

MÉTODO DE DECODIFICACIÓN Y DISPOSITIVO ELECTRÓNICO

Esta solicitud reivindica la prioridad de la Solicitud de patente china n.º 202310843541.X, presentada ante la Administración Nacional de Propiedad Intelectual de China el 10 de julio de 2023 y que se titula "MÉTODO DE DECODIFICACIÓN Y DISPOSITIVO ELECTRÓNICO"

CAMPO TÉCNICO DE LA INVENCION

Las modalidades de esta solicitud se refieren al campo de codificación y decodificación, y en particular, a un método de decodificación y un dispositivo electrónico.

ANTECEDENTES DE LA INVENCION

Para cumplir con los requisitos de audio inmersivo para aplicaciones de realidad extendida (Extended Reality, XR), el estándar de audio MPEG-I proporciona una tecnología de audio de 6 grados de libertad (Degree of Freedom, DoF) basada en ambisónicos de orden superior de múltiples puntos (Higher-order Ambisonics, HOA). En primer lugar, se despliegan varias fuentes de sonido y micrófonos HOA en una pluralidad de puntos en una habitación. Posteriormente, las fuentes de sonido reproducen señales de audio, y los micrófonos HOA en la pluralidad de puntos capturan sonidos en la sala para obtener señales HOA de la pluralidad de puntos. Luego, se obtiene un flujo de bits después de que las señales HOA de la pluralidad de puntos pasan a través de un codificador de audio. Después de decodificar el flujo de bits, un decodificador de audio lleva a cabo la reproducción con realismo con base en información tal como una señal obtenida a través de la decodificación, metadatos HOA y una posición y postura del usuario, para obtener una señal reproducida con realismo binaural.

BREVE DESCRIPCIÓN DE LA INVENCION

Dado lo anterior, esta solicitud proporciona un método de decodificación y un dispositivo electrónico. En el método de decodificación, un lado del decodificador puede realizar un análisis sintáctico un flujo de bits para obtener metadatos de una señal de audio de escena. De esta manera, el lado del decodificador no necesita obtener los metadatos de la señal de audio de escena de otro dispositivo electrónico, y no necesita generar localmente los metadatos de la señal de audio de escena, de modo que se puede reducir el tiempo para

obtener una señal reproducida con realismo binaural a través de la reproducción con realismo, mejorando así la experiencia del usuario.

De acuerdo con un primer aspecto, una modalidad de esta solicitud proporciona un método de codificación. El método incluye: primero obtener una señal de audio de escena y metadatos de la señal de audio de escena, donde la señal de audio de escena describe un campo de sonido de una fuente de sonido en una escena, la escena incluye una pluralidad de micrófonos, los metadatos incluyen primeros metadatos, los primeros metadatos se utilizan en combinación con una distancia entre dos de la pluralidad de micrófonos para determinar un radio de altavoz virtual, el radio de altavoz virtual y una señal de altavoz virtual se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo, y la señal de altavoz virtual se genera con base en una señal de audio reconstruida de la señal de audio de escena; y luego codificar la señal de audio de escena y los metadatos para obtener un flujo de bits. De esta manera, después de que el flujo de bits se transmite a un lado del decodificador, el lado del decodificador puede realizar un análisis sintáctico del flujo de bits para obtener los metadatos de la señal de audio de escena y aplicar los metadatos de la señal de audio de escena a un proceso posterior de obtención de la señal de audio reproducida con realismo a través de la reproducción con realismo. Además, el lado del decodificador no necesita obtener los metadatos de la señal de audio de escena de otro dispositivo electrónico, y no necesita generar localmente los metadatos de la señal de audio de escena, mejorando la eficiencia de obtención de los metadatos de la señal de audio de escena por el lado del decodificador, y reduciendo una cantidad de cálculo de generación de los metadatos de la señal de audio de escena. Por lo tanto, se puede reducir el tiempo para obtener la señal de audio reproducida con realismo a través de la reproducción con realismo, mejorando así la experiencia del usuario.

Por ejemplo, la señal de audio reproducida con realismo puede ser una señal de audio que se emite a un módulo de reproducción (por ejemplo, un auricular) para su reproducción, y puede ser una señal reproducida con realismo binaural.

Por ejemplo, hay una distancia igual entre dos micrófonos cualesquiera en la pluralidad de micrófonos.

Por ejemplo, la señal de audio de escena puede incluir señales HOA de una pluralidad de puntos (donde las señales HOA pueden incluir una señal HOA tridimensional y una señal HOA bidimensional (la cual también se puede denominar señal HOA plana)) y una señal de audio tridimensional. La señal de audio tridimensional puede ser una señal de audio distinta

de la señal HOA en la señal de audio de escena. En esta solicitud, se utiliza para la descripción un ejemplo en el cual la señal de audio de escena son señales HOA de una pluralidad de puntos.

5 Por ejemplo, el micrófono puede ser un micrófono HOA. Varias fuentes de sonido y micrófonos HOA en una pluralidad de puntos se pueden desplegar primero en una habitación (es decir, una escena real). Posteriormente, las fuentes de sonido reproducen señales de audio, y los micrófonos HOA en la pluralidad de puntos capturan sonidos en la sala para obtener señales HOA de la pluralidad de puntos.

10 Por ejemplo, la señal de audio de escena puede ser una señal de audio que se codificará en un lado del codificador, y la señal de audio reconstruida de la señal de audio de escena es una señal de audio obtenida por un lado del decodificador o un módulo de decodificación en el lado del codificador con base en un flujo de bits obtenido codificando la señal de audio de escena. La señal de audio reconstruida también puede denominarse señal de audio de escena reconstruida.

15 Cabe señalar que una secuencia de codificación de la señal de audio de la escena y la codificación de los metadatos no está limitada en esta solicitud.

20 De acuerdo con el primer aspecto, el método además incluye: escribir un primer identificador en el flujo de bits, donde el primer identificador indica complejidad de representación, y la complejidad de representación se utiliza para determinar una cantidad de altavoces virtuales. De esta manera, el lado del decodificador puede llevar a cabo la reproducción con realismo con base en el número de altavoces virtuales, el radio del altavoz virtual y la señal del altavoz virtual, para obtener la señal de audio reproducida con realismo.

25 Por ejemplo, la calidad de audio de una señal reproducida con realismo binaural obtenida a través de la reproducción con realismo varía correspondientemente con la complejidad de la reproducción con realismo. Un valor del primer identificador puede definirse por el usuario, o puede determinarse con base en una capacidad informática del lado del decodificador. De esta manera, el lado del decodificador puede obtener, a través de la reproducción con realismo, una señal reproducida con realismo binaural cuya calidad de audio coincida con factores tales como un requisito del usuario y una capacidad informática de un dispositivo electrónico.

30

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del

primer aspecto, se escribe un segundo identificador en el flujo de bits, donde el segundo identificador indica si el flujo de bits incluye los metadatos de la señal de audio de escena.

Puede haber una pluralidad de reproductores con realismo (cada reproductor con realismo utiliza un método de reproducción con realismo diferente) en el lado del decodificador.

5 La pluralidad de reproductores con realismo incluye un reproductor con realismo que lleva a cabo la reproducción con realismo con base en los metadatos de la señal de audio de escena y un reproductor con realismo que lleva a cabo la reproducción con realismo con base en otra información. Además, el segundo identificador también puede indicar si realizar la representación con base en los metadatos de la señal de audio de escena. De esta manera,
10 el lado del decodificador puede determinar, con base en un valor del segundo identificador obtenido del flujo de bits a través del análisis sintáctico, cual reproductor con realismo o cual algoritmo de reproducción con realismo se utilizará para la reproducción con realismo.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, el segundo identificador también indica si el flujo de bits además incluye el
15 primer identificador.

La pluralidad de reproductores con realismo incluidos en el lado del decodificador puede incluir un reproductor con realismo que lleva a cabo la reproducción con realismo con base en los metadatos de la señal de audio de escena y el primer identificador y un reproductor con realismo que lleva a cabo la reproducción con realismo con base en otra información.
20 Además, el segundo identificador también puede indicar si realizar la representación con base en los metadatos de la señal de audio de escena y el primer identificador. De esta manera, el lado del decodificador puede determinar, con base en un valor del segundo identificador obtenido del flujo de bits a través del análisis sintáctico, cual reproductor con realismo o qué algoritmo de reproducción con realismo se utilizará para la reproducción con realismo.

25 De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, el método además incluye: escribir un tercer identificador en el flujo de bits, donde el tercer identificador indica si el flujo de bits incluye el primer identificador. En este caso, el segundo identificador puede indicar solo si el flujo de bits incluye los metadatos.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del
30 primer aspecto, obtener los metadatos de la señal de audio de escena incluye: obtener información de descripción de escena de la escena; y determinar los metadatos con base en la información de descripción de escena de la escena. De esta manera, una señal reproducida

con realismo binaural obtenida a través de la reproducción con realismo basada en los metadatos de la señal de audio de la escena está más cerca de una señal de audio real escuchada en un entorno real en el cual se encuentra un usuario.

Por ejemplo, la información de descripción de escena puede ser información para describir la escena y, por ejemplo, puede incluir, entre otros: identificadores, información de posición, características, direcciones y similares de micrófonos HOA en una pluralidad de puntos en una escena real; coordenadas, un material y similares de un objeto incluido en la escena real; un identificador, información de posición, una característica, una dirección y similares de una fuente de sonido en la escena real; una distancia crítica de la sala en la escena real; un parámetro de reverberación acústica (por ejemplo, una relación de energía directa a reverberación (Direct-to-Reverberation energy Ratio, DRR), tiempo de reverberación (Reverberation Time) (tal como T20 (tiempo de atenuación de energía de 20 dB), T30 o T60) y claridad (Clarity) (tal como C50 (una relación de energía de sonido dentro de 50 ms a energía de sonido después de 50 ms) o C80)); y similares. Esto no está limitado en esta solicitud.

Por ejemplo, la distancia crítica de la sala puede significar una distancia desde un punto en el cual una densidad de energía sonora directa es igual a una densidad de energía sonora de reverberación hasta un centro de una fuente de sonido en una dirección específica cuando la fuente de sonido irradia de forma estable una onda sonora en el interior (en un entorno real), y también se denomina distancia de campo de difusión. Cuando la directividad de la radiación de la fuente de sonido es esférica, las distancias críticas en todas las direcciones son iguales, y también se denominan radio de reverberación en este caso.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, determinar los metadatos con base en la información de descripción de escena de la escena incluye: cuando la información de descripción de escena incluye la distancia crítica de sala de la escena, determinar los primeros metadatos con base en un producto de la distancia crítica de sala y un valor preestablecido.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, determinar los metadatos con base en la información de descripción de escena de la escena incluye: cuando la información de descripción de escena incluye el parámetro de reverberación acústica de la escena, determinar un tipo de reverberación de la escena con base en el parámetro de reverberación acústica de la escena; y determinar los primeros metadatos con base en el tipo de reverberación de la escena.

Puede haber una pluralidad de tipos de reverberación, y diferentes tipos de reverberación corresponden a diferentes primeros metadatos. Por ejemplo, los primeros metadatos correspondientes a un primer tipo de reverberación (correspondientes a una escena sonora seca o una habitación no reverberante) son 0.5, y los primeros metadatos correspondientes a un segundo tipo de reverberación (correspondientes a una escena sonora húmeda o una habitación reverberante) son 1.5.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, los metadatos además incluyen segundos metadatos, y determinar los metadatos con base en la información de descripción de escena de la escena incluye: cuando la información de descripción de escena incluye información de posición de la pluralidad de micrófonos, usar la información de posición de la pluralidad de micrófonos como los segundos metadatos, donde los segundos metadatos se usan para determinar la distancia entre dos de la pluralidad de micrófonos.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, los metadatos además incluyen segundos metadatos, y determinar los metadatos con base en la información de descripción de escena de la escena incluye: cuando la información de descripción de escena incluye información de posición de la pluralidad de micrófonos, determinar la distancia entre dos de la pluralidad de micrófonos con base en la información de posición de la pluralidad de micrófonos; y usar la distancia entre dos de la pluralidad de micrófonos como los segundos metadatos.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, obtener los metadatos de la señal de audio de escena incluye: obtener los metadatos que son de la señal de audio de escena y que son introducidos por el usuario. En otras palabras, el usuario puede personalizar los metadatos de la señal de audio de escena, de modo que una señal reproducida con realismo binaural obtenida posteriormente por el lado del decodificador a través de la reproducción con realismo pueda coincidir con un requisito personalizado del usuario.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, el flujo de bits incluye un elemento de sintaxis de los metadatos de la señal de audio de escena y la información de descripción correspondiente.

Por ejemplo, el elemento de sintaxis de los metadatos de la señal de audio de escena y la información de descripción correspondiente pueden agregarse a una estructura de

sintaxis original (o denominarse tabla de elementos de sintaxis de alto nivel) del flujo de bits. De esta manera, el lado del codificador puede escribir los metadatos de la señal de audio de escena en el flujo de bits con base en el elemento de sintaxis de los metadatos de la señal de audio de escena y la información de descripción correspondiente en la estructura de sintaxis del flujo de bits.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, el flujo de bits incluye un elemento de sintaxis del primer identificador y la información de descripción correspondiente.

Por ejemplo, el elemento de sintaxis del primer identificador y la información de descripción correspondiente se pueden agregar a la estructura de sintaxis original del flujo de bits. De esta manera, el lado del codificador puede escribir el primer identificador en el flujo de bits con base en el elemento de sintaxis del primer identificador y la información de descripción correspondiente en la estructura de sintaxis del flujo de bits.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, el flujo de bits incluye un elemento de sintaxis del segundo identificador y la información de descripción correspondiente.

Por ejemplo, el elemento de sintaxis del segundo identificador y la información de descripción correspondiente se pueden agregar a la estructura de sintaxis original del flujo de bits. De esta manera, el lado del codificador puede escribir el segundo identificador en el flujo de bits con base en el elemento de sintaxis del segundo identificador y la información de descripción correspondiente en la estructura de sintaxis del flujo de bits.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, el flujo de bits incluye un elemento de sintaxis del tercer identificador y la información de descripción correspondiente.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, la señal de audio de escena incluye señales HOA ambisónicas de orden superior de una pluralidad de puntos.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, un elemento de sintaxis de los primeros metadatos es distanceFactor.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del

primer aspecto, el elemento de sintaxis del primer identificador es complexityLevel.

De acuerdo con cualquiera del primer aspecto o las implementaciones anteriores del primer aspecto, el elemento de sintaxis del segundo identificador es hoaGroupHasLowProfileConfig.

5 De acuerdo con un segundo aspecto, una modalidad de esta solicitud proporciona un método de decodificación. El método incluye: recibir un flujo de bits; realizar un análisis sintáctico del flujo de bits para obtener metadatos de una señal de audio de escena; y obtener una señal de audio reconstruida de la señal de audio de escena con base en el flujo de bits, donde la señal de audio de escena describe un campo de sonido de una fuente de sonido en
10 una escena, la escena incluye una pluralidad de micrófonos, los metadatos incluyen primeros metadatos, los primeros metadatos se utilizan en combinación con una distancia entre dos de la pluralidad de micrófonos para determinar un radio de altavoz virtual, el radio de altavoz virtual y una señal de altavoz virtual se utilizan para reproducirla de forma realista para obtener una señal de audio reproducida de forma realista, y la señal de altavoz virtual se genera con
15 base en la señal de audio reconstruida. De esta manera, un lado del decodificador no necesita obtener los metadatos de la señal de audio de escena de otro dispositivo electrónico, y no necesita generar localmente los metadatos de la señal de audio de escena, mejorando la eficiencia de obtención de los metadatos de la señal de audio de escena, y reduciendo una cantidad de cálculo de generación de los metadatos de la señal de audio de escena. Por lo
20 tanto, se puede reducir el tiempo para obtener una señal reproducida de forma realista biaural a través de la reproducción realista, mejorando así la experiencia del usuario.

De acuerdo con el segundo aspecto, el método además incluye: realizar un análisis sintáctico del flujo de bits para obtener un primer identificador, donde el primer identificador indica complejidad de representación, y la complejidad de representación se utiliza para
25 determinar una cantidad de altavoces virtuales.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el método además incluye: realizar un análisis sintáctico del flujo de bits para obtener un segundo identificador, donde un valor del segundo identificador que es un primer valor de identificador preestablecido indica que el flujo de bits incluye los metadatos.
30 En otras palabras, cuando el valor del segundo identificador es el primer valor de identificador preestablecido, se puede determinar que el lado del decodificador lleva a cabo la reproducción con realismo con base en los metadatos de la señal de audio de escena, y se determina que el flujo de bits incluye los metadatos de la señal de audio de escena. Además, el flujo de bits

puede continuar analizándose para obtener los metadatos de la señal de audio de escena.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el valor del segundo identificador que es el primer valor de identificador preestablecido indica que el flujo de bits además incluye el primer identificador. En otras palabras, cuando el valor del segundo identificador es el primer valor de identificador preestablecido, se puede determinar que el lado del decodificador realiza la representación con base en los metadatos de la señal de audio de escena y el primer identificador, y también se puede determinar que el flujo de bits además incluye el primer identificador. Además, el flujo de bits puede continuar analizándose para obtener el primer identificador.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el método además incluye: realizar un análisis sintáctico del flujo de bits para obtener un tercer identificador, donde un valor del tercer identificador que es un segundo valor de identificador preestablecido indica que el flujo de bits incluye el primer identificador.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, los metadatos además incluyen segundos metadatos, y los segundos metadatos se utilizan para determinar la distancia entre dos de la pluralidad de micrófonos.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, la señal de audio de escena incluye señales HOA ambisónicas de orden superior de una pluralidad de puntos, y la señal de audio reconstruida incluye señales de audio reconstruidas de las señales HOA de la pluralidad de puntos.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el flujo de bits incluye un elemento sintáctico de los metadatos y la información de descripción correspondiente; y realizar un análisis sintáctico del flujo de bits para obtener los metadatos de la señal de audio de escena incluye: realizar un análisis sintáctico del flujo de bits para obtener los metadatos con base en el elemento sintáctico de los metadatos y la información de descripción correspondiente.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el flujo de bits incluye un elemento sintáctico del primer identificador y la información de descripción correspondiente; y realizar un análisis sintáctico del flujo de bits para obtener el primer identificador incluye: realizar un análisis sintáctico del flujo de bits para obtener el primer identificador con base en el elemento sintáctico del primer identificador y la

información de descripción correspondiente.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el flujo de bits incluye un elemento sintáctico del segundo identificador y la información de descripción correspondiente; y realizar un análisis sintáctico del flujo de bits para obtener el segundo identificador incluye: realizar un análisis sintáctico del flujo de bits para obtener el segundo identificador con base en el elemento sintáctico del segundo identificador y la información de descripción correspondiente.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el flujo de bits incluye un elemento sintáctico del tercer identificador y la información de descripción correspondiente; y realizar un análisis sintáctico del flujo de bits para obtener el tercer identificador incluye: realizar un análisis sintáctico del flujo de bits para obtener el tercer identificador con base en el elemento sintáctico del tercer identificador y la información de descripción correspondiente.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, un elemento de sintaxis de los primeros metadatos es distanceFactor.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el elemento de sintaxis del primer identificador es complexityLevel.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el elemento de sintaxis del segundo identificador es hoaGroupHasLowProfileConfig.

