



(12) 发明专利

(10) 授权公告号 CN 113016190 B

(45) 授权公告日 2023.06.13

(21) 申请号 201980074887.X

(22) 申请日 2019.09.30

(65) 同一申请的已公布的文献号  
申请公布号 CN 113016190 A

(43) 申请公布日 2021.06.22

(30) 优先权数据  
62/739,713 2018.10.01 US  
62/869,703 2019.07.02 US

(85) PCT国际申请进入国家阶段日  
2021.05.13

(86) PCT国际申请的申请数据  
PCT/US2019/053830 2019.09.30

(87) PCT国际申请的公布数据  
W02020/072364 EN 2020.04.09

(73) 专利权人 杜比实验室特许公司  
地址 美国加利福尼亚州

(72) 发明人 S·戴利 P·A·C·克拉姆  
E·D·吉特曼 S·M·鲁杰里

(74) 专利代理机构 北京律盟知识产权代理有限  
责任公司 11287

专利代理师 陈甜甜

(51) Int.Cl.

H04N 21/442 (2006.01)

H04N 21/462 (2006.01)

H04N 21/41 (2006.01)

H04N 21/414 (2006.01)

H04N 21/422 (2006.01)

H04N 21/4223 (2006.01)

H04N 21/434 (2006.01)

H04N 21/4402 (2006.01)

H04N 21/236 (2006.01)

H04N 21/84 (2006.01)

H04N 21/854 (2006.01)

(56) 对比文件

US 2005223237 A1, 2005.10.06

CN 101543086 A, 2009.09.23

US 2010086204 A1, 2010.04.08

US 8401248 B1, 2013.03.19

US 2017125059 A1, 2017.05.04

WO 2017120469 A1, 2017.07.13

US 9652676 B1, 2017.05.16

US 2009094629 A1, 2009.04.09

审查员 王莹莹

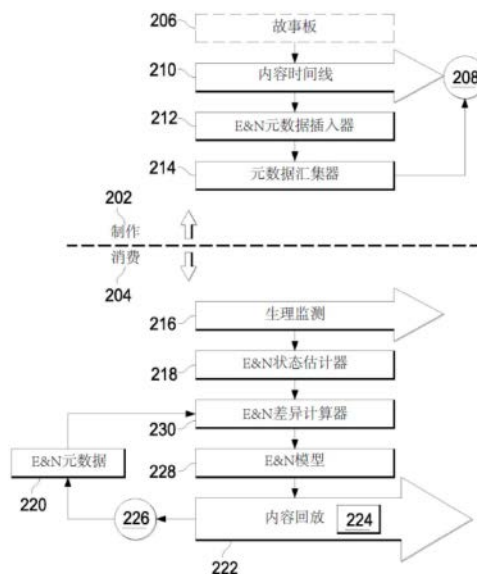
权利要求书2页 说明书40页 附图19页

(54) 发明名称

经由生理监测的创作意图可扩展性

(57) 摘要

接收描述与媒体内容相关的情绪预期及叙事信息的创作意图输入。基于所述创作意图输入产生与所述媒体内容相关的预期生理上可观察状态。将视听内容信号与所述媒体内容及包括所述生理上可观察状态的媒体元数据提供到回放设备。所述视听内容信号致使回放装置使用生理监测信号关于观看者确定与所述媒体内容相关的经评估生理上可观察状态且基于所述预期生理上可观察状态及所述经评估生理上可观察状态产生将向所述观看者呈现的经修改媒体内容。



1. 一种计算机实施方法,其包括:

接收描述与媒体内容的一或多个部分相关的情绪预期及叙事信息的创作意图输入,其中所述创作意图输入含有创意人员的意图的语义表达;

基于所述创作意图输入产生与所述媒体内容的所述一或多个部分相关的一或多个预期生理上可观察状态;

将视听内容信号与所述媒体内容及包括所述媒体内容的所述一或多个部分的所述一或多个预期生理上可观察状态的媒体元数据提供到回放设备,其中所述媒体元数据包括以下各者中的一者:用于在消费阶段中导出一组非语义信号修改选项的所述语义表达或在制作阶段基于所述语义表达产生的所述一组非语义信号修改选项;

其中所述视听内容信号致使回放装置(a)使用一或多个生理监测信号关于观看者确定与所述媒体内容的所述一或多个部分相关的一或多个经评估生理上可观察状态且(b)基于所述一或多个预期生理上可观察状态及所述一或多个经评估生理上可观察状态从所述媒体内容产生经修改媒体内容同时调整且向所述观看者呈现从所述媒体内容产生的所述经修改媒体内容,其中所述回放装置从所述一组信号修改选项选择一或多个特定信号修改选项来对所述媒体内容执行一或多个媒体内容调整,以响应于确定所述一或多个预期生理上可观察状态与所述一或多个经评估生理上可观察状态之间的差异大于差异阈值而最小化所述差异。

2. 根据权利要求1所述的方法,其中所述创作意图输入表示其致使所述媒体内容及所述媒体元数据在制作阶段产生的创意人员的创作意图。

3. 一种计算机实施方法,其包括:

接收视听内容信号与媒体内容及媒体元数据,其中所述媒体元数据包括所述媒体内容的一或多个部分的一或多个预期生理上可观察状态,其中所述媒体元数据包括用于响应于检测一或多个经评估生理上可观察状态与所述一或多个预期生理上可观察状态之间的差异而修改所述媒体内容的所述一或多个部分的一或多个信号修改选项;

其中与所述媒体内容的所述一或多个部分相关的所述一或多个预期生理上可观察状态是基于描述与媒体内容的一或多个部分相关的情绪预期及叙事信息的创作意图输入产生;

使用一或多个生理监测信号关于观看者确定与所述媒体内容的所述一或多个部分相关的所述一或多个经评估生理上可观察状态;

基于所述一或多个预期生理上可观察状态及所述一或多个经评估生理上可观察状态从所述媒体内容产生并呈现经修改媒体内容同时调整且向所述观看者呈现从所述媒体内容产生的所述经修改媒体内容,其中所述一或多个信号修改选项用于在所述媒体内容的内容回放中最小化关于所述观看者的所述一或多个经评估生理上可观察状态与所述一或多个预期生理上可观察状态之间的所述差异。

4. 根据权利要求3所述的方法,其中所述一或多个经评估生理上可观察状态包括所述观看者的经评估情绪状态,其中所述观看者的所述一或多个预期生理上可观察状态包括所述观看者的预期情绪状态,其是与所述观看者的所述经评估情绪状态相同类型的情绪状态。

5. 根据权利要求3或权利要求4所述的方法,其中所述一或多个经评估生理上可观察状

态包括所述观看者的经评估叙事状态,其中所述观看者的所述一或多个预期生理上可观察状态包括所述观看者的预期叙事状态,其是与所述观看者的所述经评估叙事状态相同类型的叙事状态。

6. 根据权利要求3或权利要求4所述的方法,其中所述一或多个经评估生理上可观察状态包括所述观看者的经评估注意点,其中所述一或多个预期生理上可观察状态包括所述观看者的预期注意点。

7. 根据权利要求3或4所述的方法,其中所述一或多个信号修改选项的至少一个信号修改包括用于对以下各者中的一或多者实施媒体内容修改的指令:亮度、空间分辨率、对比度、色彩饱和度、色相、色调映射、视场、色域、亮度动态范围、位深度、空间滤波、图像刷新率、缩放因子、图像转向、非视觉特性、运动呈现特性、枢轴、亮度映射的斜率及偏移、特定图像区域、特定对象、特定人物、背景中的亮度分布、亮度、音频对象的位置、频率均衡、混响、音色、相位、扬声器数目、扬声器配置、扬声器的频率范围、扬声器的相位失真、喇叭选择、音量、实际音频通道配置、用于选择单个扬声器呈现且用于选择多扬声器内插的捕捉容差选项、音频对象位置、音频对象大小、音频对象半径、音频对象方向、音频对象轨迹、对话音量、非对话音量、对话增强、音频动态范围、特定喇叭选择、特定喇叭配置、回声特性、延迟、信号攻击时间、或信号释放时间。

8. 根据权利要求3或4所述的方法,其中所述一或多个生理监测信号由以下各者的一或多者产生:基于显示器的传感器、可见光波长相机传感器、同步定位及映射传感器、热成像器、头戴式显示器传感器、入耳式传感器、手腕传感器、注视位置传感器、瞳孔直径传感器、面部表情传感器、头位置传感器、观看距离传感器、效价传感器、唤醒传感器、脑电图传感器、明确定位的电极、热传感器、光学传感器、眼电图传感器、呼吸传感器、基于体积描记法心率的传感器、皮肤电响应传感器、气体传感器、CO<sub>2</sub>含量传感器、R3COH含量传感器、或基于座椅的传感器。

9. 根据权利要求3或4所述的方法,其中所述一或多个信号修改选项是至少部分基于回放装置特性化数据及呈现环境特性化数据产生。

10. 一种非暂时性计算机可读存储介质,其存储当由一或多个处理器执行时致使执行根据权利要求1到9中任一者所述的方法的软件指令。

11. 一种包括一或多个处理器及一或多个存储媒体的设备,所述存储媒体存储当由一或多个处理器执行时致使执行根据权利要求1到9中任一者所述的方法的一组指令。

## 经由生理监测的创作意图可扩展性

[0001] 相关申请案的交叉参考

[0002] 本申请案主张2019年7月2日申请的第62/869,703号美国临时专利申请案及2018年10月1日申请的第62/739,713号美国临时专利申请案的优先权权益,所述两个申请案特此以其全文引用方式并入。

### 技术领域

[0003] 本发明大体上涉及视听技术,且特定来说,涉及经由生理监测的跨回放装置的创作意图可扩展性。

### 背景技术

[0004] 当今视听生态系统包含观众对同一源视听内容的体验可大幅改变的各种各样的相异回放装置(例如,图像及/或声学再现等)。在许多情况中,观众针对不同回放装置的体验的显著变化引起基于其创作视听内容的创作意图的扭曲。

[0005] 此段中描述的方法是可采用方法,但并不一定是先前已经设想或采用的方法。因此,除非另外指示,否则不应假定此段中描述的方法中的任何方法仅凭借其包含于此段中就有资格作为现有技术。类似地,除非另外指示,否则关于一或多种方法识别的问题不应假设已经基于此段在任何现有技术中辨识出。

### 附图说明

[0006] 本发明在附图的图式中通过实例而非通过限制来说明,且其中相似的参考编号指示类似元件且其中:

[0007] 图1A描绘媒体内容递送管线的实例过程;图1B说明实例回放装置及在呈现环境中与回放装置一起操作的其它装置;图1C说明根据实施实例的实例音频编码及解码;

[0008] 图2A到图2C说明实例制作及消费阶段;图2D说明实例情绪预期元数据及叙事元数据;图2E到图2G说明实例生理监测及评估;图2H说明媒体内容的实例媒体特性及对媒体内容的媒体内容调整/修改;

[0009] 图3A到图3G说明由回放装置进行的针对生理上可观察状态及对应元数据的实例媒体呈现处理;

[0010] 图4A及图4B说明实例过程流;及

[0011] 图5说明本文中所描述的计算机或计算装置可在其上实施的实例硬件平台。

### 具体实施方式

[0012] 本文中描述涉及经由生理监测的创作意图可扩展性的实例实施实例。在以下描述中,出于解释的目的,陈述众多特定细节以便提供本发明的详尽理解。然而,应明白,可在没有这些特定细节的情况下实践本发明。在其它例子中,未详尽描述众所周知的结构及装置以便避免不必要地遮挡、模糊或混淆本发明。

[0013] 本文中根据以下提纲描述实例实施例：

- [0014] 1. 总体概述
- [0015] 2. 实例媒体内容递送处理管线
- [0016] 3. 创作意图
- [0017] 4. 情绪及表示
- [0018] 5. 生理监测及评估
- [0019] 6. 元数据控制及生理监测
- [0020] 7. 内容及元数据制作及消费
- [0021] 8. 情绪预期及叙事元数据格式
- [0022] 9. 信号分离与融合
- [0023] 10. 媒体内容调整或修改
- [0024] 11. 实例内容调整过程
- [0025] 12. 实例过程流程
- [0026] 13. 实施机制-硬件概述
- [0027] 14. 等效物、扩展、替代物及其它

[0028] 1. 总体概述

[0029] 此概述呈现本发明的实例实施例的一些方面的基本描述。应注意，此概述不是实例实施例的各方面的广泛或详尽概要。此外，应注意，此概述既不希望被理解为识别实例实施例的任何特别重要的方面或要素，也不希望被理解为描绘特定来说实例实施例或一般来说本发明的任何范围。此概述仅以精简及简化格式呈现了与实例实施例相关的一些概念，且应被理解为仅为以下实例实施例的更详细描述的前言。注意，尽管本文中讨论了单独的实施例，但本文中论述的实施例及/或部分实施例的任何组合可经组合以形成另外实施例。

[0030] 本文中所描述的技术可用于出于保留基于其创作视听内容的创作意图的目的而修改或调适由回放装置向观众或观看者呈现的视听内容。

[0031] 更明确来说，视听内容的呈现受对观看者的内部生理状态（例如情绪、认知、注意点等）的评估影响或适应于所述评估，所述评估通过对观看者的生理方面进行各种类型的监测获得或推断。观看者的内部生理状态的此评估与由视听内容的创意人员预期及/或希望且在视听内容及元数据制作阶段（或后制作阶段）期间被插入到视听内容的媒体元数据中的叙事及情绪预期组合。

[0032] 所插入到媒体元数据中或由媒体元数据所表示的视听内容的创作意图包含视听内容的创作者的情绪及叙事目标。另外、任选地或替代地，媒体元数据包含用于修改由回放装置从其接收媒体内容及元数据的接收到的视听信号的指令。

[0033] 在本文中所描述的技术下，例如人工情绪智力（或情绪AI）的情感计算可用于辨识、解译、模拟、估计或预测人类情绪、理解、行为等。计算模型（例如算法、方法、程序、操作等）可用于整合多个生理监测信号源，还可用于与用于含有向观看者描绘视听内容的媒体内容的源信号的最终呈现、再现及/或转导的媒体元数据及回放装置进行交互。因此，这些技术允许在媒体内容及元数据中所表示的创作意图能够跨许多类型的回放系统尽可能地最好地扩展。如本文中使用，可扩展性意味着本文中所描述的技术可跨各种各样的不同装置工作-例如小型智能手表装置、平板计算机、移动电话、膝上型计算机、高端回放装置、基

于大型剧院的系统、基于电影院的系统等-以防止或最小化与创作意图的偏离。

[0034] 如本文中使用时,呈现是指向单个观看者观众或多观看者观众呈现图像及/或音频内容的图像及/或音频处理操作。实例图像处理操作包含(无限制)时空、色彩、深度、裁剪、根据需要跨多个回放装置操纵图像信号等。实例音频处理操作包含(无限制)位置(例如方向、空间等)、均衡、混响、音色、相位、喇叭选择、音量等。本文中所描述的图像及音频处理两者都可为线性的、非线性的及/或自适应的。

[0035] 本文中描述的实例实施例涉及编码及/或提供用于从表示视听内容的媒体信号的回放来优化创作意图的媒体内容及元数据。接收描述与媒体内容的一或多个部分相关的情绪预期及叙事信息的创作意图输入。至少部分基于所述创作意图输入产生与所述媒体内容的所述一或多个部分相关的一或多个预期生理上可观察状态。将视听内容信号与所述媒体内容及包括所述媒体内容的所述一或多个部分的所述一或多个预期生理上可观察状态的媒体元数据一起提供到回放设备。所述视听内容信号致使所述回放装置(a)使用一或多个生理监测信号关于观看者确定与所述媒体内容的所述一或多个部分相关的一或多个经评估生理上可观察状态且(b)至少部分基于所述一或多个预期生理上可观察状态及所述一或多个经评估生理上可观察状态从所述媒体内容产生经修改媒体内容同时调整且向所述观看者呈现从所述媒体内容产生的经修改媒体内容。

[0036] 本文中描述的实例实施例涉及解码及/或消费经产生用于从表示视听内容的媒体信号的回放来优化创作意图的媒体内容及元数据。接收视听内容信号与媒体内容及媒体元数据。所述媒体元数据包括媒体内容的一或多个部分的一或多个预期生理上可观察状态。与所述媒体内容的所述一或多个部分相关的所述一或多个预期生理上可观察状态是至少部分基于描述与媒体内容的一或多个部分相关的情绪预期及叙事信息的创作意图输入而产生。使用一或多个生理监测信号关于观看者确定与所述媒体内容的所述一或多个部分相关的一或多个经评估生理上可观察状态。至少部分基于所述一或多个预期生理上可观察状态及所述一或多个经评估生理上可观察状态从所述媒体内容产生并呈现经修改媒体内容同时调整且向所述观看者呈现从所述媒体内容产生的所述经修改媒体内容。

[0037] 所属领域的技术人员将容易地明白对优选实施例的各种修改及本文中描述的一般原理及特征。因此,本发明并不希望限于展示的实施例,而是被赋予与本文中描述的原理及特征一致的最宽范围。

[0038] 2. 实例媒体内容递送处理管线

[0039] 图1A描绘展示从媒体内容捕获/产生到回放装置的各个阶段的媒体内容递送管线100的实例过程。媒体内容递送管线(100)的一些或全部处理块可用一或多个计算机装置实施于硬件、软件、硬件与软件的组合等等中。

[0040] 本文中所描述的实例回放装置可包含(但不限于)移动装置、基于剧院的装置、增强现实(AR)装置、虚拟现实(VR)装置、计算机游戏装置、TV、家庭影院、头戴式装置、穿戴式装置等。

