



(12)发明专利

(10)授权公告号 CN 107925840 B

(45)授权公告日 2020.06.16

(21)申请号 201680050986.0

(22)申请日 2016.08.30

(65)同一申请的已公布的文献号
申请公布号 CN 107925840 A

(43)申请公布日 2018.04.17

(30)优先权数据
15183790.3 2015.09.04 EP

(85)PCT国际申请进入国家阶段日
2018.03.02

(86)PCT国际申请的申请数据
PCT/EP2016/070343 2016.08.30

(87)PCT国际申请的公布数据
W02017/037032 EN 2017.03.09

(73)专利权人 皇家飞利浦有限公司
地址 荷兰艾恩德霍芬

(72)发明人 W.P.J.德布勒因

(74)专利代理机构 中国专利代理(香港)有限公司 72001

代理人 刘红 陈岚

(51)Int.Cl.
H04S 7/00(2006.01)
H04N 13/332(2018.01)
H04N 13/398(2018.01)
H04N 13/128(2018.01)

(56)对比文件
CN 103699364 A,2014.04.02,
CN 103426163 A,2013.12.04,
CN 104054126 A,2014.09.17,
CN 103609105 A,2014.02.26,
EP 2871842 A1,2015.05.13,
EP 2475193 A1,2012.07.11,

审查员 刘贺

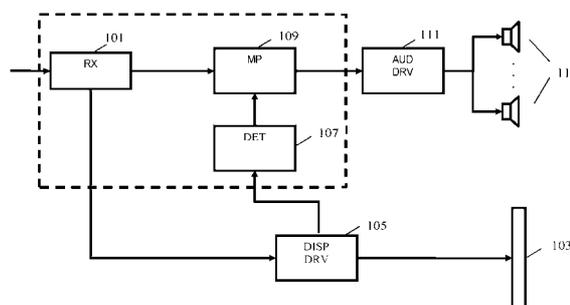
权利要求书2页 说明书20页 附图4页

(54)发明名称

用于处理音频信号的方法和装置

(57)摘要

音频信号处理装置包括接收器(101),其接收音频信号,音频信号包括与三维图像相关联的至少第一音频对象的音频数据。音频信号也包括指示用于第一音频对象的目标深度位置的深度位置数据。确定器(107)确定用于呈现三维图像的目标三维显示器的视觉渲染深度范围,以及映射器(109)用于将目标深度位置映射至用于音频对象的渲染深度位置,其中映射取决于视觉渲染深度范围。视觉渲染深度范围特定地可以是其中三维显示器能够精确地渲染对象的深度范围,并且映射器(109)可以修正音频对象声源的位置,以致这些与利用三维显示器所呈现的对应视觉对象的深度位置相匹配。



1. 一种音频信号处理装置,包括:

接收器(101),用于接收包括至少第一音频对象的音频数据的音频信号,第一音频对象在利用三维图像视觉表示的场景中表示音频源,所述音频信号进一步包括指示用于第一音频对象的第一深度位置的深度位置数据;

确定器(107),用于确定用于呈现所述三维图像的第一三维显示器的深度渲染属性,所述深度渲染属性指示第一三维显示器的视觉渲染深度范围;

映射器(109),用于将第一深度位置映射至用于第一音频对象的渲染深度位置;所述映射取决于所述视觉渲染深度范围。

2. 根据权利要求1所述的音频信号处理装置,其中所述音频信号进一步包括用于所述三维图像的第二显示器的深度渲染属性的指示;以及其中所述映射器(109)被安排成适配所述映射,以响应用于第二显示器的深度渲染属性。

3. 根据权利要求2所述的音频信号处理装置,其中所述映射是用于第二显示器的视觉渲染深度范围至用于第一三维显示器的视觉渲染深度范围的单调映射。

4. 根据权利要求1所述的音频信号处理装置,其中所述映射器被安排成:针对指示用于具有第一显示尺寸的第一三维显示器的第一视觉渲染深度范围的视觉渲染深度范围和针对指示用于具有第一显示尺寸的第一三维显示器的第二视觉渲染深度范围的视觉渲染深度范围,不同地映射第一深度位置,第一视觉渲染深度范围不同于第二视觉渲染深度范围。

5. 根据权利要求1所述的音频信号处理装置,其中所述映射围绕屏幕深度水平而不是不对称的,所述屏幕深度水平对应于零差距。

6. 根据权利要求1所述的音频信号处理装置,其中所述映射是非线性映射。

7. 根据权利要求1所述的音频信号处理装置,其中所述音频信号包括指示是否所述音频对象是屏幕相关音频对象的用于第一音频对象的标志,屏幕相关音频对象是对其而言渲染处理取决于第一三维显示器的视觉渲染属性的音频对象;以及其中所述映射器(109)被安排成仅在所述标志指示至少一个音频对象是屏幕相关音频对象时才将所述映射应用于第一音频对象的第一深度位置。

8. 根据权利要求1所述的音频信号处理装置,进一步包括用于从多个二维图像中生成所述三维图像的转换器,所述三维图像的生成包括确定用于所述三维图像的深度映射;以及其中所述确定器(107)被安排成确定所述深度渲染属性,以响应所述深度映射。

9. 根据权利要求1所述的音频信号处理装置,其中所述音频信号进一步包括指示所述深度渲染属性的显示数据,以及所述确定器被安排成:提取所述显示数据;和确定所述深度渲染属性,以响应所述显示数据。

10. 根据权利要求1所述的音频信号处理装置,进一步包括选择器(401),其被安排成选择利用所述三维图像表示的三维场景的子集,以便呈现在第一三维显示器的完全显示区域上;以及其中所述映射器(109)被安排成适配所述映射,以响应所述子集的深度渲染属性。

11. 根据权利要求10所述的音频信号处理装置,其中所述映射器(109)被安排成将所述映射适配成所述子集的视觉渲染深度范围至用于第一三维显示器的视觉渲染深度范围的单调映射。

12. 根据权利要求11所述的音频信号处理装置,其中用于第一音频对象的第一深度位置在所述子集之外。

13. 一种处理音频信号的方法,所述方法包括:

接收包括至少第一音频对象的音频数据的音频信号,第一音频对象在利用三维图像视觉表示的场景中表示音频源,所述音频信号进一步包括指示用于第一音频对象的第一深度位置的深度位置数据;

确定用于呈现所述三维图像的第一三维显示器的深度渲染属性,所述深度渲染属性指示第一三维显示器的视觉渲染深度范围;

将第一深度位置映射至用于第一音频对象的渲染深度位置;所述映射取决于所述视觉渲染深度范围。

14. 一种计算机程序产品,其包括计算机程序代码装置,所述计算机程序代码装置被适配成当所述程序被运行在计算机上时执行权利要求13的所有步骤。

用于处理音频信号的方法和装置

技术领域

[0001] 本发明涉及用于处理音频信号的方法和装置,并且尤其但不完全涉及三维音频信号的处理以便与相关联的三维视频一起呈现。

背景技术

[0002] 音频的渲染已变得越来越多样,并且一系列的渲染方案 and 用户体验已被引入。例如,作为音频视觉(audio visual)体验的一部分的空间音频尤其采用环绕声的形式而已变得普及。在这样的系统中,图像或视频与正被创建的相关联的空间音频环境一起进行呈现。

[0003] 为了在空间音频渲染中支持变化和灵活性,用于表示(represent)空间音频的许多格式已被开发。

[0004] 最近的格式是MPEG Surround(环绕)格式。然而,虽然这为许多应用提供合适的格式,但是对于其他的应用来说,它仍然不像所希望的那样灵活。例如,仍然针对特定的扬声器设置例如ITU 5.1扬声器设置来产生和发射音频。在不同设置上以及在非标准(即,灵活的或用户定义的)扬声器设置上的再现没有被规定。

[0005] 为了提供音频的更灵活表示,正在开发将个别音频源表示为个别音频对象的格式。因而,已提议提供各自表示特定音频源(包括例如背景、扩散和环境声源)的个别音频对象,而不是利用与特定(标称或参考)位置相对应的音频通道来表示音频场景。典型地,可以给音频对象提供(可选的)位置信息,其指示音频对象在声场(sound stage)中的目标位置。因而,在这样的方案中,音频源可以被表示为独立的且单个的音频对象,而不是利用其对于与特定的、预定的位置相关联的音频通道所作出的贡献来表示。

[0006] 为了支持这样的方案,MPEG已标准化被称为“Spatial Audio Object Coding(空间音频对象编码)”(ISO/IEC MPEG-D SAOC)的格式。与多通道音频编码系统诸如DTS、Dolby Digital和MPEG Surround相比而言,SAOC提供个别音频对象而非音频通道的高效编码。而在MPEG Surround中,每一个扬声器通道能够被认为起源于声音对象的不同混合,SAOC虑及个别声音对象在多通道混合中的位置的交互式操纵。

[0007] 与MPEG Surround相类似地,SAOC也创建单声道或立体声下混(downmix)。另外,对象参数被计算并被包括。在解码器侧上,用户可以操纵这些参数来控制个别对象的各种特性,诸如位置、水平、均衡乃至应用诸如混响之类的效果。

[0008] SAOC允许更灵活的方案并且尤其除了只有再现通道之外通过发射音频对象也允许更多的基于渲染的适应性。如果空间被扬声器充分覆盖的话,这允许解码器侧将音频对象放置在空间中的任意位置上。这样一来,在所发射的音频与再现或渲染设置之间没有关系,因此能够使用任意扬声器设置。这对于例如其中扬声器几乎从不在预期位置上的典型起居室中的家庭影院设置而言是有利的。在SAOC中,在解码器侧上决定这些对象在声音场景中被放置在哪里。然而,鉴于音频对象的渲染侧操纵被支持,典型地希望能够渲染音频而不要求用户输入,同时仍提供合适的声场。特别地,当与链接的视频信号一起提供音频时,希望在与图像中的位置相对应的位置上渲染音频源。相应地,时常可以给音频对象提供指

示针对个别音频对象所建议的渲染位置的目标位置数据。

[0009] 基于音频对象的格式的其他示例包括MPEG-H3D Audio [ISO/IEC 23008-3 (DIS): Information technology--High efficiency coding and media delivery in heterogeneous environments--Part3: 3D audio, 2014.]、ADM [EBU Tech 3364 "Audio Definition Model Ver.1.0", 2014]和专有标准诸如Dolby Atmos [SMPTE TC-25CSS10 WG on "Interoperable Immersive Sound Systems for Digital Cinema", 2014]与DTS-MDA [ETSI document TS 103 223, "The Multi-Dimensional Audio (MDA) Content Creation Format Specification with Extensions for Consumer Environments", 2014]。

[0010] 基于对象的音频制作和再现的概念提供超越传统的基于通道的方案的许多优点。特别地,将空间中的特定位置分配给个别声音对象的可能性对于交互性提供很大程度的灵活性、可伸缩性和新的可能性。

