente con relación a objetos físicos tales como edificios o puntos geográficos de interés. Esto permitiría al usuario interactuar con una experiencia de audio virtualizada que esté relacionada con el mundo real, es decir, de realidad aumentada.

Las implantaciones permiten también la mezcla ascendente espacial, realizando una mezcla ascendente perfeccionada leyendo los metadatos sólo si los datos de audio de los objetos no están disponibles. Conocer la posición de todos los objetos y sus tipos permite al mezclador ascendente diferenciar mejor los elementos dentro de las pistas basadas en canal. Los algoritmos de mezcla ascendente existentes tienen que inferir información tal como el tipo de contenido de audio (voz, música, efectos ambientales), así como la posición de diferentes elementos dentro del torrente de audio, para crear una mezcla ascendente de alta calidad con artefactos audibles mínimos o nulos. Muchas veces la información inferida puede ser incorrecta o inapropiada. Con el audio adaptativo, la información adicional disponible de los metadatos relacionada, por ejemplo, con el tipo de contenido de audio, la posición espacial, la velocidad, el tamaño del objeto de audio, etc., puede ser utilizada por un algoritmo de mezcla ascendente para crear un resultado de reproducción de alta calidad. El sistema también hace coincidir espacialmente el audio con el vídeo al posicionar con precisión el objeto de audio de la pantalla para los elementos visuales. En este caso, es posible una experiencia de reproducción de audio/vídeo convincente, particularmente con tamaños de pantalla más grandes, si la ubicación espacial reproducida de algunos elementos de audio coincide con los elementos de la imagen en la pantalla. Un ejemplo es que el diálogo en una película o programa de televisión coincida espacialmente con una persona o personaje que está hablando en la pantalla. Con el audio basado en canales del altavoz normal, no existe un método fácil para determinar dónde debe posicionarse espacialmente el diálogo para que coincida con la ubicación de la persona o personaje en la pantalla. Con la información de audio disponible con audio adaptativo, se puede conseguir tal alineamiento audio/visual. El alineamiento visual posicional y espacial de audio también se puede usar para objetos que no sean personajes/diálogos, tales como automóviles, camiones, animación, etc.

El sistema 100 facilita un procesamiento de enmascaramiento espacial, ya que el conocimiento de la intención espacial de una mezcla a través de los metadatos de audio adaptativos significa que la mezcla se puede adaptar a cualquier configuración de altavoz. Sin embargo, se corre el riesgo de realizar una mezcla descendente de objetos en la misma o casi en la misma ubicación debido a las limitaciones del sistema de reproducción. Por ejemplo, un objeto destinado a ser panoramizado en la parte trasera izquierda podría mezclarse de manera descendente en el frente izquierdo si los canales envolventes no están presentes, pero si se produce un elemento más ruidoso en el frente izquierdo al mismo tiempo, el objeto mezclado de manera descendente se enmascarará y desaparecerá de la mezcla. Usando metadatos de audio adaptativos, el renderizador puede anticipar el enmascaramiento espacial, y pueden ajustarse los parámetros espaciales y/o de mezcla descendente de intensidad sonora de cada objeto para que todos los elementos de audio de la mezcla permanezcan tan perceptibles como en la mezcla original. Debido a que el renderizador entiende la relación espacial entre la mezcla y el sistema de reproducción, tiene la capacidad de "ajustar" objetos a los altavoces más cercanos en lugar de crear una imagen fantasma entre dos o más altavoces. Si bien esto puede distorsionar ligeramente la representación espacial de la mezcla, también permite que el renderizador evite una imagen fantasma no intencionada. Por ejemplo, si la posición angular del altavoz izquierdo de la etapa de mezcla no se corresponde con la posición angular del altavoz izquierdo del sistema de reproducción, el uso de la función de ajuste al altavoz más cercano podría evitar que el sistema de reproducción reprodujera una imagen fantasma constante del canal izquierdo de la etapa de mezcla.

Con respecto al procesamiento de contenido, el sistema 100 de audio adaptativo permite al creador de contenido crear individuales objetos de audio y añadir información sobre el contenido que se puede transferir al sistema de reproducción. Esto permite una gran flexibilidad en el procesamiento del audio anterior a la reproducción. Desde el punto de vista del procesamiento y de la renderización de contenido, el sistema de audio adaptativo permite adaptar el procesamiento al tipo de objeto. Por ejemplo, el perfeccionamiento del diálogo se puede aplicar sólo a los objetos del diálogo. El perfeccionamiento del diálogo se refiere a un método de procesamiento de audio que contiene un diálogo tal que la audibilidad y/o inteligibilidad del diálogo aumente y/o mejore. En muchos casos, el procesamiento de audio que se aplica al diálogo es inapropiado para contenido de audio que no sea de diálogo (es decir, música, efectos ambientales, etc.) y puede dar como resultado artefactos audibles objetables. Con el audio adaptativo, un objeto de audio podría contener sólo el diálogo en un fragmento de contenido, y se puede etiquetar en consecuencia, para que una solución de renderización pudiera aplicar selectivamente el perfeccionamiento del

diálogo sólo al contenido del diálogo. Además, si el objeto de audio es sólo diálogo (y no una mezcla de diálogo y otro contenido, como suele ser el caso), entonces el procesamiento de perfeccionamiento del diálogo puede procesar el diálogo exclusivamente (limitando por ello cualquier procesamiento que se realice en cualquier otro contenido). Asimismo, la gestión de graves (filtrado, atenuación, ganancia) puede dirigirse a objetos específicos en base a su tipo. La gestión de graves se refiere a aislar y procesar selectivamente sólo las frecuencias graves (o más bajas) de un fragmento de contenido en particular. Con los sistemas de audio y los mecanismos de entrega actuales, éste es un proceso “ciego” que se aplica a todo el audio. Con el audio adaptativo, los metadatos pueden identificar objetos de audio específicos para los que es apropiada la gestión de graves, y el procesamiento de renderización se puede aplicar de forma adecuada.

El sistema 100 de audio adaptativo también proporciona compresión de intervalo dinámico basada en objeto y mezcla ascendente selectiva. Las pistas de audio tradicionales tienen la misma duración que el contenido en sí, mientras que un objeto de audio puede aparecer durante un tiempo limitado en el contenido. Los metadatos asociados con un objeto pueden contener información sobre su amplitud de señal promedio y pico, así como su tiempo de comienzo o ataque (particularmente para material transitorio). Esta información permitiría a un compresor adaptar mejor su compresión y constantes de tiempo (ataque, liberación, etc.) para adaptarse mejor al contenido. Para la mezcla ascendente selectiva, los creadores de contenido pueden optar por indicar en el torrente de bits de audio adaptativo si un objeto debe mezclarse o no. Esta información permite al renderizador de audio adaptativo y al mezclador superior distinguir qué elementos de audio se pueden mezclar de forma segura, respetando la intención del creador.

Las implantaciones también permiten que el sistema de audio adaptativo seleccione un algoritmo preferido de renderización entre varios algoritmos de renderización disponibles y/o formatos de sonido envolvente. Los ejemplos de algoritmos de renderización disponibles incluyen: binaural, dipolo estéreo, ambisónica, síntesis de campo de ondas (WFS), panoramización multicanal, vástagos sin procesar con metadatos de posición. Otros incluyen equilibrio dual y panoramización de amplitud basada en vectores.

El formato de distribución binaural usa una representación de dos canales de un campo de sonido en términos de la señal presente en los oídos izquierdo y derecho. La información binaural se puede crear mediante grabación en el oído o modelos sintetizados utilizando la HRTF. La reproducción de una representación binaural se hace típicamente a través de auriculares o mediante el empleo de anulación de intercomunicación. La reproducción en una configuración de altavoz arbitraria requeriría un análisis de la señal para determinar el campo de sonido asociado y/o la/s fuente/s de señal.

El método de renderización de dipolo estéreo es un proceso de anulación de diafonía transaural para hacer que las señales binaurales puedan reproducirse a través de altavoces estéreo (por ejemplo, a + y - 10 grados fuera del centro).

La ambisónica es un (formato de distribución y un método de representación) que está codificado con una forma de cuatro canales denominada formato B. El primer canal, W, es la señal de presión no direccional; el segundo canal, X, es el gradiente de presión direccional que contiene la información frontal y posterior; el tercer canal, Y, contiene la izquierda y la derecha, y el Z la de arriba y abajo. Estos canales definen una muestra de primer orden del campo sonoro completo en un punto. La ambisónica utiliza todos los altavoces disponibles para recrear el campo de sonido muestreado (o sintetizado) dentro de la matriz de altavoces tal que cuando algunos altavoces empujan, otros tiran.

La síntesis de campo de ondas es un método de renderización de reproducción de sonido, basado en la construcción precisa del campo de ondas deseado por fuentes secundarias. La WFS se basa en el principio de Huygens y se implanta como matrices de altavoces (decenas o cientos) que suenan en el espacio de escucha y funcionan de manera coordinada y en fases para recrear cada onda de sonido individual.

La panoramización multicanal es un formato de distribución y/o un método de renderización, y puede denominarse audio basado en canales. En este caso, el sonido se representa como una serie de fuentes discretas que se reproducirán a través de un número igual de altavoces en ángulos definidos del oyente. El creador/mezclador de contenido puede crear imágenes virtuales panoramizando señales entre canales adyacentes para proporcionar señales de dirección; las reflexiones adelantadas, la reverberación, etc., se pueden mezclar en muchos canales para proporcionar apuntes de dirección y ambientales.

Los vástagos sin procesar con metadatos de posición son un formato de distribución, y también pueden denominarse audio basado en objetos. En este formato, se representan distintas fuentes de sonido “con micrófono cercano” junto con metadatos de posición y ambientales. Las fuentes virtuales se renderizan en base a los metadatos, el equipo de reproducción y el entorno de escucha.

