



(12) 发明专利

(10) 授权公告号 CN 114600188 B

(45) 授权公告日 2025. 07. 08

(21) 申请号 202080072214.3

(22) 申请日 2020.10.08

(65) 同一申请的已公布的文献号
申请公布号 CN 114600188 A

(43) 申请公布日 2022.06.07

(30) 优先权数据
19202935.3 2019.10.14 EP

(85) PCT国际申请进入国家阶段日
2022.04.14

(86) PCT国际申请的申请数据
PCT/EP2020/078297 2020.10.08

(87) PCT国际申请的公布数据
W02021/074007 EN 2021.04.22

(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 P·H·A·迪伦 F·M·J·德邦特
J·G·H·科庞

(74) 专利代理机构 永新专利商标代理有限公司
72002
专利代理师 孟杰雄

(51) Int.Cl.
G10L 19/008 (2013.01)
H04N 21/233 (2011.01)
H04N 21/235 (2011.01)

(56) 对比文件
WO 2018047667 A1, 2018.03.15
US 2012310654 A1, 2012.12.06

审查员 严明蒙

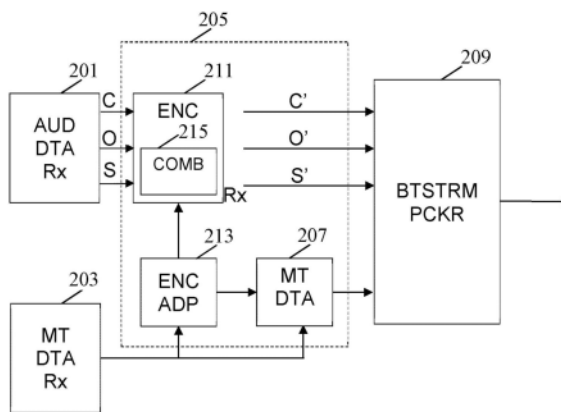
权利要求书2页 说明书15页 附图3页

(54) 发明名称

用于音频编码的装置和方法

(57) 摘要

一种音频编码装置包括接收表示音频场景的音频项目的音频接收器(201),并且元数据接收器(203)接收描述用于对音频项目的渲染的呈现约束的音频项目的输入呈现元数据。呈现约束对当渲染音频项目时可以调整的渲染参数进行约束。音频编码器(205)通过对多个音频项目进行编码来生成针对音频场景的编码音频数据,其中,所述编码响应于输入呈现元数据而被调整。元数据电路(207)根据输入呈现元数据来生成输出呈现元数据。输出呈现元数据包括针对编码音频项目的数据,所述数据约束当渲染编码音频项目时渲染的可调整参数可以被调整的程度。输出(209)生成包括编码音频数据和输出呈现元数据的编码音频数据流。



1. 一种音频编码装置,包括:

音频接收器(201),其被布置为接收表示音频场景的多个输入音频项目;

元数据接收器(203),其被布置为接收针对所述多个输入音频项目的输入呈现元数据,所述输入呈现元数据描述用于对所述多个输入音频项目的渲染的呈现约束;

音频编码器(205),其被布置为通过对所述多个输入音频项目进行编码来生成针对所述音频场景的编码音频数据,所述音频编码器(205)包括组合器(215),所述组合器被布置为通过响应于针对所述多个输入音频项目中的第一音频项目的输入呈现元数据和针对所述多个输入音频项目中的第二音频项目的输入呈现元数据而至少组合所述第一音频项目和所述第二音频项目来生成组合音频项目,并且其中,所述音频编码器(205)被布置为通过对所述组合音频项目进行编码来生成针对所述第一音频项目和所述第二音频项目的组合音频编码数据,并且将所述组合音频编码数据包括在所述编码音频数据中;

元数据电路(207),其被布置为根据所述输入呈现元数据来生成输出呈现元数据,所述输出呈现元数据包括针对所述编码音频数据的数据,所述针对所述编码音频数据的数据约束当渲染所述编码音频数据时客户端处的渲染的可调整参数能够被调整的程度;以及

输出电路(209),其被布置为生成要发送到所述客户端的包括所述编码音频数据和所述输出呈现元数据的编码音频数据流;并且

其中,所述输出呈现元数据包括以下各项中的至少一项:

混响约束;以及

音频项目位置约束。

2. 根据权利要求1所述的音频编码装置,其中,所述组合器(215)被布置为响应于针对所述第一音频项目和所述第二音频项目的所述输入呈现元数据而从所述多个输入音频项目中选择所述第一音频项目和所述第二音频项目。

3. 根据权利要求1或2所述的音频编码装置,其中,所述组合器(215)被布置为响应于对至少一些针对所述第一音频项目的输入呈现元数据和针对所述第二音频项目的输入呈现元数据满足相似性准则的确定而选择所述第一音频项目和所述第二音频项目。

4. 根据权利要求1所述的音频编码装置,其中,针对所述第一音频项目的所述输入呈现元数据和针对所述第二音频项目的所述输入呈现元数据包括增益约束和位置约束中的至少一项。

5. 根据权利要求1所述的音频编码装置,其中,所述音频编码器(205)还被布置为响应于针对所述第一音频项目的所述输入呈现元数据和针对所述第二音频项目的所述输入呈现元数据而生成针对所述组合音频项目的组合呈现元数据;并且将所述组合呈现元数据包括在所述输出呈现元数据中。

6. 根据权利要求5所述的音频编码装置,其中,所述音频编码器(205)被布置为生成至少一些组合呈现元数据以反映针对所述组合音频项目的呈现参数的约束,所述约束被确定为满足由针对所述第一音频项目的输入呈现元数据指示的所述第一音频项目的约束和由针对所述第二音频项目的输入呈现元数据指示的所述第二音频项目的约束两者的约束。

7. 根据权利要求1所述的音频编码装置,其中,所述音频编码器(205)被布置为响应于针对第二音频项目的输入呈现元数据而调整对第一音频项目的压缩。

8. 根据权利要求7所述的音频编码装置,其中,所述音频编码器(205)被布置为响应于

针对所述第二音频项目的输入呈现元数据而估计由所述第二音频项目对所述第一音频项目的掩蔽效果,并且响应于所述掩蔽效果而调整对所述第一音频项目的所述压缩。

9. 根据权利要求8所述的音频编码装置,其中,所述音频编码器(205)被布置为响应于由针对所述第二音频项目的所述输入呈现元数据指示的所述第二音频项目的增益约束和位置约束中的至少一项而估计由所述第二音频项目对所述第一音频项目的所述掩蔽效果。

10. 根据权利要求7所述的音频编码装置,其中,所述音频编码器(205)还被布置为响应于针对所述第一音频项目的输入呈现元数据而调整对所述第一音频项目的所述压缩。

11. 根据权利要求1所述的音频编码装置,其中,所述输入呈现元数据包括针对至少一些音频项目的优先级数据,并且所述编码器被布置为响应于所述输入呈现元数据中针对所述第一音频项目的优先级指示而调整针对第一音频项目的压缩。

12. 根据权利要求1所述的音频编码装置,其中,所述音频编码器(205)被布置为生成指示如何响应于所述输入呈现元数据而调整所述编码的编码调整数据,并且将所述编码调整数据包括在所述编码音频数据流中。

13. 一种编码音频的方法,所述方法包括:

接收表示音频场景的多个输入音频项目;

接收针对所述多个输入音频项目的输入呈现元数据,所述输入呈现元数据描述用于对所述多个输入音频项目的渲染的呈现约束;

通过对所述多个输入音频项目进行编码来生成针对所述音频场景的编码音频数据,所述编码通过以下操作来响应于所述输入呈现元数据:通过响应于针对所述多个输入音频项目中的第一音频项目的输入呈现元数据和针对所述多个输入音频项目中的第二音频项目的输入呈现元数据而至少组合所述第一音频项目和所述第二音频项目来生成组合音频项目;以及通过对所述组合音频项目进行编码,来生成针对所述第一音频项目和所述第二音频项目的组合音频编码数据,并且将所述组合音频编码数据包括在所述编码音频数据中;

根据所述输入呈现元数据来生成输出呈现元数据,所述输出呈现元数据包括针对所述编码音频数据的数据,所述针对编码音频数据的数据约束当渲染所述编码音频数据时客户端处的渲染的可调整参数能够被调整的程度;并且