[0011] 如果使用合适的音频渲染技术,基于对象的音频使之能够以感知现实的方式在3D空间中的任何位置上定位对象,其包括相对于收听者的方位、高程(elevation)和距离的精确定位。这样的渲染技术的一些示例是:双耳耳机再现、反式耳(trans-aural)扬声器再现、Wave Field Synthesis(波场合成)扬声器再现以及在某种程度上VBAP扬声器再现。

[0012] 典型地,与在视频显示器上渲染的对应视频内容一起呈现基于对象的音频内容。如果音频对象对应于在屏幕上呈现的视觉对象,通常所希望的是:在所感知的听觉和视觉对象位置之间具有某种空间同步或一致,即,对象的声音和图像在空间中相匹配。如果这样的同步不存在,即,如果听觉和对应的视觉对象的感知位置是明显不同的,这可能让用户感到困惑并降低视听(audio-visual)呈现的整体感知质量或沉浸感(immersion)。

[0013] 然而,由于渲染设置、特别是视频渲染设置典型地明显变化,所以实现紧的(tight)空间同步能够是困难的,并且这在许多情形中可能导致降级的用户体验。特别地,不同显示器的能力和渲染特征可能明显变化,并且这可能在不同的情景中引起不同的渲染。

[0014] 因此,用于处理空间音频信号以便渲染的改进方案将是有利的,并且尤其,允许增加的灵活性、便利的操作、降低的复杂度和/或资源需求、针对相关联视频而言改进的空间同步和/或改进的用户体验的方案将是有利的。

发明内容

[0015] 相应地,本发明寻求单个地或在任何组合中优选地减轻、缓解或消除上述缺点之中的一个或多个缺点。

[0016] 根据本发明的一方面,提供一种音频信号处理装置,其包括:接收器(101),用于接收包括至少第一音频对象的音频数据的音频信号,第一音频对象在利用三维图像视觉表示的场景中表示音频源,音频信号进一步包括指示第一音频对象的第一深度位置的深度位置数据;确定器(107),用于确定用于呈现三维图像的第一三维显示器的深度渲染属性,深度渲染属性指示第一三维显示器的视觉渲染深度范围;映射器(109),用于将第一深度位置映射至第一音频对象的渲染深度位置;映射取决于视觉渲染深度范围。

[0017] 本发明在与三维图像一起渲染音频对象时可以允许改进的用户体验。特定地,场

景可以被呈现为三维图像(或提供三维视频序列的图像集合)以及与场景中的音频源相对应的音频的视听渲染。组合的视频和音频呈现可以提供改进的用户体验,并且映射音频对象的位置的方案可以导致改进的空间同步并且特定地导致观众的音频与视觉感知之间更紧密的一致性。

[0018] 发明人已意识到:尤其三维显示器在渲染能力方面可能明显变化并且特定地可能具有非常不同的深度渲染属性。发明人已进一步意识到:这样的深度渲染变化可能导致所感知的用户体验方面的明显变化。尤其,他们已意识到:深度渲染变化对于所感知的音频与视频之间的空间同步的感知效果是明显的并能够导致明显降级的视听用户体验。他们已进一步意识到:这可以通过对取决于显示器的空间深度渲染特征的音频对象深度位置应用重新映射来减轻和/或缓解。

[0019] 映射可以提供渲染深度位置作为第一深度位置的函数,第一深度位置也将被称为目标深度位置。第一三维显示器也将被称为目标(三维)显示器。映射函数可以响应于利用深度渲染属性所指示的视觉渲染深度范围来确定。该函数在许多实施例中可以是单调函数,即,映射可以提供渲染深度位置作为目标深度位置的单调函数。

[0020] 第一/目标深度位置可以提供用于第一音频对象的渲染的建议位置。它特定地可以是用于音频对象的渲染的标称位置。标称位置可以是用于渲染的可选位置,并且渲染器可以相应地改变这个位置。例如在捕获场景时,建议位置可以例如被生成并被确定为内容制作的一部分。因而,音频信号可以包括建议应该在其上面渲染第一音频对象的深度的数据。将认识到:位置数据典型地也可以包括允许非深度位置被确定的位置信息,诸如x或y位置(其中深度位置被表示为z位置)或方位和高程位置。在许多实施例中,目标深度位置可以作为指示用于第一音频对象的三维目标位置的目标位置的一部分。

[0021] 将认识到:任何合适的参考可以用于深度位置。例如,在许多实施例中,深度位置可以相对于屏幕水平(screen level)深度来参考,并且特定地可以是针对深度位置的参考,其中深度位置对应于针对不同的视角即分别针对右眼和左眼视图而言在图像中的显示(x,y)位置之间没有差距(disparity)。

[0022] 渲染深度位置类似地可以是指示用于第一音频对象的预期三维渲染位置的目标渲染位置的一部分。渲染深度位置可以被视为修改的目标位置,并因而可以是用于第一音频对象的渲染的修改的目标位置。渲染深度位置提供在利用空间音频渲染器进行渲染时应该从哪里感知第一音频对象的深度位置的指示。渲染深度位置可以通过第一深度位置的映射而生成的第二深度位置。它特定地可以是用于第一音频对象的修改的可选的标称位置。

[0023] 目标三维显示器可以是适合于渲染/呈现三维图像的任何显示器,其例如包括使用例如同步眼镜来提供三维体验的自动立体显示器或立体显示器。目标三维显示器是针对其执行映射的显示器并且可以对应于被假定用于三维图像的渲染的显示器。因而,如果利用目标三维显示器来呈现三维图像并且第一音频对象被渲染在与渲染深度位置相对应的位置上,能够实现在所感知的音频与视觉深度位置之间更紧密的对应性。实际上,该方案可以在空间上链接或同步视觉和音频域。例如,当渲染第一音频对象时,与利用第一音频对象表示的声源相对应的图像对象的视觉位置可以更紧密地与音频的位置相对准。

[0024] 深度渲染属性包括指示利用三维显示器进行的深度的渲染的视觉渲染深度范围。

它在许多实施例中可以指示三维显示器的深度渲染能力,诸如例如可用的深度范围。例如,视觉渲染深度范围可以指示其中三维显示器能够渲染图像部分而这不会导致不可接受的疲劳、质量降级、模糊等等的范围(可能地,只在三维显示器前面或后面)。该范围时常可以被给定为例如以米(或任何其他距离测量)为单位测量的物理范围(在真实的空间中)。

[0025] 确定器例如可以是能够确定深度渲染属性、包括例如通过从外部源或存储器接收或检索适当数据来确定深度渲染属性的任何控制器、处理器、单元、处理、子实体等等。

[0026] 映射器例如可以是能够从目标深度位置中确定渲染深度位置的任何控制器、处理器、单元、处理、子实体等等,即它能够将指示第一位置的数据映射成指示第二位置的数据。映射器可以包括用于适配/修改映射以响应视觉渲染深度范围的适配器。

[0027] 在一些实施例中,音频处理器可以包括用于响应于渲染深度位置来渲染第一音频对象的渲染器。渲染可以将第一音频对象渲染成从与渲染深度位置相对应的位置中进行感知(特定地,以致在标称位置上的标称收听者将感知到第一音频对象位于渲染深度位置上)。

[0028] 在许多实施例中,第一音频对象可以表示用于三维图像的图像对象的音频。因而,在许多实施例中,第一音频对象可以为在三维图像中特定地可以是可视的特定图像对象提供音频。然而,在其他的情景中,第一音频对象可以提供并不直接对应于图像对象的场景中的音频源的描述。例如,它可以提供描述来自图像中例如由于太小或被遮挡而不是直接可视的实体的音频源的音频数据。

[0029] 根据本发明的可选特性,音频信号进一步包括用于三维图像的第二显示器的深度渲染属性的指示;并且其中映射器被安排成适配映射,以响应用于第二显示器的深度渲染属性。

[0030] 这在许多情景和实施例中可以允许改进的用户体验并且特定地可以提供音频与视频之间改进的空间同步。也将被称为参考显示器的第二显示器典型地可以是三维图像的制作显示器或标称显示器。参考显示器可以是被假定用于由源或内容制作侧进行的渲染的显示器。第一音频对象的目标深度位置可以对应于在其上面渲染三维图像的对应图像片段(segment)的深度位置,如果这个渲染正在使用具有与参考显示器相同的属性的显示器的话。用于参考显示器的深度渲染属性特定地可以是参考深度渲染范围。

[0031] 根据本发明的可选特性,映射是用于第二显示器的视觉渲染深度范围到用于第一三维显示器的视觉渲染深度范围的单调映射。

[0032] 这在许多实施例中在保持低复杂度的同时可以提供改进的用户体验。映射可以提供渲染深度位置作为目标深度位置的单调函数。映射可以是目标深度位置到渲染深度位置的线性或分段(piecewise)线性映射。

[0033] 深度渲染属性包括目标三维显示器的渲染深度范围的指示。

[0034] 这在许多应用中可以提供特别有效且有用的深度渲染属性。深度渲染范围可以提供用于适配位置以提供改进的空间同步的特别有利的信息。渲染深度范围可以指示相对于三维显示器的物理空间的范围。该范围可以例如以米或对应的长度测量为单位来指示。渲染深度范围特定地可以针对在屏幕前面的对象来提供,并且可以指示其中图像片段被认为利用足够的质量来渲染且对观众而言没有引起不希望的效果(诸如疲劳)的范围。

[0035] 根据本发明的可选特性,映射器被安排成:针对指示用于具有第一显示尺寸的第

一三维显示器的第一视觉渲染深度范围的视觉渲染深度范围以及针对指示用于具有第一显示尺寸的第一三维显示器的第二视觉渲染深度范围的视觉渲染深度范围,不同地映射第一深度位置,第一视觉渲染深度范围不同于第二视觉渲染深度范围。

[0036] 具有相同显示尺寸的不同显示器可以具有不同的视觉渲染深度范围。映射器可以被安排成针对这样的显示器而不同地映射第一位置,即,映射器可以被安排成针对具有相同的显示尺寸但是不同的视觉渲染深度范围的显示器而不同地映射第一位置。用于显示器的深度范围可以取决于各种各样的属性,其包括例如分辨率/像素尺寸、所生成的视图的数量、光学组件的质量、视图形成元素的属性等等。因而,即使对于具有相同显示尺寸的显示器而言,所支持的操作深度范围也可能是不同的,并且这可以通过映射来反映。

[0037] 映射器可以被安排成针对第一三维显示器的相同显示尺寸但是针对指示不同的视觉深度渲染范围的视觉深度渲染属性而不同地映射第一位置。

[0038] 根据本发明的可选特性,映射围绕屏幕深度水平而是不对称的,屏幕深度水平对应于零差距(在左眼与右眼视图之间)。

[0039] 这在许多实施例中可以提供改进的性能。特别地,映射在屏幕/显示器前面和后面可能是不同的。对于分别在三维显示器的显示/屏幕平面前面、后面渲染的图像片段来说,映射可以特定地反映三维显示器的渲染能力的差异。在许多实施例中,映射可以被安排为仅修改在显示平面深度前面的深度值。