[0041] 如图1A中说明,视听数据102使用媒体内容制作块105捕获或产生。视听数据(102)可被数字化地捕获(例如,被数码相机及/或数字音频记录器等捕获)或由计算机产生(例如,使用计算机动画及/或计算机创作/合成、使用图像呈现模型等)以在实时或非实时操作中提供初始媒体内容107。另外、任选地或替代地,视听数据(102)可经捕获及存储为记录于

有形媒体上的模拟信号。经捕获或经记录模拟信号任选地被读取及转换成数字格式以提供初始媒体内容(107)的至少一部分。

[0042] 本文中所述的实例视听数据及/或初始媒体内容可包含(但不一定仅限于)以下各者中的任一者:仅音频数据,例如音频样本或音频帧/块中的变换系数;仅视频数据,例如图像像素值或图像帧/块中的变换系数;音频与视频数据的组合,具有或不具有独立于音频数据的音频元数据、具有或不具有独立于视频数据的图像元数据、具有或不具有其它多媒体及/或文本数据等。

[0043] 如图1A中展示,初始媒体内容(107)经提供到媒体制作块115且在被递送到视频递送管线(100)中的下一处理阶段(stage/phase)之前由媒体制作块115根据创作者的意图进行编辑或变换成发行版本(例如以不同用户群体为目标的多个发行版本当中的单个发行版本等)。发行版本包括媒体元数据117-1及对应媒体内容117-2。

[0044] 媒体制作块(115)可用一或多个音频编辑或创作装置、一或多个视频编辑或创作装置、参考音频呈现装置、及/或参考视频呈现装置来实施。一些或全部这些装置可(但不限于)与电影制片厂、商业媒体制作系统、基于家庭的媒体制作系统中的创作者(例如创意人员、创意用户等)一起操作及交互。在一些操作场景中,媒体制作块(115)包括以下各者中的一或多个者:色彩分级站、参考显示器装置、音频混合器、音频编辑器、元数据生成器等。

[0045] 发行版本的创作者-包含(但不限于)电影制片厂指定的专业人员、媒体制作人员、一或多个视频/音频专业人员、业余视频/音频内容创作者等-与媒体制作块(115)交互以将(创意)用户输入或创作意图输入提供到媒体制作块(115)且致使媒体制作块(115)执行声音元素的选择、音频混合及编辑(例如,在来自现场或经记录音频元素、媒体制作块(115)可存取的声音库或工具包等的初始媒体内容(107)中)以产生媒体内容(117-2)的音频内容。同样地,发行版本的创作者可与媒体制作块(115)交互以将(创意)用户输入提供到媒体制作块(115)且致使媒体制作块(115)选择、编辑、创作、设置视觉元素的色调、饱和度、色相、及色彩(例如,在来自媒体制作块(115)可存取的视觉库或工具包等的初始媒体内容(107)中)以产生媒体内容(117-2)的视觉内容。

[0046] 本文中所述的音频元素的选择、音频混合及编辑可包含(但不一定仅限于)以下各者中的一或多个者:选择、混合及/或编辑声音元素。音频选择、混合及/或编辑可在显著或最少手动用户输入下执行(例如,在预记录音频/视听制作等的情况中)、部分或完全自动地执行(例如,在具有很少或无用户输入/交互等的情况下)、根据预定参数/算法/程序执行(例如,在现场广播等的情况中)、自动执行及/或用户辅助的音频混合及编辑操作的组合等等。实例音频或声音元素可包含(但不一定仅限于)以下各者中的任一者:声学元素、音频元素、原声、音效、对话、会话、拟音效果(Foley effects)、来自乐器的音乐或人类语音、来自对象及/或动物的声音、自然声音、人工声音、环境声音、静止声音元素、移动声音元素等。

[0047] 选择、编辑、创作、设置视觉元素的色调、饱和度、色相、及色彩可包含对视觉元素执行色彩分级(或“色彩调整”)以产生将包含于媒体内容(117-2)中的视觉内容。包含(但不限于)色彩分级的这些操作可在显著或最少手动用户输入下执行(例如,在预记录视觉/视听制作等的情况中)、部分或完全自动地执行(例如,在具有很少或无用户输入/交互等的情况下)、根据预定参数/算法/程序执行(例如,在现场广播等的情况中)、自动执行及/或用户辅助的音频混合及编辑操作的组合等等。实例视觉或图像元素可包含(但不一定仅限于)以

下各者中的任一者：视觉对象、视觉特性、图像特征、视觉效果、描绘人类的图像或图像部分、描绘对象及/或动物的图像或图像部分、现实生活图像、人工图像、背景、静止视觉元素、移动视觉元素等。

[0048] 虽然是通过创作者与媒体制作块(115)之间的交互产生,但媒体内容(117-2)的音频内容可在参考呈现/制作环境中由创作者呈现、监听及/或连续调整,直到媒体内容(117-2)的音频内容中表示的声音元素在参考呈现/再现环境中以与创作者的创作意图一致或以其它方式表达创作者的创作意图的所期望质量/效果被呈现/再现/感知。同样地,媒体内容(117-2)的视觉内容可在参考呈现/制作环境中由创作者呈现、观看及/或连续调整,直到在媒体内容(117-2)的视觉内容中表示的视觉元素在参考呈现/再现环境中以与创作者的创作意图一致或以其它方式表达创作者的创作意图的所期望质量/效果被呈现/再现/感知。

[0049] 发行版本中的媒体内容(117-2)可包含(但不一定仅限于)以下各者中的任一者:仅音频数据,例如音频样本或音频帧/块中的变换系数;仅视频数据,例如图像像素值或图像帧/块中的变换系数;音频与视频数据的组合,具有或不具有独立于音频数据的音频元数据、具有或不具有独立于视频数据的图像元数据、具有或不具有其它多媒体及/或文本数据等。实例媒体内容可包含(但不一定仅限于)以下各者中的一或多个者:电视节目、媒体节目、视听节目、现场广播、媒体流式传输会话、电影等。

[0050] 作为从初始媒体内容(107)产生发行版本的一部分,媒体制作块(115)还产生或制作对应于媒体内容(117-2)的媒体元数据(117-1)。媒体元数据(117-1)包含(但不一定仅限于)以下各者中的一些或全部:音频元数据、图像元数据、情绪预期元数据、叙事元数据等。

[0051] 媒体元数据(117-1)中的音频及/或图像元数据可包含将用于音频及/或图像处理操作中的相对低级操作参数。媒体元数据(117-1)中的音频及/或图像元数据可包含与生理监测相关(例如,直接相关等)的元数据部分,以及与生理监测不相关(例如,不直接相关等)的元数据部分。

[0052] 音频及/或图像元数据中的一些或全部操作参数的值集可为内容特定的。举例来说,(分别地)包含于用于关于特定图像、特定视觉场景、特定音频帧、特定音频场景等将执行的音频或图像处理操作的音频或图像元数据中的操作参数可经设置有取决于(分别地)特定图像、特定视觉场景、特定音频帧、特定音频场景等中的特定像素值、特定音频样本值、特定像素值分布及/或音频样本值等的值。

[0053] 另外、任选地或替代地,一些或全部操作参数的值集可为装置特定的。举例来说,(分别地)包含于用于将由特定回放装置(或与其一起操作的装置)执行的音频或图像处理操作的音频或图像元数据中的操作参数可经设置有取决于特定回放装置、其系统配置、其图像显示器或音频呈现能力、其操作、呈现及/或再现环境、连同特定回放装置一起操作的其它装置等的值。

[0054] 媒体元数据(117-1)中的情绪预期及/或叙事元数据(或简单地“E&N元数据”)包含基于通过(创意)用户输入至少部分传达到媒体制作块(115)的创作者的意图产生的时间相依预期情绪状态及/或认知状态。预期情绪状态及/或认知状态表示内容创作者预期观看者在媒体内容(117-2)被调整且通过各种回放装置向观看者呈现时将处于的或将具有的目标生理上可观察(或待监测)的状态。

[0055] 应注意,在各个实施例中,创意人员可预期针对给定时间点或所描绘场景的单个

情绪(或单个情绪类型)或数个情绪(或数个情绪类型)。举例来说,一个观看者可选择认同所描绘故事中的一个面(例如,好的一面等),而不同观看者可选择认同所描绘的同一故事中的不同面(例如,邪恶的一面等)。因此,创意人员针对给定观看者可能预期两种情绪状态,这取决于观看者处于哪些面。如果观看者选择的那一面失败了,那么创意人员将针对观看者预期的第一情绪状态可为“同情”。如果观看者碰巧选择了相反的面,那么当描绘相同故事情节信息时创意人员将针对观看者预期的第二情绪状态可为“快乐”。

[0056] 虽然至少一些E&N元数据可在创作者正创作媒体内容(117-2)的发行版本且与媒体制作块(115)交互时基于由内容创作者提供的用户输入产生,但一些或全部E&N元数据还可基于包含(但不限于)以下各者的不同创意输入传达机构产生:与发行版本中描述的故事的情绪预期或叙事信息相关的(例如非交互、非实时、脱机等)故事板。较新技术使用呈电子文本形式的数字故事板制作及脚本。而且,最初仅仅用于计算机图形的视觉化(Previs/previsualization)现在用于现场相机捕获,且相关联软件为导演发表意见提供了场所。

[0057] 应注意,与不实施本文中所描述的技术的其它方法相比,本文中所描述的技术下的媒体元数据(117-1)中的E&N元数据将(例如,相对紧密地等)与在呈现媒体元数据(117-2)时由回放装置执行的生理监测及评估操作耦合或一起使用。回放装置使用E&N元数据及生理监测及评估两者以导出媒体内容(117-2)并根据需要对所述媒体内容进行媒体内容调整或修改以在向观众呈现媒体内容(117-2)时保留创作者的意图或避免创作者的意图扭曲。

[0058] E&N元数据可包括分别针对媒体内容(117-2)中的将在一或多个时间点、一或多个时间间隔、在一或多个场景等中呈现的一或多个数据部分的一或多个元数据部分。媒体元数据(117-1)的E&N元数据中的每一元数据部分可指定将在一时间点、时间间隔、场景等处呈现的媒体内容(117-2)中的相应数据部分的例如预期情绪状态及/或预期认知状态(或简单地,E&N状态)的生理上可观察状态。

[0059] 预期(或目标)E&N状态可以一或多种监测装置特定方式指定。举例来说,预期E&N状态可经指定为预期测量/评估结果,其是皮肤电响应或GSR特定的、眼电图或EOG特定的、脑电图或EEG特定的、特定于面部表面分析、特定于测瞳术等等。观看者的“叙事”状态可(例如,通常、有时等)称为认知状态。为了支持在不同呈现环境中可与各种回放装置一起操作(或可经配置有各种回放装置)的不同监测装置或技术,可针对单个预期E&N状态指定一种以上监测装置特定(或一种以上呈现环境特定)方式。

[0060] 另外、任选地或替代地,预期E&N状态可以通用于生理监测装置或技术的方式指定。接收发行版本中的一般指定的预期E&N状态的回放装置可将预期E&N状态映射到与回放装置一起操作(或经配置有回放装置)的特定可用监测装置或技术的预期测量/评估结果。

[0061] 除了指示预期E&N状态之外,媒体元数据(117-1)中的E&N元数据还可承载内容修改元数据部分,包含(但不限于)信号修改选项、所关注(图像)区域(ROI)、所关注音频或声学对象(AOI)、伴随操作参数等。内容修改元数据部分可由回放装置用来实现基于预期E&N状态及生理监测及评估作出的媒体内容调整或修改同时调整且向观众呈现如所调整或修改的媒体内容(117-2)。

[0062] 在实例中,内容修改元数据部分可将媒体内容(117-2)的数据部分中的一或多个(例如,关键等)声音元素指示或识别为实现媒体内容调整或修改的音频处理操作可对其作

出(例如,声学、位置、扩散、音色、响度等)调整或修改的一或多个AOI。

[0063] 在另一实例中,内容修改元数据部分可将媒体内容(117-2)的数据部分中的一或多个视觉元素或区指示或识别为实现媒体内容调整或修改的音频处理操作可对其作出(例如,亮度、空间分辨率、对比度、色彩饱和度、色调映射等)调整或修改的一或多个ROI。

[0064] 在内容消费期间,响应于确定观看者的经评估E&N状态背离预期E&N状态,回放装置可使用一或多个内容修改元数据部分产生或对媒体内容(117-2)实施调整/修改以引导观看者的注意点朝向(或在一些情况中,可能远离)媒体内容(117-2)中描绘的AOI及/或ROI且因此致使观看者的(后续)经评估E&N状态收敛于E&N元数据中所指示的观看者的预期E&N状态。

[0065] 可使发行版本可用于在各种呈现/再现环境中操作的回放装置。媒体制作块(115)可在不同于回放装置在其下操作的实际呈现环境的参考呈现环境下操作。一些或全部媒体内容(117-2)及媒体元数据(117-1)可参考参考及/或零个或更多个其它(例如,目标等)呈现环境指定。在不同于发行版本是参考其创建的呈现环境的特定(或实际)呈现环境下操作的回放装置从参考呈现环境到特定呈现环境可采用发行版本中的一些或全部媒体内容(117-2)及媒体元数据(117-1)。

[0066] 本文中所描述的发行版本中的对应媒体元数据及媒体内容可经编码(例如,经编码有单独元数据或数据容器等)于一或多个经编码位流(例如视频信号灯)中、被记录在有形计算机可读存储媒体上、及/或传输或递送到接收者装置(例如,接收者回放装置、与一或多个回放装置一起操作的接收者装置等)。

[0067] 如图1A中说明,媒体编码块120从媒体制作块(115)接收包括媒体内容(117-2)及媒体元数据(117-1)的发行版本;将所述发行版本编码成经编码位流122。如本文中使用的,经编码位流可指代音频信号、视频信号、视听信号、包括一或多个子流的媒体数据流等等。媒体编码块(120)包括用于产生经编码位流(122)的一或多个音频及视频编码器,例如由ATSC、DVB、DVD、蓝光、及其它递送格式定义的一或多个音频及视频编码器。

[0068] 经编码位流(122)向下游递送到一或多个接收器或接收者装置或回放装置,包含(但不限于)解码器、媒体源装置、媒体流式传输客户端装置、电视机(例如智能TV等)、机顶盒、电影院或类似物。

[0069] 如图1A中说明,在回放装置中,经编码位流(122)由媒体解码块130解码以产生经解码媒体元数据132-1及经解码媒体内容132-2。媒体解码块(130)包括用于解码经编码位流(122)的一或多个音频及视频解码器,例如由ATSC、DVB、DVD、蓝光、及其它递送格式定义的一或多个音频及视频解码器。

[0070] 经解码媒体元数据(132-1)可包含且可等同于由媒体编码块(120)编码(例如,在无损压缩等下)到经编码位流(122)中的一些或全部媒体元数据(117-1)。经解码媒体内容(132-2)可等同于或对应于经受量化及/或通过由媒体编码块(120)执行的(例如,有损等)压缩及由媒体解码块(130)执行的解压引起的编码错误的媒体内容(117-2)。

[0071] 经解码媒体元数据(132-1)可与经解码媒体内容(132-2)一起由回放装置、或连同回放装置一起操作的音频及/或图像呈现装置135使用,以按保护发行版本已经使用其产生的创作者的意图、或最小化或避免对所述意图的扭曲方式执行生理监测、生理状态评估、媒体内容调整或修改、音频处理、视频处理、音频再现/转导、图像呈现/再现等等。

[0072] 图1B说明实例回放装置及在呈现环境中与回放装置一起操作的其它装置。这些装置及其中的组件中的任一者可使用硬件、软件、硬件与软件的组合等实施。这些装置可包含媒体解码块(例如图1A的130等)及音频及/或图像呈现装置(例如图1A的135等)。

[0073] 如展示,单独观看者(或观众)正观看基于媒体内容(例如图1A的117-2等)在例如平板计算机的回放装置的显示器上呈现的图像且正在收听基于媒体内容(117-2)通过智能耳塞装置呈现的音频(对应于或伴随经呈现图像)。一或多个生理监测组件可包含捕获将由与回放装置一起部署的面部表情分析软件分析的观看者的面部表情的相机镜头、与智能耳塞装置及/或智能手表装置一起部署的传感器及/或电极等。这些组件可经部署以在观看者的头部或身体的各个位置处进行生理监测。从生理监测产生的感官信号及/或数据可经处理以评估在媒体内容(117-2)正被调整且向观看者呈现时观看者的E&N状态。在显示器上及使用耳塞装置呈现的内容可基于由回放装置接收的媒体元数据(例如图1A的117-1等)与媒体内容(117-2)及生理监测/评估进行调整或修改。

[0074] 仅出于说明目的,已描述,可对例如图1B中说明的单独观看者执行生理监测及评估。应注意,本文中所描述的技术意在用于范围从单个观看者到剧院中的一大群观众的观众。虽然图1B展示具有嵌入于回放装置或在特定呈现环境中与回放装置一起操作的例如智能手表装置的装置的部分中的一些可能生理监测组件及传感器的观看者,但生理监测可经实施及部署于涉及具有高达十分大群体的多个观看者的观众(例如,剧院、体育场、赛事等等中的观众)的呈现环境(例如图2F中所说明等)中。

[0075] 还应注意,媒体消费可仅涉及音频,但在本发明中的所有情况中可能不会调出“观看者”与“收听者”之间的区别。观看者及/或收听者在本文中可统称为观看者。

### [0076] 3. 创作意图

[0077] 例如创作者的意图、创作意图、艺术意图、导演的意图、制作人的意图及批准者的意图及类似物的各种术语是尚未严格定义的类似术语的实例。术语“艺术意图”源于文学、绘画、雕塑及艺术哲学世界且最初用于独奏艺术家。其它术语是源于影片制作的修改,其中在整个制作中涉及规模大得多的工作人员。关于媒体产品的最终外观及声音的决策可由导演、制作人、以及配色师、电影摄影师、乐师、及针对作品的特定方面的录音师作出。SMPTE现在使用术语批准者的意图,其了解作出用于确定将发布的版本的最终决策的人当中的广泛可变性。这包含媒体内容的编辑、整体外观、及声音,以及特定观看者/收听者群体期望的变化。

