

解码方法及电子设备

5 本申请要求于 2023 年 07 月 10 日提交中国国家知识产权局、申请号为 202310843541.X、申请名称为“解码方法及电子设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

 本申请实施例涉及编解码领域，尤其涉及一种解码方法及电子设备。

10

背景技术

 面向扩展现实（Extended Reality, XR）应用的沉浸式音频需求，MPEG-I 音频标准提供了一种基于多点高阶立体混响（Higher-order Ambisonics, HOA）的 6 自由度（Degree of Freedom, DoF）音频技术：首先，在房间中部署若干声源以及多点 HOA 麦克风；接着，由声源播放音频信号，由多点 HOA 麦克风采集
15 房间内的声音，得到多点 HOA 信号；之后，多点 HOA 信号经过音频编码器后得到码流，音频解码器对码流解码后，根据解码得到的信号、HOA 的元数据、用户位置和姿态等信息进行渲染得到双耳渲染信号。

发明内容

 鉴于此，本申请提供一种解码方法及电子设备。在该解码方法中，解码端可以从码流中解析出场景音频信号的元数据，这样，解码端无需从其他电子设备获取场景音频信号的元数据，也无需在本地生成场景
20 音频信号的元数据，从而可以节省渲染得到双耳渲染信号的时间，提高用户体验。

 第一方面，本申请实施例提供一种编码方法，该方法包括：首先，获取场景音频信号和场景音频信号的元数据；其中，场景音频信号用于描述场景中声源的声场，场景包括多个麦克风，元数据包括第一元数据，第一元数据用于与多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，虚拟扬声器半径和虚拟
25 扬声器信号用于渲染以得到渲染后的音频信号，虚拟扬声器信号是根据场景音频信号的重建音频信号生成的；接着，编码场景音频信号和元数据，得到码流。这样，码流传输到解码端后，解码端可以从码流中解析出场景音频信号的元数据，将场景音频信号的元数据作用于后续渲染以得到渲染后的音频信号的过程；进而解码端无需从其他电子设备获取场景音频信号的元数据，也无需在本地生成场景音频信号的元数据，提高了解码端获取场景音频信号的元数据的效率，且降低了生成场景音频信号的元数据的计算量：从而
30 可以节省渲染得到渲染后的音频信号的时间，提高用户体验。

 示例性的，渲染后的音频信号可以是指输出至播放模块（如耳机）播放的音频信号，可以为双耳渲染信号。

 示例性的，多个麦克风中任意两个麦克风的距离均相等。

35 示例性的，场景音频信号可以包括：多点 HOA 信号（其中，HOA 信号可以包括三维 HOA 信号和二维 HOA 信号（也可以称为平面 HOA 信号））和三维音频信号；三维音频信号可以是指场景音频信号中除 HOA 信号之外的其他音频信号。本申请以场景音频信号为多点 HOA 信号为例进行说明。

 示例性的，麦克风可以为 HOA 麦克风。其中，可以先在房间（也就是真实场景）中部署若干声源以及多点 HOA 麦克风；接着，由声源播放音频信号，由多点 HOA 麦克风采集房间内的声音，得到多点 HOA
40 信号。

 示例性的，场景音频信号可以是指编码端待编码的音频信号，场景音频信号的重建音频信号是指解码端或者编码端中的解码模块根据编码场景音频信号得到的码流所获取的音频信号。其中，重建音频信号也可以称为重建场景音频信号。

 需要说明的是，本申请不限制编码场景音频信号和编码元数据的顺序。

45

 根据第一方面，该方法还包括：将第一标识写入码流；其中，第一标识用于指示渲染复杂度，渲染复杂度作用于确定虚拟扬声器的数量。这样，解码端可以根据虚拟扬声器的数量、虚拟扬声器半径和虚拟扬声器信号进行渲染，得到渲染后的音频信号。

示例性的，渲染复杂度不同，对应渲染得到的双耳渲染信号的音频质量不同；其中，第一标识的值可以由用户自定义设置，也可以根据解码端的计算能力确定，这样，解码端可以渲染得到音频质量与用户需求、电子设备的计算能力等因素匹配的双耳渲染信号。

5 根据第一方面，或者以上第一方面的任意一种实现方式，将第二标识写入码流；其中，第二标识用于指示码流中是否包含场景音频信号的元数据。

其中，解码端可能存在多个渲染器（每个渲染器采用的渲染方法不同），这多个渲染器包括：根据场景音频信号的元数据进行渲染的渲染器和根据其他信息进行渲染的渲染器；进而，第二标识还可以用于指示是否根据场景音频信号的元数据进行渲染。这样，解码端可以根据从码流中解析出的第二标识的值，确定采用哪个渲染器或哪种渲染算法进行渲染。

根据第一方面，或者以上第一方面的任意一种实现方式，第二标识还用于指示码流中是否还包含第一标识。

15 其中，解码端包括的多个渲染器可以包括：根据场景音频信号的元数据和第一标识进行渲染的渲染器和根据其他信息进行渲染的渲染器；进而第二标识还可以用于指示是否根据场景音频信号的元数据和第一标识进行渲染。这样，解码端可以根据从码流中解析出的第二标识的值，确定采用哪个渲染器或哪种渲染算法进行渲染。

20 根据第一方面，或者以上第一方面的任意一种实现方式，该方法还包括：将第三标识写入码流；其中，第三标识用于指示码流是否包含第一标识。这种情况中，第二标识可以仅用于指示码流中是否包含元数据。

根据第一方面，或者以上第一方面的任意一种实现方式，获取场景音频信号的元数据，包括：获取场景的场景描述信息；根据场景的场景描述信息，确定元数据。这样，根据场景音频信号的元数据渲染得到的双耳渲染信号，更接近用户身处真实环境中所听到的真实音频信号。

25 示例性的，场景描述信息可以用于描述场景的信息，例如可以包括但不限于：真实场景中多点 HOA 麦克风的标识、位置信息、特性和方向等，真实场景包含的对象的坐标、材质等，真实场景中声源的标识、位置信息、特性和方向等，真实场景的房间临界距离，声学混响参数（如直混比（Direct-to-Reverberation energy Ratio, DRR）、混响时间（Reverberation Time）（如 T20（能量衰减 20dB 的时间）、T30、T60 等）、清晰度（Clarity）（如 C50（50ms 前后的声能比例）、C80 等）等等，本申请对此不作限制。

30 示例性的，房间临界距离可以是指声源在室内（真实环境中）稳定地辐射声波时，在某一方向上直达声能量密度与混响声能量密度相等的点到声源中心的距离、也称扩散场距离。当声源辐射指向性为球形时，各方向的临界距离相等，这时也称混响半径。

35 根据第一方面，或者以上第一方面的任意一种实现方式，根据场景的场景描述信息，确定元数据，包括：当场景描述信息包括场景的房间临界距离时，根据房间临界距离和预设数值的乘积值，确定第一元数据。

40 根据第一方面，或者以上第一方面的任意一种实现方式，根据场景的场景描述信息，确定元数据，包括：当场景描述信息包括场景的声学混响参数时，根据场景的声学混响参数，确定场景的混响类型；根据场景的混响类型，确定第一元数据。

其中，混响类型可以包括多种，不同的混响类型对应的第一元数据不同。例如：第一混响类型（对应干声场景或无混响房间），对应的第一元数据为 0.5；第二混响类型（对应湿声场景或混响房间）对应的第一元数据为 1.5。

45 根据第一方面，或者以上第一方面的任意一种实现方式，元数据还包括第二元数据，根据场景的场景描述信息，确定元数据，包括：当场景描述信息包括多个麦克风的位置信息时，将多个麦克风的位置信息作为第二元数据；其中，第二元数据用于确定多个麦克风中的两个麦克风之间的距离。

根据第一方面，或者以上第一方面的任意一种实现方式，元数据还包括第二元数据，根据场景的场景

描述信息，确定元数据，包括：当场景描述信息包括多个麦克风的位置信息时，根据多个麦克风的位置信息，确定多个麦克风中的两个麦克风之间的距离；将多个麦克风中的两个麦克风之间的距离，作为第二元数据。

5 根据第一方面，或者以上第一方面的任意一种实现方式，获取场景音频信号的元数据，包括：获取用户输入的场景音频信号的元数据。也就是说，用户可以自定义设置场景音频信号的元数据，这样，可以使
得解码端后续渲染得到的双耳渲染信号与用户个性化需求匹配。

10 根据第一方面，或者以上第一方面的任意一种实现方式，码流包括：场景音频信号的元数据的语法元素和对应的描述信息。

示例性的，可以在码流的原始语法结构（或者称为高层语法元素表）中，添加场景音频信号的元数据的语法元素和对应的描述信息；这样，编码端可以按照码流的语法结构中，场景音频信号的元数据的语法元素和对应的描述信息，将场景音频信号的元数据写入码流中。

15 根据第一方面，或者以上第一方面的任意一种实现方式，码流包括：第一标识的语法元素和对应的描述信息。

示例性的，可以在码流的原始语法结构中，添加第一标识的语法元素和对应的描述信息；这样，编码端可以按照码流的语法结构中，第一标识的语法元素和对应的描述信息，将第一标识写入码流中。

20 根据第一方面，或者以上第一方面的任意一种实现方式，码流包括：第二标识的语法元素和对应的描述信息。

示例性的，可以在码流的原始语法结构中，添加第二标识的语法元素和对应的描述信息；这样，编码端可以按照码流的语法结构中，第二标识的语法元素和对应的描述信息，将第二标识写入码流中。

25 根据第一方面，或者以上第一方面的任意一种实现方式，码流包括：第三标识的语法元素和对应的描述信息。

30 根据第一方面，或者以上第一方面的任意一种实现方式，场景音频信号包括多点高阶立体混响 HOA 信号。

根据第一方面，或者以上第一方面的任意一种实现方式，第一元数据的语法元素为 distanceFactor。

根据第一方面，或者以上第一方面的任意一种实现方式，第一标识的语法元素为 complexityLevel。

35 根据第一方面，或者以上第一方面的任意一种实现方式，第二标识的语法元素为
hoaGroupHasLowProfileConfig。

40 第二方面，本申请实施例提供一种解码方法，该方法包括：接收码流；从码流中解析出场景音频信号的元数据：根据码流，获取场景音频信号的重建音频信号；场景音频信号用于描述场景中声源的声场，场景包括多个麦克风，元数据包括第一元数据，第一元数据用于与多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，虚拟扬声器半径和虚拟扬声器信号用于渲染以得到渲染后的音频信号，虚拟扬声器信号是根据重建音频信号生成的。这样，解码端无需从其他电子设备获取场景音频信号的元数据，也无需在本地生成场景音频信号的元数据，提高了获取场景音频信号的元数据的效率，且降低了生成场景音频信号的元数据的计算量；从而可以节省渲染得到双耳渲染信号的时间，提高用户体验。

45 根据第二方面，该方法还包括：从码流中解析出第一标识；其中，第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量。

根据第二方面，或者以上第二方面的任意一种实现方式，该方法方法还包括：从码流中解析出第二标识；其中，第二标识的值为第一预设标识值时指示码流包括元数据。也就是说，当第二标识的值为第一预设标识值，可以确定解码端是根据场景音频信号的元数据进行渲染的，以及确定码流中包含场景音频信号的元数据；进而，可以继续对码流进行解析，得到场景音频信号的元数据。

5

根据第二方面，或者以上第二方面的任意一种实现方式，第二标识的值为第一预设标识值时指示码流中还包括第一标识。也就是说，当第二标识的值为第一预设标识值时，可以解码端是确定根据场景音频信号的元数据和第一标识进行渲染的，以及还可以确定码流中还包含第一标识；进而，可以继续对码流进行解析，得到第一标识。

10

根据第二方面，或者以上第二方面的任意一种实现方式，该方法还包括：从码流中解析出第三标识；其中，第三标识的值为第二预设标识值时指示码流中包括第一标识。

根据第二方面，或者以上第二方面的任意一种实现方式，元数据还包括第二元数据，第二元数据用于确定多个麦克风中的两个麦克风之间的距离。

15

根据第二方面，或者以上第二方面的任意一种实现方式，场景音频信号包括多点高阶立体混响 HOA 信号，重建音频信号包括多点 HOA 信号的重建音频信号。

根据第二方面，或者以上第二方面的任意一种实现方式，码流包括：元数据的语法元素和对应的描述信息；从码流中解析出场景音频信号的元数据，包括：根据元数据的语法元素和对应的描述信息，从码流中解析出元数据。

20

根据第二方面，或者以上第二方面的任意一种实现方式，码流包括：第一标识的语法元素和对应的描述信息，从码流中解析出第一标识，包括：根据第一标识的语法元素和对应的描述信息，从码流中解析出第一标识。