[0040] 屏幕深度水平对应于对其而言在不同的视图(旨在用于不同的眼睛)之间在像素位置中没有差距,即,它对应于被感知为在屏幕水平上(并且不在屏幕前面或后面)的深度。

[0041] 根据本发明的可选特性,映射是非线性映射。

[0042] 这在许多实施例中可以提供改进的性能。特别地,映射可以包括映射目标深度位置,以致从显示平面深度到渲染深度位置的距离被限于不超过最大值。

[0043] 根据本发明的可选特性,音频信号包括指示是否音频对象是屏幕相关音频对象的用于第一音频对象的标志,屏幕相关音频对象是对其而言渲染处理取决于第一三维显示器的视觉渲染属性的音频对象;以及其中映射器(109)被安排成仅在标志指示至少一个音频对象是屏幕相关音频对象时才将映射应用于第一音频对象的第一深度位置。

[0044] 这在许多实施例中可以提供改进的处理并且特定地可以提供用于支持不同类型的音频的有效方案。该方案可以提供可以被适配来反映对应图像的视觉渲染的改进的音频渲染。特别地,它可以允许在所渲染的音频与视频之间更紧密的相关性。

[0045] 根据本发明的可选特性,音频信号处理装置进一步包括用于从多个二维图像中生成三维图像的转换器,三维图像的生成包括确定用于三维图像的深度映射;以及其中确定器被安排成确定深度渲染属性,以响应深度映射。

[0046] 这可以提供特别有效的系统并且特别地提供生成适合于适配音频对象的映射的深度渲染属性的特别有效的方式。

[0047] 根据本发明的可选特性,音频信号处理装置进一步包括指示深度渲染属性的显示数据,并且确定器被安排成提取显示数据和确定深度渲染属性以响应显示数据。

[0048] 这在许多实施例中可以提供用于给音频处理器提供深度渲染属性的特别有效的方案并且特别地可以促进与其他功能的接口。

[0049] 根据本发明的可选特性,音频信号处理装置进一步包括选择器,其被安排成选择

利用三维图像所表示的三维场景的子集,以便呈现在第一三维显示器的完全显示区域上;以及其中映射器被安排成响应于子集的深度属性来适配映射。

[0050] 这在提供有吸引度的空间同步性的同时在渲染中可以提供有吸引力的用户体验和附加的灵活性。该方案例如在用户放大和缩小场景时尤其可以允许在所渲染的音频场(stage)与所呈现的图像之间更紧密的对应性。

[0051] 根据本发明的可选特性,映射器被安排成将映射适配成子集的视觉渲染深度范围到用于第一三维显示器的渲染深度范围的单调映射。

[0052] 这在许多实施例中可以提供特别有效的操作和/或有利的性能。视觉深度范围可以是子集内的图像组件诸如图像对象的深度范围。视觉深度范围特定地可以是或指示子集内的差距或深度值的范围。

[0053] 根据本发明的可选特性,用于第一音频对象的第一深度位置在子集之外。

[0054] 这在许多实施例中可以提供特别有效的操作和/或有利的性能。

[0055] 根据本发明的一方面,提供一种处理音频信号的方法,该方法包括:接收音频信号,其包括在利用三维图像视觉表示的场景中表示音频源的至少第一音频对象的音频数据,音频信号进一步包括指示用于第一音频对象的第一深度位置的深度位置数据;确定用于呈现三维图像的第一三维显示器的深度渲染属性,深度渲染属性指示第一三维显示器的视觉渲染深度范围;将第一深度位置映射到用于第一音频对象的渲染深度位置;映射取决于视觉渲染深度范围。

[0056] 根据本发明的一方面,提供一种音频信号,其包括:在利用三维图像视觉表示的场景中表示音频源的至少第一音频对象的音频数据;指示用于第一音频对象的第一深度位置的深度位置数据;指示是否音频对象是屏幕相关音频对象的用于第一音频对象的标志,屏幕相关音频对象是对其而言渲染处理取决于三维显示器的视觉渲染属性的音频对象;以及包括三维图像的该三维显示器的深度渲染属性的显示数据,深度渲染属性指示三维显示器的视觉渲染深度范围。

[0057] 本发明的这些和其他方面、特性和优点从下文描述的(多个)实施例中将是显而易见的并将参考这(些)实施例来阐述。

附图说明

[0058] 本发明的实施例将参考附图、仅通过示例来描述,其中:

[0059] 图1是包括根据本发明的一些实施例的音频信号处理装置的视听渲染系统的示例的图解;

[0060] 图2是包括根据本发明的一些实施例的音频信号处理装置的视听渲染系统的示例的图解;

[0061] 图3是根据本发明的一些实施例的音频信号处理装置的示例的图解;和

[0062] 图4是包括根据本发明的一些实施例的音频信号处理装置的视听渲染系统的示例的图解。

具体实施方式

[0063] 以下描述专注于可应用于与相关联的三维视频在三维显示器上的呈现相结合来

渲染空间音频的系统的本发明的实施例。然而,将认识到:本发明并不限于这个应用而可以应用于许多其他的实现方式、系统和产品,其例如包括静止图像的呈现或音频数据解码和转码。

[0064] 图1举例说明包括根据本发明的一些实施例的音频信号处理装置的视听渲染系统。

[0065] 视听渲染系统和音频信号处理装置包括接收器101,其被安排成接收包括音频数据和视频数据的数字视听信号。视频数据包括表示至少一个3D图像的数据,并且在特定示例中表示3D视频流的帧的视频数据被包括。

[0066] 在图1的系统中,3D视频流被呈现在3D显示器103上。3D视频数据被馈送至显示驱动器105,其为3D显示器生成合适的驱动信号,以引起3D图像被呈现。例如,3D显示器103可以是完整显示器,诸如标准3D计算机监视器或TV,并且显示驱动器105可以给3D显示器103提供合适编码的视频数据。例如,在一些实施例中,显示驱动器105可以简单地从接收器101接收视频数据并且以合适的格式将这个转发到3D显示器103。在一些实施例中,显示驱动器105可以简单地提供直接控制3D显示器103的个别像素的直接电气和/或机械接口。在一些实施例中,显示驱动器105可以包括转码功能。在一些实施例中,3D显示器103简单地可以由显示驱动器105直接驱动的3D显示面板。例如,显示驱动器105在一些实施例中可以被安排成解码视频数据并且为3D显示器103的个别像素生成合适的驱动值。例如,对于其中在单个显示设备中诸如在电视中实现图1的整个视听渲染系统的实施例而言,情况可能是这样的。在这样的情况下,3D显示器103可以表示显示面板。

[0067] 将认识到:许多不同类型的三维显示器是已知的,并且任何的用于提供视觉3D渲染的合适方案可以被使用。特定地,3D显示器可以通过给观众的两只眼睛提供所观看的场景的不同视图来给观看体验添加第三维度。这能够通过让用户佩戴眼镜以分隔开所显示的两个视图来实现。然而,由于这对用户而言是相对不便的,所以在许多情景中所希望的是使用直接生成不同视图并将其投射到用户的眼睛的自动立体显示器。

[0068] 除了视频数据之外,视听信号还包括表示与视频数据相关联的音频的音频数据。音频数据因而可以提供对其而言视频数据提供视觉描述的场景的音频描述。因此,音频数据可以包括旨在与视频流的呈现一起呈现的音频的数据。

[0069] 在示例中,音频数据可以包括规定用于场景中的各种音频源的许多不同的音频对象的音频数据。这些音频对象之中的一些音频对象可以是不与任何特定位置相关联的扩散或一般声音。例如,一个或多个音频对象可以是表示背景或环境声音的环境或背景音频对象。然而,其他的音频对象可以与对应于声场中的特定位置的特定声源相关联。相应地,视听信号也可以包括提供对于特定音频对象所期望的位置的指示的位置数据,即它可以包括用于音频对象的目标位置数据。

[0070] 因而,音频信号包括场景中的至少一个(但是典型地,多个)音频源的音频数据。进一步,场景也利用3D图像来视觉表示,并因而音频对象与表示同一场景的3D图像相关联。

[0071] 在一些情景中,一个或多个音频对象可以与特定图像对象相关联。例如,音频对象可以对应于来自图像中的人类的语音。这个音频对象因而可以与对应于人类或更特定地(例如,为了特写)人类的嘴的图像对象相关联。视听信号在这样的示例中可以包括描述语音的音频数据以及指示对应图像对象(即,说话者(或说话者的嘴))在3D图像中的位置的位

置数据。因而,在许多实施例或情景中,所接收的音频信号的至少第一音频对象表示三维图像的图像对象的音频。因而,该图像提供场景中的对象的视觉表示,并且第一音频对象可以提供这个对象的音频表示。

[0072] 对于其他的音频对象来说,与三维图像的关联可能不是与特定图像对象的关联,而可能更间接地涉及利用3D图像所表示的场景。例如,音频可以起源于(例如,由于遮挡、距离或尺寸)在渲染的图像中不是可视的音频源。作为特定示例,第一音频对象可以提供在也利用3D图像视觉表示的景观场景中歌唱的鸟的描述。然而,鸟本身可能没有在图像中被视觉表示,例如,这是因为它将只不过是太小而以致在景观中没有被看见。

[0073] 应该认识到:术语图像对象可以指的是三维图像中的任何部分、地区、片段、区域或实体,并且这可能没有被明确识别或被表征为单独的对象。例如,除了像素值之外,3D图像可以被接收作为向用户呈现的左眼和右眼图像而没有内容的任何识别。然而,与场景中的实体或实体的一部分相对应的各种像素组可以被视为图像对象。因而,术语图像对象可以简单地被认为对应于图像的一部分(典型地,具有类似的特征或属于场景中的同一对象)。

[0074] 然而,也将认识到:在一些实施例中,3D图像中的一些或全部可以作为将被渲染的单独的个别对象来提供,并且3D图像通过这些个别对象的组合渲染来形成。这样的对象也可以被视为图像对象,但是所描述的方案显然不限于这样的实现方式。

[0075] 位置数据可以允许音频对象的渲染被定位在所期望的位置上,以致它与图像中的说话者的位置相匹配。此外,在其中视频是利用3D显示器103渲染的3D视频流的特定情况中,位置信息并不仅仅包括有关方位和高程的信息(或在显示器的平面中的位置,下面被称为 x (水平)和 y (垂直)方向)。相反,用于至少第一音频对象的位置数据包括在深度方向(也被称为与利用显示平面形成的平面垂直的 z 方向)上的位置的信息。因而,位置数据可以包括指示用于第一音频对象的目标深度位置的深度位置数据(目标深度位置可以例如作为包括 x 、 y 、 z 位置值的集合中的 z 值来提供)。

