



(12) 发明专利

(10) 授权公告号 CN 111149155 B

(45) 授权公告日 2023.10.10

(21) 申请号 201880060064.7

H04S 7/00 (2006.01)

(22) 申请日 2018.07.13

G06F 3/01 (2006.01)

G10L 21/0272 (2006.01)

(65) 同一申请的已公布的文献号  
申请公布号 CN 111149155 A

(56) 对比文件

(43) 申请公布日 2020.05.12

CN 103250207 A, 2013.08.14

CN 102859584 A, 2013.01.02

(30) 优先权数据

CN 102326417 A, 2012.01.18

17181488.2 2017.07.14 EP

CN 106104680 A, 2016.11.09

(85) PCT国际申请进入国家阶段日  
2020.03.16

CN 103460285 A, 2013.12.18

CN 103635964 A, 2014.03.12

(86) PCT国际申请的申请数据  
PCT/EP2018/069140 2018.07.13

CN 104244164 A, 2014.12.24

WO 2015086377 A1, 2015.06.18

WO 2019013924 A1, 2019.01.17

(87) PCT国际申请的公布数据  
W02019/012131 EN 2019.01.17

Robert L.Bleidt.development of the mPEG-H TV audio system for ATSC 3.0.《IEEE TRANSACTIONS ON BROADCASTING》.2017,

(73) 专利权人 弗劳恩霍夫应用研究促进协会  
地址 德国慕尼黑

Thiergart O..An Acoustical Zoom based on Informed Spatial Filtering.《Int. Workshop on Acoustic Signal Enhancement》

(72) 发明人 于尔根·赫勒 伊曼纽尔·哈毕兹

.2014, (续)

(74) 专利代理机构 北京清亦华知识产权代理事务所(普通合伙) 11201

审查员 张彩

专利代理师 宋融冰

(51) Int.Cl.

G10L 19/008 (2006.01)

权利要求书2页 说明书23页 附图15页

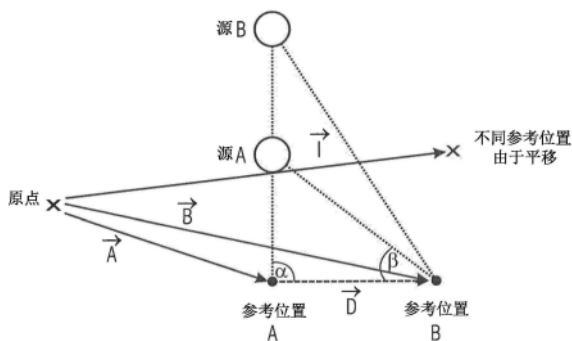
(54) 发明名称

使用多点声场描述生成经增强的声场描述的装置及方法

指示参考位置和另一参考位置或两个参考位置之一以及在两个参考位置之间延伸的向量。

(57) 摘要

一种用于生成经增强的声场描述的装置,包括:声场生成器(100,250,260),用于生成至少一个声场描述,至少一个声场描述指示关于至少一个参考位置的声场;以及元数据生成器(110),用于生成与声场的空间信息相关的元数据,其中,至少一个声场描述和元数据构成经增强的声场描述。至少一个声场描述在某个情况下包括与至少一个参考位置相关的第一声场描述以及与不同于参考位置的另一参考位置相关的第二声场描述,以及与空间信息相关的元数据,该元数据



CN 111149155 B

[转续页]

[接上页]

(56) 对比文件

彭钰林. 三维音频中若干算法研究.《中国优

秀硕士学位论文全文数据库(信息科技辑)》

.2013,

1. 一种用于生成经增强的声场描述的装置,包括:

声场生成器(100),用于生成指示声场关于第一参考位置的第一声场描述以及指示所述声场关于第二参考位置的第二声场描述,所述第二参考位置与所述第一参考位置不同;以及

元数据生成器(110),用于生成与所述声场的空间信息相关的元数据;

其中,所述元数据生成器(110)被配置为确定所述第一声场描述的第一几何描述和所述第二声场描述的第二几何描述中的至少一个作为所述元数据,以及

其中,所述第一声场描述、所述第二声场描述和包括所述第一几何描述和所述第二几何描述中的至少一个的所述元数据构成所述经增强的声场描述。

2. 如权利要求1所述的装置,

其中所述元数据生成器(110)被配置为确定所述第一参考位置和所述第二参考位置中的一个作为所述第一几何描述以及确定所述第一参考位置和所述第二参考位置中的另一个作为所述第二几何描述,或者确定所述第一参考位置与所述第二参考位置之间的距离或所述第一参考位置与所述第二参考位置之间的位置向量作为所述元数据。

3. 如权利要求2所述的装置,

其中所述第一声场描述是第一高保真度立体声响复制描述,以及其中所述第二声场描述是第二高保真度立体声响复制描述,或者其中所述第一声场描述和所述第二声场描述是高保真度立体声响复制描述或定向音频编码描述中的一个。

4. 如权利要求1所述的装置,

其中所述第一几何描述是关于从预定原点指向所述第一声场描述的所述第一参考位置的第一向量的信息;以及

其中所述第二几何描述是关于从所述预定原点指向第二声场描述的所述第二参考位置的第二向量的信息。

5. 如权利要求1所述的装置,

其中所述第一几何描述是关于从预定原点指向所述第一声场描述的所述第一参考位置的第一向量和从所述预定原点指向所述第二声场描述的所述第二参考位置的第二向量中的一个的信息,以及其中所述第二几何描述包括关于所述第一参考位置和所述第二参考位置之间的向量的信息。

6. 如权利要求1所述的装置,

其中所述第一参考位置和所述第二参考位置中的一个表示所述第一几何描述的预定原点,并且其中所述第二几何描述包括关于所述预定原点与所述第一参考位置和所述第二参考位置中的另外一个之间的向量的信息。

7. 如权利要求1所述的装置,

其中所述声场生成器(100)被配置为,使用真实麦克风设备或使用虚拟麦克风技术通过声音合成,生成所述第一声场描述或所述第二声场描述。

8. 一种生成经增强的声场描述的方法,包括:

生成指示声场关于第一参考位置的第一声场描述和指示所述声场关于第二参考位置的第二声场描述,所述第二参考位置与所述第一参考位置不同;以及

生成与所述声场的空间信息相关的元数据;

其中,生成所述元数据包括确定所述第一声场描述的第一几何描述和所述第二声场描述的第二几何描述中的至少一个作为所述元数据,以及

其中,所述第一声场描述、所述第二声场描述和包括所述第一几何描述和所述第二几何描述中的至少一个的所述元数据构成所述经增强的声场描述。

9.一种计算机可读介质,具有存储在其上的计算机程序,当在计算机或处理器上运行时,所述计算机程序用于执行如权利要求8所述的方法。

## 使用多点声场描述生成经增强的声场描述的装置及方法

### 技术领域

[0001] 本发明涉及音频处理,并且特别是,涉及关于诸如麦克风或虚拟麦克风位置的参考位置限定的声场的音频处理。

### 背景技术

[0002] 高保真度立体声响复制(Ambisonics)信号包括声场的截断球谐波分解。高保真度立体声响复制有不同的风格。存在“传统的”高保真度立体声响复制[31],其现在被称为“一阶高保真度立体声响复制(FOA,First-Order Ambisonics)”并且包括四个信号(即,一个全向信号和多达三个8字形定向信号)。最近的高保真度立体声响复制变体被称为“更高阶的高保真度立体声响复制(HOA,Higher-Order Ambisonics)”,并且以携带更多信号为代价而提供经增强的空间分辨率和更大的聆听者最佳位置区域。通常,完全定义的N阶HOA表示由 $(N+1)^2$ 个信号组成。

[0003] 与高保真度立体声响复制理念相关,已经设想到定向音频编码(DirAC, Directional Audio Coding)表示,以以更紧凑的参数样式表示FOA或HOA声音场景。更具体地,空间声音场景由一个(或多个)发送的音频声道表示,其表示声学场景的降混以及在每个时间-频率(TF)区间中的方向和扩散性的相关联辅助信息。关于DirAC的更多信息可以在[32,33]中找到。

[0004] DirAC[32]可以和不同麦克风系统及任意扬声器设置一起被使用。DirAC系统的目的是使用多声道/3D扬声器系统尽可能精确地再现现有声学环境的空间印象。在所选择的环境中,响应(连续声音或脉冲响应)由全向麦克风(W)和能够测量声音的到达方向和声音的扩散性的一组麦克风测量。一种常见的方法是应用与对应笛卡尔坐标轴对齐的三个8字形麦克风(X,Y,Z)[34]。一种实现此的方式是使用直接产生所有期望响应的声场麦克风。W、X、Y和Z信号也可以从一组离散的全向麦克风计算出来。

[0005] 在DirAC中,声音信号首先将被划分到频率信道。依据每个频率信道处的时间测量声音方向和扩散性。在传输时,一个或多个音频声道以及分析的方向和扩散数据一起被发送。在合成时,施加到扬声器的音频可以是例如全向声道W,或者针对每个扬声器的声音可以被计算为W、X、Y和Z的加权和,其形成具有针对每个扬声器的特定定向特性的信号。每个音频声道被划分成频率信道,然后依据分析的扩散性被可选择地划分为扩散串流和非扩散串流。利用一种技术,扩散串流被再现,该技术产生声音场景的扩散感知,例如,在双耳线索编码(Binaural Cue Coding)中使用的去相关(decorrelation)技术[35-37]。利用一种技术(例如,VBAP[38]),非扩散声音被再现,该技术目的是根据方向数据产生类似点的虚拟源。

[0006] 在具有有限自由度的六个自由度(6DoF)中,三种导航技术在[39]中被提出。给定单个高保真度立体声响复制信号,单个高保真度立体声响复制信号使用以下方法计算:1)在虚拟扬声器数组内仿真HOA回放和聆听者移动,2)沿着平面波计算和平移,以及3)重新扩展关于聆听者的声场。

[0007] 此外,参考(例如)于2009年11月11-13日在日本宫城县Zao举行的International Workshop on the Principles and Applications of Spatial Hearing中由V.Pulkki等人著作的出版物“Directional Audio Coding-Perception-Based Reproduction of Spatial Sound”中描述的DirAC技术。该参考文献描述了定向音频编码作为参考位置相关的声场处理的示例,特别是作为用于空间音频处理的感知激励技术。在电话会议、定向滤波、和虚拟听觉环境中,它具有在空间声音的捕获、编码和重新合成中的应用。

[0008] 声音场景的再现通常聚焦在扬声器设置上,因为这是在私人场所(例如起居室和专业环境即电影院)的典型再现。此处,场景与再现几何的关系是静态的,因为它伴随着强迫聆听者在正面方向上观看的二维图像。随后,声音和视觉对象的空间关系被限定并固定在产生时间。

[0009] 在虚拟现实(VR)中,通过允许用户在场景中自由移动而明确地实现沉浸。因此,有必要追踪用户的移动并将视觉和听觉再现调整到该用户的位置。通常,用户佩戴头戴式显示器(HMD)和耳机。对于利用耳机的沉浸式体验,音频必须被双耳化。双耳化是对人类头部、耳朵和上部躯干如何依据源的方向和距离改变源的声音进行的模拟。这通过针对它们的相对方向的信号与头部相关转移函数(HRTF)的卷积来实现[1,2]。双耳化也使声音看起来是来自场景而不是来自头部内[3]。已经成功达到的常见情况是360°视频再现[4,5]。此处,用户戴着HMD或是手持平板计算机或手机。通过移动她/他的头部或设备,用户可以在任何方向环顾四周。这是三个自由度(3DoF)场景,因为用户有三个移动度(俯仰、偏转、翻滚)。在视觉上,这通过将视频投影在用户周围的球体上来实现。音频通常用靠近视频相机的空间麦克风(例如,一阶高保真度立体声响复制(FOA))记录[6]。在高保真度立体声响复制领域,用户的头部旋转以直接的方式进行调整[7]。然后例如音频被渲染到放置在用户周围的虚拟扬声器。这些虚拟扬声器信号然后被双耳化。

[0010] 现代VR应用允许六个自由度(6DoF)。除了头部旋转之外,用户可以四处移动,从而导致在三个空间维度上平移她/他的位置。6DoF再现受到步行区域的整体尺寸的限制。在许多情况下,该区域相当小,例如传统的起居室。6DoF在VR游戏中经常遇到。这里,整个场景是利用计算机生成的图像(CGI)合成的。音频通常使用基于对象的渲染而被生成,其中基于追踪数据利用距离相依增益和与用户的相对方向来渲染每个音频对象。可以通过人工混响(artificial reverberation)和衍射来增强真实性[8,9,10]。

[0011] 关于所记录的内容,对于令人信服的视听6DoF再现,存在一些明显的挑战。在空间平移领域中空间声音操纵的早期示例是“声学变焦”技术[11,12]。此处,聆听者位置被虚拟地移动到所记录的视觉场景中,类似于放大图像。用户选择一个方向或图像部分,然后可以从平移点收听。这需要所有到达方向(DoA)相对于原始的非缩放再现而改变。

[0012] 用于已经使用空间分布记录位置的所记录的内容的6DoF再现的方法已经被提出。对于视频,相机数组可以被采用以生成光场渲染(light-field rendering)[13]。对于音频,类似的设置采用分布式麦克风数组或高保真度立体声响复制麦克风。已经表明,从这种记录可以生成放置在任意位置处的“虚拟麦克风”的信号[14]。

[0013] 为了以技术上方便的方式实现这种空间声音修改,可以采用参数声音处理或编码技术(参见[15]的概述)。定向音频编码(DirAC)[16]是一种受欢迎的方法,用于将记录转换为表示,该表示由音频频谱和关于声音方向和扩散性的参数辅助信息组成。它用于声学变

焦[11]和虚拟麦克风[14]应用。

[0014] 这里提出的方法能够从单个FOA麦克风的记录实现6DoF再现。来自单个空间位置的记录已被用于3DoF再现或声学变焦。但是,就发明人所知,到目前为止还没有提出用于从这种数据进行交互式、完全6DoF再现的方法。通过在记录中集成关于声源的距离的信息,可以实现6DoF再现。该距离信息被合并到DirAC的参数表示中,使得聆听者的所改变的角度被正确映射。

[0015] 没有任何高保真度立体声响复制声场表示(无论是常规FOA或HOA高保真度立体声响复制还是DirAC风格的参数声场表示)提供足够的信息,以允许6DoF应用所需的聆听者位置的平移,因为在声音场景中的对象距离和绝对对象位置都不是以这些格式确定的。应该注意的是,聆听者位置的移位可以转化为声音场景在相反方向上的等效移位。

[0016] 在6DoF中移动时的典型问题在图1b中示出。让我们假设使用高保真度立体声响复制在位置A处描述声音场景。在这种情况下,来自源A和源B的声音从相同方向到达,即它们具有相同的到达方向(DOA)。如果移动到位置B处,源A和源B的DOA是不同的。使用声场的标准高保真度立体声响复制描述,即没有附加信息,在给定在位置A处的高保真度立体声响复制信号时,不可能计算在位置B处的高保真度立体声响复制信号。

## 发明内容

[0017] 本发明的目的是一方面提供经增强的声场描述或另一方面提供经修改的声场描述的生成,从而允许改进的、或灵活的、或有效的处理。

[0018] 该目的通过以下实施例中描述的用于生成经增强的声场描述的装置、用于生成经修改的声场描述的装置、生成经增强的声场描述的方法,生成经修改的声场描述的方法,计算机程序或经增强的声场描述来实现。

[0019] 本发明基于以下发现:与参考位置相关的典型声场描述需要附加信息,以便这些声场描述可以被处理,使得可以计算与原始参考位置无关但是与另一参考位置相关的经修改的声场描述。至此,与该声场的空间信息相关的元数据(metadata)被生成,并且该元数据与该声场描述一起对应于经增强的声场描述,经增强的声场描述可以例如被发送或存储。为了从声场描述和元数据生成经修改的声场描述,具体地,元数据与声场描述的空间信息相关,使用该空间信息、声场描述以及指示从参考位置到不同参考位置的平移的平移信息,计算经修改的声场描述。因此,由声场描述和与声场描述下的该声场的空间信息相关的元数据所组成的经增强的声场描述被处理,以获得经修改的声场描述,该经修改的声场描述与由附加平移信息限定的不同参考位置相关,附加平移信息例如可以在解码器侧被提供或使用。

[0020] 然而,本发明不仅涉及编码器/解码器场景,也可以被应用于应用中,其中基本上在一个且同一个位置发生经增强的声场描述的生成和经修改的声场描述的生成。例如,经修改的声场描述可以是经修改的声场本身的描述,或者实际上是声道信号、双耳信号中的经修改的声场的描述,或者再次是参考位置相关的声场的描述,然而该参考位置相关的声场现在与新或不同参考位置而不是原始参考位置相关。例如,这样的应用将处于虚拟现实场景中,其中存在声场描述以及元数据以及其中聆听者从给定声场的参考位置移出并移动到不同参考位置,并且其中,然后,在虚拟区域中四处移动的聆听者的声场被计算以对应于

然而现在在用户移动到的不同参考位置处的声场。

[0021] 在特定实施例中,经增强的声场描述具有与(第一)参考位置相关的第一声场描述以及与不同于(第一)参考位置的另一(第二)参考位置相关的第二声场描述,并且元数据具有关于该参考位置和该另一参考位置的信息,诸如从预定原点指向这些参考位置的向量。可选地,元数据可以是指向该参考位置或该另一参考位置的单个向量和在该两个参考位置之间延伸的向量,两个不同声场描述与该两个参考位置相关。