El formato de audio adaptativo es un híbrido del formato de panoramización multicanal y el formato de vástagos sin procesar. El método de renderización es la panorámica multicanal. Para los canales de audio, la renderización (panoramización) se produce en el momento de la creación, mientras que para los objetos la renderización (panoramización) se produce durante la reproducción.

Formato de transmisión de audio adaptativo y metadatos

Como se indicó anteriormente, los metadatos se generan durante la etapa de creación para codificar cierta información posicional para los objetos de audio y para acompañar a un programa de audio que ayuda a renderizar el programa de audio, y, en particular, para describir el programa de audio de una manera que permita renderizar el programa de audio en una amplia variedad de equipos de reproducción y entornos de reproducción. Los metadatos se generan para un programa dado y para los editores y mezcladores que crean, recopilan, editan y manipulan el audio durante la postproducción. Una característica importante del formato de audio adaptativo es la capacidad de controlar cómo se traducirá el audio a los sistemas y entornos de reproducción que difieren del entorno de mezcla. En particular, una sala de cine dada puede tener menores capacidades que el entorno mixto.

El renderizador de audio adaptativo está diseñado para hacer el mejor uso del equipo disponible para recrear la intención del mezclador. Además, las herramientas de autoría de audio adaptativo permiten que el mezclador obtenga una vista previa y ajuste cómo se renderizará la mezcla en una variedad de configuraciones de reproducción. Todos los valores de metadatos se pueden condicionar al entorno de reproducción y a la configuración de los altavoces. Por ejemplo, se puede especificar un nivel de mezcla diferente para un elemento de audio dado en base a la configuración o al modo de reproducción. La lista de modos de reproducción condicionados es ampliable e incluye lo siguiente: (1) reproducción basada únicamente en canales: 5.1, 7.1, 7.1 (altura), 9.1; y (2) reproducción de altavoces discretos: 3D, 2D (sin altura).

Los metadatos controlan o dictan diferentes aspectos del contenido de audio adaptativo y están organizados en base a diferentes tipos que incluyen: metadatos de programa, metadatos de audio y metadatos de renderización (para canal y para objeto). Cada tipo de metadatos incluye uno o más elementos de metadatos que proporcionan valores para las características a las que hace referencia un identificador (ID). La figura 5 es una tabla que enumera los tipos de metadatos y los elementos de metadatos asociados para el sistema de audio adaptativo.

Como se muestra en la tabla 500 de la figura 5, el primer tipo de metadatos consiste en los metadatos del programa, e incluye elementos de metadatos que especifican la tasa de trama, el cómputo de pistas, la descripción del canal ampliable y la descripción de la etapa de mezcla. El elemento de metadatos de tasa de trama especifica la tasa de las tramas de contenido de audio en unidades de tramas por segundo (fps). El formato de audio sin procesar no necesita incluir el encuadre del audio o los metadatos, ya que el audio se proporciona como pistas completas (duración de un carrete o función completa) en lugar de segmentos de audio (duración de un objeto). El formato sin procesar tiene que llevar toda la información necesaria para permitir que el codificador de audio adaptativo trame el audio y los metadatos, incluyendo la tasa de trama real. La tabla 1 muestra el ID, los valores de ejemplo y la descripción del elemento de metadatos de tasa de trama.

Tabla 1

ID	Valores	Descripción 2
Tasa de trama	24, 25, 30, 48, 50, 60, 96, 100, 120, ampliable (tramas/segundo)	Indicación de la tasa de trama prevista para todo el programa. El campo proporcionará una codificación eficiente de tasas comunes, así como la capacidad de ampliarse a un campo de punto flotante ampliable con una resolución de 0,01.

El elemento de metadatos de cómputo de pistas indica el número de pistas de audio en una trama. Un decodificador/procesador de audio adaptativo de ejemplo puede soportar hasta 128 pistas de audio simultáneas, aunque que el formato de audio adaptativo soportará cualquier número de pistas de audio. La tabla 2 muestra el ID, los valores de ejemplo y la descripción del elemento de metadatos del cómputo de pistas.

Tabla 2

ID	Valores	Descripción 2
númPistas	Entero positivo, intervalo ampliable.	Indicación del número de pistas de audio en la trama.

El audio basado en canales se puede asignar a canales no estándar y el elemento de metadatos de descripción de canal ampliable permite que las mezclas utilicen nuevas posiciones de canal. Para cada canal de ampliación, se proporcionarán los siguientes metadatos de la manera que se muestra en la tabla 3:

Tabla 3

ID	Valores	Descripción 2
PosiciónCanAmp	Coordenadas x, y, z.	Posición
AnchoCanAmp	Coordenadas x, y, z.	Ancho

El elemento de metadatos de descripción de la etapa de mezcla especifica la frecuencia a la que un altavoz en particular produce la mitad de la potencia de la banda de paso. La tabla 4 muestra el ID, los valores de ejemplo y la descripción del elemento de metadatos de descripción de etapa de mezcla, donde LF = Baja frecuencia; HF = alta frecuencia; punto 3dB = borde de la banda de paso del altavoz.

Tabla 4

ID	Valores	Descripción
númAltavozMezcla	Entero positivo	
PosAltavozMezcla	Coordenadas x, y, z para cada altavoz	
TipoAltavozMezcla	{FR, LLF, Sub}, para cada altavoz	Intervalo completo, respuesta LF limitada, subwoofer
AltavozMezcla3dB	Entero positivo (Hz), para cada altavoz.	Punto LF 3dB para altavoces FR y LLF, punto HF 3dB para tipos de altavoces Sub. Puede usarse para hacer coincidir las capacidades de reproducción espectral del equipo de la etapa de mezcla
CanalMezcla	{L, C, R, Ls, Rs, Lss, Rss, Lrs, Rrs, Lts, Rts, ninguno, otro}, para cada altavoz	Altavoz -> mapeo de canales. Se usa "ninguno" para los altavoces que no están asociados
SubAltavozMezcla	Lista de pares (ganancia, número de altavoz). La ganancia es el valor real: $0 \leq \text{Ganancia} \leq 1.0$. El número de altavoz es un número entero. $0 < \text{Número de altavoz} < \text{númMezclaAltavoces}-1$	mapeo altavoz -> sub. Se utiliza para indicar el subwoofer de destino para la gestión de graves de cada altavoz. Cada altavoz se puede gestionar en graves para más de un subwoofer. La ganancia indica la parte de la señal de graves que debe ir a cada sub. Ganancia = 0 indica el final de la lista y que no aparecerá a continuación un número de altavoz. Si un altavoz no tiene gestión de graves, el primer valor de ganancia se establece en 0.
PosMezcla	Coordenadas x, y, z para la posición de mezcla	Posición nominal de mezcla
DimSalaMezcla	x, y, z para las dimensiones de la sala (metros)	Dimensiones nominales de la etapa de mezcla
SalaMezclaRT60	Valor real < 20 .	Etapas nominal de mezcla RT60
DimPantallaMezcla	x, y, z para las dimensiones de la pantalla (en metros)	
PosPantallaMezcla	x, y, z para el centro de la pantalla (en metros)	

Como se muestra en la figura 5, el segundo tipo de metadatos son los metadatos de audio. Cada elemento de audio basado en canales o en objetos consta de metadatos y esencia de audio. La esencia del audio es un torrente monofónico de audio que va en una de las muchas pistas de audio. Los metadatos asociados describen cómo se almacena la esencia de audio (metadatos de audio, por ejemplo, frecuencia de muestreo) o cómo se debe renderizar (metadatos de renderización, por ejemplo, la posición deseada de la fuente de audio). En general, las pistas de audio son continuas a lo largo de la duración del programa de audio. El editor o mezclador del programa es responsable de asignar elementos de audio a las pistas. Se espera que el uso de pistas sea escaso, es decir, que el uso medio de pistas simultáneas puede ser sólo de 16 a 32. En una implantación típica, el audio se transmitirá de manera eficiente utilizando un codificador sin pérdidas. Sin embargo, son posibles implantaciones alternativas, por ejemplo, transmitir datos de audio no codificados o datos de audio codificados con pérdidas. En una implantación típica, el formato consta de hasta 128 pistas de audio donde cada pista tiene una frecuencia de muestreo única y un sistema de codificación único. Cada pista tiene la duración de la característica (sin soporte explícito de carrete). El mapeo de objetos a pistas (multiplexación de tiempo) es responsabilidad del creador de contenido (mezclador).

Como se muestra en la figura 3, los metadatos de audio incluyen los elementos de frecuencia de muestreo, profundidad de bits y sistemas de codificación. La tabla 5 muestra el ID, los valores de ejemplo y la descripción del elemento de metadatos de frecuencia de muestreo.

Tabla 5

ID	Valores	Descripción
FrecMuestreo	16, 24, 32, 44.1, 48, 88.2 96, y ampliable (x1000 muestras/segundo)	El campo FrecMuestreo proporcionará una codificación eficiente de tasas comunes, así como la capacidad de ampliarse a un campo de punto flotante ampliable con una resolución de 0,01

La tabla 6 muestra el ID, los valores de ejemplo y la descripción del elemento de metadatos de profundidad de bits (para PCM y compresión sin pérdidas).

5

Tabla 6

ID	Valores	Descripción
ProfBit	Entero positivo hasta 32	Indicación de la profundidad de bits de la muestra. Las muestras se justificarán a la izquierda si la profundidad de bits es menor que el contenedor (es decir, LSB de llenado cero)

La tabla 7 muestra el ID, los valores de ejemplo y la descripción del elemento de metadatos del sistema de codificación.