生成要发送到所述客户端的包括所述编码音频数据和所述输出呈现元数据的编码音频数据流;并且

其中,所述输出呈现元数据包括以下各项中的至少一项:

混响约束;以及

音频项目位置约束。

14. 一种包括计算机程序代码模块的计算机程序产品,所述计算机程序代码模块适于当所述程序在计算机上运行时执行权利要求13的所有步骤。

用于音频编码的装置和方法

技术领域

[0001] 本发明涉及用于音频编码的装置和方法,特别是但不排他地,用于针对诸如虚拟现实应用的动态应用对音频进行编码的装置和方法。

背景技术

[0002] 近年来,音频和视频应用的种类和范围已经实质增加,新的服务以及利用和消费音频、图像和视频的方式不断被开发和引入。

[0003] 例如,越来越受欢迎的一项服务是以观察者能够主动和动态地与系统交互以改变渲染的参数的方式提供音频和图像。许多应用中一个非常吸引人的特征是改变有效的查看/收听位置的能力。这样的特征可以具体地允许向用户提供虚拟现实体验。

[0004] 趋势是提供越来越大的灵活性,其允许场景的渲染侧调整。为了为音频场景的渲染提供增加的渲染侧灵活性,已经提出了多个音频编码和分布方法,其中,音频场景可以由不同音频项目的组合表示。例如,音频项目可以表示单独的声音源,例如特定的发声者等。在一些方法中,所有音频项目都是相同类型的,但越来越多的系统被开发,其允许同时使用和支持多种不同的音频类型。例如,一些音频项目可以是音频通道,其他的可以是单独的音频对象,另外其他的可以是基于场景的,诸如多声道模拟立体声(Ambisonic)音频项目等。在许多系统中,元数据可以与表示音频项目的音频数据一起提供。这样的元数据可以例如指示针对音频项目的音频源的场景中的标称位置。

[0005] 这样的方法可以实现高程度的客户端/渲染侧定制和调整。例如,音频场景可以局部适于音频场景中的收听者虚拟位置的变化或个体收听者的特定偏好。

[0006] 作为特定示例,3GPP联盟目前正在开发所谓的沉浸式语音和音频服务(IVAS)编解码器。该编解码器将能够以各种配置对音频内容进行编码,例如基于通道、对象或场景(特别是多声道模拟立体声)的配置。编码的目的是用最少量的数据来传达音频信息。

[0007] IVAS编解码器还应包括渲染器,以用于将各种音频流转换为适合于在接收端再现的形式。例如,音频可以映射到已知的扬声器配置中,或者音频可以被渲染为双耳格式以经由耳机进行再现。

[0008] 在3GPP IVAS编解码器范围内,正在进行工作以收集潜在用例。对于这些,设想编解码器应该提供交互性来调制渲染。例如,耳机音频可能必须独立于头部位置和平移进行渲染,这意味着必须对头部移动进行补偿。作为另一示例,可以使得用户能够在空间上定位音频项目,例如(重新)定位承载虚拟会议中的参与者的音频的对象。

[0009] 渲染器被视为3GPP IVAS编解码器工作项的一部分,并被视为在IVAS编解码器的内部。然而,已经提出编解码器还包括直通模式。该模式将允许在解码器输出处以与其编解码器输入处输入相同的(一个或多个)配置表示音频项目(即,作为1:1对应的通道、基于对象和场景的音频项目)。经由专用的外部渲染接口,外部渲染器可以具有对这些项目的访问权并且可以实现内部IVAS渲染器的备选渲染。

[0010] 这种方法可以提供额外的灵活性并增加在接收端的定制和调整的范围。然而,该

方法也能够具有相关联的缺点。例如,在灵活性与音频质量和复杂性之间存在权衡。限制自由度从而允许内容提供者在客户端侧保留对渲染的一些控制通常可能是有用的。这不仅可以帮助渲染并产生更逼真的渲染音频场景,而且可以允许内容提供者保留关于被提供给用户的体验的一些控制。例如,其能够防止渲染器生成不真实并且可能在内容和内容提供者上反映不佳的音频场景。

[0011] 设想编码音频项目可以由元数据补充,该元数据限制了如何允许渲染器渲染音频项目。在许多情况下,这可以允许不同需求之间的改进的权衡。然而,其可能不是在所有情况下都是最佳的,并且可能例如需要增加数据速率,并且可能导致针对渲染的音频场景的降低的灵活性和/或质量。

[0012] 因此,一种改进的方法将是期望的。特别地,一种允许改进的操作、增加的灵活性、便利的实施方式、便利的操作、降低的成本、降低的复杂性、降低的数据速率、改进的感知音频质量、改进的渲染控制、改进的权衡和/或改进的性能的方法将是有利。

发明内容

[0013] 因此,本发明寻求优选地单独或以任何组合方式缓解、减轻或消除上述缺点中的一个或多个。

[0014] 根据本发明的一个方面,提供了一种音频编码装置,包括:音频接收器,其用于接收表示音频场景的多个音频项目;元数据接收器,其用于接收针对所述多个音频项目的输入呈现元数据,所述输入呈现元数据描述用于对所述多个音频项目的渲染的呈现约束,所述呈现约束约束当渲染所述多个音频项目时能够被调整的渲染参数,所述呈现约束包括来自具有以下各项的组的至少一个约束:混响约束;增益约束;以及动态范围控制约束;音频编码器,其用于通过对所述多个音频项目进行编码来生成针对所述音频场景的编码音频数据,所述编码通过包括组合器(215)来响应于所述输入呈现元数据,所述组合器用于通过至少响应于针对所述多个音频项目中的第一音频项目的输入呈现元数据和针对所述多个音频项目中的第二音频项目的输入呈现元数据而组合所述第一音频项目和所述第二音频项目来生成组合音频项目,并且所述音频编码器(205)被布置为通过对所述组合音频项目进行编码来生成针对所述第一音频项目和所述第二音频项目的组合音频编码数据,并且将所述组合音频编码数据包括在所述编码音频数据中;元数据电路,其用于根据所述输入呈现元数据来生成输出呈现元数据,所述输出呈现元数据包括针对编码音频项目的数据,所述数据约束当渲染所述编码音频项目时渲染的可调整参数能够被调整的程度;以及输出电路,其用于生成包括所述编码音频数据和所述输出呈现元数据的编码音频数据流。

[0015] 本发明可以在许多情形中提供改进的和/或更灵活的编码。在许多实施例中,该方法可以允许生成提供改进的质量与比特率比率的编码音频数据流。可以生成编码音频数据流以允许渲染的一些灵活性,同时还允许从源/编码侧对渲染的一些控制。

[0016] 针对音频项目的呈现元数据可以约束用于对音频项目的渲染的空间参数和音量参数中的至少一项,包括例如约束渲染位置、增益水平、信号水平、空间分布或混响性质。

[0017] 音频编码器可以被布置为基于输入呈现元数据,并且具体地基于针对音频项目的输入呈现元数据来调整音频项目的编码。该调整可以调整用于音频项目的编码的比特/数据(速率)压缩。可以基于输入呈现元数据来调整由对音频项目进行编码所产生的比特率。

[0018] 输入呈现元数据可以描述针对接收到的多个音频项目的呈现/渲染约束。编码音频数据可以包括针对多个编码音频项目的音频数据。可以通过对接收到的多个音频项目进行编码来生成多个编码音频项目。描述用于对多个编码音频项目的渲染的呈现/渲染约束的输出呈现元数据。

[0019] 呈现约束可以是渲染约束并且可以约束针对音频项目的渲染参数。渲染参数可以是渲染过程的参数和/或渲染信号的性质。

[0020] 输出呈现元数据具体可以是与由音频编码器生成的编码音频项目相关联/链接/为其提供的任何数据,其约束当渲染编码音频项目时呈现/渲染的一个或多个可调整/可变方面/性质/参数可以/能够被调整的程度。

[0021] 输出呈现元数据,并且特别是针对编码音频项目的数据,其约束当渲染编码音频项目时渲染的可调整参数能够被调整的程度,可以由元数据电路响应于呈现约束约束当渲染多个音频项目时可以调整的渲染参数而生成。

