(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105900169 B
(45)授权公告日 2020.01.03

(21)申请号 201580004002.0

(22)申请日 2015.01.05

(65)同一申请的已公布的文献号
申请公布号 CN 105900169 A

(43)申请公布日 2016.08.24

(30)优先权数据
P201430016 2014.01.09 ES
61/951,048 2014.03.11 US

(85)PCT国际申请进入国家阶段日
2016.07.08

(86)PCT国际申请的申请数据
PCT/US2015/010126 2015.01.05

(87)PCT国际申请的公布数据
WO2015/105748 EN 2015.07.16

(73)专利权人 杜比实验室特许公司
地址 美国加利福尼亚
专利权人 杜比国际公司

(72)发明人 D·J·布瑞巴特 陈联武 芦烈
A·M·索尔 N·R·特斯恩高斯

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038
代理人 欧阳帆

(51)Int.Cl.
G10L 19/008(2006.01)
G10L 25/48(2006.01)
H04S 3/00(2006.01)
H04S 7/00(2006.01)

(56)对比文件
CN 101859563 A,2010.10.13,
CN 101582262 A,2009.11.18,
CN 101485202 A,2009.07.15,
CN 101547000 A,2009.09.30,
GB 2459012 A,2009.10.14,
US 2008084934 A1,2008.04.10,

审查员 马杰

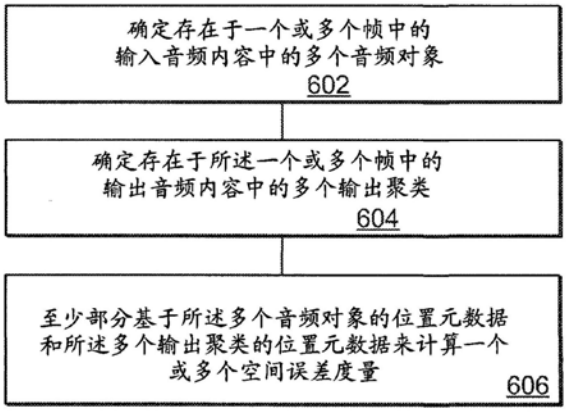
权利要求书2页 说明书19页 附图8页

(54)发明名称

音频内容的空间误差度量

(57)摘要

确定存在于一个或多个帧中的输入音频内容中的音频对象。还确定存在于所述一个或多个帧中的输出音频内容中的输出聚类。这里,输入音频内容中的音频对象被转换成输出音频内容中的输出聚类。至少部分基于音频对象的位置元数据和输出聚类的位置元数据来计算一个或多个空间误差度量。



CN 105900169 B

1.一种方法,包括:

确定存在于一个或多个帧中的输入音频内容中的多个音频对象,其中所述多个音频对象包括$N_{objects}$个音频对象,$N_{objects}>2$;

确定存在于所述一个或多个帧中的输出音频内容中的多个输出聚类,所述输入音频内容中的所述多个音频对象被转换成所述输出音频内容中的所述多个输出聚类,其中所述多个输出聚类包括$N_{clusters}$个输出聚类,$N_{objects}>N_{clusters}>1$;以及

至少部分基于所述多个音频对象的位置元数据和所述多个输出聚类的位置元数据来计算一个或多个空间误差度量,其中所述一个或多个空间误差度量至少部分取决于对象重要度;

其中,所述方法由一个或多个计算装置执行。

2.根据权利要求1所述的方法,其中,所述对象重要度是通过对以下中的一个或多个进行分析而获得的:所述多个音频对象中的音频数据、所述多个输出聚类中的音频数据、所述多个音频对象中的元数据、所述多个输出聚类中的元数据,或者其中,所述对象重要度的至少一部分是基于用户输入而确定的。

3.根据权利要求1所述的方法,其中,所述多个音频对象中的至少一个音频对象被分配到所述多个输出聚类中的两个或更多个输出聚类或者被分配到所述多个输出聚类中的一个输出聚类。

4.根据权利要求1所述的方法,还包括:

基于所述一个或多个空间误差度量来确定通过将所述输入音频内容中的所述多个音频对象转换成所述输出音频内容中的所述多个输出聚类引起的感知音频质量劣化。

5.根据权利要求4所述的方法,其中,所述感知音频质量劣化由与感知音频质量测试相关的一个或多个预测测试得分表示。

6.根据权利要求1所述的方法,其中,所述一个或多个空间误差度量包括帧内空间误差度量,所述帧内空间误差度量包括以下中的至少一个:以对象重要度加权的帧内对象位置误差度量、以对象重要度加权的帧内对象平移误差度量、经规范化的以对象重要度加权的帧内对象位置误差度量、经规范化的以对象重要度加权的帧内对象平移误差度量。

7.根据权利要求1所述的方法,其中,所述一个或多个空间误差度量包括帧间空间误差度量,所述帧间空间误差度量包括基于增益系数流并以对象重要度加权的帧间空间误差度量。

8.根据权利要求1所述的方法,其中,所述多个音频对象经由多个增益系数与所述多个输出聚类相关。

9.根据权利要求1所述的方法,其中,每个帧对应于所述输入音频内容中的第一时间段和所述输出音频内容中的第二时间段;并且其中,存在于所述输出音频内容中的第二时间段中的输出聚类被存在于所述输入音频内容中的第一时间段中的音频对象映射到。

10.根据权利要求1所述的方法,还包括:

构造一个或多个用户界面部件,所述一个或多个用户界面部件表示以下中的一个或多个:所述多个音频对象中的音频对象、收听空间中的所述多个输出聚类中的输出聚类;

使所述一个或多个用户界面部件被显示给用户。

11.根据权利要求1所述的方法,还包括:

构造一个或多个用户界面部件，所述一个或多个用户界面部件表示以下中的一个或多个：所述多个音频对象中的音频对象的各自的对象重要度、所述多个输出聚类中的输出聚类的各自的对象重要度、所述多个音频对象中的音频对象的各自的响度、所述多个输出聚类中的输出聚类的各自的响度、所述多个音频对象中的音频对象的语音或对话内容的各自的概率、所述多个输出聚类中的输出聚类的语音或对话内容的概率；

使所述一个或多个用户界面部件被显示给用户。

12.根据权利要求1所述的方法，还包括：

构造一个或多个用户界面部件，所述一个或多个用户界面部件表示以下中的一个或多个：所述一个或多个空间误差度量、至少部分基于所述一个或多个空间误差度量而确定的一个或多个预测测试得分；

使所述一个或多个用户界面部件被显示给用户。

13.根据权利要求1所述的方法，其中，所述多个输出聚类中的输出聚类包括被所述多个音频对象中的两个或更多个音频对象映射到的部分。

14.一种包括处理器并且被配置为执行权利要求1-13中所述的方法中的任何一种方法的设备。

15.一种存储有软件指令的非暂时性计算机可读存储介质，所述软件指令当被一个或多个处理器执行时使得执行权利要求1-13中所述的方法中的任何一种方法。

音频内容的空间误差度量

[0001] 相关申请的交叉引用

[0002] 本申请要求在2014年1月9日提交的西班牙专利申请No.P201430016和在2014年3月11日提交的美国临时专利申请No.61/951048的优先权，每个申请的全部内容都通过引用并入于此。

技术领域

[0003] 本发明一般涉及音频信号处理，更具体地涉及确定与音频对象的格式转换、渲染、聚类(cluster)、再混合或组合相关联的空间误差度量和音频质量劣化。

背景技术

[0004] 诸如原始创作/制作的音频内容等之类的输入音频内容可能包括分别以音频对象格式表示的大量音频对象。输入音频内容中的大量音频对象可以被用来创建空间多样化的、沉浸式的、准确的音频体验。

[0005] 然而，对包括大量音频对象的输入音频内容的编码、解码、传输、回放等可能需要高带宽、大存储缓冲区、高处理能力等。按照某些方法，输入音频内容可以被变换为包括较少音频对象的输出音频内容。同一个输入音频内容可以被用来产生与许多不同的音频内容分发、传输和回放设置对应的许多不同的输出音频内容版本，诸如与蓝光盘、广播(例如，有线的、卫星的、地面站的，等等)、移动(例如，3G、4G等)、互联网等相关的输出音频内容版本。每个输出音频内容版本可以特别地适合于相应设置，以解决该设置中对于一般性地导出的音频内容的高效率表示、处理、传输和渲染的特别挑战。

[0006] 本部分中所描述的方法是可以寻求的方法，但不一定是之前已经设想或寻求过的方法。因此，除非另有指示，否则不应仅仅因为在本部分中提到了就认为本部分中所述的任何方法是现有技术。类似地除非另有指示，否则针对一种或多种方法认定的问题不应基于本部分就认为在任何现有技术中已经认识到。

附图说明

[0007] 在附图中以举例的方式而非限制的方式例示了本发明，在附图中相似的附图标记指代相似的要素，其中：

[0008] 图1例示了音频对象聚类中所涉及的示例性的、由计算机实现的模块；

[0009] 图2例示了示例性的空间复杂度分析器；

[0010] 图3A至图3D例示了用于可视化一个或多个帧的空间复杂度的示例性用户界面；

[0011] 图4例示了两个示例性的视觉复杂度计量器实例；

[0012] 图5例示了用于计算增益流的示例性场景；

[0013] 图6例示了示例性的处理流程；以及

[0014] 图7例示了在其上可以实现本文中所描述的计算机或计算装置的示例性硬件平台。

具体实施方式

[0015] 本文中描述了与确定有关于音频对象聚类的空间误差度量和音频质量劣化相关的示例性实施例。在以下描述中，为了说明的目的，阐述了许多具体细节以便提供对本发明的透彻理解。然而，显而易见的是本发明可以在没有这些具体细节的情况下实施。在其他情况下，未详尽地描述公知的结构和设备，以避免不必要地封闭、模糊或混淆本发明。

[0016] 在本文中根据以下大纲来描述示例性实施例：

[0017] 1.总体概述

[0018] 2.音频对象聚类

[0019] 3.空间复杂度分析器

[0020] 4.空间误差度量

[0021] 4.1帧内对象位置误差

[0022] 4.2帧内对象平移误差

[0023] 4.3重要度加权的误差度量

[0024] 4.4规范化的误差度量

[0025] 4.5帧间空间误差

[0026] 5.主观音频质量的预测

[0027] 6.空间误差和空间复杂度的可视化

[0028] 7.示例性的处理流程

[0029] 8.实现机制——硬件概述

[0030] 9.等同、扩展、替代及其他

[0031] 1.总体概述

[0032] 本概述呈现了本发明的实施例的一些方面的基本描述。应注意，本概述不是对实施例的各方面的全面的或详尽的总结。另外，应注意，本概述并非意在被理解为认定实施例的任何特别重要的方面或要素，也不被理解为特别地叙述实施例的任何范围或概括地叙述本发明。本概述仅仅以扼要且简化的格式呈现了与示例性实施例相关的一些构思，并且应被理解为仅仅是以下对示例性实施例的更详细的描述的概念性前言。

[0033] 可能存在各种各样的基于音频对象的音频格式，这些基于音频对象的音频格式可以从一种格式变换、下混、转换、转码成另一种格式。在一个示例中，一种格式可以利用笛卡尔坐标系来描述音频对象或输出聚类的位置，而其他格式可以利用可能随距离增加的角度方法。在另一个示例中，为了高效地存储和传输基于对象的音频内容，可以对输入音频对象集合执行音频对象聚类以使相对较多的输入音频对象减少至相对较少的输出音频对象或输出聚类。

[0034] 本文中所描述的技术可以被用来确定与构成输入音频内容的(例如，动态的、静态的等)音频对象集合到构成输出音频内容的另一个音频对象集合的格式转换、渲染、聚类、再混合或组合等相关联的空间误差度量和/或空间质量劣化。仅仅为了说明的目的，输入音频内容中的音频对象或输入音频对象有时被简称为“音频对象”。输出音频内容中的音频对象或输出音频对象一般可被称为“输出聚类”。应注意，在各种实施例中，术语“音频对象”和“输出聚类”是与将音频对象转换到输出聚类的特定转换操作相关地使用的。例如，一个转换操作中的输出聚类很可能是后一转换操作中的输入音频对象；类似地，当前转换操作中