10

Tabla 7

ID	Valor	Descripción
Códec	PCM, sin pérdidas, ampliable	Indicación de formato de audio. A cada pista de audio se le puede asignar cualquier tipo de codificación soportado
ETAPA 1		ETAPA 2
NúmGrupo	Entero positivo	Información de agrupación de objetos. Se aplica a objetos de audio y a objetos de canal, por ejemplo, para indicar vástagos.
TipoAudio	{diálogo, música, efectos, m&e, indef, otro}	Tipo de audio. La lista será ampliable e incluirá lo siguiente: Indefinido, Diálogo, Música, Efectos, Foley, Ambiente, Otro.
TxtTipoAudio	Descripción de texto libre	

Como se muestra en la figura 5, el tercer tipo de metadatos es el de los metadatos de renderización. Los metadatos de renderización especifican valores que ayudan al renderizador a coincidir lo más posible con la intención del mezclador original, independientemente del entorno de reproducción. El conjunto de elementos de metadatos es diferente para el audio basado en canales y el audio basado en objetos. Un primer campo de metadatos de renderización selecciona entre los dos tipos de audio: basado en canales o basado en objetos, como se muestra en la tabla 8.

15

Tabla 8

ID	Valor	ETAPA 2
CanUObj	{Canal, Objeto}	Para cada elemento de audio, indicar si se describe utilizando metadatos de objeto o de canal

Los metadatos de renderización para el audio basado en canales comprenden un elemento de metadatos de posición que especifica la posición de la fuente de audio como una o más posiciones de altavoz. La tabla 9 muestra el ID y los valores del elemento de metadatos de posición para el caso basado en canales.

25

Tabla 9

ID	Valores	Descripción
PosCanal	{L, C, R, Ls, Rs, Lss, Rss, Lrs, Rrs, Lts, Rts, Lc, Rc, Crs, Cts, otro}	La posición de la fuente de audio se indica como una de un conjunto de posiciones de altavoz con nombre. El conjunto es ampliable. PosCanAmp y AnchCanAmp proporcionan la posición y la ampliación de los canales de ampliación.

Los metadatos de renderización para el audio basado en canales comprenden también un elemento de control de renderización que especifica ciertas características con respecto a la reproducción de audio basado en canales, como se muestra en la tabla 10.

30

Tabla 10

ID	Valores	Descripción
MezclaAscCan	{no, sí}	Deshabilitar (por defecto) o habilitar mezcla ascendente
ZonasMezclaAscCan	{L, C, R, Ls, Rs, Lss, Rss, Lrs, Rrs, Lts, Rts, Lc, Rc, Crs, Cts, otro}	Indicación de las zonas en las que se permite la mezcla ascendente.
VectMezclaDescCan	Valores reales positivos ≤ 1	Matrices de mezcla descendente de objetos de canal personalizados para configuraciones de canal específicas. La lista de configuración de canales será ampliable e incluirá 5.1 y Dolby Surround 7.1.
VectMezclaAscCan	Valores reales positivos ≤ 1	Matrices de mezcla ascendente de objetos de canal personalizado para configuraciones de canal específicas. La lista de configuración de canal será ampliable e incluirá 5.1, 7.1 y 9.1.
ChanSSBias		Indicación de la pantalla para solicitud envolvente. Muy útil para ajustar la renderización por defecto de los modos de reproducción alternativos (5.1, 7.1).

- 5 Para el audio basado en objetos, los metadatos incluyen elementos análogos a los del audio basado en canales. La tabla 11 proporciona el ID y los valores del elemento de metadatos de posición del objeto. La posición del objeto se describe de una de estas tres maneras: con coordenadas tridimensionales; con un plano y coordenadas bidimensionales; o con una línea y una coordenada unidimensional. El método de renderización puede adaptarse en base al tipo de información de posición.

Tabla 11

ID	Valores	Descripción
FormatPosObj	{3D, 2D, 1D}	Formato de posición
PosObj3D	coordenadas x, y, z	Posición 3D
PosObj2D	3 conjuntos de coordenadas x, y, z para definir un plano y 1 conjunto de coordenadas x, y para indicar la posición en el plano.	Plano + Posición 2D
PosObj1D	2 conjuntos de coordenadas x, y, z para definir una línea y 1 escalar para indicar la posición en la línea	Línea + Posición 1D o Curva + Posición 1D
PantallaPosObj	{sí, no}	Usar la pantalla como referencia. La información de posición se debe escalar y conmutar en base a la mezcla frente al tamaño y la posición de la pantalla de exposición.

- 10 El ID y los valores para los elementos de metadatos de control de renderización de objetos se muestran en la tabla 12. Estos valores proporcionan medios adicionales para controlar u optimizar la renderización de audio basado en objetos.

15 Tabla 12

ID	Valores	Descripción
DispersObj	x o (x, y, z), reales positivos < 1	Ancho de la función de dispersión. Los valores > 0 indican que se debe utilizar más de 1 altavoz. A medida que aumenta el valor, se utilizan más altavoces de manera creciente. La dispersión se indica como un valor único o de forma independiente para cada dimensión. Se puede utilizar para suavizar las panoramizaciones, o para crear ambigüedad de posición.
ASWObj	x o (x, y, z), reales positivos < 1	Ancho de fuente aparente. Los valores mayores indican un ancho mayor de fuente. Puede implantarse mediante el uso de anulación de correlación.
AjustObj	{sí, no}	Ajustar al altavoz más cercano. Útil cuando el timbre de fuente puntual es más importante que la precisión espacial

SuavizAjust Obj	Valor real positivo < 10 (en segundos)	Constante de tiempo de suavizado espacial para el modo "Ajustar a". Lo hace más un "Deslizar a".
TolAjustObj	Valor real positivo < 10	Ajustar a tolerancia: cuánto error espacial (en distancia normalizada, ancho de sala = 1) aceptar antes de revertir a la imagen fantasma.
AlgRendObj	{def, dualBallance, vbap, dbap, 2D, 1D, otro}	Def: elección del renderizador dualBallance: método Dolby
		vbap: panorámica de amplitud basada en vectores
		dbap: panorámica de amplitud basada en la distancia
		2D: junto con PosObj2D, usar vbap con sólo 3 posiciones de fuente (virtuales).
		1D: junto con ObjPos1D, usar la panorámica por pares entre 2 posiciones de fuente (virtuales).
ZonasObj	Valores reales positivos <= 1	Grado de contribución de cualquier zona de altavoz con nombre. Las zonas de altavoz soportadas incluyen: L, C, R, Lss, Rss, Lrs, Rrs, Lts, Rts, Lc, Rc. La lista de zonas de altavoces se podrá ampliar para soportar zonas futuras.
NivelObj	Valores reales positivos <= 2	Nivel de objeto de audio alternativo para configuraciones de canal específicas. La lista de configuración de canales será ampliable e incluirá 5.1 y Dolby Surround 7.1. El objeto puede atenuarse o eliminarse por completo cuando se renderiza en configuraciones de canal más pequeñas.
SolicitSSObj		Indicación de pantalla para solicitud de sala. Muy útil para ajustar la renderización por defecto de los modos de representación alternativos (5.1, 7.1). Se considera "opcional" porque es posible que esta característica no requiera metadatos adicionales; otros datos de renderización podrían modificarse directamente (por ejemplo, la trayectoria panorámica, la matriz de mezcla descendente).

Los metadatos descritos anteriormente e ilustrados en la figura 5 se generan y almacenan como uno o más archivos que están asociados o indexados con el contenido de audio correspondiente, de modo que los torrentes de audio son procesados por el sistema de audio adaptativo interpretando los metadatos generados por el mezclador. Cabe señalar que los metadatos descritos anteriormente son un conjunto de ejemplo de ID, valores y definiciones, y que se pueden incluir otros elementos de metadatos adicionales para su uso en el sistema de audio adaptativo.

Se asocian dos (o más) conjuntos de elementos de metadatos con cada uno de los torrentes de audio basados en canales y objetos. Se aplica un primer conjunto de metadatos a la pluralidad de torrentes de audio para una primera condición del entorno de reproducción, y se aplica un segundo conjunto de metadatos a la pluralidad de torrentes de audio para una segunda condición del entorno de reproducción. El segundo o subsiguiente conjunto de elementos de metadatos reemplaza al primer conjunto de elementos de metadatos para un torrente de audio dado en base a la condición del entorno de reproducción. La condición puede incluir factores tales como el tamaño de la sala, la forma, la composición del material dentro de la sala, la ocupación actual y la densidad de personas en la sala, las características del ruido ambiental, las características de la luz ambiental y cualquier otro factor que pueda afectar al sonido o incluso al estado de ánimo del entorno de reproducción.

Postproducción y masterización

La etapa 110 de renderización del sistema 100 de procesamiento de audio adaptativo puede incluir pasos de postproducción de audio que conducen a la creación de una mezcla final. En una aplicación de cine, las tres categorías principales de sonido que se utilizan en una mezcla de películas son el diálogo, la música y los efectos. Los efectos consisten en sonidos que no son diálogos ni música (por ejemplo, ruido ambiental, ruido de fondo/escena). El diseñador de sonido puede grabar o sintetizar los efectos de sonido o pueden obtenerse de bibliotecas de efectos. Un subgrupo de efectos que involucran fuentes de ruido específicas (por ejemplo, pasos, puertas, etc.) se conocen como Foley y son interpretados por actores de Foley. Los diferentes tipos de sonido son marcados y panoramizados en consecuencia por los ingenieros de grabación.