[0022] 声场描述可以是非参数声场描述,如一阶高保真度立体声响复制描述或更高阶的高保真度立体声响复制描述。可选地或另外地,声场描述可以是DirAC描述或其他参数声场描述,或者一个声场描述例如可以是参数声场描述而另一个声场描述例如可以是非参数声场描述。

[0023] 因此,声场描述可以针对每个声场描述生成声场的DirAC描述,具有针对不同时间-频率区间的一个或多个降混信号和各个方向数据以及可选的扩散数据。在此上下文中,元数据生成器被配置为针对两个声场描述生成几何元数据,使得可以从该元数据中识别参考位置和附加参考位置。然后,为了生成经增强或修改的声场描述,可以从两个声场描述中提取各个源并执行附加处理。

[0024] 高保真度立体声响复制已成为虚拟、增强和混合现实应用环境中用于3D音频的最常用的格式之一。已经开发各种各样的音频获取和产生工具,它们以高保真度立体声响复制格式生成输出信号。为了在交互式虚拟现实(VR)应用中呈现高保真度立体声响复制编码内容,高保真度立体声响复制格式被转换为双耳信号或声道用于再现。在上述应用中,聆听者通常能够以交互方式改变在所呈现场景中的他/她的朝向至如此程度以使得他/她能够在实现三个自由度(3DoF,即,俯仰、偏转角及翻滚)的声音场景中旋转他/她的头部并且仍然可以体验到合适的音质。这通过根据头部朝向在渲染之前旋转声音场景来实现,这可以以低计算复杂度实现并且是高保真度立体声响复制表示的优点。然而,在诸如VR的新兴应用中,期望允许用户在声音场景中自由移动而不仅仅是朝向的改变(所谓的“六个自由度”或6DoF)。结果,需要信号处理以改变声音场景的角度(即,以沿着x轴、y轴或z轴在声音场景内虚拟地移动)。然而,高保真度立体声响复制的一个主要缺点是过程从声音场景中的单个角度描述了声场。具体来说,它不包含关于声音场景中的声源的实际位置的信息,其将允许移位声音场景(“平移”),正如它是6DoF所需的。本发明的描述提供了高保真度立体声响复制的几个扩展,以克服该问题并且还促进平移,并因此实现真正的6DoF。

[0025] 一阶高保真度立体声响复制(FOA)记录可以通过耳机处理和再现。它们可以被旋转以考虑聆听者头部朝向。然而,虚拟现实(VR)系统允许聆听者以六个自由度(6DoF)移动,即三个旋转自由度加三个过渡自由度。此处,声源的视倾角(apparent angle)和距离取决于聆听者位置。一种促进6DoF的技术被描述。特别地,FOA记录使用参数模型来描述,该参数模型基于聆听者位置和关于到源的距离的信息而被修改。该方法通过听力测试、比较其中聆听者可以自由移动的合成声音场景的不同双耳渲染来进行评估。

[0026] 在进一步较佳的实施例中,经增强的声场描述通过输出界面输出,输出界面用于生成用于传输或存储的输出信号,其中,对于时间帧,输出信号包括在该时间帧内从声场和空间信息所得出的一个或多个音频信号。特别地,声场生成器在进一步的实施例中适于从声场得出方向数据,该方向数据指的是针对时间段或频率区间的声音的到达方向,并且元

数据生成器被配置为得出空间信息,作为将距离信息与方向数据相关联的数据项。

[0027] 特别地,在这样的实施例中,输出界面被配置为生成输出信号,使得时间帧的数据项被链接到不同频率区间的方向数据。

[0028] 在另一实施例中,声场生成器还被配置为针对声场的时间帧的多个频率区间生成扩散信息,其中,元数据生成器被配置为,当扩散值低于预定或自适应阈值时,仅针对频率区间生成与预定值不同的或者与无穷大不同的距离信息,或者不管怎样针对频率区间生成距离值。因此,对于具有高扩散性的时间/频率区间,根本不生成任何距离值,或者生成由解码器以某种方式解释的预定距离值。因此,确保对于具有高扩散性的时间/频率区间,任何距离相关的渲染不被执行,因为高扩散性指示对于这样的时间/频率区间,声音不是来自某个局部的源而来自任何方向,并且因此,无论是在原始参考位置还是在不同或新参考位置处感知声场,声音都是相同的。

[0029] 关于声场计算器,较佳实施例包括:用于向经修改的声场提供平移信息或指示预期聆听者的旋转的旋转信息的平移界面,用于将元数据提供给声场计算器的元数据提供器和用于将声场描述提供给声场计算器的声场供应器,以及附加的用于输出包括经修改的声场描述和经修改的元数据的经修改的声场的输出界面,该经修改的元数据是使用平移信息从元数据得出的,或该输出界面输出多个扬声器声道,每个扬声器声道与预定义扬声器位置相关,或者该输出界面输出经修改的声场的双耳表示。

[0030] 在实施例中,声场描述包括多个声场分量。多个声场分量包括全向分量和至少一个定向分量。这样的声场描述例如是具有全向分量和三个定向分量X、Y、Z的一阶高保真度立体声响复制声场描述,或者这样的声场是更高阶的高保真度立体声响复制描述,其包括全向分量、关于X、Y和Z方向的三个定向分量、以及另外的与X、Y、Z方向之外的其他方向相关的定向分量。

[0031] 在实施例中,该装置包括分析器,用于分析声场分量,以针对不同的时间或频率区间得出到达方向信息。该装置还具有平移变换器,用于使用DoA信息和元数据为每个频率或时间区间计算经修改的DoA信息,其中元数据与将距离和在两个声场描述中都包含的源相关联的深度图相关,如通过例如使用关于两个不同参考位置的两个角度和距离/位置或参考位置的三角测量处理来获得。这可以应用于时间帧的全频带表示或不同频率区间。

[0032] 此外,声场计算器具有距离补偿器,用于使用距离补偿信息计算经修改的声场,该距离补偿信息取决于使用元数据计算的距离并且取决于与时间或频率区间相关联的新距离,其中,元数据对于源的每个频率或时间区间是相同的,而源针对每个或一些时间/频率区间是不同的,新距离与经修改的DoA信息相关。

[0033] 在实施例中,声场计算器计算从参考位置指向通过声场的分析所获得的声源的第一向量。此外,声场计算器计算从不同参考位置指向声源的第二向量,并且此计算使用第一向量和平移信息来完成,其中平移信息限定从参考位置到不同参考位置的平移向量。并且然后,使用该第二向量计算从不同参考位置到声源的距离。

[0034] 此外,声场计算器被配置为除了平移信息之外还接收旋转信息,该旋转信息指示聆听者头部在由俯仰、偏转和翻滚所给出的三个旋转方向中的一个上的旋转。声场计算器然后被配置为执行旋转变换,以使用旋转信息旋转声场的经修改的到达方向数据,其中,从通过声场描述的声音分析所获得的到达方向数据和平移信息,得出经修改的到达方向数

据。

[0035] 在实施例中,声场计算器被配置为通过声音分析从声场描述以及与参考位置相关的源信号的方向确定源信号。

[0036] 然后,计算与不同参考位置相关的声源的新方向,并且此计算使用元数据来完成,并且然后计算与不同参考位置相关的声源的距离信息,以及然后,使用声源的距离信息和新方向来合成经修改的声场。

[0037] 在实施例中,通过朝由关于再现设置的新方向信息所给出的方向移动声源信号来执行声场合成,并且在执行该移动操作之前或执行该移动操作之后,使用距离信息完成声源信号的按比例缩放。

[0038] 在另一实施例中,声源信号的扩散部分被添加到声源信号的直达部分,在被添加该扩散部分之前,通过距离信息来修改该直达部分。

[0039] 特别地,较佳地,以频谱表示执行声源合成,其中针对每个频率区间计算新方向信息,其中针对每个频率区间计算距离信息,并且其中使用针对频率区间的音频信号的对每个频率区间的直达合成是使用针对频率区间的音频信号来执行的,从新方向信息得出的针对频率区间的移动增益和从针对频率区间的距离信息所得出的针对频率区间的缩放因子被执行。

[0040] 此外,使用从来自频率区间的音频信号所得出的扩散音频信号以及使用由针对频率区间的信号分析所得出的扩散参数来执行扩散合成,并且然后,直达信号和扩散信号被组合以获得针对时间或频率区间的合成音频信号,并且然后,使用针对其他时间/频率区间的音频信号,执行频率-时间转换,以获得时域合成音频信号作为经修改的声场。

[0041] 因此,通常,声场计算器被配置为针对每个声源,合成与不同参考位置相关的声场,例如,通过针对每个源使用源信号的新方向来处理源信号,以获得与该不同/新参考位置相关的源信号的声场描述。此外,使用方向信息,在处理源信号之前或处理源信号之后,该源信号被修改。并且,最后,将针对源的声场描述相加在一起以获得与该不同参考位置相关的经修改的声场。

[0042] 在进一步的实施例中,并且特别地,用于从声场描述和与该声场描述的空间信息相关的元数据生成经修改的声场描述的实施例中,声场计算器使用关于第一声场描述的空间信息、使用关于第二声场描述的空间信息、以及使用指示参考位置到不同参考位置的平移的平移信息来计算经修改的声场。特别地,元数据可以例如是指向声场描述的参考位置的向量和从相同原点指向第二声场描述的另一参考位置的另一向量。

[0043] 为了得到平移信息,通过对该第一和第二声场描述应用源分离、或波束成形、或一般任何种类的声源分析来生成对象。然后,不管这些对象是宽带对象还是针对各个时间/频率区间的对象,计算所有对象的到达方向信息。然后,将从不同声场描述中提取的对象彼此匹配,以便找到至少一个匹配对象,即在该第一和该第二声场描述中都出现的对象。举例而言,借助于使用对象信号和/或到达方向信息或其他信息的相关性或一致性计算来执行该匹配。

[0044] 因此,对于匹配对象,过程的结果是存在与参考位置相关的第一DoA信息以及与另一参考位置相关的第二DoA信息。然后,使用关于该参考位置或关联元数据中所包括的参考位置的信息,基于三角测量,计算匹配对象的位置,尤其是匹配对象到该参考位置或另一参

考位置的距离。

[0045] 该信息,以及特别是匹配对象的位置信息然后被用于基于估计位置和期望位置(即在平移之后),使用距离补偿处理来修改每个匹配对象。为了针对新聆听者位置计算新DoA信息,来自两个参考位置的旧DoA信息和平移信息被使用。基本上,因为每个匹配对象都在两个声场描述中出现,可以对两个各个声场描述执行该处理。然而,根据较佳实施例,具有最接近平移之后的新聆听者位置的参考位置的声场描述被使用。

[0046] 然后,新DoA被用于计算与不同参考位置(即,用户已经移动到的参考位置)相关的匹配对象的新声场描述。然后,为了也并入非匹配对象,但是也使用旧DoA信息计算这些对象的声场描述。并且最后,通过将所有各个声场描述加在一起生成经修改的声场。

[0047] 通过对虚拟高保真度立体声响复制信号应用单个旋转,可以实现任何朝向变化。

[0048] 因此,元数据不用于直接提供对象到参考位置的距离。相反地,元数据被提供以用于识别两个或更多个声场描述中的每一个的参考位置、并且基于例如三角测量处理步骤计算参考位置和某个匹配对象之间的距离。

## 附图说明

[0049] 本发明的较佳实施例随后参照附图描述,其中:

[0050] 图1a是用于生成经增强的声场描述的装置的较佳实施例;

[0051] 图1b是解释本发明下的示例性问题的图示说明;

[0052] 图2是用于生成经增强的声场描述的装置的较佳实施方式;

[0053] 图3a示出包括音频数据和音频数据的辅助信息的经增强的声场描述;

[0054] 图3b示出经增强的声场的进一步图示说明,该经增强的声场包括音频数据和与每个声场描述的空间信息(如几何信息)相关的元数据;

[0055] 图4a示出用于生成经修改的声场描述的装置的实施方式;

[0056] 图4b示出用于生成经修改的声场描述的装置的另一实施方式;

[0057] 图4c示出具有参考位置/地点A、另一参考位置/地点B、以及由于平移的不同参考位置的场景;

[0058] 图5示出一般意义上的空间音频的6DoF再现;

[0059] 图6a示出用于实现声场计算器的较佳实施例;

[0060] 图6b示出用于计算关于新/不同参考位置的声源的新DoA和新距离的较佳实施方式;

[0061] 图6c示出6DoF再现的较佳实施例,包括用于(例如为每个各个声场描述)生成经增强的声场描述的装置和用于为匹配源生成经修改的声场描述的装置;

[0062] 图7示出用于选择第一和第二声场描述中的一个用于宽带或窄带对象的经修改的声场的计算的较佳实施例;

[0063] 图8示出用于从如单声道信号的音频信号和到达方向数据生成声场描述的示例性设备;

[0064] 图9示出声场计算器的另一较佳实施例;

[0065] 图10示出用于生成经修改的声场描述的装置的较佳实施方式;

[0066] 图11示出用于生成经修改的声场描述的装置的另一较佳实施方式;

[0067] 图12a示出现有技术的DirAC分析实施方式;以及

[0068] 图12b示出现有技术的DirAC合成实施方式。

### 具体实施方式

[0069] 为了针对上述高保真度立体声响复制/DirAC表示实现6DoF应用,需要提供用于平移处理的缺失信息的方式扩展这些表示。应注意的是,该扩展可以例如1) 将对象的距离或位置添加到现有场景表示,和/或2) 添加能促进分离各个对象的处理的信息。

[0070] 此外,实施例的目的是保留/重新使用现有(非参数或参数)高保真度立体声响复制系统的结构,以在如下意义上提供与这些表示/系统的向后兼容性:

[0071] • 扩展表示可被转换为现有的非扩展表示(例如,用于渲染);以及

[0072] • 在使用扩展表示工作时,允许重新使用现有的软件和硬件实施方式。

[0073] 在下文中,将描述几种方法,即一种有限(但非常简单)的方法和三种不同的扩展高保真度立体声响复制格式,以实现6DoF。

[0074] 使用两个或更多个高保真度立体声响复制信号来描述声音场景,每个高保真度立体声响复制信号描述不同位置处的声音场景,或者换句话说,从不同的角度描述声音场景。假设相对位置是已知的。在声音场景中的期望位置处的经修改的高保真度立体声响复制信号是从输入的高保真度立体声响复制信号生成的。基于信号或基于参数的方法可被用于在期望位置处生成虚拟高保真度立体声响复制信号。

[0075] 多点高保真度立体声响复制表示的概念适用于传统和参数(DirAC风格)的高保真度立体声响复制。

[0076] 在基于信号的平移实施例中,使用以下步骤计算在期望位置(即,平移之后)处的虚拟高保真度立体声响复制信号:

[0077] 1. 通过将源分离应用于每个传统的高保真度立体声响复制信号来生成对象。

[0078] 2. 针对每个传统的高保真度立体声响复制信号,计算所有对象的DOA。

[0079] 3. 从一个传统的高保真度立体声响复制信号中提取的对象与从其他传统的高保真度立体声响复制信号中提取的对象进行匹配。基于对应DOA和/或信号(例如,借助于相关性/一致性)执行该匹配。

[0080] 4. 基于三角测量来估计匹配对象的位置。

[0081] 5. 基于估计位置和期望位置(即,平移之后),使用距离补偿滤波器修改每个匹配对象(单声道输入)。

[0082] 6. 针对每个匹配对象,计算在期望位置(即,平移之后)处的DOA。该DOA由DOA' 表示。

[0083] 7. 针对每个匹配对象,计算高保真度立体声响复制对象信号。该高保真度立体声响复制对象信号被生成,使得匹配对象具有到达方向DOA' 。

[0084] 8. 针对每个非匹配对象,计算高保真度立体声响复制对象信号。该高保真度立体声响复制对象信号被生成,使得非匹配对象具有到达方向DOA。

[0085] 9. 通过将所有高保真度立体声响复制对象信号加在一起来获得虚拟高保真度立体声响复制信号。

[0086] 根据另一实施例,在基于参数的平移实施例中使用以下步骤计算在期望位置(即,

平移之后)处的虚拟高保真度立体声响复制信号:

[0087] 1. 假设声场模型。声场可以被分解为一个或多个直达声音分量和扩散声音分量。直达声音分量包括信号和位置信息(例如,在极坐标或笛卡尔坐标中)。可选地,声场可以被分解为一个或多个直达/主要声音分量和残余声音分量(单声道或多声道)。

[0088] 2. 使用输入的高保真度立体声响复制信号,估计所假定声场模型的信号分量和参数。

[0089] 3. 依据在声音场景中的期望平移或期望位置,修改信号分量和/或参数。

[0090] 4. 使用经修改的信号分量和经修改的参数,生成虚拟高保真度立体声响复制信号。

[0091] 生成多点高保真度立体声响复制信号对于计算机生成及产生的内容来说,以及在经由麦克风数组或空间麦克风(例如,B格式麦克风)的自然记录的上下文中是简单的。在实施例中,较佳地在步骤2之后执行源匹配或在步骤3之前执行三角测量计算。此外,两个实施例的一个或多个步骤也可以用在对应的其他实施例中。

[0092] 通过对虚拟高保真度立体声响复制信号应用单个旋转,可以实现朝向的改变。

[0093] 图1a示出用于生成经增强的声场描述的装置,该装置包括声场(描述)生成器100,用于生成至少一个声场描述,至少一个声场描述指示关于至少一个参考位置的声场。此外,该装置包括元数据生成器110,用于生成与声场的空间信息相关的元数据。该元数据接收声场,或者替代地或另外地,接收关于声源的分离信息,作为输入。