De acuerdo con cualquiera del segundo aspecto o las implementaciones anteriores del segundo aspecto, el método además incluye: generar la señal de altavoz virtual con base en la señal de audio reconstruida de la señal de audio de escena; determinar el radio del altavoz virtual con base en los primeros metadatos y la distancia entre dos de la pluralidad de micrófonos; y llevar a cabo la reproducción con realismo con base en el radio del altavoz virtual y la señal del altavoz virtual, para obtener la señal de audio reproducida con realismo.

El segundo aspecto y cualquier implementación del segundo aspecto corresponden respectivamente al primer aspecto y cualquier implementación del primer aspecto. Para los efectos técnicos correspondientes al segundo aspecto y cualquier implementación del segundo aspecto, consulte los efectos técnicos correspondientes al primer aspecto y cualquier implementación del primer aspecto. En el presente documento no se describen los detalles

nuevamente.

De acuerdo con un tercer aspecto, una modalidad de esta solicitud proporciona un método de codificación. El método incluye: primero obtener una señal de audio de escena; y luego codificar la señal de audio de escena y un primer identificador para obtener un flujo de bits, donde el primer identificador indica complejidad de reproducción con realismo, la complejidad de reproducción con realismo se utiliza para determinar un número de altavoces virtuales, una señal de audio reconstruida de la señal de audio de escena se utiliza para generar una señal de altavoz virtual, y la señal de altavoz virtual y el número de altavoces virtuales se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo.

De esta manera, un lado del decodificador puede llevar a cabo la reproducción con realismo con base en el número de altavoces virtuales, un radio de altavoz virtual y la señal de altavoz virtual, para obtener la señal de audio reproducida con realismo. Por ejemplo, la calidad de audio de una señal reproducida con realismo binaural obtenida a través de la reproducción con realismo varía correspondientemente con la complejidad de la reproducción con realismo. Un valor del primer identificador puede definirse por el usuario, o puede determinarse con base en una capacidad informática del lado del decodificador. De esta manera, el lado del decodificador puede obtener, a través de la reproducción con realismo, una señal reproducida con realismo binaural cuya calidad de audio coincida con factores tales como un requisito del usuario y una capacidad informática de un dispositivo electrónico.

De acuerdo con el tercer aspecto, el método además incluye: escribir un segundo identificador en el flujo de bits, donde el segundo identificador indica si el flujo de bits incluye el primer identificador.

De acuerdo con cualquiera del tercer aspecto o las implementaciones anteriores del tercer aspecto, el flujo de bits incluye un elemento de sintaxis del primer identificador y la información de descripción correspondiente.

De acuerdo con cualquiera del tercer aspecto o las implementaciones anteriores del tercer aspecto, el flujo de bits incluye un elemento de sintaxis del segundo identificador y la información de descripción correspondiente.

De acuerdo con cualquiera del tercer aspecto o las implementaciones anteriores del tercer aspecto, la señal de audio de escena incluye señales HOA ambisónicas de orden

superior de una pluralidad de puntos.

De acuerdo con cualquiera del tercer aspecto o las implementaciones anteriores del tercer aspecto, el elemento de sintaxis del primer identificador es complexityLevel.

El tercer aspecto y cualquier implementación del tercer aspecto corresponden respectivamente al primer aspecto y cualquier implementación del primer aspecto. Para los efectos técnicos correspondientes al tercer aspecto y cualquier implementación del tercer aspecto, consulte los efectos técnicos correspondientes al primer aspecto y cualquier implementación del primer aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un cuarto aspecto, una modalidad de esta solicitud además proporciona un método de decodificación. El método incluye: recibir un flujo de bits; realizar un análisis sintáctico del flujo de bits para obtener un primer identificador; y obtener una señal de audio reconstruida basada en el flujo de bits, donde el primer identificador indica complejidad de reproducción con realismo, la complejidad de reproducción con realismo se utiliza para determinar un número de altavoces virtuales, la señal de audio reconstruida se utiliza para generar una señal de altavoz virtual, y la señal de altavoz virtual y el número de altavoces virtuales se utilizan para la reproducción con realismo para obtener una señal de audio reproducida con realismo.

De acuerdo con el cuarto aspecto, el método además incluye: realizar un análisis sintáctico del flujo de bits para obtener un segundo identificador, donde un valor del segundo identificador que es un valor de identificador preestablecido indica que el flujo de bits incluye el primer identificador.

De acuerdo con cualquiera del cuarto aspecto o las implementaciones anteriores del cuarto aspecto, el flujo de bits incluye un elemento sintáctico del primer identificador y la información de descripción correspondiente; y realizar un análisis sintáctico del flujo de bits para obtener el primer identificador incluye: realizar un análisis sintáctico del flujo de bits para obtener el primer identificador con base en el elemento sintáctico del primer identificador y la información de descripción correspondiente.

De acuerdo con cualquiera del cuarto aspecto o las implementaciones anteriores del cuarto aspecto, el flujo de bits incluye un elemento sintáctico del segundo identificador y la información de descripción correspondiente; y realizar un análisis sintáctico del flujo de bits

para obtener el segundo identificador incluye: realizar un análisis sintáctico del flujo de bits para obtener el segundo identificador con base en el elemento sintáctico del segundo identificador y la información de descripción correspondiente.

5 De acuerdo con cualquiera del cuarto aspecto o las implementaciones anteriores del cuarto aspecto, la señal de audio reconstruida incluye señales de audio reconstruidas de señales HOA ambisónicas de orden superior de una pluralidad de puntos.

De acuerdo con cualquiera del cuarto aspecto o las implementaciones anteriores del cuarto aspecto, el elemento de sintaxis del primer identificador es complexityLevel.

10 De acuerdo con cualquiera del cuarto aspecto o las implementaciones anteriores del cuarto aspecto, el método además incluye: generar la señal de altavoz virtual con base en la señal de audio reconstruida de una señal de audio de escena; determinar el número de altavoces virtuales con base en el primer identificador; y llevar a cabo la reproducción con realismo con base en el número de altavoces virtuales y la señal de altavoz virtual, para obtener la señal de audio reproducida con realismo.

15 El cuarto aspecto y cualquier implementación del cuarto aspecto corresponden respectivamente al primer aspecto y cualquier implementación del primer aspecto. Para los efectos técnicos correspondientes al cuarto aspecto y cualquier implementación del cuarto aspecto, consulte los efectos técnicos correspondientes al primer aspecto y cualquier implementación del primer aspecto. En el presente documento no se describen los detalles
20 nuevamente.

De acuerdo con un quinto aspecto, una modalidad de esta solicitud proporciona un aparato de codificación. El aparato de codificación incluye:

un primer módulo de obtención de datos, configurado para obtener una señal de audio de escena y metadatos de la señal de audio de escena, donde la señal de audio de escena describe un campo de sonido de una fuente de sonido en una escena, la escena incluye una pluralidad de micrófonos, los metadatos incluyen primeros metadatos, los primeros metadatos se utilizan en combinación con una distancia entre dos de la pluralidad de micrófonos para determinar un radio de altavoz virtual, el radio de altavoz virtual y una señal de altavoz virtual se utilizan para reproducir con realismo para obtener una señal de audio reproducida con
25 realismo, y la señal de altavoz virtual se genera con base en una señal de audio reconstruida de la señal de audio de escena; y
30

un primer módulo de codificación, configurado para codificar la señal de audio de escena y los metadatos para obtener un flujo de bits.

Adicionalmente, el aparato de codificación además puede incluir un módulo de comunicación.

5 Debe entenderse que el aparato de codificación además puede realizar el primer aspecto y cualquier implementación del primer aspecto. Para los efectos correspondientes, consulte los efectos técnicos correspondientes al primer aspecto y cualquier implementación del primer aspecto. En el presente documento no se describen los detalles nuevamente.

10 De acuerdo con un sexto aspecto, una modalidad de esta solicitud proporciona un aparato de decodificación. El aparato de decodificación incluye:

un primer módulo de recepción, configurado para recibir un flujo de bits;

un primer módulo de análisis sintáctico, configurado para realizar un análisis sintáctico del flujo de bits para obtener metadatos de una señal de audio de escena; y

15 un primer módulo de obtención de señal, configurado para obtener una señal de audio reconstruida de la señal de audio de escena con base en el flujo de bits, donde la señal de audio de escena describe un campo de sonido de una fuente de sonido en una escena, la escena incluye una pluralidad de micrófonos, los metadatos incluyen primeros metadatos, los primeros metadatos se utilizan en combinación con una distancia entre dos de la pluralidad de micrófonos para determinar un radio de altavoz virtual, el radio de altavoz virtual y una
20 señal de altavoz virtual se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo, y la señal de altavoz virtual se genera con base en la señal de audio reconstruida.

Adicionalmente, el aparato de decodificación además puede incluir un módulo de comunicación.

25 Debe entenderse que el aparato de decodificación además puede realizar el segundo aspecto y cualquier implementación del segundo aspecto. Para los efectos correspondientes, consulte los efectos técnicos correspondientes al segundo aspecto y cualquier implementación del segundo aspecto. En el presente documento no se describen los detalles nuevamente.

30 De acuerdo con un séptimo aspecto, una modalidad de esta solicitud proporciona un

aparato de codificación. El aparato de codificación incluye:

un segundo módulo de obtención de datos, configurado para obtener una señal de audio de escena; y

5 un segundo módulo de codificación, configurado para codificar la señal de audio de escena y un primer identificador para obtener un flujo de bits, donde el primer identificador indica complejidad de reproducción con realismo, la complejidad de reproducción con realismo se utiliza para determinar un número de altavoces virtuales, una señal de audio reconstruida de la señal de audio de escena se utiliza para generar una señal de altavoz virtual, y la señal de altavoz virtual y el número de altavoces virtuales se utilizan para reproducir con realismo
10 para obtener una señal de audio reproducida con realismo.

Adicionalmente, el aparato de decodificación además puede incluir un módulo de comunicación.

Debe entenderse que el aparato de codificación además puede realizar el tercer aspecto y cualquier implementación del tercer aspecto. Para los efectos correspondientes,
15 consulte los efectos técnicos correspondientes al tercer aspecto y cualquier implementación del tercer aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un octavo aspecto, una modalidad de esta solicitud proporciona un aparato de decodificación. El aparato de decodificación incluye:

un segundo módulo de recepción, configurado para recibir un flujo de bits;

20 un segundo módulo de análisis sintáctico, configurado para realizar un análisis sintáctico del flujo de bits para obtener un primer identificador; y

un segundo módulo de obtención de señal, configurado para obtener una señal de audio reconstruida de una señal de audio de escena basada en el flujo de bits, donde el primer identificador indica complejidad de reproducción con realismo, la complejidad de reproducción
25 con realismo se utiliza para determinar un número de altavoces virtuales, la señal de audio reconstruida se utiliza para generar una señal de altavoz virtual, y la señal de altavoz virtual y el número de altavoces virtuales se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo.

Adicionalmente, el aparato de decodificación además puede incluir un módulo de
30 comunicación.

Debe entenderse que el aparato de decodificación además puede realizar el cuarto aspecto y cualquier implementación del cuarto aspecto. Para los efectos correspondientes, consulte los efectos técnicos correspondientes al cuarto aspecto y cualquier implementación del cuarto aspecto. En el presente documento no se describen los detalles nuevamente.

5 De acuerdo con un noveno aspecto, una modalidad de esta solicitud proporciona un dispositivo electrónico, que incluye una memoria y un procesador. La memoria está acoplada al procesador. La memoria almacena las instrucciones del programa. Cuando las instrucciones de programa son ejecutadas por el procesador, el dispositivo electrónico está habilitado para realizar el método de codificación de acuerdo con el primer aspecto o cualquier
10 posible implementación del primer aspecto, o el dispositivo electrónico está habilitado para realizar el método de codificación de acuerdo con el tercer aspecto o cualquier posible implementación del tercer aspecto.

El noveno aspecto y cualquier implementación del noveno aspecto corresponden respectivamente al primer aspecto y cualquier implementación del primer aspecto, o
15 corresponden respectivamente al tercer aspecto y cualquier implementación del tercer aspecto. Para los efectos técnicos correspondientes al noveno aspecto y cualquier implementación del noveno aspecto, consulte los efectos técnicos correspondientes al primer aspecto y cualquier implementación del primer aspecto, o consulte los efectos técnicos correspondientes al tercer aspecto y cualquier implementación del tercer aspecto. En el
20 presente documento no se describen los detalles nuevamente.

De acuerdo con un décimo aspecto, una modalidad de esta solicitud proporciona un dispositivo electrónico, que incluye una memoria y un procesador. La memoria está acoplada al procesador. La memoria almacena las instrucciones del programa. Cuando las instrucciones de programa son ejecutadas por el procesador, el dispositivo electrónico está
25 habilitado para realizar el método de decodificación de acuerdo con el segundo aspecto o cualquier posible implementación del segundo aspecto, o el dispositivo electrónico está habilitado para realizar el método de decodificación de acuerdo con el cuarto aspecto o cualquier posible implementación del cuarto aspecto.

El décimo aspecto y cualquier implementación del décimo aspecto corresponden
30 respectivamente al segundo aspecto y cualquier implementación del segundo aspecto, o corresponden respectivamente al cuarto aspecto y cualquier implementación del cuarto aspecto. Para los efectos técnicos correspondientes al décimo aspecto y cualquier implementación del décimo aspecto, referirse a los efectos técnicos correspondientes al

segundo aspecto y cualquier implementación del segundo aspecto, o referirse a los efectos técnicos correspondientes al cuarto aspecto y cualquier implementación del cuarto aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un undécimo aspecto, una modalidad de esta solicitud proporciona un chip, que incluye uno o más circuitos de interfaz y uno o más procesadores. El uno o más procesadores reciben o envían datos a través del uno o más circuitos de interfaz, y cuando el uno o más procesadores ejecutan instrucciones informáticas, se realiza el método de acuerdo con el primer aspecto o cualquier posible implementación del primer aspecto, o se realiza el método de acuerdo con el tercer aspecto o cualquier posible implementación del tercer aspecto.

El undécimo aspecto y cualquier implementación del undécimo aspecto corresponden respectivamente al primer aspecto y cualquier implementación del primer aspecto, o corresponden respectivamente al tercer aspecto y cualquier implementación del tercer aspecto. Para los efectos técnicos correspondientes al undécimo aspecto y cualquier implementación del undécimo aspecto, referirse a los efectos técnicos correspondientes al primer aspecto y cualquier implementación del primer aspecto, o referirse a los efectos técnicos correspondientes al tercer aspecto y cualquier implementación del tercer aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un duodécimo aspecto, una modalidad de esta solicitud proporciona un chip, que incluye uno o más circuitos de interfaz y uno o más procesadores. El uno o más procesadores reciben o envían datos a través del uno o más circuitos de interfaz, y cuando el uno o más procesadores ejecutan instrucciones informáticas, se realiza el método de acuerdo con el segundo aspecto o cualquier posible implementación del segundo aspecto, o se realiza el método de acuerdo con el cuarto aspecto o cualquier posible implementación del cuarto aspecto.

El duodécimo aspecto y cualquier implementación del duodécimo aspecto corresponden respectivamente al segundo aspecto y cualquier implementación del segundo aspecto, o corresponden respectivamente al cuarto aspecto y cualquier implementación del cuarto aspecto. Para los efectos técnicos correspondientes al duodécimo aspecto y cualquier implementación del duodécimo aspecto, referirse a los efectos técnicos correspondientes al segundo aspecto y cualquier implementación del segundo aspecto, o referirse a los efectos técnicos correspondientes al cuarto aspecto y cualquier implementación del cuarto aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un decimotercer aspecto, una modalidad de esta solicitud proporciona un medio de almacenamiento legible por computador. El medio de almacenamiento legible por computador almacena un programa informático. Cuando el programa informático se ejecuta en un computador o un procesador, el computador o el procesador está habilitado para realizar el método de codificación de acuerdo con el primer aspecto o cualquier posible implementación del primer aspecto, o el computador o el procesador está habilitado para realizar el método de codificación de acuerdo con el tercer aspecto o cualquier posible implementación del tercer aspecto.