[0076] 将认识到:鉴于描述为了简洁和清晰起见而将专注于针对被称为第一音频对象的一个音频对象的处理,音频数据典型地将提供将被同时渲染的相对大量的音频对象的数据,其中音频对象包括为其提供特定位置数据的音频对象以及例如不与任何特定位置相关联的环境或背景音频对象二者。用于第一音频对象的位置数据包括此后也被称为目标深度位置的第一深度位置。这个特定地可以是定义用于第一音频对象(相对于本领域中将是公知的任何合适参考,例如使用显示平面/屏幕水平深度作为参考,尽管能够使用任何其他合适的参考)的深度值的数据。

[0077] 深度信息允许音频对象不仅被定位在合适的方位和高程上,而且也允许被定位在适当的深度上。例如,与说话的人相对应的音频可以被定位在与说话的人的图像对象到观众的距离相对应的深度上。

[0078] 然而,对3D显示器来说,恰当地定位音频源以匹配相关联的三维图像是特别具有挑战性的,并且特别地,这对于深度方向而言是具有挑战性的。实际上,鉴于与 x 和 y 方向上的给定位置相对应的用于音频对象的合适渲染位置是相对容易的,音频深度提示的人类感知明显是更复杂的。特别地,鉴于 x 、 y 平面中合适位置的确定典型地只是简单的线性解析几何的案例,来自3D显示器的深度提示的人类深度感知是更复杂的并且典型地不能利用纯粹

的线性几何分析来充分精确地确定。

[0079] 发明人已意识到：相应地，音频的渲染以匹配相关联的3D图像的三维渲染高度取决于那个显示器的特定深度渲染特征。

[0080] 更详细地，3D显示器在其渲染视觉对象的视觉距离的能力方面受到限制。在实践中，相对于其中能够有效渲染视觉对象的显示平面而言具有最大深度范围。在显示器前面（在显示器与用户之间）的深度范围在实践中可能尤其受到限制。其与显示平面相隔的距离超过这个范围的对象被感知成被扭曲或者查看起来是不舒服的。相应地，在许多3D视频系统中，典型地对于视觉场景或对象的渲染距离施加限制，以致所渲染的深度范围不超过屏幕的能力。

[0081] 视觉场景的深度渲染的这个限制典型地与对应音频对象的渲染无关来进行，其中音频对象因此被有效渲染在与对应的视觉对象不同的与用户相隔的距离上，从而导致视听空间不同步。此外，由于深度渲染能力可能在不同的显示器之间明显变化，所以不可能生成适合于所有3D显示器的源信号。

[0082] 这个问题可以利用其中视听渲染系统进一步包括用于将接收的音频对象的目标深度位置映射到取决于3D显示器103的深度渲染属性（并且特定地，深度渲染能力）的渲染位置的功能的图1的方案来减轻。因而，音频渲染位置在考虑显示器的视觉深度能力的情况下被确定。特定地，音频对象的深度位置的重新映射可以用于实现改进的感知空间同步。特定地，图1的音频处理器包括用于将用于音频对象的所接收的深度位置（例如距离参数、差距值或直接z轴值）变换或映射到取决于3D显示器103的深度渲染特征的渲染深度位置的功能。显示器的深度渲染特征将影响图像的各个部分的实际渲染深度，并且映射特定地可以变换音频对象的位置，以致其对应于对应视觉对象在3D显示器103的屏幕上所期望的感知视觉距离。

[0083] 更详细地，音频处理器包括显示属性确定器/处理器107，其被安排成确定用于呈现三维图像的目标三维显示器即在特定示例中用于3D显示器103的深度渲染属性。

[0084] 深度渲染属性特定地可以是3D显示器103的深度渲染能力的指示。例如，显示属性确定器107可以确定其中3D显示器103能够表示图像对象的深度范围。深度范围可以被提供作为其中显示器能够渲染图像对象的渲染环境中的物理距离，即作为其中3D显示器103能够生成导致观众感知对象位于预期位置上的视觉深度提示的物理范围（例如，以米为单位来测量）。

[0085] 深度渲染属性指示（或者实际上可以是）用于3D显示器的视觉渲染深度范围。视觉渲染深度范围可以指示用于3D显示器的操作深度范围。因而，视觉渲染深度范围可以描述其中显示器的属性允许图像对象的位置被显示的深度范围。视觉渲染深度范围可以是利用显示器的能力/特征/属性来支持的深度间隔。将认识到：用于选择或确定这样的范围的端点的确切标准或要求将取决于个别实施例的偏好和要求。因而，视觉渲染深度范围可以被确定（例如在制造或设计阶段期间）来反映其中图像对象的渲染将满足质量标准的可允许深度范围。对于不同的实施例和情景而言，确切标准可以是不同的。

[0086] 作为示例，在一些实施例中，视觉渲染深度范围可以在制造或设计处理期间由工程师手动确定。由此产生的深度范围可以被硬编码（即，被存储）在3D显示器中。在这样的情景中，确定器107可以通过与3D显示器通信以便接收为特定显示器定义视觉渲染深度范围

的数据来确定视觉渲染深度范围。视觉渲染深度范围因而可以反映3D显示器的固有属性，并且特定地可以反映3D显示器渲染深度的能力。视觉渲染深度范围因而典型地将是显示器的包括例如屏幕尺寸、分辨率、显示器的光学属性、显示器的几何属性等等的许多不同的参数和属性的结果。

[0087] 在许多实施例中，显示属性确定器107可以被耦合到显示驱动器105(和/或3D显示器103)并且可以直接接收指示显示器的渲染能力的的数据。这样的数据例如可以在3D显示器103的设计阶段期间被确定并且在制造阶段期间被包括在显示驱动器105(或3D显示器103)中。在其他的实施例中，显示属性确定器107可以接收允许视觉渲染深度范围被确定的间接指示。例如，可以给显示属性确定器107提供指示影响显示器的适当深度范围的属性的信息，例如，诸如有关分辨率、例如透镜状屏幕的几何属性、显示尺寸等等的信息。

[0088] 作为特定示例，影响显示器的深度渲染能力的显示属性是显示器的标称观看距离，其主要涉及显示尺寸和像素分辨率(但是整体取决于这二者)。在观看立体图像时，深度的错觉通过在旨在用于左眼和右眼的图像之间引入差距而获得，即用于左眼和右眼的图像被投射在显示器上横向位移的位置上。对于在显示器后面的视觉目标而言，差距为正，即用于左眼的图像被投射到用于右眼的图像的左侧。对于显示平面上的视觉目标而言，差距为零，并且两个图像重合。对于在显示器前面的视觉目标而言，差距为负，即用于左眼的图像被显示在用于右眼的图像的右侧，并因此用户不得不“交叉(cross)”其眼睛。如果这个负差距在眼睛的观看方向之间的角度方面变得太大的话，这引入不适和疲劳。对于具有较小标称观看距离的显示器来说，与具有较大标称观看距离的显示器相比而言，在屏幕前面对于较小的绝对目标距离将达到可容忍的负差距的这个极限。

[0089] 此外，所采用的立体显示技术的类型可以影响能够舒适地和/或利用足够的质量(即没有太强的模糊或“重影”伪像)来渲染的最大深度范围。例如，用于自动立体透镜状显示器的可用深度范围可能不同于用于具有主动快门式眼镜的立体显示器的可用深度范围，其中具有主动快门式眼镜的立体显示器又可能具有与具有被动(例如偏振)眼镜的立体显示器不同的可用深度范围。

[0090] 显示属性确定器107被耦合至映射器109，而映射器109进一步被耦合到它从中接收一个或典型多个音频对象的音频数据和目标位置数据的接收器101。映射器109被安排成执行所接收的音频对象的目标深度位置到渲染深度位置的映射。因而，映射器109可以将第一音频对象的z值改变为随后可以用于第一音频对象的渲染的不同值。映射取决于从显示属性确定器107接收的视觉渲染深度范围，并因而映射取决于3D显示器103的深度渲染特征。

[0091] 特定地，映射可以是这样的，以致由于映射而导致的深度位置当这个被呈现在3D显示器103上时将对应于3D图像中的对应图像对象的深度位置。这些位置可以在渲染空间中即在实际物理空间中进行匹配。

[0092] 在一些实施例中，由映射器109执行的映射可以被生成来对应于在用于3D图像的的深度值与当被渲染3D显示器103上时感知的深度距离(例如，相对于屏幕水平)之间的映射。特定地，深度渲染属性在一些实施例中可以反映在图像对象值与当图像对象利用3D显示器103来渲染时对应感知的深度位置之间的映射。由映射器109执行的映射随后可以执行所接收的音频对象深度位置与渲染深度位置的对应映射。

[0093] 例如,3D显示器103可以具有在显示器前面延长假设2米的深度渲染范围。将被呈现的图像可以利用具有在从假设0到255的范围中的深度位置值的深度映射并且利用与128的值相对应的显示平面深度来接收。在这种情况下,从0到128的范围可以相应地被映射到从显示平面深度/屏幕水平到屏幕前面两米的范围。因而,具有64的深度值的图像对象可以例如被渲染在屏幕前面一米的位置上。从屏幕水平到屏幕前面两米的渲染范围的指示可以被提供给映射器109,而映射器109可以继而执行音频源的对应映射。例如,如果在显示器前面的音频对象的目标深度位置(z 值)利用假设0-5米的范围中的值来给定,映射器109可以将0-5米中的值映射到在屏幕前面从0到2米的范围。在这样的示例中,某种深度压缩和限制可以由映射器109引入至音频对象。例如,在从0到2.5米的范围中针对音频对象所接收的所有 z 值(相对于屏幕水平)可以利用映射器109来乘以0.8,从而导致其被映射到0-2米的范围。在屏幕前面大于2.5米的所有音频对象可以被限于2米的距离。因而,0-5米的范围非线性地被映射到在屏幕前面0-2米的范围。

[0094] 映射器109在图1的示例中被耦合至音频驱动器111,而音频驱动器被安排成为一组扬声器113生成驱动信号,从而允许所接收的音频对象与3D视频的渲染一起被渲染在3D显示器103上。音频驱动器111因而形成被安排成渲染将从与通过映射所接收的用于第一音频对象的目标深度位置来确定的渲染深度位置相对应的位置中感知的第一音频对象的渲染器。

[0095] 音频驱动器111包括空间音频处理器并被安排成生成合适的驱动信号来引起音频被渲染,以致其被感知成起源于所希望的位置。确切的功能和扬声器配置将取决于个别实施例的偏好和要求,并且用于空间音频渲染的许多不同的技术和方案对于技术人员来说是已知的。

[0096] 作为示例,扬声器113可以形成典型的五或七扬声器环绕声设置,并且音频驱动器111可以通过使用平移(panning)算法(例如通过应用VBAP)确定对于每一个环绕声通道的信号值贡献而将音频对象定位在特定位置(对应于重新映射的位置)上。用于空间渲染的合适技术的其他示例可以包括例如Wave Field Synthesis(波场合成)、反耳(transaural)音频或高保真度立体声响复制(ambisonics)。

[0097] 图1的系统可以提供改进的用户体验并且特定地可以针对3D视听呈现来提供改进的音频与视频之间的空间同步。

[0098] 在一些实施例中,音频和视频处理可以是分开的并且个别地来执行。例如,在一些实施例中,音频信号可以是不包括与视听内容项中的视觉内容相关的任何信息的单独的音频数据流。实际上,在许多情况中,可以根据只涉及音频内容且不包括视频数据的音频格式或标准来提供音频。