[0094] 声场描述生成器100和元数据生成器110两者的输出构成经增强的声场描述。在一个实施例中,声场描述生成器100和元数据生成器110两者的输出可以在组合器120或输出界面120内被组合,以获得经增强的声场描述,经增强的声场描述包括由元数据生成器110生成的声场的空间元数据或空间信息。

[0095] 图1b示出本发明所解决的情况。例如,位置A是至少一个参考位置,并且声场由源A和源B生成,并且位于该位置A处的例如某个实际或虚拟麦克风检测来自源A和源B的声音。声音是来自发射声音源的声音的迭加。这表示由声场描述生成器所生成的声场描述。

[0096] 另外,通过某种实施方式,元数据生成器将得出关于源A的空间信息和关于源B的另一空间信息,诸如这些源到参考位置(诸如位置A)的距离。

[0097] 自然地,参考位置另外也可以是位置B。然后,实际或虚拟麦克风将被放置在位置B处,并且声场描述将是,例如,由一阶高保真度立体声响复制分量、或更高阶的高保真度立体声响复制分量、或任何其他声音分量所表示的声场,任何其他声音分量具有潜力以描述关于至少一个参考位置(即位置B)的声场。

[0098] 然后,元数据生成器可以生成声源A到位置B的距离或者源B到位置B的距离作为关于声源的信息。当然,关于声源的替代信息可以是关于参考位置的绝对或相对位置。参考位置可以是一般坐标系统的原点,或者可以位于与一般坐标系统的原点有限定关系的位置处。

[0099] 其他元数据可以是一个声源的绝对位置和关于第一声源的另一声源的相对位置等等。

[0100] 图2示出用于生成经增强的声场描述的装置,其中声场生成器包括用于第一声场的声场生成器250、用于第二声场的声场生成器260、以及用于一个或多个声场的任意数量

的声场生成器,该一个或多个声场例如是第三、第四等声场。另外,元数据被配置为计算并向组合器120转发关于第一声场和第二声场的信息。所有的该信息被组合器120使用,以便生成经增强的声场描述。因此,组合器120还被配置为输出界面,以生成经增强的声场描述。

[0101] 图3a示出作为数据流的经增强的声场描述,其包括第一声场描述330、第二声场描述340以及与其相关联的元数据350,该元数据350包括关于第一声场描述和第二声场描述的信息。例如,第一声场描述可以是B格式描述、或更高阶描述、或任何其他描述,该任何其他描述允许确定以全频带表示或频率选择表示的声源的定向分布。因此,例如第一声场描述330和第二声场描述340也可以是针对不同参考位置的参数声场描述,其具有例如针对不同时间/频率区间的降混信号和到达方向数据。

[0102] 然而,该第一和第二声场描述的几何信息350分别对于包括在第一声场描述330中的所有源或者对于在第二声场描述340中的源是相同的。因此,当示例性地在第一声场描述330中存在三个源和关于第一声场描述存在几何信息时,该几何信息对于第一声场描述中的三个源是相同的。类似地,例如当在第二声场描述中存在五个源时,则被包括在元数据350中的关于第二声场的几何信息对于第二声场描述中的所有源是相同的。

[0103] 图3b示出图3a的元数据350的示例性构造。在一个实施例中,参考位置351可以被包括在元数据中。然而,在参考位置信息351也可以被省略的情况下,这不一定是必要的。

[0104] 对于第一声场,给出第一几何信息,其可以例如是关于图4c中所示的向量A的信息,该向量A从原点指向与第一声场相关的参考位置/地点A。

[0105] 例如,第二几何信息可以是关于从原点指向与第二声场描述相关的第二参考位置/地点B的向量B的信息。

[0106] A和B是两个声场描述的参考位置或记录位置。

[0107] 替代的几何信息例如可以是关于在参考位置A和另一参考位置B之间延伸的向量D的信息和/或原点和从该原点指向两个点之一的向量。因此,包括在元数据中的几何信息可以包括向量A和向量D、或者可以包括向量B和向量D、或者可以包括向量A和向量B而没有向量D、或者可以包括其他信息,由此,可以在某个三维坐标系中识别参考位置A和考位置B。然而,相同的考虑附加地被应用于二维声音描述以及特别地在图4c中示出,图4c仅显示二维情况。

[0108] 图4a示出用于从声场描述以及与该声场描述的空间信息相关的元数据生成经修改的声场描述的装置的较佳实施方式。特别地,该装置包括声场计算器420,其使用元数据、声场描述、以及平移信息生成经修改的声场,该平移信息指示从参考位置到不同参考位置的平移。

[0109] 在一个实施例中,声场计算器420连接到输入界面400,用于接收经增强的声场描述,例如,关于图1a或图2所讨论的,然后输入界面400一方面分离声场描述,即由图1a的块100或图2的块210所生成的。此外,输入界面400将元数据从经增强的声场描述分离,即图3a的项350或图3b的可选的351和352至354。

[0110] 此外,平移界面410从聆听者获得平移信息和/或附加或分离的旋转信息。平移界面410的实施方式可以是头部追踪单元,其不仅追踪在虚拟现实环境中的头部的旋转,而且还追踪头部从一个位置(即图1b中的位置A)到另一个位置(即图1b中的位置B)的平移。

[0111] 图4b示出与图1a类似的另一实施方式,但与编码器/解码器场景无关,但与一般场

景相关,其中由元数据提供器402指示的元数据供应、由声场提供器404指示的声场供应在没有分离经编码或增强的声场描述的某个输入界面的情况下完成,但例如,在虚拟现实应用中存在的实际场景中全部完成。然而,本发明不限于虚拟现实应用,而是还可以在任何其他应用中实施,其中,与参考位置相关的声场的空间音频处理是有用的,以将与第一参考位置相关的声场变换至与不同的第二参考位置相关的另一声场。

[0112] 声场计算器420然后生成经修改的声场描述、或者可选地生成(虚拟)扬声器表示、或者生成诸如双声道表示的双耳表示用于耳机再现。因此,声场计算器420可以生成经修改的声场描述作为经修改的声场,经修改的声场描述基本上与原始声场描述相同,但是现在关于新参考位置。在替代实施例中,可以为诸如5.1方案的预定扬声器设置或具有更多扬声器的扬声器设置(并且特别地,具有扬声器的三维布置而不是仅二维布置,即扬声器布置具有关于用户位置而被升高的扬声器)生成虚拟或实际扬声器表示。对虚拟现实应用特别有用的其他应用是用于双耳再现的应用,即用于可应用于虚拟现实用户头部的耳机。

[0113] 示例性地,随后描述的图6示出一种情况,其中DirAC合成器仅对诸如全向或压力分量的降混分量操作,而在关于图12b示出的另一替代实施例中,DirAC合成器对整个声场数据(即在图12b中的实施例中的具有有全向分量 $w$ 和三个定向分量 $x$ 、 $y$ 、 $z$ 的场描述的全分量表示)操作。

[0114] 图4c示出本发明的较佳实施例下的场景。该图示出第一参考位置/地点A、第二参考位置/地点B、和两个不同声源A和B、以及平移向量 $l$ 。

[0115] 声源A和B两者都被包括在与参考位置A相关的声场描述和与参考位置B相关的第二声场描述中。

[0116] 为了计算源A(例如)到第一参考位置或到第二参考位置的距离,与A和B相关的两个不同声场描述都经受源分离过程,并且然后,通过这些不同声音分离过程所获得的源的匹配被获得。例如,这将产生源A。在针对第一声场描述以及也针对第二声场描述的源分离算法中找到源A。当从与参考位置A相关的第一声场描述获得时,源A的到达方向信息将是角度 $\alpha$ 。另外,但现在从与另一参考位置B相关的第二声场描述所获得的相同源A的到达方向信息将是角度 $\beta$ 。

[0117] 现在,基于已知或可计算的距离 $D$ ,即例如可从声场描述的元数据获得或计算,并且基于两个角度 $\alpha$ 和 $\beta$ ,由源A限定的三角形,参考位置A和参考位置B是被完全限定的。因此,例如通过三角测量处理操作,可以计算从源A到参考位置A的距离或从源A到参考位置B的距离或源A的一般位置,即,从原点指向源A的实际位置的向量。位置或距离两者都代表关于距离或位置的信息。

[0118] 然后可以对每个匹配源执行相同的过程,即也可对源B执行相同的过程。

[0119] 因此,每个匹配源的距离/位置信息被计算,并且然后,可以以距离/位置是完全已知的或者例如由附加元数据所给出的情况处理每个匹配源。然而,仅需要第一声场描述和第二声场描述的几何信息,而不是每个各个源的任何距离/深度信息。

[0120] 图8示出用于执行与DirAC合成器不同的合成的另一实施方式。例如,当声场分析器为每个源信号生成分离的单声道信号 $S$ 和原始到达方向时,以及当依据平移信息计算新到达方向时,则例如图8的高保真度立体声响复制信号生成器430将被用于生成声源信号(即单声道信号 $S$ )的但是针对新到达方向(DoA)数据的声场描述,新到达方向(DoA)数据由

水平角 $\theta$ 或仰角 $\theta$ 和方位角 $\Phi$ 组成的。然后,由图4b的声场计算器420执行的过程将针对具有新到达方向的每个声源生成例如一阶高保真度立体声响复制声场表示,并且然后,可以使用缩放因子来为每个声源执行进一步修改,该缩放因子取决于声场到新参考位置的距离,并且然后,来自各个源的所有声场可以相互迭加以最终获得再一次例如以与某个新参考位置相关的高保真度立体声响复制表示的经修改的声场。

[0121] 当解释由图6的DirAC分析器422、422a、422b处理的每个时间/频率区间表示某个(带宽受限的)声源时,则高保真度立体声响复制信号生成器430可以被使用取代DirAC合成器425、425a、425b,以针对每个时间/频率区间,使用作为图8的“单声道信号S”的该时间/频率区间的降混信号或压力信号或全向分量,生成完整的高保真度立体声响复制表示。然后,针对W、X、Y、Z分量中的每一个,频率-时间转换器中的各个频率-时间转换将产生声场描述,该声场描述不同于图4c中所示的声场描述。

[0122] 场景是从麦克风的视点(PoV)记录的,其位置被用作参考坐标系的原点。场景需要从聆听者的PoV再现,该聆听者在6DoF中被追踪,参见图5。这里显示单个声源用于图示说明,关系适用于每个时间-频率区间。

[0123] 图5示出空间音频的6DoF再现。声源通过麦克风被记录,该声源在相对于麦克风位置和朝向(黑线和弧线)的距离 $d_r$ 处具有DoA $r_r$ 。它需要相对于具有DoA $r_1$ 和距离 $d_1$ (虚线)的移动聆听者而被再现。这需要考虑聆听者平移 $l$ 和旋转 $o$ (点线)。该DOA被表示为指向源的具有单位长度的向量。

[0124] 在坐标 $d_r \in \mathbb{R}^3$ 的声源是从单位向量 $r_r = d_r / \|d_r\|$ 表示的到达方向(DoA)记录的。可以从记录的分析来估计该DoA。它来自距离 $d_r = \|d_r\|$ 。假设该信息可以从每个源的元数据得出,或者通常从图3b的项352、353、354得出,并且可被表示为具有从记录位置到距离(例如以米为单位给出)的任何方向 $r$ 的距离描述,如通过使用关于两个不同参考位置的两个角度和距离/位置或参考位置的三角测量处理获得的。

[0125] 聆听者在6DoF中被追踪。在给定的时间,他相对于麦克风位于位置 $l \in \mathbb{R}^3$ ,并且相对于麦克风的坐标系具有旋转 $o \in \mathbb{R}^3$ 。记录位置被选为坐标系统的原点以简化符号。

[0126] 因此,需要以导致音量改变的不同距离 $d_1$ 、及作为平移和随后旋转的结果的不同DoA  $r_1$ 再现声音。

[0127] 如以下部分所解释,一种用于基于参数表示通过专用变换从聆听者角度获得虚拟信号的方法被概述。

[0128] 所提出的方法基于用于参数空间声音编码的基本DirAC方法,参见[16]。假设在分析频谱的每个时间-频率实例中存在一个主要直达源,并且这些可以被独立地处理。使用短时傅立叶变换(STFT)将记录变换为时间-频率表示。时间帧索引用 $n$ 表示,以及频率索引用 $k$ 表示。然后,分析变换后的记录,以针对复数频谱 $P(k, n)$ 的每个时间-频率区间估计方向 $r_r(k, n)$ 和扩散性 $\psi(k, n)$ 。在合成中,信号被划分成直达部分和扩散部分。此处,通过依据扬声器位置而平移直达部分并添加扩散部分,来计算扬声器信号。

[0129] 参见图6c,在6DoF中根据聆听者角度变换FOA信号的方法可以被分为五个步骤。

[0130] 图6c示出6DoF再现的方法。B格式的记录的FOA信号被DirAC编码器处理,该编码器针对复数频谱的每个时间-频率区间计算方向和扩散值。然后,通过聆听者的追踪位置并根据每个源的距离图中所给出的距离信息(例如通过三角测量计算得出),变换方向向量。然

后,根据头部旋转,旋转所得方向向量。最后,在DirAC解码器中,合成针对8+4个虚拟扬声器声道的信号。然后,它们被双耳化。

[0131] 在实施例中,在DirAC编码器422中分析输入信号,来自距离图 $m(r)$ 的距离信息被添加,该距离图 $m(r)$ 为每个(匹配)源给出距离,然后,聆听者的追踪平移和旋转被应用在新型变换423和424中。DirAC解码器425合成针对8+4个虚拟扬声器的信号,其转而被双耳化427以用于耳机回放。应注意的是,由于在平移之后的声音场景的旋转是独立操作,所以它可选地可被应用在双耳渲染器中。针对6DoF所变换的唯一参数是方向向量。通过模型定义,扩散部分被假设为等向性和同构型的,因此保持不变。

[0132] DirAC编码器的输入是以B格式表示的FOA声音信号。其由四个信道组成,即全向声压和三个一阶空间梯度,它们在某个假设下与粒子速度成比例。参见[18],该信号以参数方式被编码。参数从作为经变换的全向信号的复数声压 $P(k, n)$ 以及对应于经变换的梯度信号的复数粒子速度向量 $U(k, n) = [U_x(k, n), U_y(k, n), U_z(k, n)]^T$ 得出。

[0133] DirAC表示由在每个时间-频率区间处的声波的信号 $P(k, n)$ 、扩散性 $\psi(k, n)$ 和方向 $r(k, n)$ 组成。为了得到后者,首先,有效声音强度向量 $I_a(k, n)$ 被计算作为压力向量与速度向量的共轭复数(由 $(\cdot)^*$ 表示)的乘积的实部(由 $\text{Re}(\cdot)$ 表示)[18]:

$$[0134] \quad I_a(k, n) = \frac{1}{2} \text{Re}(P(k, n)U^*(k, n)). \quad (1)$$

[0135] 从该向量的变异系数,估计扩散性[18]。

$$[0136] \quad \psi(k, n) = \sqrt{1 - \frac{\|E\{I_a(k, n)\}\|}{E\{\|I_a(k, n)\|\}}} \quad (2)$$

[0137] 其中 $E$ 表示沿时间帧的期望算子,实现为移动平均。

[0138] 由于预计使用基于方向的距离图来操纵声音,该距离图具有每个(匹配)源到参考位置的距离,因此方向估计的方差(variance)在可选的实施例中应该较低。由于帧通常很短,情况并非总是如此。因此,移动平均被应用以获得平滑的方向估计 $\bar{I}_a(k, n)$ 。在一个实施例中,然后计算信号的直达部分的DoA作为在相反方向的单位长度向量:

$$[0139] \quad r_r(k, n) = - \frac{\bar{I}_a(k, n)}{\|\bar{I}_a(k, n)\|} \quad (3)$$

[0140] 由于针对每个时间-频率区间将方向编码为单位长度的三维向量,因此可以直接对距离信息进行积分。方向向量与其对应图项目相乘,使得向量长度表示对应声源的距离 $d_r(k, n)$ :

$$[0141] \quad d_r(k, n) = r_r(k, n)d_r(k, n) = r_r(k, n)m(r_r(k, n)) \quad (4)$$

[0142] 其中 $d_r(k, n)$ 是从麦克风的记录位置指向在时间 $n$ 和频率区间 $k$ 有效的声源的向量。

[0143] 聆听者位置通过当前处理帧的追踪系统给出为 $l(n)$ 。参考图6b,利用源位置的向量表示,可以减去追踪位置向量 $l(n)$ 以生成具有长度 $d_l(k, n) = \|d_l(k, n)\|$ 的新平移方向向量 $d_l(k, n)$ 。从聆听者的PoV到声源的距离被得出,并且DoA在单个步骤中被调整:

$$[0144] \quad d_l(k, n) = d_r(k, n) - l(n) \quad (5)$$

[0145] 真实再现的重要方面是距离衰减。衰减假设是声源和聆听者之间距离的函数[19]。方向向量的长度用于编码衰减或放大,以用于再现。到记录位置的距离根据距离图以

$d_r(k, n)$  编码, 并且待再现的距离以  $d_1(k, n)$  编码。如果将向量归一化为单位长度, 然后乘以旧及新距离的比率, 则看出所需长度通过将  $d_1(k, n)$  除以原始向量的长度而被给出:

$$[0146] \quad \mathbf{d}_v(k, n) = \frac{\mathbf{d}_1(k, n)}{\|\mathbf{d}_1(k, n)\|} \frac{\|\mathbf{d}_1(k, n)\|}{\|\mathbf{d}_r(k, n)\|} = \frac{\mathbf{d}_1(k, n)}{\|\mathbf{d}_r(k, n)\|} \quad (6)$$