La figura 6 ilustra un flujo de trabajo de ejemplo para un proceso de postproducción en un sistema de audio adaptativo. Como se muestra en el diagrama 600, todos los componentes de sonido individuales de la música, el

diálogo, el Foley y los efectos se reúnen en la sala de doblaje durante la mezcla final 606, y el mezclador o los mezcladores 604 de regrabación usa/n las mezclas previas (también conocidas como la 'mezcla menos') junto con los objetos de sonido individuales y los datos de posición para crear vástagos como una forma de agrupar, por ejemplo, diálogos, música, efectos, Foley y sonidos de fondo. Además de formar la mezcla final 606, la música y todos los vástagos de efectos pueden usarse como base para crear versiones dobladas de la película. Cada vástago consta de un lecho basado en canales y de varios objetos de audio con metadatos. Los vástagos se combinan para formar la mezcla final. Utilizando información de panoramización de objetos tanto de la estación de trabajo de audio como de la mesa de mezclas, la unidad 608 de renderización y masterización renderiza el audio a las ubicaciones de altavoz en la sala de doblaje. Esta renderización permite a los mezcladores oír cómo se combinan los lechos basados en canal y los objetos de audio, y proporciona también la capacidad de renderizar para diferentes configuraciones. El mezclador puede usar metadatos condicionales, que por defecto son los perfiles relevantes, para controlar cómo se renderiza el contenido en los canales envolventes. De esta manera, los mezcladores mantienen el control total de cómo se reproduce la película en todos los entornos escalables. Puede incluirse un paso de monitorización después del paso 604 de regrabación y del paso 606 de mezcla final, después de uno o de ambos, para permitir que el mezclador escuche y evalúe el contenido intermedio generado durante cada una de estas etapas.

Durante la sesión de masterización, los vástagos, los objetos y los metadatos se juntan en un paquete 614 de audio adaptativo, que es producido por el maestro 610 de impresión. Este paquete también contiene la mezcla retrocompatible (versiones heredadas 5.1 o 7.1) 612 de sonido envolvente de sala. La unidad 608 de renderización/masterización (RMU) puede renderizar esta salida si se desea; se elimina por ello la necesidad de pasos adicionales de flujo de trabajo en la generación de entregables existentes basados en canal. Los archivos de audio se empaquetan utilizando el embalaje de formato de intercambio de material (MXF) estándar. El archivo maestro de mezcla de audio adaptativo también se puede utilizar para generar otros entregables, tales como mezclas estéreo o multicanal de consumo. Los perfiles inteligentes y los metadatos condicionales permiten renderizaciones controladas que pueden reducir significativamente el tiempo requerido para crear tales mezclas.

Se puede usar un sistema de empaquetado para crear un paquete de cine digital para los entregables que incluye una mezcla de audio adaptativo. Los archivos de pistas de audio pueden bloquearse juntos para contribuir a impedir errores de sincronización con los archivos de pistas de audio adaptativos. Ciertos territorios requieren la adición de archivos de pistas durante la fase de empaquetado, por ejemplo, la adición de pistas para personas con discapacidad auditiva (HI) o de narración para personas con discapacidad visual (VI-N) al archivo de pista de audio principal.

La matriz de altavoces en el entorno de reproducción puede comprender cualquier cantidad de altavoces de sonido envolvente, colocados y designados de acuerdo con los estándares de sonido envolvente establecidos. También se puede colocar cualquier cantidad de altavoces adicionales para una renderización precisa del contenido de audio basado en objetos en base a las condiciones del entorno de reproducción. Estos altavoces adicionales pueden ser configurados por un ingeniero de sonido, y esta configuración se proporciona al sistema en forma de un archivo de configuración que es utilizado por el sistema para renderizar los componentes basados en objetos del audio adaptativo a un altavoz o altavoces específico/s dentro de la matriz general de altavoces. El archivo de configuración incluye al menos una lista de designaciones de altavoces y un mapeo de canales para altavoces individuales, información en relación con la agrupación de altavoces y un mapeo del tiempo de ejecución basado en una posición relativa de los altavoces al entorno de reproducción. El mapeo en tiempo de ejecución es utilizado por una característica de ajuste del sistema que renderiza el contenido de audio basado en objetos de fuente puntual a un altavoz específico que está más cerca de la ubicación percibida del sonido según lo previsto por el ingeniero de sonido.

La figura 7 es un diagrama de un flujo de trabajo de ejemplo para un proceso de empaquetado de cine digital que utiliza archivos de audio adaptativos. Como se muestra en el diagrama 700, los archivos de audio que comprenden tanto los archivos de audio adaptativo como los archivos de audio de sonido envolvente 5.1 o 7.1 se meten en un bloque 704 de embalaje/encriptación. Tras la creación del paquete de cine digital en el bloque 706, el archivo PCM MXF (con las pistas adicionales apropiadas adjuntas) se encripta utilizando especificaciones SMPTE de acuerdo con la práctica existente. La MXF de audio adaptativo está empaquetada como un archivo de pista auxiliar y está opcionalmente encriptada usando una clave de contenido simétrica según la especificación SMPTE. Este único DCP 708 se puede enviar a cualquier servidor compatible con iniciativas de cine digital (DCI). En general, cualesquiera instalaciones que no estén adecuadamente equipadas simplemente ignorarán el archivo de pista adicional que contiene la banda sonora de audio adaptativo, y utilizarán el archivo de pista de audio principal existente para la reproducción estándar. Las instalaciones equipadas con procesadores de audio adaptativos apropiados podrán asimilar y reproducir la banda sonora de audio adaptativo cuando corresponda, volviendo a la pista de audio estándar según sea necesario. El componente 704 de embalaje/encriptación también puede proporcionar entrada directamente a un bloque 710 KDM de distribución para generar una clave de seguridad apropiada para su uso en el servidor de cine digital. Otros elementos de película o archivos, tales como subtítulos 714 e imágenes 716, se pueden embalar y encriptar junto con los archivos 702 de audio. En este caso, se pueden incluir ciertos pasos de procesamiento, tales como la compresión 712 en el caso de archivos 716 de imagen.

Con respecto a la gestión de contenido, el sistema 100 de audio adaptativo permite al creador de contenido crear objetos de audio individuales y añadir información sobre el contenido que se puede transferir al sistema de reproducción. Esto permite una gran flexibilidad en la gestión de contenido de audio. Desde el punto de vista de la gestión de contenido, los métodos de audio adaptativo permiten varias funciones diferentes. Estas incluyen cambiar el idioma del contenido reemplazando únicamente el objeto de diálogo para ahorro espacio, eficiencia de descarga, adaptación de reproducción geográfica, etc. Las películas, la televisión y otros programas de entretenimiento se distribuyen típicamente a nivel internacional. Esto a menudo requiere que se cambie el idioma del contenido según el lugar donde se reproducirá (francés para las películas que se proyectan en Francia, alemán para los programas de televisión que se proyectan en Alemania, etc.). Hoy en día, esto a menudo requiere la creación, empaquetado y distribución de una banda sonora de audio completamente independiente. Con el audio adaptativo y su concepto inherente de objetos de audio, el diálogo para una pieza de contenido podría ser un objeto de audio independiente. Esto permite que el idioma del contenido se cambie fácilmente sin actualizar o alterar otros elementos de la banda sonora de audio, como música, efectos, etc. Esto no sólo se aplicaría a idiomas extranjeros, sino también al lenguaje inapropiado para ciertos públicos (por ejemplo, programas de televisión para niños, películas de aerolíneas, etc.), publicidad dirigida, etc.

Consideraciones sobre la instalación y el equipo

El formato de archivo de audio adaptativo y los procesadores asociados permiten cambios en la forma en que se instala, calibra y mantiene el equipo de sala de cine. Con la introducción de muchas más salidas de altavoz potenciales, cada una ecualizada y equilibrada individualmente, existe la necesidad de una ecualización de sala automática inteligente y eficiente en el tiempo, que se pueda realizar mediante la capacidad de ajustar manualmente cualquier ecualización de sala automatizada. El sistema de audio adaptativo utiliza un motor optimizado de ecualización de banda de 1/12^ª octavas. Se pueden procesar hasta 64 salidas para equilibrar con mayor precisión el sonido en el cine. El sistema también permite la monitorización programada de las salidas de los altavoces individuales, desde la salida del procesador de cine hasta el sonido reproducido en el auditorio. Se pueden crear alertas locales o de red para garantizar que se tomen las medidas adecuadas. El sistema flexible de renderización puede eliminar automáticamente un altavoz o amplificador dañado de la cadena de reproducción y renderizar a su alrededor, permitiendo que el espectáculo continúe.

El procesador de cine se puede conectar al servidor de cine digital con conexiones de audio principales 8xAES existentes y una conexión Ethernet para retransmitir datos de audio adaptativo. La reproducción de contenido envolvente 7.1 o 5.1 utiliza las conexiones PCM existentes. Los datos de audio adaptativo se retransmiten a través de Ethernet al procesador de cine para decodificar y renderizar, y la comunicación entre el servidor y el procesador de cine permite identificar y sincronizar el audio. En caso de cualquier problema con la reproducción de la pista de audio adaptativo, el sonido se revierte al audio PCM Dolby Surround 7.1 o 5.1.

Aunque se han descrito ejemplos con respecto a los sistemas de sonido envolvente 5.1 y 7.1, debe tenerse en cuenta que se pueden usar muchas otras configuraciones envolventes presentes y futuras junto con los ejemplos que incluyen 9.1, 11.1 y 13.1 y más allá.

El sistema de audio adaptativo está diseñado para permitir que tanto los creadores de contenido como los expositores decidan cómo se va a renderizar el contenido de sonido en diferentes configuraciones de altavoces de reproducción. El número ideal de canales de salida de altavoz utilizados variará de acuerdo con el tamaño de la sala. De este modo, la colocación recomendada de los altavoces depende de muchos factores, tales como el tamaño, la composición, la configuración de los asientos, el entorno, la cantidad media de público, etc. Se proporcionan aquí ejemplos o configuraciones y trazados de altavoces representativos con fines ilustrativos únicamente, que no pretenden limitar el alcance de ninguna de las realizaciones reivindicadas.