[0099] 典型地,在包括一个或多个基本流的传输流中提供音频视觉内容,其中每一个基本流典型地可以是视频流或音频流。实际上,典型地每一个基本流包括用于仅仅一种类型的音频视觉内容(即,或视频或音频)的数据,并且典型地基本流能够利用单个音频或视频解码处理来解码。因而根据允许其被本身解码且不考虑传输流中的任何其他基本流的格式或标准来生成每一个基本流。特定地,音频基本流能够利用单个音频解码器来解码而不考虑任何其他的基本流。类似地,视频基本流能够利用单个视频解码器来解码而不考虑任何其他的基本流。

[0100] 因而,在许多实际的系统中,每一个基本流是单个内容类型、单独可解码的音频流。每一个基本流此外根据典型地是标准化格式的特定音频或视频编码格式来编码。因而,每一个基本流个别地利用单个标准化格式解码器来处理。

[0101] 使用这样的安排格式的内容分发方案的示例是广泛且流行的MPEG标准。

[0102] 在一些实施例中,可以采用所描述的包括多个基本流的传输流的格式来接收内容。除了下面描述的差异之外,在图2中举例说明这样的与图1的系统相对应的实施例的示例。

[0103] 在图2的系统中,传输流接收器201接收包括至少一个基本视频流和一个基本音频流的传输流。基本视频流被馈送到显示驱动器105,其在显示驱动器中利用合适的视频解码器来解码,并且响应于所解码的数据,3D显示器103被驱动来呈现3D图像。基本视频流的解码和处理完全独立于基本音频流的解码、处理和渲染。

[0104] 基本音频流被馈送至在示例中是音频解码器的音频处理器。音频处理器相应地接收音频数据流。音频数据流在它包括渲染利用音频流所表示的音频分量所需的所有数据的意义上而言是完整的。音频流此外不包括渲染(多个)对应3D图像/视频所必需的任何数据。因而,虽然音频流中的音频与在3D显示器103上呈现的3D视频相关联,但是它并不包含表示这个视频的任何视频数据。

[0105] 提供给音频处理器的音频流包括至少三种不同类型的数据。

[0106] 首先,它包括与三维图像相关联的一个但典型地明显更多的音频对象的音频数据。因而,这些音频对象旨在与3D视频的渲染一起进行渲染并提供利用该视频所表示的场景区的音频的描述。

[0107] 音频对象不是针对特定的预定或参考位置提供的音频通道分量(例如左通道、右通道和中央通道),但是反而典型地对应于场景中的个别声源。特别地,这些音频对象之中的一个或多个可以直接涉及3D图像中的一个图像对象(例如,语音音频对象可以与表示说话者的图像对象相关联)。

[0108] 相应地,音频流其次可以包括指示音频对象之中的一个或多个的目标深度位置的深度位置数据。用于音频对象的目标深度位置涉及沿着与3D显示器103的屏幕/显示平面垂直的轴(典型地,被称为z轴)的位置。将认识到:音频流可以并且典型地也将包括屏幕平面(即沿着x和y方向)中的位置值。

[0109] 用于给定音频对象的位置数据因而可以指示利用音频对象所表示的声源在场景中的位置。

[0110] 将认识到:位置数据不需要针对所有的音频对象而被包括。特别地,一些音频对象可能不对应于具有特定位置的声源,而是可以具有扩散或传播源位置。例如,一些音频对象可以对应于非旨在从特定位置中进行渲染的环境或背景声音。

[0111] 第三,音频流包括具有用于音频与之相关联的3D图像的呈现的目标三维显示器的深度渲染属性的指示的数据。在特定示例中,音频流相应地包括表示3D显示器103的深度渲染属性以及特定地视觉渲染深度范围的数据。深度渲染属性因而指示如前所述的用于显示器的视觉渲染深度范围。

[0112] 因而,在信号中,接收音频信号/流,其包括:

[0113] - 用于与三维图像相关联的至少第一音频对象的音频数据;

[0114] - 指示第一音频对象的目标深度位置的深度位置数据;和

[0115] - 包括用于呈现三维图像的目标三维显示器的深度渲染属性的显示数据。

[0116] 如图3所示,图2的音频处理器可以相应地包括接收音频流的接收器101,其中音频流包括用于音频对象的音频数据、用于音频对象的位置数据以及指示用于目标显示器的深度渲染属性的显示数据。

[0117] 深度渲染属性由显示属性确定器107从显示数据中进行确定并被馈送至映射器109,而映射器109也接收音频对象数据和位置数据。映射器109随后基于深度渲染属性并且特定地基于视觉渲染深度范围来执行深度位置的映射。

[0118] 在图2的系统中,深度渲染属性相应地不利用音频处理器与显示器或视频功能之间的直接交互来提供。相反,所需要的针对音频处理器的唯一输入是音频流本身。音频流可以根据合适的音频格式来提供并因而能够是根据音频格式操作的通用或标准音频处理器或解码器。相同的通用音频处理器相应地能够被使用在各种各样不同的应用和系统中而不要求其针对个别的实现方式进行定制或特定地进行适配或设计。

[0119] 指示深度渲染属性的数据在一些实施例中可以由传输流接收器201插入到音频流中。特定地,显示驱动器105 可以将描述3D显示器103的深度渲染特征的相关数据馈送至传输流接收器201,并且传输流接收器201可以将这个数据插入到音频流的恰当字段中。在其他实施例中,深度渲染属性可以在传输流接收器201(或任何其他处理器)没有修改音频信号的情况下直接经由单独的本地接口被提供至音频处理器。因而,在许多实施例中,所接收的音频流(例如MPEG-H基本流)可以被直接馈送至音频处理器。然而,在这样的实施例中,本地接口可以例如使用与MPEG-H相对应的格式来提供深度渲染属性的信息。

[0120] 在图2的系统中,显示属性确定器107被安排成从音频流中提取显示数据并且响应于显示数据来确定深度渲染属性。例如,它可以提取描述用于目标显示器(其在特定示例中是3D显示器103)的视觉渲染深度范围的音频流的相关数据字段并将这个馈送至映射器109。

[0121] 映射器109随后使用这个信息来确定从输入深度位置到输出深度位置的映射并且特定地使用视觉渲染深度范围来确定用于从输入深度位置中计算输出深度位置的函数。因而,映射器109被安排成响应于视觉渲染深度范围来适配映射。所确定的映射随后被应用于所接收的用于音频对象的目标深度位置,以便确定修改的深度位置。将认识到:在一些实施例中,所接收的目标深度位置可以在映射之前(例如由用户)进行修改。类似地,在一些实施例中,在渲染之前可以(例如由用户)修改由此产生的渲染位置。

[0122] 将认识到:映射可以考虑其他参数并且特定地在一些情景中也可以取决于例如屏幕平面中的位置(并且特定地取决于与音频对象相对应的图像对象的方位或高程位置)。

[0123] 在许多实施例中,映射可以只被应用于所接收的音频对象的子集,并且特定地,音频流可以包括描述是否个别音频对象应该被重新映射的信息。重新映射典型地可以限于与场景中的对象相关的音频对象并且典型地限于3D图像中的图像对象。

[0124] 在许多实施例中,音频流可以针对每一个音频对象包括标志,其指示是否它涉及三维图像中的视觉对象。如果是这样的话,用于该音频对象的深度位置将被重新映射,而否则原始深度位置将被使用。

[0125] 在一些实施例中,可以仅针对音频对象之中的一个或一些来提供标志,但是在其

他实施例中,可以针对所有的音频对象来提供标志。用于音频对象的标志可以指示是否该音频对象是屏幕相关对象。屏幕相关音频对象是对其而言渲染处理取决于三维显示器的属性的音频对象。因而,信号的音频对象可以被分类或被指定为或屏幕相关音频对象或非屏幕相关音频对象。对于被指定为屏幕相关的音频对象,音频渲染被适配以响应3D显示器在渲染3D图像时的一个或多个属性。因而,对于这些对象来说,音频渲染基于显示器的深度渲染属性来适配,并且特定地基于视觉渲染深度范围的映射被应用于这样的对象。然而,映射不被应用于被指定为非屏幕相关音频对象的音频对象。

[0126] 因而,映射器109可以被安排成响应于标志来执行映射,并且特定地,它可以被安排成仅在标志指示这个音频对象实际上是屏幕相关音频对象时才将映射应用于该音频对象。因而,如果标志指示音频对象是屏幕相关对象,则位置在音频渲染之前被映射,否则可以直接使用该位置而没有任何映射被应用。

[0127] 标志在许多实施例中本质上可以指示是否音频对象与3D图像中可视的某物相关联。设置这个标志可以被用于允许任何屏幕相关处理被应用于音频对象的最高级先决条件,其包括基于显示器的尺寸的2D重新映射、针对缩放的重新映射或者所描述的基于显示器的深度渲染属性的深度重新映射。

[0128] 标志将被设置在内容创建侧上并且是所发射的比特流的一部分以及特定地所接收的基本音频流的一部分。

[0129] 例如,在MPEG-H中,单个标志(“bsIsScreenRelativeObject”)能够被使用来区分针对各种类型的屏幕相关处理而言是候选者的对象和对其而言不应该应用这样的处理的对象。被应用于为其设置标志的音频对象的屏幕相关处理的特定类型可以取决于用于特定个别处理的相关元数据的存在。特定地,如果有关本地显示器的深度渲染属性的信息是可用的,则可以应用所描述的深度重新映射。

[0130] 作为特定示例,音频流可以是被修改成包括深度渲染属性信息的MPEG-H 3D音频流。

[0131] 因而,特定地,在一些实施例中,该方案可以通过在MPEG-H 3D Audio中在现有机制上针对所谓的“屏幕相关”对象的方位和高程数据的重新映射进行构建来实现。特定地,该方案可以重新使用一些已存在MPEG-H 3D Audio元数据项来实现。作为特定示例:

[0132] - 描述是否音频对象是用于重新映射的候选者的信息可以被实现为MPEG-H 3D Audio Audio Object Metadata (OAM) 中的现有标志“isScreenRelativeObject”

[0133] - 描述深度渲染属性的信息可以被实现为现有MPEG-H 3D Audio本地设置信息元素“LocalScreenSizeInformation()”的扩展,具有例如新字段:

[0134] bsLocalScreenDepth,其提供本地屏幕深度范围的总体指示,例如以米为单位,或

[0135] bsLocalScreenDepthFront和/或bsLocalScreenDepthBack,用于单独规定在本地屏幕前面和在本地屏幕后面的视觉对象的深度范围。

[0136] 此外,在这样的基于MPEG-H 3D Audio上下文的示例中,利用映射函数所变换的深度位置(音频对象距离参数)可以对应于现有音频对象位置数据字段position_radius[]。