[0022] 音频编码器可以生成编码音频数据以包括多个编码音频项目(通过编码多个音频项目)。

[0023] 所述音频编码器包括包括组合器,所述组合器用于通过响应于针对所述多个音频项目中的第一音频项目的输入呈现元数据和针对所述多个音频项目中的第二音频项目的输入呈现元数据而组合至少所述第一音频项目和所述第二音频项目来生成组合音频项目,并且所述音频编码器被布置为通过对所述组合音频项目进行编码来生成针对所述第一音频项目和所述第二音频项目的组合音频编码数据,并且将所述组合音频编码数据包括在所述编码音频数据中。

[0024] 这可以在许多实施例中提供特别有效的编码和/或灵活性。在许多实施例中,其可以特别地提供具有减少的感知退化的有效的比特率压缩。

[0025] 根据本发明的任选特征,所述组合器被布置为响应于针对所述第一音频项目和所述第二音频项目的所述输入呈现元数据而从所述多个音频项目中选择所述第一音频项目和所述第二音频项目。

[0026] 这可以在许多实施例中提供特别有效的编码和/或灵活性。

[0027] 根据本发明的任选特征,所述组合器被布置为响应于对至少一些针对所述第一音频项目的输入呈现元数据和针对所述第二音频项目的输入呈现元数据满足相似性准则的确定而选择所述第一音频项目和所述第二音频项目。

[0028] 这可以在许多实施例中提供特别有效的编码和/或灵活性。相似性准则可以包括对由呈现元数据约束的渲染参数的渲染约束满足相似性标准的要求。

[0029] 根据本发明的任选特征,针对所述第一音频项目的所述输入呈现元数据和针对所述第二音频项目的所述输入呈现元数据包括增益约束和位置约束中的至少一项。

[0030] 这可以在许多实施例中提供特别有效的操作。

[0031] 根据本发明的任选特征,所述音频编码器还被布置为响应于针对所述第一音频项目的所述输入呈现元数据和针对所述第二音频项目的所述输入呈现元数据而生成针对所述组合音频项目的组合呈现元数据;并且将所述组合呈现元数据包括在所述输出呈现元数据中。

[0032] 这可以在许多实施例中提供改进的操作并且尤其在许多实施例中可以允许编码

器以相同的方式处置组合音频项目和编码输入音频项目,并且实际上不知道个体音频项目是否是组合音频项目。

[0033] 根据本发明的任选特征,所述音频编码器被布置为生成至少一些组合呈现元数据以反映针对所述组合音频项目的呈现参数的约束,所述约束被确定为满足由针对所述第一音频项目的输入呈现元数据指示的所述第一音频项目的约束和由针对所述第二音频项目的输入呈现元数据指示的所述第二音频项目的约束两者的约束。

[0034] 这可以在许多情形和应用中提供改进的性能。

[0035] 根据本发明的任选特征,所述音频编码器被布置为响应于针对第二音频项目的输入呈现元数据而调整对第一音频项目的压缩。

[0036] 该方法通常可以允许音频项目的改进的压缩和编码。压缩可以是比特率降低并且增加压缩可以导致编码音频项目的降低的数据速率。压缩可以是比特率降低/压缩。音频编码可以使得表示一个或多个输入音频项目的编码音频项目由比一个或多个输入音频项目更少的比特表示。

[0037] 根据本发明的任选特征,所述音频编码器被布置为响应于针对所述第二音频项目的输入呈现元数据而估计由所述第二音频项目对所述第一音频项目的掩蔽效果,并且响应于所述掩蔽效果而调整对所述第一音频项目的所述压缩。

[0038] 这可以在许多实施例中提供特别有效的操作和改进的性能。

[0039] 根据本发明的任选特征,所述音频编码器被布置为响应于由针对所述第二音频项目的所述输入呈现元数据指示的所述第二音频项目的增益约束和位置约束中的至少一项而估计由所述第二音频项目对所述第一音频项目的所述掩蔽效果。

[0040] 这可以在许多实施例中提供特别有效的操作和改进的性能。

[0041] 根据本发明的任选特征,音频编码器还被布置为响应于针对所述第一音频项目的输入呈现元数据而调整对所述第一音频项目的所述压缩。

[0042] 这可以在许多实施例中提供特别有利的操作和/或性能。

[0043] 根据本发明的任选特征,所述输入呈现元数据包括针对至少一些音频项目的优先级数据,并且所述编码器被布置为响应于所述输入呈现元数据中针对所述第一音频项目的优先级指示而调整针对第一音频项目的压缩。

[0044] 这可以在许多实施例中提供特别有利的操作和/或性能。

[0045] 根据本发明的任选特征,所述音频编码器被布置为生成指示如何响应于所述输入呈现元数据而调整所述编码的编码调整数据,并且将所述编码调整数据包括在所述编码音频数据流中。

[0046] 这可以在许多实施例中提供特别有利的操作和/或性能。其尤其可以允许通过解码器的改进的调整以匹配编码过程。

[0047] 根据本发明的一个方面,提供了一种编码音频的方法,所述方法包括:接收表示音频场景的多个音频项目;接收针对所述多个音频项目的输入呈现元数据,所述输入呈现元数据描述用于对所述多个音频项目的渲染的呈现约束,所述呈现约束约束当渲染所述音频项目时能够被调整的渲染参数,所述呈现约束包括来自具有以下各项的组的至少一个约束:混响约束;增益约束;以及动态范围控制约束;通过对所述多个音频项目进行编码来生成针对所述音频场景的编码音频数据,所述编码通过以下操作来响应于所述输入呈现元数

据:通过至少响应于针对所述多个音频项目中的第一音频项目的输入呈现元数据和针对所述多个音频项目中的第二音频项目的输入呈现元数据而组合所述第一音频项目和所述第二音频项目来生成组合音频项目,并且通过对所述组合音频项目进行编码来生成针对所述第一音频项目和所述第二音频项目的组合音频编码数据,并且将所述组合音频编码数据包括在所述编码音频数据中;根据所述输入呈现元数据来生成输出呈现元数据,所述输出呈现元数据包括针对编码音频项目的数据,所述数据约束当渲染所述编码音频项目时渲染的可调整参数能够被调整的程度;并且生成包括所述编码音频数据和所述输出呈现元数据的编码音频数据流。

[0048] 本发明的这些和其他方面、特征和优点将根据下文描述的一个或多个实施例而显而易见并且参考下文中描述的一个或多个实施例得到阐述。

附图说明

[0049] 将参考附图仅通过示例描述本发明的实施例,其中,

[0050] 图1图示了根据本发明的一些实施例的音频分布系统的元件的示例;

[0051] 图2图示了根据本发明一些实施例的音频编码装置的元件的示例;并且

[0052] 图3图示了根据本发明一些实施例的音频解码装置的元件的示例。

具体实施方式

[0053] 以下描述将集中于可以与3GPP沉浸式语音和音频服务 (IVAS) 编解码器兼容的音频编码和解码系统,但是将意识到,所描述的原理和概念可以用于许多其他应用和实施例中。

[0054] 图1图示了音频编码系统的示例。在该系统中,音频源101向音频编码器单元103提供音频数据。音频数据包括针对表示音频场景的音频的多个音频项目的音频数据。音频项目可以被提供为不同的类型,具体包括:

[0055] 基于通道的音频项目:对于此类音频项目,1D(单声道)、2D或3D空间音频内容通常被表示为离散信号,旨在经由扬声器在相对于收听者的预定位置处呈现。众所周知的扬声器设置是例如双通道立体声(也称为“2.0”),或围绕收听者的5个通道加上低频效果通道(也称为“5.1”)。而且,双耳音频通常被认为是基于通道的音频,包括两个音频信号通道,其旨在直接呈现给收听者相应的耳朵(通常经由耳机)。

[0056] 基于对象的音频项目:对于此类音频项目,通常使用个体音频信号来表示不同的声音源。这些声音源常常与实际对象或人有关,例如电话会议的参与者。信号通常是单声道的,但也可以使用其他表示。基于对象的音频信号常常伴随有描述另外的性质的元数据,诸如对象音频的范围(空间扩展)、方向性或扩散性。

[0057] 基于场景的音频项目:对于此类音频项目,原始2D或3D空间音频场景通常表示为与某些球谐函数有关的多个音频信号。借助于组合这些基于场景的音频信号,可呈现音频信号可以在任意2D或3D位置处被构建,例如在实际扬声器在音频再现设置中的位置处。基于场景的音频的示例实施方式是多声道模拟立体声。基于场景的音频使用称为“高阶多声道模拟立体声”(HOA)的声场技术来创建独立于特定扬声器布局的实况捕获和艺术创建的声音场景两者的整体描述。