[0078] 由于本发明涉及媒体制作阶段的细节,例如涉及与技术人员及艺术工作人员的交互,所以术语“创作意图”或“创作者的意图”用于描述来自媒体作品的创意工作人员的各种目标。

[0079] 针对例如叙事影片、有声读物、音乐剧、歌剧及类似物的叙事媒体内容,创作意图可包括各种元素,但大多数落于以下种类或方面中的一者中:

[0080] -故事所需要(或将针对故事传达)的叙事信息,例如观看者是否能够感知组成叙事的各种信息元素,所描绘视觉元素是否是可见的,经呈现声音元素是否是可听的,所描绘或经呈现视觉/声音元素是否处于正确的位置,对话是否是可理解的,特定回放装置及回放装置周围的环境如何影响一些或全部视觉及/或声音元素等等。

[0081] -故事的情绪预期,例如观看者是否体验到创作者期望的预期情绪、色彩/对比度/

音色/动画在范围及准确度两个方面如何影响这些唤醒、效价等等。出于本发明的目的,例如由于刺耳时刻造成的低级生理响应/响应可被分类为一种类型的情绪。

[0082] -预期将被观看者感知且了解的由创作者所期望的审美。创作者(或艺术家)可有目的地选择特定色彩,无论是由于心血来潮、根据感觉、个人色彩和谐理解、象征意义等。本文中所描述的系统力图在各种呈现环境中匹配期望审美。在其中由创作者所期望的特定呈现选项在实际呈现环境中不可用或在技术上不可行的操作案例中,系统可确定在实际呈现环境中是否存在可用的仍满足艺术家的创作意图的其它可接受呈现选项。举例来说,改变音乐音调来适应歌手的音域作为保留或尊重创作意图的替代通常是可接受的。类似或替代视觉元素还可用作用于保留或尊重创作意图的替代。

[0083] 在一些操作案例中,审美的不对称性可能会成为一个有意义的问题。举例来说,如果回放装置的显示能力受到限制,那么在图像内容回放时减小色彩饱和度可为可接受的呈现选项;然而,即使回放装置具有比用于设置或表达创作意图的范围更大的范围,增大色彩饱和度也无法始终作为可接受的呈现选项。

[0084] -消息。故事情节、勾起的情绪、审美等的组合可用于传达消息的主要目的。然而,可存在其中创作意图仅预期集中在上文所识别的三个种类或方面且没有总体消息的情况。

[0085] -恢复。视听内容可含有指示或展现仅可被专家或鉴赏家了解的技术或创意品质的创意方面、特征及细节,专家或鉴赏家可能只代表内容创作者期望的总体内容消费者群体的限度或一小部分。

[0086] 本文中所描述的技术可以解决或可能不能解决上文所论述的全部五个种类或方面的创作意图。在一些操作案例中,本文中所描述的系统仅可以解决五个种类或方面的创作意图的子集,例如前两者:叙事信息方面及情绪预期或效果方面。另外、任选地或替代地,系统可以在认为审美对叙述故事是重要的时通过叙事信息方面或在认为审美对影响或引发情绪预期或效果是重要的时通过情绪预期或效果方面来解决第三方面,审美。

[0087] 4. 情绪及表现

[0088] 存在关于情绪的几个分类法,范围从相对小的数目(例如从埃尔克曼(Elkman)理论六种)到含有更多细微差别且包含差不多三十种不同情绪的其它数目。

[0089] 一些情绪具有对应面部表情,而其它情绪涉及对裸眼或其它视觉技术来说不具有可见标志的更深的内心感受。具有-或伴随有对应-面部表情的情绪可被视为相异集合,这是因为其在一些操作案例中可以最容易地评估。举例来说,指向观众或单独观看者的相机或面部表情分析软件可用于获得伴随对应面部表情的那些情绪的估计。

[0090] 下文表1展示四种人类情绪分类法(其相应数量在括号中)以及可从面部表情分析或确定的实例子集。

[0091] 表1

[0092]

埃尔克曼 (Elkman)	普拉切克 (Plutchik)	核心匿名 (7)	考恩(Cowan) 理论 (27)	面部表情 (9)
理论 (6)	理论 (8)			

[0093]

欢乐	欢乐	快乐(欢乐)	欢乐	快乐(欢乐)
惊讶	惊讶	惊讶	惊讶	惊讶
悲伤	悲伤	悲伤	悲伤	悲伤
愤怒	愤怒	愤怒	愤怒	愤怒
反感	反感	反感	反感	反感
恐惧	恐惧	恐惧	恐惧	恐惧
	期待		激动(可能地, 期待)	
	信赖			
		鄙视	鄙视	鄙视
			平静	中性
			厌烦	厌烦
			尴尬	
			焦虑	
			恐惧	
			浪漫	
			性欲	
			怀旧	
			困惑	
			恍惚(可能地, 惊愕)	
			娱乐	
			崇拜	
			钦佩	
			敬畏	
			审美鉴赏	
			渴望	
			感兴趣(可能地, 期待)	
			满意	
			轻松	

[0094] 未由上文表1中所列举的特定理论引述的其它熟悉的情绪可包含(无限制)警惕、悲痛、发怒、厌恶、狂喜等。在一些操作案例中,未引述的情绪可近似地被映射到上文表1的列表中的其它对应同义词。举例来说,几种情绪可被映射如下:警惕~感兴趣、悲痛~悲伤、发怒~愤怒、厌恶~鄙视、狂喜~浪漫或性欲或惊愕等,其中两种情绪之间中的“~”标示从前一情绪到后一情绪的映射。有时,针对一种情绪的词的变化仅是同一情绪的量级变化。举

例来说,悲痛是更强幅度版本的悲伤。

[0095] 如本文中使用的术语“沉浸感”可意味着,在观看-其如本文中使用的,可包含在视觉上看及/或在听觉上听以及可能地感知运动-由回放装置所呈现的媒体内容(例如音频内容、视觉内容等)时,观看者感觉好像真的处于媒体内容中描绘的故事世界中一样。

[0096] 沉浸感可通过呈现环境中的真实感图像及音频呈现能力实现,例如更宽的视场(FOV)、更宽的色域、增加的(亮度)动态范围、更高位精度、更高保真度定位声音等等。通过比较,具有相对低图像及音频呈现能力的呈现环境中的观看者可不断地从小屏幕显示器上的窄FOV展示看到图像边框,借此被阻止具有沉浸感的感受。

[0097] 因此,图像及音频呈现能力可进一步经改进以避免或减小相对容易被较低图像及音频呈现能力引起的视觉或听觉注意力分散。然而,应注意,虽然通过技术实现的沉浸感及观看者参与度通常可密切相关,但呈现能力与沉浸感之间的相关性不是绝对的且并非在所有情况下都适用。举例来说,如果书中描绘的故事是以一种扣人心弦的吸引人的方式来讲述的,那么例如书的相对“低”技术能力仍可致使读者感觉彻底沉浸在书中描绘的故事中。相反,如果VR游戏是令人反感的派生物或很无聊,那么高端专业呈现环境中的例如VR游戏的相对高技术能力仍然可能无法使游戏用户参与进来或无法致使游戏用户感觉沉浸其中。

[0098] 虽然在上文表1的情绪中未明确解释,但沉浸感在许多操作案例中作为在表1中所识别的一些或全部情绪的放大器对媒体生态系统(或呈现环境)及回放装置具有重要意义。因此,无需直接或单独地评估/测量/量化沉浸感,而是通过已经通过沉浸感放大或增强的可检测、可评估、可测量及/或可量化的情绪在这些操作案例中施加其影响。

[0099] 本文中所描述的情绪可以多种方式来表示。在一些操作案例中,可离散地表示情绪。举例来说,情绪可经特性化或评估成例如上文表1中识别的情绪类型的特定情绪类型(例如,通过面部跟踪及面部表情分析等)且具有各种强度等级,例如五(5)个等级、少于五(5)个等级、或多于五(5)个等级。情绪的各种强度等级在生理状态评估最低点处可能压缩到中性。

[0100] 在一些操作案例中,情绪可使用连续值来特性化或评估。举例来说,情绪可使用分别表示唤醒及效价的两个笛卡尔(Cartesian)轴来表示或建模。唤醒实质上是上文所论述的强度的量级或等级,而效价确定与情绪相关的人类情绪是正还是负。举例来说,使用一或多个生理监测探针或传感器所获得的效价的正值测量可指示积极情绪,而使用生理监测探针或传感器所获得的效价的负值测量可指示消极情绪。

[0101] 此基于坐标的情绪建模可用作通过生理监测探针或传感器获得的一些生理测量及/或评估,且对应计算机实施分析仅可识别、量化、测量及/或评估内在情绪的唤醒及效价等级,这取决于可用生理监测及评估技术。GSR是其中仅可在此时评估的情绪的唤醒及效价的实例。

[0102] 另外、任选地或替代地,本文中所描述的系统可连同其它类型的情绪建模或表示-例如与面部情绪的IAPS(互动情感画面系统)相关联的标准化模型一起操作。此情绪建模或表示可用于识别、量化、测量及/或评估唤醒及效价以及可能地内在情绪的其它方面(例如,优势度等)。

[0103] 5.生理监测及评估

[0104] 本文中所描述的技术可与各种各样的生理监测及/或评估技术的任何组合一起操

作以监测及/或评估观看者的情绪状态、认知状态等。

[0105] 生理监测装置/传感器/电极可包含(但不一定仅限于)以下各者中的一或多个者:具有监测装置/传感器/电极的头戴式显示器(HMD)、眼睛周围的监测装置/传感器、电极、具有(或入耳式)监测装置/传感器/电极的耳塞、EOG装置、EEG装置、视线跟踪器、气体含量监测器、测瞳术监测器、与特定回放装置一起部署的监测装置、与特定呈现环境一起部署的监测装置等等。

[0106] 一些监测/评估技术可直接并入在作为回放装置或系统的一部分的图像显示器上,而一些其它监测/评估技术可通过连同回放装置一起操作的单独、辅助、外围、智能耳塞及/或智能手表装置并入。在具有相当多的观众的媒体消费应用中,例如在电影院或剧院中,范围从面向观众的一或多个相机到放置在座位中的传感器、到整个电影院或剧院的测量(例如气体含量或温度)等的生理监测/评估技术可经实施或用于监测及/或评估多个观看者的共同及/或个别情绪状态、认知状态等。

[0107] 本文中所述的系统可产生个别生理状态评估以及群体生理状态评估。实例生理监测/评估技术包含(但不一定仅限于)以下各者中的一或多个者:经由EOG的视线跟踪、经由EEG的认知状态、经由EEG的听觉注意力、经由测瞳术的情绪及/或叙事状态等等。

[0108] 生理状态评估可分成两个种类或方面:认知及情绪,其分别可经映射到先前所识别的创作意图的叙事信息方面及情绪预期或效果方面。认知的生理状态评估涉及认知负荷,其指示观看者是否以及在多大程度上努力理解对故事情节来说重要的元素。

[0109] 参与度是对情绪及认知来说重要的内部注意力状态。内部注意力状态可(但不限于)通过眼睛跟踪器测量,例如将观看者的注视位置映射到经呈现媒体内容中的特定音频或视觉元素上。此类眼睛跟踪器可在视频显示器应用、虚拟现实(VR)应用、增强现实(AR)应用等中经构建到图像显示器(例如TV、移动显示器、计算机监测器等)中。

[0110] 观看者对所描述故事的参与度(或内部注意力状态)可使用EEG通过P300诱发的潜在响应来(定性地/定量地)评估。通过P300诱发的潜在响应所确定的电场电势减小指示观看者部分的参与度或注意力比其它情况好。

[0111] 在一些操作案例中,参与度可被视为情绪的子集。在这些操作案例中,对由回放装置呈现的各种视觉及/或音频元素的预期参与水平(或注意力水平)可在媒体元数据中指定为情绪预期元数据。

[0112] 在一些其它操作案例中,参与度(或注意力)可被视为认知的子集,而非被视为情绪的子集。对由回放装置呈现的各种视觉及/或音频元素的预期参与水平(或注意力水平)可在媒体元数据中指定为叙事信息(或认知)元数据。

[0113] 本文中所述的技术可经实施以支持分类、表示及/或测量情绪及其维度/等级/强度的不同方法。在一些操作案例中,可依据效价及唤醒(例如,唤醒及效价的连续值、唤醒及效价的连续值的范围等)监测、测量及/或评估(例如,通过生理监测装置/传感器/电极等)情绪。在一些操作案例中,可依据一组相异(但相关)情绪(例如,一组相异(但相关)情绪的离散类型值、一组相异(但相关)情绪的离散整数表示、一组相异(但相关)情绪的分类等)监测、测量及/或评估(例如,通过生理监测装置/传感器/电极等)情绪。

[0114] 可从例如通过可见光、热成像相机等获取的观看者的面部的影像读取特定情绪。一或多个面部表情方法、算法及/或程序可用于通过在获取的影像中捕获的面部表情评估

或读取观看者的内心状态或情绪。从热图像而非可见光图像读取观看者的内心状态或情绪可提供及供应比在读取可见光图像的情况下可能的达到理解程度相对更深的对观看者的内心状态或情绪的理解,这是由于可见光图像可被观看者的“扑克脸”掩饰,而热图像可能不能那么容易地被此“扑克脸”掩饰。

[0115] 为了评估非可见情绪,脑电图 (EEG) 感官数据收集方法可使用安置有在多个地方触摸观看者的头部的电极 (例如,几十个电极、仅少数电极等) 的头盔中实施。EEG 感官数据收集方法还可通过与头带、耳罩式耳机 (或头戴听筒)、帽子的一部分等一起部署、嵌入及/或安置的电极实施。在例如 VR 应用及类似物的一些应用中,多传感器 EEG 系统或组合件可经构建到头戴式显示器 (HMD) 中。而且,收集 EEG 感官数据的相对无害方式可通过放置于智能耳塞中的电极开发或实施。

[0116] 如先前提及,一些生理监测/评估技术允许、支持、读取 (例如,仅在例如优势度的其它维度下等) 唤醒及效价,例如 GSR,其也可称为外胚层活动 (EDA)、皮肤传导、皮肤电响应 (EDR)、心理电响应 (PGR)、皮肤传导响应 (SCR)、交感皮肤响应 (SSR)、皮肤传导水平 (SCL)、或类似物。心率及呼吸监测是 (例如,仅,等) 可监测或评估内在情绪的唤醒水平的生理监测/评估实例。

[0117] 6. 元数据控制及生理监测

[0118] 本文中所述的媒体呈现操作可在元数据控制下进行。如先前提及,媒体元数据可与对应媒体内容一起插入及/或嵌入于被传输及/或递送到例如回放装置的下游接收者装置的经编码位流、媒体文件等中。媒体元数据可包含例如针对杜比视界、三星 (Samsung) 的 HDR 10+、彩色印片先进 HDR、杜比 ATMOS 等产生的元数据部分的元数据部分。一些或全部媒体元数据可与媒体内容一起插入及嵌入于经编码位流、媒体文件等。接收者回放装置可使用媒体元数据调适或更改例如关于目标环境中的实际图像显示器的 (亮度) 动态范围、色彩饱和度、色相、空间滤波等,且使用音频元数据更改在部署于目标环境中的实际音频扬声器/声道配置下的音频呈现/再现。

[0119] 媒体元数据进一步包括 E&N 元数据,例如预期情绪状态、预期认知状态 (例如认知负荷等)、内容修改元数据及类似物。E&N 元数据中的情绪预期元数据可用于描述上文在表 1 中所列举的一组情绪或例如唤醒及效价的一组情绪维度。在一些操作案例中,描述于情绪预期元数据中的一组情绪的一些或全部情绪可使用面部表情提取技术监测、测量、估计、确定及/或评估。描述于情绪预期元数据中的一组情绪中的一些或全部情绪可使用 EEG、测瞳术、例如热及 GSR 的其它生理状态评估技术、不同生理状态评估技术的组合等等进行监测、测量、估计、确定及/或评估。

[0120] 在媒体内容及元数据制作阶段中,在情绪预期元数据中无需使用、包含及/或描述上文在表 1 中所列举的所有情绪。在媒体消费阶段 (例如,使用回放装置实施) 中,由媒体回放装置无需监测、测量、估计、确定及/或评估所有情绪。一些情绪可比其它情绪更适用于具有特定回放装置的特定呈现环境中。一些情绪可比其它情绪更适用于特定观看者。

[0121] 应注意,不同技术领域或学科可使用同义或在含义上具有实质重叠的不同技术术语。一些技术术语倾向于由创意人员 (或媒体内容的创作者) 使用,而一些其它技术术语倾向于由神经科学专业人员或专家使用。如与通俗说法或用法相比较,技术术语在学科或领域中可具有特异性的优点。在此档案中使用最适合与实施本文中所描述的技术的系统的每

一特定部分交互的术语的术语。因此,针对涉及元数据的插入的步骤,如由创意人员在媒体制作阶段中进行,使用创意人员更熟悉的术语。通过比较,针对涉及生理信号的处理的步骤,使用最适合神经科学的术语。

[0122] 具有重叠含义的实例术语是术语“混乱”,其与认知状态及混乱估计相关。术语“混乱”是创意人员使用的更适当术语,而具有重叠含义的术语“认知负荷”是可使用后一术语来描述或指示混乱水平的神经科学家使用的更适当术语。作为技术术语,认知负荷在神经科学中具有额外特异性,这是由于所述术语包含从非常紧张的混乱到仅需要注意的精神状态的分级。

[0123] 7. 内容及元数据制作及消费

[0124] 图2A说明其中媒体元数据是针对对应媒体内容产生的实例制作阶段202及其中经产生媒体元数据与生理监测及评估一起用于在媒体内容跨不同回放装置呈现时支持创作意图可扩展性的实例消费阶段204。

