三维音频信号编码方法、装置和编码器

本申请要求于2021年05月17日提交国家知识产权局、申请号为202110536634.9、申请名称为“三维音频信号编码方法、装置和编码器”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

本申请涉及多媒体领域，尤其涉及一种三维音频信号编码方法、装置和编码器。

背景技术

随着高性能计算机和信号处理技术的飞速发展，收听者对语音、音频体验提出了越来越高的要求，浸入式音频能够满足人们在这方面的需求。例如，三维音频技术在无线通信（例如4G/5G等等）语音、虚拟现实/增强现实和媒体音频等方面得到了广泛应用。三维音频技术是对真实世界中的声音和三维声场信息进行获取、处理、传输和渲染回放的音频技术，使声音具有强烈的空间感、包围感及沉浸感，给收听者以“身临其境”的非凡听觉体验。

通常，采集设备（如：麦克风）采集大量的数据记录三维声场信息，向回放设备（例如扬声器，耳机等）传输三维音频信号，以便于回放设备播放三维音频。由于三维声场信息的数据量较大，导致需要大量的存储空间存储数据，以及传输三维音频信号的带宽需求较高。为了解决上述问题，可以对三维音频信号进行压缩，存储或传输压缩数据。目前，编码器先遍历候选虚拟扬声器集合中的虚拟扬声器，利用选取的虚拟扬声器对三维音频信号进行压缩。但是，若连续帧选取的虚拟扬声器的结果差异较大，导致重建三维音频信号声像不稳定，降低重建三维音频信号的音质。

发明内容

本申请提供了三维音频信号编码方法、装置和编码器，由此可以增强了帧之间的方位的连续性，提高了重建后三维音频信号的声像的稳定性，确保重建后三维音频信号的音质。

第一方面，本申请提供了一种三维音频信号编码方法，该方法可以由编码器执行，具体包括如下步骤：编码器获取到三维音频信号的当前帧的第一数量个当前帧初始投票值后，根据第一数量个当前帧初始投票值，以及第六数量个虚拟扬声器与三维音频信号的在先帧对应的第六数量个在先帧最终投票值，获取第七数量个虚拟扬声器与当前帧对应的第七数量个当前帧最终投票值。其中，虚拟扬声器与当前帧初始投票值一一对应，第一数量个虚拟扬声器包括第一虚拟扬声器，第一虚拟扬声器的当前帧初始投票值用于表征对当前帧进行编码时使用第一虚拟扬声器的优先级。第七数量个虚拟扬声器包括第一数量个虚拟扬声器，且第七数量个虚拟扬声器包括第六数量个虚拟扬声器。进而，编码器根据第七数量个当前帧最终投票值，从第七数量个虚拟扬声器中选取第二数量个当前帧的代表虚拟扬声器，第二数量小于第七数量，表示第二数量个当前帧的代表虚拟扬声器是第七数量个虚拟扬声器中的部分虚拟扬声器；根据第二数

量个当前帧的代表虚拟扬声器对当前帧进行编码，得到码流。

在虚拟扬声器搜索过程中，由于真实声源的位置与虚拟扬声器的位置不一定重合，会导致虚拟扬声器不一定能够与真实声源形成一一对应关系，且由于在实际的复杂场景下，可能出现有限数量的虚拟扬声器集合无法表征声场中所有声源的情况，此时，帧与帧之间搜索到的虚拟扬声器可能会发生频繁跳变，这种跳变会明显地影响听音者的听觉感受，导致解码重建后三维音频信号中出现明显的不连续和噪声现象。本申请的实施例提供的选择虚拟扬声器的方法通过继承在先帧的代表虚拟扬声器，即对于相同编号的虚拟扬声器，用在先帧最终投票值调整当前帧初始投票值，使得编码器更倾向于选择在先帧的代表虚拟扬声器，从而降低帧与帧之间的虚拟扬声器的频繁跳变，增强了帧之间的信号方位的连续性，提高了重建后三维音频信号的声像的稳定性，确保重建后三维音频信号的音质。

示例地，若第六数量个虚拟扬声器包括第一虚拟扬声器，根据第一数量个当前帧初始投票值以及第六数量个虚拟扬声器与三维音频信号的在先帧对应的第六数量个在先帧投票值，获取第七数量个虚拟扬声器与当前帧对应的第七数量个当前帧最终投票值包括：根据第一虚拟扬声器的在先帧最终投票值更新第一虚拟扬声器的当前帧初始投票值，以获得第一虚拟扬声器的当前帧最终投票值。

在一种可能的实现方式中，若第一数量个虚拟扬声器包括第二虚拟扬声器，且第六数量个虚拟扬声器不包括第二虚拟扬声器，则第二虚拟扬声器的当前帧最终投票值等于第二虚拟扬声器的当前帧初始投票值；或，若第六数量个虚拟扬声器包括第三虚拟扬声器，且第一数量个虚拟扬声器不包括第三虚拟扬声器，则第三虚拟扬声器的当前帧最终投票值等于第三虚拟扬声器的在先帧最终投票值。

在另一种可能的实现方式中，根据第一虚拟扬声器的在先帧最终投票值更新第一虚拟扬声器的当前帧初始投票值包括：编码器根据第一调整参数调整第一虚拟扬声器的在先帧最终投票值，以得到第一虚拟扬声器的在先帧调整后投票值；根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧初始投票值。

其中，第一调整参数是根据在先帧中指向性声源的数量、对当前帧进行编码的编码速率和帧类型中至少一个确定的。从而，编码器利用第一调整参数调整第一虚拟扬声器的在先帧最终投票值，使得编码器更倾向于选择在先帧的代表虚拟扬声器，从而增强了帧之间的方位的连续性，提高了重建后三维音频信号的声像的稳定性，确保重建后三维音频信号的音质。

在另一种可能的实现方式中，根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧初始投票值包括：编码器根据第二调整参数调整第一虚拟扬声器的当前帧初始投票值，得到第一虚拟扬声器的当前帧调整后投票值；根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧调整后投票值。

其中，第二调整参数是根据第一虚拟扬声器的在先帧调整后投票值和第一虚拟扬声器的当前帧初始投票值确定的。从而，编码器利用第二调整参数调整第一虚拟扬声器的当前帧初始投票值，降低了当前帧初始投票值的频繁跳变，使得编码器更倾向于选择在先帧的代表虚拟扬声器，从而增强了帧之间的方位的连续性，提高了重建后三维音频信号的声像的稳定性，确保重建后三维音频信号的音质。

第二数量用于表征编码器选取的当前帧的代表虚拟扬声器的数量。第二数量越大表示当前帧的代表虚拟扬声器的数量越大，三维音频信号的声场信息越多；第二数量越小表示当前帧的代表虚拟扬声器的数量越小，三维音频信号的声场信息越少。因此，可通过设置第二数量控制编码器选取的当前帧的代表虚拟扬声器的数量。例如，第二数量可以是预设的，又如，第二数量可以是根据当前帧确定的。示例地，第二数量的取值可以是 1、2、4 或 8。

在另一种可能的实现方式中，获取第一数量个虚拟扬声器与三维音频信号的当前帧对应的第一数量个当前帧初始投票值包括：编码器根据当前帧的第三数量个代表系数、候选虚拟扬声器集合和投票轮数确定第一数量个虚拟扬声器，以及第一数量个当前帧初始投票值。其中，候选虚拟扬声器集合包括第五数量个虚拟扬声器，第五数量个虚拟扬声器包括第一数量个虚拟扬声器，第一数量小于或等于第五数量，投票轮数为大于或等于 1 的整数，且投票轮数小于或等于第五数量。

目前，在虚拟扬声器搜索过程中，编码器依据待编码的三维音频信号和虚拟扬声器之间的相关计算的结果作为虚拟扬声器的选择衡量指标。而且，若编码器对每一个系数传输一个虚拟扬声器，则无法达到高效数据压缩的目的，会对编码器造成沉重的计算负担。本申请实施例提供的选择虚拟扬声器的方法，编码器利用较少数量的代表系数代替当前帧的全部系数对候选虚拟扬声器集合中每个虚拟扬声器进行投票，依据投票值选取当前帧的代表虚拟扬声器。进而，编码器利用当前帧的代表虚拟扬声器对待编码的三维音频信号进行压缩编码，不仅有效地提升了对三维音频信号进行压缩编码的压缩率，而且降低了编码器搜索虚拟扬声器的计算复杂度，从而降低了对三维音频信号进行压缩编码的计算复杂度以及减轻了编码器的计算负担。

在另一种可能的实现方式中，根据当前帧的第三数量个代表系数、候选虚拟扬声器集合和投票轮数确定第一数量个虚拟扬声器，以及第一数量个当前帧初始投票值之前，方法还包括：编码器获取当前帧的第四数量个系数，以及第四数量个系数的频域特征值；根据第四数量个系数的频域特征值，从第四数量个系数中选取第三数量个代表系数，第三数量小于第四数量，表示第三数量个代表系数是第四数量个系数中的部分系数。

其中，三维音频信号的当前帧为高阶立体混响（higher order ambisonics，HOA）信号；系数的频域特征值是依据 HOA 信号的系数确定的。

如此，由于编码器从当前帧的全部系数中选取部分系数作为代表系数，利用较少数量的代表系数代替当前帧的全部系数从候选虚拟扬声器集合中选取代表虚拟扬声器，因此有效地降低了编码器搜索虚拟扬声器的计算复杂度，从而降低了对三维音频信号进行压缩编码的计算复杂度以及减轻了编码器的计算负担。

另外，编码器根据第二数量个当前帧的代表虚拟扬声器对当前帧进行编码，得到码流包括：编码器根据第二数量个当前帧的代表虚拟扬声器和当前帧生成虚拟扬声器信号；对虚拟扬声器信号进行编码得到码流。

在另一种可能的实现方式中，方法还包括：编码器获取当前帧与在先帧的代表虚拟扬声器集合的第一相关度，若第一相关度不满足复用条件，获取三维音频信号的当前帧的第四数量个系数，以及第四数量个系数的频域特征值。在先帧的代表虚拟扬声

器集合包括第六数量个虚拟扬声器，第六数量个虚拟扬声器包含的虚拟扬声器为对三维音频信号的在先帧进行编码所使用的在先帧的代表虚拟扬声器，第一相关度用于确定对当前帧进行编码时是否复用在先帧的代表虚拟扬声器集合。

如此，编码器可以先判断是否可以复用在先帧的代表虚拟扬声器集合对当前帧进行编码，如果编码器复用在先帧的代表虚拟扬声器集合对当前帧进行编码，从而，避免编码器再执行搜索虚拟扬声器的过程，有效地降低了编码器搜索虚拟扬声器的计算复杂度，因此降低了对三维音频信号进行压缩编码的计算复杂度以及减轻了编码器的计算负担。另外，还可以降低帧与帧之间的虚拟扬声器的频繁跳变，增强了帧之间的方位的连续性，提高了重建后三维音频信号的声像的稳定性，确保重建后三维音频信号的音质。如果编码器不能复用在先帧的代表虚拟扬声器集合对当前帧进行编码，编码器再选取代表系数，利用当前帧的代表系数对候选虚拟扬声器集合中每个虚拟扬声器进行投票，依据投票值选取当前帧的代表虚拟扬声器，来达到降低对三维音频信号进行压缩编码的计算复杂度以及减轻编码器的计算负担的目的。