[0058] 除了音频数据之外,音频源还可以提供针对音频项目的呈现元数据。呈现元数据可以描述用于对音频场景的渲染的呈现约束,并且因此可以提供针对多个音频项目的呈现/渲染约束。

[0059] 呈现元数据可以描述如何由渲染器执行对音频项目的渲染的约束。呈现元数据可以定义对一个或多个渲染参数/性质的约束。参数/性质可以具体地是影响音频项目的渲染的感知性质的参数/性质。约束可以是影响场景中的音频项目的空间感知和/或(相对)信号水平的约束。呈现元数据可以具体地约束针对一个或多个音频项目的空间和/或增益/信号水平参数。元数据例如可以是对每个音频项目的位置和/或增益的约束。

[0060] 元数据例如可以描述针对一个或多个音频项目的一个或多个参数的范围或一组可允许值。对(一个或多个)音频项目的渲染可以在约束内自由执行,即渲染可以使得受约束的参数具有指示的可允许值中的任何,但可能不使得受约束的参数不具有该值。

[0061] 作为示例,呈现元数据可以针对音频项目中的一个或多个来描述区域和/或(相对)增益范围。然后必须用区域内的感知位置和/或用增益范围内的增益来渲染音频项目。

[0062] 呈现元数据因此可以约束渲染,同时仍然允许一些灵活性来调整和定制局部渲染。

[0063] 可能由呈现元数据提供的参数或性质的渲染约束的示例包括:

[0064] 针对一个或多个音频项目的位置约束。例如,这可以定义音频场景中必须从其渲染音频项目的空间区域或体积。

[0065] 针对一个或多个音频项目的混响约束。这可以例如定义最小或最大混响时间。该约束可以例如确保以期望的扩散度渲染音频项目。例如,表示一般环境背景声音的音频项目可能需要以最小的混响量来渲染,而表示主发声者的音频项目可能需要以小于给定的混响阈值来渲染。

[0066] 增益约束。可以由渲染器根据渲染过程的特定偏好将对音频项目的渲染调整到更响亮或更安静。例如,相对于环境背景声音的发声者的增益在某些情况下可以基于收听者偏好来增加或减少。然而,增益约束能够约束增益可以被修改多少,例如从而确保在环境噪音上始终可以听到发声者。

[0067] 响度约束。可以由渲染器根据渲染过程的特定偏好将对音频项目的渲染调整到更响亮或更安静。例如,在一些情况下,电话会议参与者的增益能够基于收听者的偏好来增加或减少。然而,响度约束能够约束某些参与者的感知响度可以被修改多少,例如从而确保例如在存在其他发声者或背景噪音的情况下,会议主席的声音总是足够响亮。

[0068] 动态范围控制约束。音频项目的动态范围可以由渲染器调整到更响亮,例如其可能降低,使得在收听者位置处存在背景噪音的情况下,在较低水平的时段期间,音频也保持可听。例如,小提琴声音可以在低水平处自动实现为更响亮。然而,动态范围控制约束能够约束动态范围可以减少多少,因此例如确保对小提琴的正常动态的足够自然的感知。

[0069] 描述用于对多个音频项目的渲染的呈现约束的呈现元数据可以具体地是提供对渲染参数或性质的约束的数据,该渲染参数或性质可以在渲染音频项目时被调整(为其提供呈现元数据)。渲染参数或性质可以是渲染操作的参数/性质和/或生成的渲染/呈现的信号和/或音频的参数或性质。

[0070] 输入呈现元数据具体可以是与用于音频编码器205的输入音频项目相关联/链接/

为其提供的任何数据,其约束呈现/渲染的一个或多个可调整/可变方面/性质/参数当渲染输入音频项目时可以/能够被调整的程度。

[0071] 音频编码器单元103被布置为生成编码音频数据流,该编码音频数据流包括用于音频场景的编码音频数据。编码音频数据是通过对音频项目进行编码生成的(即,接收到的表示音频项目的音频数据)。此外,音频编码器单元103为编码音频项目生成输出呈现元数据,并将该元数据包括在编码音频数据流中。输出呈现元数据描述了用于编码音频项目的渲染约束。

[0072] 输出呈现元数据具体可以是与由音频编码器205生成的编码音频项目相关联/链接/为其提供的任何数据,其约束呈现/渲染的一个或多个可调整/可变方面/性质/参数当渲染编码音频项目时可以/能够被调整的程度。

[0073] 响应于(输入)呈现约束,可以由元数据电路生成输出呈现元数据,并且特别是针对约束渲染的可调整参数当渲染编码音频项目时可以被调整的程度的编码音频项目的数据,所述(输入)呈现约束约束当渲染多个(输入)音频项目时可以调整的渲染参数。

[0074] 音频编码器单元103被耦合到发射器105,发射器105被馈送编码音频数据流。发射器105在示例中被布置为将编码音频数据流发送/分布到一个或多个客户端,该一个或多个客户端可以基于编码音频数据流来渲染音频场景。

[0075] 在该示例中,编码音频数据流经由网络107分布,该网络具体可以是或可以包括因特网。发射器105可以被布置为同时支持潜在大量的客户端,并且音频数据通常可以分布到多个客户端。

[0076] 在特定示例中,编码音频数据流可以被发送到一个或多个渲染设备109。渲染设备109可以包括接收器111,其从网络107接收编码音频数据流。

[0077] 将意识到,发射器105和接收器111可以以任何合适的形式并使用任何合适的通信协议、标准、技术和功能进行通信。在示例中,发射器105和接收器111可以包括适当的网络接口功能,但是将意识到,在其他实施例中,发射器105和/或接收器111可以例如包括无线电通信功能、光纤通信功能等。

[0078] 接收器111耦合到解码器113,解码器113被馈送接收到的编码音频数据流。解码器113被布置为对编码音频数据流进行解码以重新创建音频项目。解码器113还可以从编码音频数据流中解码呈现元数据。

[0079] 解码器113耦合到渲染器115,渲染器115被馈送用于音频项目的解码音频数据和呈现元数据。渲染器115可以通过基于接收到的呈现元数据渲染音频项目来渲染音频场景。渲染器115的渲染可以针对所使用的特定音频再现系统。例如,对于5.1环绕声系统,可以生成针对个体通道的音频信号,对于耳机系统,可以使用例如HRTF滤波器等生成双耳信号。将意识到,许多不同的可能的音频渲染算法和技术是已知的,并且可以使用任何合适的方法而不偏离本发明。

[0080] 渲染器115可以具体地生成用于再现的输出音频信号,使得组合再现在由收听者感知时提供音频场景的感知。渲染器通常会根据针对个体音频项目的特定特性分别且不同地处理不同的音频项目,并且然后为每个输出通道组合所得到的信号分量。例如,对于音频对象音频项目,可以根据对应于音频对象的音频源在音频场景中的期望位置为每个输出通道生成信号分量。音频通道音频项目可以例如通过针对对应的输出再现通道的信号分量的

生成来进行渲染,或者例如如果它没有精确地映射到再现通道之一(例如,如果合适,使用平移或上混技术)则可以通过多个再现通道来进行渲染。

[0081] 由多个通常不同类型的音频项目对音频场景的表示可以允许渲染器115在对场景的渲染中具有高度的灵活性和适应性。例如,这可以由渲染器使用以调整和定制渲染的音频场景。例如,可以调整不同音频对象的相对增益和/或位置,可以修改音频项目的频率内容,可以控制音频项目的动态范围,可以改变混响性质,等等。因此,渲染器115可以生成输出,其中,音频场景适于针对当前应用/渲染的特定偏好,包括对所使用的特定再现系统和/或收听者的个人偏好的适应。例如,该方法还可以允许所渲染的音频场景有效地局部适应于音频场景中的虚拟收听位置的变化。例如,为了支持虚拟现实应用,渲染器115可以动态且连续地接收用户位置数据输入并且响应于音频场景中用户的所指示的虚拟位置的变化来调整渲染。

[0082] 渲染器115被布置为基于接收到的呈现元数据来渲染音频项目。特别地,呈现元数据可以指示对编码/解码音频项目的渲染的可变方面/性质/参数的约束,并且呈现器115可以在渲染时依从这些约束。