[0137] 将认识到:在其中信息被包括在音频流中的实施例以及在其中经由可供选择的直接本地接口来提供信息的实施例中,深度渲染属性的描述都可以遵循这个格式。

[0138] 在一些实施例中,音频流也可以包括用于三维图像的第二显示器的深度渲染属性

的指示,其中第二显示器将被称为参考显示器。参考显示器例如可以是用于3D图像/视频的制作的特定显示器或者可以例如是例如可以被期望用于3D图像/视频的渲染的标称或标准显示器。

[0139] 显示属性确定器107可以提取这个参考渲染属性并且随后据此来适配映射。

[0140] 因而,在一些实施例中,映射可以取决于3D显示器103的深度渲染属性以及所接收的3D图像的参考深度渲染属性二者。在其他实施例中,映射可以只考虑3D显示器103的深度渲染属性。这样的方案可能特别适合于其中原始制作屏幕是2D屏幕的情景。

[0141] 特别地,音频流可以包括在内容的制作期间使用、例如或明确地以信号通知(signal)或作为以信号通知默认参考距离/深度范围的标志使用的屏幕的距离/深度范围的可选指示。可选指示例如可以是制作屏幕深度范围的总体指示,例如以米为单位,或者可以单独规定在制作屏幕前面和/或后面的视觉对象的深度范围。

[0142] 作为特定示例,该指示可以被实现为现有MPEG-H 3D Audio元数据音频元素“mae_ProductionScreenSizeData()”的扩展,具有例如新字段:

[0143] hasNonStandardScreenDepth,以便以信号通知:制作显示器深度与默认参考显示器深度的差异被以信号通知,

[0144] bsScreenDepth,其提供制作显示器深度范围的总体指示,例如以米为单位,或

[0145] bsScreenDepthFront和/或bsScreenDepthBack,用于单独规定在制作显示器前面和后面的视觉对象的深度范围。

[0146] 因而,在许多实施例中,可以使用包括以下的音频信号:

[0147] - 表示用于三维图像的图像对象的音频的至少第一音频对象的音频数据,

[0148] - 指示用于第一音频对象的目标深度位置的深度位置数据,

[0149] - 指示是否至少一个音频对象的位置对应于三维图像中的可视对象的至少一个音频对象(可能地,第一音频对象)的标志;以及

[0150] - 包括用于三维图像的参考三维显示器的深度渲染属性的显示数据,深度渲染属性指示三维显示器的视觉渲染深度范围。

[0151] 例如,MPEG-H音频比特流可以包括指示是否对象是屏幕相关的标志以及描述参考屏幕的元数据。这些特性可以是在内容创建侧上定义的元数据元素。

[0152] 然而,鉴于有可能音频流也包括例如由传输流接收器201插入的描述目标三维显示器(3D显示器103)的元数据,这个信息典型地在许多实施例中被单独且外部提供给音频处理器而不是作为音频流的一部分。特定地,如图1所示,该数据可以从渲染系统的其他元素中直接提供而不是被包括在音频流中。

[0153] 因而,描述本地设置的元数据例如有关本地屏幕尺寸、用户定义的缩放区域、显示深度渲染属性等等的信息可以经由单独的本地接口而被直接馈送至音频处理(其特定地可以是MPEG-H解码器)。用于本地接口以及在这个本地接口上使用的元数据元素的语法然而可以遵循标准化方案,因而促进在不同实体之间的互通。

[0154] 将认识到:所使用的特定映射以及这个取决于深度渲染属性的方式在不同的实施例之间可以变化,以便反映个别应用的特定偏好和要求。

[0155] 在许多实施例中,映射可以是用于参考显示器的渲染深度范围到用于目标三维显示器的渲染深度范围的单调映射。实际上,映射可以被适配成取决于参考显示器深度范围

与3D显示器103深度范围之间即以信号通知的制作显示器深度范围与本地屏幕深度范围之间的比率的线性函数。

[0156] 例如,如果音频流包括指示3D图像是使用具有高达3米的在屏幕前面的深度范围的显示器来生成的数据,鉴于深度渲染属性指示用于3D显示器103的深度范围仅仅是假设1.5米,映射可以被调整成对应于从z位置到屏幕的距离与 $1.5/3=0.5$ 的因子的乘法。如果相反深度渲染属性指示3D显示器103的深度范围是假设2米,则映射因子将是 $2/3=0.67$ 。

[0157] 因而,在许多实施例中,深度渲染属性可以是能够由3D显示器103提供的深度渲染范围的指示。这可以提供特别有利的用户体验并且特定地在许多情景中可以在所呈现的音频与视频之间提供改进的空间同步。

[0158] 作为示例,映射可以包括限制音频对象深度值,以致音频对象将总是在本地屏幕的视觉深度范围的极限内进行渲染,其中具有在这个范围之外的原始深度位置的音频对象被渲染在这个范围的极限上。

[0159] 在许多实施例中,映射有利地可以围绕显示(器)深度而是不对称的。映射在一些实施例中对于在显示器前面的位置和对于在显示器后面的位置而言可以是不同的,即,在屏幕前面和在屏幕后面的映射功能可以是不同的。

[0160] 发明人已意识到:不对称的映射可以提供改进的用户体验和感知的更紧密的空间同步。这可以反映:在屏幕前面的深度范围典型地是明显更加感知敏感的,并且当对象在屏幕前面时,与当对象在屏幕后面时相比而言,显示器典型地在没有感知降级的情况下在渲染大的深度偏差时不是那么有效的。

[0161] 在许多实施例中,深度渲染属性可以是深度渲染范围,并且这对于在屏幕前面的对象和对于在屏幕后面的对象来说可能是不同的。例如,深度渲染属性可以指示:在屏幕前面的可用深度渲染范围是假设3米,而在屏幕后面的可用深度渲染范围是假设20米。

[0162] 映射随后可以为在屏幕前面的对象确定与在屏幕后面的对象不同的映射。例如,对于在屏幕前面和在屏幕后面的对象来说,可以确定个别映射因子。

[0163] 在许多实施例中,映射可以被安排成只修改在显示器前面的深度值,而不修改在显示器/屏幕平面后面的深度值。因而,在一些实施例中,映射可以只改变在屏幕前面的对象的深度值。这可以对应于被应用于3D视频图像的对应的不对称深度映射。将这样的不对称映射应用于3D视频图像的原因可能是:因为在显示器前面的图像对象的负差距属性(迫使眼睛“交叉”),所以视觉不适和疲劳的非预期效果主要针对在显示器前面的图像对象发生。

[0164] 然而,不对称映射也可以仅被应用于音频对象。这将允许音频对象与在屏幕前面的对应图像对象之间的完全空间同步,其中在保持使得“遥远的”音频对象在屏幕后面的可能性的同时,即使对应图像对象正被重新映射成更靠近显示面板,这样的同步也是最关键的。

[0165] 在许多实施例中,映射可以是非线性映射。

[0166] 例如,具有的目标距离超过显示器的物理深度范围的音频对象可以与位于显示深度范围内的音频对象不同地来对待。例如,前一类别的对象的深度位置可以被修剪(clip)至显示深度范围的极限,而在显示深度范围内的音频对象的深度位置保持不变。作为另一示例,非线性压缩曲线可以采用进一步远离显示器的深度位置(或在其前面或在其后面)与

更靠近显示器的对象的深度位置相比而言朝向显示平面被更多压缩的方式而被应用于深度位置。这样的非线性压缩曲线的示例可以是对于靠近显示器的音频对象的深度位置不应用压缩并且使得深度位置对于在与屏幕相隔不断增加的距离上的音频对象而言朝着显示深度范围的极限而渐近增加的曲线。

[0167] 再次,这个非线性映射可以对应于针对3D图像所应用的对应非线性映射。

[0168] 在一些实施例中,映射是感知加权的映射。特定地,在一些实施例中,映射可以基于人类深度视野的感知模型来确定。该模型可以反映观众的深度感知将如何取决于3D显示器103的深度渲染属性。这样的方案可以特别利用更紧密感知的空间同步来提供改进的用户体验。

[0169] 许多3D显示器诸如特定地3D电视包括用于从接收的2D视频中生成3D视频内容的功能。实际上,将认识到:对于执行2D-3D视频转换处理而言,许多不同的算法和技术对于技术人员来说将是已知的。例如,考虑不同图像对象的相对运动和视差的技术是已知的。

[0170] 在一些实施例中,系统可以包括用于从二维图像中生成三维图像、即用于执行2D-3D转换的转换器。转换包括生成包括用于例如不同的图像对象或像素的深度值的深度映射。实际上,转换可以使用现有算法,其作为转换的一部分来生成反映估计的深度值的深度映射。

[0171] 在一些实施例中,显示属性确定器107可以被安排成响应于这样的深度映射的值来确定深度渲染属性。作为特定示例,显示属性确定器107可以将深度渲染属性确定为深度映射的深度范围,即它可以将这个确定为从最高值到最低值的范围。深度值在一些实施例中可以直接被用作相对值或可以利用考虑显示器的已知物理渲染特征的转换而被转换成特定物理距离。

[0172] 深度映射信息例如可以被使用来将映射的极限值设置成对应于深度映射的极值。

[0173] 这个方案可以提供其中在重新使用已经正在被用于其他目的的某功能的同时能够实现改进的空间同步的有效系统。特别地,该方案可以利用:显示器中深度映射的本地生成典型地考虑显示器的深度渲染属性,以及由此产生的深度映射的考虑相应地可以提供控制声音对象的z位置的映射的有效方式。

[0174] 在一些实施例中,视听渲染系统包括用于3D图像的功能,或者更一般地,这样的3D场景(假使有关这个超过利用3D图像所表示的信息是可用的)被呈现成被缩放即不是必然呈现整个3D图像,图像的子集可以被选择并被呈现在3D显示器103上。因而,只有图像的子集当被呈现在3D显示器103上时才被选择成可视的。

[0175] 图4举例说明这样的系统的示例。该示例对应于图2的系统,但是附加地包括被安排成选择用于在3D显示器103的完全显示区域上呈现的(多个)3D图像的子集的选择器401。选择器401特定地可以被安排成选择该子集,以响应例如可以经由用户接口403提供的用户输入。例如,用户接口403可以与用户接口,以允许用户选择所呈现的完全3D图像的预定数量的子区域或卷(volume)之中的一个。

[0176] 选择器401被耦合到显示驱动器105,其被馈送子集选择并且在响应中选择3D图像的对应部分以及将这个呈现在3D显示器103的完全可用屏幕区域上。

[0177] 选择器401进一步接收3D图像的深度信息(例如,诸如深度映射)并且被安排成确定被选择用于呈现的3D图像的特定子集的深度属性。特定地,它可以表示被特定选择的子

集跨越 (span) 的深度范围。例如,如果利用在 $[0, 1023]$ 范围中的 z 值来给定总范围,但是在所选择的子集内只有 z 值位于假设 $[612, 784]$ 的范围中,深度属性可以被确定为这个范围。该范围随后可以由映射器109使用3d显示器103的深度渲染范围的信息转换成实际物理范围。