[0147] 聆听者朝向的更改将被应用于以下步骤。相对于作为原点的记录位置, 通过追踪所给出的朝向可以被写为由俯仰、偏转和翻滚组成的向量  $\mathbf{o}(n) = [o_x(n), o_z(n), o_y(n)]^T$ 。根据聆听者朝向旋转源方向, 这是使用2D旋转矩阵实现的:

$$[0148] \quad \mathbf{d}_p(k, n) = R_y(o_y(n)) R_z(o_z(n)) R_x(o_x(n)) \mathbf{d}_v(k, n) \quad (7)$$

[0149] 聆听者的所得DoA然后由被归一化为单位长度的向量所给出:

$$[0150] \quad \mathbf{r}_p(k, n) = \frac{\mathbf{d}_p(k, n)}{\|\mathbf{d}_p(k, n)\|} \quad (8)$$

[0151] 经变换的方向向量、扩散性和复数频谱被用于合成针对均匀分布的8+4虚拟扬声器设置的信号。八个虚拟扬声器以45°方位角步阶位于聆听者平面(仰角0°)上, 以及四个以90°十字队形位于45°的仰角上。对于每个扬声器声道  $1 \leq i \leq I$  (其中  $I=12$  是扬声器的数量), 合成被分成直达和扩散部分[16]:

$$[0152] \quad Y_i(k, n) = Y_{i,S}(k, n) + Y_{i,D}(k, n) \quad (9)$$

[0153] 对于直达部分, 边缘衰落幅度平移(EFAP)在给定虚拟扬声器几何结构的情况下被应用于从正确方向再现声音[20]。给定DoA向量  $\mathbf{r}_p(k, n)$ , 这为每个虚拟扬声器声道  $i$  提供移动增益  $G_i(r)$ 。每个DoA的距离相依增益是从方向向量  $\mathbf{d}_p(k, n)$  的所得长度得出的。声道  $i$  的直达合成变为:

$$[0154] \quad Y_{i,S}(k, n) = \sqrt{1 - \psi(k, n)} P(k, n) G_i(\mathbf{r}_p(k, n)) (\|\mathbf{d}_p(k, n)\|)^{-\gamma} \quad (10)$$

[0155] 其中指数  $\gamma$  是一个调整因子, 其通常设置为约1[19]。应注意的是, 当  $\gamma = 0$  时, 距离相依增益被关闭。

[0156] 压力  $P(k, n)$  用于生成  $I$  去相关的信号  $\tilde{P}_i(k, n)$ 。这些去相关的信号作为扩散分量被添加到各个扬声器声道。这遵循标准方法[16]:

$$[0157] \quad Y_{i,D}(k, n) = \sqrt{\psi(k, n)} \frac{1}{\sqrt{I}} \tilde{P}_i(k, n) \quad (11)$$

[0158] 每个声道的扩散和直达部分被加在一起, 并且通过逆STFT将信号变换回时域。取决于扬声器位置, 这些声道时域信号针对左耳和右耳而与HRTF进行卷积运算, 以产生双耳信号。

[0159] 图6a示出用于使用空间信息和第一和第二声场描述以及平移信息计算经修改的声场的另一较佳实施例, 该平移信息指示参考位置到不同参考位置的平移, 例如, 如关于图4c或图5中向量1所讨论的。

[0160] 图6a示出块700, 其指示声音分离的应用, 或者通常, 对于与图4c的参考位置A相关的第一声场描述和与图4c的参考位置B相关的第二声场描述的声音分析过程。

[0161] 该过程将产生一个或多个提取对象的第一群组, 以及另外, 一个或多个提取对象的第二群组。

[0162] 在块702内使用这些群组用于为所有分离的源(即, 所提取的源的第一群组和一个

或多个所提取的源的第二群组)计算到达方向信息。

[0163] 在其他实施例中,步骤700和702在单个过程内实现,一方面提供源的信号,另一方面提供源的DoA信息。对于参数过程也是如此,如DirAC的时间/频率选择过程,其中源信号是在时间/频率区间中的B格式表示的信号、或时间/频率区间的压力信号或全向信号,以及作为该特定区间的DoA参数的DoA信息。

[0164] 然后,在步骤704中,在第一群组的源和第二群组的源之间执行源匹配,并且源匹配的结果是匹配源。

[0165] 这些匹配源被用于使用新DoA和新距离为每个匹配对象计算声场,如块710中所示。此外,匹配对象的到达方向信息(即,对每个对象为两个,如源A的图4c的 $\alpha$ 和 $\beta$ )在块706中被使用以便使用例如三角测量操作计算匹配对象的位置,或者可选地或另外地,计算匹配对象的距离。

[0166] 块706的结果是每个匹配对象的位置,或者可选地或另外地,匹配对象到第一或第二参考位置A、B中的一个的距离,例如在图4c中所示的。

[0167] 另外,较佳地,不仅使用块708中的平移信息而且还使用旋转信息,来计算匹配对象的新到达方向信息和新距离。

[0168] 虽然已经概述了匹配对象的位置被输入到步骤708,但是要强调的是,其用于仅计算匹配对象的新到达方向信息、匹配对象的实际位置,或者换句话说,匹配对象的距离对于计算关于聆听者已移动到的新(不同)参考位置的新到达方向不是必需的,例如距离不是必需的。

[0169] 然而,为了使源信号适应新情况,距离是必要的。因此,当源或声音对象到新参考位置的距离变得更短时,则将算出小于1的缩放因子。然而,当距离变得更大时,则将算出高于1的比例因子,例如,如关于图6b所讨论的。因此,尽管在图6a中针对实施例示出,但不一定是计算匹配对象的明确位置、并且然后计算匹配对象的距离、并且然后针对每个匹配对象使用新到达方向和新距离计算声场的情况。相反,仅匹配对象到两个参考位置中的一个参考位置的距离通常是足够的,然后,使用新DoA和新距离计算每个匹配对象的声场。

[0170] 另外,块714示出使用通过块702获得的旧DoA信息,计算非匹配对象的声场。此外,在块712中组合在块710中获得的匹配对象的声场以及通过块714获得的非匹配对象的声场,以便获得经修改的声场描述,该经修改的声场描述例如可以是诸如一阶高保真度立体声响复制描述、更高阶的高保真度立体声响复制描述的高保真度立体声响复制描述,或者,可选地是与某个扬声器设置相关的扬声器声道描述,当然,扬声器设置对于块710和块714是相同的,使得可以在块712中执行简单的逐声道添加。

[0171] 图6b示出声场计算器420的较佳实施方式。在块1102中,对于每个源的源分离和到达方向或一般方向信息的计算被执行。然后,在块1104中,到达方向向量被乘以距离信息向量,即,从原始参考位置到声源的向量,即例如,从图5的项520到项510的向量。然后,在块1106中,平移信息,即,从图5的项520到项500的向量被考虑,以便计算新平移方向向量,该新平移方向向量是从聆听者位置500到声源位置510的向量。然后,在块1108中计算由 $dv$ 所指示的具有正确长度的新到达方向向量。该向量指向与 $dr$ 相同的方向,但具有不同长度,因为向量的长度反映了声源510被记录在具有某个声量的原始声场中的事实,以及因此, $dv$ 的长度或多或少指示响度变化。这是通过将向量 $d_l$ 除以记录距离 $dr$ 来获得的,即从麦克风520

到声源510的向量 $dr$ 的长度。如上所述,通过三角测量计算,可以得出从麦克风520到声源510的向量 $dr$ 的长度。当麦克风处于第一声场描述的参考位置时,则使用从第一声场描述的参考位置到声源的距离。然而,当麦克风处于第二声场描述的另一参考位置时,则使用从第二声场描述的另一参考位置到声源的距离。

[0172] 当如图5所示,再现距离大于记录距离时,则 $dv$ 的长度将低于单位长度(unity)。这将导致,用于在新聆听者位置处的再现的声源510的衰减。然而,当再现距离 $d_1$ 小于记录距离时,则通过块1108计算的 $dv$ 的长度将大于1,并且对应缩放因子将导致声源的放大。

[0173] 在图6a中,项710指示使用新到达方向信息和新距离来计算每个匹配对象的声场。然而,基本上,为了计算每个匹配对象的声场,通常可以使用从一个或多个所提取的源的第一群组或一个或多个所提取的源的第二群组获得的对象信号。然而,在实施例,图7中所示的特定选择被执行,以便确定在块710中使用哪个声场描述来执行声场计算。在块720中,确定从新聆听者位置到第一声场描述的第一参考位置的第一距离。关于图4c,这是在差异参考位置和参考位置A之间的距离。

[0174] 此外,在步骤722中,确定从新聆听者位置到第二声场描述的第二参考位置的第二距离。在图4c的实施例中,这是在不同参考位置(由于平移)和参考位置B之间的距离。

[0175] 看起来,从不同参考位置到参考位置B的距离低于从不同参考位置到参考位置A的差。因此,这将在步骤724中被确定。并且,在步骤726中,从具有较小距离的声场描述所得出的群组中选择对象信号。因此,为了渲染对应于图4c中的匹配源的源A和B,将使用从与另一参考位置B相关的第二声场描述所得出的声源信号。

[0176] 然而,在其他实施例中,其中平移从原点指向不同参考位置(例如,指向图4c图示说明中的左边),较小距离将从该另外的参考位置到参考位置A,然后,第一声场描述将被使用,以最终在图6b的块710中计算每个匹配对象的声场。再次,选择将通过图7中所示的过程来执行。

[0177] 图9示出另一较佳实施例。在步骤740中,执行关于第一声场描述的声场分析,例如,以图6c的块422中所示的DirAC分析的形式参数声场分析。

[0178] 这例如对于每个时间/频率区间产生第一组参数,其中每组参数包括DoA参数,并且可选地包括扩散参数。

[0179] 在步骤741中,对第二声场描述执行声场分析,并且再次,如在块740中那样以及(例如)如关于图6c的块422所讨论的那样执行DirAC分析。

[0180] 这例如针对时间/频率区间产生第二组参数。

[0181] 然后,在块746中,使用来自第一时间/频率区间的对应DoA参数和来自第二组参数的相同时间/频率区间的DoA参数,可以确定每个参数对的位置。这将产生每个参数对的位置。然而,对应时间/频率区间的扩散性在第一组参数和/或第二组参数中越低,位置将越有用。

[0182] 因此,较佳地,仅进一步使用来自在第一组参数和第二组参数两者中都产生相当低的扩散性的时间/频率区间的位置。

[0183] 另外,较佳地,还对如也由块740和块741输出的时间/频率区间中的对应信号执行相关性(correlation)。

[0184] 因此,例如,图6a中的步骤704的“源匹配”可以被完全避免并且可以被基于扩散参

数的匹配源/匹配时间/频率区间的确定代替,或者此外,可以使用来自B格式分量例如或者来自通过图6c的块422输出的压力信号或对象信号的时间/频率区间中的对应信号,执行该匹配。

[0185] 在任何情况下,块46将产生某些(选定)时间/频率区间的某些位置,其对应于在图6a的块704中所找到的“匹配对象”。

[0186] 然后,在块748中,针对由块746所获得的位置和/或例如由帽子跟踪器所获得的对应平移/旋转的经修改的参数和/或信号被计算,并且块748的输出表示针对不同时间/频率区间的经修改的参数和/或经修改的信号。

[0187] 因此,块748可以对应于块424的平移变换423和旋转变换,用于计算经修改的参数,并且,例如,经修改的信号的计算将通过图6c的块425执行,该执行较佳地在考虑从对应时间/频率区间的位置所得出的某个比例因子的情况下发生。

[0188] 最后,在块750中使用经修改的数据,执行声场描述的合成。例如,这可以通过使用第一或第二声场描述的DirAC合成来完成,或者可以通过如在块425所示的高保真度立体声响复制信号生成器来执行,并且结果将是用于传输/存储/渲染的新声场描述。

[0189] 图10示出声场计算器420的另一较佳实施方式。图10所示过程的至少部分分别对每个匹配源执行。块1120确定匹配源的距离,例如通过三角测量计算来确定。

[0190] 基于声场描述,在1100中确定全频带到达方向或每频带到达方向。这些到达方向信息表示声场的到达方向数据。基于到达方向数据,在块1110中执行平移变换。至此,块1120针对每个匹配源计算距离。基于数据,块1110生成声场的新到达方向数据,在此实施方式中,其仅取决于从参考位置到不同参考位置的平移。至此,块1110接收例如通过在虚拟现实实现的情境中的追踪而生成的平移信息。

[0191] 较佳地或替代地,也使用旋转数据。至此,块1130使用旋转信息执行旋转变换。当平移和旋转两者被执行时,则较佳地,在计算已经包括来自平移的信息以及来自块1120的源距离的声场的新DoA之后,执行旋转变换。

[0192] 然后,在块1140中,新声场描述被生成。至此,可以使用原始声场描述,或者,可选地,可以使用通过源分离算法从声场描述分离的源信号,或者可以使用任何其他应用。基本上,新声场描述可以是例如由高保真度立体声响复制生成器430获得的或由DirAC合成器425生成的定向声场描述,或者可以是在随后的双耳渲染中从虚拟扬声器表示所生成的双耳表示。

[0193] 较佳地,如图10所示,每个到达方向的距离也被用于生成新声场描述,以使某个声源的音量或响度适应新位置,即新或不同参考位置。

[0194] 尽管图10示出在平移变换之后执行旋转变换的情况,但是应注意,该顺序可以是不同的。特别地,旋转变换可以被应用于通过块1100所生成的声场的DoA,并且然后,应用由于主体从参考位置到不同参考位置的平移引起的附加平移变换。

[0195] 一旦已经通过块1100确定声场的DoA,就使用块1120从元数据中获取距离信息,该距离信息然后通过块1140中生成新声场描述而被使用,用来计算某个源关于某个参考位置的改变的距离以及因此的改变的响度。基本上,可以说在距离变大的情况下,则特定声源信号被衰减,而当距离变短时,则声源信号被放大。自然地,取决于距离的某个声源的衰减或放大与该距离变化成比例地进行,但是,在其他实施例中,可以以相当粗略的增量将较不

复杂的操作应用于声源信号的这种放大或衰减。与任何距离变化完全被忽略的情况相比，即使是这种较不复杂的实现也能提供更好的结果。

[0196] 图11示出声场计算器的另一较佳实施例。在块1200中，例如，对于每频带或全频带，确定来自声场的各个源。当每帧和频带的确定被执行时，则这可以通过DirAC分析来完成。如果全频带或子频带确定被执行，则可以通过任何类型的全频带或子频带源分离算法来完成。

[0197] 在块1210中，例如通过头部追踪，确定聆听者的平移和/或旋转。

[0198] 在块1220中，通过使用元数据并且例如通过使用三角测量计算的元数据，确定每个源的旧距离。因此，每个频带被认为是某个源（假设扩散性低于某个阈值），并且然后，确定具有低扩散值的每个时间/频率区间的某个距离。

[0199] 然后，在块1230中，每个源的新距离例如通过每频带的向量计算被获得，例如，在图6b的上下文中所讨论的。

[0200] 此外，如块1240中所示，例如，通过在DirAC分析中获得的DoA计算或例如通过源分离算法中的到达方向或方向信息分析，确定每个源的旧方向。

[0201] 然后，在块1250中，例如通过执行每频带或全频带的向量计算，确定每个源的新方向。

[0202] 然后，在块1260中，为平移和旋转的聆听者生成新声场。例如，这可以通过在DirAC合成中按比例缩放每个声道的直达部分来完成。取决于具体实施方式，除了或替代在块1260中执行距离修改，可在块1270a、1270b或1270c中完成距离修改。

[0203] 例如，当确定声场仅具有单个源时，则距离修改可以已经在块1270a中被执行。

[0204] 可选地，当通过块1200计算各个源信号时，则在实际新声场在块1260中被生成之前，在块1270b中，可以针对各个源，执行距离修改。

[0205] 另外，例如，当块1260中的声场生成不渲染扬声器设置信号或双耳信号，而是另一声场描述时，例如，使用高保真度立体声响复制编码器或计算器430，则距离修改也可以在块1260中生成之后执行，这意指在块1270c中执行。依据该实施方式，距离修改也可以被分配给多个修改器，以便最终某个声源处于某个响度，该某个响度由原始距离与新距离之间的差异所指示，该原始距离为声源和参考位置之间的距离、该新距离为声源和不同参考位置之间的距离。

[0206] 图12a示出最初公开（例如）在2009年IWPASH的早先引用的参考文献“Directional Audio Coding”中的DirAC分析器。

[0207] DirAC分析器包括一组带通滤波器1310、能量分析器1320、强度分析器1330、时间平均块1340和扩散性计算器1350以及方向计算器1360。

[0208] 在DirAC中，分析和合成两者都在频域中执行。有几种方法用于将声音划分到频带，每种方法都在不同的属性内。最常用的频率变换包括短时傅立叶变换（STFT）和正交镜像滤波器组（QMF）。除了这些之外，还可以完全自由地设计具有任意滤波器的滤波器组，该滤波器组可针对任何特定目的进行优化。定向分析的目标是在每个频带处估计声音的到达方向，以及估计声音是否同时从一个或多个方向到达。原则上，这可以用许多技术来执行，然而，声场的能量分析已经被发现是合适的，这在图12a中示出。当从单个位置捕获一维、二维或三维的压力信号和速度信号时，能量分析可以被执行。在一阶B格式信号中，全向信号