的输入音频对象很可能是前一转换操作中的输出聚类。

[0035] 如果输入音频对象相对较少或稀疏，则从输入音频对象到输出聚类的一对一映射对于输入音频对象中的至少一些输入音频对象是可能的。

[0036] 在一些实施例中，音频对象可以表示固定位置处的一个或多个声音元素(例如，音频床(audio bed)或音频床的一部分、物理声道等)。在一些实施例中，输出聚类也可以表示固定位置处的一个或多个声音元素(例如，音频床或音频床的一部分、物理声道等)。在一些实施例中，具有动态位置(或非固定位置)的输入音频对象可以被聚类成具有固定地点的输出聚类。在一些实施例中，具有固定位置的输入音频对象(例如，音频床、音频床的一部分等)可以被映射到具有固定位置的输出聚类(例如，音频床、音频床的一部分等)。在一些实施例中，所有输出聚类都具有固定位置。在一些实施例中，输出聚类中的至少一个输出聚类具有动态位置。

[0037] 当输入音频内容中的输入音频对象被转换成输出音频内容中的输出聚类时，输出聚类的数量可以少于或者可以不少于音频对象的数量。输入音频内容中的音频对象可以被分配到输出音频内容中的多于一个的输出聚类中。音频对象也可以仅被分配到可以或者可以不位于与该音频对象所在的位置相同的位置处的输出聚类。音频对象的位置到输出聚类的位置的移位引起空间误差。本文中所描述的技术可以被用来确定与由于从输入音频内容中的音频对象到输出音频内容中的输出聚类的转换而导致的空间误差相关的空间误差度量和/或音频质量劣化。

[0038] 按照如本文所描述的技术确定的空间误差度量和/或音频质量劣化可以作为测量由有损编解码器引起的编码误差、量化误差等的其他质量度量(例如，PEAQ等)的附加或替代而被使用。在示例中，空间误差度量、音频质量劣化等可以与音频对象或输出聚类中的位置元数据和其他元数据一起用于视觉地传达多声道的基于多对象的音频内容中的音频内容的空间复杂度。

[0039] 附加地、可选地或可替代地，在一些实施例中，音频质量劣化可以以基于一个或多个空间误差度量而生成的预测测试得分的形式被提供。预测测试得分可以被用作输出音频内容或输出音频内容的部分(例如，在一个帧中，等等)相对于输入音频内容的感知音频质量劣化的指示，而无需实际进行对输入音频内容和输出音频内容的感知音频质量的任何用户调查。预测测试得分可以与诸如MUSHRA(隐藏参考和基准的多刺激)测试、MOS(平均意见得分)测试等主观音频质量测试有关。在一些实施例中，一个或多个空间误差度量通过使用根据一个或多个代表性的训练音频内容数据集合确定/优化的预测参数(例如，相关因子等)而被转换为一个或多个预测测试得分。

[0040] 例如，训练音频内容数据集合中的每个元素(或摘录)可以在该元素(或摘录)中的输入音频对象被转换或映射成对应的输出聚类之前和之后经受感知音频质量的主观用户调查。根据用户调查确定的测试得分可以与基于该元素(或摘录)中的输入音频对象和对应的输出聚类计算的空间误差度量相关，以用于确定或优化预测参数的目的，预测参数然后可以被用来对不一定在训练数据集合中的音频内容预测测试得分。

[0041] 按照如本文所描述的技术的系统可以被配置为以客观的方式将空间误差度量和/或音频质量劣化提供给指导将输入音频内容(中的音频对象)转换成输出音频内容(中的输出聚类)的处理、操作、算法等的音频工程师。出于减轻或防止音频质量劣化的目的，该系统

可以被配置为接受用户输入或者从音频工程师接收反馈，以优化该处理、操作、算法等，从而使得显著地影响输出音频内容的音频质量的空间误差最小化，等等。

[0042] 在一些实施例中，对象重要度是针对单个的音频对象或输出聚类估计或确定的，并且被用于估计空间复杂度和空间误差。例如，就相对响度和位置接近度而言为静默的或者被其他音频对象遮掩的音频对象可能由于为这种音频对象分配较低的对象重要度而经受较大的空间误差。由于较不重要的音频对象与在场景中更为主导的其他音频对象截然相比是相对安静的，所以较不重要的音频对象的较大空间误差可能造成很小的听得见的噪声(artifact)。

[0043] 如本文所描述的技术可以被用来计算帧内空间误差度量以及帧间空间误差度量。帧内空间误差度量的示例包括但不限于以下中的任何一个：对象位置误差度量、对象平移误差、以对象重要度加权的空间误差度量、经规范化的以对象重要度加权的空间误差度量等。在一些实施例中，帧内空间误差度量可以基于以下方面被计算为客观质量度量：(i)音频对象中的音频样本数据，包括但不限于音频对象在它们各自的上下文下的个体对象重要度；以及(ii)转换之前的音频对象的原始位置和转换之后的音频对象的重构位置之间的差异。

[0044] 帧间空间误差度量的示例包括但不限于：与(在时间上)相邻帧中的输出聚类的增益系数差值和位置差值的乘积相关的帧间空间误差度量、与(在时间上)相邻帧中的增益系数流相关的帧间空间误差度量。帧间空间误差度量对于指示(在时间上)相邻帧中的不一致性可能特别有用；例如，由于在从一个帧到下一个帧的插值期间造成的帧间空间误差，在时间上相邻的帧之间的音频对象到输出聚类分派/分配的变化可能导致听得见的噪声。

[0045] 在一些实施例中，可以基于以下项来计算帧间空间误差度量：(i)随着时间(例如，两个相邻帧之间，等等)的与输出聚类相关的增益系数差值；(ii)输出聚类随着时间的位置变化(例如，当音频对象被平移到聚类中时，音频对象至输出聚类的相应平移矢量改变)；(iii)音频对象的相对响度；等等。在一些实施例中，可以至少部分基于输出聚类之间的增益系数流来计算帧间空间误差度量。

[0046] 如本文所描述的空间误差度量和/或音频质量劣化可以被用来驱动一个或多个用户界面与用户交互。在一些实施例中，在用户界面中提供视觉复杂度计量器以显示出音频对象集合相对于这些音频对象被转换成的输出聚类集合的空间复杂度(例如，高质量/低空间复杂度、低质量/高空间复杂度等)。在一些实施例中，视觉空间复杂度计量器显示音频质量劣化的指示(例如，与感知MOS测试、MUSHRA测试相关的预测测试得分，等等)以作为将输入音频对象转换到输出聚类的相应转换处理的反馈。空间误差度量和/或音频质量劣化的值可以通过使用VU计量器、条形图、夹灯(clip light)、数值指示符、其他视觉部件等而被可视化在显示器上的用户界面中，以视觉地传达与转换处理相关联的空间复杂度和/或空间误差度量。

[0047] 在一些实施例中，如本文所描述的机制形成媒体处理系统的一部分，所述媒体处理系统包括但不限于以下中的任何一个：手持装置、游戏机、电视、家庭影院系统、机顶盒、平板、移动装置、膝上型计算机、上网本计算机、蜂窝无线电电话、电子书阅读器、销售点终端、台式计算机、计算机工作站、计算机亭、各种其他种类的终端和媒体处理单元等。

[0048] 对于本领域技术人员来说，本文中所描述的优选实施例的各种变型以及一般性原

理和特征将是容易明白的。因此，本公开并非意在限于所示出的实施例，而应被赋予与本文中所描述的原理和特征一致的最宽泛的范围。

[0049] 如本文所描述的任何实施例可以单独使用或者按任何组合与另一个实施例一起使用。尽管各种实施例可能是由在本说明书中的一个或多个地方可能讨论或暗示的现有技术的各种缺陷驱使的，但是实施例不一定解决这些缺陷中的任何缺陷。换句话说，不同实施例可以解决在本说明书中可能讨论的不同缺陷。一些实施例可以仅部分解决在本说明书中可能讨论的一些缺陷或者仅仅一个缺陷，并且一些实施例可能不解决这些缺陷中的任何缺陷。

[0050] 2.音频对象聚类

[0051] 音频对象可以被认为是可以被感知为源自于收听空间(或环境)中的特定物理地点或多个特定物理地点的单个声音元素或声音元素集。音频对象的示例包括但不仅限于以下中的任何一个：音频制作会话中的音轨等。音频对象可以是静态的(例如，静止的)或动态的(例如，运动的)。音频对象包括与表示一个或多个声音元素的音频样本数据分开的元数据。该元数据包括定义声音元素中的一个或多个声音元素在给定时间点(例如，在一个或多个帧中、在帧的一个或多个部分中、等等)的一个或多个位置(例如，动态的或固定的形心(centroid)位置、扬声器在收听空间中的固定位置、一组表示周围效果的一个、两个或更多个动态的或固定的位置等)。在一些实施例中，当音频对象被回放时，它是通过使用存在于实际回放环境中的扬声器并根据其位置元数据被渲染的，而不是一定被输出到由上游音频编码器采取的参考音频声道配置的预定义物理声道，所述上游音频编码器将音频对象编码为有利于下游音频解码器的音频信号。

[0052] 图1例示了用于音频对象聚类的示例性的、由计算机实现的模块。如图1中所示，共同表示输入音频内容的输入音频对象102通过音频对象聚类处理106被转换成输出聚类104。在一些实施例中，输出聚类104共同表示输出音频内容并且构成输入音频内容的比输入音频对象更紧凑的表示(例如，更少的音频对象等)，从而使得可以降低存储要求和传输要求并且降低用于再现输入音频内容的计算要求和存储器要求，尤其对于具有有限处理能力、有限电池功率、有限通信能力、有限再现能力等的消费者领域装置。然而，音频对象聚类导致一定量的空间误差，因为并非所有输入音频对象在与其他音频对象聚合时都可以保持空间保真度，尤其是在存在大量稀疏分布的输入音频对象的实施例中。

[0053] 在一些实施例中，音频对象聚类处理106至少部分基于根据输入音频对象的样本数据、音频对象元数据等中的一个或多个产生的对象重要度108来对输入音频对象102进行聚类。样本数据、音频对象元数据等被输入到对象重要度估计器110，对象重要度估计器110产生供音频对象聚类处理106使用的对象重要度108。

[0054] 如本文所描述的，对象重要度估计器110和音频对象聚类处理106可以作为时间的函数执行。在一些实施例中，用输入音频对象102编码的音频信号或者用根据输入音频对象102产生的输出聚类104编码的相应音频信号可以被分割为单个的帧(例如，持续时间为诸如20毫秒的单元，等)。这种分割可以被应用于时域波形上，但是也是通过使用滤波器组，或者可以应用于任何其他的变换域上。对象重要度估计器(110)可以被配置为产生输入音频对象(102)关于输入音频对象(102)的一个或多个特性上的各个对象重要度，所述特性包括但不限于内容类型、局部响度等。

[0055] 如本文所描述的局部响度可以表示音频对象在一组、一批、一群、多个、一簇音频对象等的上下文下根据心理声学原理的(相对)响度。音频对象的局部响度可以用于确定音频对象的对象重要度,以便当渲染系统不具有足以单个地渲染所有音频对象的能力时选择性地渲染音频对象,等等。

[0056] 音频对象在给定时间(例如,逐个帧地、在一个或多个帧中、在帧的一个或多个部分中,等等)可以被分类为若干种(例如,定义的)内容类型之一,诸如对话、音乐、周围环境、特殊效果等。音频对象可以在其整个持续时间期间改变内容类型。(例如,一个或多个帧、帧的一个或多个部分等中的)音频对象可以被分配该音频对象在帧中为特定内容类型的概率。在示例中,持续对话类型的音频对象可以被表示为百分之百的概率。在另一个示例中,从对话类型变换为音乐类型的音频对象可以被表示为50%对话/50%音乐、或者对话和音乐类型的不同百分比组合。

[0057] 音频对象聚类处理106或按音频对象聚类处理106操作的模块可以被配置为逐个帧地确定音频对象的内容类型(例如,被表示为具有含布尔值的分量的矢量等)以及音频对象的内容类型的概率(例如,被表示为具有百分比值的分量的矢量等)。基于音频对象的内容类型,音频对象聚类处理106可以被配置为:逐个帧地、在一个或多个帧中、在帧的一个或多个部分中,将音频对象聚类到特定输出聚类中,分配音频对象和输出聚类之间的相互一对一映射等。

[0058] 出于例示的目的,存在于第m帧中的多个音频对象(例如,输入音频对象102等)当中的第i音频对象可以用相应的函数$x_i(n,m)$表示,其中,n是表示第m帧中的多个音频数据样本当中的第n音频数据样本的索引。诸如第m帧等的帧中的音频数据样本的总数取决于音频信号被采样以创建音频数据样本的采样速率(例如,48kHz等)。