El trazado recomendado de los altavoces para un sistema de audio adaptativo sigue siendo compatible con los sistemas de cine existentes, lo cual es vital como para no comprometer la reproducción de los formatos basados en canal 5.1 y 7.1 existentes. Con el fin de preservar la intención del ingeniero de sonido de audio adaptativo y la intención de los mezcladores de contenido 7.1 y 5.1, las posiciones de los canales de pantalla existentes no deben alterarse de manera demasiado radical en un esfuerzo por reforzar o acentuar la introducción de nuevas ubicaciones de altavoces. En contraste con el uso de los 64 canales de salida disponibles en su totalidad, el formato de audio adaptativo es capaz de ser renderizado con precisión en el cine para configuraciones de altavoces tales como 7.1, por lo que incluso permite que el formato (y los beneficios asociados) se utilicen en salas de cine existentes sin cambios en amplificadores o altavoces.

Las diferentes ubicaciones de los altavoces pueden tener una eficacia diferente dependiendo del diseño de la sala de cine, por lo que en la actualidad no existe un número o ubicación ideal de canales especificados por la industria. El audio adaptativo está pensado para ser verdaderamente adaptable y capaz de reproducirse con precisión en diferentes auditorios, ya sea que tengan un número limitado de canales de reproducción o muchos canales con configuraciones altamente flexibles.

La figura 8 es una vista aérea 800 de un trazado de ejemplo de ubicaciones de altavoces sugeridas, para su uso con un sistema de audio adaptativo en un auditorio típico, y la figura 9 es una vista frontal 900 del trazado de ejemplo de ubicaciones sugeridas para los altavoces en la pantalla del auditorio. La posición de referencia a la que se hace referencia a continuación corresponde a una posición de 2/3 de la distancia desde la pantalla hasta la pared trasera, en la línea central de la pantalla. Los altavoces 801 de pantalla estándar se muestran en sus posiciones habituales con respecto a la pantalla. Los estudios de la percepción de la elevación en el plano de la pantalla han demostrado que los altavoces adicionales 804 detrás de la pantalla, tales como los altavoces de pantalla del centro izquierdo (Lc) y del centro derecho (Rc) (en las ubicaciones de los canales izquierda extra y derecha extra en formatos de película de 70 mm), puede resultar beneficioso para crear panorámicas más suaves en la pantalla. De este modo, se recomiendan tales altavoces opcionales, especialmente en auditorios con pantallas de más de 12 m (40 pies) de ancho. Todos los altavoces de la pantalla deben inclinarse de tal modo que apunten hacia la posición de referencia. La ubicación recomendada del subwoofer 810 detrás de la pantalla debe permanecer sin cambios, incluido el mantenimiento de la ubicación asimétrica de la caja, con respecto al centro de la sala, para impedir la estimulación de ondas estacionarias. Se pueden colocar subwoofers 816 adicionales en la parte trasera de la sala de cine.

Los altavoces envolventes 802 deben cablearse individualmente de nuevo al bastidor del amplificador, y amplificarse individualmente cuando sea posible con un canal dedicado de amplificación de potencia que coincida con el manejo de potencia del altavoz de acuerdo con las especificaciones del fabricante. Idealmente, los altavoces de sonido envolvente deben especificarse para manejar un SPL aumentado para cada altavoz individual, y también con una respuesta de frecuencia más amplia cuando sea posible. Como regla general para una sala de cine de tamaño medio, el espacio entre los altavoces envolventes debe estar entre 2 y 3 m (6'6" y 9'9"), con los altavoces envolventes izquierdo y derecho colocados simétricamente. Sin embargo, la separación de los altavoces de sonido envolvente se considera más eficazmente como ángulos subtendidos de un oyente determinado entre altavoces adyacentes, en lugar de utilizar distancias absolutas entre altavoces. Para una reproducción óptima en todo el auditorio, la distancia angular entre los altavoces adyacentes debe ser de 30 grados o menos, referenciada desde cada una de las cuatro esquinas del área de escucha principal. Se pueden conseguir buenos resultados con un espaciado de hasta 50 grados. Para cada zona de sonido envolvente, los altavoces deben mantener el espaciado lineal adyacente igual al área de asientos siempre que sea posible. El espaciado lineal más allá del área de escucha, por ejemplo entre la primera fila y la pantalla, puede ser ligeramente mayor. La figura 11 es un ejemplo de un posicionamiento de los altavoces envolventes superiores 808 y de los altavoces envolventes laterales 806 con respecto a la posición de referencia.

Los altavoces de sonido envolvente laterales adicionales 806 deben montarse más cerca de la pantalla de lo que dicta la práctica actualmente recomendada de empezar aproximadamente a un tercio de la distancia que media con la parte posterior del auditorio. Estos altavoces no se utilizan como envolventes laterales durante la reproducción de pistas de sonido Dolby Surround 7.1 o 5.1, pero permitirán una transición suave y una mejor coincidencia de timbres al panoramizar objetos desde los altavoces de la pantalla a las zonas envolventes. Para maximizar la impresión de espacio, las matrices de sonido envolvente deben colocarse lo más bajo posible, sujetas a las siguientes restricciones: la colocación vertical de los altavoces de sonido envolvente en la parte frontal de la matriz debe estar razonablemente cerca de la altura del centro acústico del altavoz de la pantalla, y lo suficientemente alto para mantener una buena cobertura en todo el área de los asientos de acuerdo con la directividad del altavoz. La colocación vertical de los altavoces de sonido envolvente debe ser tal que formen una línea recta de adelante a atrás y (típicamente) inclinados hacia arriba para que la elevación relativa de los altavoces de sonido envolvente por encima de los oyentes se mantenga hacia la parte posterior del cine a medida que aumenta la elevación de los asientos, como se muestra en la figura 10, que es una vista lateral de un trazado de ejemplo de ubicaciones de altavoces sugeridas para su uso con un sistema de audio adaptativo en un auditorio típico. En la práctica, esto se puede conseguir de la manera más simple eligiendo la elevación para los altavoces envolventes laterales más delanteros y más traseros, y colocando los altavoces restantes en una línea entre estos puntos.

Con el fin de proporcionar una cobertura óptima para cada altavoz a través del área de asientos, el envolvente lateral 806 y los altavoces traseros 816 y los envolventes superiores 808 deben apuntar hacia la posición de referencia en la sala de cine, bajo pautas definidas con respecto a espaciado, posición, ángulo, etc.

El formato y el sistema de cine de audio adaptativo descrito consiguen niveles mejorados de inmersión y participación del público con respecto a los sistemas actuales al ofrecer nuevas y potentes herramientas de autoría a los mezcladores, y un nuevo procesador de cine con un motor de renderización flexible que optimiza la calidad de audio y los efectos envolventes de la banda sonora en relación con el trazado y las características de los altavoces de cada sala. Además, el sistema mantiene la retrocompatibilidad y minimiza el impacto en los flujos de trabajo de producción y distribución actuales.

Aunque se han descrito realizaciones con respecto a ejemplos e implantaciones en un entorno de cine en el que el contenido de audio adaptativo está asociado con el contenido de la película para su uso en sistemas de procesamiento de cine digital, cabe señalar que las realizaciones también se pueden implantar en entornos de cine. El contenido de audio adaptativo que comprende audio basado en objetos y audio basado en canales puede usarse junto con cualquier contenido relacionado (audio, video, gráfico, etc. asociado), o puede constituir contenido de audio independiente. El entorno de reproducción puede ser cualquier entorno de escucha apropiado, desde

auriculares o monitores de campo cercano hasta salas pequeñas o grandes, automóviles, estadios al aire libre, salas de conciertos, etc.

Aspectos del sistema 100 pueden implantarse en un entorno de red de procesamiento de sonido basado en ordenador apropiado, para procesar archivos de audio digitales o digitalizados. Partes del sistema de audio adaptativo pueden incluir una o más redes que comprenden cualquier número deseado de máquinas individuales, incluyendo uno o más enrutadores (no mostrados) que sirven para almacenar en memoria intermedia y enrutar los datos transmitidos entre los ordenadores. Tal red puede estar construida sobre diversos protocolos de red diferentes, y puede ser Internet, una red de área amplia (WAN), una red de área local (LAN) o cualquier combinación de las mismas. En una realización en la que la red comprende Internet, pueden configurarse una o más máquinas para acceder a Internet a través de programas de navegador web.

Pueden implantarse uno o más de los componentes, bloques, procesos u otros componentes funcionales a través de un programa informático que controla la ejecución de un dispositivo informático basado en procesador del sistema. También debe tenerse en cuenta que las diversas funciones divulgadas en el presente documento pueden describirse utilizando cualquier número de combinaciones de hardware, firmware y/o como datos y/o instrucciones incorporadas en diversos medios legibles por máquina o por ordenador, en términos de su comportamiento, transferencia de registro, componente lógico y/u otras características. Los medios legibles por ordenador en los que se pueden incorporar tales datos y/o instrucciones formateados incluyen, pero no se limitan a, medios de almacenamiento físicos (no transitorios), no volátiles en diversas formas, tales como medios de almacenamiento ópticos, magnéticos o semiconductores.

A menos que el contexto requiera claramente lo contrario, a lo largo de la descripción y las reivindicaciones, las palabras “comprende”, “que comprende” y similares deben interpretarse en un sentido inclusivo en oposición a un sentido exclusivo o exhaustivo; que es como decir que deben interpretarse en el sentido de “que incluyen, pero no se limitan a”. Las palabras que usan el número singular o plural también incluyen el número plural o singular respectivamente. Además, las palabras “aquí”, “a continuación”, “anteriormente”, “más adelante” y palabras de similar importancia se refieren a esta solicitud en su conjunto y no a partes particulares de esta solicitud. Cuando la palabra “o” se usa en referencia a una lista de dos o más elementos, esa palabra cubre todas las siguientes interpretaciones de la palabra: cualquiera de los elementos de la lista, todos los elementos de la lista y cualquier combinación de los elementos de la lista.