25

根据第二方面，或者以上第二方面的任意一种实现方式，码流包括：第二标识的语法元素和对应的描述信息，从码流中解析出第二标识，包括：根据第二标识的语法元素和对应的描述信息，从码流中解析出第二标识。

30

根据第二方面，或者以上第二方面的任意一种实现方式，码流包括：第三标识的语法元素和对应的描述信息，从码流中解析出第三标识，包括：根据第三标识的语法元素和对应的描述信息，从码流中解析出第三标识。

根据第二方面，或者以上第二方面的任意一种实现方式，第一元数据的语法元素为 distanceFactor。

35

根据第二方面，或者以上第二方面的任意一种实现方式，第一标识的语法元素为 complexityLevel。

根据第二方面，或者以上第二方面的任意一种实现方式，第二标识的语法元素为 hoaGroupHasLowProfileConfig。

40

根据第二方面，或者以上第二方面的任意一种实现方式，该方法还包括：根据场景音频信号的重建音频信号，生成虚拟扬声器信号；根据第一元数据与多个麦克风中的两个麦克风之间的距离，确定虚拟扬声器半径；根据虚拟扬声器半径和虚拟扬声器信号进行渲染，得到渲染后的音频信号。

第二方面以及第二方面的任意一种实现方式分别与第一方面以及第一方面的任意一种实现方式相对应。第二方面以及第二方面的任意一种实现方式所对应的技术效果可参见上述第一方面以及第一方面的任意一种实现方式所对应的技术效果，此处不再赘述。

45

第三方面，本申请实施例提供一种编码方法，该方法包括：首先，获取场景音频信号；接着，编码场景音频信号和第一标识，以得到码流；第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量，场景音频信号的重建音频信号用于生成虚拟扬声器信号，虚拟扬声器信号和虚拟扬声器的数量用于渲染以得到渲染后的音频信号。

5 这样，解码端可以根据虚拟扬声器的数量、虚拟扬声器半径和虚拟扬声器信号进行渲染，得到渲染后的音频信号。示例性的，渲染复杂度不同，对应渲染得到的双耳渲染信号的音频质量不同；其中，第一标识的值可以由用户自定义设置，也可以根据解码端的计算能力确定，这样，解码端可以渲染得到音频质量与用户需求、电子设备的计算能力等因素匹配的双耳渲染信号。

10 根据第三方面，该方法还包括：将第二标识写入码流；其中，第二标识用于指示码流是否包含第一标识。

根据第三方面，或者以上第三方面的任意一种实现方式，码流包括：第一标识的语法元素和对应的描述信息。

15 根据第三方面，或者以上第三方面的任意一种实现方式，码流包括：第二标识的语法元素和对应的描述信息。

20 根据第三方面，或者以上第三方面的任意一种实现方式，场景音频信号包括多点高阶立体混响 HOA 信号。

根据第三方面，或者以上第三方面的任意一种实现方式，第一标识的语法元素为 complexityLevel。

25 第三方面以及第三方面的任意一种实现方式分别与第一方面以及第一方面的任意一种实现方式相对应。第三方面以及第三方面的任意一种实现方式所对应的技术效果可参见上述第一方面以及第一方面的任意一种实现方式所对应的技术效果，此处不再赘述。

30 第四方面，本申请实施例还提供一种解码方法，该方法包括：接收码流；从码流中解析出第一标识；根据码流，获取重建音频信号，其中，第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量，重建音频信号用于生成虚拟扬声器信号，虚拟扬声器信号和虚拟扬声器的数量用于渲染以得到渲染后的音频信号。

根据第四方面，该方法还包括：从码流中解析出第二标识；第二标识的值为预设标识值时指示码流中包括第一标识。

35 根据第四方面，或者以上第四方面的任意一种实现方式，码流包括：第一标识的语法元素和对应的描述信息，从码流中解析出第一标识，包括：根据第一标识的语法元素和对应的描述信息，从码流中解析出第一标识。

40 根据第四方面，或者以上第四方面的任意一种实现方式，码流包括：第二标识的语法元素和对应的描述信息，从码流中解析出第二标识，包括：根据第二标识的语法元素和对应的描述信息，从码流中解析出第二标识。

根据第四方面，或者以上第四方面的任意一种实现方式，重建音频信号包括多点高阶立体混响 HOA 信号的重建音频信号。

45 根据第四方面，或者以上第四方面的任意一种实现方式，第一标识的语法元素为 complexityLevel。

根据第四方面，或者以上第四方面的任意一种实现方式，该方法还包括：根据场景音频信号的重建音频信号，生成虚拟扬声器信号；根据第一标识，确定虚拟扬声器的数量；根据虚拟扬声器的数量和虚拟扬

声器信号进行渲染，得到渲染后的音频信号。

第四方面以及第四方面的任意一种实现方式分别与第一方面以及第一方面的任意一种实现方式相对应。第四方面以及第四方面的任意一种实现方式所对应的技术效果可参见上述第一方面以及第一方面的任意一种实现方式所对应的技术效果，此处不再赘述。

5

第五方面，本申请实施例提供一种编码装置，该编码装置包括：

第一数据获取模块，用于获取场景音频信号和场景音频信号的元数据；其中，场景音频信号用于描述场景中声源的声场，场景包括多个麦克风，元数据包括第一元数据，第一元数据用于与多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，虚拟扬声器半径和虚拟扬声器信号用于渲染以得到渲染后的音频信号，虚拟扬声器信号是根据场景音频信号的重建音频信号生成的；

10

第一编码模块，用于编码场景音频信号和元数据，得到码流。

此外，编码装置还可以包括通信模块。

应该理解的是，编码装置还可以执行上述第一方面以及第一方面的任意一种实现方式，对应的效果也可以参见上述第一方面以及第一方面的任意一种实现方式所对应的技术效果，此处不再赘述。

15

第六方面，本申请实施例提供一种解码装置，该解码装置包括：

第一接收模块，用于接收码流；

第一解析模块，用于从码流中解析出场景音频信号的元数据；

第一信号获取模块，用于根据码流，获取场景音频信号的重建音频信号；场景音频信号用于描述场景中声源的声场，场景包括多个麦克风，元数据包括第一元数据，第一元数据用于与多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，虚拟扬声器半径和虚拟扬声器信号用于渲染以得到渲染后的音频信号，虚拟扬声器信号是根据重建音频信号生成的。

20

此外，解码装置还可以包括通信模块。

应该理解的是，解码装置还可以执行上述第二方面以及第二方面的任意一种实现方式，对应的效果也可以参见上述第二方面以及第二方面的任意一种实现方式所对应的技术效果，此处不再赘述。

25

第七方面，本申请实施例提供一种编码装置，该编码装置包括：

第二数据获取模块，用于获取场景音频信号；

第二编码模块，用于编码场景音频信号和第一标识，以得到码流；第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量，场景音频信号的重建音频信号用于生成虚拟扬声器信号，虚拟扬声器信号和虚拟扬声器的数量用于渲染以得到渲染后的音频信号。

30

此外，编码装置还可以包括通信模块。

应该理解的是，编码装置还可以执行上述第三方面以及第三方面的任意一种实现方式，对应的效果也可以参见上述第三方面以及第三方面的任意一种实现方式所对应的技术效果，此处不再赘述。

35

第八方面，本申请实施例提供一种解码装置，该解码装置包括：

第二接收模块，用于接收码流；

第二解析模块，用于从码流中解析出第一标识；

第二信号获取模块，用于根据码流，获取场景音频信号的重建音频信号；第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量，重建音频信号用于生成虚拟扬声器信号，虚拟扬声器信号和虚拟扬声器的数量用于渲染以得到渲染后的音频信号。

40

此外，解码装置还可以包括通信模块。

应该理解的是，解码装置还可以执行上述第四方面以及第四方面的任意一种实现方式，对应的效果也可以参见上述第四方面以及第四方面的任意一种实现方式所对应的技术效果，此处不再赘述。

45

第九方面，本申请实施例提供一种电子设备，包括：存储器和处理器，存储器与处理器耦合；存储器存储有程序指令，当程序指令由处理器执行时，使得电子设备执行第一方面或第一方面的任意可能的实现方式中的编码方法；或者，使得电子设备执行第三方面或第三方面的任意可能的实现方式中的编码方法。

第九方面以及第九方面的任意一种实现方式分别与第一方面以及第一方面的任意一种实现方式相对应，或者，分别与第三方面以及第三方面的任意一种实现方式相对应。第九方面以及第九方面的任意一种实现方式所对应的技术效果可参见上述第一方面以及第一方面的任意一种实现方式所对应的技术效果，或者，可参见上述第三方面以及第三方面的任意一种实现方式所对应的技术效果此处不再赘述。

5

第十方面，本申请实施例提供一种电子设备，包括：存储器和处理器，存储器与处理器耦合；存储器存储有程序指令，当程序指令由处理器执行时，使得电子设备执行第二方面或第二方面的任意可能的实现方式中的解码方法；或者，使得电子设备执行第四方面或第四方面的任意可能的实现方式中的解码方法。

第十方面以及第十方面的任意一种实现方式分别与第二方面以及第二方面的任意一种实现方式相对应，或者，分别与第四方面以及第四方面的任意一种实现方式相对应。第十方面以及第十方面的任意一种实现方式所对应的技术效果可参见上述第二方面以及第二方面的任意一种实现方式所对应的技术效果，或者，可参见上述第四方面以及第四方面的任意一种实现方式所对应的技术效果此处不再赘述。

第十一方面，本申请实施例提供一种芯片，包括一个或多个接口电路和一个或多个处理器；一个或多个处理器通过一个或多个接口电路接收或发送数据，当一个或多个处理器执行计算机指令时，使得第一方面或第一方面的任意可能的实现方式中的方法被执行；或者，使得第三方面或第三方面的任意可能的实现方式中的方法被执行。

第十一方面以及第十一方面的任意一种实现方式分别与第一方面以及第一方面的任意一种实现方式相对应，或者，分别与第三方面以及第三方面的任意一种实现方式相对应。第十一方面以及第十一方面的任意一种实现方式所对应的技术效果可参见上述第一方面以及第一方面的任意一种实现方式所对应的技术效果，或者，可参见上述第三方面以及第三方面的任意一种实现方式所对应的技术效果此处不再赘述。

第十二方面，本申请实施例提供一种芯片，包括一个或多个接口电路和一个或多个处理器；一个或多个处理器通过一个或多个接口电路接收或发送数据，当一个或多个处理器执行计算机指令时，使得第二方面或第二方面的任意可能的实现方式中的方法被执行；或者，使得第四方面或第四方面的任意可能的实现方式中的方法被执行。

第十二方面以及第十二方面的任意一种实现方式分别与第二方面以及第二方面的任意一种实现方式相对应，或者，分别与第四方面以及第四方面的任意一种实现方式相对应。第十二方面以及第十二方面的任意一种实现方式所对应的技术效果可参见上述第二方面以及第二方面的任意一种实现方式所对应的技术效果，或者，可参见上述第四方面以及第四方面的任意一种实现方式所对应的技术效果此处不再赘述。

第十三方面，本申请实施例提供一种计算机可读存储介质，计算机可读存储介质存储有计算机程序，当计算机程序运行在计算机或处理器上时，使得计算机或处理器执行第一方面或第一方面的任意可能的实现方式中的编码方法；或者，使得计算机或处理器执行第三方面或第三方面的任意可能的实现方式中的编码方法。

第十三方面以及第十三方面的任意一种实现方式分别与第一方面以及第一方面的任意一种实现方式相对应，或者，分别与第三方面以及第三方面的任意一种实现方式相对应。第十三方面以及第十三方面的任意一种实现方式所对应的技术效果可参见上述第一方面以及第一方面的任意一种实现方式所对应的技术效果，或者，可参见上述第三方面以及第三方面的任意一种实现方式所对应的技术效果此处不再赘述。

40

第十四方面，本申请实施例提供一种计算机可读存储介质，计算机可读存储介质存储有计算机程序，当计算机程序运行在计算机或处理器上时，使得计算机或处理器执行第二方面或第二方面的任意可能的实现方式中的解码方法；或者，使得计算机或处理器执行第四方面或第四方面的任意可能的实现方式中的解码方法。

第十四方面以及第十四方面的任意一种实现方式分别与第二方面以及第二方面的任意一种实现方式相对应，或者，分别与第四方面以及第四方面的任意一种实现方式相对应。第十四方面以及第十四方面的任意一种实现方式所对应的技术效果可参见上述第二方面以及第二方面的任意一种实现方式所对应的技术效果，或者，可参见上述第四方面以及第四方面的任意一种实现方式所对应的技术效果此处不再赘述。