[0059] 在一些实施例中,如以下表达式中所示,(例如,在音频对象聚类处理中,等等)第m帧中的多个音频对象基于线性运算而被聚类成多个输出聚类$y_j(n,m)$:

[0060] $$y_j(n,m)=\sum_i g_{ij}x_i(n,m) \quad (1)$$

[0061] 其中,$g_{ij}(m)$表示对象i到聚类j的增益系数。为了避免输出聚类$y_j(n,m)$中的不连续,可以在加窗的、部分重叠的帧上执行聚类操作以对跨帧的$g_{ij}(m)$的变化进行插值。如本文所使用的,增益系数表示特定输入音频对象的一部分到特定输出聚类的分配。在一些实施例中,音频对象聚类处理(106)被配置为产生用于根据表达式(1)将输入音频对象映射成输出聚类的多个增益系数。可替代地、附加地或可选地,增益系数$g_{ij}(m)$可以跨样本(n个)进行插值以创建插值增益系数$g_{ij}(m,n)$。可替代地,增益系数可以是频率相关的。在这种实施例中,输入音频也通过使用合适的滤波器组被划分为频带,并且可能不同的增益系数集合被应用于每个划分的音频。

[0062] 3.空间复杂度分析器

[0063] 图2例示了示例性的空间复杂度分析器200,空间复杂度分析器200包括若干个由计算机实现的模块,诸如帧内空间误差分析器204、帧间空间误差分析器206、音频质量分析器208、用户界面模块210等。如图2中所示,空间复杂度分析器200被配置为接收/收集音频对象数据202,将针对关于输入音频对象集合(例如,图1的102,等等)和这些输入音频对象被转换成的输出聚类集合(例如,图1的104,等等)的空间误差和音频质量劣化来分析所述音频对象数据202。音频对象数据202包括以下中的一个或多个:用于输入音频对象(102)的

元数据、用于输出聚类(104)的元数据、如表达式(1)中所示的将输入音频对象(102)映射到输出聚类(104)的增益系数、输入音频对象(102)的局部响度、输入音频对象(102)的对象重要度、输入音频对象(102)的内容类型、输入音频对象(102)的内容类型的概率等。

[0064] 在一些实施例中,帧内空间误差分析器(204)被配置为逐个帧地基于音频对象数据(202)确定一种或多种类型的帧内空间误差度量。在一些实施例中,对于每个帧,帧内空间误差分析器(204)被配置为:(i)从音频对象数据(202)提取增益系数、输入音频对象(102)的位置元数据、输出聚类(102)的位置元数据等;(ii)基于从帧中的输入音频对象中的音频对象数据(202)提取的数据,针对帧中的每个输入音频对象分别计算所述一种或多种类型的帧内空间误差度量中的每个帧内空间误差度量;等等。

[0065] 帧内空间误差分析器(204)可以被配置为基于针对输入音频对象(102)分别计算的空间误差来对所述一种或多种类型的帧内空间误差度量中的对应类型计算总体每帧空间误差度量等。总体每帧空间误差度量可以通过用权重因子对单个音频对象的空间误差进行加权来计算,所述权重因子诸如是帧中的输入音频对象(102)的各自的对象重要度等。附加地、可选地或可替代地,总体每帧空间误差度量可以用与权重因子之和相关的规范化因子来规范化,所述权重因子之和诸如是指示帧中的输入音频对象(102)的各自的对象重要度等的值之和。

[0066] 在一些实施例中,帧间空间误差分析器(206)被配置为基于两个或更多个相邻帧的音频对象数据(202)来确定一种或多种类型的帧间空间误差度量。在一些实施例中,对于两个相邻帧,帧间空间误差分析器(206)被配置为:(i)从音频对象数据(202)提取增益系数、输入音频对象(102)的位置元数据、输出聚类(102)的位置元数据等;(ii)基于从帧中的输入音频对象中的音频对象数据(202)提取的数据,针对帧中的每个输入音频对象分别计算所述一种或多种类型的帧间空间误差度量中的每个帧间空间误差度量;等等。

[0067] 帧间空间误差分析器(206)可以被配置为对于两个或更多个相邻帧,基于针对帧中的输入音频对象(102)分别计算的空间误差来对所述一种或多种类型的帧间空间误差度量中的对应类型计算总体空间误差度量等。总体空间误差度量可以通过用权重因子对单个的音频对象的空间误差进行加权来计算得到,所述权重因子诸如是帧中的输入音频对象(102)的各自的对象重要度等。附加地、可选地或可替代地,总体空间误差度量可以用规范化因子(例如,与帧中的输入音频对象(102)的各自的对象重要度相关的规范化因子)来规范化。

[0068] 在一些实施例中,音频质量分析器(208)被配置为基于例如由帧内空间误差分析器(204)或帧间空间误差分析器(206)产生的帧内空间误差度量或帧间空间误差度量中的一个或多个来确定感知音频质量。在一些实施例中,感知音频质量由基于所述一个或多个空间误差度量产生的一个或多个预测测试得分指示。在一些实施例中,预测测试得分中的至少一个与对音频质量的主观评估测试(诸如MUSHRA测试、MOS测试等)相关。音频质量分析器(208)可以用根据一个或多个训练数据集合等预先确定的预测参数(例如,相关因子等)来配置。在一些实施例中,音频质量分析器(208)被配置为基于预测参数来将所述一个或多个空间误差度量转换为一个或多个预测测试得分。

[0069] 在一些实施例中,空间复杂度分析器(200)被配置为将根据本文所描述的技术确定的空间误差度量、音频质量劣化、空间复杂度等中的一个或多个作为输出数据212提供给

用户或其他装置。附加地、可选地或可替代地，在一些实施例中，空间复杂度分析器(200)可以被配置为接收用户输入214，用户输入214向在将输入音频内容转换为输出音频内容时使用的处理、算法、操作参数等提供反馈或改变。这种反馈的示例是对象重要度。附加地、可选地或可替代地，在一些实施例中，空间复杂度分析器(200)可以被配置为例如基于在用户输入214中接收到的反馈或改变或者基于估计的空间音频质量来将控制数据216发送给在将输入音频内容转换为输出音频内容时使用的处理、算法、操作参数等。

[0070] 在一些实施例中，用户界面模块(210)被配置为通过一个或多个用户界面与用户交互。用户界面模块(210)可以被配置为通过用户界面向用户呈现或者使得向用户显示描绘输出数据212中的一些或全部的用户界面部件。用户界面模块(210)可以被进一步配置为通过所述一个或多个用户界面接收用户输入214中的一些或全部。

[0071] 4.空间误差度量

[0072] 可以基于单个帧或多个相邻帧中的总体空间误差来计算多个空间误差度量。在确定/估计总体空间误差度量和/或总体音频质量劣化时，对象重要度可以起到主要作用。相比于在当前场景中占主导的音频对象，(例如，就响度、空间邻近度等而言)静默的、相对静默的或者被其他音频对象(部分)遮掩的音频对象可以经受更大的空间误差，直到音频对象聚类的噪声变得可听见。出于例示的目的，在一些实施例中，具有索引i的音频对象具有各自的对象重要度(其被表示为N_i)。该对象重要度可以由对象重要度估计器(图1的110)基于若干个性质产生，所述性质包括但不仅限于以下中的任何一个：根据感知响度模型的相对于音频床和其他音频对象的局部响度的音频对象的局部响度、语义信息(诸如是对话的概率)等。考虑到音频内容的动态本性，第i音频对象的对象重要度$N_i(m)$典型地作为时间的函数而变化，例如，作为帧索引m的函数(帧索引m逻辑地表示或者映射到诸如媒体回放时间等之类的时间)而变化。另外，对象重要度度量可以依赖于对象的元数据。这种依赖性的示例是基于对象的位置或运动速度而对对象重要度进行的修改。

[0073] 对象重要度可以被定义为时间和频率的函数。如本文所描述的，转码、重要度估计、音频对象聚类等可以通过使用任何合适的变换(诸如离散傅立叶变换(DFT)、正交镜像滤波器(QMF)组、(修正)离散余弦变换(MDCT)、听觉滤波器组、类似的变换处理等)而在频带中执行。不失一般性地，第m帧(或者具有帧索引m的帧)包括在时域中的或者在合适的变换域中的音频样本集合。

[0074] 4.1帧内对象位置误差

[0075] 帧内空间误差度量中的一个帧内空间误差度量与对象位置误差相关，并且可以被表示为帧内对象位置误差度量。

[0076] 表达式(1)中的每个音频对象(例如，第i音频对象等)对于每个帧(例如，m等)具有相关联的位置矢量(例如，$\vec{p}_i(m)$等)。类似地，表达式(1)中的每个输出聚类(例如，第j输出聚类等)也具有相关联的位置矢量(例如，$\vec{p}_j(m)$等)。这些位置矢量可以由空间复杂度分析器(例如，200等)基于音频对象数据(202)中的位置元数据来确定。音频对象的位置误差可以用该音频对象的位置和被分配到输出聚类的该音频对象的质心的位置之间的距离表示。在一些实施例中，第i音频对象的质心的位置被确定为该音频对象被分配到的输出聚类的位置与充当权重因子的增益系数$g_{ij}(m)$的加权和。音频对象的位置和被分配到输出聚类的

该音频对象的质心的位置之间的距离的平方可以用如下表达式计算：

[0077] $$E_i^2(m) = \left|\vec{p}_i(m) - \frac{\sum_j g_{ij}(m)\vec{p}_j(m)}{\sum_j g_{ij}(m)}\right|^2 \quad (2)$$

[0078] 表达式右侧(RHS)的输出聚类的位置的加权和表示第i音频对象的被感知位置。$E_i(m)$可以被称为第i音频对象在帧m中的帧内对象位置误差。

[0079] 在示例性实现中，增益系数(例如，$g_{ij}(m)$等)通过优化用于每个音频对象(例如，第i音频对象等)的成本函数而被确定。被用来获得表达式(1)中的增益系数的成本函数的示例包括但不限于以下中的任何一个：$E_i(m)$、不同于$E_i(m)$的L2范数。应注意，本文所描述的技术可以被配置为使用通过用不同于$E_i(m)$的其他类型的成本函数进行优化而获得的增益系数。

[0080] 在一些实施例中，由$E_i(m)$表示的帧内对象位置误差仅对于位置在输出聚类的凸包外部的音频对象才会很大，而对于位置在凸包内部的音频对象为零。

[0081] 4.2帧内对象平移误差

[0082] 即使在如表达式(2)中表示的音频对象的位置误差为零(例如，在输出聚类的凸包内，等等)的情况下，与在没有聚类的情况下直接渲染该音频对象相比，该音频对象在聚类和渲染之后也仍可能听起来显著不同。如果聚类形心的地点都不在音频对象的位置附近，则这种情况可能会出现，因此音频对象(例如，样本数据部分、表示音频对象的信号等)被分布在各种输出聚类之间。与第i音频对象在帧m中的帧内对象平移误差相关的误差度量可以用如下表达式表示：

[0083] $$F_i^2(m) = \sum_j g_{ij}^2(m)\left|\vec{p}_i(m) - \vec{p}_j(m)\right|^2 \quad (3)$$

[0084] 在通过质心优化来计算表达式(1)中的增益系数$g_{ij}(m)$的一些实施例中，如果输出聚类之一(例如，第j输出聚类等)的位置$\vec{p}_j$与对象位置$\vec{p}_i$重合，则表达式(3)中的误差度量$F_i^2(m)$为零。然而，在没有这种重合的情况下，将对象平移到输出聚类的形心导致$F_i^2(m)$为非零值。

[0085] 4.3重要度加权的误差度量

[0086] 在一些实施例中，空间复杂度分析器(200)被配置为用(例如，基于局部响度N_i等确定的)各自的对象重要度来对场景中的每个音频对象的单个对象误差度量(例如，E_i、F_i等)进行加权。对象重要度、局部响度N_i等可以由空间复杂度分析器(200)根据接收的音频对象数据(202)来估计或确定。用各自的对象重要度加权的对象误差度量可以被总计，以产生如以下表达式中所示的关于所有音频对象的总体误差度量：

[0087] $$A_{E_i}(m) = \sum_i E_i(m)N_i(m)$$

[0088] $$A_{F_i}(m) = \sum_i F_i(m)N_i(m) \quad (4)$$

[0089] 可替代地、附加地或可选地，场景中的每个音频对象的单个误差度量(例如，E_i、F_i

等)可以被总计,以产生如以下表达式中所示的关于场景中的所有音频对象的在平方域中的总体误差度量:

[0090] $$A_{E_i^2}(m) = \sqrt{\sum_i E_i^2(m)N_i^2(m)}$$

[0091] $$A_{F_i^2}(m) = \sqrt{\sum_i F_i^2(m)N_i^2(m)} \quad (5)$$

[0092] 4.4规范化的误差度量

[0093] 如以下表达式中所示,表达式(4)和(5)中的未规范化的误差度量可以用总体响度或对象重要度来规范化:

[0094] $$A'_{E_i}(m) = \frac{\sum_i E_i(m)N_i(m)}{\sum_i N_i(m)+N_0} \quad (6)$$

[0095] $$A'_{F_i}(m) = \frac{\sum_i F_i(m)N_i(m)}{\sum_i N_i(m)+N_0}$$

[0096] $$A'_{E_i^2}(m) = \sqrt{\frac{\sum_i E_i^2(m)N_i^2(m)}{\sum_i N_i^2(m)+N_0^2}} \quad (7)$$

[0097] $$A'_{F_i^2}(m) = \sqrt{\frac{\sum_i F_i^2(m)N_i^2(m)}{\sum_i N_i^2(m)+N_0^2}}$$

[0098] 其中,N_0是用于防止当局部响度之和或经平方的局部响度之和接近零时(例如,当音频内容的一部分是安静的或近乎安静的时,等等)可能出现的数值不稳定的数值稳定因子。控制复杂度分析器(200)可以用针对局部响度之和或经平方的局部响度之和的特定阈值(例如,最小安静程度等)来配置。如果所述和处于或低于该特定阈值,则稳定因子可以被插入到表达式(7)中。应注意,本文所描述的技术也可以被配置为在计算未规范化的或规范化的误差度量时与防止数值不稳定的其他方式(诸如减幅等)一起工作。

[0099] 在一些实施例中,空间误差度量针对每个帧m被计算,随后被低通滤波(例如,利用具有诸如500ms等之类的时间常数的一阶低通滤波器);空间误差度量的最大值、均值、中间值等可以被用作帧的音频质量的指示。

[0100] 4.5帧间空间误差

[0101] 在一些实施例中,与相邻帧的在时间上的变化相关的空间误差度量可以被计算,并且在本文中可以被称为帧间空间误差度量。这些帧间空间误差可以但不限于被用在相邻帧中的每个帧中的空间误差(例如,帧内空间误差)可能非常小或者甚至为零的情况中。即使帧内空间误差很小,跨帧的对象到聚类分配的变化仍也可能例如由于在从一个帧到下一个帧的插值期间造成的空间误差而导致听得见的噪声。

[0102] 在一些实施例中,如本文所描述的音频对象的帧间空间误差基于一个或多个空间误差相关因子而产生,所述空间误差相关因子包括但不仅限于以下中的任何一个:音频对象被聚类或平移到的输出聚类形心的位置变化、相对于音频对象被聚类或平移到的输出聚类的增益系数变化、音频对象的位置变化、音频对象的相对或局部响度等。

[0103] 如以下表达式中所示,示例性的帧间空间误差可以基于音频对象的增益系数的变化以及音频对象被聚类或平移到的输出聚类的位置变化而产生:

[0104] $$\sum_j |g_{ij}(m) - g_{ij}(m+1)||\vec{p}_j(m) - \vec{p}_j(m+1)| \quad (8)$$

[0105] 如果(1)音频对象的增益系数显著地变化，和/或(2)音频对象被聚类或平移到的输出聚类的位置显著地变化，则以上度量提供大的误差。此外，如以下表达式中所示，以上度量可以用音频对象的特定对象重要度(诸如局部响度等)进行加权：

[0106]

$$A_i^2(m \to m+1) = \sum_j N_i(m)N_i(m+1)|g_{ij}(m) - g_{ij}(m+1)||\vec{p}_j(m) - \vec{p}_j(m+1)| \quad (9)$$

[0107] 因为该度量涉及从一个帧到另一个帧的转变，所以可以使用两个帧的响度值的乘积，以使得如果第m帧或第(m+1)帧中的对象的响度为零，则所得到的以上误差度量的值也将为零。这可以被用来处理音频对象在这两个帧中的后一个帧中开始存在或不再存在的情况；这种音频对象对以上误差度量的贡献为零。

[0108] 针对音频对象，另一个示例性的帧间空间误差可以不仅基于音频对象的增益系数的变化和音频对象被聚类或平移到的输出聚类的位置变化而且还基于该音频对象在第一帧(例如，第m帧等)中被渲染成的输出聚类的第一配置和该音频对象在第二帧(例如，第(m+1)帧等)中被渲染成的输出聚类的第二配置之间的差异或距离而产生，如图5中所示。在图5所描绘的示例中，输出聚类2的形心跳到或移到新的位置；结果，音频对象(被表示为三角形)的渲染矢量和增益系数(或增益系数分布)相应地变化。然而，在这个示例中，即使输出聚类2的形心跳过很长距离，对于特定音频对象(三角形)来说，它仍可以通过使用输出聚类3的4的两个形心而被很好地表示/渲染。仅考虑输出聚类的位置变化(或形心变化)的跳跃或差异可能过高估计帧间空间误差或者在与相邻帧(例如，第m帧和第(m+1)帧，等等)相关的变化之间引起的潜在噪声。这种过高估计可以通过在确定与相邻帧相关的帧间空间误差时计算并且考虑作为相邻帧的增益系数分布的变化的基础的增益流来减轻。

[0109] 在一些实施例中，音频对象在第m帧中的增益系数可以用增益矢量[$g_1(m)$，$g_2(m)$，...，$g_N(m)$]表示，其中，该增益矢量的每个分量(例如，1、2、……N等)对应于被用来将音频对象渲染到多个输出聚类(例如，N个输出聚类等)中的相应输出聚类(例如，第1个输出聚类、第2个输出聚类、……、第N个输出聚类等)中的增益系数。仅仅出于例示的目的，在增益矢量的分量中忽略了音频对象在增益系数中的索引。音频对象在第(m+1)帧中的增益系数可以用增益矢量[$g_1(m+1)$，$g_2(m+1)$，...，$g_N(m+1)$]表示。类似地，第m帧中的多个输出聚类的形心的位置可以用矢量$[\vec{p}_1(m),\ \vec{p}_2(m),\ldots,\vec{p}_N(m)]$表示。第(m+1)帧中的多个输出聚类的形心的位置可以用矢量$[\vec{p}_1(m+1),\ \vec{p}_2(m+1),\ldots,\vec{p}_N(m+1)]$表示。音频对象的从第m帧到第(m+1)帧的帧间空间误差可以如以下表达式中所示那样计算得到(音频对象的响度、对象重要度等目前被忽略，并且稍后可以被应用)：

[0110] $D(m \to m+1) = \sum_i \sum_j g_{i \to j} d_{i \to j}$ (10)

[0111] 其中，i是第m帧中的输出聚类的形心的索引，j是第(m+1)帧中的输出聚类的形心的索引。$g_{i \to j}$是从第m帧中的第i输出聚类的形心到第(m+1)帧中的第j输出聚类的形心的增

益流的值。$d_{i\to j}$是第m帧中的第i输出聚类的形心和第(m+1)帧中的第j输出聚类的形心之间的距离(例如,增益流等),并且可以如以下表达式中所示的那样直接计算:

[0112] $$d_{i\to j} = \left|\vec{p}_i(t) - \vec{p}_j(t+1)\right| \quad (11)$$

[0113] 在一些实施例中,增益流值$g_{i\to j}$用包括以下步骤的方法估计:

[0114] 1.将$g_{i\to j}$初始化为零。如果$g_i(m)$和$g_j(m+1)$大于零(0),则针对每对(i,j)计算$d_{i\to j}$。按升序对$d_{i\to j}$进行排序。

[0115] 2.选择具有最小距离的形心对(i^*,j^*),其中,形心对(i^*,j^*)在之前未被选择过。

[0116] 3.按照$g_{i^*\to j^*} = \min(g_{i^*}, g_{j^*})$计算增益流值。

[0117] 4.更新$g_{i^*} = g_{i^*} - g_{i^*\to j^*}$、$g_{j^*} = g_{j^*} - g_{i^*\to j^*}$。

[0118] 5.如果经更新的g_i、g_j全都为零,则停止。否则,跳到上面的步骤2。

[0119] 在图5中所描绘的示例中,通过应用以上方法而获得的非零增益流为:$g_{1\to 1}=0.5$,$g_{2\to 3}=0.2$,$g_{2\to 4}=0.2$,并且$g_{2\to 1}=0.1$。因此,音频对象(在图5中被表示为三角形)的帧间空间误差可以如下计算:

[0120] $D(m\to m+1) = g_{1\to 1}*d_{1\to 1}+g_{2\to 3}*d_{2\to 3}+g_{2\to 4}*d_{2\to 4}+g_{2\to 1}*$

[0121] $d_{2\to 1}$

[0122] $=0.5*d_{1\to 1}+0.2*d_{2\to 3}+0.2*d_{2\to 4}+0.1*d_{2\to 1}$

[0123] (12)

[0124] 相比之下,基于表达式(8)计算的帧间空间误差如下:

[0125] $$D(m\to m+1) = |g_2(m) - g_2(m+1)| * |\vec{p}_2(m) - \vec{p}_2(m+1)| = 0.5 * |\vec{p}_2(t) - \vec{p}_2(t+1)| \quad (13)$$

[0126] 在表达式(12)和(13)中可以看出,表达式(13)中计算的仅取决于$|\vec{p}_2(m) - \vec{p}_2(m+1)|$的帧间空间误差可能过高估计实际的空间误差,因为输出聚类2的形心的运动由于邻近的输出聚类3和4的存在而不会引起音频对象上的大空间误差,邻近的输出聚类3和4可以容易地(并且就空间误差而言相对精确地)占据增益系数的之前被渲染到第m帧中的输出聚类2的部分(或增益流)。

[0127] 音频对象k的帧间空间误差可以被表示为D_k。在一些实施例中,总体帧间空间误差可以如下计算:

[0128] $E_{inter}(m\to m+1) = \sum_k D_k(m\to m+1)$ (14)

[0129] 通过考虑音频对象的各自的对象重要度(诸如局部响度等),总体帧间空间误差可以如下进一步计算:

[0130] $E_{inter}(m\to m+1) = \sum_k N_k(m)N_k(m+1)D_k(m\to m+1)$ (15)

其中,$N_k(m)$和$N_k(m+1)$分别是音频对象k在第m帧和第(m+1)帧中的对象重要度,诸如局部响度等。

[0131] 在一些实施例中,在音频对象还在运动的情况下,音频对象的运动在计算帧间空

间误差时被补偿，例如，如以下表达式中所示：

[0132] $E_{inter}(m\rightarrow m+1)=\sum_k N_k(m)N_k(m+1)\max\{D_k(m\rightarrow m+1)-O_k(m\rightarrow m+1),0\}$ (16)

[0133] 其中，$O_k(m\rightarrow m+1)$是音频对象从第m帧到第(m+1)帧的实际运动。

[0134] 5.主观音频质量的预测

[0135] 在一些实施例中，如本文所描述的空间误差度量中的一个、一些或全部可以被用来预测用于计算空间误差度量的一个或多个帧的感知音频质量(例如，与诸如MUSHRA测试、MOS测试等之类的感知音频质量测试相关)。训练数据集(例如，代表性的音频内容元素或摘录的集合，等等)可以被用来确定空间误差度量和从多个用户收集的主观音频质量的测量结果之间的相关性(例如，反映空间误差越高导致利用用户测量的主观音频质量越低的负值)。基于训练数据集确定的相关性可以被用来确定预测参数。这些预测参数可以被用来基于从一个或多个帧(例如，非训练数据等)计算的空间误差度量产生所述一个或多个帧的感知音频质量的一个或多个指示。在其中多个空间误差度量(例如，帧内对象位置误差、帧内对象平移误差等)被用来预测主观音频质量的一些实施例中，与(例如，基于训练数据集通过针对多个用户进行MUSHRA测试而测量得到的，等等)主观音频质量的相关性相对较高的空间误差度量(例如，帧内对象平移误差度量等)(例如，具有相对较大量值的负值等)可以被给予所述多个空间误差度量(例如，帧内对象位置误差、帧内对象平移误差等)当中的相对较高的权重。应注意，本文中所描述的技术可以被配置为与基于通过这些技术确定的一个或多个空间误差度量来预测音频质量的其他方式一起工作。