可选地，方法还包括：编码器还可以采集三维音频信号的当前帧，以便于对三维音频信号的当前帧进行压缩编码得到码流，将码流传输至解码端。

第二方面，本申请提供了一种三维音频信号编码装置，所述装置包括用于执行第一方面或第一方面任一种可能设计中的三维音频信号编码方法的各个模块。例如，三维音频信号编码装置包括虚拟扬声器选择模块和编码模块。所述虚拟扬声器选择模块，用于获取第一数量个虚拟扬声器与三维音频信号的当前帧对应的第一数量个当前帧初始投票值，虚拟扬声器与当前帧初始投票值一一对应，第一数量个虚拟扬声器包括第一虚拟扬声器，第一虚拟扬声器的当前帧初始投票值用于表征对当前帧进行编码时使用第一虚拟扬声器的优先级；所述虚拟扬声器选择模块，还用于根据第一数量个当前帧初始投票值以及第六数量个虚拟扬声器与三维音频信号的在先帧对应的第六数量个在先帧最终投票值，获取第七数量个虚拟扬声器与当前帧对应的第七数量个当前帧最终投票值，第七数量个虚拟扬声器包括第一数量个虚拟扬声器，且第七数量个虚拟扬声器包括第六数量个虚拟扬声器；所述虚拟扬声器选择模块，还用于根据第七数量个当前帧最终投票值，从第七数量个虚拟扬声器中选取第二数量个当前帧的代表虚拟扬声器，第二数量小于第七数量；所述编码模块，用于根据第二数量个当前帧的代表虚拟扬声器对当前帧进行编码，得到码流。这些模块可以执行上述第一方面方法示例中的相应功能，具体参见方法示例中的详细描述，此处不做赘述。

第三方面，本申请提供一种编码器，该编码器包括至少一个处理器和存储器，其中，所述存储器用于存储一组计算机指令；当处理器执行所述一组计算机指令时，执行第一方面或第一方面任一种可能实现方式中的三维音频信号编码方法的操作步骤。

第四方面，本申请提供一种系统，系统包括如第三方面所述的编码器，以及解码器，所述编码器用于执行第一方面或第一方面任一种可能实现方式中的三维音频信号编码方法的操作步骤，所述解码器用于解码所述编码器生成的码流。

第五方面，本申请提供一种计算机可读存储介质，包括：计算机软件指令；当计算机软件指令在编码器中运行时，使得编码器执行如第一方面或第一方面任意一种可能的实现方式中所述方法的操作步骤。

第六方面，本申请提供一种计算机程序产品，当计算机程序产品在编码器上运行时，使得编码器执行如第一方面或第一方面任意一种可能的实现方式中所述方法的操作步骤。

本申请在上述各方面提供的实现方式的基础上，还可以进行进一步组合以提供更多实现方式。

附图说明

图 1 为本申请实施例提供的一种音频编解码系统的结构示意图；

图 2 为本申请实施例提供的一种音频编解码系统的场景示意图；

图 3 为本申请实施例提供的一种编码器的结构示意图；

图 4 为本申请实施例提供的一种三维音频信号编解码方法的流程示意图；

图 5 为本申请实施例提供的一种选择虚拟扬声器方法的流程示意图；

图 6 为本申请实施例提供的一种三维音频信号编码方法的流程示意图；

图 7 为本申请实施例提供的另一种选择虚拟扬声器方法的流程示意图；

图 8 为本申请实施例提供的一种调整投票值的方法的流程示意图；

图 9 为本申请实施例提供的另一种选择虚拟扬声器方法的流程示意图；

图 10 为本申请提供的一种编码装置的结构示意图；

图 11 为本申请提供的一种编码器的结构示意图。

具体实施方式

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。

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。

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。

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。

三维音频技术是指将人耳以外的空间假设为一个系统，耳膜处接收到的信号为声源发出的声音经过耳朵以外系统滤波输出的三维音频信号。例如，人耳以外的系统可以定义为系统冲击响应 h(n)，任意一个声源可以定义为 x(n)，耳膜处接收到的信号为 x(n)和 h(n)的卷积结果。本申请实施例所述的三维音频信号可以是指高阶立体混响（higher order ambisonics，HOA）信号。三维音频也可以称为三维音效、空间音频、三维声场重建、虚拟 3D 音频或双耳音频等。

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。

$$\nabla^2 p+k^2 p=0 \qquad \text{公式(1)}$$

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)。

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

其中，r 表示球半径，θ 表示水平角，φ 表示俯仰角，k 表示波数，s 表示理想平面波的幅度，m 表示三维音频信号的阶数序号（或称为 HOA 信号的阶数序号）。$j^m j_m^{kr}(kr)$ 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，$(2m+1)j^m j_m^{kr}(kr)$ 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。三维音频信号系数满足公式(3)。

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

将公式(3)代入公式(2)，公式(2)可以变形为公式(4)。

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

其中，$B_{m,n}^{\sigma}$ 表示 N 阶的三维音频信号系数，用于近似描述声场。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。比如，N 的取值范围为 2 至 6 的整数。本申请的实施例所述的三维音频信号的系数可以是指 HOA 系数或环境立体声（ambisonic）系数。

三维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。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，可以将三维音频信号描述的声场使用多个平面波的叠加来表达，并通过三维音频信号系数重建声场。

相对 5.1 声道的音频信号或 7.1 声道的音频信号，由于 N 阶的 HOA 信号有 $(N+1)^2$ 个声道，则 HOA 信号包括用于描述声场的空间信息的数据量较多。若采集设备（比如：麦克风）将该三维音频信号传输到回放设备（比如：扬声器），需要消耗较大的带宽。目前，编码器可以利用空间压缩环绕音频编码（spatial squeezed surround audio coding，S3AC）或定向音频编码（directional audio coding，DirAC）对三维音频信号进行压缩编码得到码流，向回放设备传输码流。回放设备对码流进行解码，并重建三

维音频信号，播放重建后三维音频信号。从而降低向回放设备传输三维音频信号的数据量，以及带宽的占用。但是，编码器对三维音频信号进行压缩编码的计算复杂度较高，占用编码器过多的计算资源。因此，如何降低对三维音频信号进行压缩编码的计算复杂度是一个亟待解决的问题。

本申请实施例提供一种音频编解码技术，尤其是提供一种面向三维音频信号的三维音频编解码技术，具体提供一种采用较少的声道表示三维音频信号的编解码技术，以改进传统的音频编解码系统。音频编码（或通常称为编码）包括音频编码和音频解码两部分。音频编码在源侧执行，通常包括处理（例如，压缩）原始音频以减少表示该原始音频所需的数据量，从而更高效地存储和/或传输。音频解码在目的侧执行，通常包括相对于编码器作逆处理，以重建原始音频。编码部分和解码部分也合称为编解码。下面将结合附图对本申请实施例的实施方式进行详细描述。

图 1 为本申请实施例提供的一种音频编解码系统的结构示意图。音频编解码系统 100 包括源设备 110 和目的设备 120。源设备 110 用于对三维音频信号进行压缩编码得到码流，向目的设备 120 传输码流。目的设备 120 对码流进行解码，并重建三维音频信号，播放重建后三维音频信号。

具体地，源设备 110 包括音频获取器 111、预处理器 112、编码器 113 和通信接口 114。

音频获取器 111 用于获取原始音频。音频获取器 111 可以是任意类型的用于捕获现实世界声音的音频采集设备，和/或任意类型的音频生成设备。音频获取器 111 例如是用于生成计算机音频的计算机音频处理器。音频获取器 111 也可以为存储音频的任意类型的内存或存储器。音频包括现实世界声音、虚拟场景（如：VR 或增强现实（augmented reality，AR））声音和/或其任意组合。

预处理器 112 用于接收音频获取器 111 采集的原始音频，并对原始音频进行预处理，得到三维音频信号。例如，预处理器 112 执行的预处理包括声道转换、音频格式转换或去噪声等。

编码器 113 用于接收预处理器 112 生成的三维音频信号，对三维音频信号进行压缩编码得到码流。示例地，编码器 113 可以包括空间编码器 1131 和核心编码器 1132。空间编码器 1131 用于根据三维音频信号从候选虚拟扬声器集合选取（或称为搜索）虚拟扬声器，根据三维音频信号和虚拟扬声器生成虚拟扬声器信号。虚拟扬声器信号也可以称为回放信号。核心编码器 1132 用于对虚拟扬声器信号进行编码，得到码流。

通信接口 114 用于接收编码器 113 生成的码流，通过通信信道 130 向目的设备 120 发送码流，以便于目的设备 120 根据码流重建三维音频信号。

目的设备 120 包括播放器 121、后处理器 122、解码器 123 和通信接口 124。

通信接口 124 用于接收通信接口 114 发送的码流，并将码流传输给解码器 123。以便于解码器 123 根据码流重建三维音频信号。

通信接口 114 和通信接口 124 可用于通过源设备 110 与目的设备 120 之间的直连通信链路，例如直接有线或无线连接等，或者通过任意类型的网络，例如有线网络、无线网络或其任意组合、任意类型的私网和公网或其任意类型的组合，发送或接收原始音频的相关数据。

通信接口 114 和通信接口 124 均可配置为如图 1 中从源设备 110 指向目的设备 120 的对应通信信道 130 的箭头所指示的单向通信接口，或双向通信接口，并且可用于发送和接收消息等，以建立连接，确认并交换与通信链路和/或例如编码后的码流传输等数据传输相关的任何其它信息，等等。

解码器 123 用于对码流进行解码，并重建三维音频信号。示例地，解码器 123 包括核心解码器 1231 和空间解码器 1232。核心解码器 1231 用于对码流进行解码，得到虚拟扬声器信号。空间解码器 1232 用于根据候选虚拟扬声器集合和虚拟扬声器信号重建三维音频信号，得到重建后三维音频信号。

后处理器 122 用于接收解码器 123 生成的重建后三维音频信号，对重建后三维音频信号进行后处理。例如，后处理器 122 执行的后处理包括音频渲染、响度归一化、用户交互、音频格式转换或去噪声等。

播放器 121 用于根据重建后三维音频信号播放重建的声音。

需要说明的是，音频获取器 111 和编码器 113 可以集成在一个物理设备上，也可以设置在不同的物理设备上，不予限定。示例地，如图 1 所示的源设备 110 包括音频获取器 111 和编码器 113，表示音频获取器 111 和编码器 113 集成在一个物理设备上，则源设备 110 也可称为采集设备。源设备 110 例如是无线接入网的媒体网关、核心网的媒体网关、转码设备、媒体资源服务器、AR 设备、VR 设备、麦克风或者其他采集音频设备。若源设备 110 不包括音频获取器 111，表示音频获取器 111 和编码器 113 是两个不同的物理设备，源设备 110 可以从其他设备（如：采集音频设备或存储音频设备）获取原始音频。

另外，播放器 121 和解码器 123 可以集成在一个物理设备上，也可以设置在不同的物理设备上，不予限定。示例地，如图 1 所示的目的设备 120 包括播放器 121 和解码器 123，表示播放器 121 和解码器 123 集成在一个物理设备上，则目的设备 120 也可称为回放设备，目的设备 120 具有解码和播放重建音频的功能。目的设备 120 例如是扬声器、耳机或其他播放音频的设备。若目的设备 120 不包括播放器 121，表示播放器 121 和解码器 123 是两个不同的物理设备，目的设备 120 对码流解码重建三维音频信号后，将重建后三维音频信号传输给其他播放设备（如：扬声器或耳机），由其他播放设备回放重建后三维音频信号。

此外，图 1 示出了源设备 110 和目的设备 120 可以集成在一个物理设备上，也可以设置在不同的物理设备上，不予限定。

示例地，如图 2 中的(a)所示，源设备 110 可以是录音棚中的麦克风，目的设备 120 可以是扬声器。源设备 110 可以采集各种乐器的原始音频，将原始音频传输至编解码设备，编解码设备对原始音频进行编解码处理，得到重建后三维音频信号，由目的设备 120 回放重建后三维音频信号。又示例地，源设备 110 可以是终端设备中的麦克风，目的设备 120 可以是耳机。源设备 110 可以采集外界的声音或终端设备合成的音频。