第十五方面，本申请实施例提供一种计算机程序产品，计算机程序产品包括计算机指令，当计算机指令被计算机或处理器执行时，使得计算机或处理器执行第一方面或第一方面的任意可能的实现方式中的编码方法，或者，使得计算机或处理器执行第三方面或第三方面的任意可能的实现方式中的编码方法。

第十五方面以及第十五方面的任意一种实现方式分别与第一方面以及第一方面的任意一种实现方式相对应，或者，分别与第三方面以及第三方面的任意一种实现方式相对应。第十五方面以及第十五方面的任意一种实现方式所对应的技术效果可参见上述第一方面以及第一方面的任意一种实现方式所对应的技术效果，或者，可参见上述第三方面以及第三方面的任意一种实现方式所对应的技术效果此处不再赘述。

第十六方面，本申请实施例提供一种计算机程序产品，计算机程序产品包括计算机指令，当计算机指令被计算机或处理器执行时，使得计算机或处理器执行第二方面或第二方面的任意可能的实现方式中的编码方法，或者，使得计算机或处理器执行第四方面或第四方面的任意可能的实现方式中的编码方法。

第十六方面以及第十六方面的任意一种实现方式分别与第二方面以及第二方面的任意一种实现方式相对应，或者，分别与第四方面以及第四方面的任意一种实现方式相对应。第十六方面以及第十六方面的任意一种实现方式所对应的技术效果可参见上述第二方面以及第二方面的任意一种实现方式所对应的技术效果，或者，可参见上述第四方面以及第四方面的任意一种实现方式所对应的技术效果此处不再赘述。

第十七方面，本申请实施例还提供一种XR设备，XR设备包括：处理器和音频播放模块；
处理器用于执行如上述第二方面以及第二方面的任意一种实现方式中的步骤；
音频播放模块用于播放第二方面以及第二方面的任意一种实现方式中涉及的渲染后的音频信号。

第十七方面以及第十七方面的任意一种实现方式分别与第二方面以及第二方面的任意一种实现方式相对应。第十七方面以及第十七方面的任意一种实现方式所对应的技术效果可参见上述第二方面以及第二方面的任意一种实现方式所对应的技术效果，此处不再赘述。

第十八方面，本申请实施例还提供一种XR设备，XR设备包括：处理器和音频播放模块；
处理器用于执行如上述第四方面以及第四方面的任意一种实现方式中的步骤；
音频播放模块用于播放第四方面以及第四方面的任意一种实现方式中涉及的渲染后的音频信号。

第十八方面以及第十八方面的任意一种实现方式分别与第四方面以及第四方面的任意一种实现方式相对应。第十八方面以及第十八方面的任意一种实现方式所对应的技术效果可参见上述第四方面以及第四方面的任意一种实现方式所对应的技术效果，此处不再赘述。

附图说明

图1A为示例性示出的应用场景示意图；
图1B为示例性示出的应用场景示意图；
图1C为示例性示出的应用场景示意图；
图1D为示例性示出的应用场景示意图；
图2A为示例性示出的场景音频信号的压缩框架的示意图；
图2B为示例性示出的场景音频信号的转码框架示意图；
图2C为示例性示出的渲染过程示意图；
图2D为示例性示出的场景示意图；
图2E为示例性示出的场景示意图；
图3A为示例性示出的编码过程示意图；
图3B为示例性示出的码流结构示意图；
图4A为示例性示出的解码过程示意图；
图4B为示例性示出的双耳渲染信号的生成过程示意图；
图5A为示例性示出的编码过程示意图；
图5B为示例性示出的码流结构示意图；
图6A为示例性示出的解码过程示意图；
图6B为示例性示出的双耳渲染信号的生成过程示意图；

- 图 7A 为示例性示出的编码过程示意图;
- 图 7B 为示例性示出的码流结构示意图;
- 图 8 为示例性示出的解码过程示意图;
- 图 9A 为示例性示出的编码过程示意图;
- 5 图 9B 为示例性示出的码流结构示意图;
- 图 10 为示例性示出的解码过程示意图;
- 图 11 为示例性示出的编码过程示意图;
- 图 12 为示例性示出的解码过程示意图;
- 图 13 示例性示出的双耳渲染信号的生成过程示意图;
- 10 图 14 为示例性示出的编码装置示意图;
- 图 15 为示例性示出的解码装置示意图;
- 图 16 为示例性示出的编码装置示意图;
- 图 17 为示例性示出的解码装置示意图;
- 图 18 为示例性示出的装置的结构示意图。

15

具体实施方式

下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

20

本文中术语“和/或”,仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A 和/或 B,可以表示:单独存在 A,同时存在 A 和 B,单独存在 B 这三种情况。

本申请实施例的说明书和权利要求书中的术语“第一”和“第二”等是用于区别不同的对象,而不是用于描述对象的特定顺序。例如,第一目标对象和第二目标对象等是用于区别不同的目标对象,而不是用于描述目标对象的特定顺序。

25

在本申请实施例中,“示例性的”或者“例如”等词用于表示作例子、例证或说明。本申请实施例中被描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言,使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。

在本申请实施例的描述中,除非另有说明,“多个”的含义是指两个或两个以上。例如,多个处理单元是指两个或两个以上的处理单元;多个系统是指两个或两个以上的系统。

30

为了下述各实施例的描述清楚简洁,首先给出相关技术的简要介绍。

声音(sound)是由物体振动产生的一种连续的波。产生振动而发出声波的物体称为声源。声波通过介质(如:空气、固体或液体)传播的过程中,人或动物的听觉器官能感知到声音。

35

声波的特征包括音调、音强和音色。音调表示声音的高低。音强表示声音的大小。音强也可以称为响度或音量。音强的单位是分贝(decibel, dB)。音色又称为音品。

声波的频率决定了音调的高低。频率越高音调越高。物体在一秒钟之内振动的次数称为频率,频率单位是赫兹(hertz, Hz)。人耳能识别的声音的频率在 20 Hz~20000 Hz 之间。

声波的幅度决定了音强的强弱。幅度越大音强越大。距离声源越近,音强越大。

声波的波形决定了音色。声波的波形包括方波、锯齿波、正弦波和脉冲波等。

40

根据声波的特征,声音可以分为规则声音和无规则声音。无规则声音是指声源无规则地振动发出的声音。无规则声音例如是影响人们工作、学习和休息等的噪声。规则声音是指声源规则地振动发出的声音。规则声音包括语音和乐音。声音用电表示时,规则声音是一种在时频域上连续变化的模拟信号。该模拟信号可以称为音频信号。音频信号是一种携带语音、音乐和音效的信息载体。

45

由于人的听觉具有辨别空间中声源的位置分布的能力,则听音者听到空间中的声音时,除了能感受到声音的音调、音强和音色外,还能感受到声音的方位。

随着人们对听觉系统体验的关注和品质要求与日俱增,为了增强声音的纵深感、临场感和空间感,则三维音频技术应运而生。从而听音者不仅感受到来自前、后、左和右的声源发出的声音,而且感受到自己所处空间被这些声源产生的空间声场(简称“声场”(sound field))所包围的感觉,以及声音向四周扩

散的感觉，营造出一种使听音者置身于影院或音乐厅等场所的“身临其境”的音响效果。

众所周知，声波在理想介质中传播，波数为 $k = w/c$ ，角频率为 $w = 2\pi f$ ，其中， f 为声波频率， c 为声速。声压 p 满足公式(1)， ∇^2 为拉普拉斯算子。

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

- 5 假设人耳以外的空间系统是一个球形，听音者处于球的中心，从球外传来的声音在球面上有一个投影，过滤掉球面以外的声音，假设声源分布在这个球面上，用球面上的声源产生的声场来拟合原始声源产生的声场，即三维音频技术就是一个拟合声场的方法。具体地，在球坐标系下求解公式(1)等式方程，在无源球形区域内，该公式(1)方程解为如下公式(2)。

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

- 10 其中， r 表示球半径， θ 表示水平角信息（或者称为方位角信息）， φ 表示俯仰角信息（或称为仰角信息）， k 表示波数， s 表示理想平面波的幅度， m 表示 HOA 信号的阶数序号（或称为 HOA 信号的阶数序号）。 $j_m^{kr} j_m^{kr}$ 表示球贝塞尔函数，球贝塞尔函数又称为径向基函数，其中，第一个 j 表示虚数单位， $(2m+1) j_m^{kr} j_m^{kr}$ 不随角度变化。 $Y_{m,n}^{\sigma}(\theta, \varphi)$ 表示 θ, φ 方向的球谐函数， $Y_{m,n}^{\sigma}(\theta_s, \varphi_s)$ 表示声源方向的球谐函数。HOA 信号满足公式(3)。

$$15 \quad B_{m,n}^{\sigma} = s \cdot Y_{m,n}^{\sigma}(\theta_s, \varphi_s) \quad (3)$$

将公式(3)代入公式(2)，公式(2)可以变形为公式(4)。

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

- 其中，将 m 截断到第 N 项，即 $m=N$ ，以 $B_{m,n}^{\sigma}$ 作为对声场的近似描述；此时， $B_{m,n}^{\sigma}$ 可以称为 HOA 系数（可以用于表示 N 阶 HOA 信号）。声场是指介质中有声波存在的区域。 N 为大于或等于 1 的整数， N 阶 HOA 信号包括 $(N+1)^2$ 个通道的音频信号。

公式(4)表明声场可以在球面上按球谐函数展开，即声场可以分解为多个平面波的叠加。因此，可以将 HOA 信号描述的声场使用多个平面波的叠加来表达，并通过 HOA 系数重建声场。

- 25 本申请实施例涉及的场景音频信号是一种携带声场中声源的空间位置信息的信息载体，描述了空间中听音者的声场，或者说可以是指用于描述声源声场的信号。其中，场景音频信号可以包括：多点 HOA 信号（其中，HOA 信号可以包括三维 HOA 信号和二维 HOA 信号（也可以称为平面 HOA 信号））和三维音频信号；三维音频信号可以是指场景音频信号中除 HOA 信号之外的其他音频信号。本申请以场景音频信号为多点 HOA 信号为例进行说明。

- 30 图 1A 为示例性示出的应用场景示意图。在图 1A 中，示出的是用户使用增强现实（Augmented Reality, AR）设备体验 AR 音乐会的应用场景。应该理解的是，用户还可以使用 AR 设备体验其他 AR 项目如 AR 科学讲堂、AR 影院等等，本申请对此不作限制。

其中，图 1A 中用户所处的真实场景与用户体验的 AR 项目所对应虚拟场景是同一场景，即用户所处的真实场景为音乐厅，用户体验的 AR 项目所对应的虚拟场景为虚拟音乐厅。

- 35 参照图 1A，示例性的，当用户期望在空闲的音乐厅内体验 AR 音乐会时，可以佩戴 AR 设备；接着，用户可以在 AR 设备中选取 AR 音乐厅作为期望体验的 AR 项目，以及选取期望体验的音乐文件并执行播放操作。随后，AR 设备可以对用户所选取音乐文件进行处理如渲染，得到渲染后的音频信号（例如可以是空间音频信号（即双耳渲染信号），本申请以渲染后的音频信号为双耳渲染信号为例进行说明）并通过耳机播放该双耳渲染信号；以及 AR 设备还可以通过显示屏显示演奏画面，使得用户通过显示屏看到演奏者在音乐厅舞台上的对应位置演奏。这样，通过视觉和听觉结合，使得用户可以沉浸式体验音乐会。

图 1B 为示例性示出的应用场景示意图。在图 1B 中，示出的是用户使用虚拟现实（Virtual Reality，VR）设备体验 VR 电影院的应用场景。应该理解的是，用户还可以使用 VR 设备体验其他 VR 项目如 VR 音乐厅、VR 科学讲堂等等，本申请对此不作限制。

其中，图 1B 中用户所处的真实场景与用户体验的 VR 项目所对应的虚拟场景不是同一场景，即用户所处的真实场景为客厅，用户体验的 VR 项目所对应的虚拟场景为虚拟电影院。

参照图 1B，示例性的，当用户期望在客厅体验 VR 影院时，可以佩戴 VR 设备；接着，用户可以在 VR 设备中选取 VR 电影院作为期望体验的 VR 项目，以及选取期望体验的电影文件并执行播放操作。随后，VR 设备可以对电影文件中所包含的音频文件进行处理如渲染，得到双耳渲染信号并通过耳机播放该双耳渲染信号；以及 VR 设备还可以在显示屏显示电影画面；这样，通过视觉和听觉结合，使得用户可以沉浸式体验电影。

图 1C 为示例性示出的应用场景示意图。在图 1C 中，示出的是用户使用（手机+耳机）体验音乐会的应用场景。

应该理解的是，用户还可以使用（手机+耳机）体验其他听音项目如相声、电影等，本申请对此不作限制。此外，手机仅是本申请的一个示例，本申请还可以采用平板电脑、笔记本电脑、个人计算机、智能手表等终端设备+耳机体验听音。