[0136] 6.空间误差和空间复杂度的可视化

[0137] 在一些实施例中，根据本文中所描述的技术针对一个或多个帧确定的一个或多个空间误差度量可以与所述一个或多个帧中的音频对象和/或输出聚类的性质(例如，响度、位置等)一起用于提供所述一个或多个帧中的音频内容的空间复杂度在显示器(例如，计算机屏幕、网页等)上的可视化。可视化可以通过多种多样的图形用户界面部件(诸如VU计量器(例如，2D、3D等))、音频对象和/或输出聚类的可视化、条形图、其他合适的手段等来提供。在一些实施例中，例如当空间创作或转换处理正在被执行时、在这种处理被执行之后、等等，空间复杂度的总体指示被提供在显示器上。

[0138] 图3A至图3D例示了用于可视化一个或多个帧中的空间复杂度的示例性用户界面。用户界面可以由空间复杂度分析器(例如，图2的200等)或用户界面模块(例如，图2的210等)、混合工具、格式转换工具、音频对象聚类工具、独立分析工具等提供。用户界面可以被用来当输入音频内容中的音频对象被压缩成输出音频内容中的数量更少的(例如，少得多的，等等)输出聚类时提供可能的音频质量劣化和其他相关信息的可视化。可能的音频质量劣化和其他相关信息的可视化可以与从同一源音频内容生成一个或多个版本的基于对象的音频内容同时提供。

[0139] 在一些实施例中，如图3A中所示，用户界面包括3D显示部件302，该3D显示部件302可视化音频对象和输出聚类在示例性的3D收听空间中的位置。如用户界面中所描绘的音频对象或输出聚类中的零个、一个或多个可以具有收听环境中的动态位置或固定位置。

[0140] 在一些实施例中，用户或收听者在3D收听空间的地平面的中间。在一些实施例中，如图3B中所示，用户界面包括3D收听空间的不同的2D视图，诸如表示3D收听空间的不同投影的顶视图、侧视图、后视图等。

[0141] 在一些实施例中，如图3C中所示，用户界面还包括条形图304和306，这些条形图分别对(例如，基于响度、语义对话概率等确定/估计的)对象重要度和对象响度L(以方为单位)进行可视化。“输入索引”表示音频对象(或输出聚类)的索引。输入索引的每个值处的竖条的高度指示语音或对话的概率。纵轴“L”表示可被用作确定对象重要度等的基础的局部响度。纵轴“P”表示语音或对话内容的概率。条形图304和306中的竖条(表示音频对象或输出聚类的语音或对话内容的单个的局部响度和概率)可以随着帧不同而起伏。

[0142] 在一些实施例中，如图3D中所示，用户界面包括与帧内空间误差相关的第一空间复杂度计量器308和与帧间空间误差相关的第二空间复杂度计量器310。在一些实施例中，音频内容的空间复杂度可以由根据帧内空间误差度量、帧间空间误差度量等中的一个或多个(例如，不同的组合等)产生的空间误差度量或预测音频质量测试得分来量化或表示。在一些实施例中，基于训练数据确定的预测参数可以被用来基于一个或多个空间误差度量预测音频质量劣化。所预测的感知音频质量劣化可以由参照主观感知音频质量测试(诸如MUSHRA测试、MOS测试等)的一个或多个预测的感知测试得分来表示。在一些实施例中，可以分别至少部分基于帧内空间误差和帧间空间误差来预测两组感知测试得分。至少部分基于帧内空间误差产生的第一组感知测试得分可以被用来驱动第一空间复杂度计量器308的显示。至少部分基于帧间空间误差产生的第二组感知测试得分可以被用来驱动第二空间复杂度计量器310的显示。

[0143] 在一些实施例中，“听得见的误差”指示器灯可以被描绘在用户界面中，以指示由空间复杂度计量器(例如，308、310等)中的一个或多个表示的所预测的音频质量劣化(例如，在0至10的值范围内，等等)已经越过了所配置的“令人讨厌的”阈值(例如，10，等等)。在一些实施例中，如果空间复杂度计量器(例如，308、310等)均未越过所配置的“令人讨厌的”阈值(例如，其数值为10，等等)，则“听得见的误差”指示器灯不被描绘，但是可以在空间复杂度计量器之一越过所配置的“令人讨厌的”阈值时被触发。在一些实施例中，空间复杂度计量器(例如，308、310等)中的所预测的音频质量劣化的不同子范围可以由不同颜色带表示(例如，0-3的子范围被映射到指示极小的音频质量劣化的绿色带，8-10的子范围被映射到指示严重的音频质量劣化的红色带，等等)。

[0144] 音频对象在图3A和图3B中被描绘为圆圈。然而，在各种实施例中，音频对象或输出聚类可以使用不同的形状描绘。在一些实施例中，表示音频对象或输出聚类的形状的大小可以指示(例如，可以与下述项成比例，等等)音频对象的对象重要度、音频对象或输出聚类的绝对或相对响度等。不同的颜色编码方案可以被用来给用户界面中的用户界面部件上色。例如，音频对象可以被上绿色，而输出聚类可以被上非绿色。相同颜色的不同形状可以被用来区分音频对象的性质的不同值。音频对象的颜色可以基于音频对象的性质、音频对象的空间误差、音频对象相对于该音频对象被分配或分配到的输出聚类的距离等而改变。

[0145] 图4例示了VU计量器形式的视觉复杂度计量器的两个示例性实例402和404。VU计量器可以是图3A至图3D中所描绘的用户界面的一部分或者是与图3A至图3D中所描绘的用户界面不同的用户界面(例如，由图2的用户界面模块210等提供)。视觉复杂度计量器的第一实例402指示与低空间误差对应的高音频质量和低空间复杂度。视觉复杂度计量器的第二实例404指示与高空间误差对应的低音频质量和高空间复杂度。在VU计量器中指示的复杂度度量值可以是帧内空间误差、帧间空间误差、基于帧内空间误差预测/确定的感知音频

质量测试得分、基于帧间空间误差预测/确定的预测音频质量测试得分等。附加地、可选地或可替代地，VU计量器可以包括/实现“峰值保持”函数，该函数被配置为显示在某个(例如，过去的，等等)时间间隔内出现的最低质量和最高复杂度。该时间间隔可以是固定的(例如，最后10秒，等等)，或者可以是可变的且是相对于正被处理的音频内容的开头的。此外，复杂度度量值的数值显示可以与VU计量器显示结合使用，或者替代VU计量器显示使用。

[0146] 如图4中所示，复杂度夹灯可以被显示在表示复杂度计量器的垂直标度的下面。如果复杂度值已经达到/越过某个临界阈值，则该夹灯可以变为工作。这可以通过点亮、改变颜色、可以被视觉地感知的任何其他变化来可视化。在一些实施例中，作为显示复杂度标签(例如，高、良好、中等和低质量等)的替代或附加，垂直标度也可以是数值的(例如，从0至10等)以指示复杂度或音频质量。

[0147] 7.示例性的处理流程

[0148] 图6例示了示例性的处理流程。在一些实施例中，一个或多个计算装置或单元(例如，图2的空间复杂度分析器200等)可以执行该处理流程。

[0149] 在块602中，空间复杂度分析器200(例如，如图2等中所示)确定存在于一个或多个帧中的输入音频内容中的多个音频对象。

[0150] 在块604中，空间复杂度分析器(200)确定存在于所述一个或多个帧中的输出音频内容中的多个输出聚类。这里，输入音频内容中的所述多个音频对象被转换成输出音频内容中的所述多个输出聚类。

[0151] 在块606中，空间复杂度分析器(200)至少部分基于所述多个音频对象的位置元数据和所述多个输出聚类的位置元数据来计算一个或多个空间误差度量。

[0152] 在实施例中，所述多个音频对象中的至少一个音频对象被分配到所述多个输出聚类中的两个或更多个输出聚类。

[0153] 在实施例中，所述多个音频对象中的至少一个音频对象被分配到所述多个输出聚类中的一个输出聚类。

[0154] 在实施例中，空间复杂度分析器(200)被进一步配置为基于所述一个或多个空间误差度量来确定通过将输入音频内容中的多个音频对象转换到输出聚类中的多个输出聚类而引起的感知音频质量劣化。

[0155] 在实施例中，感知音频质量劣化由与感知音频质量测试相关的一个或多个预测测试得分表示。

[0156] 在实施例中，所述一个或多个空间误差度量包括以下中的至少一个：帧内空间误差度量、帧间空间误差度量。

[0157] 在实施例中，帧内空间误差度量包括以下中的至少一个：帧内对象位置误差度量、帧内对象平移误差度量、重要度加权的帧内对象位置误差度量、重要度加权的帧内对象平移误差度量、规范化的帧内对象位置误差度量、规范化的帧内对象平移误差度量等。

[0158] 在实施例中，帧间空间误差度量包括以下中的至少一个：基于增益系数流的帧间空间误差度量、不基于增益系数流的帧间空间误差度量等。

[0159] 在实施例中，每个帧间空间误差度量是关于两个不同的帧而被计算的。

[0160] 在实施例中，所述多个音频对象经由多个增益系数而与所述多个输出聚类相关。

[0161] 在实施例中，每个帧对应于输入音频内容中的时间段和输出音频内容中的第二时

间段;存在于输入音频内容中的第一时间段中的音频对象被映射到存在于输出音频内容中的第二时间段中的输出聚类。

[0162] 在实施例中,所述一个或多个帧包括两个连续的帧。

[0163] 在实施例中,空间复杂度分析器(200)被进一步配置为执行:重构一个或多个用户界面部件,该一个或多个用户界面部件表示以下中的一个或多个:所述多个音频对象中的音频对象、收听空间中的所述多个输出聚类中的输出聚类,等等;并且使所述一个或多个用户界面部件被显示给用户。

[0164] 在实施例中,所述一个或多个用户界面部件中的用户界面部件表示所述多个音频对象中的音频对象;音频对象被映射到所述多个输出聚类中的一个或多个输出聚类;并且用户界面部件的至少一个视觉特性表示与将音频对象映射到所述一个或多个输出聚类相关的一个或多个空间误差的总量。

[0165] 在实施例中,所述一个或多个用户界面部件包括收听空间的3维(3D)形式的表示。

[0166] 在实施例中,所述一个或多个用户界面部件包括收听空间的2维(2D)形式的表示。

[0167] 在实施例中,空间复杂度分析器(200)被进一步配置为执行:构造一个或多个用户界面部件,该一个或多个用户界面部件表示以下中的一个或多个:所述多个音频对象中的音频对象的各自的对象重要度、所述多个输出聚类中的输出聚类的各自的对象重要度、所述多个音频对象中的音频对象的各自的响度、所述多个输出聚类中的输出聚类的各自的响度、所述多个音频对象中的音频对象的语音或对话内容的各自的概率、所述多个输出聚类中的输出聚类的语音或对话内容的概率等;并且使所述一个或多个用户界面部件被显示给用户。

[0168] 在实施例中,空间复杂度分析器(200)被进一步配置为执行:构造一个或多个用户界面部件,该一个或多个用户界面部件表示以下中的一个或多个:一个或多个空间误差度量、至少部分基于一个或多个空间误差度量而确定的一个或多个预测的测试得分等;并且使所述一个或多个用户界面部件被显示给用户。

[0169] 在实施例中,转换处理将存在于输入音频内容中的时间相关的音频对象转换成构成输出聚类的时间相关的输出聚类;并且所述一个或多个用户界面部件包括在包含并且长至一个或多个帧的过去时间间隔内在转换处理中出现最差音频质量劣化的视觉指示。

[0170] 在实施例中,所述一个或多个用户界面部件包括在包含并且长至一个或多个帧的过去时间间隔内在转换处理中出现的音频质量劣化已经超过音频质量劣化阈值的视觉指示。

[0171] 在实施例中,所述一个或多个用户界面部件包括其高度指示所述一个或多个帧中的音频质量劣化的竖条,并且其中,该竖条基于所述一个或多个帧中的音频质量劣化而被颜色编码。