又示例地，如图 2 中的(b)所示，源设备 110 和目的设备 120 集成在虚拟现实（virtual reality，VR）设备、增强现实（Augmented Reality，AR）设备、混合现实（Mixed Reality，MR）设备或扩展现实（Extended Reality，XR）设备中，则 VR/AR/MR/XR 设备具备采集原始音频、回放音频和编解码的功能。源设备 110 可以采集用户发出的声音和用

户所处的虚拟环境中虚拟物体发出的声音。

在这些实施例中，源设备 110 或其对应功能和目的设备 120 或其对应功能可以使用相同硬件和/或软件或通过单独的硬件和/或软件或其任意组合来实现。根据描述，图 1 所示的源设备 110 和/或目的设备 120 中的不同单元或功能的存在和划分可能根据实际设备和应用而有所不同，这对技术人员来说是显而易见的。

上述音频编解码系统的结构只是示意性说明，在一些可能的实现方式中，音频编解码系统还可以包括其他设备，例如，音频编解码系统还可以包括端侧设备或云侧设备。源设备 110 采集到原始音频后，对原始音频进行预处理，得到三维音频信号；并将三维音频传输至端侧设备或云侧设备，由端侧设备或云侧设备实现对三维音频信号进行编解码的功能。

本申请实施例提供的音频信号编解码方法主要应用于编码端。结合图 3 对编码器的结构进行详细说明。如图 3 所示，编码器 300 包括虚拟扬声器配置单元 310、虚拟扬声器集合生成单元 320、编码分析单元 330、虚拟扬声器选择单元 340、虚拟扬声器信号生成单元 350 和编码单元 360。

虚拟扬声器配置单元 310 用于根据编码器配置信息生成虚拟扬声器配置参数，以便得到多个虚拟扬声器。编码器配置信息包括但不限于：三维音频信号的阶数（或通常称为 HOA 阶数），编码比特率，用户自定义信息，等。虚拟扬声器配置参数包括但不限于：虚拟扬声器的数量，虚拟扬声器的阶数，虚拟扬声器的位置坐标，等。虚拟扬声器的数量例如是 2048、1669、1343、1024、530、512、256、128 或 64 等。虚拟扬声器的阶数可以是 2 阶至 6 阶中任一个。虚拟扬声器的位置坐标包括水平角和俯仰角。

虚拟扬声器配置单元 310 输出的虚拟扬声器配置参数作为虚拟扬声器集合生成单元 320 的输入。

虚拟扬声器集合生成单元 320 用于根据虚拟扬声器配置参数生成候选虚拟扬声器集合，候选虚拟扬声器集合包括多个虚拟扬声器。具体地，虚拟扬声器集合生成单元 320 根据虚拟扬声器的数量确定了候选虚拟扬声器集合包括的多个虚拟扬声器，以及根据虚拟扬声器的位置信息（如：坐标）和虚拟扬声器的阶数确定虚拟扬声器的系数。示例地，虚拟扬声器的坐标确定方法包括但不限于：按等距规则产生多个虚拟扬声器，或者根据听觉感知原理生成非均匀分布的多个虚拟扬声器；然后，根据虚拟扬声器的数量生成虚拟扬声器的坐标。

根据上述三维音频信号的生成原理也可以生成虚拟扬声器的系数。将公式(3)中的 θ_s 和 φ_s 分别设置为虚拟扬声器的位置坐标，$B_{m,n}^{\sigma}$ 表示 N 阶的虚拟扬声器的系数。虚拟扬声器的系数也可以称作 ambisonics 系数。

编码分析单元 330 用于对三维音频信号进行编码分析，例如分析三维音频信号的声场分布特征，即三维音频信号的声源数量、声源的方向性和声源的弥散度等特征。

虚拟扬声器集合生成单元 320 输出的候选虚拟扬声器集合包括的多个虚拟扬声器的系数作为虚拟扬声器选择单元 340 的输入。

编码分析单元 330 输出的三维音频信号的声场分布特征作为虚拟扬声器选择单元 340 的输入。

虚拟扬声器选择单元 340 用于根据待编码的三维音频信号、三维音频信号的声场分布特征和多个虚拟扬声器的系数确定与三维音频信号匹配的代表虚拟扬声器。

不限定的是，本申请实施例的编码器 300 还可以不包括编码分析单元 330，即编码器 300 可以不对输入信号进行分析，虚拟扬声器选择单元 340 采用一种默认配置确定代表虚拟扬声器。例如，虚拟扬声器选择单元 340 仅根据三维音频信号和多个虚拟扬声器的系数确定与三维音频信号匹配的代表虚拟扬声器。

其中，编码器 300 可以将从采集设备获取的三维音频信号或采用人工音频对象合成的三维音频信号作为编码器 300 的输入。另外，编码器 300 输入的三维音频信号可以是时域三维音频信号也可以是频域三维音频信号，不予限定。

虚拟扬声器选择单元 340 输出的代表虚拟扬声器的位置信息和代表虚拟扬声器的系数作为虚拟扬声器信号生成单元 350 和编码单元 360 的输入。

虚拟扬声器信号生成单元 350 用于根据三维音频信号和代表虚拟扬声器的属性信息生成虚拟扬声器信号。代表虚拟扬声器的属性信息包括代表虚拟扬声器的位置信息、代表虚拟扬声器的系数和三维音频信号的系数中至少一个。若属性信息为代表虚拟扬声器的位置信息，根据代表虚拟扬声器的位置信息确定代表虚拟扬声器的系数；若属性信息包括三维音频信号的系数，根据三维音频信号的系数获取代表虚拟扬声器的系数。具体地，虚拟扬声器信号生成单元 350 根据三维音频信号的系数和代表虚拟扬声器的系数计算虚拟扬声器信号。

示例地，假设矩阵 A 表示虚拟扬声器的系数，矩阵 X 表示 HOA 信号的 HOA 系数。矩阵 X 为矩阵 A 的逆矩阵。采用最小二乘方法求得理论的最优解 w，w 表示虚拟扬声器信号。虚拟扬声器信号满足公式(5)。

$$w = A^{-1}X \qquad \text{公式(5)}$$

其中，A^{-1} 表示矩阵 A 的逆矩阵。矩阵 A 的大小为 $(M \times C)$，C 表示代表虚拟扬声器的数量，M 表示 N 阶的 HOA 信号的声道的数量，a 表示代表虚拟扬声器的系数，矩阵 X 的大小为 $(M \times L)$，L 表示 HOA 信号的系数的数量，x 表示 HOA 信号的系数。代表虚拟扬声器的系数可以是指代表虚拟扬声器的 HOA 系数或代表虚拟扬声器的 ambisonics 系数。例如，$A = \begin{bmatrix} a_{11} & \cdots & a_{1C} \\ \vdots & & \vdots \\ a_{M1} & \cdots & a_{MC} \end{bmatrix}$，$X = \begin{bmatrix} x_{11} & \cdots & x_{1L} \\ \vdots & & \vdots \\ x_{M1} & \cdots & x_{ML} \end{bmatrix}$。

虚拟扬声器信号生成单元 350 输出的虚拟扬声器信号作为编码单元 360 的输入。

编码单元 360 用于对虚拟扬声器信号进行核心编码处理，得到码流。核心编码处理包括但不限于：变换、量化、心理声学模型、噪声整形、带宽扩展、下混、算数编码、码流产生等。

值得注意的是，空间编码器 1131 可以包括虚拟扬声器配置单元 310、虚拟扬声器集合生成单元 320、编码分析单元 330、虚拟扬声器选择单元 340 和虚拟扬声器信号生成单元 350，即虚拟扬声器配置单元 310、虚拟扬声器集合生成单元 320、编码分析单元 330、虚拟扬声器选择单元 340 和虚拟扬声器信号生成单元 350 实现了空间编码器

1131 的功能。核心编码器 1132 可以包括编码单元 360，即编码单元 360 实现了核心编码器 1132 的功能。

图 3 所示的编码器可以生成一个虚拟扬声器信号，也可以生成多个虚拟扬声器信号。多个虚拟扬声器信号可以由图 3 所示的编码器多次执行得到，也可以由图 3 所示的编码器一次执行得到。

接下来，结合附图对三维音频信号的编解码过程进行说明。图 4 为本申请实施例提供的一种三维音频信号编解码方法的流程示意图。在这里由图 1 中源设备 110 和目的设备 120 执行三维音频信号编解码过程为例进行说明。如图 4 所示，该方法包括以下步骤。

S410、源设备 110 获取三维音频信号的当前帧。

如上述实施例所述，若源设备 110 携带音频获取器 111，源设备 110 可以通过音频获取器 111 采集原始音频。可选地，源设备 110 也可以接收其他设备采集的原始音频；或者从源设备 110 中的存储器或其他存储器获取原始音频。原始音频可以包括实时采集的现实世界声音、设备存储的音频和由多个音频合成的音频中至少一种。本实施例对原始音频的获取方式以及原始音频的类型不予限定。

源设备 110 获取到原始音频后，根据三维音频技术和原始音频生成三维音频信号，以便于回放原始音频时，为听音者提供“身临其境”的音响效果。生成三维音频信号的具体方法可以参考上述实施例中预处理器 112 的阐述和现有技术的阐述。

另外，音频信号是一个连续的模拟信号。在音频信号处理过程中，可以先对音频信号进行采样，生成帧序列的数字信号。帧可以包括多个采样点。帧也可以指采样得到的采样点。帧也可以包括对帧划分得到的子帧。帧也可以指对帧划分得到的子帧。例如一帧长度为 L 个采样点，划分为 N 个子帧，那么每个子帧对应 L/N 个采样点。音频编解码通常是指处理包含多个采样点的音频帧序列。

音频帧可以包括当前帧或在先帧。本申请的各个实施例所述的当前帧或在先帧可以是指帧或是子帧。当前帧是指在当前时刻进行编解码处理的帧。在先帧是指在当前时刻之前时刻已进行编解码处理的帧。在先帧可以是当前时刻的前一时刻或者前多个时刻的帧。本申请的实施例中，三维音频信号的当前帧是指在当前时刻进行编解码处理的一帧三维音频信号。在先帧是指在当前时刻之前时刻已进行编解码处理的一帧三维音频信号。三维音频信号的当前帧可以是指三维音频信号的待编码当前帧。三维音频信号的当前帧可以简称为当前帧。三维音频信号的在先帧可以简称为在先帧。

S420、源设备 110 确定候选虚拟扬声器集合。

在一种情形下，源设备 110 的存储器中预先配置有候选虚拟扬声器集合。源设备 110 可以从存储器中读取候选虚拟扬声器集合。候选虚拟扬声器集合包括多个虚拟扬声器。虚拟扬声器表示空间声场中虚拟存在的扬声器。虚拟扬声器用于根据三维音频信号计算虚拟扬声器信号，以便于目的设备 120 回放重建后三维音频信号。

在另一种情形下，源设备 110 的存储器中预先配置有虚拟扬声器配置参数。源设备 110 根据虚拟扬声器配置参数生成候选虚拟扬声器集合。可选地，源设备 110 根据自身的计算资源（如：处理器）能力和当前帧的特征（如：信道和数据量）实时生成候选虚拟扬声器集合。