[0125] 媒体制作块(例如图1A的115等)使用在图2A中说明为沿着正或递增时间方向的单向箭头的(预期)内容时间线210产生媒体内容(例如图1A的117-2等)。媒体内容(117-2)由例如多个音频帧、多个图像帧等的多个数据部分组成。内容时间线(210)指示预期媒体内容中的多个数据部分沿着其被回放装置回放的时间线。更明确来说,多个数据部分中的每一数据部分可经指定(例如,由创意人员等指定)以由各个回放装置在沿着内容时间线(210)的多个时间点中的特定时间点回放。应注意,由特定回放装置所实施的实际回放时间线可在某一程度上按在媒体内容(117-2)的实际呈现中存在的时间漂移、时钟差异、用户或装置动作(例如暂停、快进、重绕、重载等)等从预期内容时间线(210)偏离或在预期内容时间线(210)周围波动。

[0126] 在制作阶段(202)中,媒体制作块(115)或其中的E&N元数据插入器212可与创意人员(例如,制作人员中的创意人员等)交互以获得由创意人员通过一或多个用户接口提供的用户输入。用户输入描述针对媒体内容(117-2)的一或多个数据部分的情绪预期及叙事信息(例如,关键点等)。

[0127] 另外、任选地或替代地,在制作阶段(202)中,媒体制作块(115)评估含有针对媒体内容(117-2)中的数据部分的叙事信息(例如,数字化故事信息等)及情绪预期的故事板206。故事板(206)提供表示于媒体内容(117-2)中的一或多个媒体程序的相对高级映射图或描述。当可用时,故事板(206)可由媒体制作块(115)处理以提取与媒体内容(117-2)相关的叙事信息、情绪预期、主要人物、所关注区域、故事情节连贯性等。

[0128] 至少部分基于从用户输入接收及/或从故事板(206)提取的叙事信息及情绪预期,媒体制作块(115)产生针对媒体内容(117-2)的媒体元数据(例如图1A的117-1等)的一或多个元数据部分。媒体元数据(117-1)中的一或多个元数据部分可包括描述媒体内容(117-2)的多个数据部分中的一或多个数据部分的情绪预期(例如,预期唤醒水平、预期效价水平、将预期的其它情绪维度等)及叙事关键点(例如,预期认知状态、预期参与度或注意力水平等)的一或多个E&N元数据部分。另外、任选地或替代地,媒体元数据(117-1)中的一或多个E&N元数据部分可包括将媒体内容(117-2)的数据部分中的声音或视觉元素指示或识别为AOI或ROI的零个或更多内容修改元数据部分。另外、任选地或替代地,媒体元数据(117-1)可包括相对低级音频元数据、图像元数据等等。

[0129] 可贯穿内容时间线(210)针对媒体内容(117-2)的数据部分重复地执行本文中所描述的元数据产生。媒体制作块(115)或其中的元数据汇集器214可通过元数据到内容联编操作208(例如,由图1A的编码块120等执行)将媒体内容(117-1)的各个元数据部分与媒体内容(117-2)的对应数据部分汇集、格式化及联编/多路复用于经编码位流(例如图1A的122等)中。

[0130] 消费阶段(204)实施或包含(但不一定仅限于)两个(例如,主要、关键等)时间相依过程。两个时间相依过程中的步骤在媒体内容(117-2)正被调整且通过回放装置向单独观看者或具有多个观看者的观众呈现时由回放装置(或连同其一起操作的装置)执行。

[0131] 消费阶段(204)中的两个时间相依过程中的第一者包含沿着在制作阶段(202)中所指定且由回放装置进一步实施的内容时间线210的观看者(上文提及的单独观看者或上文提及的观众中的观看者)或多个观看者观众的生理监测216(例如,通过经汇总观众响应等)。观看者的生理监测(216)理想地在时间上是连续的但可取决于在呈现环境中与回放装置一起操作的生理监测组件精细或粗糙地进行取样。

[0132] 回放装置或其中的E&N状态估计器218处理来自观看者的生理监测(216)的生理监测信号且使用所述生理监测信号估计或评估观看者关于媒体内容(117-2)的已经呈现的数据部分的E&N状态。在实例中,观看者的经评估E&N状态可表示通过例如唤醒、效价、优势情绪等的一或多个情绪维度描述的经评估情绪。在另一实例中,观看者的经评估E&N状态可表示指示媒体内容(117-2)的已经呈现的数据部分中的有效叙事信息(例如关键点等)如何被观看者传达或理解的经评估认知状态。

[0133] 消费阶段(204)中的两个时间相依过程中的第二者包含沿着与在制作阶段(202)中指定且由回放装置进一步实施的相同的内容时间线(210)的媒体内容(117-2)的内容回放222及修改224。

[0134] 作为内容回放(222)的一部分,回放装置执行元数据提取操作226(例如,作为图1A的解码/多路分用块130的一部分等)以提取与来自经编码位流(122)的媒体内容(117-2)的对应数据部分一起联编的媒体元数据(117-1)的各个元数据部分的一些或全部。

[0135] 针对媒体内容(117-2)的数据部分将在其处向观看者呈现的特定时间点,回放装置的E&N差异计算器230接收从E&N状态估计器(218)所估计的观看者的经评估E&N状态。E&N差异计算器(230)还评估或接收使用经编码位流(122)编码的媒体元数据(117-1)中的对应于媒体内容(117-2)的数据部分的E&N元数据部分且使用所述E&N元数据部分确定观看者在同一时间点的预期E&N状态。

[0136] E&N差异计算器(230)确定观看者的预期E&N状态与观看者的经评估E&N状态之间的差异。举例来说,如果观看者的预期E&N状态及观看者的经评估E&N状态从属于观看者的情绪状态,那么E&N差异计算器(230)确定由观看者的预期E&N状态所指示的观看者的预期情绪状态与由观看者的经评估E&N状态所指示的观看者的预期情绪状态之间的差异。另一方面,如果观看者的预期E&N状态及观看者的经评估E&N状态从属于观看者的认知状态,那么E&N差异计算器(230)确定由观看者的预期E&N状态所指示的观看者的预期认知状态与由观看者的经评估E&N状态所指示的观看者的预期认知状态之间的差异。

[0137] 接着,观看者的预期E&N状态与观看者的经评估E&N状态之间的差异可提供为到E&N内容修改模型228的输入且用于从E&N内容修改模型(228)产生呈对媒体内容的数据部分

的内容修改224的形式的输出以在给定时间点向观看者呈现。如果差异未超过E&N状态差异阈值(例如,效价差异阈值、唤醒差异阈值、注意力水平差异阈值等),那么内容修改(224)可为零(或空)修改。如果差异超过E&N状态差异阈值,那么内容修改(224)可为非零(或非空)修改。内容修改(224)的量级及/或类型可定性或定量地取决于预期E&N状态与经评估E&N状态之间的差异。

[0138] 可针对将在制作阶段(202)中所指定的内容时间线(210)的其它时间点处呈现且在消费阶段(204)中实施的媒体内容(117-2)的其它数据部分中的每一者重复前述步骤(或操作)。

[0139] 用于最小化经评估状态与预期状态之间的差异的媒体内容的已经呈现的部分的无效性可通过从E&N元数据(220)所确定或提取的观看者的预期E&N状态与通过生理监测(216)所确定或估计的观看者的经评估E&N状态之间的相对大不一致(例如,超过唤醒差异阈值的唤醒差异、超过效价差异阈值的效价差异、通过面部表情分析检测的偏离情绪类型等)来指示或测量。预期E&N状态可由本文中所描述的系统用作用于最小化差异的基于反馈的控制处理的情绪及叙事目标。

[0140] E&N内容修改模型(228)可用于基于观看者的预期E&N状态与观看者的经评估E&N状态之间的差异产生输出。经产生输出可包括将沿着内容时间线(210)呈现的数据部分的媒体内容修改(或对驱动音频或图像呈现操作的信号的修改)。媒体内容修改明确地经实施以降低关于描述或体现于E&N元数据(220)中的创作意图(例如情绪预期、叙事状态、注意点等)所测量的媒体内容(117-2)的已经呈现的数据部分的任何检测到的无效性。

[0141] 图2B说明实例媒体内容及元数据制作阶段(例如图2A的202等)。如展示,故事板(例如图2A或图2B的206等)含有与沿着内容时间线(例如图2A或图2B的210等)的多个个别时间点或间隔对准的多个个别故事板页面206-1到206-7。E&N元数据插入器(例如图2A或图2B的212等)与媒体内容(例如图1A的117-2等)的创意人员交互以接收描述媒体内容(117-2)的一或多个数据部分的情绪预期及叙事信息(例如关键点等)的用户输入。如图2B中展示,用户输入中所描述的情绪预期及叙事信息包括多个内容时间线编辑(例如,其中一者可为图2B的234,以指示场景的开始或结束等)、描绘于媒体内容(117-2)中的故事中的多个关键时刻212-1到212-6及弧线232-1到232-3(例如,场景中的关键时刻、观看者的情绪预期、将向观看者传达的叙事关键点等)等等。

[0142] 如说明,每一个别内容时间线编辑对应于沿着内容时间线(210)的相应时间点或间隔。同样地,每一关键时刻及弧线对应于沿着内容时间线(210)的相应时间点或间隔。

[0143] 在一些操作案例中,可存在比故事板页面更多的编辑。此外,编辑及故事板页面可或不沿着内容时间线(210)对准。另外、任选地或替代地,媒体元数据(例如图1的117-1等)中的媒体元数据部分212-1到212-6(例如,仅等)可被插入到媒体内容(117-2)中或在编辑之间的关键场景期间通过元数据汇集器(例如图2A或图2B的214等)与媒体内容(117-2)一起联编于内容回放文件、视频信号、经编码位流(例如图1A的122等)及类似物中。

[0144] 图2C说明实例媒体内容及元数据消费阶段(例如图2B的204等)。在此阶段中,用于生理监测216-1的信号可在媒体内容(117-2)正被调整且通过回放装置236向观看者呈现时由组件(例如,以面部表情分析软件等一起操作的电极、传感器、相机)产生。在一些操作案例中,一些或全部生理监测组件经配置为或经包含作为回放装置(236)的一部分。在一些其

它操作案例中,用于生理监测(216-1)的一些或全部组件单独或与回放装置(236)分开地配置或部署且连同回放装置(236)一起操作。生理监测组件将生理监测信号提供或传输到回放装置(236)以进行关于观看者的生理状态评估。

[0145] 如图2C中说明,E状态估计器218-1到N状态估计器218-2分别实施情绪状态估计模型及认知状态(或认知负荷)估计模型、使用生理监测信号作为到估计模型的输入、且在媒体内容(117-2)正被调整且通过回放装置(236)向观看者呈现时将接收到的生理监测信号转换成作为来自估计模型的输出的观看者的(经评估)情绪状态及认知状态。

[0146] 同时,回放装置(236)接收或继续接收媒体元数据(117-1)(的将呈现的部分)与媒体内容(117-2)。媒体元数据(117-1)中的E&N元数据220可由回放装置(236)用于在内容回放222-1中的各个时间点获得观看者的预期情绪状态及认知状态。

[0147] 从情绪状态估计模型及/或认知状态估计模型输出的观看者的(经评估)情绪状态及认知状态与指定于媒体元数据(117-1)中的预期情绪状态及认知状态一起作用反馈,以帮助执行实时内容回放及修改操作244。一些或全部这些内容回放及修改操作(244)可由回放装置(236)实施为时间相依过程。

[0148] 在一些操作案例中,为了执行内容回放及修改操作(244),回放装置(236)实施情绪状态内容修改模型228-1及认知状态内容修改模型228-2、使用观看者的经评估及预期情绪状态及认知状态作为到内容修改模型(228-1及228-2)的输入、产生预期状态与经评估状态之间的差异、及在媒体内容(117-2)正被修改且通过回放装置(236)向观看者呈现时使用所述差异(或根据创作意图的预期状态与实际状态之间的差异)产生作为来自内容修改模型的输出的相对高级修改信号。

[0149] 从内容修改模型(228-1及228-2)输出的高级修改信号可至少部分基于例如载送用于音频或图像处理操作的操作参数的相对低级信号域元数据的媒体元数据(117-1)的非E&N元数据242经转换成所选择的内容修改信号224-1到224-5。

[0150] 所选择的内容修改信号(224-1到224-5)在内容回放(222-1)的不同时间点作用于媒体内容(117-2)且在内容回放(222-1)期间出于最小化创作意图与观看者的经评估状态之间的差异的目的致使对媒体内容(117-2)作出特定内容修改。对媒体内容(117-2)的特定内容修改可为涉及在非E&N元数据(242)中识别的用以如由创意人员期望地那样致使观看者的生理状态朝向体验预期情绪移动或理解由媒体内容(117-2)描绘的故事中的关键点的一些或全部AOI或ROI的媒体内容调整或修改。本文中所描述的信号修改(224-1到224-5中的任一者)可通常保持恒定、相对平滑地改变、或可在适用时间间隔内(例如,在创意人员的编辑之间等)改变。

[0151] 另外、任选地或替代地,回放装置特性化数据238及/或周围环境特性化数据240可用于回放装置(236)的内容回放及修改操作(244)中。可使回放装置特性化数据(238)及/或周围环境特性化数据(240)可在回放装置(236)处存取或本地存储(例如,配置数据或文件、能力数据或文件、静态元数据、可配置元数据等)。回放装置特性化数据(238)涉及或描述包含(但不限于)以下各者中的一或多者的回放装置(236)的音频及视频处理能力及/或限制:回放装置的类型(例如,小型、基于家庭的、基于电影院的等)、(亮度)动态范围、色域、与回放装置一起操作的图像显示器的空间分辨率、用于音频呈现/转导的扬声器所支持的媒体信号的位深度、数目、配置、频率范围、及/或频率/相位失真、位置呈现能力等。周围环境特

性化数据(240)涉及或描述其中回放装置(236)在其中进行操作的呈现环境的特性,包含(但不限于)以下各者中的一或多者:物理大小、呈现环境的几何结构及/或特性、环境声音、环境照明、白噪声级、虚拟环境中的杂波特性等。

[0152] 8. 情绪预期及叙事元数据格式

[0153] 图2D说明在内容及元数据制作阶段(例如202等)中基于创意输入246产生且在内容及元数据消费阶段(例如204等)中在内容回放(例如222等)时使用的实例E&N元数据。

[0154] 在制作阶段(202)中,包括E状态元数据部分248及N状态元数据部分250的各个E&N元数据部分可基于创意输入(246)在沿着预期视听内容时间线(例如210等)的多个时间间隔内的多个时间点产生。E状态元数据部分(248)及叙事元数据部分(250)可或不沿着内容时间线(210)对准。E状态元数据部分(248)及叙事元数据部分(250)的特定元数据部分的起始及结束位置可例如通过创意输入(246)中所提供的内容时间线编辑(例如234等)设置、配置或特定。

[0155] 在消费阶段(204)的内容回放(222)中,包括E状态元数据部分(248)及叙事元数据部分(250)的一些或全部E&N元数据部分可经提取且与生理监测及评估一起用于在必要时沿着回放时间线-例如,由回放装置在内容回放(222-以传达在制作阶段(202)中E&N元数据针对其产生的对应媒体内容(例如图1A的117-2等)的创作意图)所实施的内容时间线产生媒体内容调整或修改。

[0156] 如图2D中展示,E&N元数据分解成分别包括针对情绪的数字字段或容器及针对叙事信息的(例如,单独等)数据字段或容器的E状态元数据部分(248)及叙事元数据部分(250)。针对E状态元数据部分(248)中的情绪的数据字段或容器被细分成预期状态(例如,预期情绪及量级等)及预期修改(例如,校正信号修改等)。同样地,针对叙事元数据部分(250)中的叙事信息的数据字段或容器被细分成预期状态(例如,叙事ROI、AOI、混乱指数等)及期望修改(例如,校正信号修改等)。

[0157] 叙事元数据部分(250)可在范围从例如语义水平的相对高级到例如特定所关注图像区域(每帧跟踪或跨场景跟踪)、所关注音频对象、混乱指数等等的相对低级的多种不同抽象级中的一者下指定。

[0158] 混乱指数预期被稀疏地使用但当对应(例如,重要、关键、主要等)故事情节信息将被观看者(例如,充分、完全、良好等)理解时作为元数据被插入。混乱指数可经设置以区分例如混乱动作场景的预期混乱与(例如,重要、关键、主要等)故事情节信息的非所要混乱。当需要时,针对内容时间线(210)的给定时间点或给定时间间隔存在混乱指数,且与元数据相关联(例如,识别为元数据中的所关注对象)的音频或视觉对象无需继续存在(例如,如果其未被使用等)。

[0159] 在一些操作案例中,E状态或N状态元数据状态可在编辑结(例如,前一场景、前一媒体内容部分等)处插入且跨例如视频或音频帧的媒体内容部分继续存在直到下一编辑结(例如,在下一场景、下一媒体内容部分等之前)。在一些操作案例中,使旗标可用于(例如,在经编码位流中、在元数据部分中、在音频或视频帧的标头中等)出于避免引发承载每帧的重复元数据的附加位的目的发信号通知回放装置继续使用在先前接收到的元数据部分中所指定的信息。在场景的开头或中间插入的旗标或元数据部分可继续存在到下一场景。旗标或元数据部分可在帧级、场景级、场景细分级、序列级等下插入。举例来说,在一些操作案

例中,如果需要的话,划定不同媒体部分及/或不同元数据部分的编辑结可处于帧分辨率下。另外、任选地或替代地,旗标或数据字段的不同值之间的斜坡或转变周期可经实施于媒体内容中。另外、任选地或替代地,校正信号修改选项可经包含为本文中所描述的一或多个E状态或N状态元数据部分的一部分。