被称为W信号,其已经被2的平方根缩小。声压可以被估计为在STFT域中表示的 $P = \sqrt{2} * W$ 。

[0209] X、Y和Z声道具有沿笛卡尔轴定向的偶极子的定向模式,其一起形成向量 $U = [X, Y, Z]$ 。向量估计声场速度向量,并且也在STFT域中被表示。声场的能量E被计算。可以利用定向麦克风的重合定位或者利用紧密间隔的全向麦克风组,获得B格式信号的捕获。在一些应用中,麦克风信号可以在计算域中形成,即模拟。

[0210] 声音的方向被定义为强度向量I的相反方向。在发送的元数据中,该方向被表示为对应角度方位角和仰角值。还使用强度向量和能量的期望算子来计算声场的扩散性。该公式的结果是在0和1之间的实数值,其特征在于声音能量是从单个方向到达(扩散性是零)还是从所有方向到达(扩散性是1)。在全3D或较小维度的速度信息可用的情况下,该过程是适当的。

[0211] 图12b示出DirAC合成,再一次,其具有一组带通滤波器1370、虚拟麦克风块1400、直达/扩散合成器块1450、以及某个扬声器设置或虚拟预期扬声器设置1460。另外,使用扩散性-增益变换器1380、基于向量的幅度平移(VBAP)增益表块1390、麦克风补偿块1420、扬声器增益平均块1430和用于其他声道的分配器1440。

[0212] 在利用扬声器的这种DirAC合成中,图12b中所示的高质量版本的DirAC合成接收所有B格式信号,为此针对扬声器设置1460的每个扬声器方向,计算虚拟麦克风信号。所使用的定向模式通常是偶极子。然后,依据元数据,虚拟麦克风信号以非线性方式被修改。DirAC的低比特率版本未在图12b中显示,然而,在这种情况下,如图6所示,仅一个音频声道被发送。处理的不同之处在于所有虚拟麦克风信号都将被所接收音频的单个声道所取代。虚拟麦克风信号被分成两个串流:扩散串流和非扩散串流,它们被分开处理。

[0213] 通过使用基于向量的幅度平移(VBAP),非扩散声音将被再现为点源。在平移中,在与扬声器特定的增益因子相乘之后,单声道声音信号将被应用于扬声器的子集。使用扬声器设置的信息和指定平移方向,计算增益因子。在低比特率版本中,输入信号被简单地朝元数据所隐含的方向平移。在高质量版本中,每个虚拟麦克风信号与对应增益因子相乘,这会产生与平移相同的效果,然而它不太容易出现任何非线性伪声。

[0214] 在许多情况下,定向元数据经受突然的时间变化的影响。为了避免伪声,利用VBAP计算的扬声器的增益因子通过利用频率相依时间常数的时间积分而被平滑,该时间常数等于每个频带处的约50个循环周期。这有效地消除了伪声,然而,在大多数情况下,方向的变化不会被感知为比未进行平均的情况更慢。

[0215] 扩散声音的合成的目的是创造围绕聆听者的声音的感知。在低比特率版本中,通过去相关输入信号并将其从每个扬声器再现,扩散串流被再现。在高质量版本中,扩散串流的虚拟麦克风信号在某种程度上已经不连贯,并且它们需要仅稍微去相关。与低比特率版本相比,这种方法为环绕声混响和环境声音提供了更好的空间质量。

[0216] 对于利用耳机的DirAC合成,针对非扩散串流采用围绕聆听者的某个数量的虚拟扬声器以及针对扩散串流采用某个数量的扬声器规划DirAC。虚拟扬声器被实现为输入信号与测量的头部相关转移函数(HRTF)的卷积。

[0217] 尽管已经在装置的上下文中描述了一些方面,但是显然这些方面也表示相应方法的描述,其中块或设备对应于方法步骤或方法步骤的特征。类似地,在方法步骤的上下文中描述的方面也表示相应装置的相应块或项目或特征的描述。

[0218] 本发明的经增强的声场描述可以存储在数字存储介质或非暂时性存储介质上,或者可以在诸如无线传输介质、或诸如因特网的有线传输介质的传输介质上传输。

[0219] 根据某些实施要求,本发明的实施例可以用硬件或软件实现。该实施可以使用其上存储有电子可读取控制信号的数字存储介质来执行,例如软性磁盘、DVD、CD、ROM、PROM、EPROM、EEPROM或FLASH内存,电子可读取控制信号与(或者能够与)可编程计算机系统协作,以便执行相应的方法。

[0220] 根据本发明的一些实施例包括具有电子可读取控制信号的非暂时性数据载体,电子可读取控制信号能够与可编程计算机系统协作,从而执行本文所述的方法之一。

[0221] 通常,本发明的实施例可被实现为具有程序代码的计算机程序产品,该程序代码可操作用于在计算机程序产品在计算机上运行时执行这些方法之一。该程序代码可以例如存储在机器可读载体上。

[0222] 其他实施例包括用于执行存储在机器可读载体上的本文所述方法之一的计算机程序。

[0223] 换句话说,本发明方法的实施例因此是具有程序代码的计算机程序,当该计算机程序在计算机上运行时,该程序代码用于执行本文所述的方法之一。

[0224] 因此,本发明方法的另一实施例是数据载体(或数字存储介质、或计算机可读介质),数据载体包括记录在其上的用于执行本文所述方法之一的计算机程序。

[0225] 因此,本发明方法的另一实施例是表示用于执行本文所述方法之一的计算机程序的数据串流或信号序列。数据串流或信号序列可以例如被配置为经由数据通信连接传输,例如经由因特网传输。

[0226] 另一实施例包括处理装置,例如计算机或可编程逻辑设备,其被配置为或适于执行本文所述的方法之一。

[0227] 另一实施例包括计算机,其上安装有用于执行本文所述方法之一的计算机程序。

[0228] 在一些实施例中,可编辑逻辑设备(例如现场可编程门阵列)可用于执行本文所述方法的一些或全部功能。在一些实施例中,现场可编程门阵列可以与微处理器协作,以便执行本文描述的方法之一。通常,方法较佳地由任何硬件装置执行。

[0229] 上述实施例仅代表本发明的原理的说明。应理解,本文所述的布置和细节的任何修改和变化对于本领域其他技术人员是明显的。因此,意欲仅由随附专利权利要求的范围而非由通过示例和本文的实施例的解释所表示地详细细节限制。

[0230] 参考文献

[0231] [1]Liitola,T.,Head phone sound externalization,Ph.D.thesis,Helsinki University of Technology.Department of Electrical and Communications Engineering Laboratory of Acoustics and Audio Signal Processing.,2006年.

[0232] [2]Blauert,J.,Spatial Hearing-Revised Edition:The Psychophysics of Human Sound Localization,The MIT Press,1996年,ISBN0262024136.

[0233] [3]Zhang,W.,Samarasinghe,P.N.,Chen,H.,and Abhayapala,T.D.,“Surround by Sound:ARe-view of Spatial Audio Recording and Reproduction,”Applied Sciences,7(5),页532,2017年.

[0234] [4]Bates,E.and Boland,F.,“Spatial Music,Virtual Reality,and

360Media,” in Audio Eng.Soc.Int.Conf.on Audio for Virtual and Augmented Reality, Los Angeles, CA, U.S.A., 2016年.

[0235] [5] Anderson, R., Gallup, D., Barron, J.T., Kontkanen, J., Snavely, N., Esteban, C.H., Agarwal, S., and Seitz, S.M., “Jump: Virtual Reality Video,” ACM Transactions on Graphics, 35(6), 页198, 2016年.

[0236] [6] Merimaa, J., Analysis, Synthesis, and Perception of Spatial Sound: Binaural Localization Modeling and Multichannel Loudspeaker Reproduction, Ph.D. thesis, Helsinki University of Technology, 2006年.

[0237] [7] Kronlachner, M. and Zotter, F., “Spatial Transformations for the Enhancement of Ambisonics Recordings,” in 2<sup>nd</sup> International Conference on Spatial Audio, 埃朗根, 德国, 2014年.

[0238] [8] Tsingos, N., Gallo, E., and Drettakis, G., “Perceptual Audio Rendering of Complex Virtual Environments,” ACM Transactions on Graphics, 23(3), 页249-258, 2004年.

[0239] [9] Taylor, M., Chandak, A., Mo, Q., Lauterbach, C., Schissler, C., and Manocha, D., “Guided multi-view ray tracing for fast auralization,” IEEE Trans. Visualization & Comp. Graphics, 18, 页1797-1810, 2012年.

[0240] [10] Rungta, A., Schissler, C., Rewkowski, N., Mehra, R., and Manocha, D., “Diffraction Kernels for Interactive Sound Propagation in Dynamic Environments,” IEEE Trans. Visualization & Comp. Graphics, 24(4), 页1613-1622, 2018年.

[0241] [11] Thiergart, O., Kowalczyk, K., and Habets, E.A.P., “An Acoustical Zoom based on Informed Spatial Filtering,” in Int. Workshop on Acoustic Signal Enhancement, 页109-113, 2014年.

[0242] [12] Khaddour, H., Schimmel, J., and Rund, F., “A Novel Combined System of Direction Estimation and Sound Zooming of Multiple Speakers,” Radioengineering, 24(2), 2015年. [13] Ziegler, M., Keinert, J., Holzer, N., Wolf, T., Jaschke, T., op het Veld, R., Zakeri, F.S., and Foessel, S., “Immersive Virtual Reality for Live-Action Video using Camera Arrays,” in IBC, 阿姆斯特丹, 荷兰, 2017年.

[0243] [14] Thiergart, O., Galdo, G.D., Taseska, M., and Habets, E.A.P., “Geometry-Based Spatial Sound Acquisition using Distributed Microphone Arrays,” IEEE Trans. Audio, Speech, Language Process., 21(12), 页2583-2594, 2013年.

[0244] [15] Kowalczyk, K., Thiergart, O., Taseska, M., Del Galdo, G., Pulkki, V., and Habets, E.A.P., “Parametric Spatial Sound Processing: A Flexible and Efficient Solution to Sound Scene Acquisition, Modification, and Reproduction,” IEEE Signal Process. Mag., 32(2), 页31-42, 2015年.

[0245] [16] Pulkki, V., “Spatial Sound Reproduction with Directional Audio Coding,” J. Audio Eng. Soc., 55(6), 页503-516, 2007年.

- [0246] [17]International Telecommunication Union,“ITU-RBS.1534-3,Method for the subjective assessment of intermediate quality level of audio systems,”2015年.
- [0247] [18]Thiergart,O.,Del Galdo,G.,Kuech,F.,and Prus,M.,“Three-Dimensional Sound Field Analysis with Directional Audio Coding Based on Signal Adaptive Parameter Estimators,”in Audio Eng.Soc.Conv.Spatial Audio:Sense the Sound of Space,2010年.
- [0248] [19]Kuttruff,H.,Room Acoustics,Taylor&Francis,第4版,2000年.
- [0249] [20]Borß,C.,“A polygon-based panning method for 3D loudspeaker setups,”in Audio Eng.Soc.Conv.,页343-352,洛杉矶,加州,美国,2014年.
- [0250] [21]Rummukainen,O.,Schlecht,S.,Plinge,A.,and Habets,E.A.P.,“Evaluating Binaural Reproduction Systems from Behavioral Patterns in a Virtual Reality-A Case Study with Impaired Binaural Cues and Tracking Latency,”in Audio Eng.Soc.Conv.143,纽约,纽约州,美国,2017年.
- [0251] [22]Engelke,U.,Darcy,D.P.,Mulliken,G.H.,Bosse,S.,Martini,M.G.,Arndt,S.,Antons,J.-N.,Chan,K.Y.,Ramzan,N.,andBrunnström,K.,“Psychophysiology-Based QoE Assessment:A Survey,”IEEE Selected Topics in Signal Processing,11(1),页6-21,2017年.
- [0252] [23]Schlecht,S.J.and Habets,E.A.P.,“Sign-Agnostic Matrix Design for Spatial Artificial Reverberation with Feedback Delay Networks,”in Proc.Audio Eng.Soc.Conf.,页1-10-accepted,东京,日本,2018年.
- [0253] [31]M.A.Gerzon,“Periphony:With-height sound reproduction,”J.Acoust.Soc.Am.,卷21,110.1,页2-10,1973年.
- [0254] [32]V.Pulkki,“Directional audio coding in spatial sound reproduction and stereo upmixing,”in Proc.of the 28th AES International Conference,2006年.
- [0255] [33]--,“Spatial sound reproduction with directional audio coding,”Journal Audio Eng.Soc.,卷55,no.6,页503-516,2007年6月.
- [0256] [34]C.G.and G.M.,“Coincident microphone simulation covering three dimensional space and yielding various directional outputs,”美国专利4 042 779,1977年.
- [0257] [35]C.Faller and F.Baumgarte,“Binaural cue coding-part ii:Schemes and applications,”IEEE Trans.Speech Audio Process.,卷11,编号6,2003年11月.
- [0258] [36]C.Faller,“Parametric multichannel audio coding:Synthesis of coherence cues,”IEEE Trans.Speech Audio Process.,卷14,编号1,2006年1月.
- [0259] [37]H.P.J.E.E.Schuijers,J.Breebaart,“Low complexity parametric stereo coding,”in Proc.of the 116th AES Convention,柏林,德国,2004年.
- [0260] [38]V.Pulkki,“Virtual sound source positioning using vector base amplitude panning,”J.Acoust.Soc.A m.,卷45,编号6,页456-466,1997年6月.
- [0261] [39]J.G.Tylka and E.Y.Choueiri,“Comparison of techniques for binaural

navigation of higher-order ambisonics sound fields,"in Proc.of the AES International Conference on Audio for Virtual and Augmented Reality,纽约,2016年9月.