生成候选虚拟扬声器集合的具体方法可以参考现有技术，以及上述实施例中虚拟扬声器配置单元 310 和虚拟扬声器集合生成单元 320 的阐述。

S430、源设备 110 根据三维音频信号的当前帧，从候选虚拟扬声器集合中选取当前帧的代表虚拟扬声器。

源设备 110 根据当前帧的系数与虚拟扬声器的系数对虚拟扬声器进行投票，根据虚拟扬声器的投票值从候选虚拟扬声器集合中选择当前帧的代表虚拟扬声器。从候选虚拟扬声器集合中搜索有限数量的当前帧的代表虚拟扬声器，作为待编码的当前帧的最佳匹配虚拟扬声器，从而实现对待编码的三维音频信号进行数据压缩的目的。

图 5 为本申请实施例提供的一种选择虚拟扬声器方法的流程示意图。图 5 所述的方法流程是对图 4 中 S430 所包括的具体操作过程的阐述。在这里由图 1 所示的源设备 110 中编码器 113 执行选择虚拟扬声器过程为例进行说明。具体地实现虚拟扬声器选择单元 340 的功能。如图 5 所示，该方法包括以下步骤。

S510、编码器 113 获取当前帧的代表系数。

代表系数可以是指频域代表系数或时域代表系数。频域代表系数也可以称为频域代表频点或频谱代表系数。时域代表系数也可以称为时域代表采样点。获取当前帧的代表系数的具体方法可以参考下述图 7 中 S6101 和 S6102 的阐述。

S520、编码器 113 根据当前帧的代表系数对候选虚拟扬声器集合中虚拟扬声器的投票值，从候选虚拟扬声器集合中选取当前帧的代表虚拟扬声器。执行 S440 至 S460。

编码器 113 根据当前帧的代表系数与虚拟扬声器的系数对候选虚拟扬声器集合中的虚拟扬声器进行投票，根据虚拟扬声器的当前帧最终投票值从候选虚拟扬声器集合中选择（搜索）当前帧的代表虚拟扬声器。选取当前帧的代表虚拟扬声器的具体方法可以参考下述图 8 和图 7 中 S6103 的阐述。

需要说明的是，编码器先遍历候选虚拟扬声器集合包含的虚拟扬声器，利用从候选虚拟扬声器集合中选取的当前帧的代表虚拟扬声器对当前帧进行压缩。但是，若连续帧选取的虚拟扬声器的结果差异较大，会导致重建后三维音频信号的声像不稳定，降低重建后三维音频信号的音质。在本申请的实施例中，编码器 113 可以依据在先帧的代表虚拟扬声器的在先帧最终投票值对候选虚拟扬声器集合包含的虚拟扬声器的当前帧初始投票值进行更新处理，得到虚拟扬声器的当前帧最终投票值，则根据虚拟扬声器的当前帧最终投票值从候选虚拟扬声器集合中选取当前帧的代表虚拟扬声器。从而，通过参考在先帧的代表虚拟扬声器来选取当前帧的代表虚拟扬声器，使编码器对当前帧选择当前帧的代表虚拟扬声器时倾向于选择与在先帧的代表虚拟扬声器相同的虚拟扬声器，增加连续帧之间的方位的连续性，克服了连续帧选取的虚拟扬声器的结果差异较大的问题。因此，本申请的实施例还可以包括 S530。

S530、编码器 113 根据在先帧的代表虚拟扬声器的在先帧最终投票值调整候选虚拟扬声器集合中虚拟扬声器的当前帧初始投票值，获得虚拟扬声器的当前帧最终投票值。

编码器 113 根据当前帧的代表系数与虚拟扬声器的系数对候选虚拟扬声器集合中的虚拟扬声器进行投票，得到虚拟扬声器的当前帧初始投票值后，根据在先帧的代表虚拟扬声器的在先帧最终投票值调整候选虚拟扬声器集合中虚拟扬声器的当前帧初始

投票值，获得虚拟扬声器的当前帧最终投票值。在先帧的代表虚拟扬声器为编码器 113 对在先帧进行编码时使用的虚拟扬声器。调整候选虚拟扬声器集合中虚拟扬声器的当前帧初始投票值的具体方法可以参考下述图 6 中 S620 至 S630，以及图 8 中的 S810 至 S840 的阐述。

在一些实施例中，若当前帧是原始音频中第一帧，编码器 113 执行 S510 至 S520。若当前帧是原始音频中第二帧以上的任意一帧，编码器 113 可以先判断是否复用在先帧的代表虚拟扬声器对当前帧进行编码或判断是否进行虚拟扬声器搜索，确保连续帧之间的方位的连续性，并降低编码复杂度。本申请的实施例还可以包括 S540。

S540、编码器 113 根据在先帧的代表虚拟扬声器和当前帧判断是否进行虚拟扬声器搜索。

若编码器 113 确定进行虚拟扬声器搜索，执行 S510 至 S530。可选地，编码器 113 可以先执行 S510，即编码器 113 获取当前帧的代表系数，编码器 113 根据当前帧的代表系数和在先帧的代表虚拟扬声器的系数判断是否进行虚拟扬声器搜索，若编码器 113 确定进行虚拟扬声器搜索，再执行 S520 至 S530。

若编码器 113 确定不进行虚拟扬声器搜索，执行 S550。

S550、编码器 113 确定复用在先帧的代表虚拟扬声器对当前帧进行编码。

编码器 113 复用在先帧的代表虚拟扬声器和当前帧生成虚拟扬声器信号，对虚拟扬声器信号进行编码得到码流，向目的设备 120 发送码流，即执行 S450 和 S460。

判断是否进行虚拟扬声器搜索的具体方法可以参考下述图 9 中 S650 至 S680 的阐述。

S440、源设备 110 根据三维音频信号的当前帧和当前帧的代表虚拟扬声器生成虚拟扬声器信号。

源设备 110 根据当前帧的系数和当前帧的代表虚拟扬声器的系数生成虚拟扬声器信号。生成虚拟扬声器信号的具体方法可以参考现有技术，以及上述实施例中虚拟扬声器信号生成单元 350 的阐述。

S450、源设备 110 对虚拟扬声器信号进行编码得到码流。

源设备 110 可以对虚拟扬声器信号进行变换或量化等编码操作，生成码流，从而实现对待编码的三维音频信号进行数据压缩的目的。生成码流的具体方法可以参考现有技术，以及上述实施例中编码单元 360 的阐述。

S460、源设备 110 向目的设备 120 发送码流。

源设备 110 可以对原始音频全部编码完成后，向目的设备 120 发送原始音频的码流。或者，源设备 110 也可以以帧为单位，实时对三维音频信号进行编码处理，对一帧编码完成后发送一帧的码流。发送码流的具体方法可以参考现有技术，以及上述实施例中通信接口 114 和通信接口 124 的阐述。

S470、目的设备 120 对源设备 110 发送的码流进行解码，重建三维音频信号，得到重建后三维音频信号。

目的设备 120 接收到码流后，对码流进行解码得到虚拟扬声器信号，再根据候选虚拟扬声器集合和虚拟扬声器信号重建三维音频信号，得到重建后三维音频信号。目的设备 120 回放重建后三维音频信号。或者，目的设备 120 将重建后三维音频信号传

输给其他播放设备，由其他播放设备播放重建后三维音频信号，使得听音者置身于影院、音乐厅或虚拟场景等场所的“身临其境”的音响效果更加逼真。

为了增加连续帧之间的方位的连续性，克服了连续帧选取的虚拟扬声器的结果差异较大的问题，编码器 113 根据在先帧的代表虚拟扬声器的在先帧最终投票值调整候选虚拟扬声器集合中虚拟扬声器的当前帧初始投票值，获得虚拟扬声器的当前帧最终投票值。如图 6 所示，为本申请实施例提供的另一种选择虚拟扬声器方法的流程示意图。在这里由图 1 中源设备 110 中编码器 113 执行选择虚拟扬声器过程为例进行说明。其中，图 6 所述的方法流程是对图 5 中 S530 所包括的具体操作过程的阐述。如图 6 所示，该方法包括以下步骤。

S610、编码器 113 获取三维音频信号的当前帧的第一数量个当前帧初始投票值。

编码器 113 可以利用当前帧的代表系数对候选虚拟扬声器集合中每个虚拟扬声器进行投票，获取虚拟扬声器的当前帧初始投票值，依据投票值选取当前帧的代表虚拟扬声器，从而减小虚拟扬声器搜索的计算复杂度，以及减轻编码器的计算负担。

图 7 为本申请实施例提供的另一种三维音频信号编码方法的流程示意图。在这里由图 1 中源设备 110 中编码器 113 执行选择虚拟扬声器过程为例进行说明。其中，图 7 所述的方法流程是对图 5 中 S510 和 S520 所包括的具体操作过程的阐述。如图 7 所示，该方法包括以下步骤。

S6101、编码器 113 获取三维音频信号的当前帧的第四数量个系数，以及第四数量个系数的频域特征值。

假设三维音频信号是 HOA 信号，编码器 113 可以对 HOA 信号的当前帧进行采样，得到 $L\cdot(N+1)^2$ 个采样点，即得到第四数量个系数。N 表示 HOA 信号的阶数。示例地，假设 HOA 信号的当前帧的时长为 20 毫秒，编码器 113 根据 48KHz 频率对当前帧进行采样，得到时域上的 $960\cdot(N+1)^2$ 个采样点。采样点也可以称为时域系数。

三维音频信号的当前帧的频域系数可以是根据三维音频信号的当前帧的时域系数进行时频转换得到。时域转变为频域的方法不予限定。时域转变为频域的方法例如是修正的离散余弦变换（Modified Discrete Cosine Transform，MDCT），则可以得到频域上 $960\cdot(N+1)^2$ 个频域系数。频域系数也可以称为频谱系数或频点。

采样点的频域特征值满足 p(j)=norm(x(j))，其中，j=1,2…L，L 表示采样时刻的数量，x 表示三维音频信号的当前帧的频域系数，例如 MDCT 系数，norm 为求取二范数运算；x(j)表示第 j 个采样时刻的 $(N+1)^2$ 个采样点的频域系数。

S6102、编码器 113 根据第四数量个系数的频域特征值，从第四数量个系数中选取第三数量个代表系数。

编码器 113 将第四数量个系数指示的频谱范围划分为至少一个子带。其中，编码器 113 将第四数量个系数指示的频谱范围划分为一个子带，可以理解的，该一个子带的频谱范围等于第四数量个系数指示的频谱范围，相当于编码器 113 未对第四数量个系数指示的频谱范围进行划分。

如果编码器 113 将第四数量个系数指示的频谱范围划分为至少两个频带子带，在一种情形下，编码器 113 将第四数量个系数指示的频谱范围等分为至少两个子带，至少两个子带中每个子带包含相同数量的系数。

在另一情形下，编码器 113 对第四数量个系数指示的频谱范围进行不等分，划分得到的至少两个子带包含的系数的数量不同，或者划分得到的至少两个子带中每个子带包含的系数的数量均不同。例如，编码器 113 可以根据第四数量个系数指示的频谱范围中的低频范围、中频范围和高频范围，对第四数量个系数指示的频谱范围进行不等分，使得低频范围、中频范围和高频范围中每个频谱范围包括至少一个子带。低频范围内的至少一个子带中每个子带包含相同数量的系数。中频范围内的至少一个子带中每个子带包含相同数量的系数。高频范围内的至少一个子带中每个子带包含相同数量的系数。低频范围、中频范围和高频范围中三个频谱范围内的子带可以包含不同数量的系数。

进一步地，编码器 113 根据第四数量个系数的频域特征值，从第四数量个系数指示的频谱范围包含的至少一个子带选取代表系数，得到第三数量个代表系数。第三数量小于第四数量，第四数量个系数包含第三数量个代表系数。

例如，编码器 113 根据第四数量个系数指示的频谱范围包含的至少一个子带中每个子带中系数的频域特征值的从大到小的顺序，分别从所述每个子带中选取 Z 个代表系数，组合至少一个子带中的 Z 个代表系数，得到第三数量个代表系数，Z 为正整数。