#### [0160] 9. 信号分离与融合

[0161] 图2E说明针对具有单独观看者的观众的实例生理监测及评估。生理监测组件或传感器可经配置有或放在各个位置处,例如(例如,手持式等)图像显示器、耳塞装置、智能手表装置、TV等。这些位置中的每一者供应或提供特定感测阵列。在一些实施例中,回放装置(例如图2C的236等)包括图像显示器及音频或声音源,其可与耳塞装置无线连接。智能手表可为或可不为回放装置的一部分且可被视为或配置为与回放装置一起操作的辅助组件。

[0162] 如图2E中展示,针对单独观看者的生理监测传感器可包含以下各者中的一或多个者:基于显示器的传感器,例如可见光波长相机传感器、结构化光或SLAM(同时定位及映射)传感器、热成像器、HMD传感器等;入耳式传感器、手腕传感器;等等。

[0163] 可见光波长相机传感器可用于监测观看者的注视位置、瞳孔直径、面部表情等。结构化光或SLAM传感器可用于监测观看者的头部位置、观看距离、面部表情等。热成像器可用于监测观看者的效价、唤醒、面部表情等。HMD传感器可用于产生关于观看者的基于EEG的生理监测信号。例如电极、热传感器、光学传感器等的入耳式传感器可用于产生关于观看者的基于EOG(例如,出于注视位置监测目的等)、基于EEG、基于呼吸及/或基于体积描记-HR的生理监测信号。手腕传感器可用于产生关于观看者的基于HR及/或基于GSR的生理监测信号。

[0164] 图2E中所展示的(五角形)传感器融合与分离块可用于处理来自一些或全部生理传感器的生理监测信号。传感器融合与分离块可使用用于将接收到的生理监测信号转换到情绪状态及认知状态的一或多个模型(例如,算法、方法、程序、操作等)实施。

[0165] 传感器融合与分离块使接收到的生理监测信号分离成不同群组的生理监测信号。这些不同群组的生理监测信号可用于评估不同类型的状态。举例来说,如图2G中说明,第一群组的生理监测信号可用于估计或评估一或多个E状态,第二群组的生理监测信号可用于估计或评估一或多个N状态,第三群组的生理监测信号可用于估计或评估观看者的注意点,第四群组的生理监测信号可用于估计或评估观看者的一些或全部前述生理上可观察方面,等等。

[0166] 传感器融合与分离块将类似或重复生理监测信号(在接收到的生理监测信号中)组合或汇集成整体生理监测信号。若干整体生理监测信号可由传感器融合与分离块从全部接收到的生理监测信号产生(generate/produce)。

[0167] 在实例中,使用多种不同类型的生理监测技术、组件或传感器产生的信号可经组合或汇集整体生理监测信号以进行面部表情分析。在另一实例中,使用多种不同类型的生理监测技术、组件或传感器产生的信号可经组合或汇集成整体生理监测信号以进行心率测量或确定。

[0168] 实施于先前所提及的传感器融合与分离块中的状态估计模型可包含用于确定创意人员认为是重要的有效叙事信息如何向观看者转移或传达的认知状态估计模型(或叙事转移估计模型)。将如本文中所描述那样从媒体内容转移到观看者的叙事信息可包含(但不限于)以下各者中的一或多个者:所描绘场景中的信息(例如,犯罪场景中遗留的鞋子等)、人

物之间的对话、所关注图像区域、所关注音频或声学对象等。出于理解媒体内容中描绘的故事情节的目的有效地转移到观看者的叙事转移-或叙事信息-可使用参与度、注意点、眼睛注视、伴随的情绪响应等测量。在一些操作案例中,观看者的认知状态包括叙事转移评估的两种单独的关键元素,其是观看者的认知负荷及观看者的注意点(观看者正在关注的内容)。

[0169] 注意力可被视为认知的子集。在一些操作案例中,基于注意力的生理监测及内容调整过程经压缩到或经实施作为基于认知的过程或处理块的一部分;因此,注意力经包含为创作意图中的叙事状态的一部分。在一些操作案例中,基于注意力的生理监测及内容调整过程至少部分与基于认知的过程或处理块分离;因此,注意力可为创作意图中除了情绪及叙事状态之外的单独方面。

[0170] 在一些呈现环境中,观看者的注意点或注意位置可使用通过由与例如ROI的视觉对象的呈现相关或同步的这些传感器产生的视线跟踪或瞳孔方向监测信号监测观看者对视觉对象的注意力的传感器来确定。在一些操作案例中,观看者可能正在关注观看者的近中心凹外部的图像区域或运动活动;因此,观看者的注视可能不与注意点重合。

[0171] 另外、任选地或替代地,观看者的注意点还可通过非注视跟踪监测传感器监测。例如,观看者的注意点或注意位置可使用通过由与观看者的近中心凹外部的音频对象、图像对象的呈现等相关或同步的这些传感器产生的EOG及/或EEG监测信号监测观看者对音频对象的注意力、观看者的近中心凹外部的移动对象等的脑电活动监测传感器确定。

[0172] 在一些实施例中,在给定时间点估计的观看者的认知状态包含在给定时间点的观看者的部分的认知负荷及点或区域-例如,媒体内容(117-2)的图像呈现的图像区域、媒体内容(117-2)的音频呈现的声场中的声场区域-观看者正在关注所述区域。

[0173] 因此,传感器融合与分离块可在媒体内容(例如图1A的117-2等)正基于用使用估计模型的信号分离及汇集操作所处理的接收到的生理监测信号被调整且通过回放装置(例如图2C的236等)向单独观看者呈现时在各个时间点产生观看者的情绪状态、认知状态(或叙事转移状态)等。

[0174] 图2F说明针对一群观众的实例生理监测及评估。一群观众可包含(但不一定仅限于)剧院中的一大群观众或家里的房间或空间中的一小群-或多个观看者。生理监测组件或传感器可经配置有或放置在整体房间或场地中的各个位置处,例如座椅、TV等,以共同及/或同时监测观众中的一些或全部观看者。

[0175] 如图2F中展示,针对一群观众的生理监测传感器可包含基于房间的传感器,例如可见光波长相机传感器、热成像器、气体传感器等;基于座椅的传感器;等等。可见光波长相机传感器及/或热成像器可经安置于面向观众的位置中且用于定位一群观众成员的面部、监测一群观众的面部表情、及产生包含(但不限于)一群观众的整体E&N状态的面部表情群组统计。气体传感器可用于监测呈现环境中的CO<sub>2</sub>(例如,以确定由CO<sub>2</sub>含量指示的唤醒等)及R3COH(例如,以确定观看者是否可能在醉酒状态下观看喜剧且使对话对此观看者清晰等)气体水平以监测一群观众的呼吸及陶醉水平(如果存在,其可影响认知以及情绪)。基于座椅的传感器可与一群个别观众成员就坐在其上的个别座椅一起安置且用于产生关于一群观众的基于呼吸及/或基于体积描记-HR的生理监测信号。

[0176] 类似于先前在图2E中展示的内容,在一群观众的情况中,图2F中所展示的(五角

形)传感器融合与分离块用于处理来自一些或全部生理监测传感器的生理监测信号。传感器融合与分离块可实施或使用用于将接收到的生理监测信号转换到情绪状态及认知状态的一或多个模型(例如,算法、方法、程序、操作等)。传感器融合与分离块将接收到的生理监测信号分离成不同群组的生理监测信号。这些不同群组的生理监测信号可用于评估不同类型的状态。传感器融合与分离块将类似或重复生理监测信号(在接收到的生理监测信号中)组合或汇集成整体生理监测信号(例如,在从所有接收到的信号产生的若干整体信号当中)。

[0177] 如在单个观看者观众的情况中,在一群观众的情况中实施于传感器融合与分离块中的状态估计模型也可包含用于确定创意人员认为是重要的有效叙事信息如何向一群观众转移或传达的认知状态估计模型(或叙事转移估计模型)。在一些实施例中,在给定时间点估计的一群观众的认知状态包含在给定时间点的一群观众的部分的认知负荷及点或区域-例如,媒体内容(117-2)的图像呈现的图像区域、媒体内容(117-2)的音频呈现的声场中的声场区域-一群观众正在关注所述区域。

[0178] 因此,传感器融合与分离块可在媒体内容(例如图1A的117-2等)正基于用使用估计模型的信号分离及汇集操作所处理的接收到的生理监测信号被调整且通过回放装置(例如图2C的236等)向一群观众呈现时在各个时间点产生一群观众的情绪状态、认知状态(或叙事转移状态)等。

[0179] 图2G进一步说明针对单独观看者观众的实例传感器融合与分离。应注意,一些或全部此描述类似地适用于或可容易地扩展以覆盖具有一或多个观看者的一群观众。

[0180] 来自图2E中所说明的不同传感器或组件的生理监测信号可用于估计或评估例如能效及唤醒的观看者的情绪状态以及指示叙事转移的持续成功的观看者的认知状态。

[0181] 来自回放装置的给定组件的传感器可促进生理监测信号用于评估一些或全部观看者的情绪状态、认知负荷及注意点。在给定状态估计下,不同传感器可存在重复,例如经由基于显示器的相机的眼睛注视位置以及来自耳塞的EOG信号。这些多个信号可如图2G中展示那样用实心及空心圆汇集。

[0182] 下文表2及3说明生理监测信号的实例列表,其是就其生理位置、传感器类型、及使用所述生理监测信号的估计器类型而言的。通过实例但无限制,表2含有图2E中针对单独观众所说明的传感器,而表3含有图2E中针对一群观众所说明的传感器。

[0183] 表2

[0184]

生理监测信号	位置	传感器	估计器
注视位置	显示器(例如电话、TV、平板计算机)	可见光波长相机	注意点
瞳孔直径	显示器	可见光波长相机	注意点 &认知负荷
面部表情	显示器	可见光波长相机	情绪状态
头部位置	显示器	结构化光或 SLAM	认知负荷 (&视觉阈值)
观看距离	显示器	结构化光或 SLAM	认知负荷 (&视觉阈值)
面部表情	显示器	结构化光或 SLAM	情绪状态
效价	显示器	热相机	情绪状态
唤醒	显示器	热相机	情绪状态
EEG	显示器	HMD 传感器	情绪状态 &认知负荷
EOG 注视位置	入耳式(例如智能耳塞)	耳塞偶极电极	注意点
EEG	入耳式	耳塞偶极电极	情绪状态

[0185]

			&认知负荷
呼吸	入耳式	耳塞麦克风或加速度计	情绪状态
心率(体积描记)	入耳式	耳塞麦克风、加速度计、无源红外 (PIR)	情绪状态
心率	手腕(例如智能手表)	PPG (光电传感器)	情绪状态
皮肤电响应	手腕	皮肤电导传感器	情绪状态-觉醒

[0186]

表3

[0187]

生理监测信号	位置	传感器	估计器
面部表情群组统计	房间	可见相机	情绪状态
面部表情群组统计	房间	热相机	情绪状态
CO2	房间	气体传感器	情绪状态

R3COH	房间	气体传感器	注意点
呼吸	座椅	呼吸传感器	情绪状态
心率	座椅	心率传感器	情绪状态

[0188] 关于可插入哪种E&N元数据以及哪种信号修改可包含于E&N元数据中可存在许多选项。在一些操作案例中,用于将经评估E&N状态收敛到预期E&N状态的一些或全部信号修改由创意人员例如在媒体内容及元数据制作阶段(例如图2A或图2B的202、2D元数据等)处确定。在一些操作案例中,一些或全部信号修改使用决定应作出哪些修改的信号修改方法/算法(或模型)确定。这些信号修改方法/算法(或模型)可视情况针对特定类型的E&N状态及/或针对特定类型的E&N状态的经评估状态与预期状态之间的任何差异的特定量级(范围)产生信号修改。

#### [0189] 10. 媒体内容调整或修改

[0190] 图3A到图3G说明实例操作案例中的情绪、认知、及注意元数据及对应信号修改的实例。应注意,这些是非限制性实例。在一些回放应用(例如,教育媒体内容、信息式媒体内容等)中,叙事或认知状态(例如,使用认知负荷所测量、使用注意力所测量、使用观看者参与的时间长度所测量等)对基于生理监测的生理监测及媒体内容调整或修改很关键或很重要。在一些回放应用(例如,游戏媒体内容、娱乐媒体内容等)中,情绪状态(例如,通过效价及唤醒所测量、使用面部表情所测量、使用情绪类型的离散分类的所测量等)对生理监测及基于生理监测的媒体内容调整或修改可能是相对重要的。在一些回放应用中,情绪及叙事状态两者可能对生理监测及基于生理监测的媒体内容调整或修改很重要。在一些回放应用中,情绪与叙事状态的其它组合可能对生理监测及基于生理监测的媒体内容调整或修改很重要。

[0191] 图3A说明由回放装置(例如图2C的236等)针对情绪状态及对应元数据进行的实例媒体呈现处理。

[0192] 媒体元数据(例如图1A的117-1等)中的E&N元数据(例如图2A的220等)与媒体内容(例如图1A的117-2等)一起被回放装置(236)接收。在此实例中,E&N元数据(220)包括至少情绪预期元数据或E状态元数据。E状态元数据包括沿着与媒体内容(172-2)的内容回放222一起实施的内容时间线的给定时间点(例如,给定场景、给定时间间隔等)的预期情绪状态及当针对给定时间点所估计或预测的观看者的经评估情绪状态与在E&N元数据(220)中所指示的给定时间点的预期状态有差异时可由回放装置(236)应用的信号修改选项。

[0193] 生理监测信号可在媒体内容(172)正被调整且向观看者呈现时(例如,实时、近实时、在严格延时预算内等)产生。通过实例但无限制,一个生理监测信号可使用具有面部表情分析软件的相机产生,而另一生理监测信号可使用EEG电极产生。这两个生理监测信号经处理以提供面部情绪估计及基于EEG的情绪估计,其可由传感器融合与分离块(例如装置、模块等)汇集成单个情绪状态估计(在图3A中标示为“经估计状态”)。情绪状态估计(或经评估情绪状态)与来自E状态元数据的预期状态(作为创作意图的部分在元数据中指定)进行比较以产生情绪状态差异。接着,将此差异馈送到情绪状态内容修改模型228-1中以基于情绪状态差异与来自E状态元数据的可能信号修改选项一起产生或识别信号修改224-1。

[0194] 在一些操作案例中,内容修改模型(228-1)用于例如基于状态差异的量级等确定信号修改(224-1)的特定操作参数的量级或值。到内容修改模型(228-1)的其它输入可包含

媒体元数据(117-1)中的叙事元数据,例如所关注图像区域(ROI)及所关注音频对象(AOI)。从这些输入,确定信号修改(224-1)的操作参数,且接着将其用于在给定时间点将媒体内容部分修改到将在给定时间点回放(通过图像或音频处理、或两者)的实际媒体内容部分。

[0195] 图3B说明图3A中所展示的媒体呈现处理的特定实例。正被调整且通过回放装置向观看者呈现的媒体内容(117-2)可为具有关键场景的电影,其中中心人物可正在讲述一件事情,但人物的面部表情掩饰不同情绪。

[0196] 观看者正在使用智能耳塞收听场景的音频且正在观看正被调整且在被保持在使得观看者的视场(FOV)较小的距观看者的某一距离处的例如平板计算机的移动显示器上呈现的场景。因此,由于感知分辨率限制(例如,超过视觉截止频率的像素奈奎斯特频率等)而无法看见人物的微妙面部表情。

[0197] 如E状态元数据中指定的预期情绪状态指示观看者的预期情绪是“强烈的悲伤”。如果观看者的预期情绪状态与经评估情绪状态的不同超过量级阈值,那么如在E状态元数据中指定的信号修改选项指示特定所关注区域(ROI)的放大或缩小是建议的信号修改选项。

[0198] 平板计算机上的显示器相机可用于获取观看者的面部的图像用于面部表情分析。与智能耳塞一起部署的电极可定位在与观看者的头部接触的不同位置处且用于从观看者获取EEG信号以进行基于EEG的情绪估计。

[0199] 在本实例中,从生理监测信号估计的情绪状态是冲突的。基于显示器相机的面部表情估计指示观看者处于“平静”情绪状态,而基于EEG的情绪估计指示观看者处于“兴趣”情绪状态。本文中所描述的回放装置汇集这两个情绪状态估计以沿着比由创意人员所期望的预期情绪状态更小(例如,在唤醒、效价等方面)的所关注中性情绪向量输出整体信号分级。接着,情绪状态差异可经导出及提供为到内容修改模型(228-1)的输入。

[0200] 叙事元数据具有关于图像ROI的信息,其是人物的面部的像素位置或图像区域,而如在E状态元数据中指定的特定情绪状态差异的信号修改选项包含图像ROI。另外、任选地或替代地,叙事元数据可具有关于所关注音频对象(AOI)的相对排名的信息,其与图像ROI相关。仅出于说明的目的,可不使用关于音频对象的相对排名的信息。

[0201] 信号修改的内容修改模型(228-1)采用情绪状态差异的量级、叙事元数据中的ROI信息、及/或放大如在E状态元数据中指定的ROI的信号修改选项,以确定观看者的(将评估的)情绪状态可能受到影响或根据创作意图通过放大人物的面部从“中性兴趣”增加到“强烈的悲伤”。接着,从内容修改模型(228-1)输出的此信息可在内容回放(222)中用于实施特定信号修改(224-1),例如通过放大居中在 $I(x,y)$ 处的(人物的面部的)像素位置。

[0202] 图3C说明由回放装置(例如图2C的236等)进行的与认知状态及对应元数据(或其中的变化)相关的实例媒体呈现处理。

[0203] 媒体元数据(例如图1A的117-1等)中的E&N元数据(例如图2A的220等)与媒体内容(例如图1A的117-2等)一起被回放装置(236)接收。E&N元数据(220)包括至少部分与沿着与媒体内容(172-2)的内容回放222一起实施的内容时间线的给定时间点(例如,给定场景、给定时间间隔等)的混乱指数一起的预期认知状态的叙事元数据(或N状态元数据)及当针对给定时间点所估计或预测的观看者的经评估叙事状态在给定时间点与在E&N元数据(220)中指定的预期叙事状态有差异时可由回放装置(236)应用的信号修改选项。