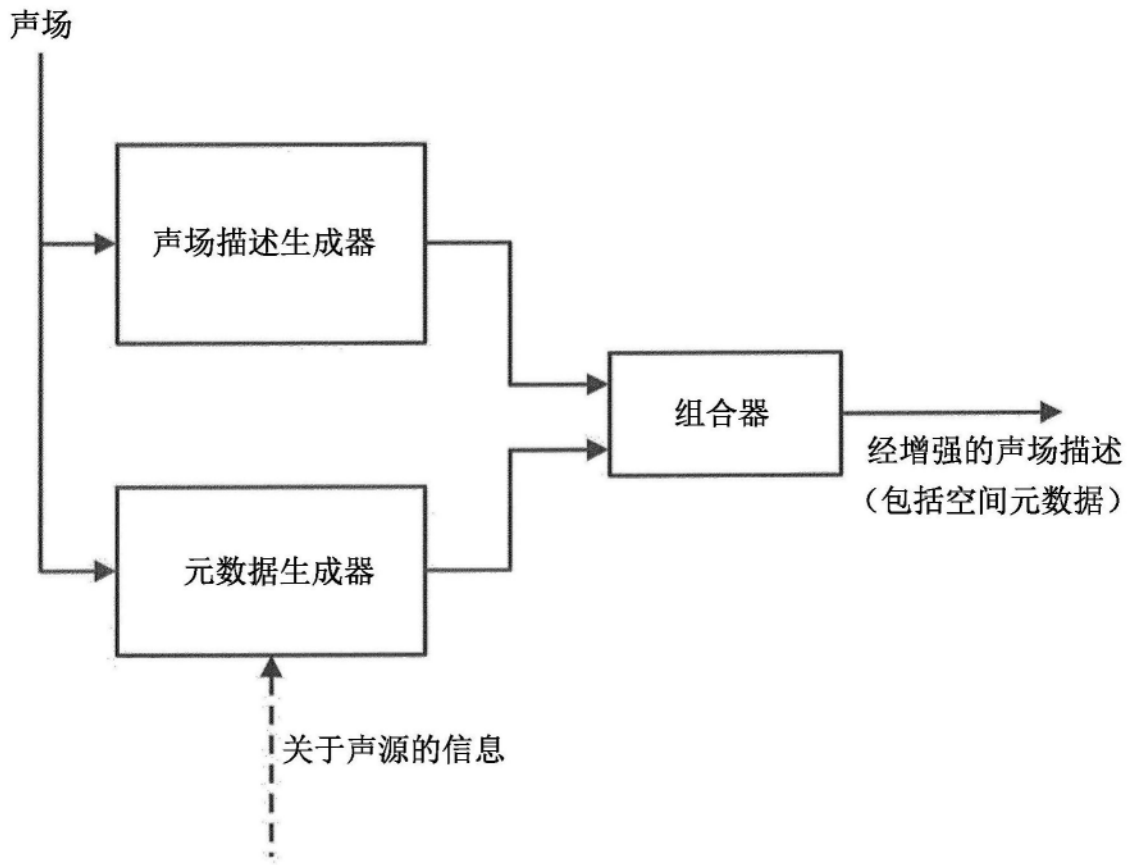


图1a

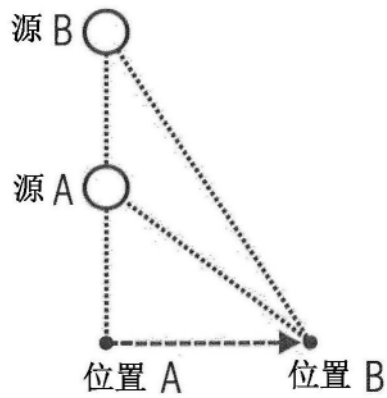


图1b

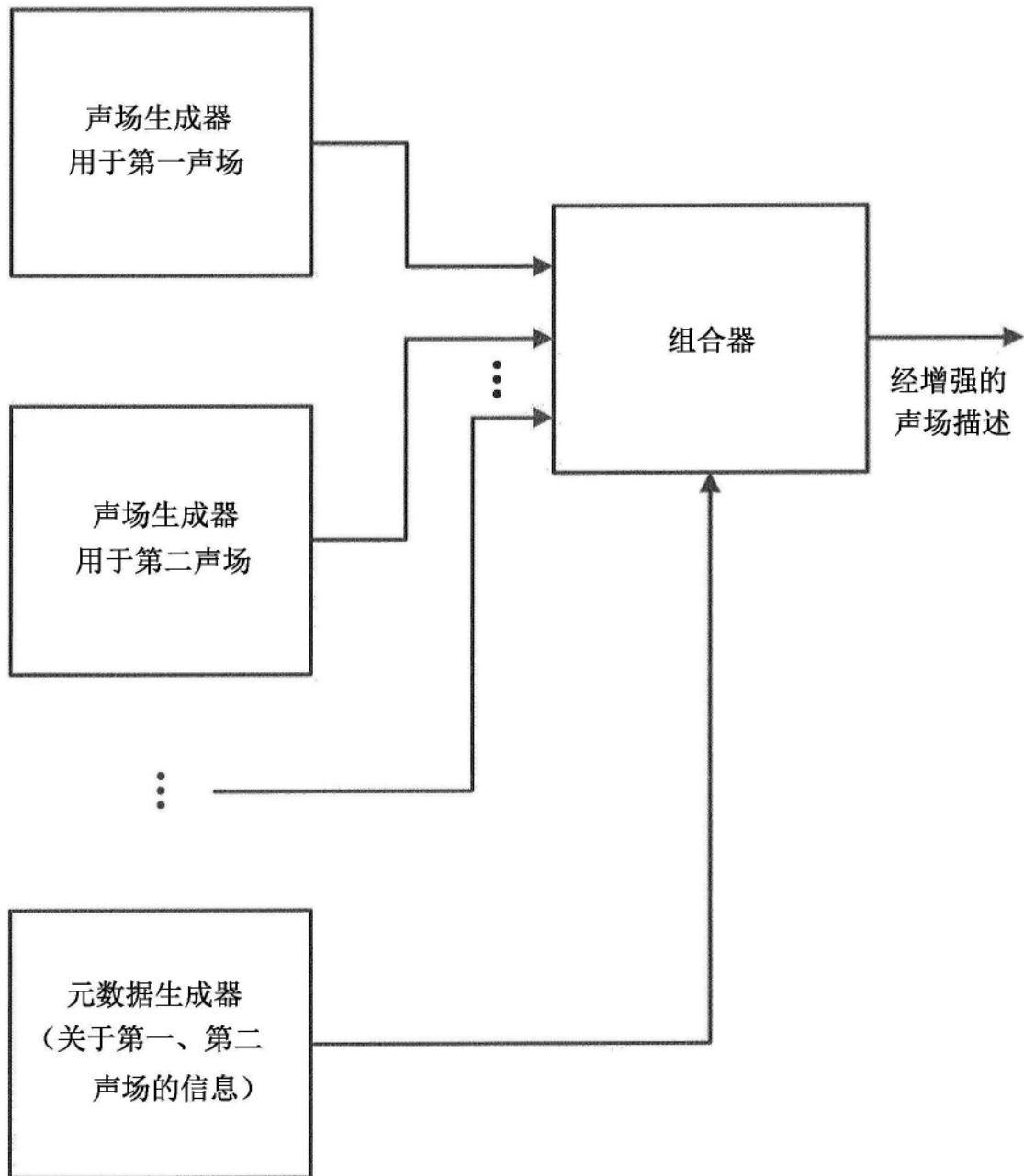


图2

第一声场描述	第二声场描述	关于第一和 第二声场的信息
--------	--------	------------------

图3a

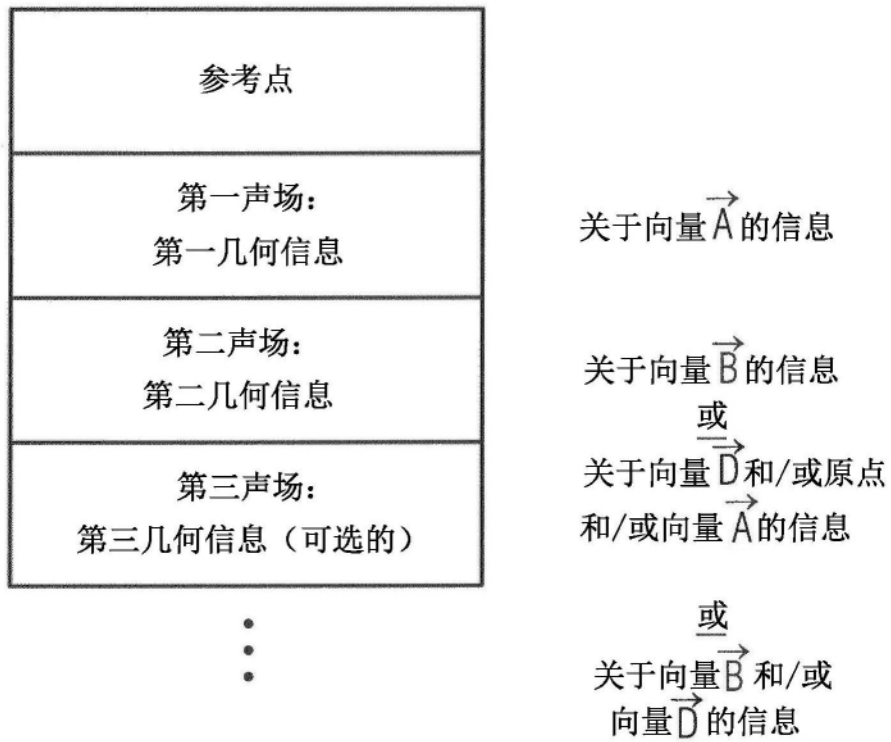


图3b

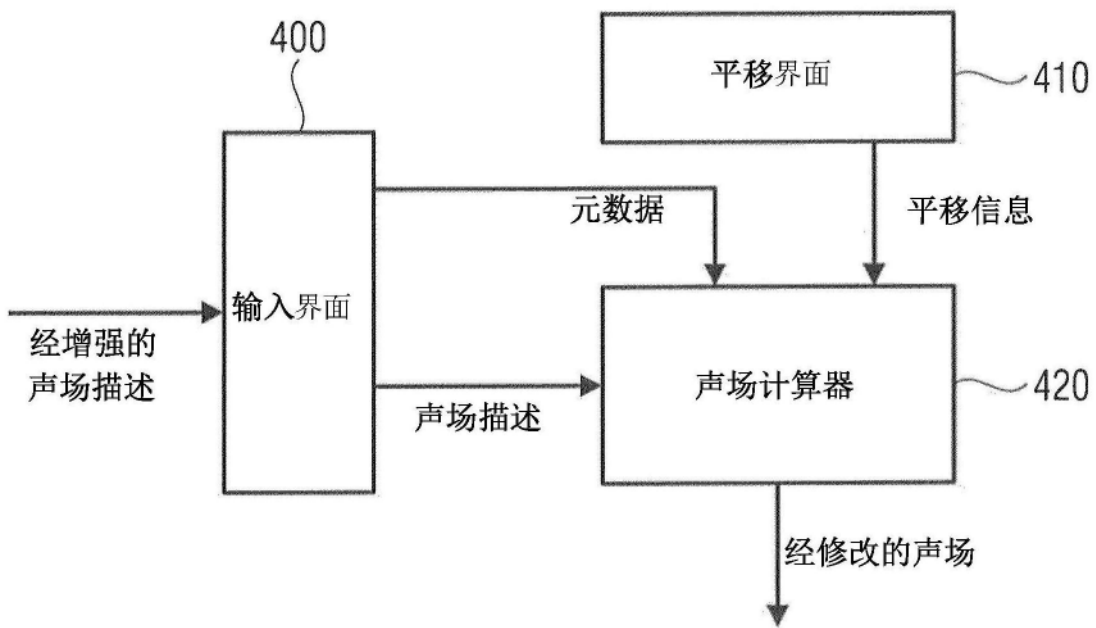


图4a

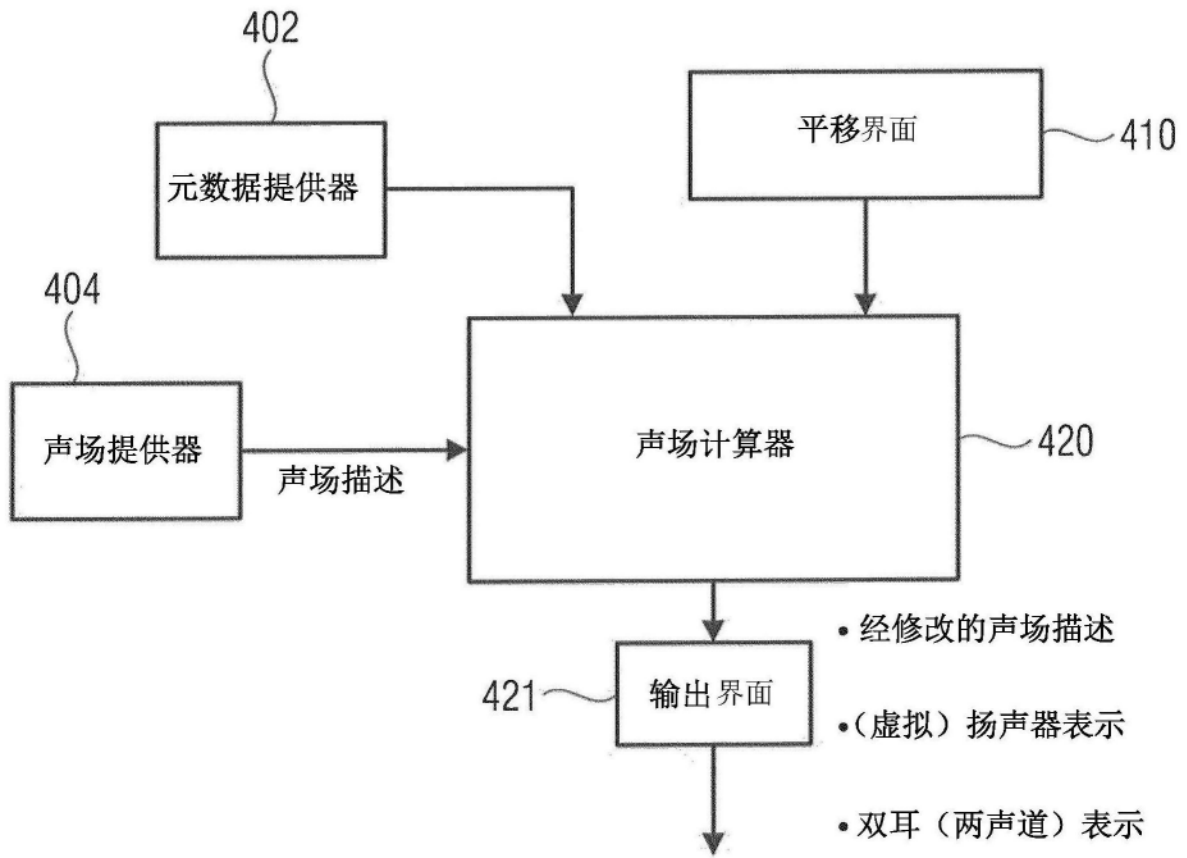


图4b

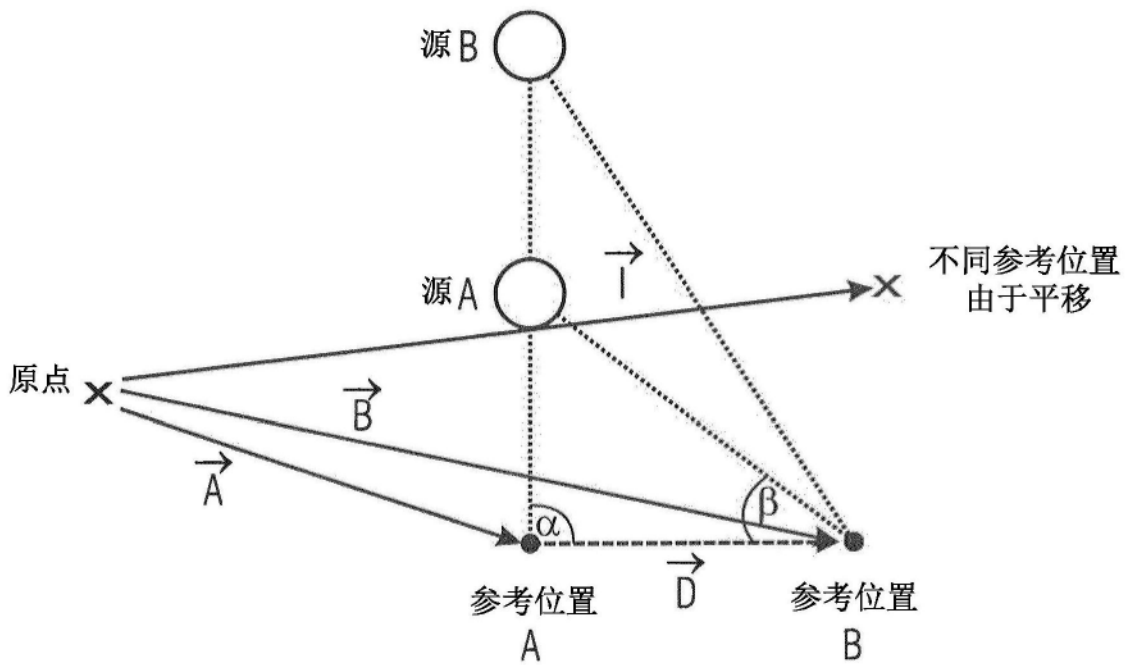


图4c

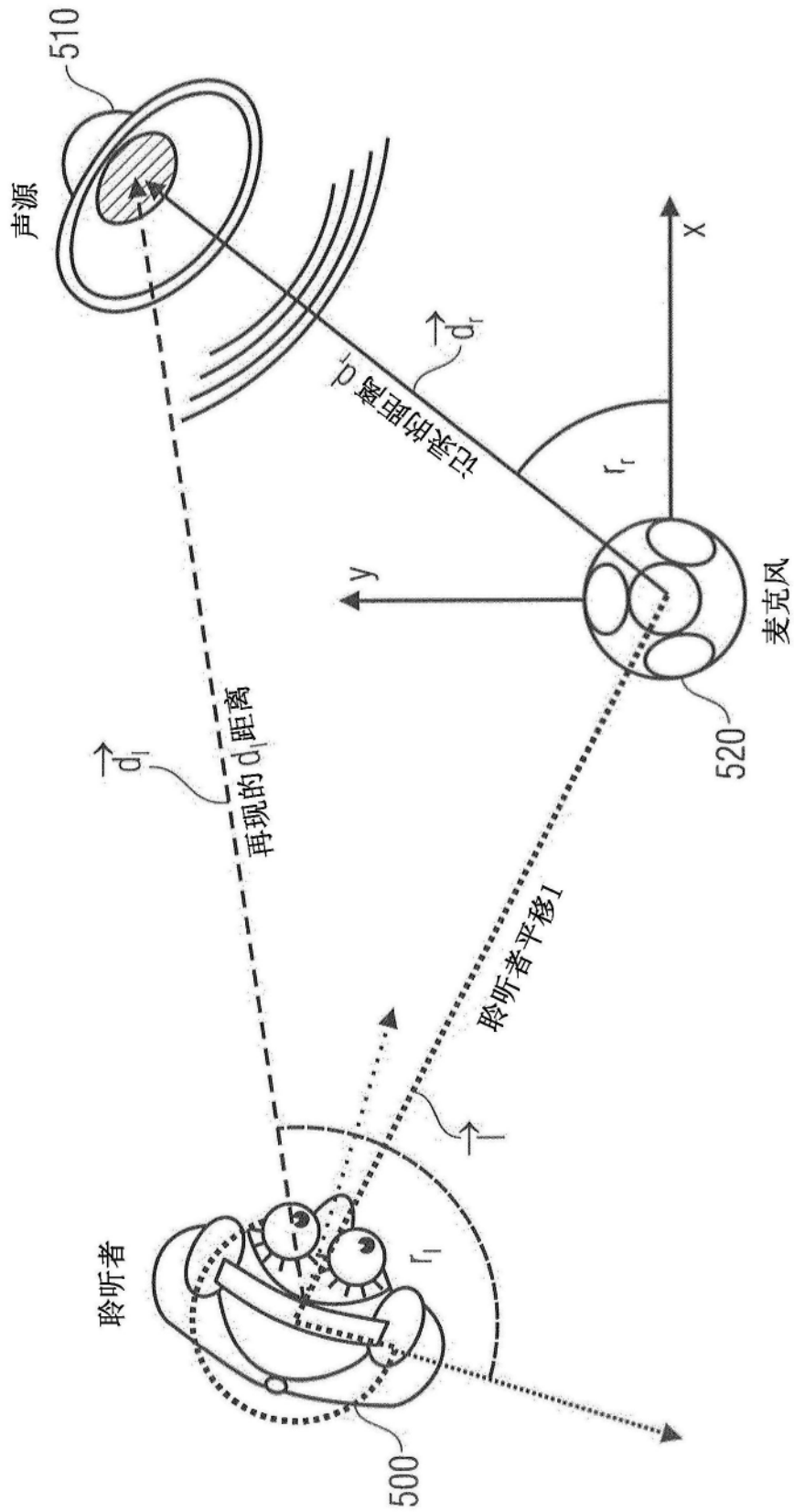