El decimotercer aspecto y cualquier implementación del decimotercer aspecto corresponden respectivamente al primer aspecto y cualquier implementación del primer aspecto, o corresponden respectivamente al tercer aspecto y cualquier implementación del tercer aspecto. Para los efectos técnicos correspondientes al decimotercer aspecto y cualquier implementación del decimotercer aspecto, referirse a los efectos técnicos correspondientes al primer aspecto y cualquier implementación del primer aspecto, o referirse a los efectos técnicos correspondientes al tercer aspecto y cualquier implementación del tercer aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un decimocuarto aspecto, una modalidad de esta solicitud proporciona un medio de almacenamiento legible por computador. El medio de almacenamiento legible por computador almacena un programa informático. Cuando el programa informático se ejecuta en un computador o un procesador, el computador o el procesador están habilitados para realizar el método de decodificación de acuerdo con el segundo aspecto o cualquier posible implementación del segundo aspecto, o el computador o el procesador están habilitados para realizar el método de decodificación de acuerdo con el cuarto aspecto o cualquier posible implementación del cuarto aspecto.

El decimocuarto aspecto y cualquier implementación del decimocuarto aspecto corresponden respectivamente al segundo aspecto y cualquier implementación del segundo aspecto, o corresponden respectivamente al cuarto aspecto y cualquier implementación del cuarto aspecto. Para los efectos técnicos correspondientes al decimocuarto aspecto y cualquier implementación del decimocuarto aspecto, consulte los efectos técnicos correspondientes al segundo aspecto y cualquier implementación del segundo aspecto, o consulte los efectos técnicos correspondientes al cuarto aspecto y cualquier implementación del cuarto aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un decimoquinto aspecto, una modalidad de esta solicitud proporciona

un producto de programa informático. El producto de programa informático incluye instrucciones informáticas. Cuando las instrucciones informáticas son ejecutadas por un computador o un procesador, el computador o el procesador está habilitado para realizar el método de codificación de acuerdo con el primer aspecto o cualquier posible implementación del primer aspecto, o el computador o el procesador está habilitado para realizar el método de codificación de acuerdo con el tercer aspecto o cualquier posible implementación del tercer aspecto.

El decimoquinto aspecto y cualquier implementación del decimoquinto aspecto corresponden respectivamente al primer aspecto y cualquier implementación del primer aspecto, o corresponden respectivamente al tercer aspecto y cualquier implementación del tercer aspecto. Para los efectos técnicos correspondientes al decimoquinto aspecto y cualquier implementación del decimoquinto aspecto, referirse a los efectos técnicos correspondientes al primer aspecto y cualquier implementación del primer aspecto, o referirse a los efectos técnicos correspondientes al tercer aspecto y cualquier implementación del tercer aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un decimosexto aspecto, una modalidad de esta solicitud proporciona un producto de programa informático. El producto de programa informático incluye instrucciones informáticas. Cuando las instrucciones informáticas son ejecutadas por un computador o un procesador, el computador o el procesador está habilitado para realizar el método de decodificación de acuerdo con el segundo aspecto o cualquier posible implementación del segundo aspecto, o el computador o el procesador está habilitado para realizar el método de decodificación de acuerdo con el cuarto aspecto o cualquier posible implementación del cuarto aspecto.

El decimosexto aspecto y cualquier implementación del decimosexto aspecto corresponden respectivamente al segundo aspecto y cualquier implementación del segundo aspecto, o corresponden respectivamente al cuarto aspecto y cualquier implementación del cuarto aspecto. Para los efectos técnicos correspondientes al decimosexto aspecto y cualquier implementación del decimosexto aspecto, referirse a los efectos técnicos correspondientes al segundo aspecto y cualquier implementación del segundo aspecto, o referirse a los efectos técnicos correspondientes al cuarto aspecto y cualquier implementación del cuarto aspecto. En el presente documento no se describen los detalles nuevamente.

De acuerdo con un decimoséptimo aspecto, una modalidad de esta solicitud además proporciona un dispositivo XR. El dispositivo XR incluye un procesador y un módulo de

reproducción de audio.

El procesador está configurado para realizar las etapas en el segundo aspecto y cualquier implementación del segundo aspecto.

5 El módulo de reproducción de audio está configurado para reproducir la señal de audio reproducida con realismo en el segundo aspecto y cualquier implementación del segundo aspecto.

10 El decimoséptimo aspecto y cualquier implementación del decimoséptimo aspecto corresponden respectivamente al segundo aspecto y cualquier implementación del segundo aspecto. Para los efectos técnicos correspondientes al decimoséptimo aspecto y cualquier implementación del decimoséptimo aspecto, consulte los efectos técnicos correspondientes al segundo aspecto y cualquier implementación del segundo aspecto. En el presente documento no se describen los detalles nuevamente.

15 De acuerdo con un decimoctavo aspecto, una modalidad de esta solicitud proporciona un dispositivo XR. El dispositivo XR incluye un procesador y un módulo de reproducción de audio.

El procesador está configurado para realizar las etapas en el cuarto aspecto y cualquier implementación del cuarto aspecto.

El módulo de reproducción de audio está configurado para reproducir la señal de audio reproducida con realismo en el cuarto aspecto y cualquier implementación del cuarto aspecto.

20 El decimoctavo aspecto y cualquier implementación del decimoctavo aspecto corresponden respectivamente al cuarto aspecto y cualquier implementación del cuarto aspecto. Para los efectos técnicos correspondientes al decimoctavo aspecto y cualquier implementación del decimoctavo aspecto, consulte los efectos técnicos correspondientes al cuarto aspecto y cualquier implementación del cuarto aspecto. En el presente documento no se describen los detalles nuevamente.

25

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La FIG. 1A es un diagrama de un escenario de solicitud de ejemplo;

la FIG. 1B es un diagrama de un escenario de solicitud de ejemplo;

la FIG. 1C es un diagrama de un escenario de solicitud de ejemplo;

la FIG. 1D es un diagrama de un escenario de solicitud de ejemplo;

la FIG. 2A es un diagrama de una estructura de compresión de ejemplo de una señal de audio de escena;

5 la FIG. 2B es un diagrama de una estructura de transcodificación de ejemplo de una señal de audio de escena;

la FIG. 2C es un diagrama de un proceso de reproducción realista de ejemplo;

la FIG. 2D es un diagrama de una escena de ejemplo;

la FIG. 2E es un diagrama de una escena de ejemplo;

10 la FIG. 3A es un diagrama de un proceso de codificación de ejemplo;

la FIG. 3B es un diagrama de una estructura de flujo de bits de ejemplo;

la FIG. 4A es un diagrama de un proceso de decodificación de ejemplo;

la FIG. 4B es un diagrama de un proceso de ejemplo para generar una señal de reproducción realista binaural;

15 la FIG. 5A es un diagrama de un proceso de codificación de ejemplo;

la FIG. 5B es un diagrama de una estructura de flujo de bits de ejemplo;

la FIG. 6A es un diagrama de un proceso de decodificación de ejemplo;

la FIG. 6B es un diagrama de un proceso de ejemplo para generar una señal de reproducción realista binaural;

20 la FIG. 7A es un diagrama de un proceso de codificación de ejemplo;

la FIG. 7B es un diagrama de una estructura de flujo de bits de ejemplo;

la FIG. 8 es un diagrama de un proceso de decodificación de ejemplo;

la FIG. 9A es un diagrama de un proceso de codificación de ejemplo;

la FIG. 9B es un diagrama de una estructura de flujo de bits de ejemplo;

la FIG. 10 es un diagrama de un proceso de decodificación de ejemplo;

la FIG. 11 es un diagrama de un proceso de codificación de ejemplo;

la FIG. 12 es un diagrama de un proceso de decodificación de ejemplo;

5 la FIG. 13 es un diagrama de un proceso de ejemplo para generar una señal reproducida con realismo binaural;

la FIG. 14 es un diagrama de un aparato de codificación de ejemplo;

la FIG. 15 es un diagrama de un aparato de decodificación de ejemplo;

la FIG. 16 es un diagrama de un aparato de codificación de ejemplo;

10 la FIG. 17 es un diagrama de un aparato de decodificación de ejemplo; y

la FIG. 18 es un diagrama de una estructura de ejemplo de un aparato.

DESCRIPCIÓN DETALLADA DE LA INVENCION

15 A continuación se describen de manera clara y completa las soluciones técnicas en las modalidades de la presente solicitud con referencia a los dibujos adjuntos en las modalidades de la presente solicitud. Es claro que las modalidades descritas son solo algunas, y no todas, las modalidades de esta solicitud. Todas las modalidades obtenidas por una persona normalmente versada en la técnica basadas en las modalidades de esta solicitud sin esfuerzos creativos caerán dentro del alcance de la protección de esta solicitud.

20 El término "y/o" en esta memoria descriptiva es simplemente una relación de asociación entre objetos asociados e indica que pueden existir tres relaciones. Por ejemplo, A y/o B pueden indicar los siguientes tres casos: Solo existe A, existen tanto A como B, y solo existe B.

25 Los términos "primero" y "segundo" en la memoria descriptiva y las reivindicaciones de las modalidades de esta solicitud se utilizan para distinguir entre diferentes objetos, pero no se utilizan para describir una secuencia específica de objetos. Por ejemplo, un primer objeto objetivo y un segundo objeto objetivo se utilizan para distinguir entre diferentes objetos objetivo, pero no se utilizan para describir una secuencia específica de los objetos objetivo.

En modalidades de esta solicitud, las palabras tales como "ejemplo" o "por ejemplo" se utilizan para representar ejemplos, ilustraciones o descripciones. Cualquier modalidad o solución de diseño descrita como "ejemplo" o "por ejemplo" en modalidades de esta solicitud no debe interpretarse como más preferida o ventajosa que otras modalidades o soluciones de diseño. Específicamente, las palabras como "ejemplo" o "por ejemplo" se utilizan para presentar conceptos relacionados de una manera específica.

En las descripciones de las modalidades de esta solicitud, a menos que se indique lo contrario, "una pluralidad de" significa dos o más. Por ejemplo, una pluralidad de unidades de procesamiento se refiere a dos o más unidades de procesamiento, y una pluralidad de sistemas se refiere a dos o más sistemas.

Para una descripción clara y breve de las siguientes modalidades, primero se proporcionan descripciones breves de las tecnologías relacionadas.

Un sonido (sound) es una onda continua generada por la vibración de un objeto. El objeto que vibra y emite la onda de sonido se conoce como fuente de sonido. En un proceso en el cual la onda sonora se propaga a través de un medio (por ejemplo, aire, sólido o líquido), un órgano auditivo de un ser humano o un animal puede sentir el sonido.

Las características de la onda de sonido incluyen un tono, intensidad y un timbre. El tono indica un nivel del sonido. La intensidad indica el volumen del sonido. La intensidad también se puede denominar intensidad acústica o volumen. Una unidad de la intensidad es el decibelio (decibel, dB). El timbre también se conoce como calidad de sonido.

Una frecuencia de la onda sonora determina el nivel del tono. Una frecuencia más alta indica un tono más alto. La cantidad de veces de vibración del objeto en 1 segundo se denomina frecuencia. Una unidad de la frecuencia es el hercio (hertz, Hz). La frecuencia de un sonido que puede ser reconocido por un oído humano está entre 20 Hz y 20000 Hz.

Una amplitud de la onda de sonido determina la intensidad. Una mayor amplitud indica una mayor intensidad. Una distancia más corta de la fuente de sonido indica una mayor intensidad.

Una forma de onda de la onda de sonido determina el timbre. Las formas de onda de las ondas de sonido incluyen una onda cuadrada, una onda en diente de sierra, una onda sinusoidal, una onda de pulso y similares.

Según las características de las ondas sonoras, los sonidos pueden clasificarse en un sonido regular y un sonido irregular. El sonido irregular es un sonido emitido por una fuente de sonido que vibra irregularmente. El sonido irregular es, por ejemplo, el ruido que afecta el trabajo, el estudio, el descanso y similares de las personas. El sonido regular es un sonido emitido por una fuente de sonido que vibra regularmente. Los sonidos regulares incluyen una voz y un sonido de música. Cuando un sonido se representa eléctricamente, el sonido regular es una señal analógica que cambia continuamente en el dominio del tiempo-frecuencia. La señal analógica puede denominarse señal de audio. La señal de audio es un soporte de información que porta una voz, música y efectos de sonido.

Debido a que el sentido auditivo humano tiene la capacidad de distinguir la distribución de la posición de una fuente de sonido en el espacio, al escuchar un sonido en el espacio, un oyente puede sentir una dirección y una posición del sonido además de un tono, intensidad y un timbre del sonido.

A medida que las personas prestan cada vez más atención a la experiencia del sistema auditivo y tienen requisitos cada vez más altos de calidad, surge una tecnología de audio tridimensional para mejorar la sensación de profundidad, la sensación de inmersión y la sensación de espacio de un sonido. De esta manera, un oyente no solo siente los sonidos producidos por las fuentes de sonido frontal, posterior, izquierda y derecha, sino que también siente que el espacio en el cual se encuentra el oyente está rodeado por un campo de sonido espacial ("campo de sonido" (sound field) para abreviar) producido por las fuentes de sonido, y siente que el sonido se propaga. Esto crea un efecto de sonido "inmersivo" que permite al oyente sentir que está en un lugar como una sala de cine o una sala de conciertos.

Es bien sabido que cuando una onda de sonido se propaga en un medio ideal, un número de onda es $k = \omega/c$, y una frecuencia angular es $\omega = 2\pi f$, donde f es una frecuencia de onda de sonido, y c es una velocidad de sonido. La presión acústica p cumple con la Fórmula (1) y ∇^2 es un operador de Laplace.

$$\nabla^2 p + k^2 p = 0 \quad (1)$$

Se supone que un sistema espacial fuera del oído humano es una esfera, y un oyente está en el centro de la esfera. Un sonido transmitido desde el exterior de la esfera tiene una proyección en una superficie esférica, y un sonido fuera de la superficie esférica se filtra. Se supone que una fuente de sonido se distribuye en la superficie esférica, y un campo de sonido

producido por la fuente de sonido en la superficie esférica se ajusta a un campo de sonido producido por una fuente de sonido original. Es decir, la tecnología de audio tridimensional es un método de ajuste de campo sonoro. Específicamente, la ecuación en la Fórmula (1) se resuelve en un sistema de coordenadas esféricas. En un área esférica pasiva, una solución a la ecuación en la Fórmula (1) es la siguiente Fórmula (2):

$$p(r, \theta, \varphi, k) = s \sum_{m=0}^{\infty} (2m+1) j^m j_m^{kr}(kr) \sum_{0 \leq n \leq m, \sigma=\pm 1} Y_{m,n}^{\sigma}(\theta_s, \varphi_s) Y_{m,n}^{\sigma}(\theta, \varphi) \quad (2)$$

r representa un radio de la esfera, θ representa información de ángulo horizontal (o denominada información de azimuth), φ representa información de ángulo de elevación (o denominada información de elevación), k representa un número de onda, s representa una amplitud de una onda plana ideal y m representa un número de serie de orden de una señal HOA (o se denomina un número de serie de orden de una señal HOA). $j^m j_m^{kr}(kr)$ representa una función de Bessel esférica, y la función de Bessel esférica también se conoce como función de base radial, donde la primera j representa una unidad imaginaria y $(2m+1) j^m j_m^{kr}(kr)$ no cambia con un ángulo. $Y_{m,n}^{\sigma}(\theta, \varphi)$ representa una función armónica esférica en una dirección θ, φ y $Y_{m,n}^{\sigma}(\theta_s, \varphi_s)$ representa una función armónica esférica en una dirección de una fuente de sonido. La señal HOA cumple con la Fórmula (3).

$$B_{m,n}^{\sigma} = s \cdot Y_{m,n}^{\sigma}(\theta_s, \varphi_s) \quad (3)$$

La Fórmula (3) se sustituye en la Fórmula (2), y la Fórmula (2) se puede transformar en la Fórmula (4).

$$p(r, \theta, \varphi, k) = \sum_{m=0}^{\infty} j^m j_m^{kr}(kr) \sum_{0 \leq n \leq m, \sigma=\pm 1} B_{m,n}^{\sigma} Y_{m,n}^{\sigma}(\theta, \varphi) \quad (4)$$

m se trunca a un $N^{\text{ésimo}}$ elemento, es decir, $m=N$, y $B_{m,n}^{\sigma}$ se usa como una descripción aproximada para un campo sonoro. En este caso, $B_{m,n}^{\sigma}$ puede denominarse coeficiente HOA (que puede representar una señal HOA de $N^{\text{ésimo}}$ orden). El campo sonoro es un área en la cual existe una onda sonora en un medio. N es un número entero mayor o igual a 1, y la señal HOA de $N^{\text{ésimo}}$ orden incluye una señal de audio de $(N+1)^2$ canales.

La Fórmula (4) indica que el campo sonoro puede expandirse en una superficie esférica con base en una función armónica esférica. En otras palabras, el campo sonoro puede descomponerse en la superposición de una pluralidad de ondas planas. Por lo tanto, el campo sonoro descrito por la señal HOA puede expresarse mediante el uso de la superposición de una pluralidad de ondas planas, y el campo sonoro se reconstruye mediante el uso del coeficiente HOA.

Una señal de audio de escena en modalidades de esta solicitud es un portador de información que porta información de posición espacial de una fuente de sonido en un campo de sonido, y describe un campo de sonido de un oyente en el espacio, o puede ser una señal para describir un campo de sonido de una fuente de sonido. La señal de audio de escena puede incluir señales HOA de una pluralidad de puntos (donde las señales HOA pueden incluir una señal HOA tridimensional y una señal HOA bidimensional (la cual también se puede denominar señal HOA plana)) y una señal de audio tridimensional. La señal de audio tridimensional puede ser una señal de audio distinta de la señal HOA en la señal de audio de escena. En esta solicitud, se utiliza para la descripción un ejemplo en el cual la señal de audio de escena son señales HOA de una pluralidad de puntos.