[0178] 在该系统中,映射器109被安排成响应于子集的深度属性来适配映射。例如,它可以适配音频对象位置映射,以致用于所选择的子集的渲染范围被映射到用于3D显示器103的整个可用深度渲染范围。这样的方案因而可以反映:图像的缩放不仅影响对象在 x 和 y 方向上的大小,而且影响图像对象的深度位置。该方案相应地可以提供即使图像缩放被支持也能够更好保持的改进的空间同步性。

[0179] 该方案在越来越重要和感兴趣的许多应用中可以提供明显改进的性能。例如,用户交互性在许多虚拟现实和增强现实多媒体应用中是越来越重要的元素。在这样的应用中,用户能够在虚拟或增强视听场景中导航,并且也能够主动操纵场景。除了用户角度的旋转和平移之外,这样的操纵也可以包括缩放到特定感兴趣区域。

[0180] 这样的缩放交互性的相关示例当前能够例如在 MPEG-DASH 的“Spatial Relationship Descriptor (SRD)”特性中并且也在 MPEG-H 3D Audio (3DA) 中找到。

[0181] 在 SRD 中,高分辨率视觉场景(例如足球比赛的镜头)由大量的瓷砖(tile)组成,并且用户能够缩放至场景的一部分(例如球当前所在的场地的部分)。在这种情况下,所选择的缩放区域(瓷砖)被“分解”来填充整个屏幕。在 SRD 中,也具有规定视觉场景(瓷砖)的空间部分与音频对象之间的空间关系的可能性。当用户缩放至于选择区域时,这些空间关系可以被使用来适配音频再现的方位和高程,以反映缩放的透视图 (perspective)。

[0182] 在 3DA 中,提供使得用户能够缩放至包含涉及用户在屏幕上看见什么的音频对象的视听场景的类似机制。通过选择缩放区域,选择的视频区域被扩展来填充整个屏幕,并且在缩放区域中包含的音频对象的位置被重新映射来也填充屏幕区域,即 x 和 y 位置可以被改变。

[0183] 然而,这些方案基于低复杂度 2D 应用并且针对低复杂度 2D 应用来开发,而且不考虑深度缩放。深度缩放是特别困难的,这是因为其(与只取决于显示尺寸的传统 2D 缩放相比而言)高度取决于个别 3D 显示器的特定特征。图 4 的系统然而可以支持这样的视觉深度缩放,同时仍保持高度的空间同步性。

[0184] 实际上,图 1 的映射器 109 被安排成:针对指示用于具有给定显示尺寸的三维显示器的第一视觉深度渲染范围的视觉渲染深度范围以及针对指示同一显示尺寸的第二不同的视觉深度渲染范围的视觉渲染深度范围,不同地映射位置。图 1 的映射器 109 也可以被安排成对于为不同的显示尺寸提供的视觉渲染深度范围的相同指示应用相同的映射。

[0185] 因而,该方案不仅仅是音频的显示尺寸相关的调整,而在渲染深度时提供映射针对显示器的特定能力的适配。这样的渲染能力可以部分取决于显示尺寸,但是可以进一步取决于许多其他的参数和属性。因而,不同尺寸的显示器可以具有相同的渲染深度能力,或者相同尺寸的显示器可以具有不同的深度渲染能力。这可以通过映射器 109 的映射来反映,并因而能够实现渲染的音频针对特定显示特征的改进适配。

[0186] 在许多实施例中,映射器 109 可以被安排成将映射适配成在选择子集中表示的深度范围到用于目标三维显示器即用于 3D 显示器 103 的渲染深度范围的单调映射。如前所述,

深度范围可以由选择器401来确定,并且映射器可以将这个与3D显示器103的深度渲染属性相关而且修改音频对象的映射,以对应于用于该子集的深度范围到3D显示器103的整个深度范围的映射。将认识到:在一些实施例中,映射只可以被应用于(3D显示器103的子集和渲染范围二者的)范围的子集。例如,只有在屏幕前面的对象的值可以被映射。

[0187] 映射特定地可以被应用于与选择的子集内的位置相对应的音频对象。因而,与选择的视觉子集相关联的音频源可以被重新映射到在缩放被应用时反映这些位置的渲染位置的位置。特定地,如果音频对象被链接到在选择的子集内的特定图像对象,音频对象可以被映射到与这个图像对象在缩放视图被渲染时的位置相对应的位置。

[0188] 然而,在许多实施例中,映射可供选择地或附加地可以被应用于具有的位置在选择的子集之外的音频对象。因而,映射不但被应用于与可视选择相对应的音频源,而且也被应用于其他的音频对象。

[0189] 在这样的实施例中,位于缩放的子集之外或没有被指示成视频相关的音频对象可以被渲染,以致其渲染位置相对于缩放的子集来适配。例如,在未缩放的场景中在用户的角度与选择的子集之间的对象在缩放的渲染中可以被渲染成在用户的后面。

[0190] 在一些实施例中,映射可供选择地或附加地可以被应用于没有与视频中的特定属性相链接的音频对象,例如,诸如环境或背景声音。然而,在许多实施例中,音频环境组件(混响、人群噪声等等)将在未被适配的情况下被渲染,即与未缩放的场景中相同,有可能在用户也旋转缩放的区域时才具有旋转。

[0191] 将认识到:上面的描述为了简洁起见而参考不同的功能电路、单元和处理器描述了本发明的实施例。然而,可以使用不同的功能电路、单元或处理器之间任何合适的功能分布而不偏离本发明,这将是显而易见的。例如,被举例说明成由单独的处理器或控制器执行的功能可以由相同的处理器或控制器来执行。因此,对于特定的功能单元或电路的引用只被视为对于用于提供所述功能的合适装置的引用而不是指示严格的逻辑或物理结构或组织。

[0192] 本发明能够采用任何合适的包括硬件、软件、固件或这些的任何组合的形式来实现。本发明可选地可以至少部分被实现为在一个或多个数据处理器和/或数字信号处理器上运行的计算机软件。本发明的实施例的元素和组件可以物理地、功能地和逻辑地采用任何合适的方式来实现。实际上,功能可以在单个单元中、在多个单元中或作为其他功能单元的一部分来实现。因此,本发明可以被实现在单个单元中或可以物理地和功能地被分布在不同的单元、电路和处理器之间。

[0193] 虽然结合一些实施例描述了本发明,但是本发明并不旨在限于本文阐述的特定形式。相反,本发明的范畴仅利用伴随权利要求书来限制。另外,虽然特性可能看起来结合特别实施例进行描述了,但是本领域技术人员将认出:所述实施例的各种特性根据本发明可以进行组合。在权利要求书中,术语包括并不排除其他元素或步骤的存在。

[0194] 此外,虽然被个别列出,但是多个装置、元素、电路或方法步骤可以利用例如单个电路、单元或处理器来实现。另外,虽然个别特性可以被包括在不同的权利要求中,但是这些也有可能可以有利地进行组合,并且在不同权利要求中的包含并不暗示:特性的组合不是可行的和/或有利的。特性在一种类别的权利要求中的包含并不暗示针对这个类别的限制,而是指示该特性同样酌情可应用于其他的权利要求类别。此外,特性在权利要求书中的

顺序并不暗示这些特性必须按照其来工作的任何特定顺序,并且特别地,个别步骤在方法权利要求中的顺序并不暗示:必须按照这个顺序来执行这些步骤。相反,这些步骤可以按照任何合适的顺序来执行。此外,单数引用并不排除复数。因而,对于“一”、“一个”、“第一”、“第二”等等的引用并不排除多个。权利要求书中的参考符号仅仅作为澄清示例来提供而不应被解释为以任何方式来限制权利要求书的范畴。