图5

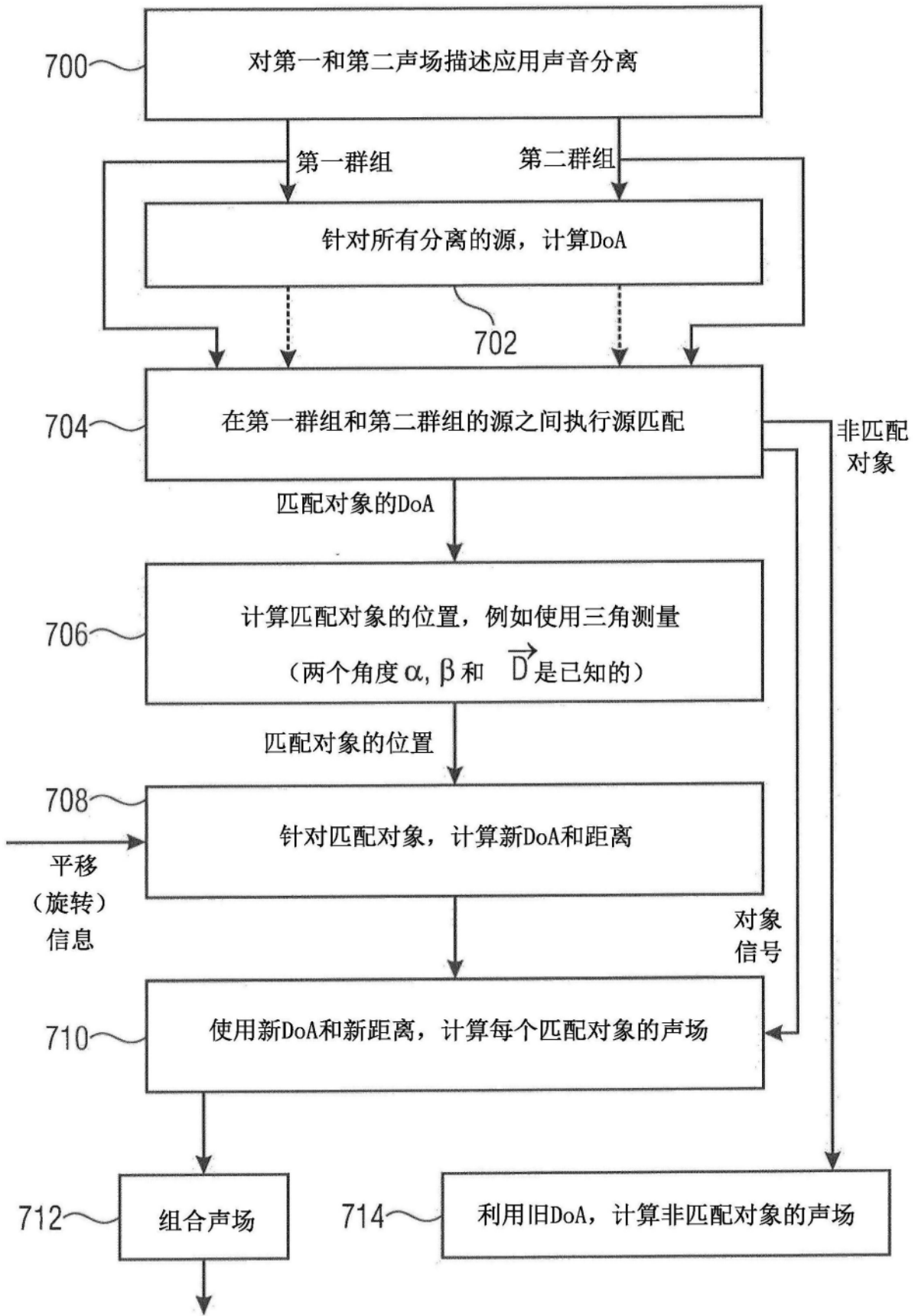


图6a

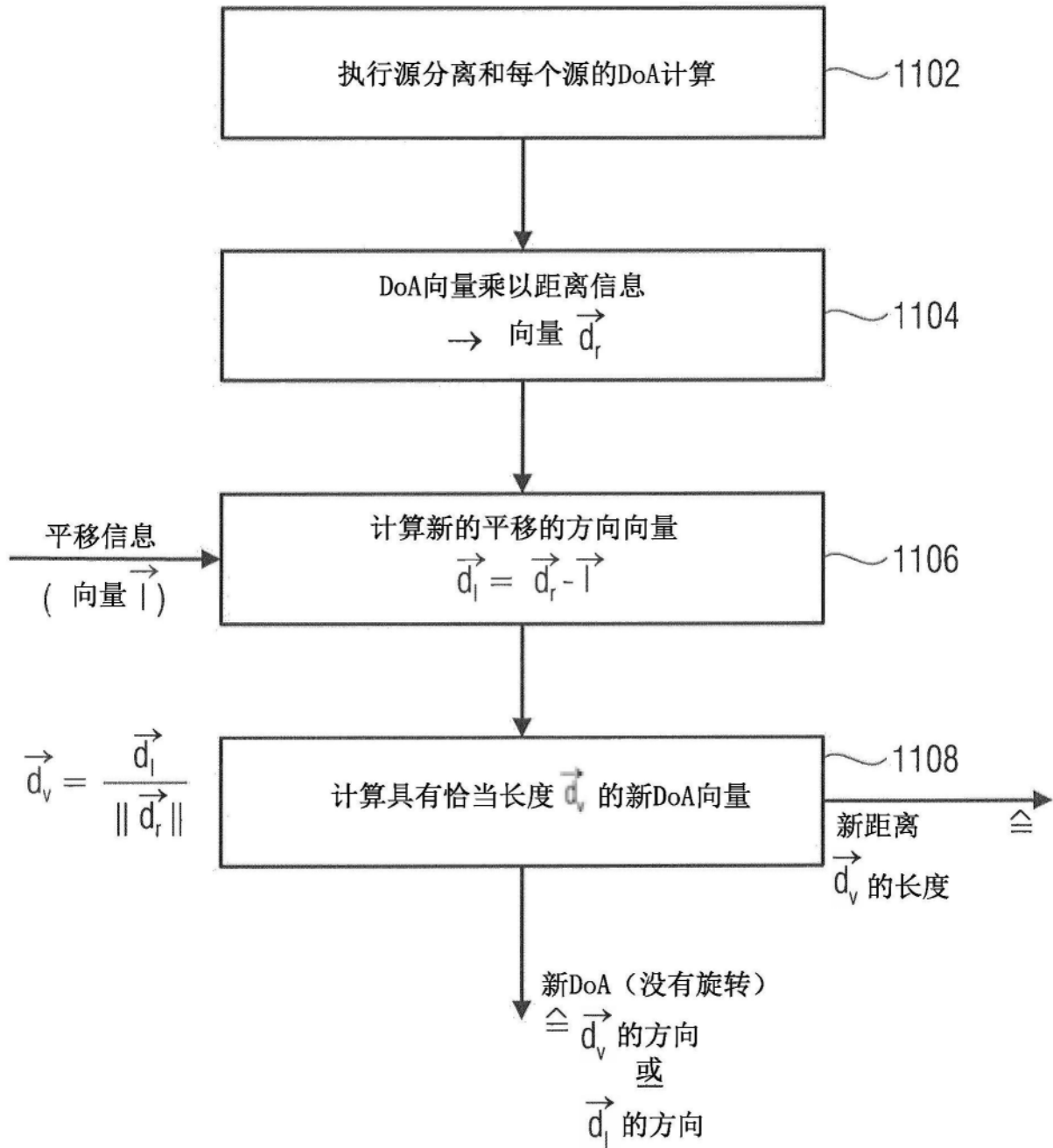


图6b

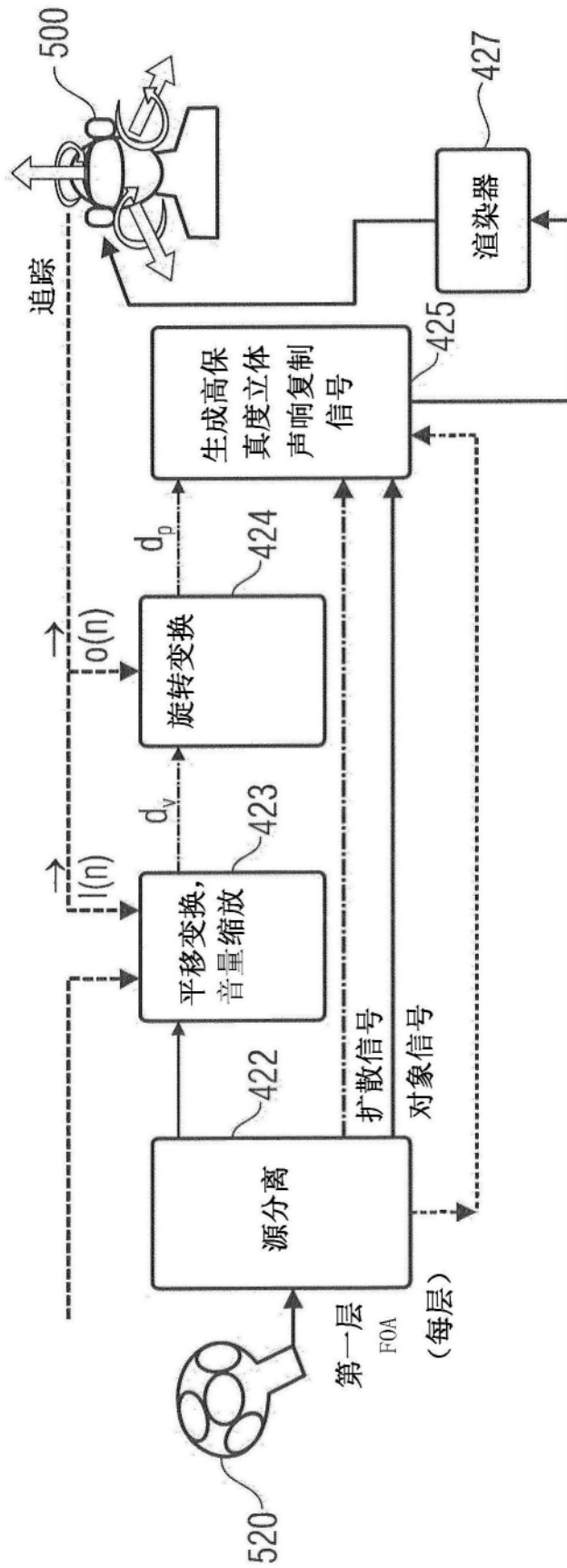


图6c

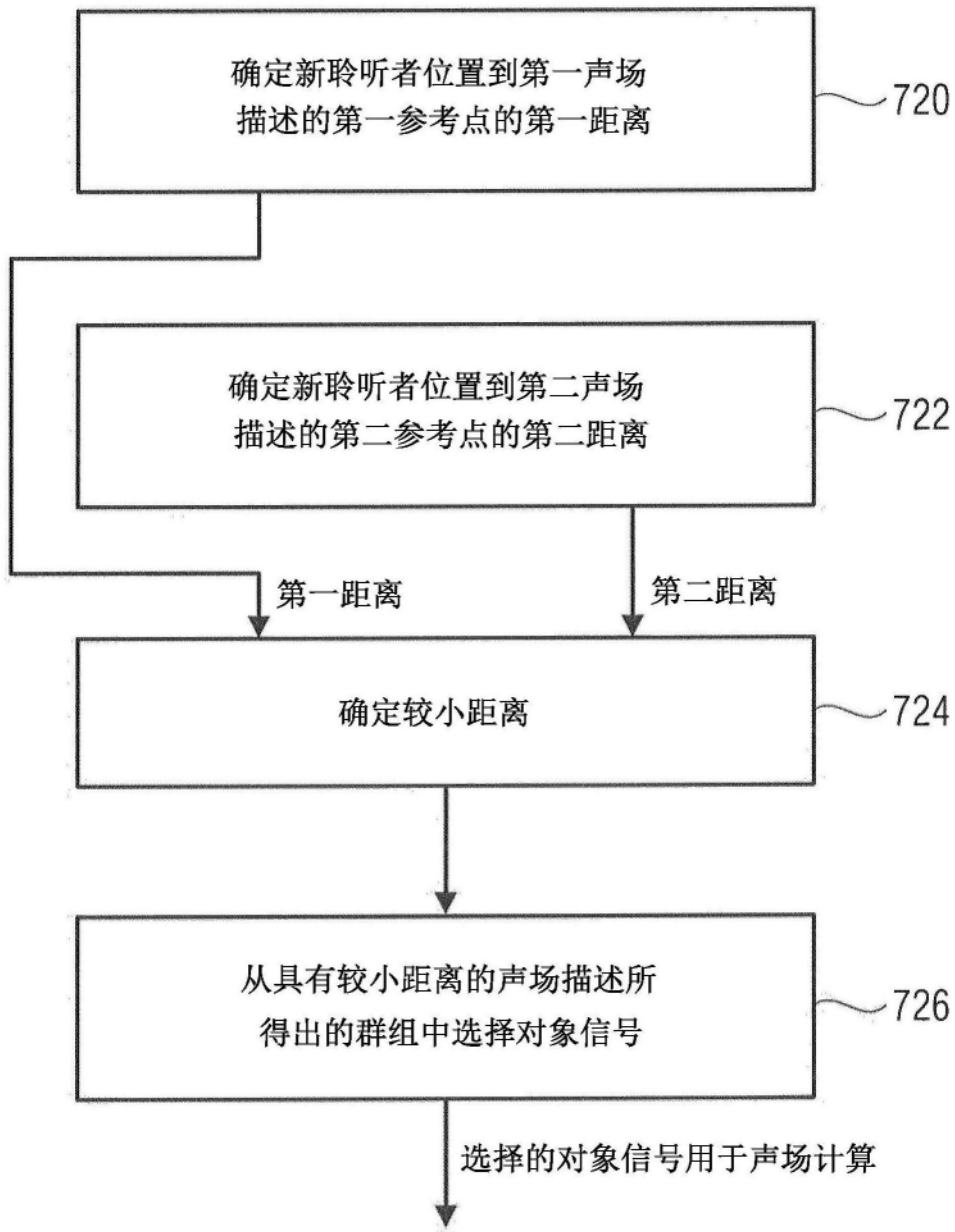
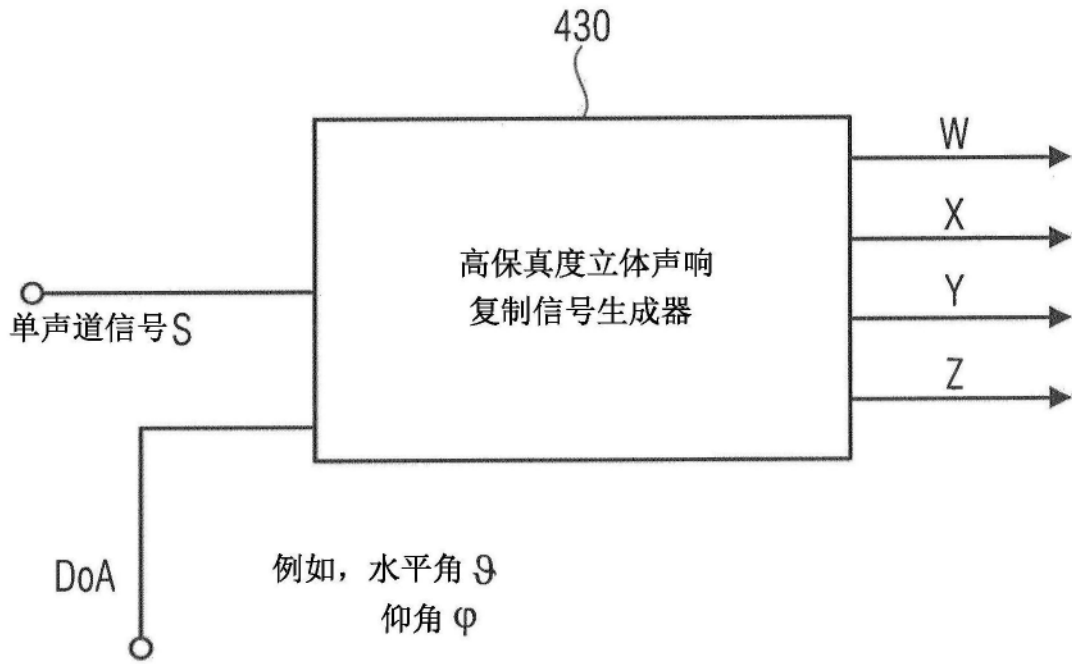


图7



全向分量  $W = S \cdot \frac{1}{\sqrt{2}}$

定向分量  $\begin{cases} X = S \cdot \cos \theta \cos \varphi \\ Y = S \cdot \sin \theta \cos \varphi \\ Z = S \cdot \sin \varphi \end{cases}$