La FIG. 1A es un diagrama de un escenario de solicitud de ejemplo. La FIG. 1A muestra un escenario de solicitud en el cual un usuario utiliza un dispositivo de realidad aumentada (Augmented Reality, AR) para experimentar un concierto de AR. Debe entenderse que el usuario además puede usar el dispositivo de AR para explorar otra experiencia de AR, como una sala de conferencias de ciencias de AR o una sala de cine de AR. Esto no está limitado en esta solicitud.

En la FIG. 1A, una escena real en la cual se encuentra el usuario y una escena virtual correspondiente a la experiencia de AR explorada por el usuario son una misma escena, es decir, la escena real en la cual se encuentra el usuario es una sala de conciertos, y la escena virtual correspondiente a la experiencia de AR explorada por el usuario es una sala de conciertos virtual.

Con referencia a la FIG. 1A. Por ejemplo, cuando el usuario espera experimentar un concierto de AR en una sala de conciertos vacía, el usuario puede usar el dispositivo de AR; y luego, el usuario puede seleccionar una sala de conciertos de AR en el dispositivo de AR como una experiencia de AR que el usuario espera explorar, seleccionar un archivo de música que el usuario espera experimentar y realizar una operación de reproducción. Posteriormente, el dispositivo de AR puede procesar, por ejemplo, reproducir con realismo, el archivo de

música seleccionado por el usuario, para obtener una señal de audio reproducida con realismo (que puede ser, por ejemplo, una señal de audio espacial (es decir, una señal reproducida con realismo biaural), donde un ejemplo en el cual la señal de audio reproducida con realismo es una señal reproducida con realismo biaural se utiliza para la descripción en esta solicitud), y reproducir la señal reproducida con realismo biaural a través de un auricular. Adicionalmente, el dispositivo de AR además puede mostrar una imagen de reproducción en una pantalla, de modo que el usuario pueda ver, en la pantalla, que un jugador toca en una posición correspondiente en un escenario de la sala de conciertos. De esta manera, el usuario puede experimentar el concierto de una manera inmersiva a través de la combinación de visión y audición.

La FIG. 1B es un diagrama de un escenario de solicitud de ejemplo. La FIG. 1B muestra un escenario de solicitud en el cual un usuario utiliza un dispositivo de realidad virtual (Virtual Reality, VR) para experimentar una sala de cine de realidad virtual. Debe entenderse que el usuario puede usar el dispositivo de VR para explorar otras experiencias de VR, como una sala de conciertos de VR o una sala de conferencias de ciencias de VR. Esto no está limitado en esta solicitud.

En la FIG. 1B, una escena real en la cual se encuentra el usuario y una escena virtual correspondiente a la experiencia VR explorada por el usuario no son una misma escena, es decir, la escena real en la cual se encuentra el usuario es una sala de estar, y la escena virtual correspondiente a la experiencia VR explorada por el usuario es una sala de cine virtual.

Con referencia a la FIG. 1B. Por ejemplo, cuando el usuario espera experimentar una sala de cine de VR en la sala de estar, el usuario puede usar el dispositivo de VR; y luego, el usuario puede seleccionar una sala de cine de VR en el dispositivo de VR como experiencia de VR que el usuario espera explorar, seleccionar un archivo de película que el usuario espera experimentar y realizar una operación de reproducción. Posteriormente, el dispositivo de VR puede procesar, por ejemplo, reproducir con realismo, un archivo de audio incluido en el archivo de película, para obtener una señal reproducida con realismo biaural y reproducir la señal reproducida con realismo biaural a través de un auricular. Adicionalmente, el dispositivo de VR además puede mostrar una imagen de película en una pantalla. De esta manera, el usuario puede experimentar la película de una manera inmersiva a través de la combinación de visión y audición.

La FIG. 1C es un diagrama de un escenario de solicitud de ejemplo. La FIG. 1C muestra un escenario de solicitud en el cual un usuario usa (un teléfono móvil + un auricular)

para experimentar un concierto.

Debe entenderse que el usuario puede usar más (el teléfono móvil + los auriculares) para explorar otra experiencia auditiva, tal como una conversación cruzada o una película. Esto no está limitado en esta solicitud. Adicionalmente, el teléfono móvil es simplemente un ejemplo en esta solicitud. En esta solicitud, un dispositivo terminal como una tableta, un computador portátil, un computador personal o un reloj inteligente y un auricular se pueden usar alternativamente para escuchar.

En la FIG. 1C, una escena real en la cual se encuentra el usuario y una escena virtual correspondiente a la experiencia auditiva explorada por el usuario son una misma escena, es decir, la escena real en la cual se encuentra el usuario es una sala de conciertos, y la escena virtual correspondiente a la experiencia auditiva explorada por el usuario es una sala de conciertos virtual.

Con referencia a la FIG. 1C. Por ejemplo, cuando el usuario espera experimentar un concierto en una sala de conciertos vacía, el usuario puede usar los auriculares y conectarlos al teléfono móvil; y luego, el usuario puede seleccionar un concierto en el teléfono móvil como una experiencia auditiva que el usuario espera explorar, seleccionar un archivo de música que el usuario espera experimentar y realizar una operación de reproducción. Posteriormente, el teléfono móvil puede procesar, por ejemplo, reproducir con realismo, el archivo de música seleccionado por el usuario, para obtener una señal reproducida con realismo binaural; y luego, puede enviar la señal reproducida con realismo binaural al auricular para su reproducción. Adicionalmente, el teléfono móvil puede reproducir además un video de reproducción de concierto. De esta manera, el usuario puede experimentar el concierto en la sala de conciertos de una manera inmersiva utilizando el teléfono móvil.

La FIG. 1D es un diagrama de un escenario de solicitud de ejemplo. La FIG. 1D muestra un escenario de solicitud en el cual un usuario usa (un teléfono móvil + un auricular) para experimentar una conversación cruzada.

Debe entenderse que el usuario puede usar más (el teléfono móvil + los auriculares) para explorar otra experiencia auditiva, tal como un concierto o una película. Esto no está limitado en esta solicitud. Adicionalmente, el teléfono móvil es simplemente un ejemplo en esta solicitud. En esta solicitud, un dispositivo terminal como una tableta, un computador portátil, un computador personal o un reloj inteligente y un auricular se pueden usar alternativamente para escuchar.

En la FIG. 1D, una escena real en la cual se encuentra el usuario y una escena virtual correspondiente a la experiencia auditiva explorada por el usuario no son una misma escena, es decir, la escena real en la cual se encuentra el usuario es una sala de estar, y la escena virtual correspondiente a la experiencia auditiva explorada por el usuario es un estudio virtual.

5 Con referencia a la FIG. 1D. Por ejemplo, cuando el usuario espera experimentar una conversación cruzada en la sala de estar, el usuario puede usar los auriculares y conectarlos al teléfono móvil; y luego, el usuario puede seleccionar una conversación cruzada en el teléfono móvil como experiencia auditiva que el usuario espera explorar, seleccionar un
10 archivo de conversación cruzada que el usuario espera experimentar y realizar una operación de reproducción. Posteriormente, el teléfono móvil puede procesar, por ejemplo, reproducir con realismo, el archivo de conversación cruzada seleccionado por el usuario, para obtener una señal reproducida con realismo binaural; y luego, puede enviar la señal reproducida con realismo binaural al auricular para su reproducción. Además, el teléfono móvil puede reproducir
15 además un video de conversación cruzada. De esta manera, el usuario puede experimentar la conversación cruzada en la sala de estar de manera inmersiva utilizando el teléfono móvil.

Para facilitar la descripción, el dispositivo terminal en la FIG. 1A a la FIG. 1D se conoce como un primer dispositivo electrónico.

En la FIG. 1A a la FIG. 1D, un archivo de música de música, un archivo de audio incluido en un archivo de película y un archivo de conversación cruzada que se encuentran
20 en el primer dispositivo electrónico (el cual también se puede denominar lado del decodificador) se obtienen de un segundo dispositivo electrónico (el cual también se puede denominar lado del codificador), y son señales de audio de escena codificadas, es decir, flujos de bits. Además, el primer dispositivo electrónico necesita decodificar el flujo de bits para obtener una señal de audio reconstruida, y luego reproducir con realismo la señal de audio reconstruida para
25 obtener una señal reproducida con realismo binaural.

Por ejemplo, varias fuentes de sonido y micrófonos HOA en una pluralidad de puntos pueden desplegarse primero en una habitación (es decir, una escena real). Posteriormente, las fuentes de sonido reproducen señales de audio, y los micrófonos HOA en la pluralidad de puntos capturan sonidos en la sala para obtener señales HOA de la pluralidad de puntos. En
30 este caso, las señales HOA de la pluralidad de puntos pueden describir un campo de sonido de la fuente de sonido en la escena real. Hay una distancia igual entre dos micrófonos HOA cualesquiera en los micrófonos HOA en la pluralidad de puntos. De esta manera, el segundo dispositivo electrónico puede obtener las señales HOA de la pluralidad de puntos de los

micrófonos HOA en la pluralidad de puntos; y luego, codificar las señales HOA de la pluralidad de puntos para obtener un flujo de bits y enviar el flujo de bits al primer dispositivo electrónico.

Adicionalmente, para la escena en la FIG. 1A y la escena en la FIG. 1B, el segundo dispositivo electrónico además puede realizar un modelado basado en la escena real para
5 obtener una escena virtual. Luego, el segundo dispositivo electrónico puede codificar la escena virtual para obtener un flujo de bits y enviar el flujo de bits al primer dispositivo electrónico. Debe entenderse que un tercer dispositivo electrónico puede realizar alternativamente el modelado y enviar, al primer dispositivo electrónico o al segundo dispositivo electrónico, un flujo de bits obtenido mediante la codificación de la escena virtual.

10 La escena real es una escena objetivamente existente, por ejemplo, una sala de cine, una sala de conciertos o una sala de conferencias. La escena virtual es una escena construida mediante el uso de una tecnología virtual, y no es una escena objetivamente existente, por ejemplo, una sala de cine virtual, una sala de conciertos virtual o una sala de conferencias virtual.

15 La FIG. 2A es un diagrama de un ejemplo de estructura de compresión de una señal de audio de escena.

Con referencia a la FIG. 2A. Por ejemplo, un primer dispositivo electrónico puede incluir un primer módulo de captura de audio, un primer módulo de codificación de audio de escena, un primer módulo de codificación de canal, un primer módulo de decodificación de canal, un
20 primer módulo de decodificación de audio de escena, un primer módulo de representación y un primer módulo de reproducción de audio. Debe entenderse que el primer dispositivo electrónico puede incluir más o menos módulos que los mostrados en la FIG. 2A. Esto no está limitado en esta solicitud. El primer módulo de representación también puede denominarse primer reproductor con realismo, el primer módulo de codificación de audio de escena también
25 puede denominarse primer codificador de audio, y el primer módulo de decodificación de audio de escena también puede denominarse primer decodificador de audio.

Con referencia a la FIG. 2A. Por ejemplo, un segundo dispositivo electrónico puede incluir un segundo módulo de captura de audio, un segundo módulo de codificación de audio de escena, un segundo módulo de codificación de canal, un segundo módulo de
30 decodificación de canal, un segundo módulo de decodificación de audio de escena, un segundo módulo de representación y un segundo módulo de reproducción de audio. Debe entenderse que el segundo dispositivo electrónico puede incluir más o menos módulos que

los mostrados en la FIG. 2A. Esto no está limitado en esta solicitud. El segundo módulo de reproducción con realismo también puede denominarse segundo reproductor con realismo, el segundo módulo de codificación de audio de escena también puede denominarse segundo codificador de audio, y el segundo módulo de decodificación de audio de escena también puede denominarse segundo decodificador de audio.

Por ejemplo, un proceso en el cual el primer dispositivo electrónico codifica una señal de audio de escena y transmite la señal de audio de escena al segundo dispositivo electrónico y el segundo dispositivo electrónico realiza la decodificación y reproducción de audio puede ser el siguiente: El primer módulo de captura de audio (por ejemplo, un micrófono HOA conectado al primer dispositivo electrónico) puede capturar audio y emitir la señal de audio de escena al primer módulo de codificación de audio de escena. Posteriormente, el primer módulo de codificación de audio de escena puede codificar la señal de audio de escena y emitir un flujo de bits al primer módulo de codificación de canal. Luego, el primer módulo de codificación de canal puede realizar la codificación de canal en el flujo de bits y transmitir, al segundo dispositivo electrónico a través de un dispositivo de comunicación de red inalámbrico o cableado, un flujo de bits obtenido a través de la codificación de canal. Luego, el segundo módulo de decodificación de canal del segundo dispositivo electrónico puede realizar la decodificación de canal en los datos recibidos para obtener un flujo de bits y emitir el flujo de bits al segundo módulo de decodificación de audio de escena. Posteriormente, el segundo módulo de decodificación de audio de escena puede decodificar el flujo de bits para obtener una señal de audio reconstruida; y luego emitir la señal de audio reconstruida al segundo módulo de representación. El segundo módulo de reproducción con realismo reproduce con realismo la señal de escena reconstruida para obtener una señal reproducida con realismo binaural; y luego envía la señal reproducida con realismo binaural al segundo módulo de reproducción de audio. El segundo módulo de reproducción de audio realiza la reproducción de audio.

Debe entenderse que un proceso en el cual el segundo dispositivo electrónico codifica una señal de audio de escena y transmite la señal de audio de escena al primer dispositivo electrónico y el primer dispositivo electrónico realiza la decodificación y la reproducción de audio es similar al proceso anterior en el cual el primer dispositivo electrónico transmite la señal de audio de escena al segundo dispositivo electrónico y el segundo dispositivo electrónico realiza la reproducción de audio. En el presente documento no se describen los detalles nuevamente.

Por ejemplo, el primer dispositivo electrónico y el segundo dispositivo electrónico pueden incluir, entre otros, un computador personal, una estación de trabajo informática, un teléfono inteligente, una tableta, un servidor, una cámara inteligente, un vehículo inteligente, otro tipo de teléfono celular, un dispositivo de consumo de medios, un dispositivo portátil, un decodificador, una consola de juegos y similares.

Por ejemplo, esta solicitud se puede aplicar específicamente a una escena de VR/AR. De una manera posible, el primer dispositivo electrónico es un servidor, y el segundo dispositivo electrónico es un dispositivo de VR/AR. De una manera posible, el segundo dispositivo electrónico es un servidor, y el primer dispositivo electrónico es un dispositivo de VR/AR.

Cabe señalar que, cuando el primer dispositivo electrónico y el segundo dispositivo electrónico son dispositivos VR/AR, el primer módulo de reproducción de audio y el segundo módulo de reproducción de audio pueden ser auriculares integrados de los dispositivos VR/AR; o cuando el primer dispositivo electrónico y el segundo dispositivo electrónico son dispositivos terminales distintos de los dispositivos VR/AR, el primer módulo de reproducción de audio y el segundo módulo de reproducción de audio pueden ser auriculares conectados externamente a los otros dispositivos terminales.

Por ejemplo, cuando el primer dispositivo electrónico codifica una señal de audio de escena y el segundo dispositivo electrónico obtiene una señal de audio reconstruida a través de la decodificación, el primer dispositivo electrónico puede denominarse lado del codificador, y el segundo dispositivo electrónico puede denominarse lado del decodificador. Cuando el segundo dispositivo electrónico codifica una señal de audio de escena y el primer dispositivo electrónico obtiene una señal de audio reconstruida a través de la decodificación, el segundo dispositivo electrónico puede denominarse lado del codificador, y el primer dispositivo electrónico puede denominarse lado del decodificador.

La FIG. 2B es un diagrama de un ejemplo de estructura de transcodificación de una señal de audio de escena.

Con referencia a la FIG. 2B(1). Por ejemplo, un dispositivo inalámbrico o de red central puede incluir un módulo de decodificación de canal, otro módulo de decodificación de audio, un módulo de codificación de audio de escena y un módulo de codificación de canal. El dispositivo inalámbrico o de red central puede configurarse para realizar la transcodificación de audio.

Por ejemplo, un escenario de solicitud específico en la FIG. 2B(1) puede ser de la siguiente manera: Cuando un primer dispositivo electrónico no está provisto de un módulo de codificación de audio de escena y está provisto solo de otro módulo de codificación de audio, y un segundo dispositivo electrónico está provisto solo de un módulo de decodificación de audio de escena y no está provisto de otro módulo de decodificación de audio, el dispositivo inalámbrico o de red central puede utilizarse para la transcodificación, de modo que el segundo dispositivo electrónico puede decodificar y reproducir una señal de audio de escena codificada por el primer dispositivo electrónico utilizando el otro módulo de codificación de audio.

Específicamente, el primer dispositivo electrónico codifica la señal de audio de escena usando el otro módulo de codificación de audio para obtener un primer flujo de bits, y envía el primer flujo de bits al dispositivo inalámbrico o de red central después de realizar la codificación de canal en el primer flujo de bits. Posteriormente, el módulo de decodificación de canales del dispositivo inalámbrico o de red central puede realizar la decodificación de canales y emitir, al otro módulo de decodificación de audio, el primer flujo de bits obtenido a través de la decodificación de canales. Luego, el otro módulo de decodificación de audio decodifica el primer flujo de bits para obtener la señal de audio de escena y emite la señal de audio de escena al módulo de codificación de audio de escena. Luego, el módulo de codificación de audio de escena puede codificar la señal de audio de escena para obtener un segundo flujo de bits y emitir el segundo flujo de bits al módulo de codificación de canal. El módulo de codificación de canal envía el segundo flujo de bits al segundo dispositivo electrónico después de realizar la codificación de canal en el segundo flujo de bits. De esta manera, el segundo dispositivo electrónico puede invocar el módulo de decodificación de audio de escena para decodificar el segundo flujo de bits obtenido a través de la decodificación de canal, para obtener una señal de audio reconstruida; y posteriormente, puede reproducir la señal de audio reconstruida.

Con referencia a la FIG. 2B(2). Por ejemplo, un dispositivo inalámbrico o de red central puede incluir un módulo de decodificación de canal, un módulo de decodificación de audio de escena, otro módulo de codificación de audio y un módulo de codificación de canal. El dispositivo inalámbrico o de red central puede configurarse para realizar la transcodificación de audio.