[0083] 来自渲染器115/渲染设备109的输出音频信号是由应用于由解码器113从接收到的编码音频数据流生成的解码音频项目的渲染操作产生的。渲染操作可能具有一些参数,所述参数可以在外部或局部地进行调整,并且在感知上影响渲染的输出音频(的方面)。描述用于渲染的呈现约束的呈现元数据可以具体地是限制渲染参数当渲染时可以被调整的集合(即,对于连续可调整参数,为值范围,或对于枚举参数,为离散值的集合)的数据。

[0084] 图2更详细地示出了音频编码器单元103的元件的示例。在该示例中,音频编码器单元103包括音频接收器201,其接收描述场景的输入音频数据。在本示例中,音频场景由三种不同类型的音频数据表示,即基于通道的音频项目C、基于对象的音频项目O和基于场景的音频项目S。音频项目由可以采取任何合适的形式的音频数据提供。音频数据可以例如提供音频项目作为原始WAV文件或作为根据任何合适格式编码的音频。通常,输入音频项目将处于高音频质量和高数据速率。

[0085] 音频编码器单元103还包括元数据接收器203,其被布置为接收针对输入音频项目的呈现元数据。如前所述,呈现元数据可以提供对音频项目的渲染的约束。

[0086] 音频接收器201和元数据接收器203耦合到音频编码器205,音频编码器205被布置为通过对接收到的音频项目进行编码来生成用于音频场景的编码音频数据。该示例中的音频编码器205具体生成编码音频项目,即由编码音频数据表示的音频项目。如对于输入音频项目,输出/编码音频项目也可以是不同类型的音频项目,并且在具体示例中具体可以是基于通道的音频C'、基于对象的音频项目O'和基于场景的音频项目S'。

[0087] 一个、一些或所有编码音频项目可以通过独立地编码输入音频项目来生成,即编码音频项目可以是编码输入音频项目。然而,在一些情形中,编码音频项目中的一个或多个可以被生成以表示多个输入音频项目,或者输入音频项目可以在多个编码音频项目中表示/由多个编码音频项目表示。

[0088] 将意识到,许多编码算法和技术是已知的并且可以使用任何合适的算法、标准和方法。还将意识到,不同的算法和技术可以用于不同的音频项目。例如,对应于音乐的音频项目可以使用AAC编码方法来编码,对应于语音的音频项目可以使用CELP编码方法来编码,

等等。对于已经以编码格式接收的音频项目,由音频编码器205编码可以是到不同编码格式的代码转换,或者可以例如只是数据速率转换(例如,通过修改量化和/或削波水平)。通常,编码包括比特率压缩,并且编码音频项目由比输入音频项目更少的比特表示。

[0089] 音频编码器单元103还包括元数据电路207,元数据电路207被布置为生成针对编码音频项目的输出呈现元数据。呈现元数据电路207被布置为根据接收到的输入呈现元数据生成该输出呈现元数据。实际上,对于许多音频项目,输出呈现元数据可能与输入呈现元数据相同。对于一个或多个音频项目,可以修改输出呈现元数据,如稍后将更详细描述。

[0090] 音频编码器205和元数据电路207耦合到输出电路209,输出电路209被布置为生成包括编码音频数据和输出呈现元数据的编码音频数据流。输出电路209具体可以是比特流打包器,其生成包括编码音频数据和输出元数据两者的编码音频数据流。可以根据标准化格式生成编码音频数据流,从而允许它被一系列接收器解释。

[0091] 因此,输出电路209操作为比特流打包器,其接受比特率降低/编码音频项目和输出呈现元数据,并将它们组合成可以通过合适的通信通道(诸如例如通过5G网络)传达的比特流。

[0092] 图3图示了可以接收和处理来自音频编码器单元103的编码音频数据流的渲染设备109的元件的具体示例。渲染设备109包括比特流解包器形式的接收器111,其接收来自音频编码器单元103的编码音频数据流,并从接收到的数据流中分离出不同的数据。具体地,接收器111可以分离出针对编码音频项目的个体音频数据,并将这些馈送到解码器113。

[0093] 解码器113具体被布置为对接收到的编码音频项目进行解码,以生成基于通道、对象和场景的音频项目的通常未编码的表示。

[0094] 对于许多音频项目,解码器113可以反转由音频编码器205执行的编码。对于其他音频项目,解码可以例如仅部分反转编码操作。例如,如果音频编码器205已经将音频项目组合成单个组合音频项目,则解码器113可以仅对组合音频项目进行解码而不完全生成个体音频项目。将意识到,可以根据个体实施例的特定偏好和要求使用任何合适的解码算法和技术。

[0095] 解码音频项目被馈送到渲染器115,渲染器115被布置为通过如前所述渲染音频项目例如作为双耳信号或环绕声信号来渲染音频场景。

[0096] 渲染设备109还包括元数据控制器/电路301,其被馈送来自接收器111的呈现元数据。在该示例中,元数据控制器301还可以接收例如可以反映局部偏好或要求的局部呈现元数据,诸如例如个人用户偏好或所使用的再现系统的性质。

[0097] 因此,除了从接收到的比特流中解包的音频呈现元数据之外,渲染设备109还可以接受局部音频呈现元数据,其例如可以经由一个或多个输入接口提供。该数据可以提供关于在其中呈现音频的背景的信息,所述信息在编码器侧不可用,诸如例如:

[0098] -期望的呈现(扬声器)配置;

[0099] -用户偏好(例如,虚拟会议中参与者音频的音频水平和取向);

[0100] -局部声学的性质,例如房间的回响。这可以允许渲染器确定哪些环境效果和性质应用于音频项目;

[0101] -局部音频信号(例如,当为音频项目选择增益时考虑);

[0102] -收听者位置;以及

[0103] -收听者头部取向。

[0104] 元数据控制器301可以将接收到的元数据和局部元数据合并在一起并将其提供给渲染器115,渲染器115可以前进到根据呈现元数据的约束来渲染音频项目。

[0105] 渲染器115可以将由解码器113生成的音频项目“C”、“O”和“S”组合成期望的呈现配置(例如双耳或环绕声)的可呈现音频。

[0106] 渲染器115可以具体地根据从元数据控制器301接收的元数据并且在所渲染的音频由接收到的呈现元数据的约束(即,从编码器侧约束)所约束的情况下来生成音频呈现。这提供了源侧/内容提供者对音频渲染和呈现的音频场景的控制,同时仍然允许客户端侧的一些灵活性。这能够例如用于提供服务或应用,其中,内容作者保留对沉浸式应用的控制,该沉浸式应用被设计为向终端用户等提供一定的有限控制。

[0107] 更详细地,元数据控制器301可以处理接收到的元数据,例如音频项目的抑制,相应地局部元数据。元数据控制器301可以例如限制局部元数据,例如旋转或仰角的范围,相应地接收到的元数据。

[0108] 在一些实施例中,渲染器115可以是与渲染设备109不同的设备或功能实体。例如,诸如设想的3GPP IVAS编解码器的标准可以规定解码器113的操作但允许渲染器115是专有的并且可更自由地调整。在一些实施例中,元数据控制器301可以是不同设备或功能实体的部分。

[0109] 在这样的实施例中,因此需要外部渲染器来处理 and 解释解码音频项目“O”、“C”、“S”和接收到的呈现元数据。外部渲染器的渲染操作仍然必须符合由呈现元数据提供的约束。

[0110] 呈现元数据因此可以是由源侧/内容提供者用于控制客户端处的渲染操作的数据。渲染必须根据呈现元数据进行调整/限制。

[0111] 然而,除了用于控制客户端侧渲染器115的渲染的呈现元数据之外,音频编码器单元103的音频编码器205还被布置为响应于输入的呈现元数据来调整编码。输入的呈现元数据被馈送到音频编码器205,并且这可以基于呈现元数据修改一个或多个音频项目的编码(通常用于一个或多个音频项目)。音频编码器205因此是响应于与音频项目一起接收的呈现元数据的可调整编码器。

[0112] 音频编码器205具体包括编码电路211和编码适配器213,编码电路211被布置为执行对音频项目的编码,编码适配器213被布置为基于呈现元数据来调整编码电路211的编码。