Aunque se han descrito una o más implantaciones a modo de ejemplo y en términos de las realizaciones específicas, debe entenderse que una o más implantaciones no se limitan a las realizaciones divulgadas.

REIVINDICACIONES

1. Un sistema para procesar señales de audio, que comprende un componente (106) de autoría configurado para

5 recibir una pluralidad de señales de audio de un programa de audio, definiéndose un programa de audio como un conjunto completo de canales de altavoz y/o canales de objeto y metadatos asociados que describe una presentación de audio espacial deseada, en la que cada canal de altavoz y cada canal de objeto comprende una señal de audio;

10 generar una mezcla (208) de audio adaptativo para el programa de audio que comprende una pluralidad de torrents monofónicos de audio y metadatos asociados con cada uno de la pluralidad de torrents monofónicos de audio y que especifica una ubicación de reproducción de un respectivo torrente de audio monofónico; en el que al menos uno de la pluralidad de torrents monofónicos de audio se identifica como audio basado en canales y en el que los otros, incluyendo al menos uno, de la pluralidad de torrents monofónicos de audio se identifican como audio
15 basado en objetos, y en el que la ubicación de reproducción del audio basado en canales comprende designaciones de altavoz de altavoces (801, 802, 804, 806, 808, 810, 816) en una matriz de altavoces, y la ubicación de reproducción del audio basado en objetos comprende una ubicación en el espacio tridimensional; y

20 encapsular la pluralidad de torrents monofónicos de audio y los metadatos en un torrente de bits para su transmisión a un sistema (110) de renderización configurado para renderizar la pluralidad de torrents monofónicos de audio a una pluralidad de alimentaciones de altavoces correspondientes a altavoces (801, 802, 804, 806, 808, 810, 816) en un entorno de reproducción, y en el que elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican si uno o más componentes de sonido se renderizan a una alimentación de altavoz para reproducción a través de un altavoz más cercano a una ubicación de reproducción
25 prevista del componente de sonido, de tal manera que un respectivo torrente de audio monofónico basado en objetos es renderizado eficazmente por el altavoz más cercano a la ubicación de reproducción prevista,

en el que los elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican adicionalmente un umbral de distorsión espacial, y en el que el elemento de metadatos que indica si
30 un respectivo componente de sonido se renderiza por el altavoz más cercano a la ubicación de reproducción prevista se ignora si la distorsión espacial resultante de renderizar el respectivo componente de sonido por el altavoz más cercano a la ubicación de reproducción prevista excede el umbral de distorsión espacial.

2. Un sistema para procesar señales de audio, que comprende un sistema (110) de renderización configurado para

35 recibir un torrente de bits que encapsula una pluralidad de torrents monofónicos de audio y metadatos asociados con cada uno de la pluralidad de torrents monofónicos de audio y que especifica una ubicación de reproducción de un respectivo torrente monofónico de audio; en el que al menos uno de la pluralidad de torrents monofónicos de audio se identifica como audio basado en canales y en el que los otros, incluyendo al menos uno, de la pluralidad de
40 torrents monofónicos de audio se identifican como audio basado en objetos, y en el que la ubicación de reproducción del audio basado en canales comprende designaciones de altavoz de altavoces (801, 802, 804, 806, 808, 810, 816) en una matriz de altavoces, y la ubicación de reproducción del audio basado en objetos comprende una ubicación en el espacio tridimensional; y

45 renderizar la pluralidad de torrents monofónicos de audio a una pluralidad de alimentaciones de altavoz correspondientes a altavoces (801, 802, 804, 806, 808, 810, 816) en un entorno de reproducción, y en el que elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican si uno o más componentes de sonido se renderizan a una alimentación de altavoz para reproducción a través de un altavoz más cercano a una ubicación de reproducción prevista del componente de sonido, de tal manera que un
50 respectivo torrente de audio monofónico basado en objetos es renderizado eficazmente por el altavoz más cercano a la ubicación de reproducción prevista,

en el que los elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican adicionalmente un umbral de distorsión espacial, y en el que el elemento de metadatos que indica si
55 un respectivo componente de sonido se renderiza por el altavoz más cercano a la ubicación de reproducción prevista se ignora si la distorsión espacial resultante de renderizar el respectivo componente de sonido por el altavoz más cercano a la ubicación de reproducción prevista excede el umbral de distorsión espacial.

3. El sistema de una cualquiera de las reivindicaciones anteriores, en el que los elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican adicionalmente un parámetro de tasa de fundido cruzado, y en el que, cuando el altavoz más cercano a la ubicación de reproducción prevista para el
60 componente de sonido cambia de un primer altavoz a un segundo altavoz, la tasa a la que el componente de sonido cambia del primer altavoz al segundo altavoz se controla en respuesta al parámetro de tasa de fundido cruzado.

65 4. El sistema de una cualquiera de las reivindicaciones anteriores, en el que los elementos de metadatos asociados con cada torrente de audio monofónico basado en objetos indican adicionalmente parámetros espaciales que

controlan la reproducción de un correspondiente componente de sonido que comprende uno o más de: posición del sonido, ancho del sonido y velocidad del sonido.

5. El sistema de una cualquiera de las reivindicaciones anteriores, en el que la ubicación de reproducción para cada uno de la pluralidad de torrentes de audio monofónicos basados en objetos comprende una posición espacial relativa a una pantalla dentro del entorno de reproducción, o una superficie que engloba el entorno de reproducción, y en el que la superficie comprende un plano frontal, un plano posterior, un plano izquierdo, un plano derecho, un plano superior y un plano inferior.

6. Un método de autoría de señales de audio para su renderización, que comprende:

recibir una pluralidad de señales de audio de un programa de audio, definiéndose un programa de audio como un conjunto completo de canales de altavoz y/o canales de objeto y metadatos asociados que describe una presentación de audio espacial deseada, en la que cada canal de altavoz y cada canal de objeto comprende una señal de audio;

generar una mezcla (208) de audio adaptativo para el programa de audio que comprende una pluralidad de torrentes monofónicos de audio y metadatos asociados con cada uno de la pluralidad de torrentes monofónicos de audio y que especifica una ubicación de reproducción de un respectivo torrente de audio monofónico; en el que al menos uno de la pluralidad de torrentes monofónicos de audio se identifica como audio basado en canales y en el que los otros, incluyendo al menos uno, de la pluralidad de torrentes monofónicos de audio se identifican como audio basado en objetos, y en el que la ubicación de reproducción del audio basado en canales comprende designaciones de altavoz de altavoces (801, 802, 804, 806, 808, 810, 816) en una matriz de altavoces, y la ubicación de reproducción del audio basado en objetos comprende una ubicación en el espacio tridimensional; y

encapsular la pluralidad de torrentes monofónicos de audio y los conjuntos de metadatos en un torrente de bits para su transmisión a un sistema (110) de renderización configurado para renderizar la pluralidad de torrentes monofónicos de audio a una pluralidad de alimentaciones de altavoces correspondientes a altavoces (801, 802, 804, 806, 808, 810, 816) en un entorno de reproducción, y en el que elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican si uno o más componentes de sonido se renderizan a una alimentación de altavoz para reproducción a través de un altavoz más cercano a una ubicación de reproducción prevista del componente de sonido, de tal manera que un respectivo torrente de audio monofónico basado en objetos es renderizado eficazmente por el altavoz más cercano a la ubicación de reproducción prevista,

en el que los elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican adicionalmente un umbral de distorsión espacial, y en el que el elemento de metadatos que indica si un respectivo componente de sonido se renderiza por el altavoz más cercano a la ubicación de reproducción prevista se ignora si la distorsión espacial resultante de renderizar el respectivo componente de sonido por el altavoz más cercano a la ubicación de reproducción prevista excede el umbral de distorsión espacial.

7. Un método para renderizar señales de audio, que comprende:

recibir un torrente de bits que encapsula una pluralidad de torrentes monofónicos de audio y metadatos asociados con cada uno de la pluralidad de torrentes monofónicos de audio y que especifica una ubicación de reproducción de un respectivo torrente monofónico de audio; en el que al menos uno de la pluralidad de torrentes monofónicos de audio se identifica como audio basado en canales y en el que los otros, incluyendo al menos uno, de la pluralidad de torrentes monofónicos de audio se identifican como audio basado en objetos, y en el que la ubicación de reproducción del audio basado en canales comprende designaciones de altavoz de altavoces (801, 802, 804, 806, 808, 810, 816) en una matriz de altavoces, y la ubicación de reproducción del audio basado en objetos comprende una ubicación en el espacio tridimensional; y

renderizar la pluralidad de torrentes monofónicos de audio a una pluralidad de alimentaciones de altavoz correspondientes a altavoces (801, 802, 804, 806, 808, 810, 816) en un entorno de reproducción, y en el que elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican si uno o más componentes de sonido se renderizan a una alimentación de altavoz para reproducción a través de un altavoz más cercano a una ubicación de reproducción prevista del componente de sonido, de tal manera que un respectivo torrente de audio monofónico basado en objetos es renderizado eficazmente por el altavoz más cercano a la ubicación de reproducción prevista,

en el que los elementos de metadatos asociados con cada respectivo torrente de audio monofónico basado en objetos indican adicionalmente un umbral de distorsión espacial, y en el que el elemento de metadatos que indica si un respectivo componente de sonido se renderiza por el altavoz más cercano a la ubicación de reproducción prevista se ignora si la distorsión espacial resultante de renderizar el respectivo componente de sonido por el altavoz más cercano a la ubicación de reproducción prevista excede el umbral de distorsión espacial.