其中，图 1C 中用户所处的真实场景与用户体验的听音项目所对应的虚拟场景是同一场景，即用户所处的真实场景为音乐厅，用户体验的听音项目所对应的虚拟场景为虚拟音乐厅。

参照图 1C，示例性的，当用户期望在空闲的音乐厅内体验音乐会时，可以佩戴耳机并将耳机与手机连接；接着，用户可以在手机中选取音乐会作为期望体验的听音项目，以及选取期望体验的音乐文件并执行播放操作。随后，手机可以对用户所选取音乐文件进行处理如渲染，得到双耳渲染信号；接着，可以将双耳渲染信号发送至耳机播放；以及，手机还可以播放音乐会演奏视频。这样，用户可以在音乐厅使用手机沉浸式体验音乐会。

图 1D 为示例性示出的应用场景示意图。在图 1D 中，示出的是用户使用（手机+耳机）体验相声的应用场景。

应该理解的是，用户还可以使用（手机+耳机）体验其他听音项目如音乐会、电影等，本申请对此不作限制。此外，手机仅是本申请的一个示例，本申请还可以采用平板电脑、笔记本电脑、个人计算机、智能手表等终端设备+耳机听音。

其中，图 1D 中用户所处的真实场景与用户体验的听音项目所对应的虚拟场景不是同一场景，即用户所处的真实场景为卧室，用户体验的听音项目所对应的虚拟场景为虚拟演播厅。

参照图 1D，示例性的，当用户期望在卧室体验相声时，可以佩戴耳机并将耳机与手机连接；接着，用户可以在手机中选取相声作为期望体验的听音项目，以及选取期望体验的相声文件并执行播放操作。随后，手机可以对用户所选取相声文件进行处理如渲染，得到双耳渲染信号；接着，可以将双耳渲染信号发送至耳机播放；以及，手机还可以播放相声视频。这样，用户可以在卧室使用手机沉浸式体验相声。

为了便于说明，将图 1A~图 1D 中的终端设备称为第一电子设备。

其中，图 1A~图 1D 中第一电子设备（也可以称为解码端）中的音乐的音乐文件、电影文件中所包含的音频文件、相声文件均是从第二电子设备（也可以称为编码端）获取的，且是编码后的场景音频信号，即码流。进而，第一电子设备需要对码流进行解码，得到重建音频信号后，再对重建音频信进行渲染，得到双耳渲染信号。

示例性的，可以先在房间（也就是真实场景）中部署若干声源以及多点 HOA 麦克风；接着，由声源播放音频信号，由多点 HOA 麦克风采集房间内的声音，得到多点 HOA 信号；此时，多点 HOA 信号可以用于描述真实场景中声源的声场；其中，多点 HOA 麦克风中的任意两个 HOA 麦克风之间的距离相等。这样，第二电子设备可以从多点 HOA 麦克风获取多点 HOA 信号；接着，对多点 HOA 信号进行编码以得到码流，发送至第一电子设备。

此外，针对图 1A 中的场景和图 1B 中的场景，第二电子设备还可以基于真实场景进行建模，得到虚拟场景。之后，第二电子设备可以对虚拟场景进行编码以得到码流，发送至第一电子设备。应该理解的是，也可以是由第三电子设备进行建模，并将编码虚拟场景所得到的码流发送至第一电子设备或第二电子设备。

其中，真实场景是指客观存在的场景，例如，电影院、音乐厅、报告厅等；虚拟场景是采用虚拟技术

构建的场景，并不是客观存在的场景，如虚拟电影院、虚拟音乐厅、虚拟报告厅等。

图 2A 为示例性示出的场景音频信号的压缩框架的示意图。

参照图 2A，示例性的，第一电子设备可以包括第一音频采集模块、第一场景音频编码模块、第一信道编码模块、第一信道解码模块、第一场景音频解码模块、第一渲染模块和第一音频回放模块。应该理解的是，第一电子设备可以包括比图 2A 所示的更多或更少的模块，本申请对此不作限制。其中，第一渲染模块也可以称为第一渲染器，第一场景音频编码模块也可以称为第一音频编码器，第一场景音频解码模块也可以称为第一音频解码器。

参照图 2A，示例性的，第二电子设备可以包括第二音频采集模块、第二场景音频编码模块、第二信道编码模块、第二信道解码模块、第二场景音频解码模块、第二渲染模块和第二音频回放模块。应该理解的是，第二电子设备可以包括比图 2A 所示的更多或更少的模块，本申请对此不作限制。其中，第二渲染模块也可以称为第二渲染器，第二场景音频编码模块也可以称为第二音频编码器，第二场景音频解码模块也可以称为第二音频解码器。

示例性的，第一电子设备编码并传输场景音频信号至第二电子设备，由第二电子设备解码以及音频回放的过程可以如下：第一音频采集模块（如与第一电子设备连接的 HOA 麦克风）可以进行音频采集，输出场景音频信号至第一场景音频编码模块。接着，第一场景音频编码模块可以对场景音频信号进行编码，输出码流至第一信道编码模块。之后，第一信道编码模块可以对码流进行信道编码，并将信道编码后的码流通过无线或有线网络通信设备传输到第二电子设备。然后，第二电子设备的第二信道解码模块可以对接收到的数据进行信道解码，以得到码流并将码流输出至第二场景音频解码模块。接着，第二场景音频解码模块可以对该码流进行解码，以得到重建音频信号；然后将该重建音频信号输出至第二渲染模块，由第二渲染模块对重建场景信号进行渲染，得到双耳渲染信号；接着将双耳渲染信号发送至第二音频回放模块，由第二音频回放模块进行音频回放。

应该理解的是，第二电子设备编码并传输场景音频信号至第一电子设备，由第一电子设备解码以及音频回放的过程，与上述第一电子设备传输场景音频信号至第二电子设备，由第二电子设备进行音频回放的过程类似，在此不再赘述。

示例性的，第一电子设备和第二电子设备均可以包括但不限于：个人计算机、计算机工作站、智能手机、平板电脑、服务器、智能摄像头、智能汽车或其他类型蜂窝电话、媒体消费设备、可穿戴设备、机顶盒、游戏机等。

示例性的，本申请具体可以应用于 VR/AR 场景。一种可能的方式中，第一电子设备为服务器，第二电子设备为 VR/AR 设备。一种可能的方式中，第二电子设备为服务器，第一电子设备为 VR/AR 设备。

需要说明的是，当第一电子设备、第二电子设备为 VR/AR 设备时，第一音频回放模块和第二音频回放模块可以是 VR/AR 设备的内置耳机；当第一电子设备、第二电子设备是除 VR/AR 设备之外的其他终端设备时，第一音频回放模块和第二音频回放模块可以是其他终端设备外接的耳机。

示例性的，当由第一电子设备编码场景音频信号，第二电子设备解码得到重建音频信号时，第一电子设备可以称为编码端，第二电子设备可以称为解码端。当由第二电子设备编码场景音频信号，第一电子设备解码得到重建音频信号时，第二电子设备可以称为编码端，第一电子设备可以称为解码端。

图 2B 为示例性示出的场景音频信号的转码框架示意图。

参照图 2B (1)，示例性的，无线或核心网设备可以包括：信道解码模块、其他音频解码模块、场景音频编码模块和信道编码模块。其中，无线或核心网设备可以用于音频转码。

示例性的，图 2B (1) 的具体应用场景可以是：在第一电子设备未设有场景音频编码模块，仅设有其他音频编码模块；而第二电子设备仅设有场景音频解码模块，未设有其他音频解码模块的情况下，为了实现第二电子设备能够解码并回放第一电子设备采用其他音频编码模块编码场景音频信号，可以使用无线或核心网设备进行转码。

具体的，第一电子设备采用其他音频编码模块对场景音频信号进行编码，得到第一码流；并将第一码流进行信道编码后发送给无线或核心网设备。接着，无线或核心网设备的信道解码模块可以进行信道解码，将信道解码出的第一码流输出至其他音频解码模块。之后，其他音频解码模块对第一码流进行解码，得到场景音频信号并将场景音频信号输出至场景音频编码模块。然后，场景音频编码模块可以对场景音频信号进行编码，以得到第二码流并将第二码流输出至信道编码模块，由信道编码模块对第二码流进行信道编码

后，发送至第二电子设备。这样，第二电子设备可以调用场景音频解码模块，对信道解码得到第二码流进行解码，得到重建音频信号；后续即可对重建音频信号进行音频回放。

参照图 2B (2)，示例性的，无线或核心网设备可以包括：信道解码模块、场景音频解码模块、其他音频编码模块和信道编码模块。其中，无线或核心网设备可以用于音频转码。

5 示例性的，图 2B (2) 的具体应用场景可以是：在第一电子设备仅设有场景音频编码模块，未设有其他音频编码模块；而第二电子设备未设有场景音频解码模块，仅设有其他音频解码模块的情况下，为了实现第二电子设备能够解码并回放第一电子设备采用场景音频编码模块编码场景音频信号，可以使用无线或核心网设备进行转码。

10 具体的，第一电子设备采用场景音频编码模块对场景音频信号进行编码，得到第一码流；并将第一码流进行信道编码后发送给无线或核心网设备。接着，无线或核心网设备的信道解码模块可以进行信道解码，将信道解码出的第一码流输出至场景音频解码模块。之后，场景音频解码模块对第一码流进行解码，得到场景音频信号并将场景音频信号输出至其他音频编码模块。然后，其他音频编码模块可以对场景音频信号进行编码，以得到第二码流并将第二码流输出至信道编码模块，由信道编码模块对第二码流进行信道编码后，发送至第二电子设备。这样，第二电子设备可以调用其他音频解码模块，对信道解码得到第二码流进行解码，得到重建音频信号；后续即可对重建音频信号进行音频回放。

基于图 2A 和图 2B，对渲染得到双耳渲染信号的过程进行说明。

图 2C 为示例性示出的渲染过程示意图。在图 2C 中示出的是渲染器（可以是第一渲染器，也可以是第二渲染器）的渲染过程示意图。

20 参照图 2C，解码端根据码流获取重建音频信号之后，可以将重建音频信号、场景音频信号的元数据、用户位置和用户姿态输入至渲染器中，由渲染器根据场景音频信号的元数据、用户位置和用户姿态，对重建音频信号进行渲染，得到双耳渲染信号。其中，用户姿态可以是指用户头部姿态。

示例性的，当解码端为 AR/VR 设备时，AR/VR 设备中的传感器可以获得用户位置和用户姿态。当解码端为其他终端设备时，可以由解码端中的定位模块获取用户位置，由与解码端连接的耳机获取用户姿态。

25 示例性的，渲染器可以采用虚拟扬声器对象（Virtual Loudspeaker Objects, VLO）算法进行渲染，得到双耳渲染信号。具体地，可以先将重建音频信号，转换为虚拟扬声器信号；具体地，可以参照如下公式 (5)：

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

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

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

30 公式 (5) 中，S 为虚拟扬声器信号，其表达方式如公式 (7) 所示，N 大于 0；也就是说，将重建音频信号，转换为 L 个虚拟扬声器的虚拟扬声器信号。公式 (5) 中，X 为重建音频信号（即 HOA 信号的重建音频信号），其表达方式如公式 (6) 所示。公式 (5) 中，D 为解码矩阵，尺寸为 L*N。

接着，可以根据场景音频信号的元数据、用户位置和用户位姿，对虚拟扬声器信号进行渲染，得到双耳渲染信号；其中，场景音频信号的元数据可以包括第一元数据和第二元数据。生成双耳渲染信号的过程可以参照如下 S11~S16：

S11，从 L 个虚拟扬声器的虚拟扬声器信号中，确定与用户位置距离最近的 Y 个虚拟扬声器信号 S_n...S_m。

其中，Y=m-n。

S12，根据第二元数据，确定两点 HOA 麦克风之间的距离 d_l。

40 一种可能的方式中，第二元数据为场景中多点 HOA 麦克风的位置信息，此时，可以根据场景中多点 HOA 麦克风的位置信息，计算任意两点 HOA 麦克风之间的距离 d_l。

一种可能的方式中，第二元数据为多点 HOA 麦克风中的任意两点 HOA 麦克风之间的距离 d_l。

图 2D 为示例性示出的场景示意图。

示例性的，在图 2D 中示出了场景中的两点 HOA 麦克风，且 $Y=4$ ；也就是说，针对一点 HOA 信号（即一点 HOA 麦克风接收到的声源（吉他）的音频信号），可以采用 4 个虚拟扬声器的虚拟扬声器信号来表示。其中，两点 HOA 麦克风之间的距离 $d1$ 如图 2D 所示。