图8

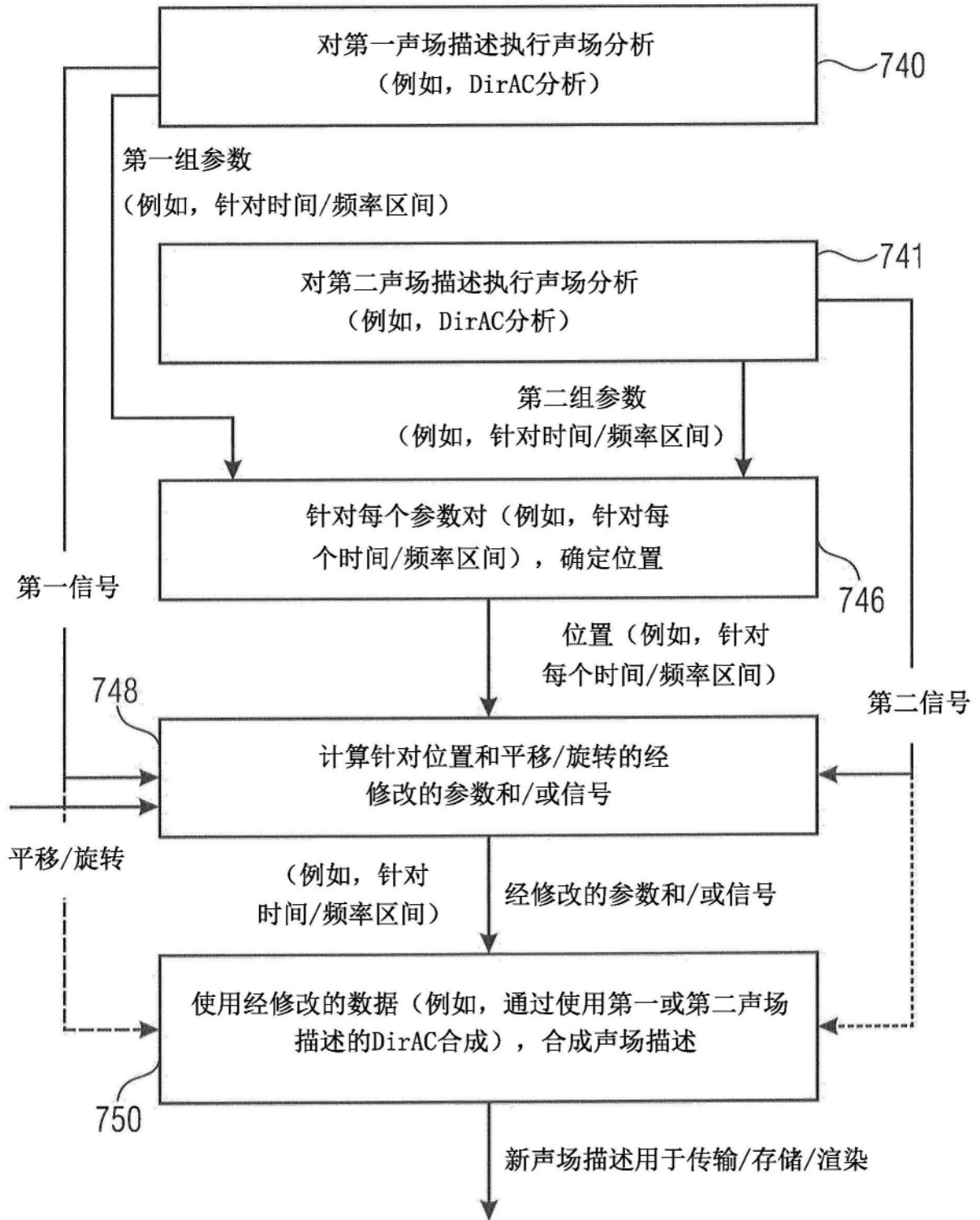


图9

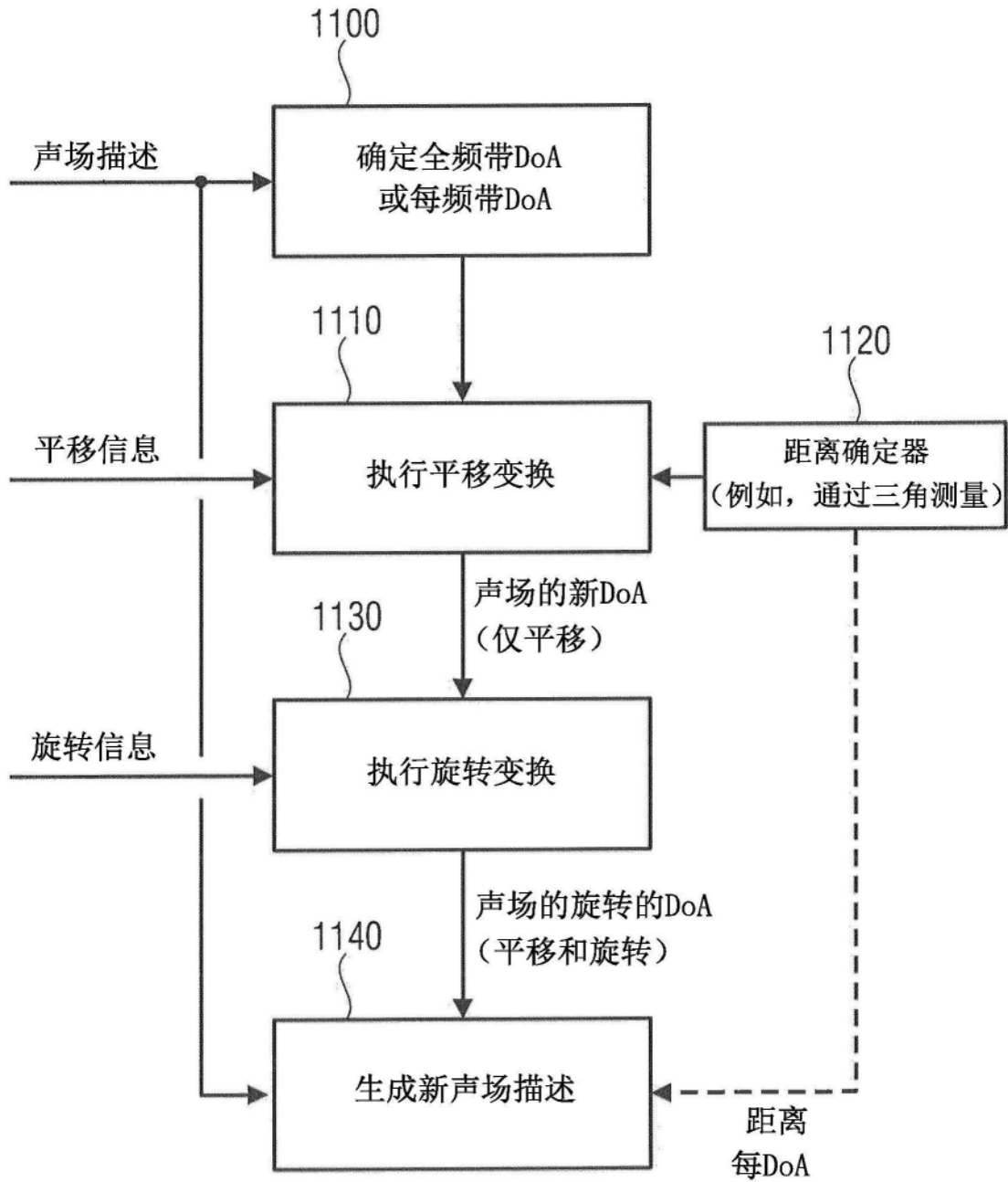
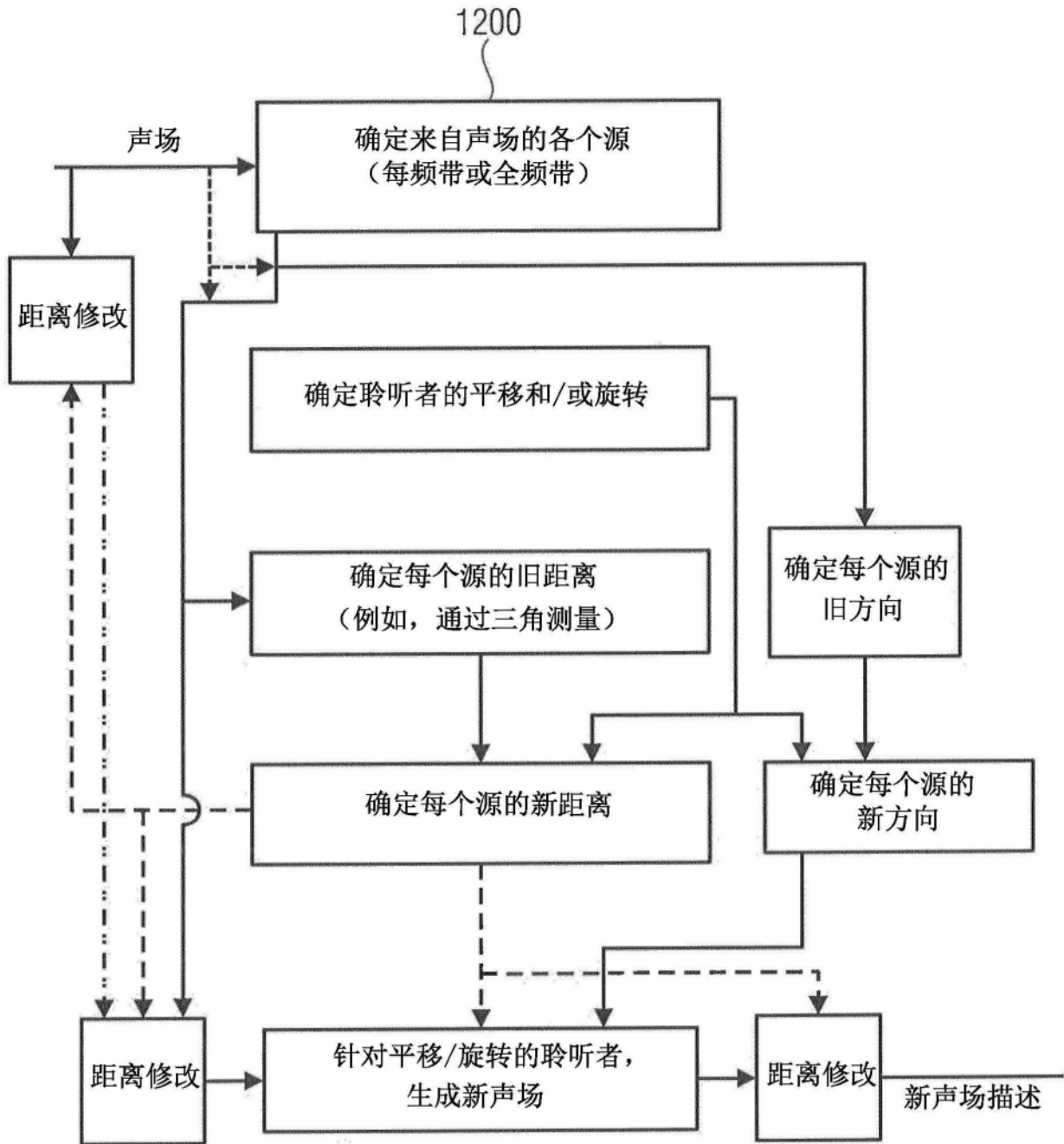


图10



例如，对原始声场，或对来自第一群组或第二群组的分离源；或对每个各个新高保真度立体声响复制信号，或对来自所有源的添加的高保真度立体声响复制信号的距离修改

图11

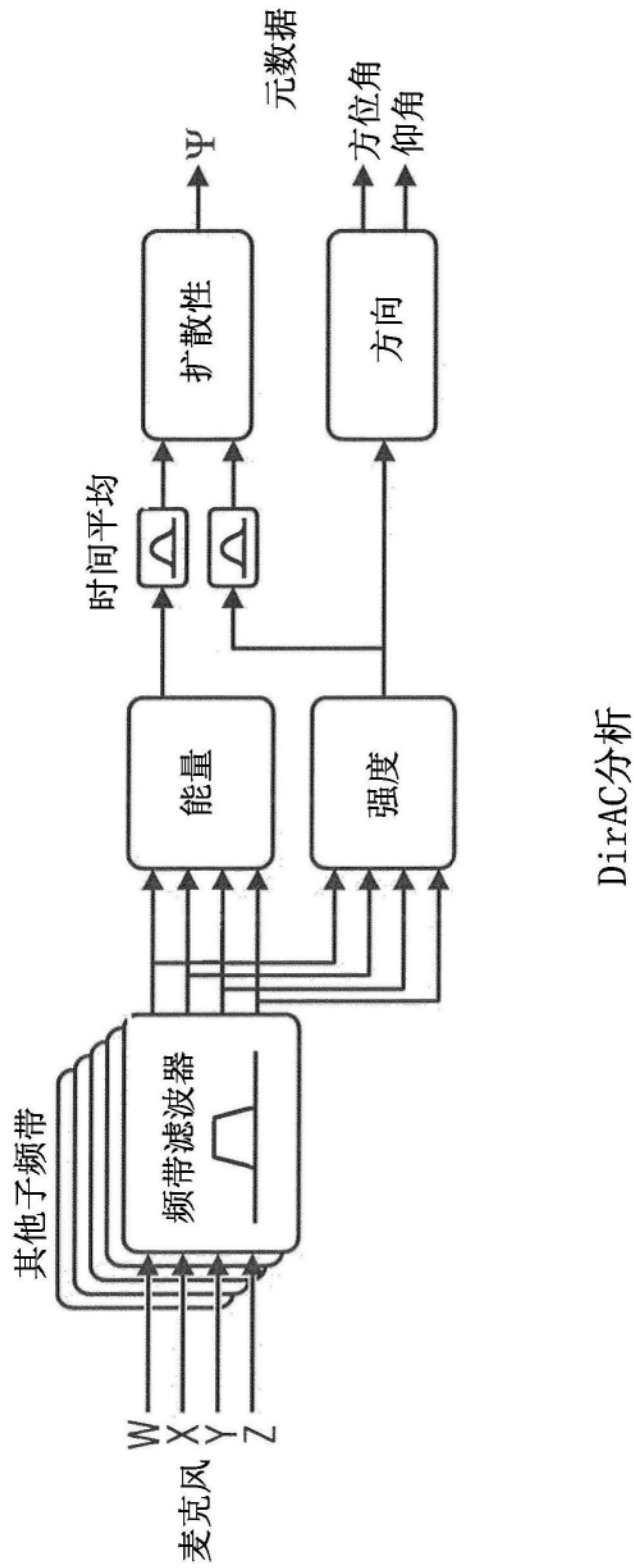
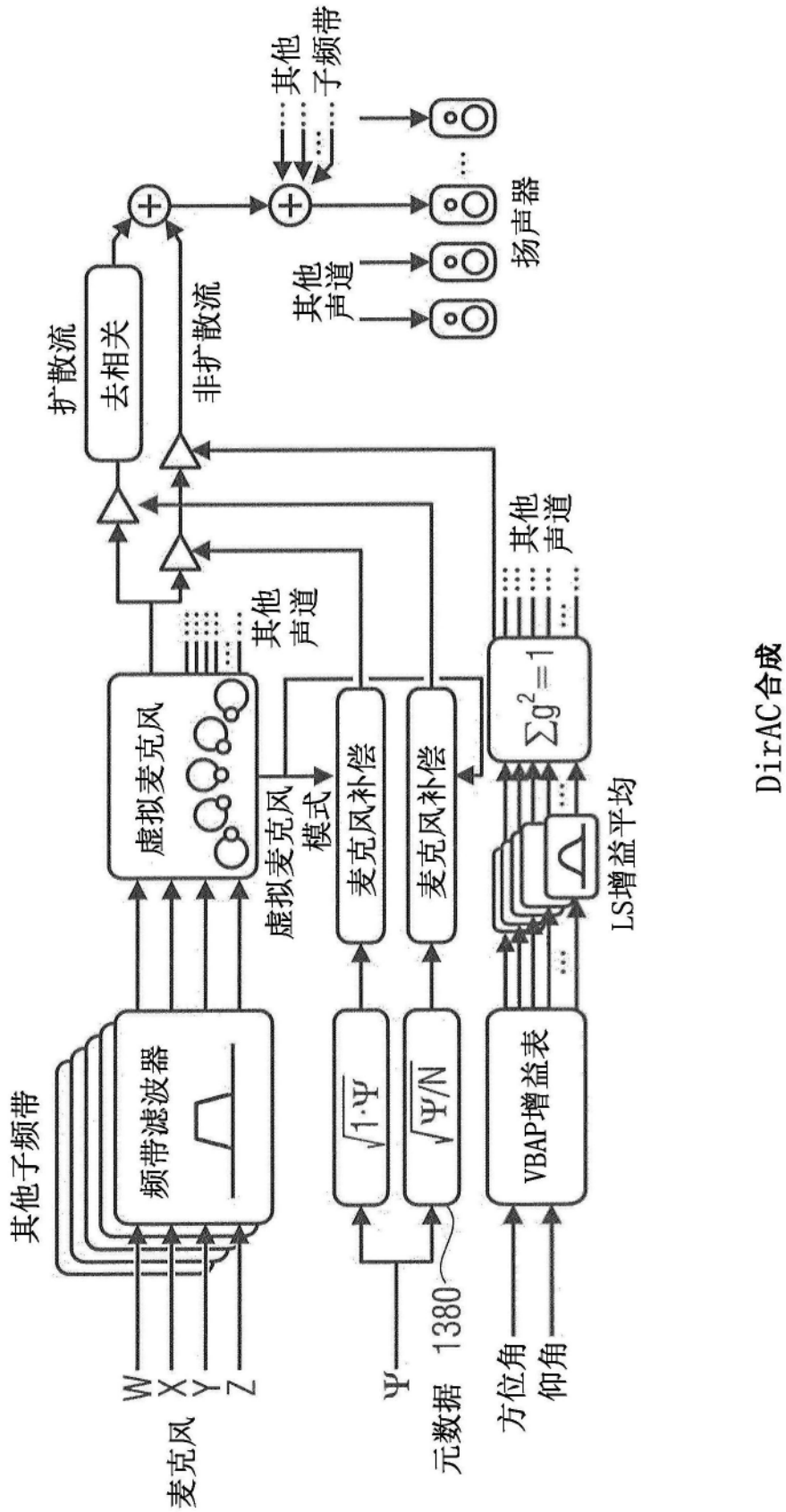


图12a



DirAC合成

图12b