Por ejemplo, un escenario de solicitud específico en la FIG. 2B(2) puede ser de la siguiente manera: Cuando un primer dispositivo electrónico está provisto solo de un módulo

de codificación de audio de escena y no está provisto de otro módulo de codificación de audio, y un segundo dispositivo electrónico no está provisto de un módulo de decodificación de audio de escena y está provisto solo de otro módulo de decodificación de audio, el dispositivo inalámbrico o de red central puede utilizarse para la transcodificación, de modo que el
5 segundo dispositivo electrónico puede decodificar y reproducir una señal de audio de escena codificada por el primer dispositivo electrónico utilizando el módulo de codificación de audio de escena.

Específicamente, el primer dispositivo electrónico codifica la señal de audio de escena utilizando el módulo de codificación de audio de escena para obtener un primer flujo de bits, y envía el primer flujo de bits al dispositivo inalámbrico o de red central después de realizar la
10 codificación de canal en el primer flujo de bits. Posteriormente, el módulo de decodificación de canales del dispositivo inalámbrico o de red central puede realizar la decodificación de canales y emitir, al módulo de decodificación de audio de escena, el primer flujo de bits obtenido a través de la decodificación de canales. A continuación, el módulo de decodificación de audio de escena decodifica el primer flujo de bits para obtener la señal de audio de escena y emite la señal de audio de escena al otro módulo de codificación de audio. Luego, el otro
15 módulo de codificación de audio puede codificar la señal de audio de escena para obtener un segundo flujo de bits y emitir el segundo flujo de bits al módulo de codificación de canal. El módulo de codificación de canal envía el segundo flujo de bits al segundo dispositivo electrónico después de realizar la codificación de canal en el segundo flujo de bits. De esta
20 manera, el segundo dispositivo electrónico puede invocar al otro módulo de decodificación de audio para decodificar el segundo flujo de bits obtenido a través de la decodificación de canal, para obtener una señal de audio reconstruida; y posteriormente, puede reproducir la señal de audio reconstruida.

Con base en la FIG. 2A y la FIG. 2B, se describe un proceso de obtención de una señal reproducida con realismo binaural a través de la reproducción con realismo.

La FIG. 2C es un diagrama de un proceso de representación de ejemplo. La FIG. 2C es un diagrama de un proceso de reproducción con realismo de un reproductor con realismo (que puede ser un primer reproductor con realismo o un segundo reproductor con realismo).

Con referencia a la FIG. 2C. Después de obtener una señal de audio reconstruida basada en un flujo de bits, un lado del decodificador puede ingresar la señal de audio reconstruida, los metadatos de una señal de audio de escena, una posición del usuario y una postura del usuario al reproductor con realismo. El reproductor con realismo reproduce con
30

realismo la señal de audio reconstruida con base en los metadatos de la señal de audio de la escena, la posición del usuario y la postura del usuario, para obtener una señal reproducida con realismo biaural. La postura del usuario puede ser una postura de la cabeza del usuario.

Por ejemplo, cuando el lado del decodificador es un dispositivo AR/VR, un sensor en el dispositivo AR/VR puede obtener la posición del usuario y la postura del usuario. Cuando el lado del decodificador es otro dispositivo terminal, un módulo de posicionamiento en el lado del decodificador puede obtener la posición del usuario, y un auricular conectado al lado del decodificador obtiene la postura del usuario.

Por ejemplo, el reproductor con realismo puede llevar a cabo la reproducción con realismo mediante el uso de un algoritmo de objeto de altavoz virtual (Virtual Loudspeaker Object, VLO), para obtener la señal reproducida con realismo biaural. Específicamente, la señal de audio reconstruida se puede convertir primero en una señal de altavoz virtual. Para más detalles, consulte la siguiente Fórmula (5):

$$S = DX \quad (5)$$

$$X = [x_1, \dots, x_L]^T, L = (N + 1)^2 \quad (6)$$

$$S = [s_1, \dots, s_L]^T, L = (N + 1)^2 \quad (7)$$

En la Fórmula (5), S es la señal del altavoz virtual, y una forma de expresión de la misma se muestra en la Fórmula (7), donde N es mayor que 0. En otras palabras, la señal de audio reconstruida se convierte en señales de altavoces virtuales de L altavoces virtuales. En la Fórmula (5), X se muestra la señal de audio reconstruida (es decir, una señal de audio reconstruida de una señal HOA), y una forma de expresión de la misma se muestra en la Fórmula (6). En la Fórmula (5), D es una matriz de decodificación cuyo tamaño es $L \times L$.

Posteriormente, la señal del altavoz virtual se puede reproducir con realismo con base en los metadatos de la señal de audio de la escena, la posición del usuario y la postura del usuario, para obtener la señal reproducida con realismo biaural. Los metadatos de la señal de audio de escena pueden incluir primeros metadatos y segundos metadatos. Para un proceso de generación de la señal reproducida con realismo biaural, consulte los siguientes S11 a S16.

S11: Determinar, a partir de las señales de altavoces virtuales de los altavoces virtuales L , las señales de altavoces virtuales Y $S_n \dots S_m$ que están más cerca de la posición

del usuario.

$$Y=m-n.$$

S12: Determinar una distancia $d1$ entre los micrófonos HOA en dos puntos con base en los segundos metadatos.

5 De una manera posible, los segundos metadatos son información de posición de micrófonos HOA en una pluralidad de puntos en una escena. En este caso, se puede calcular una distancia $d1$ entre micrófonos HOA en dos puntos cualesquiera con base en la información de posición de los micrófonos HOA en la pluralidad de puntos de la escena.

10 De una manera posible, los segundos metadatos son la distancia $d1$ entre los micrófonos HOA en dos puntos cualesquiera en los micrófonos HOA en la pluralidad de puntos.

La FIG. 2D es un diagrama de una escena de ejemplo.

15 Por ejemplo, la FIG. 2D muestra micrófonos HOA en dos puntos de la escena, y $Y=4$. En otras palabras, una señal HOA de un punto (es decir, una señal de audio de una fuente de sonido (guitar) recibida por un micrófono HOA en un punto) puede representarse mediante el uso de señales de altavoces virtuales de cuatro altavoces virtuales. La distancia $d1$ entre los micrófonos HOA en dos puntos se muestra en la FIG. 2D.

S13: Determinar un radio de altavoz virtual $r1$ con base en los primeros metadatos y la distancia $d1$ entre los micrófonos HOA en dos puntos.

20 Por ejemplo, el radio del altavoz virtual $r1$ puede mostrarse en la FIG. 6B, y una fórmula de cálculo del radio del altavoz virtual $r1$ se puede mostrar como la siguiente Fórmula (8):

$$r = kd \quad (8)$$

En la Fórmula (8), k representa los primeros metadatos. En otras palabras, cuando k no se cambia, un mayor d indica un mayor r .

25 La FIG. 2E es un diagrama de una escena de ejemplo. Para la FIG. 2E, consultar las descripciones de la FIG. 2D. En el presente documento no se describen los detalles nuevamente. Una diferencia entre la FIG. 2E y la FIG. 2D se encuentra en la distancia $d1$ entre los micrófonos HOA en dos puntos. La distancia $d1$ entre los micrófonos HOA en dos

puntos en la FIG. 2E es mayor que la distancia $d1$ entre los micrófonos HOA en dos puntos en la FIG. 2D. En consecuencia, el radio del altavoz virtual $r1$ es mayor en la FIG. 2E.

5 S14: Seleccionar, de una biblioteca de respuesta al impulso relacionado con el cabezal (Head-Related Impulse Response, HRIR) basada en la postura del usuario y el radio del altavoz virtual $r1$, las respuestas al impulso relacionadas con el cabezal $HRIRn...HRIRm$ correspondientes a las señales del altavoz virtual $S_n...S_m$.

S_n corresponde a $HRIRn$, S_{n+1} corresponde a $HRIR_{n+1}$, ..., y por analogía, S_m corresponde a $HRIRm$.

10 S15: Realizar el cálculo basado en $S_n...S_m$ y $HRIRn...HRIRm$ correspondiente para obtener una señal reproducida con realismo binaural de cada uno de los altavoces virtuales Y .

S16: Agregar señales reproducida con realismo binaurales de todos los altavoces virtuales Y para obtener la señal reproducida con realismo binaural.

15 Con referencia a la FIG. 2D o la FIG. 2E. Se añaden señales reproducida con realismo binaurales de dos altavoces virtuales para obtener una señal reproducida con realismo binaural en la posición del usuario.

20 En modalidades de esta solicitud, un lado del codificador puede obtener metadatos de una señal de audio de escena y escribir los metadatos en un flujo de bits. De esta manera, un lado del decodificador puede realizar un análisis sintáctico del flujo de bits para obtener los metadatos de la señal de audio de escena, y aplicar los metadatos de la señal de audio de escena a un proceso posterior de obtención de una señal reproducida con realismo binaural a través de la reproducción con realismo. Para un proceso de obtención de los metadatos de la señal de audio de escena y un proceso de codificación y decodificación, consulte las siguientes modalidades.

La FIG. 3A es un diagrama de un proceso de codificación de ejemplo.

25 S301: Obtenga una señal de audio de escena y metadatos de la señal de audio de escena.

Por ejemplo, un lado del codificador puede obtener señales HOA de una pluralidad de puntos de micrófonos HOA en la pluralidad de puntos como la señal de audio de escena.

Por ejemplo, en esta solicitud, los metadatos de la señal de audio de escena incluyen

información relacionada sobre la señal de audio de escena, y los metadatos de la señal de audio de escena se aplican a un proceso de obtención de una señal reproducida con realismo biaural a través de la reproducción con realismo.

Por ejemplo, de una manera posible de S301, un codificador puede obtener información de descripción de escena de una escena y luego determinar los metadatos de la señal de audio de escena con base en la información de descripción de escena de la escena. Posteriormente se describe un proceso específico.

Por ejemplo, la información de descripción de escena puede ser información para describir la escena y, por ejemplo, puede incluir, entre otros: identificadores, información de posición, características, direcciones y similares de micrófonos HOA en una pluralidad de puntos en una escena real; coordenadas, un material y similares de un objeto incluido en la escena real; un identificador, información de posición, una característica, una dirección y similares de una fuente de sonido en la escena real; una distancia crítica de la habitación en la escena real (la distancia crítica de la habitación significa una distancia desde un punto en el cual una densidad de energía de sonido directa es igual a una densidad de energía de sonido de reverberación hasta un centro de una fuente de sonido en una dirección específica cuando la fuente de sonido irradia de manera estable una onda de sonido en el interior (en la escena real), y también se denomina distancia de campo de difusión; y cuando la directividad de irradiación de la fuente de sonido es esférica, las distancias críticas en todas las direcciones son iguales, y también se denominan radio de reverberación en este caso); un parámetro de reverberación acústica (por ejemplo, una relación de energía directa a reverberación (Direct-to-Reverberation energy Ratio, DRR), tiempo de reverberación (Reverberation Time) (como T20 (tiempo de atenuación de energía de 20 dB), T30 o T60) y claridad (Clarity) (como C50 (una relación de energía sonora dentro de 50 ms a energía sonora después de 50 ms) o C80)); y similares. Esto no está limitado en esta solicitud.

Por ejemplo, de una manera posible de S301, se pueden obtener los metadatos que son de la señal de audio de escena y que son ingresados por un usuario, es decir, los metadatos de la señal de audio de escena pueden ser definidos por el usuario.

En la modalidad de la FIG. 3A, los metadatos obtenidos de la señal de audio de escena incluyen solo los primeros metadatos. Los primeros metadatos se pueden utilizar en combinación con una distancia entre dos de una pluralidad de micrófonos para determinar un radio de altavoz virtual, y los micrófonos pueden ser micrófonos HOA.

S302: Codifique la señal de audio de la escena y los metadatos de la señal de audio de la escena para obtener un flujo de bits.

Un proceso de codificación de la señal de audio de escena puede ser el siguiente:

5 En un primer esquema de codificación posible, la señal de audio de escena puede codificarse directamente para obtener datos codificados de la señal de audio de escena, y los datos codificados se escriben en el flujo de bits.

10 En un segundo esquema de codificación posible, un altavoz virtual objetivo puede seleccionarse primero de un conjunto de altavoces virtuales preestablecidos con base en una primera señal de audio de escena; posteriormente, se genera una primera señal de altavoz virtual con base en la primera señal de audio de escena e información de atributos del altavoz virtual objetivo; luego, se determina una segunda señal de audio de escena utilizando la información de atributos del altavoz virtual objetivo y la primera señal de altavoz virtual; posteriormente, se genera una señal residual con base en la primera señal de audio de escena y la segunda señal de audio de escena; y luego, la primera señal de altavoz virtual y la señal residual se codifican para obtener datos codificados de la señal de audio de escena, y los datos codificados se escriben en el flujo de bits. La primera señal de audio de escena es una señal de audio de escena original que se va a codificar.

20 En un tercer esquema de codificación posible, la información de atributos de un altavoz virtual objetivo puede determinarse con base en la señal de audio de escena, donde la señal de audio de escena incluye una señal de audio de L canales, y L es un número entero positivo; y una primera señal de audio en la señal de audio de escena y la información de atributos del altavoz virtual objetivo se codifican para obtener datos codificados de la señal de audio de escena, y los datos codificados se escriben en el flujo de bits, donde la primera señal de audio es una señal de audio de G canales en la señal de audio de escena, y G es un número entero positivo menor o igual a L.

30 Por ejemplo, en la modalidad en la FIG. 3A, en esta solicitud, un elemento de sintaxis de los metadatos de la señal de audio de escena y la información de descripción correspondiente pueden agregarse por adelantado a una estructura de sintaxis original (o denominarse tabla de elementos de sintaxis de alto nivel) del flujo de bits. De esta manera, el lado del codificador puede escribir los metadatos de la señal de audio de escena en el flujo de bits con base en el elemento de sintaxis de los metadatos de la señal de audio de escena y la información de descripción correspondiente en la estructura de sintaxis del flujo de bits.

Por ejemplo, la estructura sintáctica del flujo de bits se puede mostrar en la siguiente Tabla 1.

Tabla 1

Sintaxis	N.º de bits	Mnemónico
Syntax element 1		
{		
Syntax element 2	1	bslbf
Syntax element 3	3	bslbf
...		
Syntax element n	4	uimsbf
}		
...		
Syntax element n+h		
distanceFactor	8	uimsbf

- 5 Ver la Tabla 1. En la Tabla 1, la sintaxis representa un elemento de sintaxis, el número de bits representa una cantidad de bits de datos identificados por el elemento de sintaxis y el mnemónico representa una observación o una explicación del elemento de sintaxis.

Ver la Tabla 1. Por ejemplo, la estructura sintáctica del flujo de bits incluye una pluralidad de elementos sintácticos: un elemento sintáctico 1, un elemento sintáctico 2, un elemento sintáctico 3, ..., un elemento sintáctico n, ..., un elemento sintáctico n+h y un elemento sintáctico recién agregado "distanceFactor" en esta solicitud, donde
5 "distanceFactor" es el elemento sintáctico de los metadatos de la señal de audio de escena, y n y h son enteros positivos.

Refiérase aún a la Tabla 1. "1" en la cuarta fila en la Tabla 1 indica que los datos identificados por el elemento de sintaxis 2 en el flujo de bits se representan mediante el uso de 1 bit, y "bslbf" en la cuarta fila en la Tabla 1 indica que los datos identificados por el
10 elemento de sintaxis 2 en el flujo de bits es una cadena de bits, bit izquierdo primero.

"3" en la quinta fila en la Tabla 1 indica que los datos identificados por el elemento de sintaxis 3 en el flujo de bits se representan mediante el uso de 3 bits, y "bslbf" en la quinta fila en la Tabla 1 indica que los datos identificados por el elemento de sintaxis 3 en el flujo de bits es una cadena de bits, bit izquierdo primero.

"4" en la séptima fila en la Tabla 1 indica que los datos identificados por el elemento de sintaxis n en el flujo de bits se representan mediante el uso de 4 bits, y "uimsbf" en la séptima fila en la Tabla 1 indica que los datos identificados por el elemento de sintaxis n en el flujo de bits es un entero sin signo, el bit más significativo primero.
15

"8" en la última fila de la Tabla 1 indica que los metadatos de la señal de audio de escena en el flujo de bits se pueden representar usando 8 bits, y "uimsbf" en la última fila de la Tabla 1 indica que los metadatos de la señal de audio de escena en el flujo de bits es un entero sin signo, el bit más significativo primero. Debe entenderse que la Tabla 1 es meramente un ejemplo de esta solicitud. Los metadatos de la señal de audio de escena pueden representarse alternativamente mediante el uso de otro número de bits, por ejemplo,
20 16 bits. Esto no está limitado en esta solicitud. Además, adicionalmente a representarse mediante el uso de "distanceFactor", el elemento de sintaxis de los metadatos de la señal de audio de escena puede representarse alternativamente mediante el uso de otra cadena de caracteres. Esto no está limitado en esta solicitud.
25

Por ejemplo, si la estructura sintáctica del flujo de bits se muestra en la Tabla 1, en un proceso de realización de S302, los metadatos de la señal de audio de escena se pueden convertir en un número binario de 8 bits; y luego el número binario de 8 bits se escribe después de los datos identificados por el elemento sintáctico n+h en el flujo de bits.
30

Por ejemplo, S301 y S302 pueden ser realizados por un codificador.

La FIG. 3B es un diagrama de una estructura de flujo de bits de ejemplo. Una estructura de flujo de bits del flujo de bits en la modalidad de la FIG. 3A se puede mostrar en la FIG. 3B, es decir, los metadatos de la señal de audio de escena son posteriores a los datos codificados de la señal de audio de escena.