5 S13，根据第一元数据和两点 HOA 麦克风之间的距离 $d1$ ，确定虚拟扬声器半径 $r1$ 。

示例性的，虚拟扬声器半径 $r1$ 可以如图 6B 所示，虚拟扬声器半径 $r1$ 计算公式可以如下公式 (8) 所示：

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

公式 (8) 中， k 表示第一元数据。也就是说， k 不变时， d 越大， r 越大。

10 图 2E 为示例性示出的场景示意图。图 2E 可以参照图 2D 的描述，在此不再赘述。其中，图 2E 和图 2D 的区别在于，两点 HOA 麦克风之间的距离 $d1$ 不同；图 2E 中两点 HOA 麦克风之间的距离 $d1$ 大于图 2D 中两点 HOA 麦克风之间的距离 $d1$ ，对应的，图 2E 中虚拟扬声器半径 $r1$ 也就更大。

S14，从头相关脉冲响应（Head Related Impulse Response, HRIR）库中，根据用户姿态和虚拟扬声器半径 $r1$ ，选择与虚拟扬声器信号 $S_n \cdots S_m$ 所对应的头相关脉冲响应 $HRIR_n \cdots HRIR_m$ 。

15 其中， S_n 与 $HRIR_n$ 对应， S_{n+1} 与 $HRIR_{n+1}$ 对应，……，以此类推， S_m 与 $HRIR_m$ 对应。

S15，根据 $S_n \cdots S_m$ 和对应的 $HRIR_n \cdots HRIR_m$ 进行计算，得到 Y 个虚拟扬声器中每个虚拟扬声器的双耳渲染信号。

S16，将 Y 个虚拟扬声器中每个虚拟扬声器的双耳渲染信号相加，得到双耳渲染信号。

20 参照图 2D 或图 2E，将 2 个虚拟扬声器的双耳渲染信号相加，可以得到用户所在位置处的双耳渲染信号。

在本申请实施例中，可以由编码端获取场景音频信号的元数据并写入码流中，这样，解码端可以从码流中解析出场景音频信号的元数据，将场景音频信号的元数据作用于后续渲染得到双耳渲染信号的过程中。其中，场景音频信号的元数据的获取过程以及编解码过程可以参照如下实施例。

25 图 3A 为示例性示出的编码过程示意图。

S301，获取场景音频信号和场景音频信号的元数据。

示例性的，编码端可以从多点 HOA 麦克风获取多点 HOA 信号，作为场景音频信号。

示例性的，本申请中场景音频信号的元数据包括场景音频信号的相关信息，场景音频信号的元数据作用于渲染以得到双耳渲染信号的过程。

30 示例性的，S301 的一种可能的方式中，编码器可以获取场景的场景描述信息，接着，根据场景的场景描述信息，确定场景音频信号的元数据；具体过程在后续进行说明。

示例性的，场景描述信息可以用于描述场景的信息，例如可以包括但不限于：真实场景中多点 HOA 麦克风的标识、位置信息、特性和方向等，真实场景包含的对象的坐标、材质等，真实场景中声源的标识、位置信息、特性和方向等，真实场景的房间临界距离（房间临界距离是指声源在室内（真实场景中）稳定地辐射声波时，在某一方向上直达声能量密度与混响声能量密度相等的点到声源中心的距离、也称扩散场距离。当声源辐射指向性为球形时，各方向的临界距离相等，这时也称混响半径），声学混响参数（如直混比（Direct-to-Reverberation energy Ratio, DRR）、混响时间（Reverberation Time）（如 T_{20} （能量衰减 20dB 的时间）、 T_{30} 、 T_{60} 等）、清晰度（Clarity）（如 C_{50} （50ms 前后的声能比例）、 C_{80} 等）等等，本申请对此不作限制。

40 示例性的，S301 的一种可能的方式中，可以获取用户输入的场景音频信号的元数据；也就是说，场景音频信号的元数据可以由用户自定义设置。

在图 3A 的实施例中，获取的场景音频信号的元数据仅包括第一元数据，其中，第一元数据可以用于与多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，该麦克风可以为 HOA 麦克风。

S302，编码场景音频信号和场景音频信号的元数据，得到码流。

45 场景音频信号的编码过程可以如下：

第一种可能的编码方式中，可以直接对场景音频信号进行编码，得到场景音频信号的编码数据并写入码流。

第二种可能的编码方式中，可以先根据第一场景音频信号从预设的虚拟扬声器集合中选择出目标虚拟扬声器；接着，根据第一场景音频信号和目标虚拟扬声器的属性信息生成第一虚拟扬声器信号；之后，使用目标虚拟扬声器的属性信息和第一虚拟扬声器信号，确定第二场景音频信号；随后，根据第一场景音频信号和第二场景音频信号生成残差信号；然后，对第一虚拟扬声器信号和残差信号进行编码，得到场景音频信号的编码数据并写入码流。其中，第一场景音频信号是指原始待编码的场景音频信号。

第三种可能的编码方式中，可以基于场景音频信号，确定目标虚拟扬声器的属性信息，场景音频信号包括L个通道的音频信号，L为正整数；编码场景音频信号中第一音频信号和目标虚拟扬声器的属性信息，得到场景音频信号的编码数据并写入码流；其中，第一音频信号为场景音频信号中G个通道的音频信号，G为小于或等于L的正整数。

示例性的，在图3A的实施例中，本申请可以预先在码流的原始语法结构（或者称为高层语法元素表）中，添加场景音频信号的元数据的语法元素和对应的描述信息。这样，编码端可以根据码流的语法结构中场景音频信号的元数据的语法元素和对应的描述信息，将场景音频信号的元数据，写入码流中。

例如，码流的语法结构可以参照如下表1所示：

表 1

Syntax	No. of bits	Mnemonic
语法元素 1		
{		
语法元素 2;	1	bslbf
语法元素 3;	3	bslbf
.....		
语法元素 n;	4	uimsbf
}		
.....		
语法元素 n+h		
distanceFactor	8	uimsbf

参照表1，表1中Syntax表示语法元素，No. of bits表示语法元素所标识的数据的比特位数，Mnemonic表示语法元素备注或解释。

参照表1，示例性的，码流的语法结构中包括多个语法元素：语法元素1、语法元素2、语法元素3、.....、语法元素n、.....、语法元素n+h，以及本申请新增的语法元素“distanceFactor”；其中，“distanceFactor”为场景音频信号的元数据的语法元素，n和h为正整数。

继续参照表1，表1中第四行中的“1”表示码流中语法元素2所标识的数据采用1比特表示，表1中第四行中的“bslbf”表示码流中语法元素2所标识的数据是比特串，且左位在先。

表1中第五行中的“3”表示码流中语法元素3所标识的数据采用3比特表示，表1中第五行中的“bslbf”表示码流中语法元素3所标识的数据是比特串，且左位在先。

表1中第七行中的“4”表示码流中语法元素n所标识的数据采用4比特表示，表1中第七行中的“uimsbf”表示码流中语法元素n所标识的数据是无符号整数，且高位在先。

表1中最后一行中的“8”表示码流中场景音频信号的元数据可以采用8比特表示；表1中最后一行中的“uimsbf”表示码流中场景音频信号的元数据是无符号整数，且高位在先。应该理解的是，表1仅是本申请的一个示例，场景音频信号的元数据还可以采用其他数量的比特位表示如16位，本申请对此不作限制；此外，场景音频信号的元数据的语法元素除了可以采用“distanceFactor”表示之外，还可以采用其他字符串表示，本申请对此也不作限制。

示例性的，若码流的语法结构如表 1 所示，则在执行 S302 的过程中，可以将场景音频信号的元数据，转换为 8 个 bit 的二进制数；接着，将该 8 个 bit 的二进制数，写入码流中语法元素 n+h 所标识的数据之后。

示例性的，S301~S302 可以由编码器执行。

图 3B 为示例性示出的码流结构示意图。其中，图 3A 实施例码流的码流结构可以如图 3B 所示，即场景音频信号的元数据位于场景音频信号的编码数据之后。

应该理解的是，本申请的码流中，场景音频信号的元数据也可以位于场景音频信号的编码数据之前，本申请对此不作限制；且本申请的码流中还可以包括其他数据，本申请对此不作限制。

需要说明的是，本申请的码流中，场景音频信号的元数据与场景音频信号的编码数据可以相邻也可以不相邻，本申请对此不作限制。

图 4A 为示例性示出的解码过程示意图。图 4A 的解码过程与图 3A 的编码过程对应。

S401，接收码流。

S402，从码流中解析出场景音频信号的元数据。

S403，根据码流，获取场景音频信号的重建音频信号。

其中，S401~S402 由解码器执行。本申请不限制 S402 和 S403 的执行顺序。

示例性的，根据码流，获取场景音频信号的重建音频信号的过程可以如下：

第一种解码方式（对应上述的第一种编码方式）中，可以直接从码流中解码出场景音频信号的重建音频信号。此时，S403 可以由解码器执行。

第二种解码方式（对应上述的第二种编码方式）中，可以从码流中解码出虚拟扬声器信号和残差信号；接着，根据目标虚拟扬声器的属性信息、残差信号以及虚拟扬声器信号，得到场景音频信号的重建音频信号。此时，解码操作可以由解码器执行，其他步骤可以由其他模块执行。

第三种解码方式（对应上述的第三种编码方式）中，可以解码码流，以得到第一重建信号和目标虚拟扬声器的属性信息；接着，基于属性信息和第一重建信号，生成目标虚拟扬声器对应的虚拟扬声器信号；之后，基于属性信息和虚拟扬声器信号进行重建，以得到场景音频信号的重建音频信号。此时，解码操作可以由解码器执行，其他步骤可以由其他模块执行。

示例性的，解码端可以根据码流的语法结构中，场音频信号的元数据的语法元素和对应的描述信息，从码流中解析出场景音频信号的元数据；其中，在图 4A 的实施例中，场景音频信号的元数据可以仅包括第一元数据。之后，解码端可以从码流中解析出的第一元数据和根据码流获取的重建音频信号进行渲染，以得到双耳渲染信号。这种情况中，解码端可以从其他设备获取第二元数据或者在本地生成第二元数据，本申请对此不作限制；以及，渲染过程中虚拟扬声器的数量 Y 可以是用户自定义设置的值，也可以是默认值，本申请对此也不作限制。

图 4B 为示例性示出的双耳渲染信号的生成过程示意图。图 4B 是以图 3A 和图 4A 为基础示出的。

参照图 4B，示例性的，编码端编码场景音频信号和场景音频信号的元数据得到码流之后，将码流传输至解码端。解码端接收到码流之后，可以根据码流获取场景音频信号的重建音频信号，以及从码流中解析出场景音频信号的元数据；接着，根据场景音频信号的元数据、用户位置和用户姿态，对场景音频信号的重建音频信号进行渲染，得到双耳渲染信号。

示例性的，本申请还可以将用于指示渲染复杂度的第一标识，写入码流；其中，渲染复杂度用于在渲染过程中确定虚拟扬声器的数量 Y；也就是说，解码端可以根据场景音频信号的元数据、第一标识指示的渲染复杂度、用户位置和用户位姿，对虚拟扬声器信号进行渲染，得到双耳渲染信号。这样，可以渲染得到音频质量与用户需求、电子设备的计算能力等因素匹配的双耳渲染信号。

图 5A 为示例性示出的编码过程示意图。在图 5A 的实施例中，场景音频信号的元数据是根据场景描述信息确定的，且场景音频信号的元数据可以包括第一元数据和第二元数据。

S501，获取场景音频信号和场景的场景描述信息。

示例性的，编码端可以从本地获取场景的场景描述信息，也可以从第三电子设备获取场景的场景描述信息，本申请对此不作限制。

S502，根据场景的场景描述信息，确定场景音频信号的元数据。

示例性的，场景音频信号的元数据可以包括第一元数据和第二元数据。

一种可能的方式中，第一元数据可以根据房间临界距离和预设数值的乘积值确定；具体地，可以参照如下公式（9）：

$$k = \frac{1.5}{\text{房间临界距离}} \quad (9)$$

其中，公式（9）中的 k 为第一元数据， $k \in [0.5, 1.5]$ ；公式（9）中 1.5 为预设数值。

5 一种可能的方式中，可以根据场景中的声学混响参数确定场景的混响类型；接着，根据场景的混响类型，确定第一元数据。具体地，可以预先针对不同的混响类型，设置对应的第一元数据；例如，针对第一混响类型（对应干声场景或无混响房间），将对应的第一元数据设置为 0.5，针对第二混响类型（对应湿声场景或混响房间），将对应的第一元数据设置为 1.5。这样，当确定场景的混响类型为第一混响类型时，可以将 0.5 作为第一元数据；当确定场景的混响类型为第二混响类型时，可以将 1.5 作为第一元数据。