[0172] 在实施例中,所述多个输出聚类中的输出聚类包括所述多个音频对象中的两个或更多个音频对象所映射到的部分。

[0173] 在实施例中,所述多个音频对象中的音频对象或所述多个输出聚类中的输出聚类中的至少一个具有随着时间变化的动态位置。

[0174] 在实施例中,所述多个音频对象中的音频对象或所述多个输出聚类中的输出聚类中的至少一个具有不随着时间变化的固定位置。

[0175] 在实施例中，输入音频内容和输出音频内容中的至少一个是仅音频信号和视听信号之一的一部分。

[0176] 在实施例中，空间复杂度分析器(200)被进一步配置为执行：接收指定对于将输入音频内容转换为输出音频内容的转换处理的改变的用户输入；并且响应于接收到该用户输入，引起对于将输入音频内容转换为输出音频内容的转换处理的所述改变。

[0177] 在实施例中，如上所述的方法中的任何一个是在转换处理将输入音频内容转换为输出音频内容时同时执行的。

[0178] 实施例包括一种被配置为执行本文中所描述的方法中的任何一个的媒体处理系统。

[0179] 实施例包括一种设备，该设备包括处理器并且被配置为执行前述方法中的任何一个。

[0180] 实施例包括存储有软件指令的非暂时性计算机可读存储介质，这些软件指令当被一个或多个处理器执行时引起执行前述方法中的任何一个。注意，尽管本文中讨论了单独的实施例，但是本文中所讨论的实施例和/或部分实施例的任何组合可以被组合来形成另外的实施例。

[0181] 8.实现机制——硬件概述

[0182] 根据一个实施例，本文中所描述的技术由一个或多个专用计算装置实现。专用计算装置可以被硬连线以执行这些技术，或者可以包括被持久性地编程为执行这些技术的数字电子装置(诸如一个或多个专用集成电路(ASIC)或现场可编程门阵列(FPGA))，或者可以包括按照固件、存储器、其他储存器或组合中的程序指令执行这些技术的一个或多个通用硬件处理器。这种专用计算装置还可以结合具有自定义编程的自定义的硬连线逻辑、ASIC、或FPGA来实现这些技术。专用计算装置可以是台式计算机系统、便携式计算机系统、手持装置、联网装置、或包含硬连线逻辑和/或程序逻辑来实现这些技术的任何其他装置。

[0183] 例如，图7是例示了在其上可以实现本发明的实施例的计算机系统700的框图。计算机系统700包括用于传送信息的总线702或其他通信机制、以及与总线702耦接的用于对信息进行处理的硬件处理器704。硬件处理器704可以是例如专用微处理器。

[0184] 计算机系统700还包括耦接到总线702的用于存储信息和将由处理器704执行的指令的主存储器706，诸如随机存取存储器(RAM)或其他动态存储装置。主存储器706还可以用于在将由处理器704执行的指令的执行期间存储临时变量或其他中间信息。这种指令在被存储在处理器704可访问的非暂时性存储介质中时使得计算机系统700成为装置特定于执行这些指令中所指定的操作的专用机器。

[0185] 计算机系统700还包括耦接到总线702的用于存储用于处理器704的静态信息和指令的只读存储器(ROM)708或其他静态存储装置。存储装置710(诸如磁盘或光学盘)被提供并且耦接到总线702，以用于存储信息和指令。

[0186] 计算机系统700可以经由总线702耦接到用于向计算机用户显示信息的显示器712，诸如液晶显示器(LCD)。包括字母数字键和其他键的输入装置714耦接到总线702，以用于将信息和命令选择传送给处理器704。另一种类型的用户输入装置是用于将方向信息和命令选择传送给处理器704并且用于控制显示器712上的光标移动的光标控制器716，诸如鼠标、轨迹球、或光标方向键。该输入装置典型地具有两个轴(第一轴(例如，x)和第二轴(例

如,y))上的两个自由度,这允许装置可以指定平面中的位置。

[0187] 计算机系统700可以使用与该计算机系统组合使计算机系统700成为专用机器或者将计算机系统700编程为专用机器的装置特定的硬连线逻辑、一个或多个ASIC或FPGA、固件和/或程序逻辑来实现本文中所描述的技术。根据一个实施例,本文中的技术由计算机系统700响应于执行主存储器706中包含的一个或多个指令的一个或多个序列的处理器704来执行。这种指令可以从另一个存储介质(诸如存储装置710)读取到主存储器706中。主存储器706中包含的指令序列的执行使处理器704执行本文中所描述的处理步骤。在替代实施例中,硬连线的电路系统可以被用来代替软件指令或者与软件指令组合使用。

[0188] 本文中所使用的术语“存储介质”是指存储使机器以特定方式运行的数据和/或指令的任何非暂时性介质。这种存储介质可以包括非易失性介质和/或易失性介质。非易失性介质例如包括光学盘或磁性盘,诸如存储装置710。易失性介质包括动态存储器,诸如主存储器706。存储介质的常见形式例如包括软盘、柔性盘、硬盘、固态驱动器、磁带或任何其他磁性数据存储介质、CD-ROM、任何其他光学数据存储介质、具有孔图案的任何物理介质、RAM、PROM、以及EPROM、FLASH-EPROM、NVRAM、任何其他存储器芯片或盒。

[0189] 存储介质不同于传输介质,但是可以与传输介质结合使用。传输介质参与在存储介质之间传递信息。例如,传输介质包括同轴电缆、铜线和光纤,包括包含总线702的电线。传输介质还可以采取声波或光波的形式,诸如在无线电波和红外数据通信期间产生的声波或光波。

[0190] 在将一个或多个指令的一个或多个序列传载到处理器704以便执行时可以涉及各种形式的介质。例如,指令可以首先承载在远程计算机的磁盘或固态驱动器上。远程计算机可以将指令加载到其动态存储器中,并且使用调制解调器通过电话线发送这些指令。计算机系统700本地的调制解调器可以接收电话线上的数据,并且使用红外发射器来将该数据转换为红外信号。红外探测器可以接收红外信号中所承载的数据,并且适当的电路系统可以将该数据放置在总线702上。总线702将数据传载到主存储器706,处理器704从主存储器706取得并执行这些指令。主存储器706接收的指令可选地可以在被处理器704执行之前或之后存储在存储装置710上。

[0191] 计算机系统700还包括耦接到总线702的通信接口718。通信接口718提供耦接到网络链路720的双向数据通信,网络链路720连接到本地网络722。例如,通信接口718可以是综合服务数字网络(ISDN)卡、电缆调制解调器、卫星调制解调器、或者用于提供与相应类型的电话线的数据通信连接的调制解调器。作为另一个示例,通信接口718可以是用于提供与可兼容局域网(LAN)的数据通信连接的LAN卡。还可以实现无线链接。在任何这种实现中,通信接口718发送和接收承载有表示各种类型的信息的数字数据流的电信号、电磁信号或光学信号。

[0192] 网络链接720典型地通过一个或多个网络提供与其他数据装置的数据通信。例如,网络链接720可以通过局域网722提供与主机724或由互联网服务提供商(ISP)726运营的数据设备的连接。ISP 726继而通过全球分组数据通信网络(现在通常被称为“因特网”728)提供数据通信服务。本地网络722和互联网728都使用承载数字数据流的电信号、电磁信号或光学信号。通过各种网络的信号、以及网络链接720上的通过通信接口718的信号是传输介质的示例形式,这些信号承载了来去计算机系统700的数字数据。

[0193] 计算机系统700可以通过网络、网络链接720和通信接口718来发送消息和接收包括程序代码的数据。在因特网示例中，服务器730可以通过因特网728、ISP 726、本地网络722和通信接口718发送被请求的应用程序代码。

[0194] 所接收的代码可以在其被接收时被执行、和/或被存储在存储装置710或其他非易失性储存器中以供以后执行。

[0195] 9.等同、扩展、替代及其他

[0196] 在前面的说明书中，已经参照随着实现不同而有所变化的许多特定细节描述了本发明的实施例。因此，本发明是什么、申请人意图本发明是什么的唯一且排他的指示是从本申请发表的特定形式的一套权利要求，包括任何后续修正，这样的权利要求以该特定形式发布。在本文中对于这种权利要求中所包含的术语明确阐述的任何定义应决定这样的术语在权利要求中所使用的意义。因此，在权利要求中没有明确记载的限制、元件、性质、特征、优点或属性均不得以任何方式限制这种权利要求的范围。说明书和附图因此要从例示性而非限制性的意义上来看待。