Debe entenderse que, en el flujo de bits en esta solicitud, los metadatos de la señal de audio de escena pueden estar alternativamente antes de los datos codificados de la señal de audio de escena. Esto no está limitado en esta solicitud. Adicionalmente, el flujo de bits en esta solicitud además puede incluir otros datos. Esto no está limitado en esta solicitud.

Cabe señalar que, en el flujo de bits en esta solicitud, los metadatos de la señal de audio de escena pueden ser adyacentes o no a los datos codificados de la señal de audio de escena. Esto no está limitado en esta solicitud.

La FIG. 4A es un diagrama de un proceso de decodificación de ejemplo. El proceso de decodificación en la FIG. 4A corresponde al proceso de codificación en la FIG. 3A.

S401: Recibir un flujo de bits.

S402: Análisis sintáctico del flujo de bits para obtener metadatos de una señal de audio de escena.

S403: Obtener una señal de audio reconstruida de la señal de audio de la escena con base en el flujo de bits.

S401 y S402 se realizan mediante un decodificador. Una secuencia de realización de S402 y S403 no está limitada en esta solicitud.

Por ejemplo, un proceso de obtención de la señal de audio reconstruida de la señal de audio de escena con base en el flujo de bits puede ser el siguiente:

En un primer esquema de decodificación (correspondiente al primer esquema de codificación anterior), la señal de audio reconstruida de la señal de audio de escena puede decodificarse directamente a partir del flujo de bits. En este caso, el decodificador puede realizar S403.

En un segundo esquema de decodificación (correspondiente al segundo esquema de

codificación anterior), una señal de altavoz virtual y una señal residual pueden decodificarse a partir del flujo de bits; y luego, la señal de audio reconstruida de la señal de audio de escena se obtiene con base en la información de atributos de un altavoz virtual objetivo, la señal residual y la señal de altavoz virtual. En este caso, el decodificador puede realizar una
5 operación de decodificación, y otro módulo puede realizar otra etapa.

En un tercer esquema de decodificación (correspondiente al tercer esquema de codificación anterior), el flujo de bits puede decodificarse para obtener una primera señal reconstruida e información de atributos de un altavoz virtual objetivo; posteriormente, se genera una señal de altavoz virtual correspondiente al altavoz virtual objetivo con base en la
10 información de atributos y la primera señal reconstruida; y luego, se realiza la reconstrucción con base en la información de atributos y la señal de altavoz virtual para obtener la señal de audio reconstruida de la señal de audio de escena. En este caso, el decodificador puede realizar una operación de decodificación, y otro módulo puede realizar otra etapa.

Por ejemplo, un lado del decodificador puede realizar un análisis sintáctico del flujo de
15 bits para obtener los metadatos de la señal de audio de escena con base en un elemento sintáctico de los metadatos de la señal de audio de escena y la información de descripción correspondiente en una estructura sintáctica del flujo de bits. En la modalidad de la FIG. 4A, los metadatos de la señal de audio de escena pueden incluir solo los primeros metadatos. Luego, el lado del decodificador puede llevar a cabo la reproducción con realismo con base
20 en los primeros metadatos obtenidos del flujo de bits a través del análisis sintáctico y la señal de audio reconstruida obtenida con base en el flujo de bits, para obtener una señal reproducida con realismo binaural. En este caso, el lado del decodificador puede obtener segundos metadatos de otro dispositivo o generar localmente segundos metadatos. Esto no está limitado en esta solicitud. Además, una cantidad Y de altavoces virtuales en un proceso de
25 reproducción con realismo puede ser un valor definido por el usuario, o puede ser un valor predeterminado. Esto no está limitado en esta solicitud.

La FIG. 4B es un diagrama de un proceso de ejemplo para generar una señal reproducida con realismo binaural. La FIG. 4B se muestra con base en la FIG. 3A y la FIG. 4A.

Con referencia a la FIG. 4B. Por ejemplo, después de codificar una señal de audio de
30 escena y metadatos de la señal de audio de escena para obtener un flujo de bits, un lado del codificador transmite el flujo de bits a un lado del decodificador. Después de recibir el flujo de bits, el lado del decodificador puede obtener una señal de audio reconstruida de la señal de audio de escena con base en el flujo de bits, y realizar un análisis sintáctico del flujo de bits

para obtener los metadatos de la señal de audio de escena; y luego reproducir con realismo la señal de audio reconstruida de la señal de audio de escena con base en los metadatos de la señal de audio de escena, una posición del usuario y una postura del usuario, para obtener una señal reproducida con realismo binaural.

5 Por ejemplo, en esta solicitud, un primer identificador que indica la complejidad de la representación puede escribirse adicionalmente en el flujo de bits. La complejidad de reproducción con realismo se utiliza para determinar una cantidad Y de altavoces virtuales en un proceso de reproducción con realismo. En otras palabras, el lado del decodificador puede reproducir con realismo una señal de altavoz virtual con base en los metadatos de la señal de
10 audio de escena, la complejidad de reproducción con realismo indicada por el primer identificador, la posición del usuario y la postura del usuario, para obtener una señal reproducida con realismo binaural. De esta manera, se puede obtener una señal reproducida con realismo binaural cuya calidad de audio coincida con factores tales como un requisito del usuario y una capacidad informática de un dispositivo electrónico a través de la reproducción
15 con realismo.

La FIG. 5A es un diagrama de un proceso de codificación de ejemplo. En la modalidad de la FIG. 5A, los metadatos de una señal de audio de escena se determinan con base en la información de descripción de escena, y los metadatos de la señal de audio de escena pueden incluir primeros metadatos y segundos metadatos.

20 S501: Obtener una señal de audio de escena e información de descripción de escena de una escena.

Por ejemplo, un lado del codificador puede obtener localmente la información de descripción de escena de la escena, o puede obtener la información de descripción de escena de la escena de un tercer dispositivo electrónico. Esto no está limitado en esta solicitud.

25 S502: Determinar los metadatos de la señal de audio de la escena con base en la información de descripción de la escena.

Por ejemplo, los metadatos de la señal de audio de escena pueden incluir primeros metadatos y segundos metadatos.

De una manera posible, los primeros metadatos pueden determinarse con base en un
30 producto de una distancia crítica de la sala y un valor preestablecido. Para más detalles, consulte la siguiente Fórmula (9):

$$k = \frac{1.5}{\text{Distancia crítica de la habitación}} \quad (9)$$

k en la Fórmula (9) son los primeros metadatos, y $k \in [0.5, 1.5]$. 1.5 en la Fórmula (9) es un valor preestablecido.

De una manera posible, se puede determinar un tipo de reverberación de la escena con base en un parámetro de reverberación acústica en la escena, y luego se determinan los primeros metadatos con base en el tipo de reverberación de la escena. Específicamente, los primeros metadatos correspondientes pueden estar preestablecidos para diferentes tipos de reverberación. Por ejemplo, para un primer tipo de reverberación (correspondiente a una escena de sonido seco o una habitación no reverberante), los primeros metadatos correspondientes se establecen en 0.5, y para un segundo tipo de reverberación (correspondiente a una escena de sonido húmedo o una habitación reverberante), los primeros metadatos correspondientes se establecen en 1.5. De esta manera, cuando se determina que el tipo de reverberación de la escena es el primer tipo de reverberación, se puede usar 0.5 como los primeros metadatos; o cuando se determina que el tipo de reverberación de la escena es el segundo tipo de reverberación, se puede usar 1.5 como los primeros metadatos.

De una manera posible, la información de posición de los micrófonos HOA en una pluralidad de puntos en la escena incluida en la información de descripción de escena se puede utilizar como los segundos metadatos.

De una manera posible, se puede determinar una distancia entre los micrófonos HOA en dos puntos en los micrófonos HOA en la pluralidad de puntos con base en la información de posición de los micrófonos HOA en la pluralidad de puntos en la escena incluida en la información de descripción de escena; y luego la distancia entre los micrófonos HOA en los dos puntos en los micrófonos HOA en la pluralidad de puntos se utiliza como los segundos metadatos.

Por ejemplo, cuando los primeros metadatos son ingresados por un usuario, un rango de valores de los primeros metadatos puede limitarse a $k \in [0.25, 4]$.

Debe entenderse que los primeros metadatos además pueden tener otra manera de generación, y los metadatos de la señal de audio de escena además pueden incluir otros metadatos. Esto no está limitado en esta solicitud.

S503: Codifique la señal de audio de la escena y los metadatos de la señal de audio de la escena para obtener un flujo de bits.

Por ejemplo, para S503, consulte las descripciones anteriores de S302. En el presente documento no se describen los detalles nuevamente.

5 S504: Escribir un primer identificador en el flujo de bits.

Por ejemplo, se puede preestablecer una pluralidad de niveles de complejidad de representación, como se muestra en la Tabla 2.

Tabla 2

Nivel de complejidad de reproducción con realismo	Número de altavoces virtuales	Número de circunvoluciones	Complejidad de reproducción con realismo	Calidad de audio
0	$(N + 1)^2$	$2(N+1)^2$	Dependiendo de una señal HOA	Dependiendo de una señal HOA
1	4	8	Más bajo	Más bajo
2	9	18	Bajo	Bajo
3	16	32	Alto	Alto
4	25	50	Más alto	Más alto

10 El número de altavoces virtuales en la Tabla 2 es Y.

Ver la Tabla 2. Por ejemplo, la complejidad de la representación se puede clasificar en

cinco niveles. En un nivel 1 a un nivel 4, un mayor número de altavoces virtuales indica un mayor número de convoluciones en un proceso de reproducción con realismo, una mayor complejidad de reproducción con realismo y una mayor calidad de audio.

Ver la Tabla 2. Por ejemplo, para un nivel 0, N es un orden de una señal HOA, un
5 número de altavoces virtuales es $(N + 1)^2$, un número de convoluciones es $2(N + 1)^2$, y tanto la complejidad de representación como la calidad de audio dependen de la señal HOA.

Debe entenderse que la Tabla 2 es simplemente un ejemplo en esta solicitud, y los niveles de complejidad de representación pueden establecerse de acuerdo con un requisito. Esto no está limitado en esta solicitud.

10 De una manera posible, un nivel de complejidad de representación objetivo se puede seleccionar de la pluralidad de niveles de complejidad de representación en la Tabla 2 con base en una capacidad informática de un lado del decodificador; y un valor del primer identificador se establece en el nivel de complejidad de representación objetivo.

15 De una manera posible, el usuario puede ingresar un nivel de complejidad de representación objetivo basado en la pluralidad de niveles de complejidad de representación en la Tabla 2, y un valor del primer identificador se establece en el nivel de complejidad de representación objetivo que ingresa el usuario.

20 Por ejemplo, en la modalidad en la FIG. 5A, en esta solicitud, un elemento de sintaxis de los metadatos de la señal de audio de escena y la información de descripción correspondiente, y un elemento de sintaxis del primer identificador y la información de descripción correspondiente se pueden agregar por adelantado a una estructura de sintaxis original del flujo de bits. De esta manera, el lado del codificador puede escribir el primer identificador en el flujo de bits con base en el elemento sintáctico del primer identificador y la información de descripción correspondiente en la estructura sintáctica del flujo de bits; y
25 escribir los metadatos de la señal de audio de escena en el flujo de bits con base en el elemento sintáctico de los metadatos de la señal de audio de escena y la información de descripción correspondiente en la estructura sintáctica del flujo de bits.

Cabe señalar que una secuencia de realización de S503 y S504 no está limitada en esta solicitud.

30 Por ejemplo, la estructura sintáctica del flujo de bits se puede mostrar en la siguiente

Tabla 3.

Tabla 3

Sintaxis	N.º de bits	Mnemónico
Syntax element 1		
{		
Syntax element 2	1	bslbf
Syntax element 3	3	bslbf
...		
Syntax element n	4	uimsbf
}		
...		
Syntax element n+1		
distanceFactor	8	uimsbf
complexityLevel	3	uimsbf

Ver la Tabla 3. En la modalidad de la FIG. 5A, los elementos de sintaxis "distanceFactor" y "complexityLevel" se agregan nuevamente en esta solicitud, donde

"distanceFactor" es el elemento de sintaxis de los metadatos de la señal de audio de la escena, y "complexityLevel" es el elemento de sintaxis del primer identificador.

"3" en la última fila en la Tabla 3 indica que el primer identificador en el flujo de bits se puede representar mediante el uso de 3 bits, y "uimsbf" en la última fila en la Tabla 3 indica que el primer identificador en el flujo de bits es un entero sin signo, el bit más significativo primero. Debe entenderse que la Tabla 3 es meramente un ejemplo de esta solicitud. El primer identificador puede representarse alternativamente mediante el uso de otro número de bits, por ejemplo, 4 bits. Esto no está limitado en esta solicitud. Además, además de representarse mediante el uso de "complexityLevel", el elemento de sintaxis del primer identificador puede representarse alternativamente mediante el uso de otra cadena de caracteres. Esto no está limitado en esta solicitud.

Para obtener explicaciones de otros elementos de sintaxis en la Tabla 3, consulte las descripciones en la Tabla 1. En el presente documento no se describen los detalles nuevamente.

Por ejemplo, si la estructura sintáctica del flujo de bits se muestra en la Tabla 3, en un proceso de realización de S503, los metadatos de la señal de audio de escena se pueden convertir en un número binario de 8 bits; y luego el número binario de 8 bits se escribe después de los datos identificados por el elemento sintáctico n+1 en el flujo de bits. Adicionalmente, en un proceso de realización de S504, el primer identificador se puede convertir en un número binario de 3 bits; y luego el número binario de 3 bits se escribe después de los metadatos de la señal de audio de escena en el flujo de bits.

La FIG. 5B es un diagrama de una estructura de flujo de bits de ejemplo. Una estructura de flujo de bits en la modalidad de la FIG. 5A se puede mostrar en la FIG. 5B, es decir, los metadatos de la señal de audio de escena son después de los datos codificados de la señal de audio de escena, y el primer identificador es después de los metadatos de la señal de audio de escena.

Debe entenderse que las posiciones de los metadatos de la señal de audio de escena, los datos codificados de la señal de audio de escena y el primer identificador en el flujo de bits no están limitados en esta solicitud.

Debe entenderse que, en el flujo de bits en esta solicitud, dos cualesquiera de los metadatos de la señal de audio de escena, los datos codificados de la señal de audio de

escena y el primer identificador pueden ser adyacentes o pueden no ser adyacentes. Esto no está limitado en esta solicitud.

La FIG. 6A es un diagrama de un proceso de decodificación de ejemplo. El proceso de decodificación en la FIG. 6A corresponde al proceso de codificación en la FIG. 5A.

5 S601: Recibir un flujo de bits.

S602: Obtener una señal de audio reconstruida de una señal de audio de escena basada en el flujo de bits.

S603: Análisis sintáctico del flujo de bits para obtener metadatos de la señal de audio de la escena.

10 S604: Análisis sintáctico del flujo de bits para obtener un primer identificador.

Por ejemplo, una secuencia de realización de S602, S603 y S604 no está limitada en esta solicitud.

Por ejemplo, después de recibir el flujo de bits, un lado del decodificador puede obtener la señal de audio reconstruida de la señal de audio de escena con base en el flujo de bits.

15 Para más detalles, consulte las descripciones en la modalidad en la FIG. 4A. En el presente documento no se describen los detalles nuevamente.

Por ejemplo, el lado del decodificador puede realizar un análisis sintáctico adicionalmente el flujo de bits para obtener los metadatos de la señal de audio de escena con base en un elemento sintáctico de los metadatos de la señal de audio de escena y la información de descripción correspondiente en una estructura sintáctica del flujo de bits. Adicionalmente, el lado del decodificador puede realizar un análisis sintáctico adicionalmente el flujo de bits para obtener el primer identificador con base en un elemento sintáctico del primer identificador y la información de descripción correspondiente en la estructura sintáctica del flujo de bits.

25 Posteriormente, el lado del decodificador puede llevar a cabo la reproducción con realismo con base en la señal reconstruida de la señal de audio de escena, los primeros metadatos, los segundos metadatos y el primer identificador, para obtener una señal reproducida con realismo binaural. En este caso, el lado del decodificador puede realizar un análisis sintáctico del flujo de bits para obtener los segundos metadatos y los primeros metadatos. Adicionalmente, en un proceso de representación, se puede determinar una

30

cantidad Y de altavoces virtuales seleccionados con base en el primer identificador.

La FIG. 6B es un diagrama de un proceso de ejemplo para generar una señal reproducida con realismo biaural. La FIG. 6B se muestra con base en la FIG. 5A y la FIG. 6A.

Con referencia a la FIG. 6B. Por ejemplo, después de codificar una señal de audio de
5 escena, metadatos de la señal de audio de escena y un primer identificador en un flujo de bits,
un lado del codificador transmite el flujo de bits a un lado del decodificador. Después de recibir
el flujo de bits, el lado del decodificador puede obtener una señal de audio reconstruida de la
señal de audio de escena con base en el flujo de bits, realizar un análisis sintáctico del flujo
de bits para obtener los metadatos de la señal de audio de escena y el primer identificador, y
10 luego reproducir con realismo la señal de audio reconstruida de la señal de audio de escena
con base en los metadatos de la señal de audio de escena, un nivel de complejidad de
reproducción con realismo indicado por el primer identificador, una posición del usuario y una
postura del usuario, para obtener una señal reproducida con realismo biaural.

Cabe destacar que en la FIG. 6B, el primer identificador, el nivel de complejidad de
15 reproducción con realismo indicado por el primer identificador o una cantidad Y de altavoces
virtuales correspondientes al nivel de complejidad de reproducción con realismo indicado por
el primer identificador se pueden ingresar en un reproductor con realismo. Esto no está
limitado en esta solicitud.