又如，至少一个子带包括至少两个子带时，编码器 113 根据至少两个子带中每个子带内的第一候选系数的频域特征值确定每个子带各自的权重；并根据每个子带各自的权重分别调整每个子带内的第二候选系数的频域特征值，得到每个子带内的第二候选系数的调整后频域特征值，第一候选系数和第二候选系数为子带内的部分系数。编码器 113 根据至少两个子带内的第二候选系数的调整后频域特征值，以及至少两个子带内除第二候选系数之外的系数的频域特征值，确定第三数量个代表系数。

由于编码器从当前帧的全部系数中选取部分系数作为代表系数，利用较少数量的代表系数代替当前帧的全部系数从候选虚拟扬声器集合中选取代表虚拟扬声器，因此有效地降低了编码器搜索虚拟扬声器的计算复杂度，从而降低了对三维音频信号进行压缩编码的计算复杂度以及减轻了编码器的计算负担。

S6103、编码器 113 根据当前帧的第三数量个代表系数、候选虚拟扬声器集合和投票轮数确定第一数量个虚拟扬声器和第一数量个投票值。

投票轮数用于限定对虚拟扬声器进行投票的次数。投票轮数为大于或等于 1 的整数，且投票轮数小于或等于候选虚拟扬声器集合包含的虚拟扬声器的数量，以及投票轮数小于或等于编码器传输的虚拟扬声器信号的数量。例如，候选虚拟扬声器集合包括第五数量个虚拟扬声器，第五数量个虚拟扬声器包括第一数量个虚拟扬声器，第一数量小于或等于第五数量，投票轮数为大于或等于 1 的整数，且投票轮数小于或等于所述第五数量。虚拟扬声器信号也是指当前帧对应的当前帧的代表虚拟扬声器的传输通道。通常情况下虚拟扬声器信号的数量小于或等于虚拟扬声器的数量。

在一种可能的实现方式中，投票轮数可以是预先配置的，也可以是根据编码器的计算能力确定的，比如，投票轮数是根据编码器的编码速率和/或编码应用场景确定的。

在另一种可能的实现方式中，投票轮数是根据当前帧中指向性声源的数量确定的。例如，当声场中指向性声源数量为 2 时，设置投票轮数为 2。

本申请实施例提供了确定第一数量个虚拟扬声器和第一数量个投票值的三种可能

实现方式，下面对三种方式分别进行详述。

在第一种可能的实现方式中，投票轮数等于 1，编码器 113 采样到多个代表系数后，获取当前帧的每个代表系数对候选虚拟扬声器集合中所有虚拟扬声器的投票值，累加相同编号的虚拟扬声器的投票值，得到第一数量个虚拟扬声器和第一数量个投票值。可理解的，候选虚拟扬声器集合包括第一数量个虚拟扬声器。第一数量等于候选虚拟扬声器集合包括的虚拟扬声器的数量。假设候选虚拟扬声器集合包括第五数量个虚拟扬声器，则第一数量等于第五数量。第一数量个投票值包括候选虚拟扬声器集合中所有虚拟扬声器的投票值。编码器 113 可以将第一数量个投票值作为第一数量个虚拟扬声器的当前帧初始投票值，执行 S620 至 S640。

其中，虚拟扬声器与投票值一一对应，即一个虚拟扬声器对应一个投票值。例如，第一数量个虚拟扬声器包括第一虚拟扬声器，第一数量个投票值包括第一虚拟扬声器的投票值，第一虚拟扬声器与第一虚拟扬声器的投票值对应。第一虚拟扬声器的投票值用于表征对当前帧进行编码时使用第一虚拟扬声器的优先级。优先级也可以替换描述为倾向性，即第一虚拟扬声器的投票值用于表征对当前帧进行编码时使用第一虚拟扬声器的倾向性。可理解的，第一虚拟扬声器的投票值越大，表示第一虚拟扬声器的优先级越高或倾向性越高，相对于候选虚拟扬声器集合中比第一虚拟扬声器的投票值小的虚拟扬声器，编码器 113 更倾向选择第一虚拟扬声器对当前帧进行编码。

在第二种可能的实现方式中，与上述第一种可能的实现方式的区别在于，编码器 113 获取当前帧的每个代表系数对候选虚拟扬声器集合中所有虚拟扬声器的投票值后，从每个代表系数对候选虚拟扬声器集合中所有虚拟扬声器的投票值中选取部分投票值，累加部分投票值对应的虚拟扬声器中相同编号的虚拟扬声器的投票值，得到第一数量个虚拟扬声器和第一数量个投票值。可理解的，候选虚拟扬声器集合包括第一数量个虚拟扬声器。第一数量小于或等于候选虚拟扬声器集合包括的虚拟扬声器的数量。第一数量个投票值包括候选虚拟扬声器集合包括的部分虚拟扬声器的投票值，或者，第一数量个投票值包括候选虚拟扬声器集合包括的全部虚拟扬声器的投票值。

在第三种可能的实现方式中，与上述第二种可能的实现方式的区别在于，投票轮数为大于或等于 2 的整数，对于当前帧的每个代表系数，编码器 113 对候选虚拟扬声器集合中所有虚拟扬声器进行至少 2 轮次投票，每轮选择最大投票值的虚拟扬声器。对当前帧的每个代表系数对所有虚拟扬声器进行至少 2 轮次投票后，累加相同编号的虚拟扬声器的投票值，得到第一数量个虚拟扬声器和第一数量个投票值。

S620、编码器 113 根据第一数量个当前帧初始投票值，以及第六数量个在先帧最终投票值，获取第七数量个虚拟扬声器与当前帧对应的第七数量个当前帧最终投票值。

编码器 113 可以上述 S610 所述的方法，根据三维音频信号的当前帧、候选虚拟扬声器集合和投票轮数确定第一数量个虚拟扬声器和第一数量个投票值，进而，将第一数量个投票值作为第一数量个虚拟扬声器的当前帧初始投票值。

虚拟扬声器与当前帧初始投票值一一对应，即一个虚拟扬声器对应一个当前帧初始投票值。例如，第一数量个虚拟扬声器包括第一虚拟扬声器，第一数量个当前帧初始投票值包括第一虚拟扬声器的当前帧初始投票值，第一虚拟扬声器与第一虚拟扬声器的当前帧初始投票值对应。第一虚拟扬声器的当前帧初始投票值用于表征对当前帧

进行编码时使用第一虚拟扬声器的优先级。

第六数量个虚拟扬声器可以是编码器 113 对三维音频信号的在先帧进行编码所使用的在先帧的代表虚拟扬声器。在 S650 中，编码器 113 获取三维音频信号的当前帧与在先帧的代表虚拟扬声器集合的第一相关度时，在先帧的代表虚拟扬声器集合包括第六数量个虚拟扬声器。

具体地，编码器 113 根据第六数量个在先帧最终投票值，更新第一数量个当前帧初始投票值，即编码器 113 计算第一数量个虚拟扬声器与第六数量个虚拟扬声器中相同编号的虚拟扬声器的当前帧初始投票值和在先帧最终投票值之和，获取第七数量个虚拟扬声器与当前帧对应的第七数量个当前帧最终投票值。

在第一种可能的情形下，第一数量个虚拟扬声器包括第六数量个虚拟扬声器，且第一数量等于第六数量，第一数量个虚拟扬声器的编号与第六数量个虚拟扬声器的编号相同。可理解的，编码器 113 获取的第一数量个虚拟扬声器是第六数量个虚拟扬声器，第六数量个虚拟扬声器的在先帧最终投票值为第一数量个虚拟扬声器的在先帧最终投票值。编码器 113 可以利用第六数量个虚拟扬声器的在先帧最终投票值更新第一数量个虚拟扬声器的当前帧初始投票值。因此，第七数量个虚拟扬声器也是第一数量个虚拟扬声器，第七数量个当前帧最终投票值为第一数量个虚拟扬声器的在先帧最终投票值和第一数量个虚拟扬声器的当前帧初始投票值之和。

例如，假设第六数量个虚拟扬声器包括第一虚拟扬声器，第一数量个虚拟扬声器包括第一虚拟扬声器，第六数量个虚拟扬声器和第一数量个虚拟扬声器均不包含其他虚拟扬声器。编码器 113 可以根据第一虚拟扬声器的在先帧最终投票值，更新第一虚拟扬声器的当前帧初始投票值，以获得第一虚拟扬声器的当前帧最终投票值。第一虚拟扬声器的当前帧最终投票值为第一虚拟扬声器的在先帧最终投票值与第一虚拟扬声器的当前帧初始投票值之和。

在第二种可能的情形下，第一数量个虚拟扬声器包括第六数量个虚拟扬声器，且第一数量大于第六数量。可理解的，第一数量个虚拟扬声器还包括除了第六数量个虚拟扬声器之外的其他虚拟扬声器。编码器 113 可以利用第六数量个虚拟扬声器的在先帧最终投票值更新第一数量个虚拟扬声器中与第六数量个虚拟扬声器的编号相同的虚拟扬声器的当前帧初始投票值。因此，第七数量个虚拟扬声器包括第一数量个虚拟扬声器，且第七数量等于第一数量，第七数量个虚拟扬声器的编号与第一数量个虚拟扬声器的编号相同。第七数量个当前帧最终投票值包括第一数量个虚拟扬声器中与第六数量个虚拟扬声器的编号相同的虚拟扬声器的当前帧最终投票值，以及第一数量个虚拟扬声器中与第六数量个虚拟扬声器的编号不同的虚拟扬声器的当前帧最终投票值。

第一数量个虚拟扬声器中与第六数量个虚拟扬声器的编号相同的虚拟扬声器的当前帧最终投票值为第六数量个虚拟扬声器的在先帧最终投票值与第一数量个虚拟扬声器的当前帧初始投票值之和。第一数量个虚拟扬声器中与第六数量个虚拟扬声器的编号不同的虚拟扬声器的当前帧最终投票值为第一数量个虚拟扬声器中与第六数量个虚拟扬声器的编号不同的虚拟扬声器的当前帧初始投票值。

例如，假设第一数量个虚拟扬声器包括第一虚拟扬声器和第二虚拟扬声器，第六数量个虚拟扬声器包括第一虚拟扬声器，且第六数量个虚拟扬声器不包括第二虚拟扬

声器，则第二虚拟扬声器的当前帧最终投票值等于第二虚拟扬声器的当前帧初始投票值。编码器 113 可以根据第一虚拟扬声器的在先帧最终投票值，更新第一虚拟扬声器的当前帧初始投票值，以获得第一虚拟扬声器的当前帧最终投票值。第一虚拟扬声器的当前帧最终投票值为第一虚拟扬声器的在先帧最终投票值与第一虚拟扬声器的当前帧初始投票值之和。

在第三种可能的情形下，第一数量个虚拟扬声器包括第六数量个虚拟扬声器中部分虚拟扬声器，第六数量个虚拟扬声器还包括与第一数量个虚拟扬声器的编号不同的其他虚拟扬声器。因此，第七数量个虚拟扬声器包括第一数量个虚拟扬声器，以及第六数量个虚拟扬声器中与第一数量个虚拟扬声器的编号不同的虚拟扬声器。第七数量个当前帧最终投票值包括第一数量个虚拟扬声器的当前帧最终投票值，以及第六数量个虚拟扬声器中与第一数量个虚拟扬声器的编号不同的虚拟扬声器的当前帧最终投票值。

第一数量个虚拟扬声器的当前帧最终投票值包括第一数量个虚拟扬声器中与第六数量个虚拟扬声器的编号相同的虚拟扬声器的当前帧最终投票值。可选地，第一数量个虚拟扬声器的当前帧最终投票值还可以包括第一数量个虚拟扬声器中与第六数量个虚拟扬声器的编号不同的虚拟扬声器的当前帧最终投票值。