[0204] 一般来说,叙事元数据还可包含叙事ROI及AOI。然而,在此实例中,仅出于说明目的,叙事元数据中的ROI及AOI不用于信号修改。

[0205] 生理监测信号可在媒体内容(172)正被调整且向观看者呈现时(例如,实时、近实时、在严格延时预算内等)产生。一个生理监测信号可使用显示器位置的相机(例如,定位在回放装置或平板计算机的相同的面向观看者的表面上)中的(例如,硬件及/或软件等实施的)眼睛跟踪器产生,而另一生理监测信号可使用EEG电极产生。这两个生理监测信号由回放装置处理以提供或产生基于瞳孔直径的认知状态估计及基于EEG的认知状态估计。这两个认知状态估计可进一步由传感器融合与分离块(例如,装置、模块等)汇集成单个认知状态估计(在图3C中标示为“经估计状态”)。经估计或经评估认知状态与来自叙事元数据的预期认知状态(作为创作意图的部分指定为元数据中的混乱指数)进行比较以产生认知状态差异。接着,将此差异馈送回到认知(或叙事)状态内容修改模型228-2以基于认知状态差异连同来自叙事元数据的可能信号修改选项一起产生或识别信号修改224-2。

[0206] 内容修改模型(228-2)可由回放装置用于例如基于状态差异的量级等确定信号修改(224-2)的特定操作参数的量级或值。到内容修改模型(228-2)的其它输入可包含媒体元数据(117-1)中的情绪元数据。在一些操作案例中,情绪元数据中的信息可被认为是内容修改模型(228-2)中的次级或次要促成物。从一些或全部这些输入,确定信号修改(224-2)的操作参数,且接着将其用于在给定时间点将媒体内容部分修改成将在给定时间点回放(通过图像或音频处理、或两者)的实际媒体内容部分。

[0207] 图3D说明图3C中所展示的媒体呈现处理的特定实例。正被调整且向观看者呈现的媒体内容(117-2)可为具有一场景的电影,其中人物解释在对话中解释对所描绘故事很关键的某事。然而,场景具有大量辅助声音。观看者正使用智能耳塞在平板计算机观看,但呈现环境足够嘈杂使得耳塞未充分地阻挡外部声音。随后,对话中的对故事情节很关键的部分缺失。

[0208] 混乱指数在制作阶段(202)中在媒体元数据(117-1)中被设置到零,这是由于所述场景是创意人员希望观看者完全理解的重要对话场景。

[0209] 如在叙事元数据中指定的预期认知状态指示观看者的预期混乱指数由创意人员设置到零。创意人员期望或希望观看者完全理解场景或对话。应注意,在许多情况中,观看者的预期混乱指数默认为零。然而,可存在其中观看者的预期混乱指数是针对比零更高的值设置的特定场景,例如在复杂性应极高的场景(例如,动作场景、具有许多争论声音的政治戏剧等)中。

[0210] 如果观看者的预期认知状态与经评估认知状态的不同超过差异量级阈值,例如当通过生理监测评估的观看者的混乱指数相对于预指定混乱指数零较高时,那么如在叙事元数据中指定的信号修改选项进一步指示增加说话声音的音量是建议的信号修改选项。

[0211] 平板计算机上的显示器相机可用于获取用于基于瞳孔直径的认知负荷估计的观看者的面部图像。与智能耳塞一起部署的电极可定位在与观看者的头部接触的不同位置处且用于从观看者获取EEG信号以进行基于EEG的认知负荷估计。

[0212] 从生理监测信号估计的认知负荷可经汇集以输出指示观看者的混乱指数高于由创意人员所期望的认知状态中的预期混乱指数的整体认知负荷。接着,认知状态差异(例如,预期混乱指数与经评估混乱指数之间的差异等)可经导出及提供为到内容修改模型

(228-2)的输入。

[0213] 信号修改的内容修改模型(228-2)采用如从生理监测产生的认知状态差异的量级及/或如在叙事元数据中指定的信号修改选项,且产生或选择指示出于减小观看者的经评估混乱指数的目的调制对话音量相对于原声的其它音频对象的会话音量的增加的信号修改选项。从内容修改模型(228-2)输出的信号修改选项用于在内容回放(222)中实施特定信号修改(224-2),例如改变对话音频对象的音量与对应于Foley声音及背景音乐的非音频对象的音量之比。

[0214] 在本实例中,存在具有经设置以指示预期的预期情绪的旗标的情绪预期元数据(或E状态元数据)、及如果观看者的预期情绪状态与经评估情绪状态之间的差异大于情绪状态差异阈值那么就增加图像对比度的补偿步骤(或信号修改选项)。然而,由于生理监测信号指示观看者可能不理解场景的事实,所以在此实例中不使用预期情绪状态及如在E状态元数据中指定的信号修改选项。因此,情绪状态差异的信号修改选项不影响用于改进观看者的认知状态或增加观看者对场景的理解的所得信号修改。

[0215] 图1C说明音频处理的实例配置,其中音频编码器在制作阶段(例如图2A的202等)中作为媒体编码块(例如图1A的120等)的一部分操作且音频回放块作为包括媒体解码及呈现块(例如图1A的130及135等)的媒体回放装置的一部分操作。

[0216] 经解码媒体元数据(132-1)可与经解码媒体内容(132-2)一起由回放装置、或连同回放装置一起操作的音频及/或图像呈现装置135使用,以按保护发行版本已经使用其产生的创作者的意图、或最小化或避免对所述意图的扭曲的方式执行生理监测、生理状态评估、媒体内容调整或修改、音频处理、视频处理、音频再现/转导、图像呈现/再现等等。

[0217] 作为内容回放(222)的一部分,回放装置执行元数据提取操作226(例如,作为图1A的解码/多路分用块130的一部分等)以提取与来自经编码位流(122)的媒体内容(117-2)的对应数据部分一起联编的媒体元数据(117-1)的各个元数据部分的一些或全部。

[0218] 针对媒体内容(117-2)的数据部分将向观看者呈现的特定时间点,回放装置的E&N差异计算器230接收从E&N状态估计器(218)所估计的观看者的经评估E&N状态。E&N差异计算器(230)还接收使用经编码位流(122)编码的媒体元数据(117-1)中的对应于媒体内容(117-2)的数据部分的E&N元数据部分且使用所述E&N元数据部分确定观看者在同一时间点的预期E&N状态。

[0219] 音频编码器包括对话增强(DE)分析块、音频编码块等。如图1C中说明,音频编码器接收多个输入声道及对话输入。在此处,对话输入表示纯对话。另外、任选地或替代地,一些或全部输入声道包括非对话音频内容(例如,音乐、风噪声、源于非人类对象、背景、环境的声音等)、除了对话输入之外的混合对话或语音内容元素等。

[0220] DE分析块使用对话输入及含有混合对话/语音内容元素的输入声道产生用于对话增强的操作参数(标示为“DE参数”)。实例DE参数可包含(但不一定仅限于)使用应用到对话输入及含有混合对话/语音内容元素的输入声道的最小均方误差(MMSE)优化算法产生或预测的参数。对话输入、多个输入声道、DE参数、配置参数(例如,针对对话增强的最大电平移位或增益等)、重构参数等可在音频编码器中经处理(例如,下混、上混、空间化、动态范围控制等)及编码到整体经编码位流中的音频位流(例如,AC-4位流等)的一或多个经编码通道中。

[0221] 在消费阶段中,音频回放装置接收包括经编码声道与对话内容的音频位流,且将(通过音频解码块)接收到的音频位流解码成DE参数、配置参数(例如,针对对话增强的最大电平移位或增益等)、重构参数等。响应于接收与从在经编码位流中的媒体内容正被调整且向观看者呈现时执行的生理监测(例如,通过注视跟踪等)产生的认知负荷评估(例如经评估认知状态或注意点等)相关的(实时)信号修改(例如图3C或图3D的224-2等)及针对对话增强的信号修改选项(混响减小等),音频回放可实施信号修改(224-2)且产生(通过DE块)具有增强对话(例如,增强对话音量或引发的对话标准化、减小的混响、表示对话内容的相对精确的音频对象位置、增加的信噪比等)的一或多个输出音频通道。

[0222] 图3E说明由回放装置(例如图2C的236等)进行的与使用注意点(或观看者的注意力)所评估的叙事状态及对应元数据(或其中的变化)相关的实例媒体呈现处理。

[0223] 在此特定实例中,媒体元数据(例如图1A的117-1等)中的E&N元数据(例如图2A的220等)与媒体内容(例如图1A的117-2等)一起被回放装置(236)接收。E&N元数据(220)包括叙事元数据(或N状态元数据),但不包括情绪预期元数据(或E状态元数据),如由创意人员指定。叙事元数据包括由观看者对其的注意力沿着与媒体内容(172-2)的内容回放222一起实施的内容时间线的给定时间点(例如,给定场景、给定时间间隔等)监测/评估的特定图像ROI及AOI的一或多个预期注意点表示的预期叙事状态及当针对给定时间点所估计或预测的观看者的经评估状态(例如,经估计状态、经预测状态等)在给定时间点与预期状态有差异时可由回放装置(236)应用的信号修改选项。

[0224] 生理监测信号可在媒体内容(172)正被调整且向观看者呈现时(例如,实时、近实时、在严格延时预算内等)产生。出于说明目的,生理监测信号包含来自不同传感器的描述如映射到内容图像(内容图像中的位置或图像区域)的观看者的注视位置的两个生理监测信号。分别由两个注视位置生理监测信号产生的两个注视位置由传感器融合与分离块(例如装置、模块等)汇集到单个(经评估)注视位置,接着,其与来自叙事元数据的期望或预期图像ROI进行比较。假定出于一些原因观看者在视觉上正注视在场景中呈现的图像的不重要的部分,且因此经汇集经评估注视位置在与对应于如指叙事元数据中指代的特定ROI的预期注视位置进行比较时会导致差异。可将此差异提供为到叙事状态(或注意点)内容修改模型228-2的输入以基于情绪状态差异连同来自叙事元数据的可能信号修改选项一起产生或识别所选择的信号修改224-3。所述差异用于控制希望将观看者的注视朝向ROI移位回去的所选择的信号修改(224-3)。

[0225] 在一些操作案例中,内容修改模型(228-2)用于至少部分基于状态差异或注视位置差异确定所选择的信号修改(224-3)的特定操作参数的量级或值。所选择的信号修改(224-3)的操作参数可用于在给定时间点将媒体内容部分修改成将在给定时间点回放(通过图像或音频处理、或两者)的实际媒体内容部分。

[0226] 图3F说明图3E中所展示的媒体呈现处理的特定实例。用于补偿注视位置与ROI失配的叙事元数据将应用居在中在ROI处的局部锐化滤波器。

[0227] 在回放装置(236)上具有基于显示器的相机的眼睛跟踪器可用于提供关于观看者的注视位置估计(标示为位置1(x<sub>2</sub>,y<sub>2</sub>))。与眼镜、耳塞等一起操作的EOG模块可用于从观看者获取EOG信号以进行关于观看者的注视位置估计(标示为位置1(x<sub>3</sub>,y<sub>3</sub>))。

[0228] 从生理监测信号估计的注视位置可经汇集以输出整体注视位置(或经评估注意

点;标示为位置1(x4,y4))且与由被创意人员所期望的叙事元数据中的叙事状态指定的预期注视位置(或预期注意点;(标示为位置1(x1,y1)))进行比较。接着,注意点差异(例如,预期注视位置与经评估注视位置之间的差异等)可经导出及提供为到内容修改模型(228-2)的输入。

[0229] 信号修改的内容修改模型(228-2)采用如从生理监测产生的注意点(或叙事状态)差异的量级及/或如在叙事元数据中指定的信号修改选项,且产生或选择指示出于将观看者的经评估注意点移位到叙事元数据中指定的ROI的目的控制局部锐化滤波器的强度、展度及/或羽化(渐变)的信号修改选项。接着,从内容修改模型(228-2)输出的所选择的信号修改选项可用于在内容回放(222)中实施特定信号修改(224-3)。举例来说,锐化滤波器可应用在位置1(x1,y1)处的预期注意点,而模糊滤波器可应用在位置1(x4,y4)处的经评估注意点。锐化及模糊滤波器的区域大小及/或羽化可至少部分关于通过生理监测及E&N元数据确定的注意点差异的量级进行控制。举例来说,用户的眼睛趋向于被吸引或转向到图像的相对锐利的空间区域。

[0230] 图3G说明图3E中所展示的媒体呈现处理的另一特定实例。在此实例中,观看者正在具有使用站立的玻璃振动从媒体内容(117-2)中直接从具有3x3位置网格分辨率(例如,水晶声音技术等)的图像显示器的屏幕接收到的音频数据再现或产生声音的完整沉浸感声音系统(例如ATMOS声音系统等)及大型高端图像显示器(例如105英寸图像显示器等)的家庭影院中观看。

[0231] 在许多操作案例中,在沉浸式音频处理中采用低容差捕捉选项。术语“捕捉”意味着捕捉到最近定位的扬声器的音频对象位置(或以从最近定位的扬声器发出音频对象的声音)。在此低容差捕捉选项下,单个扬声器的使用-如与配合平移或内插使用多个扬声器相反-在沉浸式音频处理中是有利的(或有可能被选择)。单个扬声器的使用更好地保留声音的音色方面或质量但牺牲了被描绘为发出声音的音频对象的位置准确度。

[0232] 在本实例中,正向观看者呈现的媒体内容(117-2)是具有烛光场景的电影,其中牛顿(Newton)(或人物)正在进行炼金术实验,更明确来说,正在探索金属赘生。深夜在大教堂中的烛光场景描绘大理石地板散布着全都在运动中的复杂的晶状银色纹理,伴随着金属的起皱声。复杂的晶状银色纹理的一部分是将形状从晶状改成生物形态的树突形状,而由所关注音频对象表示或描绘为从所关注音频对象发出的-对应声音从那个活动变为具有更多微妙的人类语音低音的流体弯音(暗示牛顿正在寻求“重要精神”)。更明确来说,这些声音经定位到描绘复杂的晶状银色纹理的上文提及的部分中的树突生长的异常区域的图像区域。

[0233] 在大型显示器呈现环境中,在相机缓慢地方大异常区域以最终展示牛顿的入射面的凸面反射之前,描绘于图像区域中的异常区域仅占据在大型图像显示器(或屏幕)上呈现的一小部分图像且因此可能容易被忽略。由于图像显示器相对较大,即使观看者正在树突生长区域的大致邻域中寻找,观看者的注视位置仍稍微偏离使得异常区域(或预期注意点)刚好落在观看者的近中心凹外部。因为观看者对观看者的近中心凹外部的视觉对象的视觉分辨率不太敏锐,所以晶状与更多生物形态的纹理之间的区别在观看者的视觉中无法区分。

[0234] 相同生理监测信号及图3F中使用的E&N元数据(或其中的数据字段)的叙事状态部

分中的相同ROI及AOI元数据可用于图3G中所说明的本实例中。然而,在此本实例中,创意人员已在E&N元数据中描绘及指定用于使观看者的经评估注意点重定向的信号修改选项是通过音频处理。

[0235] 如先前提及,观看者的经评估注意点与观看者的预期注意点之间的偏差或差异可在媒体内容(117-2)正被调整且在此大型显示器呈现环境中向观看者呈现时通过生理监测检测。

[0236] 响应于确定观看者的经评估注意点从由创意人员使用ROI及/或AOI指示的预期注意点偏离(例如,在观看者的近中心凹外部等),观看者的注意点可通过音频处理引导到预期注意点或其中神秘生长树突区域看起来是活跃的异常区域。

[0237] 在一些操作案例中,元数据规范(例如SMPTE ST 2098等)可用于陈述或指定媒体元数据(117-1)的数据字段。媒体元数据(117-1)的数字字段中的一者可用于描述或指示音色或音频对象位置在沉浸式音频呈现中是否是相对重要的。

[0238] 在本实例中,根据创作意图,如果ROI或AOI落在观看者的近中心凹外,那么精确定位的AOI比保留AOI的声音音色更重要。创作意图可指示高容差捕捉选项-如与在其它沉浸式音频处理案例中受到青睐的低容差捕捉选项相比-在媒体元数据(117-1)的上文提及的数据字段中。

[0239] 给定当观看者的经评估注意点从观看者的预期注意点偏离时在媒体元数据(117-1)的叙事元数据中指定为信号修改选项的高容差捕捉选项,使用高容差捕捉选项(或设置)会导致声音通过九(或3x3)个扬声器元件在音频对象的准确位置(异常区域)下呈现,如与被放置到屏幕上的九个位置中的一者中相反(对应于玻璃面板扬声器的3x3声音网格中的九个扬声器元件位置)。高容差捕捉选项避免或防止在单个位置处离散化成可能会导致音频呈现中的音频对象位置(或AOI的位置)将与描绘应该会发出相同声音的屏幕上的异常区域的相对小图像区域失配的单个扬声器元件。

[0240] 在本实例中,准确地放置音频对象位置及容忍音色质量劣化的捕捉选项已经被创意人员设置或选择为在各种可能信号修改选项的排名中的信号修改选项。然而,应注意,在除了本实例之外的许多其它操作案例中(例如,在多播放器场景中的乐器等),音频处理可偏向以更将声音更准确地放置在准确屏幕位置处为代价出于防止音色失真的目的使用单个扬声器。

[0241] 另外、任选地或替代地,由于混响(reverb/reverbetation)还引起声音位置扩散,所以创意人员可指定本实例中的混响从其默认设置减小的意图,这是因为混响将由于所描绘场景中的教堂设置而相对较高。