Por ejemplo, puede haber una pluralidad de reproductores con realismo (cada
20 reproductor con realismo utiliza un método de reproducción con realismo diferente) en un lado
del decodificador. La pluralidad de reproductores con realismo incluye un reproductor con
realismo que lleva a cabo la reproducción con realismo con base en los metadatos de una
señal de audio de escena (por ejemplo, un reproductor con realismo que lleva a cabo la
reproducción con realismo utilizando el método de S11 a S16) y un reproductor con realismo
25 que lleva a cabo la reproducción con realismo con base en otra información. Además, se
puede establecer un valor de identificador correspondiente para cada reproductor con
realismo. Por ejemplo, para un reproductor con realismo A, un valor de identificador
correspondiente se establece en "0"; para un reproductor con realismo B, un valor de
identificador correspondiente se establece en "1"; y para un reproductor con realismo C, un
30 valor de identificador correspondiente se establece en "2". Cuando se determina que el lado
del decodificador necesita llevar a cabo la reproducción con realismo con base en los
metadatos de la señal de audio de escena, de una manera posible, un valor de identificador
(el cual posteriormente se denomina primer valor de identificador) del reproductor con

realismo que lleva a cabo la reproducción con realismo con base en los metadatos de la señal de audio de escena puede usarse como un valor de identificador de un segundo identificador; o de una manera posible, un segundo valor de identificador (por ejemplo, "0" o "1", el cual puede establecerse de acuerdo con un requisito) puede usarse como un valor de identificador de un segundo identificador; y luego el segundo identificador se agrega a un flujo de bits. El segundo valor de identificador indica que el lado del decodificador realizará la representación con base en los metadatos de la señal de audio de la escena. En otras palabras, el segundo valor identificador indica que el flujo de bits incluye los metadatos de la señal de audio de escena. De esta manera, es conveniente que el lado del decodificador determine qué reproductor con realismo o qué algoritmo de reproducción con realismo se utilizará para la reproducción con realismo.

La FIG. 7A es un diagrama de un proceso de codificación de ejemplo. Cabe destacar que la FIG. 7A se muestra con base en la FIG. 3A. En la modalidad de la FIG. 7A, los metadatos de una señal de audio de escena se determinan con base en la información de descripción de escena.

S701: Obtener una señal de audio de escena.

S702: Codificar la señal de audio de la escena para obtener un flujo de bits.

Por ejemplo, para S701 y S702, consulte las descripciones anteriores. En el presente documento no se describen los detalles nuevamente.

S703: Escribir un segundo identificador en el flujo de bits.

Tanto el primer valor de identificador como el segundo valor de identificador descritos anteriormente pueden denominarse primer valor de identificador preestablecido. En otras palabras, cuando se determina que un lado del decodificador necesita usar metadatos de la señal de audio de escena para la representación, un valor del segundo identificador se puede establecer en el primer valor de identificador preestablecido, y luego el primer valor de identificador preestablecido se escribe en el flujo de bits. Cuando se determina que un lado del decodificador no necesita usar metadatos de la señal de audio de escena para la representación, un valor del segundo identificador se puede establecer en otro valor de identificador (el otro valor de identificador es diferente del primer valor de identificador preestablecido), y luego el otro valor de identificador preestablecido se escribe en el flujo de bits.

S704: Obtener información de descripción de escena de una escena.

S705: Determine los metadatos de la señal de audio de la escena con base en la información de descripción de la escena.

S706: Escribir los metadatos de la señal de audio de la escena en el flujo de bits.

5 Por ejemplo, para S704 a S706, consulte las descripciones anteriores de S501 a S503. En el presente documento no se describen los detalles nuevamente.

10 Por ejemplo, en la modalidad en la FIG. 7A, en esta solicitud, un elemento de sintaxis de los metadatos de la señal de audio de escena y la información de descripción correspondiente, y un elemento de sintaxis del segundo identificador y la información de descripción correspondiente se pueden agregar por adelantado a una estructura de sintaxis original del flujo de bits. De esta manera, un lado del codificador puede escribir el segundo identificador en el flujo de bits con base en el elemento de sintaxis del segundo identificador y la información de descripción correspondiente en la estructura de sintaxis del flujo de bits.

15 Por ejemplo, la estructura sintáctica del flujo de bits se puede mostrar en la siguiente Tabla 4.

Tabla 4

Sintaxis	N.º de bits	Mnemónico
Syntax element 1		
{		
Syntax element 2	1	bslbf
Syntax element 3	3	bslbf
...		

Sintaxis	N.º de bits	Mnemónico
Syntax element n	4	uimsbf
}		
...		
Syntax element n+1		
hoaGroupHasLowProfileConfig	1	bslbf
if (hoaGroupHasLowProfileConfig) {		
distanceFactor	3	uimsbf
}		

Ver la Tabla 4. En la modalidad de la FIG. 7A, los elementos de sintaxis "distanceFactor" y "hoaGroupHasLowProfileConfig" se agregan recientemente en esta solicitud, donde "distanceFactor" es el elemento de sintaxis de los metadatos de la señal de audio de escena, y "hoaGroupHasLowProfileConfig" es el elemento de sintaxis del segundo identificador.

"1" en la cuarta a la última fila en la Tabla 4 indica que el segundo identificador en el flujo de bits se puede representar usando 1 bit, y "bslbf" en la cuarta a la última fila en la Tabla 4 indica que el segundo identificador en el flujo de bits es una cadena de bits, bit izquierdo primero. Debe entenderse que la Tabla 4 es meramente un ejemplo de esta solicitud. El segundo identificador puede representarse alternativamente mediante el uso de otro número de bits, por ejemplo, 2 bits. Esto no está limitado en esta solicitud. Adicionalmente, el elemento de sintaxis del segundo identificador puede representarse alternativamente mediante el uso

de otra cadena de caracteres del segundo identificador. Esto no está limitado en esta solicitud.

Para obtener explicaciones de otros elementos de sintaxis en la Tabla 4, consulte las descripciones en la Tabla 1. En el presente documento no se describen los detalles nuevamente.

Por ejemplo, si la estructura sintáctica del flujo de bits se muestra en la Tabla 4, en un proceso de realización de S703, el segundo identificador se puede convertir en un número binario de 1 bit; y luego el número binario de 1 bit se escribe después de los datos identificados por el elemento sintáctico n+1 en el flujo de bits. Además, en un proceso de realización de S706, los metadatos de la señal de audio de escena se pueden convertir en un número binario de 8 bits; y luego el número binario de 8 bits se escribe después del segundo identificador en el flujo de bits. Cuando la estructura sintáctica del flujo de bits se muestra en la Tabla 4, el valor del segundo identificador es "1".

De una manera posible, un primer valor de identificador de un reproductor con realismo que lleva a cabo la reproducción con realismo con base en los metadatos de la señal de audio de escena puede usarse alternativamente como el valor del segundo identificador. Si el primer valor de identificador del reproductor con realismo que lleva a cabo la reproducción con realismo con base en los metadatos de la señal de audio de escena es "3", la estructura sintáctica del flujo de bits se puede mostrar en la Tabla 5.

Por ejemplo, la estructura sintáctica del flujo de bits se puede mostrar en la siguiente Tabla 5.

Tabla 5

Sintaxis	N.º de bits	Mnemónico
Syntax element 1		
{		
Syntax element 2	1	bslbf

Sintaxis	N.º de bits	Mnemónico
Syntax element 3	3	bslbf
...		
Syntax element n	4	uimsbf
}		
...		
Syntax element n+1		
hoaGroupHasLowProfileConfig	2	bslbf
if (hoaGroupHasLowProfileConfig==3) {		
distanceFactor	3	uimsbf
}		

La FIG. 7B es un diagrama de una estructura de flujo de bits de ejemplo. Una estructura de flujo de bits en la modalidad de la FIG. 7A se puede mostrar en la FIG. 7B, es decir, el segundo identificador es posterior a los datos codificados de la señal de audio de escena, y los metadatos de la señal de audio de escena son posteriores al segundo identificador.

Debe entenderse que las posiciones de los metadatos de la señal de audio de escena, los datos codificados de la señal de audio de escena y el segundo identificador en el flujo de bits no están limitados en esta solicitud.

Debe entenderse que, en el flujo de bits en esta solicitud, dos cualesquiera de los metadatos de la señal de audio de escena, los datos codificados de la señal de audio de escena y el segundo identificador pueden ser adyacentes o pueden no ser adyacentes. Esto no está limitado en esta solicitud.

5 La FIG. 8 es un diagrama de un ejemplo de proceso de decodificación. El proceso de decodificación en la FIG. 8 corresponde al proceso de codificación en la FIG. 7A.

S801: Recibir un flujo de bits.

S802: Obtener una señal de audio reconstruida de una señal de audio de escena basada en el flujo de bits.

10 Por ejemplo, para S801 y S802, consulte las descripciones anteriores. En el presente documento no se describen los detalles nuevamente.

S803: Análisis sintáctico del flujo de bits para obtener un segundo identificador.

S804: Determinar si un valor del segundo identificador es un primer valor de identificador preestablecido.

15 Por ejemplo, el flujo de bits se puede realizar un análisis sintáctico primero para obtener el segundo identificador; y luego, se determina si el segundo identificador es el primer valor de identificador preestablecido, para determinar si los metadatos de la señal de audio de escena se pueden obtener del flujo de bits a través del análisis sintáctico.

20 Por ejemplo, cuando el valor del segundo identificador es el primer valor de identificador preestablecido, indica que el flujo de bits incluye los metadatos de la señal de audio de escena. En este caso, se puede realizar S805. Cuando el valor del segundo identificador es otro valor de identificador, indica que el flujo de bits no incluye los metadatos de la señal de audio de la escena. En este caso, se decodifica un campo después del segundo identificador en el flujo de bits.

25 Por ejemplo, cuando se muestra una estructura de flujo de bits del flujo de bits en la Tabla 4, se puede determinar si el valor del segundo identificador es 1. Cuando el valor del segundo identificador es 1, indica que el flujo de bits incluye los metadatos de la señal de audio de escena; o cuando el valor del segundo identificador no es 1, indica que el flujo de bits no incluye los metadatos de la señal de audio de escena.

Por ejemplo, cuando se muestra una estructura de flujo de bits del flujo de bits en la Tabla 5, se puede determinar si el valor del segundo identificador es 3. Cuando el valor del segundo identificador es 3, indica que el flujo de bits incluye los metadatos de la señal de audio de escena; o cuando el valor del segundo identificador no es 3, indica que el flujo de bits no incluye los metadatos de la señal de audio de escena.

S805: Análisis sintáctico del flujo de bits para obtener los metadatos de la señal de audio de la escena.

Por ejemplo, puede haber una pluralidad de reproductores con realismo (cada reproductor con realismo utiliza un método de reproducción con realismo diferente) en un lado del decodificador. La pluralidad de reproductores con realismo incluye un reproductor con realismo que lleva a cabo la reproducción con realismo con base en los metadatos de una señal de audio de escena y un primer identificador (por ejemplo, un reproductor con realismo que lleva a cabo la reproducción con realismo utilizando el método de S11 a S16) y un reproductor con realismo que lleva a cabo la reproducción con realismo con base en otra información. Además, se puede establecer un valor de identificador correspondiente para cada reproductor con realismo. Por ejemplo, para un reproductor con realismo A, un valor de identificador correspondiente se establece en "0"; para un reproductor con realismo B, un valor de identificador correspondiente se establece en "1"; y para un reproductor con realismo C, un valor de identificador correspondiente se establece en "2". Cuando se determina que el lado del decodificador necesita llevar a cabo la reproducción con realismo con base en los metadatos de la señal de audio de escena y el primer identificador, de una manera posible, un valor de identificador (el cual posteriormente se denomina primer valor de identificador) del reproductor con realismo que lleva a cabo la reproducción con realismo con base en los metadatos de la señal de audio de escena y el primer identificador se puede utilizar como un valor de identificador de un segundo identificador; o de una manera posible, un segundo valor de identificador (por ejemplo, "0" o "1", el cual se puede establecer de acuerdo con un requisito) se puede utilizar como un valor de identificador de un segundo identificador; y luego el segundo identificador se agrega a un flujo de bits. El segundo valor de identificador indica que el lado del decodificador realizará la representación con base en los metadatos de la señal de audio de la escena y el primer identificador. En otras palabras, el segundo valor identificador indica que el flujo de bits incluye los metadatos de la señal de audio de escena y el primer identificador. De esta manera, es conveniente que el lado del decodificador determine qué reproductor con realismo o qué algoritmo de reproducción con realismo se utilizará para la reproducción con realismo.

Sintaxis	N.º de bits	Mnemónico
Syntax element 1		
{		
Syntax element 2	1	bslbf
Syntax element 3	3	bslbf
...		
Syntax element n	4	uimsbf
}		
...		
Syntax element n+1		
hoaGroupHasLowProfileConfig	1	bslbf
if (hoaGroupHasLowProfileConfig) {		
distanceFactor	3	uimsbf
complexityLevel	3	uimsbf
}		

La FIG. 9B es un diagrama de una estructura de flujo de bits de ejemplo. Una estructura de flujo de bits en la modalidad de la FIG. 9A se puede mostrar en la FIG. 9B, es decir, el segundo identificador es posterior a los datos codificados de la señal de audio de escena, los metadatos de la señal de audio de escena son posteriores al segundo identificador y el primer identificador es posterior a los metadatos de la señal de audio de escena.

Debe entenderse que las posiciones de los metadatos de la señal de audio de escena, los datos codificados de la señal de audio de escena, el primer identificador y el segundo identificador en el flujo de bits no están limitados en esta solicitud.

Debe entenderse que, en el flujo de bits en esta solicitud, dos cualesquiera de los metadatos de la señal de audio de escena, los datos codificados de la señal de audio de escena, el primer identificador y el segundo identificador pueden ser adyacentes o pueden no ser adyacentes. Esto no está limitado en esta solicitud.

La FIG. 10 es un diagrama de un ejemplo de proceso de decodificación. El proceso de decodificación en la FIG. 10 corresponde al proceso de codificación en la FIG. 9A.

S1001: Recibir un flujo de bits.

S1002: Obtener una señal de audio reconstruida de una señal de audio de escena basada en el flujo de bits.

S1003: Análisis sintáctico del flujo de bits para obtener un segundo identificador.

S1004: Determinar si un valor del segundo identificador es un primer valor de identificador preestablecido.

S1005: Análisis sintáctico del flujo de bits para obtener metadatos de la señal de audio de la escena.

S1006: Análisis sintáctico del flujo de bits para obtener un primer identificador.

Por ejemplo, para S1001 a S1006, consultar las descripciones en las modalidades anteriores. En el presente documento no se describen los detalles nuevamente.

Cabe destacar que en las modalidades en la FIG. 5A a la FIG. 10, en el proceso de codificación, un tercer identificador puede escribirse adicionalmente en el flujo de bits, y el

tercer identificador puede indicar si el flujo de bits incluye el primer identificador. En este caso, el segundo identificador puede indicar solo si el flujo de bits incluye los metadatos de la señal de audio de escena. Por ejemplo, cuando el primer identificador se escribe en el flujo de bits, un valor del tercer identificador se puede establecer en un segundo valor de identificador preestablecido, y luego el segundo valor de identificador preestablecido se escribe en el flujo de bits.

En consecuencia, en esta solicitud, un elemento de sintaxis de los metadatos de la señal de audio de escena y la información de descripción correspondiente, un elemento de sintaxis del segundo identificador y la información de descripción correspondiente, un elemento de sintaxis del primer identificador y la información de descripción correspondiente, y un elemento de sintaxis del tercer identificador y la información de descripción correspondiente se pueden agregar por adelantado a una estructura de sintaxis original del flujo de bits.

Por ejemplo, la estructura sintáctica del flujo de bits se puede mostrar en la siguiente Tabla 7.

Tabla 7

Sintaxis	N.º de bits	Mnemónico
Syntax element 1		
{		
Syntax element 2	1	bslbf
Syntax element 3	3	bslbf
...		
Syntax element n	4	uimbsf

Sintaxis	N.º de bits	Mnemónico
}		
...		
Syntax element n+1		
hoaGroupHasLowProfileConfig	1	bslbf
if (hoaGroupHasLowProfileConfig) {		
distanceFactor	3	uimsbf
}		
hoaGroupHascomplexityConfig		
if (hoaGroupHascomplexityConfig) {		
complexityLevel	3	uimsbf
}		

En la Tabla 7, "hoaGroupHascomplexityConfig" es el elemento de sintaxis del tercer identificador, y el segundo valor de identificador preestablecido es "1".

De esta manera, un lado del decodificador puede decodificar el flujo de bits para obtener el tercer identificador. Al determinar que el valor del tercer identificador es el segundo valor de identificador preestablecido, el lado del decodificador puede determinar que el flujo de bits incluye el primer identificador; y luego decodificar el flujo de bits para obtener el primer

identificador.

Por ejemplo, en esta solicitud, solo se puede escribir el primer identificador en el flujo de bits. En este caso, el lado del decodificador puede obtener los metadatos de la señal de audio de escena de otro dispositivo electrónico o generar localmente los metadatos de la señal de audio de escena.

La FIG. 11 es un diagrama de un proceso de codificación de ejemplo.

S1101: Obtener una señal de audio de escena.

S1102: Codificar la señal de audio de escena y un primer identificador para obtener un flujo de bits, donde el primer identificador indica complejidad de reproducción con realismo, la complejidad de reproducción con realismo se utiliza para determinar un número de altavoces virtuales, una señal de audio reconstruida de la señal de audio de escena se utiliza para generar una señal de altavoz virtual, y la señal de altavoz virtual y el número de altavoces virtuales se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo.

Para S1101 y S1102, consulte las descripciones anteriores. En el presente documento no se describen los detalles nuevamente.

Se puede agregar un elemento de sintaxis del primer identificador y la información de descripción correspondiente a una estructura de sintaxis del flujo de bits.

Por ejemplo, la estructura sintáctica del flujo de bits se puede mostrar en la siguiente Tabla 8.