10 一种可能的方式中，可以将场景描述信息包含的场景中多点 HOA 麦克风的位置信息，作为第二元数据。

一种可能的方式中，可以根据场景描述信息包含的场景中多点 HOA 麦克风的位置信息，确定多点 HOA 麦克风中两点 HOA 麦克风之间的距离；然后，将多点 HOA 麦克风中两点 HOA 麦克风之间的距离，作为第二元数据。

15 示例性的，当第一元数据是用户输入的时，第一元数据的取值范围可以限定为 $k \in [0.25, 4]$ 。

应该理解的是，第一元数据还可以包括其他生成方式，以及场景音频信号的元数据还可以包括其他元数据，本申请对此不作限制。

S503，编码场景音频信号和场景音频信号的元数据，得到码流。

示例性的，S503 可以参照上述 S302 的描述，在此不再赘述。

20 S504，将第一标识写入码流。

示例性的，可以预先设置多个渲染复杂度等级，可以如表 2 所示：

表 2

渲染复杂度等级	虚拟扬声器数量	卷积次数	渲染复杂度	音频质量
0	$(N+1)^2$	$2(N+1)^2$	取决于 HOA 信号	取决于 HOA 信号
1	4	8	Lowest	Lowest
2	9	18	Low	Low
3	16	32	High	High
4	25	50	Highest	Highest

表 2 中的虚拟扬声器数量即为 Y。

25 参照表 2，示例性的，可以将渲染复杂度划分为 5 个等级，其中，等级 1~等级 4 中，虚拟扬声器数量越多时，渲染过程涉及到的卷积次数越多，渲染的复杂度越高，音频质量越好。

参照表 2，示例性的，针对等级 0，N 为 HOA 信号的阶数，虚拟扬声器数量为 $(N+1)^2$ ，卷积次数为 $2(N+1)^2$ ；渲染复杂度和音频质量均取决于 HOA 信号。

应该理解的是，表 2 仅是本申请的一个示例，可以按照需求设置渲染复杂度等级，本申请对此不作限制。

30 一种可能的方式中，可以根据解码端的计算能力，从表 2 中的多个渲染复杂度等级中，选取目标渲染复杂度等级；将第一标识的值设置为目标渲染复杂度等级。

一种可能的方式中，用户可以根据表 2 中的多个渲染复杂度等级，输入目标渲染复杂度等级，将第一标识的值设置为用户输入的目标渲染复杂度等级。

35 示例性的，在图 5A 的实施例，本申请可以预先在码流的原始语法结构中，添加场景音频信号的元数据的语法元素和对应的描述信息，以及添加第一标识的语法元素和对应的描述信息。这样，编码端可以

根据码流的语法结构中第一标识的语法元素和对应的描述信息，将第一标识写入码流；以及根据码流的语法结构中场景音频信号的元数据的语法元素和对应的描述信息，将场景音频信号的元数据，写入码流中。

需要说明的是，本申请不限制 S503 和 S504 的执行顺序。

例如，码流的语法结构可以参照如下表 3 所示：

5

表 3

Syntax	No. of bits	Mnemonic
语法元素 1		
{		
语法元素 2;	1	bslbf
语法元素 3;	3	bslbf
.....		
语法元素 n;	4	uimsbf
}		
.....		
语法元素 n+1		
distanceFactor	8	uimsbf
complexityLevel	3	uimsbf

参照表 3，在图 5A 的实施例中，本申请新增的语法元素“distanceFactor”和“complexityLevel”；其中，“distanceFactor”为场景音频信号的元数据的语法元素，“complexityLevel”为第一标识的语法元素。

表 3 中最后一行中的“3”表示码流中第一标识可以采用 3 比特表示；表 3 中最后一行中的“uimsbf”表示码流中第一标识是无符号整数，且高位在先。应该理解的是，表 3 仅是本申请的一个示例，第一标识还可以采用其他数量的比特位表示如 4 位，本申请对此不作限制；此外，第一标识的语法元素除了可以采用“complexityLevel”表示之外，还可以采用其他字符串表示，本申请对此也不作限制。

10

其中，表 3 中其他语法元素的解释可以参照表 1 中的描述，在此不再赘述。

示例性的，若码流的语法结构如表 3 所示，则在执行 S503 的过程中，可以将场景音频信号的元数据，转换为 8 个 bit 的二进制数；接着，将该 8 个 bit 的二进制数，写入码流中语法元素 n+1 所标识的数据之后。以及在执行 S504 的过程中，可以将第一标识转换为 3 个 bit 的二进制数；接着，将该 3 个 bit 的二进制数，写入码流中场景音频信号的元数据之后。

15

图 5B 为示例性示出的码流结构示意图。其中，图 5A 实施例中的码流结构可以如图 5B 所示，即场景音频信号的元数据位于场景音频信号的编码数据之后，第一标识位于场景音频信号的元数据之后。

20

应该理解的是，本申请不限制码流中场景音频信号的元数据、场景音频信号的编码数据和第一标识的位置。

应该理解的是，本申请的码流中场景音频信号的元数据、场景音频信号的编码数据和第一标识中的任意两种数据，可以相邻也可以不相邻，本申请对此不作限制。

25

图 6A 为示例性示出的解码过程示意图。图 6A 的解码过程与图 5A 的编码过程对应。

S601，接收码流。

S602，根据码流获取场景音频信号的重建音频信号。

S603，从码流中解析场景音频信号的元数据。

S604，从码流中解析出第一标识。

示例性的，本申请不限制 S602、S603 和 S604 的执行顺序。

示例性的，解码端接收到码流之后，可以根据码流获取场景音频信号的重建音频信号；具体可以参照图 4A 实施例中的描述，在此不再赘述。

示例性的，解码端还可以根据码流的语法结构中场景音频信号的元数据的语法元素和对应的描述信息，从码流中解析出场景音频信号的元数据。以及，解码端还可以根据码流的语法结构中第一标识的语法元素和对应的描述信息，从码流中解析出第一标识。

后续，解码端可以根据场景音频信号的重建信号、第一元数据、第二元数据和第一标识进行渲染，以得到双耳渲染信号。这种情况中，解码端可以从码流中解析出第二元数据和第一元数据；以及，渲染过程中，选取的虚拟扬声器的数量 Y 可以根据第一标识确定的。

图 6B 为示例性示出的双耳渲染信号的生成过程示意图。图 6B 是以图 5A 和图 6A 为基础示出的。

参照图 6B，示例性的，编码端将场景音频信号、场景音频信号的元数据和第一标识编入码流之后，将码流传输至解码端。解码端接收到码流之后，可以根据码流获取场景音频信号的重建音频信号；以及从码流中解析出场景音频信号的元数据和第一标识；接着，根据场景音频信号的元数据、第一标识指示的渲染复杂度等级、用户位置和用户姿态，对场景音频信号的重建音频信号进行渲染，得到双耳渲染信号。

需要说明的是，在图 6B 中，输入至渲染器的可以是第一标识，也可以是第一标识指示的渲染复杂度等级或者第一标识指示的渲染复杂度等级对应的虚拟扬声器的数量 Y；本申请对此不作限制。

示例性的，解码端可能存在多个渲染器（每个渲染器采用的渲染方法不同），其中，这多个渲染器包括：根据场景音频信号的元数据进行渲染的渲染器（例如，采用上述 S11~S16 中的方法进行渲染的渲染器）和根据其他信息进行渲染的渲染器。进而，针对每个渲染器可以设置对应的标识值，例如，针对渲染器 A，设置对应的标识值为“0”；针对渲染器 B，设置对应的标识值为“1”；针对渲染器 C，设置对应的标识值为“2”等等。当确定解码端需要根据场景音频信号的元数据进行渲染时，一种可能的方式中，可以根据场景音频信号的元数据进行渲染的渲染器的标识值（后续称为第一标识值），作为第二标识的标识值；一种可能的方式中，可以将第二标识值（如“0”或“1”，可以按照需求设置）作为第二标识的标识值；然后将第二标识添加在码流中。其中，第二标识值用于指示解码端根据场景音频信号的元数据进行渲染，或者说，第二标识值用于指示码流中包括场景音频信号的元数据。这样，便于解码端确定采用哪个渲染器或哪种渲染算法进行渲染。

图 7A 为示例性示出的编码过程示意图。需要说明的是，图 7A 是在图 3A 的基础上示出的，在图 7A 的实施例中，场景音频信号的元数据是根据场景描述信息确定的。

S701，获取场景音频信号。

S702，编码场景音频信号，得到码流。

示例性的，S701~S702 可以参照上述描述，在此不再赘述。

S703，将第二标识写入码流。

其中，上述的第一标识值和第二标识值，均可以称为第一预设标识值。也就是说，当确定解码端需要使用场景音频信号的元数据进行渲染时，可以将第二标识的值设置为第一预设标识值，然后将第一预设标识值写入码流。否则，可以将第二标识的值设置为其他标识值（其他标识值与第一预设标识值不同），然后将第一预设标识值写入码流。

S704，获取场景的场景描述信息。

S705，根据场景的场景描述信息，确定场景音频信号的元数据。

S706，将场景音频信号的元数据，写入码流。

示例性的，S704~S706 可以参照上述 S501~S503 的描述，在此不再赘述。

示例性的，在图 7A 的实施例中，本申请可以预先在码流的原始语法结构中，添加场景音频信号的元数据的语法元素和对应的描述信息，以及添加第二标识的语法元素和对应的描述信息。这样，编码端可以根据码流的语法结构中第二标识的语法元素和对应的描述信息，将第二标识写入码流中。

例如，码流的语法结构可以参照如下表 4 所示：

表 4

Syntax	No. of bits	Mnemonic
--------	-------------	----------

语法元素 1		
{		
语法元素 2;	1	bslbf
语法元素 3;	3	bslbf
.....		
语法元素 n;	4	uimsbf
}		
.....		
语法元素 n+1		
hoaGroupHasLowProfileConfig;	1	bslbf
if (hoaGroupHasLowProfileConfig) {		
distanceFactor;	3	uimsbf
}		

参照表 4，在图 7A 的实施例中，本申请新增的语法元素“distanceFactor”和“hoaGroupHasLowProfileConfig”；其中，“distanceFactor”为场景音频信号的元数据的语法元素，“hoaGroupHasLowProfileConfig”为第二标识的语法元素。

表 4 中倒数第四行中的“1”表示码流中第二标识可以采用 1 比特表示；表 4 中倒数第四行中的“bslbf”表示码流中第二标识是比特串，且左位在先。应该理解的是，表 4 仅是本申请的一个示例，第二标识还可以采用其他数量的比特位表示如 2 位，本申请对此不作限制；此外，第二标识的语法元素也可以采用其字符串表示，本申请对此不作限制。

其中，表 4 中其他语法元素的解释可以参照表 1 中的描述，在此不再赘述。

示例性的，若码流的语法结构如表 4 所示，则在执行 S703 的过程中，可以将第二标识，转换为 1 个 bit 的二进制数；接着，将该 1 个 bit 的二进制数，写入码流中语法元素 n+1 所标识的数据之后。以及在执行 S706 的过程中，可以将场景音频信号的元数据转换为 8 个 bit 的二进制数；接着，将该 8 个 bit 的二进制数，写入码流中第二标识之后。当码流的语法结构如表 4 所示时，第二标识的值为“1”。

一种可能的方式中，还可以将根据场景音频信号的元数据进行渲染的渲染器的第一标识值，作为第二标识的值。假设，根据场景音频信号的元数据进行渲染的渲染器的第一标识值为“3”则，码流的语法结构可以如表 5 所示：

例如，码流的语法结构可以参照如下表 5 所示：

表 5

Syntax	No. of bits	Mnemonic
语法元素 1		
{		
语法元素 2;	1	bslbf
语法元素 3;	3	bslbf

.....		
语法元素 n;	4	uimsbf
}		
.....		
语法元素 n+1		
hoaGroupHasLowProfileConfig;	2	bslbf
if (hoaGroupHasLowProfileConfig==3) {		
distanceFactor;	3	uimsbf
}		

图 7B 为示例性示出的码流结构示意图。其中，图 7A 实施例中的码流结构可以如图 7B 所示，即第二标识位于场景音频信号的编码数据之后，场景音频信号的元数据位于第二标识之后。