[0113] 编码适配器213可以被布置为基于针对给定音频项目的呈现元数据来设置针对该音频项目的编码的参数。例如,其可以被布置为基于例如由呈现元数据指示为音频项目可允许的增益范围或位置范围来设置比特率分配/目标、量化水平、掩蔽阈值、频率范围等。

[0114] 在许多实施例中,编码电路211是比特率压缩器,其被布置为用与接收到的输入音频项目相比减少数量的比特对音频项目进行编码。编码因此可以是比特率压缩,从而允许更有效和更容易分布要生成的编码音频数据流。在这样的实施例中,编码适配器213可以基于呈现元数据来调整编码电路211的比特率降低(以便根据合适的优化准则/算法来优化渲染音频的质量)。

[0115] 编码适配器213可以例如执行编码分析过程,该过程分析呈现元数据并决定如何最好地执行各种输入音频项目的比特率降低。可由编码适配器213执行的操作和调整的示

例包括：

[0116] -发信号通知编码电路211的(最小)掩蔽水平以遵守比特率降低。编码适配器213具有与哪些音频项目被共同呈现以及在什么水平和在哪个取向有关的信息。这可以允许其为个体音频项目调整掩蔽水平,然后由编码使用掩蔽水平。

[0117] -转换音频项目,例如将音频对象移动到基于通道或场景的音频中。

[0118] -选择用于下混的音频项目(具有相关联的上混参数),其中,可以上混下混以在解码器侧重建沉浸式音频,同时确保参数下混编码的伪影被一起呈现的各种音频项目充分掩蔽。作为另外的改进,编码适配器213可以

[0119] -优化下混/上混增益以获得最大性能/最小伪影;

[0120] -选择具有最佳时间/频率特性的上混参数。

[0121] -不可逆地将音频项目组合成组合音频项目,其然后可以由渲染器115渲染为单个音频项目。这可以利用在渲染侧不需要使所有音频信息单独可用的固有需要。例如,如果不允许对一些输入音频项目进行单独调整(例如,可能要求它们以相同的位置渲染),则不必使音频项目个体地可用。例如,具有相似取向和增益调整约束的多个输入音频对象可以组合成一个基于场景的音频项目,其中,在渲染整个的场景的增益和取向期间仍然可以调整,但先前对象将具有固定的相对音频水平和场景中的固定相对位置。

[0122] -根据针对音频项目的呈现元数据,将不同的比特率预算分配给不同的音频项目。例如,比特率可以基于它们各自表示的未掩蔽信息的量来分配给音频项目。

[0123] 编码电路211然后可以根据由编码适配器213生成的编码控制数据来采用音频项目的编码。例如,编码电路211可以生成一些基于通道、对象和场景的音频项目的比特率降低(例如,量化、参数化等)版本。此外,由于例如作为不同音频项目的编码的一部分的组合或转换,编码音频项目中的至少一些能够表示与输入音频项目不同的音频信息,即输入音频项目和编码音频项目之间可能没有直接对应关系。

[0124] 在一些实施例中,音频编码器205可以具体包括组合器215,其被布置为将输入音频项目组合成一个或多个组合音频项目。组合器215可以具体地将第一和第二输入音频项目组合成组合音频项目。然后可以对组合音频项目进行编码以生成组合编码音频项目,并且该组合编码音频项目可以包括在编码音频数据流中,通常替换第一和第二音频项目。因此,不是个体编码第一和第二音频项目,组合器215可以将它们组合成单个编码音频项目,其然后被包括在编码音频数据流中,而对于第一或第二音频项目不分别包括个体编码音频数据。

[0125] 响应于接收到的呈现元数据而执行音频项目的组合。在许多实施例中,被选择用于组合的音频项目是基于呈现元数据来选择的。例如,编码适配器213可以响应于包括对音频项目的约束满足相似性准则的要求的准则而选择针对组合的音频项目。

[0126] 例如,对于要组合的音频项目,可能要求如由呈现元数据指示的音频项目的约束不必相互矛盾,即其必须可能满足两个约束。因此,可能需要由呈现元数据指示的约束不冲突,并且例如约束至少具有交叠,使得存在至少一个渲染参数值,该值允许满足对被组合的两个(或所有)音频项目的渲染约束。编码适配器213能够要求呈现元数据不描述对公共渲染参数的不兼容约束。

[0127] 例如,呈现元数据可以描述对音频场景中的音频项目的位置的约束。在这种情况下

下,可能要求位置约束必须交叠,并且必须存在一些公共允许位置。

[0128] 要组合的音频项目的选择可以基于针对音频项目的呈现元数据。因此,用于组合的第一和第二音频项目的选择可以基于针对第一和第二音频项目的呈现元数据。例如,如上所述,可能要求针对第一和第二音频项目的呈现元数据不定义冲突约束。

[0129] 在一些实施例中,第一和第二音频项目可以例如被选为具有对例如最相似的同参数的约束的音频项目。例如,可以选择具有基本相同位置约束的音频项目。

[0130] 具体而言,可以确定针对两个音频项目的相似性量度以反映可允许位置之间的交叠。例如,相似性量度可以被生成为交叠的可允许位置的区域的音量相对于针对两个音频项目的个体可允许位置的音量之和之间的比率。

[0131] 作为另一示例,满足针对其位置调整约束的相似性准则的多个音频对象,即使在相应位置范围或空间体积可能不交叠时,也可以组合成基于场景的音频项目,其中,音频源将从其上在基于场景的音频中具有彼此固定的相对取向(即不可单独调整),但它们的取向仍然可以作为整体一起调整。

[0132] 作为另一示例,可以生成相似性量度以反映两个音频项目的交叠增益范围的大小。公共可允许增益范围越大,相似度越大。

[0133] 编码适配器213可以针对不同的音频项目对来评价这样的相似性量度并选择例如相似性量度高于给定阈值的对。然后将可以将这些音频项目组合成单个组合音频项目。

[0134] 在许多实施例中,编码适配器213还被布置为从输入呈现元数据生成针对组合音频项目的组合呈现元数据。然后将该呈现元数据馈送到比特流打包器209,该打包器将其包括在输出编码音频数据流中。

[0135] 元数据电路207可以具体地生成组合呈现元数据,其链接到组合音频项目并且为组合呈现元数据提供渲染约束。然后将可以将生成的具有相关联的组合呈现元数据的组合音频项目视为任何其他音频项目,并且实际上客户端/解码器/渲染器甚至可能不知道组合音频项目确实是通过由音频编码器205对输入音频项目的组合生成的。相反,组合音频项目和相关联的呈现元数据对于客户端侧而言可能与输入音频项目和相关联的呈现元数据无法区分,并且可以被渲染为任何其他音频项目。

[0136] 在许多实施例中,例如可以生成组合呈现元数据以反映针对组合音频项目的呈现参数的约束。可以确定约束,使得其满足被组合的音频项目的个体约束,如由针对这些音频项目的输入呈现元数据所指示的。具体地,针对第一和第二音频项目的组合音频项目的约束可以被确定为满足由针对第一音频的输入呈现元数据指示的第一音频项目的约束和由针对第二音频项目的输入呈现元数据指示的第二音频项目的约束两者的约束。因此,生成组合呈现元数据以提供一个或多个约束,如果满足组合的约束,则确保满足对个体音频项目的个体约束。

[0137] 例如,对于作为音频对象的第一音频项目,输入呈现元数据可以指示它必须以例如从-6dB到0dB的范围内的相对增益并且在(例如,([0,100],[-40,60],[0.5,1.5])的坐标体积内的方位角,仰角,半径)内的位置进行渲染。对于作为音频对象的第二音频项目,输入呈现元数据可以指示它必须以例如从-3dB到3dB的范围内的相对增益并且在例如([-100,80],[-20,70],[0.2,1.0])的(方位角,仰角,半径)的坐标体积内的位置处进行渲染。在这种情况下,可以生成组合呈现元数据以指示作为音频对象的组合音频项目必须以例如从-

3dB到0dB的范围内的相对增益并且在例如([0, 80], [-20, 60], [-0.5, 1.0])坐标体积(方位角, 仰角, 半径)内的位置处进行渲染。这将确保以第一音频项目和第二音频项目两者将可接受的方式渲染组合音频项目。