Tabla 8

Sintaxis	N.º de bits	Mnemónico
Syntax element 1		
{		

Sintaxis	N.º de bits	Mnemónico
Syntax element 2	1	bslbf
Syntax element 3	3	bslbf
...		
Syntax element n	4	uimsbf
}		
...		
Syntax element n+h		
complexityLevel	3	uimsbf

La FIG. 12 es un diagrama de un ejemplo de proceso de decodificación. El proceso de decodificación en la FIG. 12 corresponde al proceso de codificación en la FIG. 11.

S1201: Recibir un flujo de bits.

5 S1202: Obtener una señal de audio reconstruida de una señal de audio de escena basada en el flujo de bits.

10 S1203: Análisis sintáctico del flujo de bits para obtener un primer identificador, donde el primer identificador indica la complejidad de la reproducción con realismo, la complejidad de la reproducción con realismo se utiliza para determinar un número de altavoces virtuales, la señal de audio reconstruida se utiliza para generar una señal de altavoz virtual, y la señal de altavoz virtual y el número de altavoces virtuales se utilizan para la reproducción con realismo

para obtener una señal de audio reproducida con realismo.

Por ejemplo, para S1201 a S1203, consulte las descripciones anteriores. En el presente documento no se describen los detalles nuevamente.

Adicionalmente, con base en la FIG. 11 y la FIG. 12, un lado del codificador puede escribir además un segundo identificador en el flujo de bits, donde el segundo identificador indica si el flujo de bits incluye el primer identificador. Por ejemplo, cuando el primer identificador se escribe en el flujo de bits, un valor del segundo identificador se puede establecer en un valor de identificador preestablecido, y luego el valor de identificador preestablecido se escribe en el flujo de bits.

Por ejemplo, la estructura sintáctica del flujo de bits se puede mostrar en la siguiente Tabla 9.

Tabla 9

Sintaxis	N.º de bits	Mnemónico
Syntax element 1		
{		
Syntax element 2	1	bslbf
Syntax element 3	3	bslbf
...		
Syntax element n	4	uimsbf
}		

Sintaxis	N.º de bits	Mnemónico
...		
Syntax element n+1		
hoaGroupHascomplexityConfig		
if (hoaGroupHascomplexityConfig) {		
complexityLevel	3	uimsbf
}		

En la Tabla 9, "hoaGroupHascomplexityConfig" es un elemento de sintaxis del segundo identificador, y el valor de identificador preestablecido es "1".

De esta manera, un lado del decodificador puede decodificar el flujo de bits para obtener el segundo identificador. Al determinar que el valor del segundo identificador es el valor de identificador preestablecido, el lado del decodificador puede determinar que el flujo de bits incluye el primer identificador; y luego decodificar el flujo de bits para obtener el primer identificador.

Adicionalmente, el elemento de sintaxis del segundo identificador y la información de descripción correspondiente se pueden agregar adicionalmente a la estructura de sintaxis del flujo de bits. Para más detalles, consulte las descripciones anteriores. En el presente documento no se describen los detalles nuevamente.

La FIG. 13 es un diagrama de un proceso de ejemplo para generar una señal reproducida con realismo binaural. La modalidad en la FIG. 13 se realiza por un dispositivo terminal, y el dispositivo terminal puede incluir un decodificador y un reproductor con realismo. La FIG. 13 se muestra con base en la FIG. 4A.

S1301: Recibir un flujo de bits.

S1302: Análisis sintáctico del flujo de bits para obtener metadatos de una señal de audio de escena.

5 S1303: Obtener una señal de audio reconstruida de la señal de audio de la escena con base en el flujo de bits.

Por ejemplo, para S1301 a S1303, consulte las descripciones anteriores de S401 a S403. En el presente documento no se describen los detalles nuevamente.

S1304: Generar una señal de altavoz virtual basada en la señal de audio reconstruida de la señal de audio de la escena.

10 S1305: Determinar un radio de altavoz virtual con base en los primeros metadatos y una distancia entre dos de una pluralidad de micrófonos.

S1306: Realizar la reproducción con realismo con base en el radio del altavoz virtual y la señal del altavoz virtual, para obtener una señal de audio reproducida con realismo.

15 Para S1304 a S1306, consulte las descripciones anteriores de S11 a S16. En el presente documento no se describen los detalles nuevamente.

Cabe señalar que S1305 y S1306 pueden ser realizados por el reproductor con realismo. De una manera posible, el decodificador puede realizar S1301 a S1303. De una manera posible, S1301 y S1302 pueden ser realizados por el decodificador, y S1303 puede ser implementado de forma colaborativa por el decodificador y otro módulo.

20 La FIG. 14 es un diagrama de un ejemplo de aparato de codificación. El aparato de codificación puede estar configurado para realizar el método en las modalidades anteriores. Por lo tanto, para conocer los efectos beneficiosos que se pueden lograr mediante el aparato de codificación, consulte los efectos beneficiosos en el método correspondiente proporcionado anteriormente. En el presente documento no se describen los detalles
25 nuevamente.

Un primer módulo de obtención de datos 1401 está configurado para obtener una señal de audio de escena y metadatos de la señal de audio de escena, donde la señal de audio de escena describe un campo de sonido de una fuente de sonido en una escena, la escena incluye una pluralidad de micrófonos, los metadatos incluyen primeros metadatos, los

5 primeros metadatos se utilizan en combinación con una distancia entre dos de la pluralidad de micrófonos para determinar un radio de altavoz virtual, el radio de altavoz virtual y una señal de altavoz virtual se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo, y la señal de altavoz virtual se genera con base en una señal de audio reconstruida de la señal de audio de escena.

Un primer módulo de codificación 1402 está configurado para codificar la señal de audio de escena y los metadatos para obtener un flujo de bits.

10 La FIG. 15 es un diagrama de un ejemplo de aparato de decodificación. El aparato de decodificación puede estar configurado para realizar el método en las modalidades anteriores. Por lo tanto, para los efectos beneficiosos que se pueden lograr mediante el aparato de decodificación, consulte los efectos beneficiosos en el método correspondiente proporcionado anteriormente. En el presente documento no se describen los detalles nuevamente.

Un primer módulo de recepción 1501 está configurado para recibir un flujo de bits.

15 Un primer módulo de análisis sintáctico 1502 está configurado para realizar un análisis sintáctico del flujo de bits para obtener metadatos de una señal de audio de escena.

20 Un primer módulo de obtención de señal 1503 está configurado para obtener una señal de audio reconstruida de la señal de audio de escena con base en el flujo de bits, donde la señal de audio de escena describe un campo de sonido de una fuente de sonido en una escena, la escena incluye una pluralidad de micrófonos, los metadatos incluyen primeros metadatos, los primeros metadatos se utilizan en combinación con una distancia entre dos de la pluralidad de micrófonos para determinar un radio de altavoz virtual, el radio de altavoz virtual y una señal de altavoz virtual se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo, y la señal de altavoz virtual se genera con base en la señal de audio reconstruida.

25 La FIG. 16 es un diagrama de un ejemplo de aparato de codificación. El aparato de codificación puede estar configurado para realizar el método en las modalidades anteriores. Por lo tanto, para conocer los efectos beneficiosos que se pueden lograr mediante el aparato de codificación, consulte los efectos beneficiosos en el método correspondiente proporcionado anteriormente. En el presente documento no se describen los detalles nuevamente.

30

Un segundo módulo de obtención de datos 1601 está configurado para obtener una

señal de audio de escena.

Un segundo módulo de codificación 1602 está configurado para codificar la señal de audio de escena y un primer identificador para obtener un flujo de bits, donde el primer identificador indica complejidad de reproducción con realismo, la complejidad de reproducción con realismo se utiliza para determinar un número de altavoces virtuales, una señal de audio reconstruida de la señal de audio de escena se utiliza para generar una señal de altavoz virtual, y la señal de altavoz virtual y el número de altavoces virtuales se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo.

La FIG. 17 es un diagrama de un ejemplo de aparato de decodificación. El aparato de decodificación puede estar configurado para realizar el método en las modalidades anteriores. Por lo tanto, para los efectos beneficiosos que se pueden lograr mediante el aparato de decodificación, consulte los efectos beneficiosos en el método correspondiente proporcionado anteriormente. En el presente documento no se describen los detalles nuevamente.

Un segundo módulo de recepción 1701 está configurado para recibir un flujo de bits.

Un segundo módulo de análisis sintáctico 1702 está configurado para realizar un análisis sintáctico del flujo de bits para obtener un primer identificador.

Un segundo módulo de obtención de señal 1703 está configurado para obtener una señal de audio reconstruida de una señal de audio de escena basada en el flujo de bits, donde el primer identificador indica complejidad de reproducción con realismo, la complejidad de reproducción con realismo se utiliza para determinar un número de altavoces virtuales, la señal de audio reconstruida se utiliza para generar una señal de altavoz virtual, y la señal de altavoz virtual y el número de altavoces virtuales se utilizan para reproducir con realismo para obtener una señal de audio reproducida con realismo.

En un ejemplo, la FIG. 18 es un diagrama de bloques de un aparato 1800 de acuerdo con una modalidad de esta solicitud. El aparato 1800 puede incluir un procesador 1801 y un transceptor/pin de transceptor 1802, y opcionalmente, además incluye una memoria 1803.

Los componentes del aparato 1800 se acoplan entre sí mediante el uso de un bus 1804. Además de un bus de datos, el bus 1804 también incluye un bus de alimentación, un bus de control y un bus de señal de estado. Sin embargo, en aras de claridad, los diversos buses están referidos como el bus 1804 en la figura.

Opcionalmente, la memoria 1803 puede configurarse para almacenar instrucciones en las modalidades de método anteriores. El procesador 1801 puede configurarse para ejecutar las instrucciones en la memoria 1803, controlar un PIN de recepción para recibir una señal y controlar un PIN de envío para enviar una señal.

- 5 El aparato 1800 puede ser el dispositivo electrónico o un chip en el dispositivo electrónico en las modalidades de método anteriores.

Por ejemplo, el dispositivo electrónico puede ser un dispositivo terminal o un servidor.

- 10 Todo el contenido relacionado de cada etapa en la modalidad del método anterior puede citarse a una descripción de función de un módulo funcional correspondiente, y los detalles no se describen nuevamente en la presente.

- 15 Una modalidad de esta solicitud además proporciona un chip, que incluye uno o más circuitos de interfaz y uno o más procesadores. El uno o más procesadores reciben o envían datos a través del uno o más circuitos de interfaz. Cuando uno o más procesadores ejecutan instrucciones informáticas, se realizan las etapas para implementar el método en las modalidades anteriores en las etapas del método relacionados anteriores. El circuito de interfaz es el transceptor/pin del transceptor 1802.

- 20 Una modalidad adicional proporciona un medio de almacenamiento legible por computador. El medio de almacenamiento legible por computador almacena las instrucciones del computador. Cuando las instrucciones informáticas se ejecutan en un dispositivo electrónico, el dispositivo electrónico está habilitado para realizar las etapas del método relacionadas anteriores para implementar el método en las modalidades anteriores.

- 25 Una modalidad además proporciona un producto de programa informático. El producto de programa informático incluye instrucciones informáticas, y cuando las instrucciones informáticas son ejecutadas por un computador o un procesador, el computador puede realizar las etapas relacionadas anteriores para implementar el método en las modalidades anteriores.

- 30 Adicionalmente, una modalidad de esta solicitud además proporciona un aparato. El aparato puede ser específicamente un chip, un componente o un módulo. El aparato puede incluir un procesador y una memoria que estén conectados. La memoria está configurada para almacenar instrucciones ejecutables por computador. Cuando el aparato se ejecuta, el procesador puede ejecutar las instrucciones ejecutables por computador almacenadas en la

memoria, de modo que el chip realice el método en las modalidades de método anteriores.

El dispositivo electrónico, el medio de almacenamiento legible por computador, el producto de programa informático o el chip proporcionado en las modalidades están configurados para realizar el método correspondiente proporcionado anteriormente. Por lo tanto, para conocer los efectos beneficiosos que se pueden lograr, consulte los efectos beneficiosos en el método correspondiente proporcionado anteriormente. En el presente documento no se describen los detalles nuevamente.

Una persona normalmente versada en la técnica puede aprender a partir de la descripción anterior de las implementaciones que, por conveniencia y brevedad de la descripción, la división en solo los módulos funcionales anteriores se utiliza como un ejemplo para la descripción. En la aplicación real, las funciones anteriores se pueden asignar a diferentes módulos funcionales para su implementación de acuerdo con un requisito, es decir, una estructura interna de un aparato se divide en diferentes módulos funcionales, para completar todas o algunas de las funciones descritas anteriormente.

En las diversas modalidades proporcionadas en esta solicitud, debe entenderse que el aparato y el método descritos pueden implementarse de otras maneras. Por ejemplo, las modalidades del aparato descritas son meramente ejemplos. Por ejemplo, la división en los módulos o unidades es simplemente una división de función lógica. En la implementación real, puede haber otra forma de división. Por ejemplo, se pueden combinar o integrar una pluralidad de unidades o componentes en otro aparato, o se pueden ignorar o no realizar algunas funciones. Adicionalmente, los acoplamientos mutuos o acoplamientos directos o conexiones de comunicación mostrados o expuestos pueden implementarse a través de algunas interfaces, y los acoplamientos indirectos o conexiones de comunicación entre aparatos o unidades pueden implementarse de forma eléctrica, mecánica u otra forma.

Las unidades descritas como partes separadas pueden o no estar físicamente separadas, y las partes mostradas como unidades pueden ser una o más unidades físicas, pueden estar ubicadas en un lugar, o pueden distribuirse en una pluralidad de lugares diferentes. Algunas o todas las unidades se pueden seleccionar de acuerdo con un requerimiento real para lograr los objetivos de las soluciones en las modalidades.

Adicionalmente, las unidades funcionales en las modalidades de la presente solicitud pueden integrarse en una unidad de procesamiento, o cada unidad puede existir físicamente sola, o dos o más unidades pueden integrarse en una unidad. La unidad integrada puede

implementarse en forma de *hardware*, o puede implementarse en forma de una unidad funcional de *software*.

Cualquier contenido en las modalidades de esta solicitud y cualquier contenido en una misma modalidad se pueden combinar libremente. Cualquier combinación del contenido anterior se encuentra dentro del alcance de esta solicitud.

Cuando la unidad integrada se implementa en una forma de una unidad funcional de *software* y se venden o utilizan como un producto independiente, la unidad integrada se puede almacenar en un medio de almacenamiento legible. Con base en tal entendimiento, las soluciones técnicas de las modalidades de esta solicitud esencialmente, o la parte que contribuye a la tecnología convencional, o todas o algunas de las soluciones técnicas pueden implementarse en forma de producto de *software*. El producto de *software* se almacena en un medio de almacenamiento e incluye varias instrucciones para indicar a un dispositivo (el cual puede ser un microcomputador de un solo chip, un chip o similares) o a un procesador (processor) que realice todos o algunos de las etapas de los métodos descritos en las modalidades de esta solicitud. El medio de almacenamiento anterior incluye cualquier medio que pueda almacenar código de programa, tal como una unidad flash USB, un disco duro extraíble, una memoria de solo lectura (read-only memory, ROM), una memoria de acceso aleatorio (memoria de acceso aleatorio, RAM), un disco magnético o un disco óptico.

Las modalidades de esta solicitud se describieron anteriormente con referencia a los dibujos adjuntos. Sin embargo, esta solicitud no se limita a las implementaciones específicas anteriores. Las implementaciones específicas anteriores son meramente un ejemplo, pero no una limitación. Muchas formas que puede realizar un experto en la técnica sin apartarse del principio de esta solicitud y el alcance de protección de las reivindicaciones caen dentro del alcance de protección de esta solicitud.

Las etapas del método o algoritmo descritos con referencia al contenido descrito en las modalidades de esta solicitud pueden implementarse de una manera de *hardware*, o pueden implementarse de una manera en la cual un procesador ejecute instrucciones de *software*. La instrucción de *software* puede incluir un módulo de *software* correspondiente. El módulo de *software* puede almacenarse en una memoria de acceso aleatorio (Random Access Memory, RAM), una memoria flash, una memoria de solo lectura (Read-Only Memory, ROM), una memoria de solo lectura programable borrable (Erasable Programmable ROM, EPROM), una memoria de solo lectura programable borrable eléctricamente (Electrically EPROM, EEPROM), un registro, un disco duro, un disco duro extraíble, una memoria de solo

lectura de disco compacto (CD-ROM) o cualquier otra forma de medio de almacenamiento bien conocido en la técnica. Un ejemplo de medio de almacenamiento está acoplado a un procesador, de modo que el procesador puede leer información del medio de almacenamiento y puede escribir información en el medio de almacenamiento. Ciertamente, el medio de almacenamiento puede ser alternativamente una parte del procesador. El procesador y el medio de almacenamiento pueden estar ubicados en un ASIC.

Una persona normalmente versada en la técnica debe ser consciente de que en los uno o más ejemplos anteriores, las funciones descritas en las modalidades de esta solicitud pueden implementarse mediante *hardware*, *software*, *firmware* o cualquier combinación de estos. Cuando se implementan mediante *software*, estas funciones pueden almacenarse en un medio legible por computador o transmitirse como una o más instrucciones o código en el medio legible por computador. El medio legible por computador incluye un medio de almacenamiento legible por computador y un medio de comunicación. El medio de comunicación incluye cualquier medio que facilite la transferencia de un programa informático de un lugar a otro. El medio de almacenamiento puede ser cualquier medio utilizable accesible a un computador de propósito general o dedicada.

Las modalidades de esta solicitud se describieron anteriormente con referencia a los dibujos adjuntos. Sin embargo, esta solicitud no se limita a las implementaciones específicas anteriores. Las implementaciones específicas anteriores son meramente un ejemplo, pero no una limitación. Muchas formas que puede realizar un experto en la técnica sin apartarse del principio de esta solicitud y el alcance de protección de las reivindicaciones caen dentro del alcance de protección de esta solicitud.