应该理解的是，本申请不限制码流中，场景音频信号的元数据、场景音频信号的编码数据和第二标识的位置。

5 应该理解的是，本申请的码流中，场景音频信号的元数据、场景音频信号的编码数据和第二标识中的任意两种数据，可以相邻也可以不相邻，本申请对此不作限制。

图 8 为示例性示出的解码过程示意图。图 8 的解码过程与图 7A 的编码过程对应。

S801，接收码流。

10 S802，根据码流获取场景音频信号的重建音频信号。

示例性的，S801~S802 可以参照上文中的描述，在此不再赘述。

S803，从码流中解析第二标识。

S804，判断第二标识的值是否为第一预设标识值。

15 示例性的，可以先从码流中解析出第二标识；接着，判断第二标识是否为第一预设标识值，来判断是否可以从码流中解析出场景音频信号的元数据。

示例性的，当第二标识的值为第一预设标识值时，说明码流中包含场景音频信号的元数据，此时，可以执行 S805。当第二标识的值是其他标识值时，说明码流中不包含场景音频信号的元数据，此时，从对码流中第二标识之后的字段进行解码。

20 示例性的，当码流的码流结构如表 4 所示时，则可以判断第二标识的值是否为 1，当第二标识的值为 1 时，说明码流中包含场景音频信号的元数据；当第二标识的值不为 1 时，说明码流中不包含场景音频信号的元数据。

示例性的，当码流的码流结构如表 5 所示时，可以判断第二标识的值是否为 3，当第二标识的值为 3 时，说明码流中包含场景音频信号的元数据；当第二标识的值不为 3 时，说明码流中不包含场景音频信号的元数据。

25 S805，从码流中解析出场景音频信号的元数据。

30 示例性的，解码端可能存在多个渲染器（每个渲染器采用的渲染方法不同），其中，这多个渲染器包括：根据场景音频信号的元数据和第一标识进行渲染的渲染器（例如，采用上述 S11~S16 中的方法进行渲染的渲染器）和根据其他信息进行渲染的渲染器。进而，针对每个渲染器可以设置对应的标识值，例如，针对渲染器 A，设置对应的标识值为“0”；针对渲染器 B，设置对应的标识值为“1”；针对渲染器 C，设置对应的标识值为“2”等等。当确定解码端需要根据场景音频信号的元数据和第一标识进行渲染时，一种可能的方式中，可以将根据场景音频信号的元数据和第一标识进行渲染的渲染器的标识值（后续称为

第一标识值)，作为第二标识的标识值；一种可能的方式中，可以将第二标识值（如“0”或“1”，可以按照需求设置）作为第二标识的标识值；然后将第二标识添加在码流中。其中，第二标识值用于指示解码端根据场景音频信号的元数据和第一标识进行渲染，或者说，第二标识值用于指示码流中包括场景音频信号的元数据和第一标识。这样，便于解码端确定采用哪个渲染器或哪种渲染算法进行渲染。

5 图 9A 为示例性示出的编码过程示意图。需要说明的是，图 9A 是在图 5A 的基础上示出的，在图 9A 的实施例中，场景音频信号的元数据是根据场景描述信息确定的。

 S901，获取场景音频信号。

 S902，编码场景音频信号，得到码流。

 示例性的，S901~S902 可以参照上文中的描述，在此不再赘述。

10 S903，将第二标识写入码流。

 S904，获取场景的场景描述信息。

 S905，根据场景的场景描述信息，确定场景音频信号的元数据。

 S906，将场景音频信号的元数据，写入码流。

 S907，将第一标识，写入码流。

15 示例性的，本申请不限制 S906 和 S907 的执行顺序。

 示例性的，在图 9A 的实施例中，本申请可以预先在码流的原始语法结构中，添加场景音频信号的元数据的语法元素和对应的描述信息，添加第二标识的语法元素和对应的描述信息，以及添加第一标识的语法元素和对应的描述信息。

 例如，码流的语法结构可以参照如下表 6 所示：

20 表 6

Syntax	No. of bits	Mnemonic
语法元素 1		
{		
语法元素 2;	1	bslbf
语法元素 3;	3	bslbf
.....		
语法元素 n;	4	uimsbf
}		
.....		
语法元素 n+1		
hoaGroupHasLowProfileConfig;	1	bslbf
if (hoaGroupHasLowProfileConfig) {		
distanceFactor;	3	uimsbf
complexityLevel;	3	uimsbf
}		

图 9B 为示例性示出的码流结构示意图。图 9B 为示例性示出的码流结构示意图。其中，图 9A 实施例中的码流结构可以如图 9B 所示，即第二标识位于场景音频信号的编码数据之后，场景音频信号的元数据位于第二标识之后，第一标识位于场景音频信号的元数据之后。

应该理解的是，本申请不限制码流中，场景音频信号的元数据、场景音频信号的编码数据、第一标识和第二标识的位置。

应该理解的是，本申请的码流中，场景音频信号的元数据、场景音频信号的编码数据、第一标识和第二标识中的任意两种数据，可以相邻也可以不相邻，本申请对此不作限制。

5

图 10 为示例性示出的解码过程示意图。图 10 的解码过程与图 9A 的编码过程对应。

S1001，接收码流。

S1002，根据码流获取场景音频信号的重建音频信号。

S1003，从码流中解析第二标识。

10 S1004，判断第二标识的值是否为第一预设标识值。

S1005，从码流中解析出场景音频信号的元数据。

S1006，从码流中解析第一标识。

示例性的，S1001~S1006 可以参照上述实施例的描述，在此不再赘述。

15 需要说明的是，图 5A~图 10 的实施例中，在编码过程中，还可以将第三标识写入码流，第三标识可以用于指示码流中是否包括第一标识；这种情况中，第二标识可以仅用于指示码流中是否包括场景音频信号的元数据。示例性的，当在码流中写入第一标识时，可以将第三标识的值设置为第二预设标识值，然后将第二预设标识值写入码流。

20 对应的，本申请可以预先在码流的原始语法结构中，添加场景音频信号的元数据的语法元素和对应的描述信息，添加第二标识的语法元素和对应的描述信息，添加第一标识的语法元素和对应的描述信息，以及添加第三标识的语法元素和对应的描述信息。

例如，码流的语法结构可以参照如下表 7 所示：

表 7

Syntax	No. of bits	Mnemonic
语法元素 1		
{		
语法元素 2;	1	bslbf
语法元素 3;	3	bslbf
.....		
语法元素 n;	4	uimsbf
}		
.....		
语法元素 n+1		
hoaGroupHasLowProfileConfig;	1	bslbf
if (hoaGroupHasLowProfileConfig) {		
distanceFactor;	3	uimsbf
}		
hoaGroupHascomplexityConfig;		

if (hoaGroupHascomplexityConfig) {		
complexityLevel;	3	uimsbf
}		

在表 7 中，“hoaGroupHascomplexityConfig”为第三标识的语法元素；且第二预设标识值为“1”。

这样，解码端可以从码流中解码出第三标识；当确定第三标识的值为第二预设标识值时，可以确定码流中包括第一标识；然后从码流中解码出第一标识。

5 示例性的，本申请还可以仅将第一标识写入码流中，这种情况下，解码端可以从其他电子设备获取场景音频信号的元数据或在本地生成场景音频信号的元数据。

图 11 为示例性示出的编码过程示意图。

S1101，获取场景音频信号；

10 S1102，编码场景音频信号和第一标识，以得到码流；其中，第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量，场景音频信号的重建音频信号用于生成虚拟扬声器信号，虚拟扬声器信号和虚拟扬声器的数量用于渲染以得到渲染后的音频信号。

S1101~S1102，可以参照上述的描述，在此不再赘述。

其中，可以将第一标识的语法元素和对应的描述信息添加至码流的语法结构中。

例如，码流的语法结构可以如下表 8 所示：

15

表 8

Syntax	No. of bits	Mnemonic
语法元素 1		
{		
语法元素 2;	1	bslbf
语法元素 3;	3	bslbf
.....		
语法元素 n;	4	uimsbf
}		
.....		
语法元素 n+h		
complexityLevel;	3	uimsbf

图 12 为示例性示出的解码过程示意图。图 12 的解码过程与图 11 的编码过程对应。

S1201，接收码流；

S1202，根据码流获取场景音频信号的重建音频信号。

20 S1203，从码流中解析出第一标识，其中，第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量，重建音频信号用于生成虚拟扬声器信号，虚拟扬声器信号和虚拟扬声器的数量用于渲染以得到渲染后的音频信号。

示例性的，S1201~S1203，可以参照上述的描述，在此不再赘述。

此外，在图 11 和图 12 的基础上，编码端还可以将第二标识写入码流中，其中，第二标识用于指示码

流是否包含第一标识；示例性的，当在码流中写入第一标识时，可以将第二标识的值设置为预设标识值，然后将预设标识值写入码流。

例如，码流的语法结构可以参照如下表 9 所示：

表 9

Syntax	No. of bits	Mnemonic
语法元素 1		
{		
语法元素 2;	1	bslbf
语法元素 3;	3	bslbf
.....		
语法元素 n;	4	uimsbf
}		
.....		
语法元素 n+1		
hoaGroupHascomplexityConfig;		
if (hoaGroupHascomplexityConfig) {		
complexityLevel;	3	uimsbf
}		

- 5在表 9 中，“hoaGroupHascomplexityConfig”为第二标识的语法元素；且预设标识值为“1”。
- 这样，解码端可以从码流中解码出第二标识；当确定第二标识的值为预设标识值时，可以确定码流中包括第一标识；然后从码流中解码出第一标识。
- 此外，还可以将第二标识的语法元素和对应的描述信息添加至码流的语法结构中。具体可以参照上文中的描述，在此不再赘述。

10

图 13 为示例性示出的双耳渲染信号生成过程示意图。图 13 实施例由终端设备执行，该终端设备可以包括解码器和渲染器；图 13 是在图 4A 的基础上示出的。

S1301，接收码流。

S1302，从码流中解析出场景音频信号的元数据。

- 15S1303，根据码流，获取场景音频信号的重建音频信号。

示例性的，S1301~S1303 可以参照上述 S401~S403 的描述，在此不再赘述。

S1304，根据场景音频信号的重建音频信号，生成虚拟扬声器信号。

S1305，根据第一元数据与多个麦克风中的两个麦克风之间的距离，确定虚拟扬声器半径。

S1306，根据虚拟扬声器半径和虚拟扬声器信号进行渲染，得到渲染后的音频信号。

- 20其中，S1304~S1306 可以参照上述 S11~S16 的描述，在此不再赘述。

需要说明的是，S1305~S1306 可以由渲染器执行；一种可能的方式中，S1301~S1303 可以由解码器执行；一种可能的方式中，S1301~S1302 可以由解码器执行，S1303 可以由解码器和其他模块协同实现。

图 14 为示例性示出的编码装置示意图。其中，编码装置可以用于执行前述实施例的方法，因此，其

所能达到的有益效果可参考上文所提供的对应的方法中的有益效果，此处不再赘述。

第一数据获取模块 1401，用于获取场景音频信号和场景音频信号的元数据；其中，场景音频信号用于描述场景中声源的声场，场景包括多个麦克风，元数据包括第一元数据，第一元数据用于与多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，虚拟扬声器半径和虚拟扬声器信号用于渲染以得到渲染后的音频信号，虚拟扬声器信号是根据场景音频信号的重建音频信号生成的；

第一编码模块 1402，用于编码场景音频信号和元数据，得到码流。

图 15 为示例性示出的解码装置示意图。其中，解码装置可以用于执行前述实施例的方法，因此，其所能达到的有益效果可参考上文所提供的对应的方法中的有益效果，此处不再赘述。

第一接收模块 1501，用于接收码流；

第一解析模块 1502，用于从码流中解析出场景音频信号的元数据；

第一信号获取模块 1503，用于根据码流，获取场景音频信号的重建音频信号；场景音频信号用于描述场景中声源的声场，场景包括多个麦克风，元数据包括第一元数据，第一元数据用于与多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，虚拟扬声器半径和虚拟扬声器信号用于渲染以得到渲染后的音频信号，虚拟扬声器信号是根据重建音频信号生成的。

图 16 为示例性示出的编码装置示意图。其中，编码装置可以用于执行前述实施例的方法，因此，其所能达到的有益效果可参考上文所提供的对应的方法中的有益效果，此处不再赘述。

第二数据获取模块 1601，用于获取场景音频信号；

第二编码模块 1602，用于编码场景音频信号和第一标识，以得到码流；第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量，场景音频信号的重建音频信号用于生成虚拟扬声器信号，虚拟扬声器信号和虚拟扬声器的数量用于渲染以得到渲染后的音频信号。

图 17 为示例性示出的解码装置示意图。其中，解码装置可以用于执行前述实施例的方法，因此，其所能达到的有益效果可参考上文所提供的对应的方法中的有益效果，此处不再赘述。

第二接收模块 1701，用于接收码流；

第二解析模块 1702，用于从码流中解析出第一标识；

第二信号获取模块 1703，用于根据码流，获取场景音频信号的重建音频信号；第一标识用于指示渲染复杂度，渲染复杂度用于确定虚拟扬声器的数量，重建音频信号用于生成虚拟扬声器信号，虚拟扬声器信号和虚拟扬声器的数量用于渲染以得到渲染后的音频信号。