[0138] 在一些实施例中, 音频编码器205可以被布置为基于针对另一个音频项目的呈现元数据来调整对一个音频项目的压缩。

[0139] 作为低复杂度示例, 一个音频项目的压缩可能取决于针对另一音频项目的接近度和增益/水平。例如, 如果针对当前音频项目的呈现元数据指示位置范围和水平范围, 则可以将这与第二音频项目的位置范围和水平范围进行比较。如果第二音频项目被约束为靠近第一音频项目定位并且被约束为以比第一音频项目实质更高的水平渲染, 则第一音频项目很可能仅被收听者勉强感知。因此, 与不存在其他音频项目的情况相比, 第一音频项目的编码可以具有更高的压缩/比特率降低。具体地, 用于对第一音频项目的编码的比特率分配可以取决于到一个或多个其他音频项目的距离和水平。

[0140] 在一些实施例中, 编码适配器213可以被布置为从第二音频项目估计对第一音频项目的掩蔽效果。掩蔽效果可以由掩蔽量度表示, 所述掩蔽量度指示从对第二音频项目的渲染引入到第一音频项目的掩蔽程度。因此, 掩蔽量度指示在存在第二音频项目的情况下第一音频项目的感知重要性。

[0141] 掩蔽量度可以具体地根据由呈现元数据指示的约束生成为当渲染第二音频项目时从第二音频项接收到的声音的水平相对于从第一音频项目接收到的声音的水平的指示。

[0142] 例如, 可以采用最低增益处的第一音频项目对其最高增益处的第二音频项目的掩蔽效果来估计第二项目的掩蔽水平, 并且反之亦然。

[0143] 作为另一示例, 可以确定第一和第二音频项目之间的最远(或例如平均)距离并且估计它们之间的衰减。然后可以基于对衰减的补偿之后的相对水平差异来估计掩蔽效果。

[0144] 作为另一个示例, 如果系统采用标称收听位置, 则分别来自第一和第二音频项目的收听位置处的信号水平可以基于相对增益水平或信号水平以及与声音源的位置的衰减差来确定。音频项目位置可以从可允许位置中选择, 例如使得掩蔽效果被最小化(针对第一音频项目的最近可允许位置和针对第二音频项目的最远位置)。

[0145] 因此, 编码适配器213可以基于由针对第二音频项目的输入呈现元数据所指示的第二音频项目的增益/水平约束和位置约束并且常常还基于由针对第一音频项目的输入呈现元数据指示的第一音频项目的增益/水平约束和位置约束来估计从第二音频项目对第一音频项目的掩蔽效果。

[0146] 在一些实施例中, 编码适配器213可以基于针对第二音频项目的呈现元数据来直接确定针对第一音频项目的掩蔽阈值, 并且编码电路211可以前进到使用所确定的掩蔽阈值对第一音频项目进行编码。

[0147] 在一些实施例中, 由音频编码器205对编码的调整可以是内部过程, 而没有其他功能被相应地调整。例如, 可以执行将多个音频项目不可逆地组合成组合音频项目, 其中, 组合音频项目被包括在编码音频数据流中并且没有如何创建组合音频项目的指示, 即没有渲染设备执行组合音频项目的任何特定处理。

[0148] 然而, 在许多实施例中, 音频编码器205可以生成编码调整数据, 其指示如何响应于输入呈现元数据而调整编码。该编码调整数据然后可以被包括在编码音频数据流中。在

该方法中,渲染设备109可以相应地具有编码调整的信息并且可以被布置为相应地调整解码和/或渲染。

[0149] 例如,音频编码器205可以生成指示声学环境数据的哪些音频项目实际上是组合音频项目的数据。其还可以指示组合的一些参数并且实际上这些在许多实施例中可以允许渲染设备109生成被组合的原始音频项目的表示。实际上,在一些实施例中,可以将组合音频项目生成为输入音频项目的下混,并且音频编码器205可以生成参数上混数据并将这包括在编码音频数据流中,从而使渲染设备能够执行合理的上混。

[0150] 作为另一示例,这样的解码可能不被调整,但是该信息可以用于与收听者/终端用户的交互。例如,在其调整约束中被称为“接近”的多个音频对象可以由编码器组合成一个基于场景的音频项目,而它们作为“虚拟对象”的存在在编码调整数据中被发信号通知给解码器。然后可以向用户该出该信息,并且可以向用户提供手动控制“虚拟声音源”(尽管仅作为整体,因为它们已经组合在基于场景的音频中),而不是被告知/知道基于场景的音频作为针对虚拟对象的载体。

[0151] 在一些实施例中,呈现元数据可以包括针对一个或多个音频项目的优先级数据,并且音频编码器205可以被布置为响应于针对第一音频项目的优先级指示而调整针对第一音频项目的压缩。

[0152] 优先级指示可以是指示音频场景中的音频项目的感知意义或重要性的渲染优先级指示。例如,其可以用于指示表示主发声者的音频项目比表示例如背景中鸟鸣的音频项目更重要。

[0153] 渲染器115可以基于优先级指示来调整渲染。例如,对于具有降低的听力的收听者,渲染器115可以增加高优先级主对话相对于低优先级背景噪声的增益,从而使语音更易懂。

[0154] 此外,音频编码器205可以增加压缩以降低优先级。例如,为了组合音频项目,可能需要优先级水平必须低于给定水平。作为另一示例,音频编码器205可以组合优先级水平低于给定水平的所有音频项目。

[0155] 在一些实施例中,针对每个音频项目的比特分配可以取决于优先级水平。例如,对不同音频项目的比特分配可以基于考虑包括优先级的多个参数的算法或公式。针对给定音频项目的比特分配可能随着优先级的增加而单调增加。

[0156] 将意识到,为了清楚起见,以上描述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,显而易见,可以使用不同功能电路、单元或处理器之间的任何合适的功能分布而不偏离本发明。例如,图示为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执行。因此,对特定功能单元或电路的引用仅被视为对用于提供所描述的功能的适当模块的引用,而不是指示严格的逻辑或物理结构或组织。

[0157] 本发明可以以任何合适的形式实施,包括硬件、软件、固件或这些的任何组合。本发明可以任选地至少部分地实施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件和部件可以以任何合适的方式在物理上、功能上和逻辑上实施。实际上,该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实施。这样一来,本发明可以在单个单元中实施,或者可以在物理上和功能上分布在不同的单元、电路和处理器之间。

[0158] 通常,音频编码装置、音频编码方法以及实施该方法的计算机程序产品的示例由以下实施例指示。

[0159] 尽管已经结合一些实施例描述了本发明,但并不意在将其限制于本文所阐述的特定形式。相反,本发明的范围仅由所附权利要求限制。此外,尽管特征能够表现为结合特定实施例来描述,但是本领域技术人员将认识到,所描述的实施例的各种特征可以根据本发明进行组合。在权利要求中,术语包括不排除其他元件或步骤的存在。

[0160] 此外,尽管个体地列出,但多个模块、元件、电路或方法步骤可以通过例如单个电路、单元或处理器实施。此外,尽管个体特征可以包括在不同的权利要求中,但这些特征可以可能有利地组合,并且不同权利要求中的包括并不意味着特征的组合是不可行的和/或有利的。而且,在一类权利要求中包括一个特征并不意味着对该类别的限制,而是指示该特征在适当时同样适用于其他权利要求类别。此外,权利要求中的特征的顺序并不意味着任何特定的顺序,其中,特征必须工作,并且特别是方法权利要求中的个体步骤的顺序并不意味着必须以该顺序执行步骤。相反,可以以任何合适的顺序执行步骤。此外,单数引用不排除复数。因此,对“一”、“一个”、“第一”、“第二”等的提及不排除复数。权利要求中的附图标记仅作为阐明示例而提供,而不应解释为以任何方式限制权利要求的范围。