第六数量个虚拟扬声器中与第一数量个虚拟扬声器的编号不同的虚拟扬声器的当前帧最终投票值为第六数量个虚拟扬声器中与第一数量个虚拟扬声器的编号不同的虚拟扬声器的在先帧最终投票值。

例如，假设第六数量个虚拟扬声器包括第一虚拟扬声器和第三虚拟扬声器，第一数量个虚拟扬声器包括第一虚拟扬声器，且第一数量个虚拟扬声器不包括第三虚拟扬声器，则第三虚拟扬声器的当前帧最终投票值等于第三虚拟扬声器的在先帧最终投票值。编码器 113 可以根据第一虚拟扬声器的在先帧最终投票值，更新第一虚拟扬声器的当前帧初始投票值，以获得第一虚拟扬声器的当前帧最终投票值。第一虚拟扬声器的当前帧最终投票值为第一虚拟扬声器的在先帧最终投票值与第一虚拟扬声器的当前帧初始投票值之和。

在一些实施例中个，如图 8 所示，为本申请实施例提供的一种更新虚拟扬声器的当前帧初始投票值方法的流程示意图。

S810、编码器 113 根据第一调整参数调整第一虚拟扬声器的在先帧最终投票值，以得到第一虚拟扬声器的在先帧调整后投票值。

第一调整参数是根据在先帧中指向性声源的数量、对当前帧进行编码的编码速率和帧类型中至少一个确定的。第一虚拟扬声器的在先帧调整后投票值满足如下公式(6)。

$$VOTE_f_g' = VOTE_f_g \cdot w_1 \cdot w_2 \cdot w_3 \qquad \text{公式(6)}$$

其中，$VOTE_f_g'$表示在先帧调整后投票值集合，$VOTE_f_g$表示在先帧最终投票值集合，g 表示在先帧的代表虚拟扬声器集合。w_1表示与编码速率相关的参数，w_2表示与帧类型相关的参数，w_3表示与指向性声源的数量相关的参数。帧类型包括暂态帧或非暂态帧。

示例地，若编码速率小于或等于 128kbps，w_1=1；若编码速率大于 128kbps，w_1=0。若在先帧为暂态帧，w_2=1；若在先帧为非暂态帧，w_2=0。若指向性声源的数量大于预

设的虚拟扬声器信号的数量，w_3=0.8；若指向性声源的数量小于或等于预设的虚拟扬声器信号的数量，w_3=0.5。

S820、编码器 113 根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧初始投票值，获得第一虚拟扬声器的当前帧最终投票值。

第一虚拟扬声器的当前帧最终投票值为第一虚拟扬声器的在先帧调整后投票值与第一虚拟扬声器的当前帧初始投票值之和。第一虚拟扬声器的当前帧最终投票值满足如下公式(7)。

$$VOTE_M_g = VOTE_f'_g + VOTE_g \qquad 公式(7)$$

其中，$VOTE_M_g$表示当前帧最终投票值集合，$VOTE_f'_g$表示在先帧调整后投票值集合，$VOTE_g$表示当前帧初始投票值集合。

可选地，编码器 113 根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧初始投票值具体包括以下步骤。

S830、编码器 113 根据第二调整参数调整第一虚拟扬声器的当前帧初始投票值，得到第一虚拟扬声器的当前帧调整后投票值。

第一虚拟扬声器的当前帧调整后投票值满足如下公式(8)。

$$VOTE'_g = VOTE_g \cdot w_4 \qquad 公式(8)$$

其中，$VOTE'_g$表示当前帧调整后投票值集合，w_4表示第二调整参数。示例地，若$norm\left(VOTE_g\right) > norm\left(VOTE_f'_g\right)$，$w_4 = \frac{norm\left(VOTE_f'_g\right)}{norm\left(VOTE_g\right)} * 1.5$。可理解的，当前帧初始投票值大于在先帧调整后投票值的情况下，利用w_4表示放大在先帧调整后投票值。

若$norm\left(VOTE_g\right) \le norm\left(VOTE_f'_g\right)$，$w_4$=1。可理解的，当前帧初始投票值小于或等于在先帧调整后投票值的情况下，无需利用w_4表示放大在先帧调整后投票值。

第二调整参数是根据第一虚拟扬声器的在先帧调整后投票值和第一虚拟扬声器的当前帧初始投票值确定的。

S840、编码器 113 根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧调整后投票值，获得第一虚拟扬声器的当前帧最终投票值。

第一虚拟扬声器的当前帧最终投票值为第一虚拟扬声器的在先帧调整后投票值与第一虚拟扬声器的当前帧调整后投票值之和。第一虚拟扬声器的当前帧最终投票值满足如下公式(9)。

$$VOTE_M_g = VOTE_f'_g + VOTE'_g \qquad 公式(9)$$

其中，$VOTE_M_g$表示当前帧最终投票值集合，$VOTE_f'_g$表示在先帧调整后投票值集合，$VOTE'_g$表示当前帧调整后投票值集合。

S630、编码器 113 根据第七数量个当前帧最终投票值，从第七数量个虚拟扬声器中选取第二数量个当前帧的代表虚拟扬声器。

编码器 113 根据第七数量个当前帧最终投票值，从第七数量个虚拟扬声器中选取第二数量个当前帧的代表虚拟扬声器，而且，第二数量个当前帧的代表虚拟扬声器的当前帧最终投票值大于预设阈值。

编码器 113 也可以根据第七数量个当前帧最终投票值，从第七数量个虚拟扬声器

中选取第二数量个当前帧的代表虚拟扬声器。例如，按照第七数量个当前帧最终投票值的从大到小的顺序，从第七数量个当前帧最终投票值中确定第二数量个当前帧最终投票值，并将第七数量个虚拟扬声器中与第二数量个当前帧最终投票值对应的虚拟扬声器作为第二数量个当前帧的代表虚拟扬声器。

可选地，若第七数量个虚拟扬声器中不同编号的虚拟扬声器的投票值相同，且该不同编号的虚拟扬声器的投票值大于预设阈值，则编码器 113 可以将该不同编号的虚拟扬声器均作为当前帧的代表虚拟扬声器。

需要说明的是，第二数量小于第七数量。第七数量个虚拟扬声器包括第二数量个当前帧的代表虚拟扬声器。第二数量可以是预设的，或者，第二数量可以是根据当前帧的声场中声源的数量确定的，例如，第二数量可以直接等于当前帧的声场中声源的数量，或者是按照预设算法对当前帧的声场中声源的数量进行处理，将处理得到的数量作为第二数量；其中，预设算法可以根据需要进行设计，例如，预设算法可以是：第二数量=当前帧的声场中声源的数量+1，或第二数量=当前帧的声场中声源的数量-1等等。

另外，编码器 113 对当前帧的下一帧进行编码前，如果编码器 113 确定复用在先帧的代表虚拟扬声器对下一帧进行编码，编码器 113 可以将第二数量个当前帧的代表虚拟扬声器作为第二数量个在先帧的代表虚拟扬声器，利用第二数量个在先帧的代表虚拟扬声器对当前帧的下一帧进行编码。

S640、编码器 113 根据第二数量个当前帧的代表虚拟扬声器对当前帧进行编码，得到码流。

编码器 113 根据第二数量个当前帧的代表虚拟扬声器和当前帧生成虚拟扬声器信号；对虚拟扬声器信号进行编码得到码流。

在虚拟扬声器搜索过程中，由于真实声源的位置与虚拟扬声器的位置不一定重合，会导致虚拟扬声器不一定能够与真实声源形成一一对应关系，且由于在实际的复杂场景下，可能出现虚拟扬声器无法表征声场中独立声源情况，此时，帧与帧之间搜索到的虚拟扬声器可能会发生频繁跳变，这种频繁跳变会明显地影响听音者的听觉感受，导致解码重建后三维音频信号中出现明显的噪声。本申请的实施例提供的选择虚拟扬声器的方法通过继承在先帧的代表虚拟扬声器，即对于相同编号的虚拟扬声器，用在先帧最终投票值调整当前帧初始投票值，使得编码器更倾向于选择在先帧的代表虚拟扬声器，从而增强了帧之间的方位的连续性，另外调整参数确保在先帧最终投票值不会继承太过久远，避免算法无法适应声源移动等声场变化的场景。

此外，本申请实施例提供还一种选择虚拟扬声器的方法，编码器可以先判断是否可以复用在先帧的代表虚拟扬声器集合对当前帧进行编码，如果编码器复用在先帧的代表虚拟扬声器集合对当前帧进行编码，从而，避免编码器再执行虚拟扬声器搜索过程，有效地降低了编码器搜索虚拟扬声器的计算复杂度，因此降低了对三维音频信号进行压缩编码的计算复杂度以及减轻了编码器的计算负担。如果编码器不能复用在先帧的代表虚拟扬声器集合对当前帧进行编码，编码器再选取代表系数，利用当前帧的代表系数对候选虚拟扬声器集合中每个虚拟扬声器进行投票，依据投票值选取当前帧的代表虚拟扬声器，来达到降低了对三维音频信号进行压缩编码的计算复杂度以及减

轻了编码器的计算负担的目的。图 9 为本申请实施例提供的一种选择虚拟扬声器方法的流程示意图。在编码器 113 获取第一数量个虚拟扬声器与三维音频信号的当前帧对应的第一数量个当前帧初始投票值，即 S610 之前，如图 9 所示，该方法包括以下步骤。

S650、编码器 113 获取三维音频信号的当前帧与在先帧的代表虚拟扬声器集合的第一相关度。

在先帧的代表虚拟扬声器集合包含的第六数量个虚拟扬声器，第六数量个虚拟扬声器包含的虚拟扬声器为对三维音频信号的在先帧进行编码所使用的在先帧的代表虚拟扬声器。第一相关度用于表征对当前帧进行编码时复用在先帧的代表虚拟扬声器集合的优先级。优先级也可以替换描述为倾向性，即第一相关度用于确定对当前帧进行编码时是否复用在先帧的代表虚拟扬声器集合。可理解的，在先帧的代表虚拟扬声器集合的第一相关度越大，表示在先帧的代表虚拟扬声器集合的优先级越高或倾向性越高，编码器 113 更倾向选择在先帧的代表虚拟扬声器对当前帧进行编码。

S660、编码器 113 判断第一相关度是否满足复用条件。

若第一相关度不满足复用条件，表示编码器 113 更倾向进行虚拟扬声器搜索，根据当前帧的代表虚拟扬声器对当前进行编码，执行 S610，编码器 113 获取第一数量个虚拟扬声器与三维音频信号的当前帧对应的第一数量个当前帧初始投票值。

可选地，编码器 113 也可以在根据第四数量个系数的频域特征值，从第四数量个系数中选取第三数量个代表系数之后，将第三数量个代表系数中最大的代表系数作为获取第一相关度的当前帧的系数，则编码器 113 获取当前帧的第三数量个代表系数中最大的代表系数与在先帧的代表虚拟扬声器集合的第一相关度，若第一相关度不满足复用条件，执行 S6103，即编码器 113 根据第一数量个投票值，从第一数量个虚拟扬声器中选取第二数量个当前帧的代表虚拟扬声器。

若第一相关度满足复用条件，表示编码器 113 更倾向选择在先帧的代表虚拟扬声器对当前帧进行编码，编码器 113 执行 S670 和 S680。

S670、编码器 113 根据在先帧的代表虚拟扬声器集合和当前帧生成虚拟扬声器信号。

S680、编码器 113 对虚拟扬声器信号进行编码得到码流。

本申请实施例提供的选择虚拟扬声器的方法，利用当前帧的代表系数与在先帧的代表虚拟扬声器的相关度判断是否进行虚拟扬声器搜索，在确保当前帧的代表虚拟扬声器的相关度的选择准确度情况下，有效地降低了编码端的复杂度。