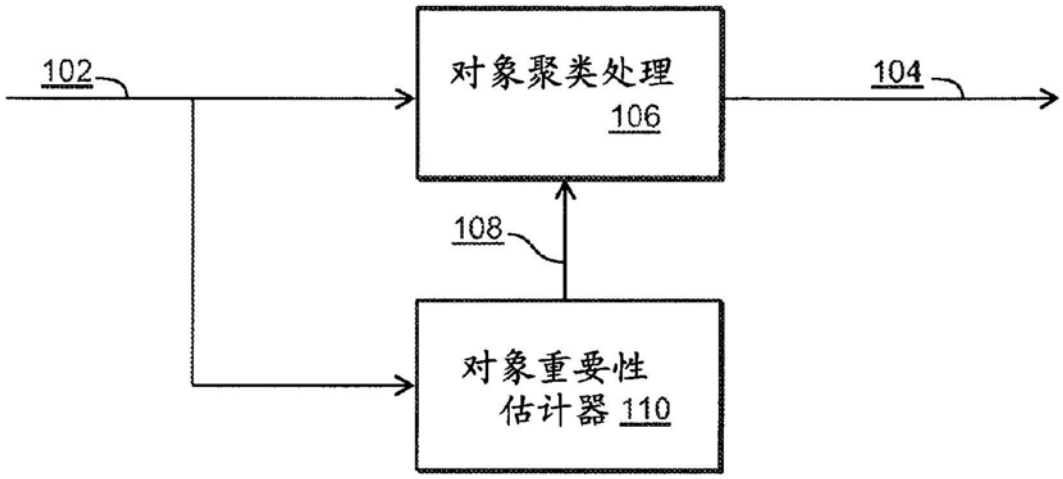

102
对象聚类处理
106
104
108
对象重要性
估计器 110

图1

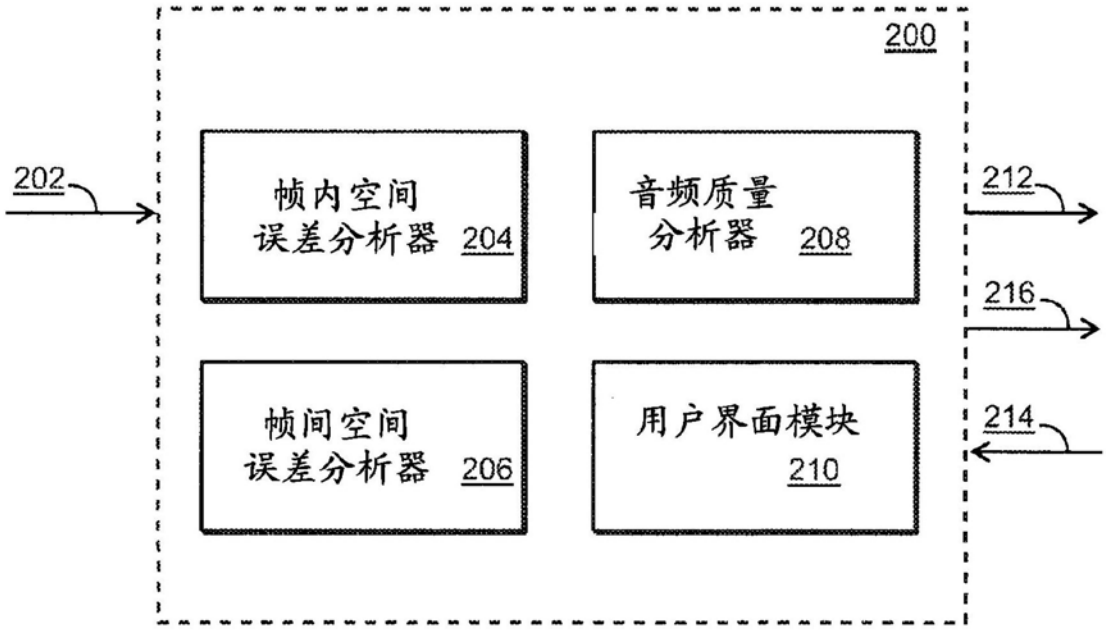

200
202
帧内空间
误差分析器 204
音频质量
分析器 208
212
216
帧间空间
误差分析器 206
用户界面模块
210
214

图2

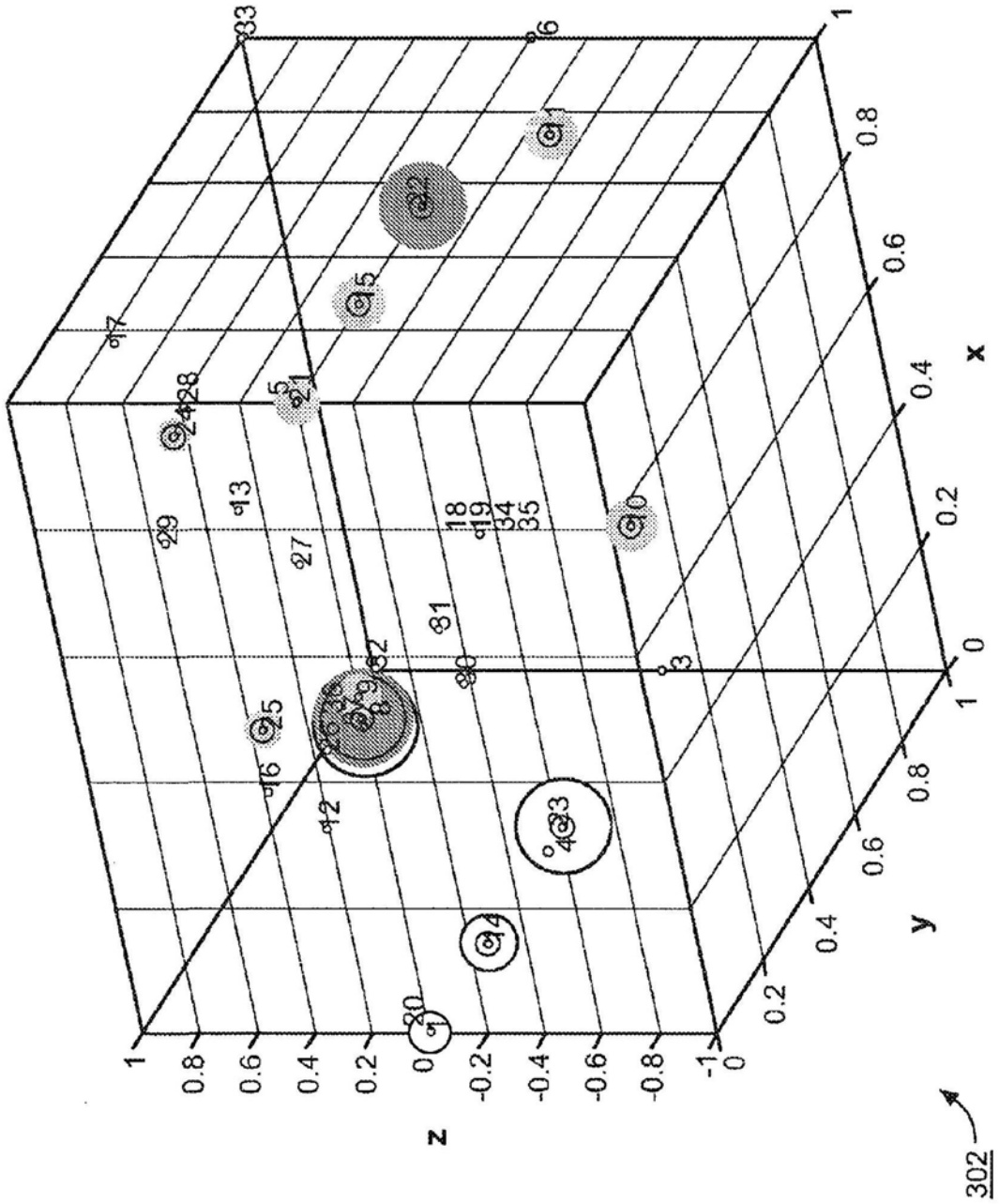


图3A

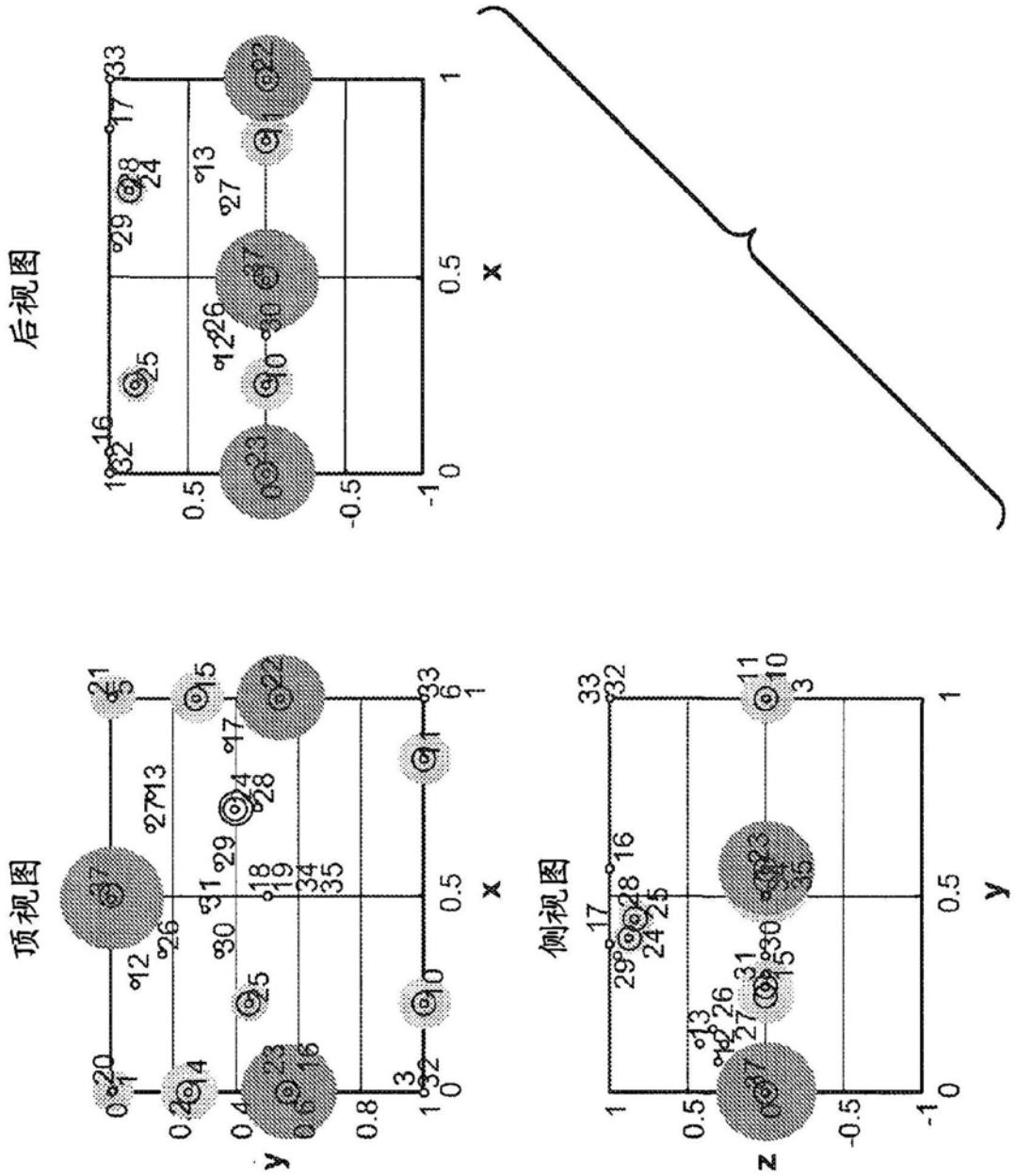


图3B

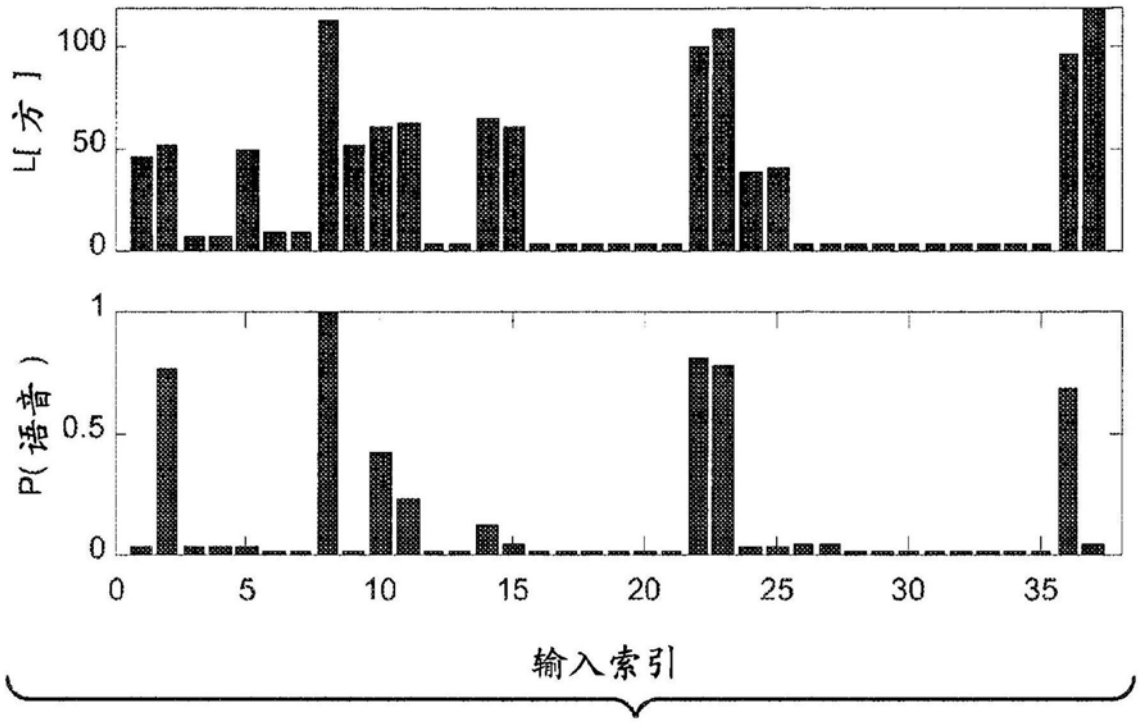

L[方]
100
50
0
P(语音)
1
0.5
0
0
5
10
15
20
25
30
35
输入索引

图3C

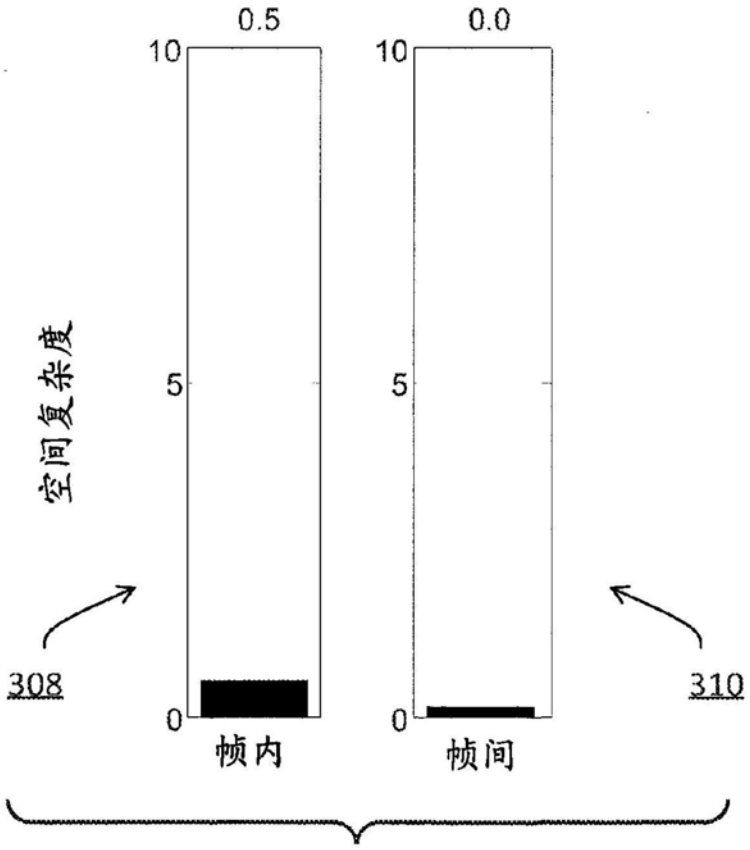

0.5
0.0
10
10
5
5
0
0
空间复杂度
308
310
帧内
帧间

图3D

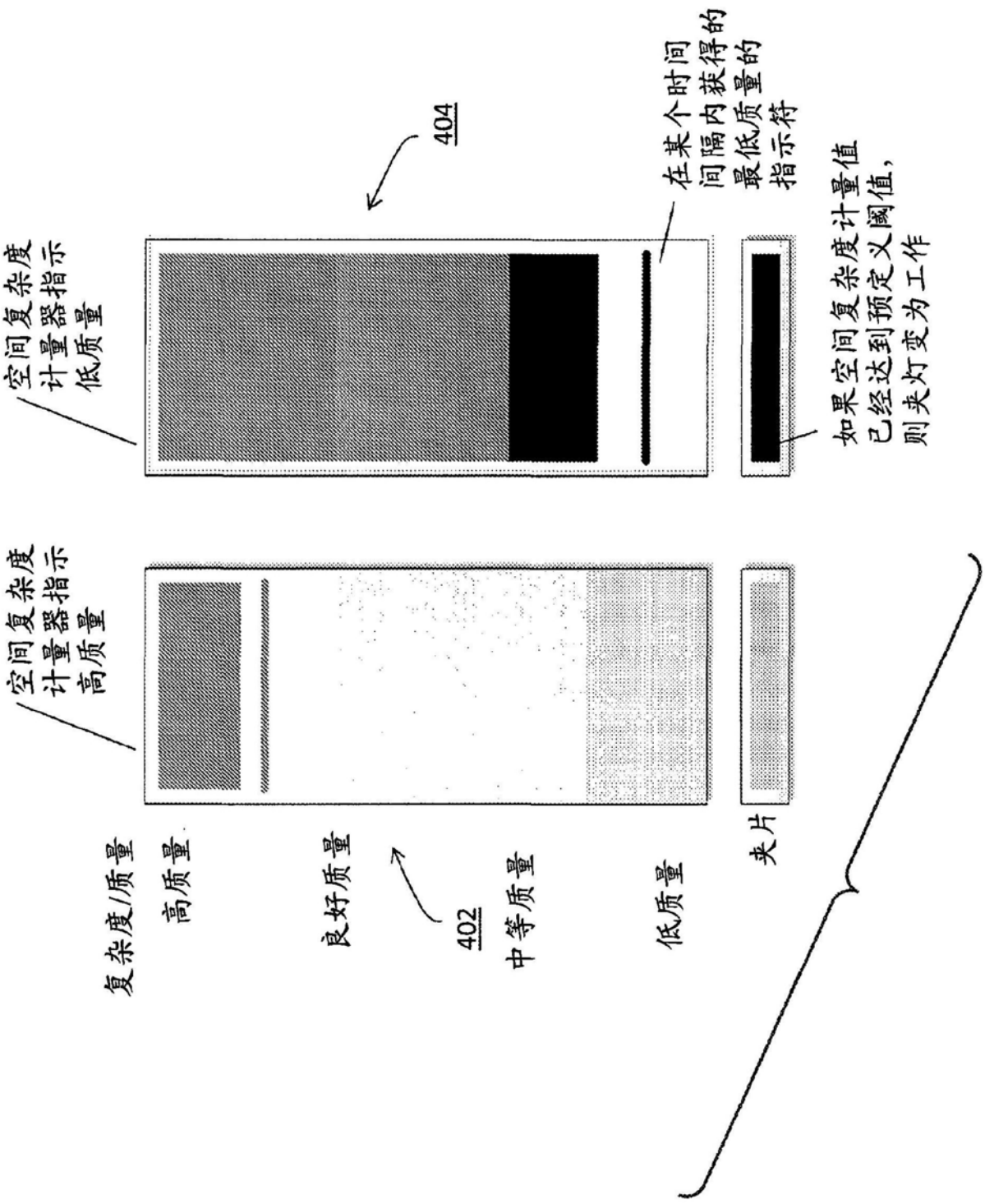