一个示例中，图 18 示出了本申请实施例的一种装置 1800 的示意性框图装置 1800 可包括：处理器 1801 和收发器/收发管脚 1802，可选地，还包括存储器 1803。

装置 1800 的各个组件通过总线 1804 耦合在一起，其中总线 1804 除包括数据总线之外，还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见，在图中将各种总线都称为总线 1804。

可选地，存储器 1803 可以用于存储前述方法实施例中的指令。该处理器 1801 可用于执行存储器 1803 中的指令，并控制接收管脚接收信号，以及控制发送管脚发送信号。

装置 1800 可以是上述方法实施例中的电子设备或电子设备的芯片。

示例性的，电子设备可以为终端设备或服务器。

其中，上述方法实施例涉及的各步骤的所有相关内容均可以援引到对应功能模块的功能描述，在此不再赘述。

本申请实施例还提供一种芯片，包括一个或多个接口电路和一个或多个处理器；一个或多个处理器通过一个或多个接口电路接收或发送数据，当一个或多个处理器执行计算机指令时，使得上述相关方法步骤实现上述实施例中的方法的步骤被执行。其中，接口电路为收发器/收发管脚 1802。

本实施例还提供一种计算机可读存储介质，该计算机可读存储介质中存储有计算机指令，当该计算机指令在电子设备上运行时，使得电子设备执行上述相关方法步骤实现上述实施例中的方法。

本实施例还提供了一种计算机程序产品，该计算机程序产品包含计算机指令，当计算机指令被计算机或处理器执行时，使得计算机执行上述相关步骤，以实现上述实施例中的方法。

另外，本申请的实施例还提供一种装置，这个装置具体可以是芯片，组件或模块，该装置可包括相连的处理器和存储器：其中，存储器用于存储计算机执行指令，当装置运行时，处理器可执行存储器存储的计算机执行指令，以使芯片执行上述各方法实施例中的方法。

其中，本实施例提供的电子设备、计算机可读存储介质、计算机程序产品或芯片均用于执行上文所提供的对应的方法，因此，其所能达到的有益效果可参考上文所提供的对应的方法中的有益效果，此处不再赘述。

通过以上实施方式的描述，所属领域的技术人员可以了解到，为描述的方便和简洁，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。

在本申请所提供的几个实施例中，应该理解到，所揭露的装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，模块或单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个装置，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是一个物理单元或多个物理单元，即可以位于一个地方，或者也可以分布到多个不同地方。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

本申请各个实施例的任意内容，以及同一实施例的任意内容，均可以自由组合。对上述内容的任意组合均在本申请的范围之内。

集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个可读存储介质中。基于这样的理解，本申请实施例的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该软件产品存储在一个存储介质中，包括若干指令用以使得一个设备（可以是单片机，芯片等）或处理器（processor）执行本申请各个实施例方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（read only memory, ROM）、随机存取存储器（random access memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

上面结合附图对本申请的实施例进行了描述，但是本申请并不局限于上述的具体实施方式，上述的具体实施方式仅仅是示意性的，而不是限制性的，本领域的普通技术人员在本申请的启示下，在不脱离本申请宗旨和权利要求所保护的范围情况下，还可做出很多形式，均属于本申请的保护之内。

结合本申请实施例公开内容所描述的方法或者算法的步骤可以硬件的方式来实现，也可以是由处理器执行软件指令的方式来实现。软件指令可以由相应的软件模块组成，软件模块可以被存放于随机存取存储器（Random Access Memory, RAM）、闪存、只读存储器（Read Only Memory, ROM）、可擦除可编程只读存储器（Erasable Programmable ROM, EPROM）、电可擦可编程只读存储器（Electrically EPROM, EEPROM）、寄存器、硬盘、移动硬盘、只读光盘（CD-ROM）或者本领域熟知的任何其它形式的存储介质中。一种示例性的存储介质耦合至处理器，从而使处理器能够从该存储介质读取信息，且可向该存储介质写入信息。当然，存储介质也可以是处理器的组成部分。处理器和存储介质可以位于 ASIC 中。

本领域技术人员应该可以意识到，在上述一个或多个示例中，本申请实施例所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。计算机可读介质包括计算机可读存储介质和通信介质，其中通信介质包括便于从一个地方向另一个地方传送计算机程序的任何介质。存储介质可以是通用或专用计算机能够存取的任何可用介质。

上面结合附图对本申请的实施例进行了描述，但是本申请并不局限于上述的具体实施方式，上述的具体实施方式仅仅是示意性的，而不是限制性的，本领域的普通技术人员在本申请的启示下，在不脱离本申请宗旨和权利要求所保护的范围情况下，还可做出很多形式，均属于本申请的保护之内。

权 利 要 求 书

- 1.一种解码方法，其特征在于，所述方法包括：
接收码流；
从所述码流中解析出场景音频信号的元数据；
- 5 根据所述码流，获取所述场景音频信号的重建音频信号；所述场景音频信号用于描述场景中声源的声场，所述场景包括多个麦克风，所述元数据包括第一元数据，所述第一元数据用于与所述多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，所述虚拟扬声器半径和虚拟扬声器信号用于渲染以得到渲染后的音频信号，所述虚拟扬声器信号是根据所述重建音频信号生成的。
- 10 2.根据权利要求1所述的方法，其特征在于，所述方法还包括：
从所述码流中解析出第一标识；
其中，所述第一标识用于指示渲染复杂度，所述渲染复杂度用于确定所述虚拟扬声器的数量。
- 15 3.根据权利要求1或2所述的方法，其特征在于，所述方法还包括：
从所述码流中解析出第二标识；
其中，所述第二标识的值为第一预设标识值时指示所述码流包括所述元数据。
- 20 4.根据权利要求3所述的方法，其特征在于，所述方法还包括：
所述第二标识的值为第一预设标识值时指示所述码流中还包括第一标识。
- 25 5.根据权利要求2所述的方法，其特征在于，所述方法还包括：
从所述码流中解析出第三标识；
其中，所述第三标识的值为第二预设标识值时指示所述码流中包括所述第一标识。
- 30 6.根据权利要求1至5任一项所述的方法，其特征在于，
所述元数据还包括第二元数据，所述第二元数据用于确定所述多个麦克风中的两个麦克风之间的距离。
- 7.根据权利要求1至6任一项所述的方法，其特征在于，
所述场景音频信号包括多点高阶立体混响 HOA 信号，所述重建音频信号包括所述多点 HOA 信号的重建音频信号。
- 35 8.根据权利要求1至7任一项所述的方法，其特征在于，所述码流包括：所述元数据的语法元素和对应的描述信息；
所述从所述码流中解析出场景音频信号的元数据，包括：
根据所述元数据的语法元素和对应的描述信息，从所述码流中解析出所述元数据。
- 40 9.根据权利要求2至8任一项所述的方法，其特征在于，所述码流包括：第一标识的语法元素和对应的描述信息，所述从所述码流中解析出第一标识，包括：
根据所述第一标识的语法元素和对应的描述信息，从所述码流中解析出所述第一标识。
- 10.根据权利要求3至9任一项所述的方法，其特征在于，所述码流包括：第二标识的语法元素和对应的描述信息，所述从所述码流中解析出第二标识，包括：
根据所述第二标识的语法元素和对应的描述信息，从所述码流中解析出所述第二标识。
- 45 11.根据权利要求5至9任一项所述的方法，其特征在于，所述码流包括：第三标识的语法元素和对应的描述信息，所述从所述码流中解析出第三标识，包括：

根据所述第三标识的语法元素和对应的描述信息，从所述码流中解析出所述第三标识。

12.根据权利要求 8 所述的方法，其特征在于，所述第一元数据的语法元素为 distanceFactor。

5 13.根据权利要求 9 所述的方法，其特征在于，所述第一标识的语法元素为 complexityLevel。

14.根据权利要求 10 所述的方法，其特征在于，所述第二标识的语法元素为
hoaGroupHasLowProfileConfig。

10 15.根据权利要求 1 至 14 任一项所述的方法，其特征在于，所述方法还包括：
根据所述场景音频信号的重建音频信号，生成虚拟扬声器信号；
根据所述第一元数据与所述多个麦克风中的两个麦克风之间的距离，确定虚拟扬声器半径；
根据所述虚拟扬声器半径和所述虚拟扬声器信号进行渲染，得到所述渲染后的音频信号。

15 16.一种解码方法，其特征在于，包括：
接收码流；
从所述码流中解析出第一标识；
根据所述码流，获取重建音频信号，其中，所述第一标识用于指示渲染复杂度，所述渲染复杂度用于
确定虚拟扬声器的数量，所述重建音频信号用于生成虚拟扬声器信号，所述虚拟扬声器信号和所述虚拟扬
20 声器的数量用于渲染以得到渲染后的音频信号。

17.根据权利要求 16 所述的方法，其特征在于，所述方法还包括：
从所述码流中解析出第二标识；
所述第二标识的值为预设标识值时指示所述码流中包括所述第一标识。

25 18.根据权利要求 16 或 17 所述的方法，其特征在于，所述码流包括：所述第一标识的语法元素和对应
的描述信息，所述从所述码流中解析出第一标识，包括：
根据所述第一标识的语法元素和对应的描述信息，从所述码流中解析出所述第一标识。

30 19.根据权利要求 17 或 18 所述的方法，其特征在于，所述码流包括：第二标识的语法元素和对应的描
述信息，所述从所述码流中解析出第二标识，包括：
根据所述第二标识的语法元素和对应的描述信息，从所述码流中解析出所述第二标识。

35 20.根据权利要求 16 至 19 任一项所述的方法，其特征在于，
所述重建音频信号包括多点高阶立体混响 HOA 信号的重建音频信号。

21.根据权利要求 18 所述的方法，其特征在于，所述第一标识的语法元素为 complexityLevel。

40 22.根据权利要求 16 至 21 任一项所述的方法，其特征在于，所述方法还包括：
根据所述场景音频信号的重建音频信号，生成虚拟扬声器信号；
根据所述第一标识，确定所述虚拟扬声器的数量；
根据所述虚拟扬声器的数量和所述虚拟扬声器信号进行渲染，得到所述渲染后的音频信号。

45 23.一种解码装置，其特征在于，所述装置包括：
第一接收模块，用于接收码流；
第一解析模块，用于从所述码流中解析出场景音频信号的元数据；
第一信号获取模块，用于根据所述码流，获取所述场景音频信号的重建音频信号；所述场景音频信号
用于描述场景中声源的声场，所述场景包括多个麦克风，所述元数据包括第一元数据，所述第一元数据用

于与所述多个麦克风中的两个麦克风之间的距离确定虚拟扬声器半径，所述虚拟扬声器半径和虚拟扬声器信号用于渲染以得到渲染后的音频信号，所述虚拟扬声器信号是根据所述重建音频信号生成的。

- 24.一种解码装置，其特征在于，所述装置包括：
- 5 第二接收模块，用于接收码流；
- 第二解析模块，用于从所述码流中解析出第一标识；
- 第二信号获取模块，用于根据所述码流，获取所述场景音频信号的重建音频信号；所述第一标识用于指示渲染复杂度，所述渲染复杂度用于确定虚拟扬声器的数量，所述重建音频信号用于生成虚拟扬声器信号，所述虚拟扬声器信号和所述虚拟扬声器的数量用于渲染以得到渲染后的音频信号。
- 10
- 25.一种扩展现实 XR 设备，其特征在于，所述 XR 设备包括：处理器和音频播放模块；
- 所述处理器用于执行如上述权利要求 1 至权利要求 22 任一项所述的解码方法；
- 所述音频播放模块用于播放如上述权利要求 1 至权利要求 22 任一项所述的渲染后的音频信号。
- 15
- 26.一种电子设备，其特征在于，包括：
- 存储器和处理器，所述存储器与所述处理器耦合；
- 所述存储器存储有程序指令，当所述程序指令由所述处理器执行时，使得所述电子设备执行如权利要求 1 至权利要求 22 中任一项所述的解码方法。
- 20
- 27.一种芯片，其特征在于，包括一个或多个接口电路和一个或多个处理器；所述一个或多个处理器通过所述一个或多个接口电路接收或发送数据，当所述一个或多个处理器执行计算机指令时，使得如权利要求 1 至权利要求 22 中任一项所述的方法的步骤被执行。
- 25
- 28.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机程序，当所述计算机程序运行在计算机或处理器上时，使得所述计算机或所述处理器执行如权利要求 1 至权利要求 22 任一项所述的解码方法。
- 29.一种计算机程序产品，其特征在于，所述计算机程序产品包含计算机指令，当所述计算机指令被计算机或处理器执行时，使得如权利要求 1 至权利要求 22 任一项所述的方法的步骤被执行。