可以理解的是，为了实现上述实施例中的功能，编码器包括了执行各个功能相应的硬件结构和/或软件模块。本领域技术人员应该很容易意识到，结合本申请中所公开的实施例描述的各示例的单元及方法步骤，本申请能够以硬件或硬件和计算机软件相结合的形式来实现。某个功能究竟以硬件还是计算机软件驱动硬件的方式来执行，取决于技术方案的特定应用场景和设计约束条件。

上文中结合图 1 至图 9，详细描述了根据本实施例所提供的三维音频信号编码方法，下面将结合图 10 和图 11，描述根据本实施例所提供的三维音频信号编码装置和编码器。

图 10 为本实施例提供的可能的三维音频信号编码装置的结构示意图。这些三维音

频信号编码装置可以用于实现上述方法实施例中编码三维音频信号的功能，因此也能实现上述方法实施例所具备的有益效果。在本实施例中，该三维音频信号编码装置可以是如图 1 所示的编码器 113，或者如图 3 所示的编码器 300，还可以是应用于终端设备或服务器的模块（如芯片）。

如图 10 所示，三维音频信号编码装置 1000 包括通信模块 1010、系数选择模块 1020、虚拟扬声器选择模块 1030、编码模块 1040 和存储模块 1050。三维音频信号编码装置 1000 用于实现上述图 6 至图 9 中所示的方法实施例中编码器 113 的功能。

通信模块 1010 用于获取三维音频信号的当前帧。可选地，通信模块 1010 也可以接收其他设备获取的三维音频信号的当前帧；或者从存储模块 1050 获取三维音频信号的当前帧。三维音频信号的当前帧为 HOA 信号；系数的频域特征值是依据依据 HOA 信号的系数确定的。

虚拟扬声器选择模块 1030 用于获取三维音频信号的当前帧的第一数量个当前帧初始投票值，第一数量个虚拟扬声器与当前帧初始投票值一一对应，第一数量个虚拟扬声器包括第一虚拟扬声器，第一虚拟扬声器的当前帧初始投票值用于表征对当前帧进行编码时使用第一虚拟扬声器的优先级。

虚拟扬声器选择模块 1030，还用于根据第一数量个当前帧初始投票值，以及第六数量个在先帧最终投票值，获取第七数量个虚拟扬声器与当前帧对应的第七数量个当前帧最终投票值，第七数量个虚拟扬声器包括第一数量个虚拟扬声器，且第七数量个虚拟扬声器包括第六数量个虚拟扬声器，第六数量个虚拟扬声器与第六数量个在先帧最终投票值一一对应，第六数量个虚拟扬声器是用于对三维音频信号的在先帧进行编码时使用的虚拟扬声器。

若第一数量个虚拟扬声器包括第二虚拟扬声器，且第六数量个虚拟扬声器不包括第二虚拟扬声器，则第二虚拟扬声器的当前帧最终投票值等于第二虚拟扬声器的当前帧初始投票值；或，若第六数量个虚拟扬声器包括第三虚拟扬声器，且第一数量个虚拟扬声器不包括第三虚拟扬声器，则第三虚拟扬声器的当前帧最终投票值等于第三虚拟扬声器的在先帧最终投票值。

当三维音频信号编码装置 1000 用于实现图 6 至图 9 所示的方法实施例中编码器 113 的功能时，虚拟扬声器选择模块 1030 用于实现 S610 至 S630，以及 S650 至 S680 的相关功能。

例如，虚拟扬声器选择模块 1030 根据第一虚拟扬声器的在先帧最终投票值更新第一虚拟扬声器的当前帧初始投票值时，具体用于：根据第一调整参数调整第一虚拟扬声器的在先帧最终投票值，以得到第一虚拟扬声器的在先帧调整后投票值；根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧初始投票值。

又如，虚拟扬声器选择模块 1030 根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧初始投票值时，具体用于：根据第二调整参数调整第一虚拟扬声器的当前帧初始投票值，得到第一虚拟扬声器的当前帧调整后投票值；根据第一虚拟扬声器的在先帧调整后投票值更新第一虚拟扬声器的当前帧调整后投票值。

第一调整参数是根据在先帧中指向性声源的数量、对当前帧进行编码的编码速率和帧类型中至少一个确定的。

第二调整参数是根据第一虚拟扬声器的在先帧调整后投票值和第一虚拟扬声器的当前帧初始投票值确定的。

当三维音频信号编码装置1000用于实现图7所示的方法实施例中编码器113的功能时，系数选择模块1020用于实现S6101和S6102的相关功能。具体地，系数选择模块1020获取当前帧的第三数量个代表系数时，具体用于：获取当前帧的第四数量个系数，以及第四数量个系数的频域特征值；根据第四数量个系数的频域特征值，从第四数量个系数中选取第三数量个代表系数，第三数量小于第四数量。

编码模块1140用于根据第二数量个当前帧的代表虚拟扬声器对当前帧进行编码，得到码流。

当三维音频信号编码装置1000用于实现图6至图9所示的方法实施例中编码器113的功能时，编码模块1140用于实现S630的相关功能。示例地，编码模块1140具体用于根据第二数量个当前帧的代表虚拟扬声器和当前帧生成虚拟扬声器信号；对虚拟扬声器信号进行编码得到码流。

存储模块1050用于存储与三维音频信号相关的系数，候选虚拟扬声器集合，在先帧的代表虚拟扬声器集合，以及选取的系数和虚拟扬声器等，以便于编码模块1040对当前帧进行编码得到码流，并将码流传输至解码器。

应理解的是，本申请实施例的三维音频信号编码装置1000可以通过专用集成电路（application-specific integrated circuit，ASIC）实现，或可编程逻辑器件（programmable logic device，PLD）实现，上述PLD可以是复杂程序逻辑器件（complex programmable logical device，CPLD），现场可编程门阵列（field-programmable gate array，FPGA），通用阵列逻辑（generic array logic，GAL）或其任意组合。也可以通过软件实现图6至图9所示的三维音频信号编码方法时，三维音频信号编码装置1000及其各个模块也可以为软件模块。

有关上述通信模块1010、系数选择模块1020、虚拟扬声器选择模块1030、编码模块1040和存储模块1050更详细的描述可以参考图6至图9所示的方法实施例中相关描述直接得到，这里不加赘述。

图11为本实施例提供的一种编码器1100的结构示意图。如图11所示，编码器1100包括处理器1110、总线1120、存储器1130和通信接口1140。

应理解，在本实施例中，处理器1110可以是中央处理器（central processing unit，CPU），该处理器1110还可以是其他通用处理器、数字信号处理器（digital signal processing，DSP）、ASIC、FPGA或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者是任何常规的处理器等。

处理器还可以是图形处理器（graphics processing unit，GPU）、神经网络处理器（neural network processing unit，NPU）、微处理器或一个或多个用于控制本申请方案程序执行的集成电路。

通信接口1140用于实现编码器1100与外部设备或器件的通信。在本实施例中，通信接口1140用于接收三维音频信号。

总线1120可以包括一通路，用于在上述组件（如处理器1110和存储器1130）之间传送信息。总线1120除包括数据总线之外，还可以包括电源总线、控制总线和状态

信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线 1120。

作为一个示例，编码器 1100 可以包括多个处理器。处理器可以是一个多核（multi-CPU）处理器。这里的处理器可以指一个或多个设备、电路、和/或用于处理数据（例如计算机程序指令）的计算单元。处理器 1110 可以调用存储器 1130 存储的与三维音频信号相关的系数，候选虚拟扬声器集合，在先帧的代表虚拟扬声器集合，以及选取的系数和虚拟扬声器等。

值得说明的是，图 11 中仅以编码器 1100 包括 1 个处理器 1110 和 1 个存储器 1130 为例，此处，处理器 1110 和存储器 1130 分别用于指示一类器件或设备，具体实施例中，可以根据业务需求确定每种类型的器件或设备的数量。

存储器 1130 可以对应上述方法实施例中用于存储与三维音频信号相关的系数，候选虚拟扬声器集合，在先帧的代表虚拟扬声器集合，以及选取的系数和虚拟扬声器等信息的存储介质，例如，磁盘，如机械硬盘或固态硬盘。

上述编码器 1100 可以是一个通用设备或者是一个专用设备。例如，编码器 1100 可以是基于 X86、ARM 的服务器，也可以为其他的专用服务器，如策略控制和计费（policy control and charging，PCC）服务器等。本申请实施例不限定编码器 1100 的类型。

应理解，根据本实施例的编码器 1100 可对应于本实施例中的三维音频信号编码装置 1100，并可以对应于执行根据图 6 至图 9 中任一方法中的相应主体，并且三维音频信号编码装置 1100 中的各个模块的上述和其它操作和/或功能分别为了实现图 6 至图 9 中的各个方法的相应流程，为了简洁，在此不再赘述。

本实施例中的方法步骤可以通过硬件的方式来实现，也可以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）、可编程只读存储器（programmable ROM，PROM）、可擦除可编程只读存储器（erasable PROM，EPROM）、电可擦除可编程只读存储器（electrically EPROM，EEPROM）、寄存器、硬盘、移动硬盘、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。另外，该 ASIC 可以位于网络设备或终端设备中。当然，处理器和存储介质也可以作为分立组件存在于网络设备或终端设备中。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机程序或指令。在计算机上加载和执行所述计算机程序或指令时，全部或部分地执行本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、网络设备、用户设备或者其它可编程装置。所述计算机程序或指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，所述计算机程序或指令可以从一个网站站点、计算机、服务器或数据中心通过有线或无线方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的

任何可用介质或者是集成一个或多个可用介质的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质，例如，软盘、硬盘、磁带；也可以是光介质，例如，数字视频光盘（digital video disc，DVD）；还可以是半导体介质，例如，固态硬盘（solid state drive，SSD）。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权 利 要 求 书

1. 一种三维音频信号编码方法，其特征在于，包括：

获取三维音频信号的当前帧的第一数量个当前帧初始投票值，其中，第一数量个虚拟扬声器与所述当前帧初始投票值一一对应，所述第一数量个虚拟扬声器包括第一虚拟扬声器，所述第一虚拟扬声器的当前帧初始投票值用于表征所述第一虚拟扬声器的优先级；

根据所述第一数量个当前帧初始投票值，以及第六数量个在先帧最终投票值，获取第七数量个虚拟扬声器与所述当前帧对应的第七数量个当前帧最终投票值，所述第七数量个虚拟扬声器包括所述第一数量个虚拟扬声器，且所述第七数量个虚拟扬声器包括所述第六数量个虚拟扬声器，第六数量个虚拟扬声器与所述第六数量个在先帧最终投票值一一对应，所述第六数量个虚拟扬声器是用于对所述三维音频信号的在先帧进行编码时使用的虚拟扬声器；

根据所述第七数量个当前帧最终投票值，从所述第七数量个虚拟扬声器中选取第二数量个当前帧的代表虚拟扬声器，所述第二数量小于所述第七数量；

根据所述第二数量个当前帧的代表虚拟扬声器对所述当前帧进行编码，得到码流。

2. 根据权利要求 1 所述的方法，其特征在于，若所述第一数量个虚拟扬声器包括第二虚拟扬声器，且所述第六数量个虚拟扬声器不包括所述第二虚拟扬声器，则所述第二虚拟扬声器的当前帧最终投票值等于所述第二虚拟扬声器的当前帧初始投票值；或