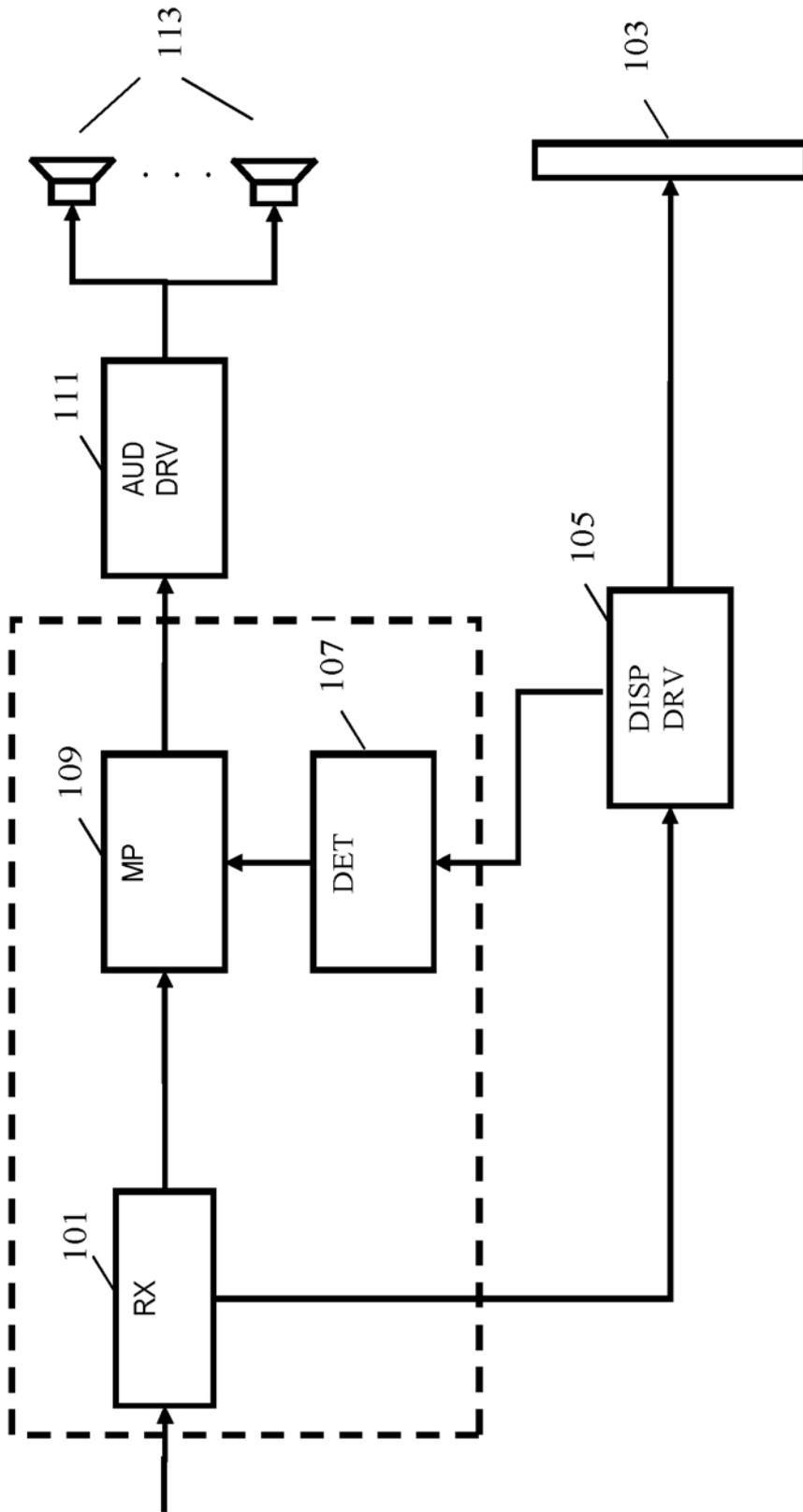


图 1

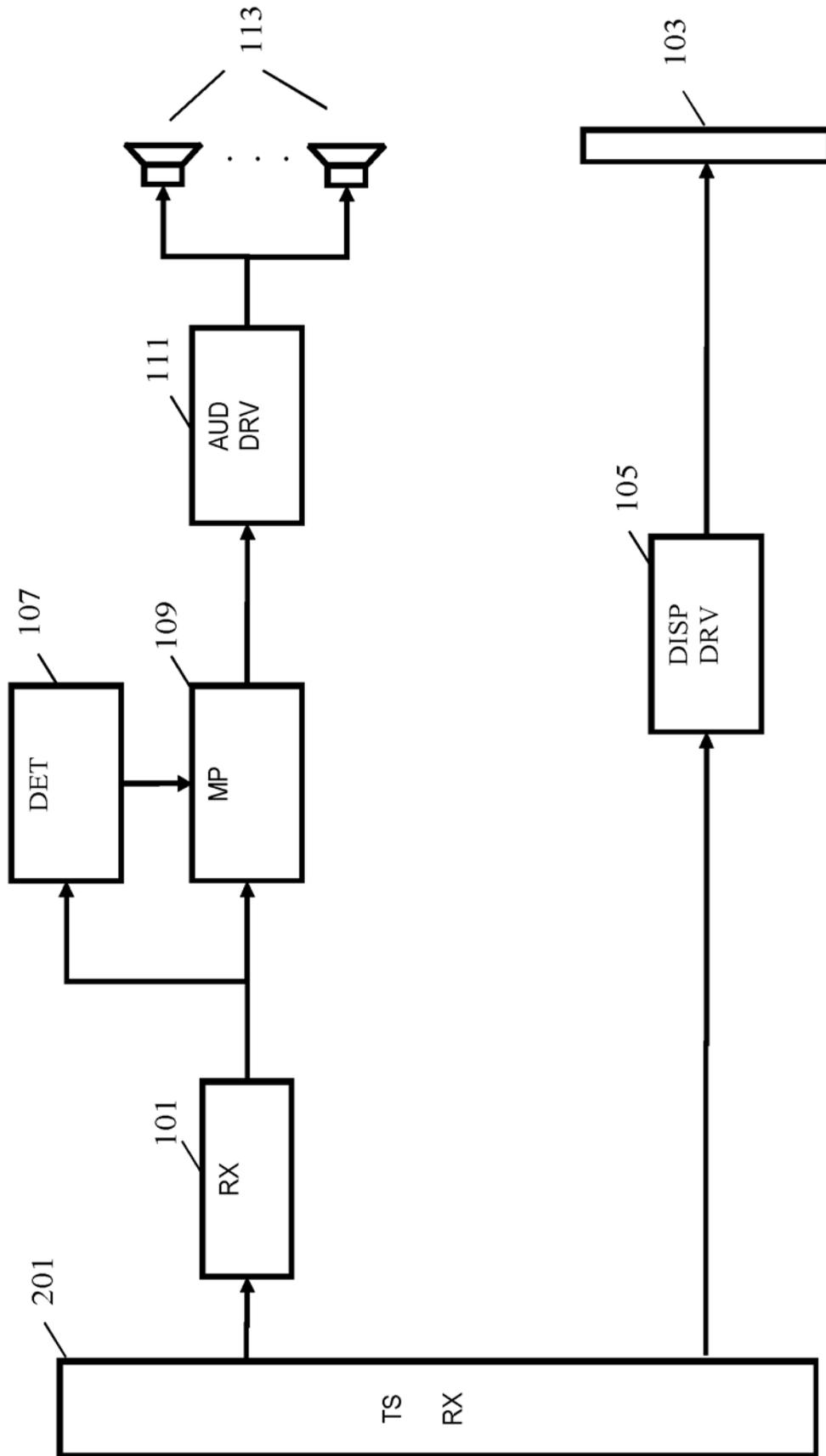


图 2

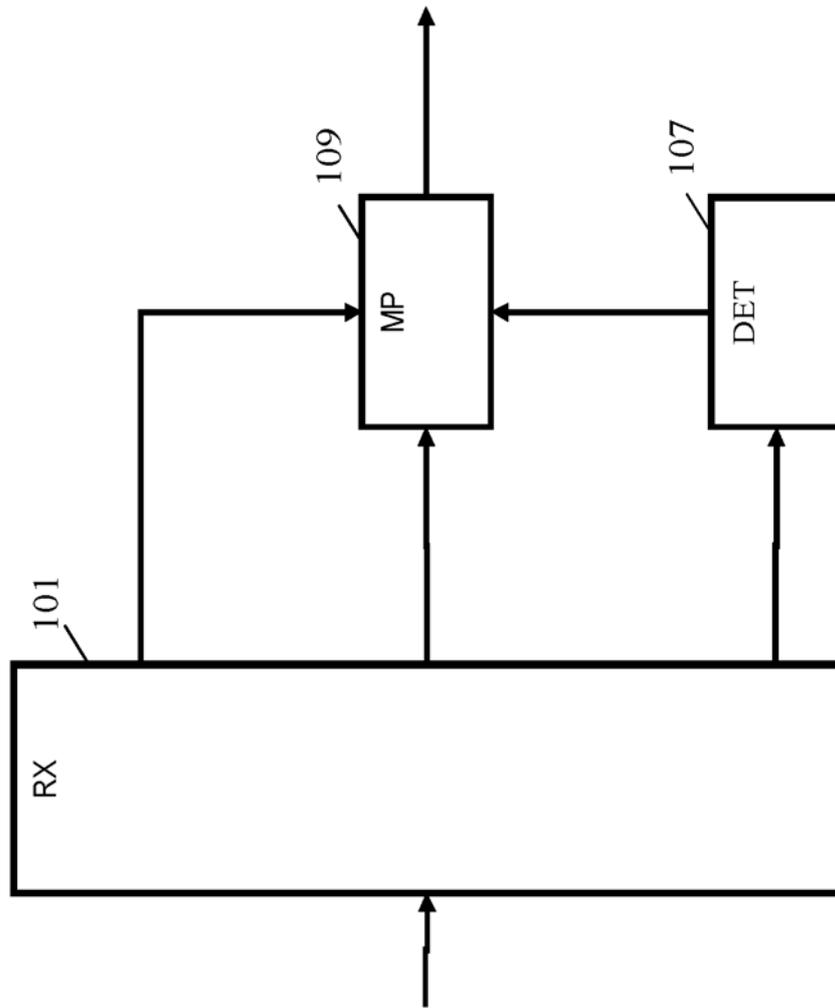


图 3

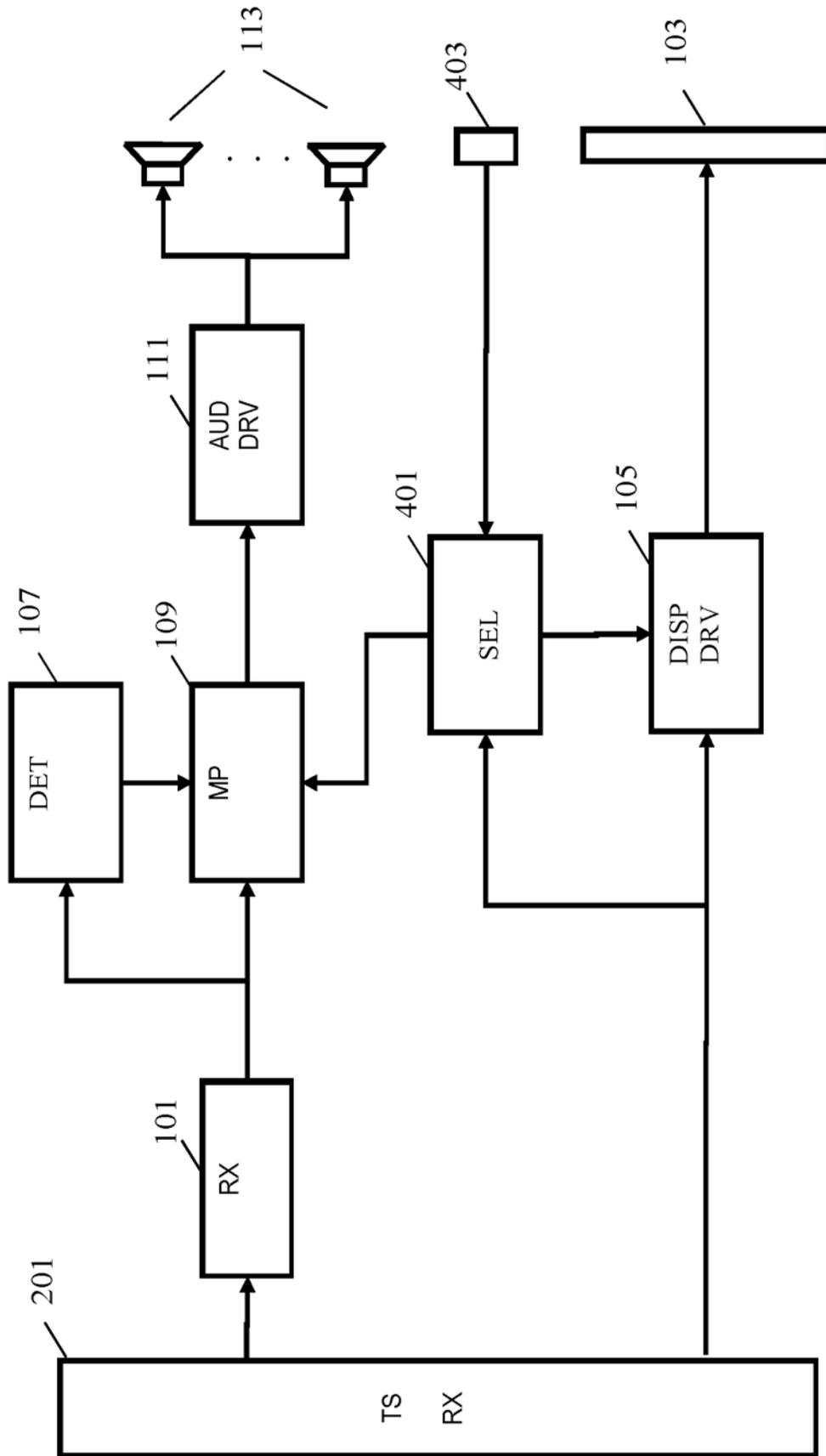


图 4