空间复杂度计量器指示低质量
404
在某个时间间隔内获得的最低质量的指示符
如果空间复杂度计量值已经达到预定义阈值，则夹灯变为工作
空间复杂度计量器指示高质量
复杂度/质量
高质量
良好质量
402
中等质量
低质量
夹片

图4

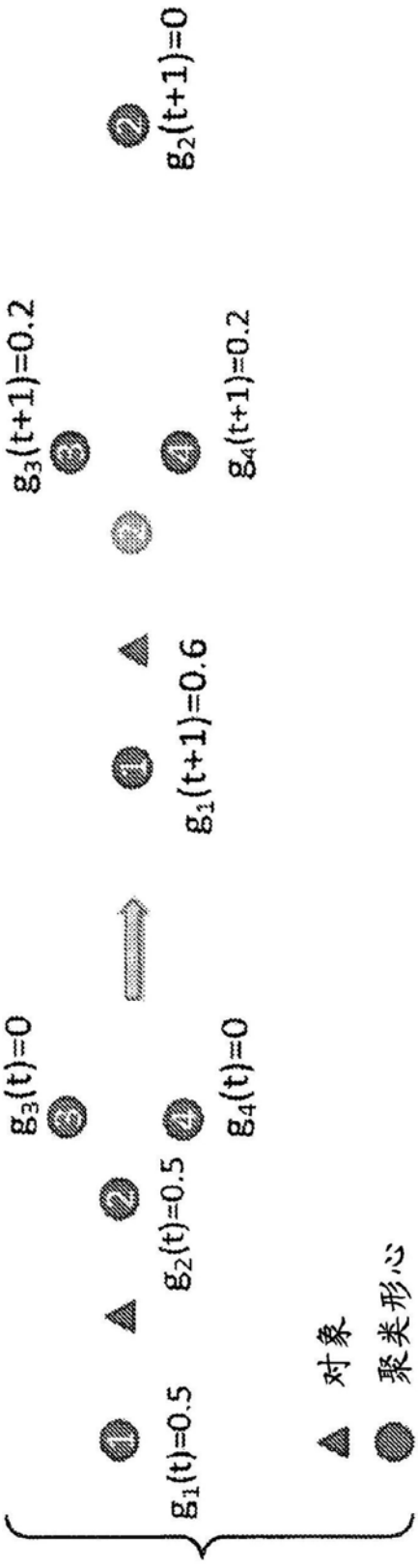


图5

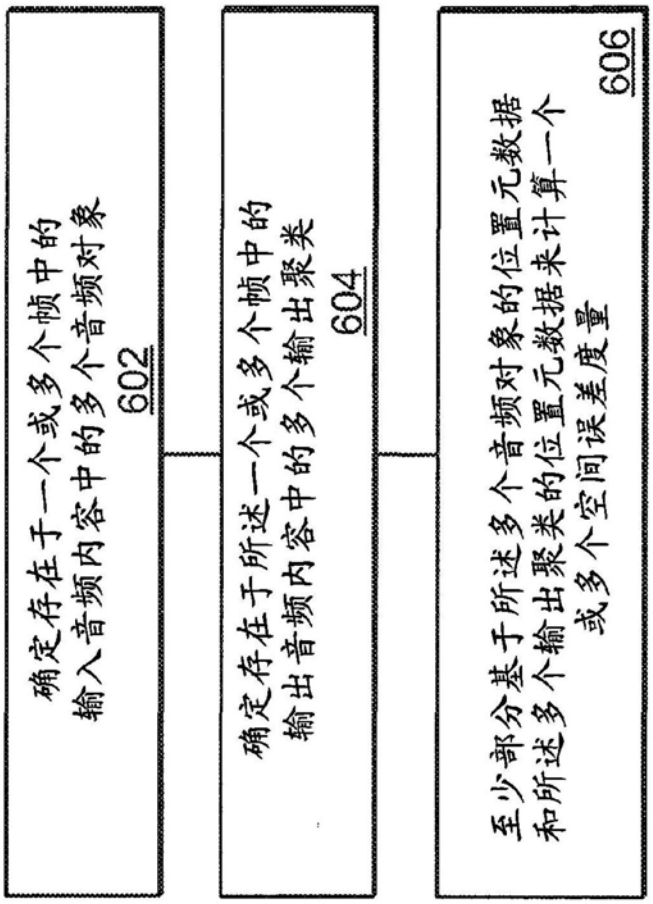

确定存在于一个或多个帧中的
输入音频内容中的多个音频对象
602
确定存在于所述一个或多个帧中的
输出音频内容中的多个输出聚类
604
至少部分基于所述多个音频对象的位置元数据
和所述多个输出聚类的位置元数据来计算一个
或多个空间误差度量
606

图6

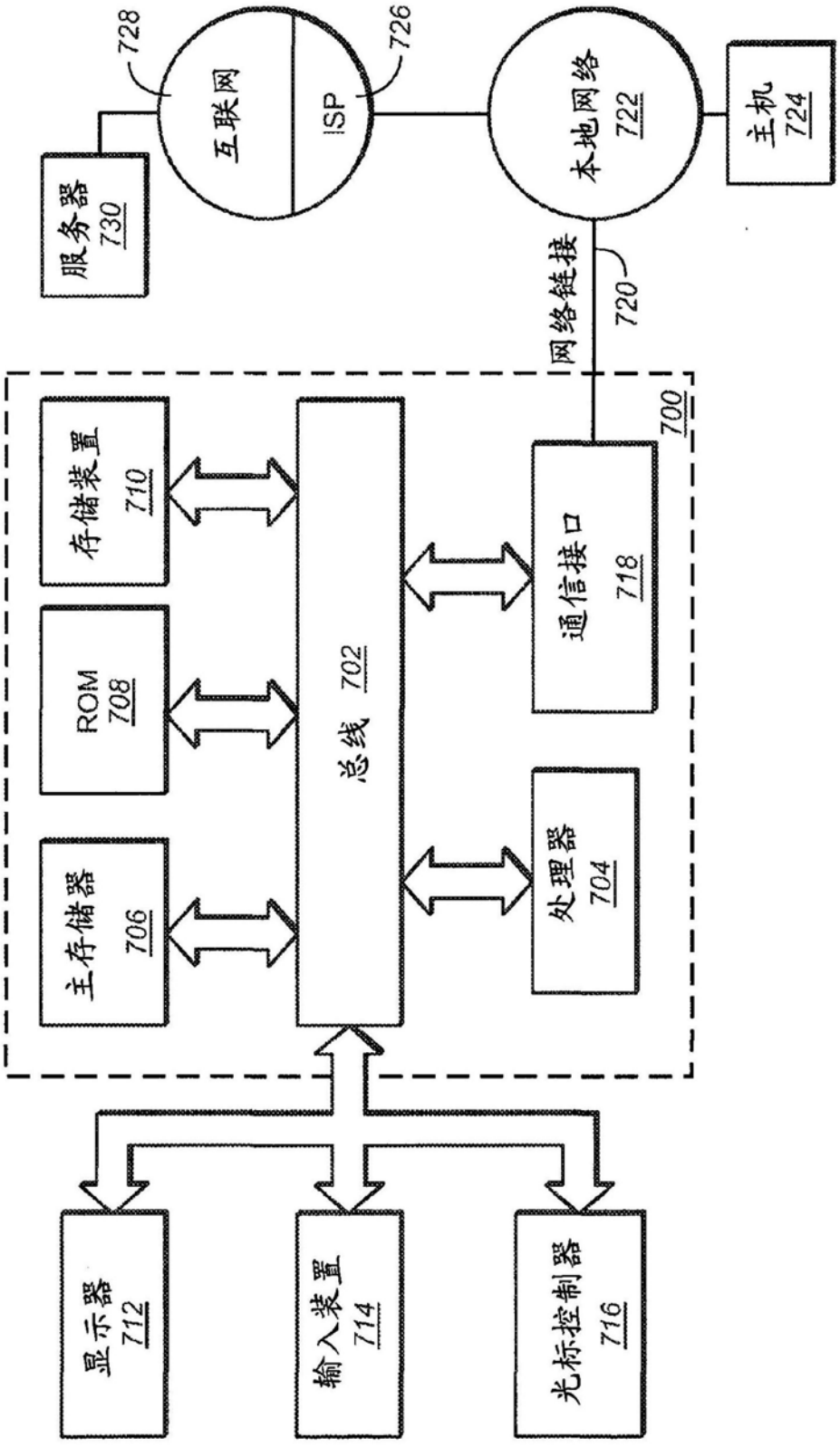

728
互联网
ISP
726
本地网络
722
主机
724
服务器
730
网络链接
720
700
存储装置
710
ROM
708
主存储器
706
总线
702
通信接口
718
处理器
704
显示器
712
输入装置
714
光标控制器
716

图7