若所述第六数量个虚拟扬声器包括第三虚拟扬声器，且所述第一数量个虚拟扬声器不包括所述第三虚拟扬声器，则所述第三虚拟扬声器的当前帧最终投票值等于所述第三虚拟扬声器的在先帧最终投票值。

3. 根据权利要求 1 或 2 所述的方法，其特征在于，若所述第六数量个虚拟扬声器包括所述第一虚拟扬声器，所述根据所述第一数量个当前帧初始投票值以及第六数量个虚拟扬声器与所述三维音频信号的在先帧对应的第六数量个在先帧投票值，获取第七数量个虚拟扬声器与所述当前帧对应的第七数量个当前帧最终投票值包括：

根据所述第一虚拟扬声器的在先帧最终投票值更新所述第一虚拟扬声器的当前帧初始投票值，以获得所述第一虚拟扬声器的当前帧最终投票值。

4. 根据权利要求 3 所述的方法，其特征在于，所述根据所述第一虚拟扬声器的在先帧最终投票值更新所述第一虚拟扬声器的当前帧初始投票值包括：

根据第一调整参数调整所述第一虚拟扬声器的在先帧最终投票值，以得到所述第一虚拟扬声器的在先帧调整后投票值；

根据所述第一虚拟扬声器的在先帧调整后投票值更新所述第一虚拟扬声器的当前帧初始投票值。

5. 根据权利要求 4 所述的方法，其特征在于，所述根据所述第一虚拟扬声器的在先帧调整后投票值更新所述第一虚拟扬声器的当前帧初始投票值包括：

根据第二调整参数调整所述第一虚拟扬声器的当前帧初始投票值，得到第一虚拟扬声器的当前帧调整后投票值；

根据所述第一虚拟扬声器的在先帧调整后投票值更新所述第一虚拟扬声器的当前

帧调整后投票值。

6. 根据权利要求 4 或 5 所述的方法，其特征在于，所述第一调整参数是根据所述在先帧中指向性声源的数量、对所述当前帧进行编码的编码速率和所述当前帧的帧类型中至少一个确定的。

7. 根据权利要求 5 所述的方法，其特征在于，所述第二调整参数是根据所述第一虚拟扬声器的在先帧调整后投票值和所述第一虚拟扬声器的当前帧初始投票值确定的。

8. 根据权利要求 1-7 中任一项所述的方法，其特征在于，所述第二数量是预设的，或者，所述第二数量是根据所述当前帧确定的。

9. 根据权利要求 1-8 中任一项所述的方法，其特征在于，所述获取第一数量个虚拟扬声器与三维音频信号的当前帧对应的第一数量个当前帧初始投票值包括：

根据所述当前帧的第三数量个代表系数、候选虚拟扬声器集合和投票轮数确定所述第一数量个虚拟扬声器，以及所述第一数量个当前帧初始投票值，所述候选虚拟扬声器集合包括第五数量个虚拟扬声器，所述第五数量个虚拟扬声器包括所述第一数量个虚拟扬声器，所述第一数量小于或等于所述第五数量，所述投票轮数为大于或等于 1 的整数，且所述投票轮数小于或等于所述第五数量。

10. 根据权利要求 9 所述的方法，其特征在于，所述根据所述当前帧的第三数量个代表系数、候选虚拟扬声器集合和投票轮数确定所述第一数量个虚拟扬声器，以及所述第一数量个当前帧初始投票值之前，所述方法还包括：

获取所述当前帧的第四数量个系数，以及所述第四数量个系数的频域特征值；

根据所述第四数量个系数的频域特征值，从所述第四数量个系数中选取所述第三数量个代表系数，所述第三数量小于所述第四数量。

11. 根据权利要求 10 所述的方法，其特征在于，所述方法还包括：

获取所述当前帧与在先帧的代表虚拟扬声器集合的第一相关度，所述在先帧的代表虚拟扬声器集合包括所述第六数量个虚拟扬声器，所述第六数量个虚拟扬声器为对所述在先帧进行编码所使用的在先帧的代表虚拟扬声器，所述第一相关度用于确定对所述当前帧进行编码时是否复用所述在先帧的代表虚拟扬声器集合；

若所述第一相关度不满足复用条件，获取所述三维音频信号的当前帧的第四数量个系数，以及所述第四数量个系数的频域特征值。

12. 根据权利要求 1-11 中任一项所述的方法，其特征在于，所述三维音频信号的当前帧为高阶立体混响 HOA 信号；所述当前帧的系数的频域特征值是依据 HOA 信号的系数确定的。

13. 一种三维音频信号编码装置，其特征在于，包括：

虚拟扬声器选择模块，用于获取三维音频信号的当前帧的第一数量个当前帧初始投票值，其中，第一数量个虚拟扬声器与所述当前帧初始投票值一一对应，所述第一数量个虚拟扬声器包括第一虚拟扬声器，所述第一虚拟扬声器的当前帧初始投票值用于表征所述第一虚拟扬声器的优先级；

所述虚拟扬声器选择模块，还用于根据所述第一数量个当前帧初始投票值，以及第六数量个在先帧最终投票值，获取第七数量个虚拟扬声器与所述当前帧对应的第七数量个当前帧最终投票值，所述第七数量个虚拟扬声器包括所述第一数量个虚拟扬声

器，且所述第七数量个虚拟扬声器包括所述第六数量个虚拟扬声器，第六数量个虚拟扬声器与所述第六数量个在先帧最终投票值一一对应，所述第六数量个虚拟扬声器是用于对所述三维音频信号的在先帧进行编码时使用的虚拟扬声器；

所述虚拟扬声器选择模块，还用于根据所述第七数量个当前帧最终投票值，从所述第七数量个虚拟扬声器中选取第二数量个当前帧的代表虚拟扬声器，所述第二数量小于所述第七数量；

编码模块，用于根据所述第二数量个当前帧的代表虚拟扬声器对所述当前帧进行编码，得到码流。

14. 根据权利要求 13 所述的装置，其特征在于，若所述第一数量个虚拟扬声器包括第二虚拟扬声器，且所述第六数量个虚拟扬声器不包括所述第二虚拟扬声器，则所述第二虚拟扬声器的当前帧最终投票值等于所述第二虚拟扬声器的当前帧初始投票值；或

若所述第六数量个虚拟扬声器包括第三虚拟扬声器，且所述第一数量个虚拟扬声器不包括所述第三虚拟扬声器，则所述第三虚拟扬声器的当前帧最终投票值等于所述第三虚拟扬声器的在先帧最终投票值。

15. 根据权利要求或 13 或 14 所述的装置，其特征在于，若所述第六数量个虚拟扬声器包括所述第一虚拟扬声器，所述虚拟扬声器选择模块根据所述第一数量个当前帧初始投票值以及第六数量个虚拟扬声器与所述三维音频信号的在先帧对应的第六数量个在先帧投票值，获取第七数量个虚拟扬声器与所述当前帧对应的第七数量个当前帧最终投票值时，具体用于：

根据所述第一虚拟扬声器的在先帧最终投票值更新所述第一虚拟扬声器的当前帧初始投票值，以获得所述第一虚拟扬声器的当前帧最终投票值。

16. 根据权利要求 15 所述的装置，其特征在于，所述虚拟扬声器选择模块根据所述第一虚拟扬声器的在先帧最终投票值更新所述第一虚拟扬声器的当前帧初始投票值时，具体用于：

根据第一调整参数调整所述第一虚拟扬声器的在先帧最终投票值，以得到所述第一虚拟扬声器的在先帧调整后投票值；

根据所述第一虚拟扬声器的在先帧调整后投票值更新所述第一虚拟扬声器的当前帧初始投票值。

17. 根据权利要求 16 所述的装置，其特征在于，所述虚拟扬声器选择模块根据所述第一虚拟扬声器的在先帧调整后投票值更新所述第一虚拟扬声器的当前帧初始投票值时，具体用于：

根据第二调整参数调整所述第一虚拟扬声器的当前帧初始投票值，得到第一虚拟扬声器的当前帧调整后投票值；

根据所述第一虚拟扬声器的在先帧调整后投票值更新所述第一虚拟扬声器的当前帧调整后投票值。

18. 根据权利要求 16 或 17 所述的装置，其特征在于，所述第一调整参数是根据所述在先帧中指向性声源的数量、对所述当前帧进行编码的编码速率和所述当前帧的帧类型中至少一个确定的。

19. 根据权利要求 17 所述的装置，其特征在于，所述第二调整参数是根据所述第一虚拟扬声器的在先帧调整后投票值和所述第一虚拟扬声器的当前帧初始投票值确定的。

20. 根据权利要求 13-19 中任一项所述的装置，其特征在于，所述第二数量是预设的，或者，所述第二数量是根据所述当前帧确定的。

21. 根据权利要求 13-20 中任一项所述的装置，其特征在于，所述虚拟扬声器选择模块获取第一数量个虚拟扬声器与三维音频信号的当前帧对应的第一数量个当前帧初始投票值时，具体用于：

根据所述当前帧的第三数量个代表系数、候选虚拟扬声器集合和投票轮数确定所述第一数量个虚拟扬声器，以及所述第一数量个当前帧初始投票值，所述候选虚拟扬声器集合包括第五数量个虚拟扬声器，所述第五数量个虚拟扬声器包括所述第一数量个虚拟扬声器，所述第一数量小于或等于所述第五数量，所述投票轮数为大于或等于 1 的整数，且所述投票轮数小于或等于所述第五数量。

22. 根据权利要求 21 所述的装置，其特征在于，所述装置还包括系数选择模块；

所述系数选择模块，用于获取所述当前帧的第四数量个系数，以及所述第四数量个系数的频域特征值；

所述系数选择模块，还用于根据所述第四数量个系数的频域特征值，从所述第四数量个系数中选取所述第三数量个代表系数，所述第三数量小于所述第四数量。

23. 根据权利要求 22 所述的装置，其特征在于，所述虚拟扬声器选择模块，还用于：

获取所述当前帧与在先帧的代表虚拟扬声器集合的第一相关度，所述在先帧的代表虚拟扬声器集合包括所述第六数量个虚拟扬声器，所述第六数量个虚拟扬声器包含的虚拟扬声器为对所述在先帧进行编码所使用的在先帧的代表虚拟扬声器，所述第一相关度用于确定对所述当前帧进行编码时是否复用所述在先帧的代表虚拟扬声器集合；

若所述第一相关度不满足复用条件，获取所述三维音频信号的当前帧的第四数量个系数，以及所述第四数量个系数的频域特征值。

24. 根据权利要求 13-23 中任一项所述的装置，其特征在于，所述三维音频信号的当前帧为高阶立体混响 HOA 信号；所述当前帧的系数的频域特征值是依据 HOA 信号的系数确定的。

25. 一种编码器，其特征在于，所述编码器包括至少一个处理器和存储器，其中，所述存储器用于存储计算机程序，使得所述计算机程序被所述至少一个处理器执行时实现如权利要求 1-12 中任一项所述的三维音频信号编码方法。

26. 一种系统，其特征在于，所述系统包括如权利要求 25 所述的编码器，以及解码器，所述编码器用于执行上述权利要求 1-12 中任一项所述的方法的操作步骤，所述解码器用于解码所述编码器生成的码流。

27. 一种计算机程序，其特征在于，所述计算机程序被执行时实现如权利要求 1-12 中任一项所述的三维音频信号编码方法。

28. 一种计算机可读存储介质，其特征在于，包括计算机软件指令；当计算机软件指令在编码器中运行时，使得所述编码器执行如权利要求 1-12 中任一项所述的三维

音频信号编码方法。

29. 一种计算机可读存储介质，其特征在于，包括如权利要求 1-12 中任一项所述的三维音频信号编码方法所获得的码流。