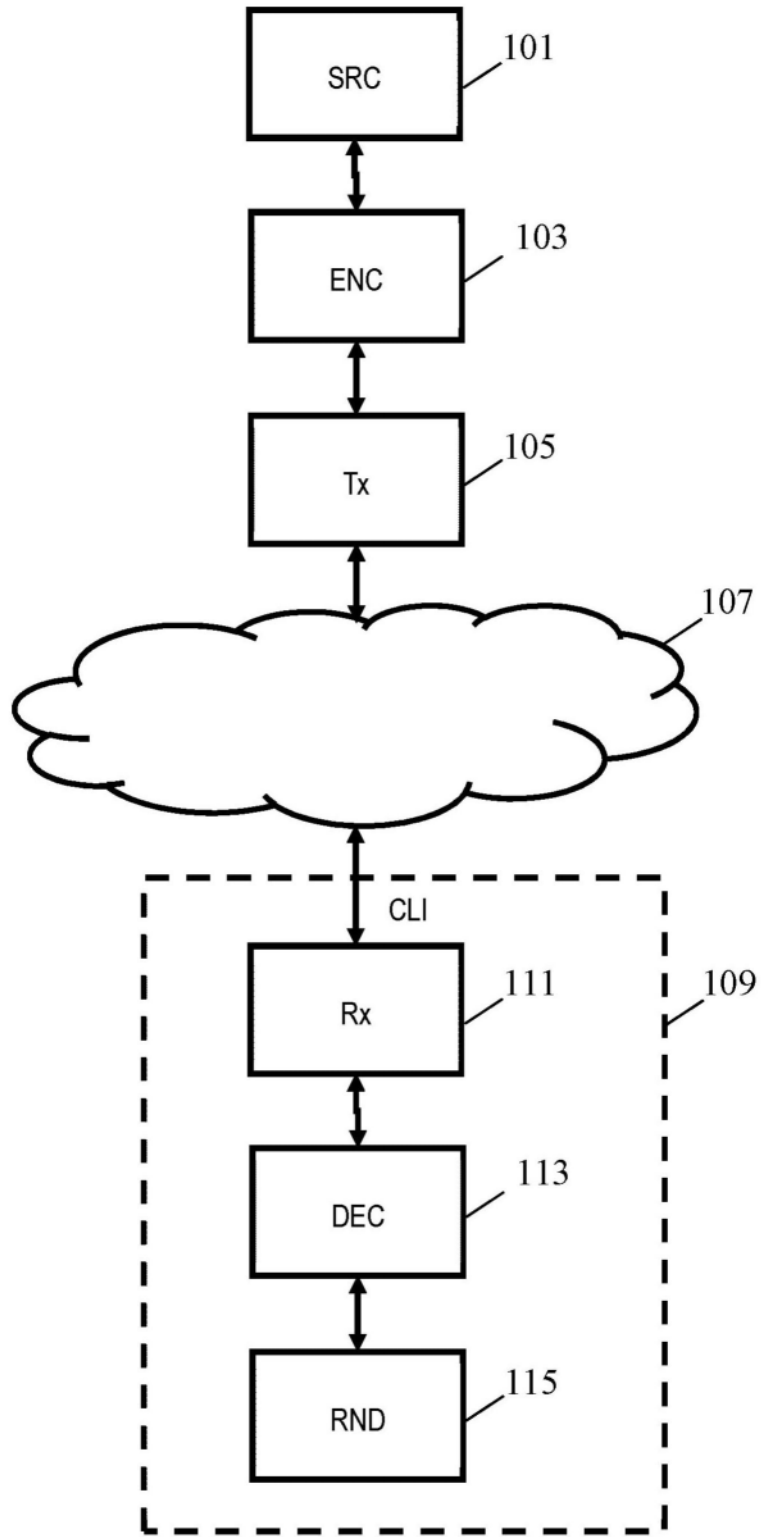
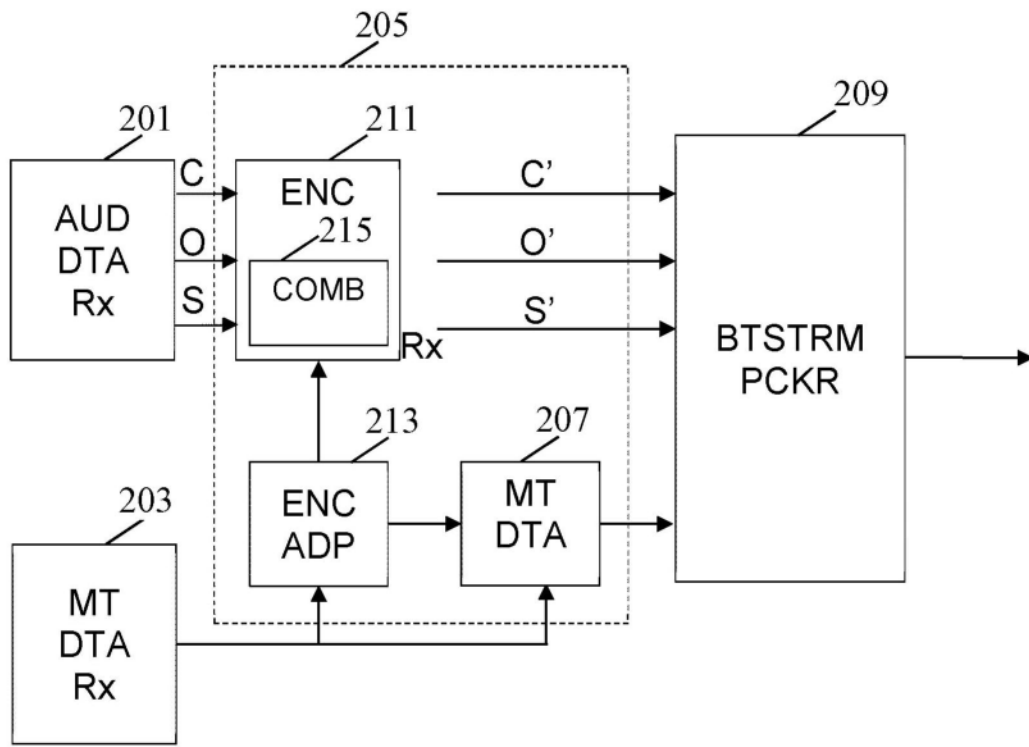


图1



103

图2

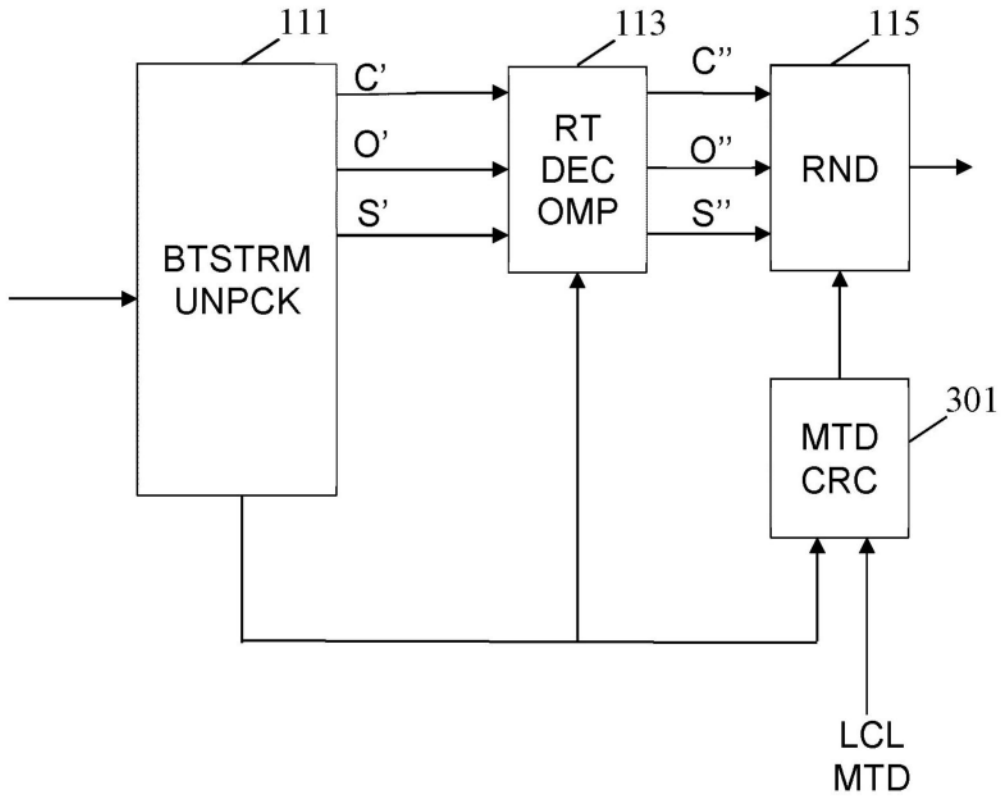


图3