[0242] 用于指示音频对象的位置的音频空间表示可标示为 $A(x, y, z)$ 。同样地,用于指示所描绘视觉对象的位置的图像空间表示可标示为 $I(x, y, z)$ 。在非限制性实例中,音频空间表示中的位置可转换到如下的图像空间表示中的对应位置中: $I(x, y, z) = A(x, z, y)$ 。即,图像空间表示中的 $z$ 维度/轴(指示深度)对应于音频空间表示中的 $y$ 维度/轴,而图像空间表示中的 $y$ 维度/轴(指示高度)对应于音频空间表示中的 $z$ 维度或轴,例如在其中SMPTE 2098用于指定本文中所描述的经编码位流中的元数据编码语法的一些操作案例中。

[0243] 如在叙事元数据中指定的音频空间表示中的音频对象的预期位置(标示为 $A(x_1, z_1, y_1)$ )可经转换到图像空间表示的对应预期位置(标示为 $I(x_1, y_1, z_1)$ )。如转换到图像空

间表示中的音频对象的预期位置表示观看者的预期注意点,且与由从图像空间表示中的显示器注视位置(标示为 $l(x_2, y_2)$ )及EOG注视位置(标示为 $l(x_3, y_3)$ )产生的经汇集注视位置估计(标示为 $l(x_4, y_4)$ )表示的观看者的经评估注意点进行比较。

[0244] 观看者的预期注意点与观看者的经评估注意点之间的差异(如使用位置的 $x$ 及 $y$ 维度确定)可由内容修改模型(228-2)用作输入以进行信号修改以产生指示减小混响且取消将所关注音频对象捕捉到屏幕后面的扬声器位置的信号修改选项-现在取消捕获的AOI被捕捉到其的扬声器可使用扬声位置内插来确定或选择-出于使观看者的经评估注意点移位到指定叙事元数据中的AOI或对应ROI(例如异常区域等)的目的。接着,从内容修改模型(228-2)输出的信号修改选项可用于在内容回放(222)中实施特定信号修改(224-4)。举例来说,特定信号修改可导致媒体呈现处理增加AOI的音量、减小AOI的混响、在 $I(x_1, y_1)$ 处取消将AOI捕捉到屏幕后面的所选择的扬声器的位置、将AOI的声音捕捉到所选择的扬声器。用于增加音量、减小混响、用于选择扬声器的位置容差的操作参数等可取决于观看者的预期注意点与观看者的经评估注意点之间的差异的量级进行设置。

[0245] 此大型显示器呈现环境可与其中观看者在小型图像显示器(或屏幕)上观看相同场景的小型显示器呈现环境形成对比。在小型显示器环境中(例如,如由针对回放装置(236)配置的环境特性化数据指示等),无论如何,针对场景的多数经呈现图像都很可能落在观看者的近中心凹内(具有相对敏锐或锐利的视觉)。异常区域(看起来是活跃的)中的树突形状很可能会被观看者注意到,而无需诉诸于将应用于大型显示器呈现环境中的先进音频补偿处理。

[0246] 如提及,许多其它实例可使用类似媒体呈现处理设计,但具有由媒体元数据(117-1)描述的不同特定情绪以及其中指定的信号修改选项。

[0247] 11. 实例内容调整过程

[0248] 图2H说明表示以下各者的实例曲线图:(i)在制作阶段(例如图2A的202等)中基于如由媒体内容(117-2)的创意人员指定的预期情绪及/或叙事状态产生的媒体内容(例如图1A的117-2等)的媒体特性(例如亮度等);(ii)在消费阶段(例如图2A的204等)中通过生理监测产生的观看者的经评估状态;及(iii)将由回放装置在消费阶段(204)中对媒体内容(117-2)的媒体特性(例如亮度等)作出的用于在观看者的预期状态与经评估状态之间实现、或试图实现零差异(标示为“ $0 \Delta$ ”)的媒体内容调整或修改(例如亮度差异等)。用于产生媒体内容及元数据、执行生理监测、及作出媒体内容调整/修改的过程可由包括回放装置及连同回放装置一起操作的一或多个生理监测装置/组件的一或多个计算装置执行。

[0249] 至少部分基于(i)使用反映或表示创意输入的E&N元数据(例如,在图1A的媒体元数据117-1等中)指示的观看者的预期情绪及/或叙事状态、(ii)从可用生理监测信号所确定/估计/预测的观看者的经评估状态、及(iii)使用E&N元数据指示的信号修改选项,媒体内容调整/修改可从用于将如在经编码位流(例如图1A的122等)中接收的媒体内容(117-2)的原始视觉及/或音频(声学)特性更改为向观看者呈现的媒体内容的经修改视觉及/或音频(声学)特性的各种各样的媒体内容调整/修改进行选择。

[0250] 将如本文中描述那样被调整/修改的实例视觉特性包含(但不一定仅限于)以下各者中的任一者:(例如最小值、最大值、平均值、高光、中间调、暗区域等)亮度、亮度动态范围、色彩、饱和度、色相、空间分辨率、图像刷新速率、放大或缩小操作、图像转向(图像经转

向以跟随观看者从房间到房间的移动)等等。这些视觉特性中的任一者、一些或全部可关于经呈现图像序列、视觉场景(由两个连续场景切换界限)、视觉场景的细分、一群组图片(GOP)、跨越多个帧的一或多个图块大小的区域、空间时间流的组块、整个图像(例如,平均图片电平或APL等)、与描绘特定人物或对象相关的(经呈现/经表示图像的)一或多个图像区域中的图像区域相关、等等进行测量。

[0251] 将如本文中描述那样被调整/修改的实例音频特性包含(但不一定仅限于)以下各者中的任一者:音频对象位置(或描绘于表示或呈现于呈现环境中的音频声场中的音频源的空间位置)、音频对象的大小/半径(例如点音频源、具有有限大小的音频源、例如风、环境声音的扩散性音频源等)、音频对象的方向及/或轨迹、对话及/或非对话音量、对话增强、音频动态范围、特定喇叭选择、特定喇叭配置、光谱均衡、音色、混响、回声、光谱/频率相依处理、相位及/或延迟、音频攻击或发行时间等等。这些音频特性中的任一者可关于音频帧/块序列、音频场景、音频场景的细分、原声、单个音频对象、音频对象集群、声音元素、整个声场、与(经呈现/经表示声场的)一或多个声场区域中的声场区域相关、描绘特定人物或对象的所关注音频或声学对象、等等进行测量。

[0252] 在运行时由回放装置选择的媒体内容调整/修改(或信号修改)可作用于(或更改)媒体内容(117-2)的一或多个视觉特性。另外、任选地或替代地,由回放装置选择的媒体内容调整/修改(或信号修改)可作用于(或更改)媒体内容(117-2)的一或多个音频特性。另外、任选地或替代地,由回放装置选择的媒体内容调整/修改(或信号修改)可作用于(或更改)媒体内容(117-2)的一或多个视觉及/或音频特性的组合。应进一步注意,在各个实施例中,可在媒体消费会话中在内容回放(例如电影、TV节目等)的不同时间点(例如不同场景等)使用不同信号修改。

[0253] 仅出于说明目的,媒体内容(例如图1A的117-2等)沿着内容时间线(例如图2A的210等)具有一个半小时的回放持续时间。在此回放持续时间期间,媒体内容(117-2)的创意人员预期或期望观看者体验随时间而变化的一或多个特定预期情绪及/或叙事状态。包含(但不限于)一或多个特定预期情绪及/或叙事状态的创意人员的意图(或创作意图)可用于产生媒体内容(117-2)及对应于媒体内容(117-2)的媒体元数据(例如图1A的117-1等)。

[0254] 本文中所描述的情绪及/或叙事状态可语义地或非语义地表示于媒体元数据(117-1)及/或媒体内容(117-2)中。如本文中使用的术语“语义地”可意味着在语义表达中使用神经科学领域、电影领域、音频领域、或相关领域中的符号、符记、术语或技术术语描述情绪及/或叙事状态。在许多操作案例中,虽然创意人员可使用语义表达(例如,“观众应理解此关键故事细节”、“在注意点不在此人物上的情况下帮助观众理解此”等)描述或定义预期情绪及/或叙事状态,但创意人员对此预期状态的描述可(例如,以编程方式、完全自动地、一旦语义表达被给定就在无用户交互或最少用户交互的情况下、在定义一或多个范围、阈值的进一步用户交互下、整体地、部分地等)经转译或转换成与经呈现图像(例如视觉场景、视觉场景的细分、个别图像、图像的部分或区域等)及/或经呈现音频(例如,经呈现声学、相关音频声场、经呈现音频对象、音频场景、音频场景的细分、个别音频帧、个别音频对象等)的潜在视觉及/或音频特性紧密相关联的非语义表示(例如,如工程过程、图1A的媒体制作块115等中所定义)。

[0255] 通过说明但无限制,在制作阶段(202)中,观看者的预期状态-例如预期唤醒,其表

示预期可测量情绪状态的预期情绪状态或维度-在随着时间的推移消费媒体内容(117-2)时在图2H中说明为粗实线的随时间而变化的原始或预调整平均图片电平(或APL)下被转译/转换成或经实施于媒体内容(117-2)中。

[0256] 在一些操作案例中,随着时间的推移语义地描述的观看者的预期状态到可修改视觉及/或音频特性(例如随着内容时间(在从制作阶段(202)输出的发行版本中)的推移非语义地描述的APL)的转译、转换及实施可为部分基于一或多个E&N状态到媒体特性转译/转换/实施模型(例如算法、方法、程序、操作等)。转译/转换/实施可为(但不一定仅限于)用于使用明确选择的视觉及/或音频特性影响观看者的特定情绪及/或叙事状态的一或多个理论及/或实验模型。这些模型可(例如,以编程方式、完全自动地、一旦语义表达被给定就在无用户交互或最少用户交互下、在定义一或多个范围、阈值的进一步用户交互下、整体地、部分地等)在例如最大值、最小值、平均亮度、其它视觉特性、非视觉特性等的额外输入情况下并入或改变输出。

[0257] 用于将语义地描述的情绪及/或叙事状态转译、转换及/或实施到(低级)非语义视觉及/或音频特性的一些或全部这些转译/转换/实施模型可基于平均观看者(例如,如由人类视觉系统或HVS等表示)的响应(例如,在一群体的不同媒体内容类型或一或多个特定媒体内容类型的子集等下收集)及/或平均收听者(例如,在平均听力及声学理解等下)实施。另外、任选地或替代地,用于将语义地描述的情绪及/或叙事状态转译、转换及/或实施到(低级)非语义视觉及/或音频特性的一些或全部这些转译/转换/实施模型可基于表示各个子集人口统计的观看者(例如,恐怖片的影迷、马术爱好者等)的响应实施。

[0258] 响应于接收媒体内容(117-2)及媒体元数据(117-1),回放装置可向观看者呈现媒体内容(117-2);使用在呈现环境中与回放装置一起操作的可用生理监测装置/传感器/处理器在观看者正消费(观看且收听)与媒体内容(117-2)一起呈现的视觉及/或音频内容时监测随时间而变化的观看者的情绪及/或叙事响应(或产生生理监测信号);使用观看者的情绪及/或叙事响应(或生理监测信号)产生随时间而变化的例如经评估唤醒的观看者的特定经评估情绪及/或叙事状态等。观看者的特定经评估情绪及/或叙事状态可具有与观看者的特定预期情绪及/或叙事状态(例如唤醒)相同的种类。通过实例但无限制,随着时间的推移的观看者的特定经评估情绪及/或叙事状态(例如唤醒)可在图2H的细实曲线中表示为随着时间的推移的百分位值。

[0259] 由于观看者可能偏离转译/转换/实施模型中用于转译或映射观看者的预期状态的平均观看者/收听者,且还由于回放装置在其中进行操作的呈现环境可能偏离媒体内容(117-2)在其处标定的参考呈现环境,所以观看者的评估状态(或图2H的细实线)可能偏离或不同于观看者的预期状态(或图2H的粗实线),例如用如媒体元数据(117-1)中指定的预期百分位值(未展示)表示的预期唤醒。

[0260] 举例来说,在内容时间线(210)的第一时间点(对应于图2H中的具有数字1的圆圈),响应于确定通过生理监测(如图2H的细实线中指示)所估计或预测的观看者的经评估状态(例如经评估唤醒)与在第一时间点如在媒体元数据(117-1)中指示的观看者的预期状态(例如经评估唤醒(未展示))相比响应不足,回放装置可应用如通过第一时间点处图2H的虚线与粗实线之间的差异表示的第一媒体内容调整/修改(或第一信号修改)来在如实施于接收到的媒体内容(117-2)中的第一时间点处将第一原始或预调整APL改变或提高到从接

收到的媒体内容(117-2)导出的经呈现媒体内容中的第一时间点处的第一经调整或经修改APL。如从原始或预调整APL提高的第一经调整或经修改APL可用于导致观看者的经评估状态或唤醒朝向观看者的预期状态或唤醒移动(朝向实现零差异或 $0\Delta$ )或变得更被唤醒。

[0261] 第一媒体内容调整/修改、或如通过图2H的虚线与粗实现之间的差异表示的APL的提高可通过基于实施于回放装置中的闭环系统中的负反馈调整亮度色调映射(例如,调整最大值、最小值及平均亮度值、调整枢轴、斜率、亮度值分布/映射中的偏移等)产生,其中其量级与观看者的预期状态与经评估状态之间的差异的量级成比例或与其相称。另外、任选地或替代地,第一媒体内容调整/修改可至少部分基于类似于用于在制作阶段(202)中转译、转换及/或实施媒体内容(117-2)中的观看者的特定预期情绪及/或叙事状态的模型的模型产生。

[0262] 在内容时间线(210)的第二时间点(对应于图2H中的具有数字2的圆圈),响应于确定通过生理监测(如图2H的细实线中指示)所估计或预测的观看者的经评估状态(例如经评估唤醒)与在第二时间点如在媒体元数据(117-1)中指定的观看者的预期状态(例如经评估唤醒(未展示))相比响应过度,回放装置可应用如通过第二时间点处图2H的虚线与粗实线之间的差异表示的第二媒体内容调整/修改(或第二信号修改)来在如实施于接收到的媒体内容(117-2)中的第二时间点处将第二原始或预调整APL改变或降低到从接收到的媒体内容(117-2)导出的经呈现媒体内容中的第二时间点处的第二经调整或经修改APL。如从原始或预调整APL降低的第二经调整或经修改APL可用于导致观看者的经评估状态或唤醒朝向观看者的预期状态或唤醒移动(朝向实现零差异或 $0\Delta$ )或变得不太被唤醒。

[0263] 第二媒体内容调整/修改、或如通过图2H的虚线与粗实现之间的差异表示的APL的降低可基于实施于回放装置中的闭环系统中的负反馈产生,其中其量级与观看者的预期状态与经评估状态之间的差异的量级成比例或与其相称。另外、任选地或替代地,第二媒体内容调整/修改可至少部分基于类似于用于在制作阶段(202)中转译、转换及/或实施媒体内容(117-2)中的观看者的特定预期情绪及/或叙事状态的模型的模型产生。

[0264] 针对内容时间线(210)的第三时间点(对应于图2H中具有数字3的圆圈),由回放装置基于生理监测、接收到的媒体内容(117-2)及/或接收到的媒体元数据(117-1)执行的媒体呈现操作类似于在第二时间点执行的媒体呈现操作。

[0265] 针对内容时间线(210)的第四时间点(对应于图2H中具有数字4的圆圈),由回放装置基于生理监测、接收到的媒体内容(117-2)及/或接收到的媒体元数据(117-1)执行的媒体呈现操作类似于在第一时间点执行的媒体呈现操作。

[0266] 针对内容时间线(210)的第五时间点(对应于图2H中具有数字5的圆圈),由回放装置基于生理监测、接收到的媒体内容(117-2)及/或接收到的媒体元数据(117-1)执行的媒体呈现操作类似于在第二或第三时间点执行的媒体呈现操作,但由于观看者的预期状态与经评估状态之间的差异小于与第二或第三时间点相关联的差异所以是在较低程度上类似。在一些操作案例中,当观看者的预期状态与经评估状态之间的差异小于E&N状态差异阈值(例如,预配置的、动态配置的、自动应地设置的等)时,在第五时间点不作出调整。

[0267] 针对内容时间线(210)的第六时间点(对应于图2H中具有数字6的圆圈),由回放装置基于生理监测、接收到的媒体内容(117-2)及/或接收到的媒体元数据(117-1)执行的媒体呈现操作类似于在第一或第四时间点执行的媒体呈现操作,但由于观看者的预期状态与

经评估状态之间的差异小于与第一或第四时间点相关联的差异所以是在较低程度上类似。在一些操作案例中,当观看者的预期状态与经评估状态之间的差异小于E&N状态差异阈值(例如,预配置的、动态配置的、自动应地设置的等)时,在第五时间点不作出调整。

[0268] 针对内容时间线(210)的第七时间点(对应于图2H中具有数字7的圆圈),响应于确定观看者的预期状态与经评估状态之间的差异小于E&N状态差异阈值(例如,预配置的、动态配置的、自动应地设置的等),回放装置不作出调整。

[0269] 如图2H中展示,如基于创作意图在媒体内容及元数据中指示、指定及/或实施的观看者的预期情绪及/或叙事状态可随内容时间线(210)中的时间或由回放装置实施的内容回放而改变。在一些时间点或时间间隔(或一些场景)处,可预期观看者更加激动,而在一些其它时间点或时间间隔(或一些其它场景)处,可预期观看者不太激动,甚至是沉默或安静的,例如以便为大规模的震惊或兴趣增强或情绪唤起热身或做准备。类似地,在一些时间点或时间间隔(或一些场景)处,可预期观看者更加参与其中,而在一些其它时间点或时间间隔(或一些其它场景)处,可预期观看者不太参与其中,甚至是放松的。