8. Un programa de ordenador que comprende instrucciones que, cuando se ejecutan mediante un dispositivo informático, hacen que el dispositivo informático realice todos los pasos del método de acuerdo con la reivindicación 6 o 7.
- 5 9. Un medio de almacenamiento legible por ordenador que almacena el programa de ordenador de acuerdo con la reivindicación 8.

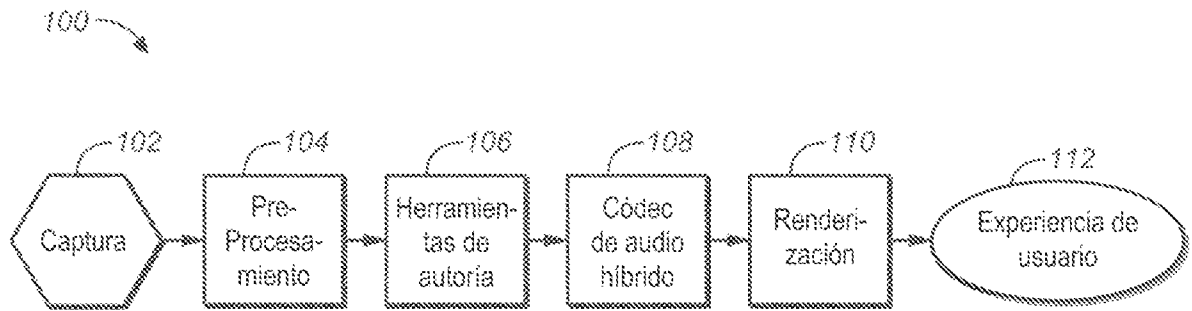


FIG. 1

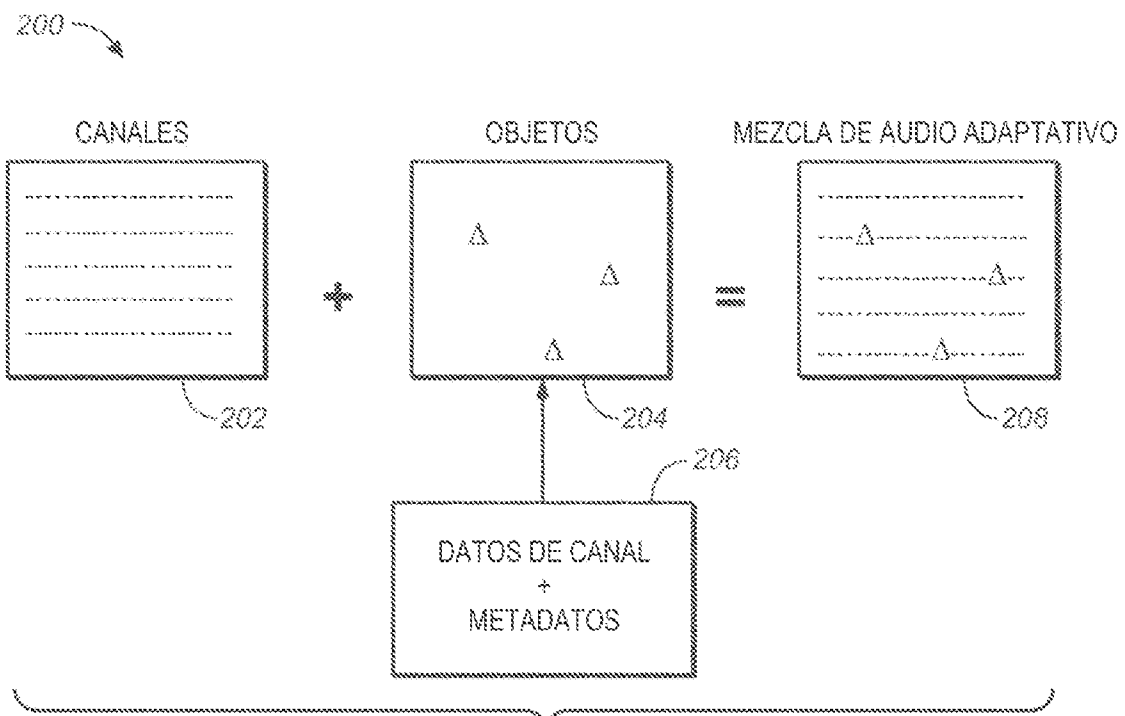


FIG. 2

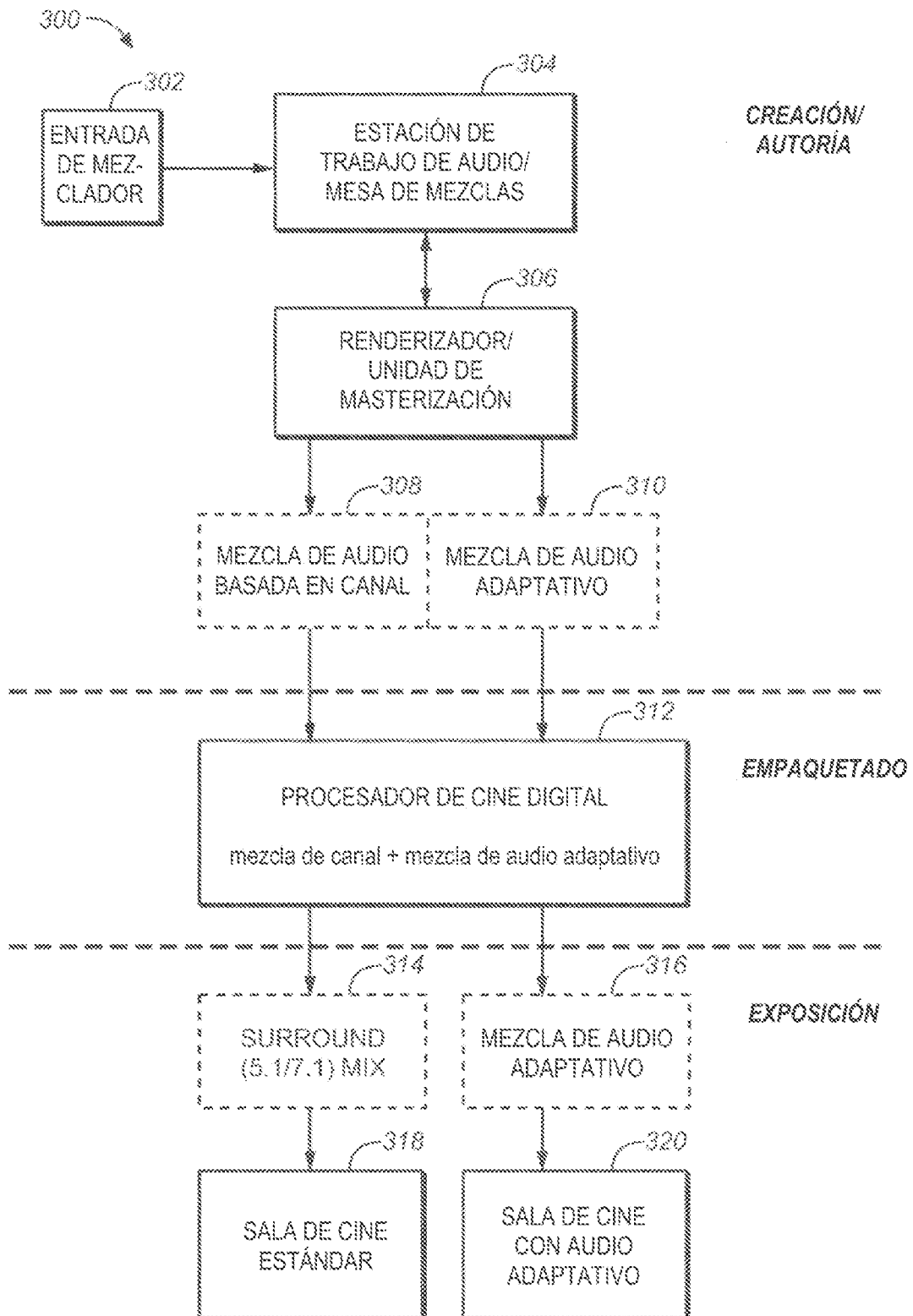


FIG. 3

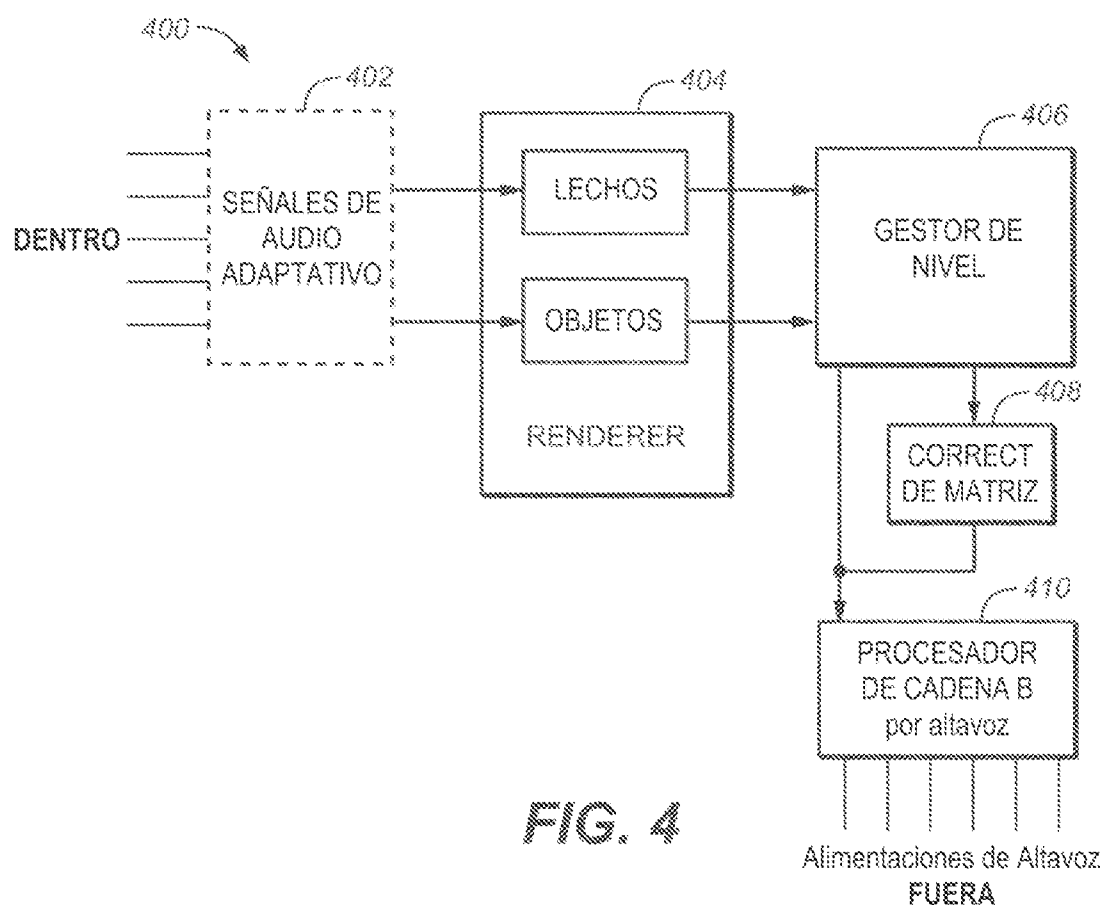


FIG. 4

500

TIPO DE METADATOS	ELEMENTOS DE METADATOS
METADATOS DE PROGRAMA	Velocidad de trama
	Cómputo de pistas
	Descripción de canal ampliable
	Descripción de etapa de mezcla
METADATOS DE AUDIO	Frecuencia de muestreo
	Profundidad de bit
	Sistemas de codificación
METADATOS DE RENDERIZACIÓN	Posición de canal
	Control de renderización de canal
	Posición de objeto
	Control de renderización de objeto

FIG. 5

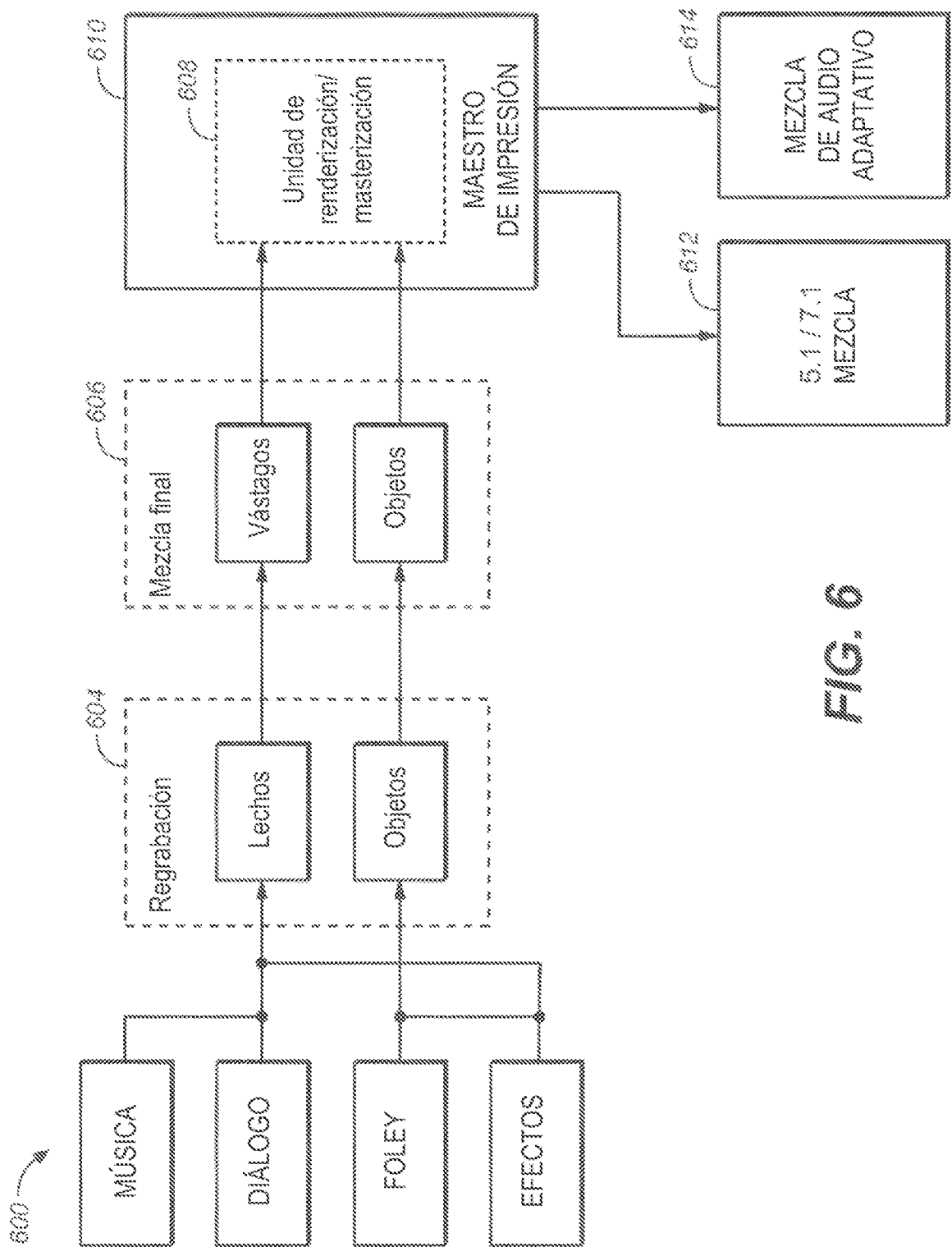


FIG. 6

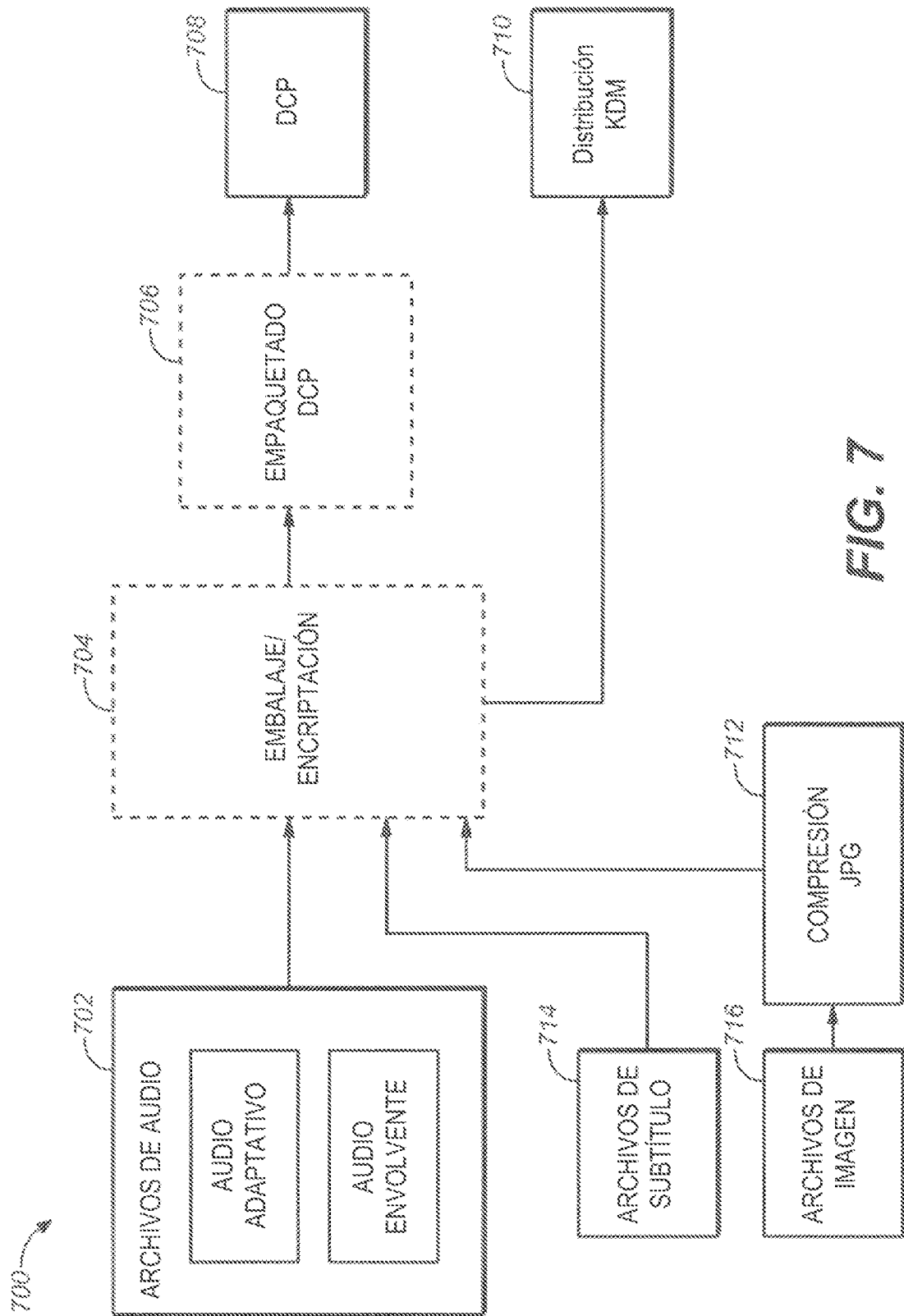


FIG. 7