[0270] 如基于创作意图在媒体内容及元数据中指示、指定及/或实施的观看者的预期状态提供由回放装置实施的闭环系统围绕其可打算或试图实现零差异的经编程(或在制作阶段(202)中是可编程的)基线。更明确来说,如先前提及,对应于观看者的预期状态的观看者的经评估状态可通过接收及处理由在呈现环境中与回放装置一起操作的可用生理装置/传感器产生的(例如实时、近实时等)生理监测信号。因此,观看者的经评估状态(例如经评估唤醒)可通过可用生理装置/传感器(例如EEG电极、GSR传感器等)产生,且与观看者的预期状态进行比较。观看者的经评估状态与预期状态之间的差异(例如经评估及预期唤醒)可由通过回放装置实施的闭环系统在内容回放中用作负反馈以试图在观看者的经评估状态与预期状态之间实现零差异,在一些操作案例中经受状态差异阈值。

[0271] 应注意,观看者的经评估或预期状态不仅仅限于如通过特定类型的生理监测装置/传感器/工具测量的观看者的经评估或预期生理响应,例如唤醒。观看者的经评估或预期状态可通过如通过其它类型的生理监测装置/传感器/工具测量的其它类型的生理响应指定、传达及/或测量。如图2C中说明,回放装置特性化数据238及/或周围环境特性化数据240可用于回放装置(236)的内容回放及修改操作(244)中。在制作阶段(202)中,创意人员可(例如,同时、深埋在媒体内容或元数据中等)指定将使用不同生理监测装置/传感器/工具在不同呈现环境中测量的不同生理响应类型。回放装置可使用回放装置特性化数据(238)及/或周围环境特性化数据(240)确定或选择一或多个特定生理监测装置/传感器/工具(在不同生理监测装置/传感器/工具当中)来监测关于观众的一或多个特定生理响应类型。

[0272] 仅出于说明目的,已描述,可关于特定亮度相关特性(例如APL)基于在制作阶段中产生的接收到的媒体内容及元数据及(例如实时、近实时等)生理监测实施(例如实时、近实时等)媒体内容调整/修改。应注意,在各个实施例中,可关于其它亮度相关特性(例如最大、最小及平均亮度值、特定图像区域、特定对象、特定人物、背景的亮度值等)基于在制作阶段中产生的接收到的媒体内容及元数据及(例如实时、近实时等)生理监测实施(例如实时、近实时等)媒体内容调整/修改。另外、任选地或替代地,可关于其它视觉特性(例如色彩精度、饱和度、色相、空间分辨率、图像刷新速率、放大及/或缩小操作等)基于在制作阶段中产生

的接收到的媒体内容及元数据及(例如实时、近实时等)生理监测实施(例如实时、近实时等)媒体内容调整/修改。另外、任选地或替代地,可关于音频特性、运动相关特性、触觉特性等基于在制作阶段中产生的接收到的媒体内容及元数据及(例如实时、近实时等)生理监测实施(例如实时、近实时等)媒体内容调整/修改及相关呈现操作。另外、任选地或替代地,支持媒体内容调整或修改的类型的不同组合及/或支持生理监测的类型的不同组合的不同发行版本可在不同呈现环境中由不同类型的回放装置制作及消费。

[0273] 实施本文中所述的技术的媒体制作系统可在不同级别下与创意人员交互以产生媒体内容(例如图1A的117-2等)及媒体元数据(例如图1A的117-1等)。举例来说,如由创意人员提供的用户输入中的语义表达(指示内容时间线的各个时间点或场景的预期状态)可经接收、提取、变换、嵌入、及/或实施于媒体内容(117-2)及媒体元数据(117-1)中。对应于从语义表达提取及/或转译的预期状态的(或与其相关联的)观看者的情绪或叙事响应可在媒体内容正被动态地调适且向观看者呈现时通过生理监测来评估。举例来说,在各个时间点或场景处的对应于从语义表达(例如,在故事板信息中、在创意人员的编辑中等)提取及/或转译的预期叙事状态的观看者的认知负荷可通过生理监测来评估以导致特别适于将观看者的经评估叙事状态收敛到预期叙事状态的媒体内容调整/修改(例如,增加对话音量、增加对话的信噪比等)。

[0274] 本文中所述的技术可用于防止盲目地进行对收敛于预期状态不必要的媒体内容修改且取决于观看者、观看者的回放装置、观看者的回放装置在其中操作的呈现环境等等个别地作出不同媒体内容修改。因此,针对具有影响第一观看者的叙事状态或认知负荷的听力问题的第一观看者,可增加对话音量。针对在嘈杂呈现环境中的第二观看者,可增加对话信噪比,而非提高对话音量以导致第二观看者感觉对话音量太响。针对使用具有有效地屏蔽环境噪声的耳机的回放装置的第三观看者,可降低对话音量。在确定媒体内容调整/修改的特定类型及特定调整量级时还可将例如环境光、混响、回声等的其它因素考虑在内。在一些操作案例中,媒体内容调整/修改的特定类型及/或特定调整量级可在无除了由创意人员提供的语义表达之外的来自创意人员的用户输入的情况下被确定或完全自动地产生。在各个实施例中,没有一个、一些或所有用于确定或产生媒体内容调整/修改的特定类型及/或特定调整量级的选择因素、选择性加入选项、选择性退出选项、缩放、阈值、下限及上限等可通过用户接口暴露给创意人员或相关联艺术及/或工程专业人员(或用户)且完全或部分由创意人员或相关联艺术及/或工程专业人员(或用户)控制。另外、任选地或替代地,可向在不同领域中工作的创意人员给出更多或更少控制。在一些操作案例中,如音频专业人员相比,更熟悉对比度、饱和度等如何影响观众/观看者的预期情绪及/或叙事状态的视频专业人员例如通过用户接口、故事板等被给予更多控制以操纵视觉特性及视觉特性及对媒体内容及元数据中表示的视觉特性的响应。

[0275] 在制作阶段(例如图2A的202等)中,媒体内容(例如图1A的117-2等)及媒体元数据(例如图1A的117-1等)可关于参考呈现环境(例如电影院、家庭影院、平板计算机、移动手机等)创建。举例来说,音频内容及相关元数据部分可以ATMOS格式经创建用于相对高端音频内容呈现环境。

[0276] 在消费阶段(例如图2A的204等)中,特定呈现环境中的回放装置(具有耳塞耳机等)可将关于参考呈现环境创建的媒体内容(117-2)及媒体元数据(117-1)调适或变换为关

于特定呈现环境将呈现的媒体内容。举例来说,在制作阶段(202)中如以ATMOS格式经创建的用于相对高端音频内容呈现环境的音频内容及相关元数据部分可经调适或变换(例如,维度缩减等)成适用于回放装置(例如耳塞等)将呈现的音频内容。

[0277] 12. 实例过程流

[0278] 图4A说明根据本发明的实例实施例的实例过程流。在一些实例实施例中,一或多个计算装置或组件可执行此过程流。在框402中,媒体制作系统接收描述与媒体内容的一或多个部分相关的情绪预期及叙事信息的用户输入。

[0279] 在框404中,媒体制作系统至少部分基于用户输入产生与媒体内容的一或多个部分相关的一或多个预期生理上可观察状态。

[0280] 在框406中,媒体制作系统将视听内容信号与媒体内容及包括媒体内容的一或多个部分的一或多个预期生理上可观察状态的媒体元数据提供到回放设备。

[0281] 在实施例中,所述视听内容信号致使所述回放装置(a)使用一或多个生理监测信号关于观看者确定与所述媒体内容的所述一或多个部分相关的一或多个经评估生理上可观察状态且(b)至少部分基于所述一或多个预期生理上可观察状态及所述一或多个经评估生理上可观察状态从媒体内容产生经修改媒体内容同时调整且向所述观看者呈现从所述媒体内容产生的所述经修改媒体内容。

[0282] 图4B说明根据本发明的实例实施例的实例过程流。在一些实例实施例中,一或多个计算装置或组件可执行此过程流。在框452中,媒体制作系统接收视听内容信号与媒体内容及媒体元数据。在实施例中,所述媒体元数据包括媒体内容的一或多个部分的一或多个预期生理上可观察状态。

[0283] 在实施例中,与所述媒体内容的所述一或多个部分相关的所述一或多个预期生理上可观察状态是至少部分基于描述与媒体内容的一或多个部分相关的情绪预期及叙事信息的用户输入产生。

[0284] 在框454中,媒体制作系统使用一或多个生理监测信号关于观看者确定与所述媒体内容的所述一或多个部分相关的一或多个经评估生理上可观察状态。

[0285] 在框456中,媒体制作系统至少部分基于所述一或多个预期生理上可观察状态及所述一或多个经评估生理上可观察状态从所述媒体内容产生并呈现经修改媒体内容同时调整且向所述观看者呈现从所述媒体内容产生的所述经修改媒体内容。

[0286] 在各个实例实施例中,一种设备、系统、设备或一或多个其它计算装置执行所描述的前述方法的任一者或一部分。在实施例中,一种非暂时性计算机可读存储媒体存储当由一或多个处理器执行时致使执行本文中所描述的方法的软件指令。

[0287] 注意,尽管本文中讨论了单独的实施例,但本文中论述的实施例及/或部分实施例的任何组合可经组合以形成另外实施例。

[0288] 13. 实施机制-硬件概述

[0289] 根据一个实施例,本文中描述的技术由一或多个专用计算装置实施。专用计算装置可经硬接线以执行所述技术,或可包含永久经编程以执行所述技术的数字电子装置,例如一或多个专用集成电路(ASIC)或现场可编程门阵列(FPGA),或可包含经编程以依据固件、存储器、其它存储装置或组合中的程序指令执行所述技术的一或多个通用硬件处理器。此类专用计算装置还可组合定制硬接线逻辑、ASIC或FPGA与定制编程以实现所述技术。专

用计算装置可为桌面计算机系统、便携式计算机系统、手持式装置、联网装置或并入硬接线及/或程序逻辑以实施所述技术的任何其它装置。

[0290] 举例来说,图5是说明本发明的实例实施例可实施于其上的计算机系统500的框图。计算机系统500包含用于传达信息的总线502或其它通信机构及与总线502耦合以处理信息的硬件处理器504。硬件处理器504可为例如通用微处理器。

[0291] 计算机系统500还包含耦合到总线502以存储将由处理器504执行的信息及指令的主存储器506,例如随机存取存储器(RAM)或另一动态存储装置。主存储器506还可用于在将由处理器504执行指令的执行期间存储临时变量或其它中间信息。此类指令当存储于处理器504可存取的非暂时性存储媒体中时使计算机系统500变成经定制以执行指定于指令中的操作的专用机器。

[0292] 计算机系统500进一步包含只读存储器(ROM)508或耦合到总线502以用于存储处理器504的静态信息及指令的另一静态存储装置。

[0293] 例如磁盘或光盘、固态RAM的存储装置510经提供且耦合到总线502以存储信息及指令。

[0294] 计算机系统500可经由总线502耦合到显示器512,例如液晶显示器,以向计算机用户显示信息。包含字母数字及其它按键的输入装置514耦合到总线502以将信息及命令选择传达给处理器504。另一类型的用户输入装置是用于将方向信息及命令选择传达给处理器504且用于控制显示器512上的光标移动的光标控件516,例如鼠标、跟踪球或光标方向按键。此输入装置在第一轴(例如x)及第二轴(例如y)的两个轴中通常具有两个自由度,此允许装置指定平面中的位置。

[0295] 计算机系统500可使用与计算机系统组合地致使或编程计算机系统500成为专用机器的定制硬接线逻辑、一或多个ASIC或FPGA、固件及/或程序逻辑实施本文中描述的技术。根据一个实施例,本文中的技术由计算机系统500响应于处理器504执行主存储器506中所含的一或多个指令的一或多个序列而执行。此类指令可从例如存储装置510的另一存储媒体读取到主存储器506中。主存储器506中所含的指令序列的执行致使处理器504执行本文中描述的过程步骤。在替代实施例中,硬接线电路系统可代替软件指令或与软件指令组合来使用。

[0296] 本文中所使用的术语“存储媒体”是指存储致使机器以特定方式操作的数据及/或指令的任何非暂时性媒体。此类存储媒体可包括非易失性媒体及/或易失性媒体。非易失性媒体包含例如光盘或磁盘,例如存储装置510。易失性媒体包含动态存储器,例如主存储器506。常见形式的存储媒体包含例如软盘(floppy disk/flexible disk)、硬盘、固态驱动、磁带或任何其它磁性数据存储媒体、CD-ROM、任何其它光学数据存储媒体、具有孔图案的任何物理媒体、RAM、PROM及EPROM、FLASH-EPROM、NVRAM、任何其它存储器芯片或卡盒。

[0297] 存储媒体不同于但可结合传输媒体使用。传输媒体参与在存储媒体之间传送信息。举例来说,传输媒体包含同轴电缆、铜线及光纤光学器件,其包含包括总线502的线。传输媒体还可采用声波或光波的形式,例如在无线电波及红外数据通信期间产生的声波或光波。

[0298] 在将一或多个指令的一或多个序列载送到处理器504以供执行时可涉及到各种形式的媒体。举例来说,指令最初可在远程计算机的磁盘或固态驱动上载送。远程计算机可将

指令加载到其动态存储器中及使用调制解调器经由电话线发送指令。在计算机系统500本地的调制解调器可接收电话线上的数据及使用红外传输器将数据转换成红外信号。红外检测器可接收红外信号中载送的数据,且适当的电路系统可将数据放置在总线502上。总线502将数据载送到主存储器506,处理器504从主存储器506检索及执行指令。由主存储器506接收的指令可任选地在由处理器504执行之前或之后经存储在存储装置510上。

[0299] 计算机系统500还包含耦合到总线502的通信接口518。通信接口518将双向数据通信耦合提供到连接到本地网络522的网络链路520。举例来说,通信接口518可为集成服务数字网络(ISDN)卡、电缆调制解调器、卫星调制解调器或将数据通信连接提供到对应类型的电话线的调制解调器等。作为另一实例,通信接口518可为将数据通信连接提供到兼容LAN的局域网(LAN)卡。还可实施无线链路。在任何此实施方案中,通信接口518发送及接收载送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。

[0300] 网络链路520通常通过一或多个网络将数据通信提供到其它数据装置。举例来说,网络链路520可通过本地网络522将连接提供到主机计算机524或由因特网服务提供商(ISP)526操作的数据设备。ISP 526又通过现在通常称为“因特网”528的全球分组数据通信网络提供数据通信服务。本地网络522及因特网528两者都使用载送数字数据流的电信号、电磁信号或光学信号。将数字数据载送到计算机系统500且载送来自计算机系统500的数字数据的通过各个网络的信号及在网络链路520上且通过通信接口518的信号是传输媒体的实例形式。

[0301] 计算机系统500可通过网络、网络链路520及通信接口518发送消息及接收数据,包含程序代码。在因特网实例中,服务器530可通过因特网528、ISP 526、本地网络522及通信接口518传输针对应用程序请求的代码。

[0302] 接收到的代码可在其被接收时由处理器504执行及/或存储于存储装置510或其它非易失性存储装置中以供稍后执行。

[0303] 14. 等效物、扩展、替代物及其它

[0304] 在前述说明书中,已参考可因实施方案而异的众多特定细节描述了本发明的实例实施例。因此,本发明的内容及申请人所期望的本发明的内容的唯一且排他性指示是从本申请案发出的权利要求的集,其呈此类权利要求的具体发出形式,包含任何后续修正。本文中针对此类权利要求中所包含的术语的明确陈述的任何定义均应决定权利要求中所使用的此类术语的含义。因此,在权利要求书中没有明确叙述的任何限制、要素、性质、特征、优点或属性都不应以任何方式限制此权利要求书的范围。因此,说明书及附图应被认为是说明性的而不是限制性的。

[0305] 枚举的示范性实施例

[0306] 本发明可已本文中描述的形式中的任一者体现,包含(但不限于)描述本发明的一些部分的结构、特征及功能性的下文枚举的实例实施例(EEE)。

[0307] EEE1. 一种计算机实施方法,其包括:

[0308] 接收描述与媒体内容的一或多个部分相关的情绪预期及叙事信息的创作意图输入;

[0309] 至少部分基于所述创作意图输入产生与所述媒体内容的所述一或多个部分相关的一或多个预期生理上可观察状态;

[0310] 将视听内容信号与所述媒体内容及包括所述媒体内容的所述一或多个部分的所述一或多个预期生理上可观察状态的媒体元数据提供到回放设备；

[0311] 其中所述视听内容信号致使所述回放装置(a)使用一或多个生理监测信号关于观看者确定与所述媒体内容的所述一或多个部分相关的一或多个经评估生理上可观察状态且(b)至少部分基于所述一或多个预期生理上可观察状态及所述一或多个经评估生理上可观察状态从所述媒体内容产生经修改媒体内容同时调整且向所述观看者呈现从所述媒体内容产生的所述经修改媒体内容。

[0312] EEE2.根据EEE 1所述的方法,其中所述创作意图输入表示致使所述媒体内容及所述媒体元数据在制作阶段产生的创意人员的创作意图。

[0313] EEE3.根据EEE 1或EEE 2所述的方法,其中所述创作意图输入含有创意人员的意图的语义表达,其中所述媒体元数据包括以下各者的一者:用于在消费阶段中导出一组非语义信号修改选项的所述语义表达或在制作阶段基于所述语义表达产生的所述一组非语义信号修改选项,且其中所述回放装置从所述一组信号修改选项选择一或多个特定信号修改选项来对所述媒体内容执行一或多个媒体内容调整来响应于确定所述一或多个预期生理上可观察状态与所述一或多个经评估生理上可观察状态的差异大于差异阈值来最小化所述差异。

[0314] EEE4.一种计算机实施方法,其包括:

[0315] 接收视听内容信号与媒体内容及媒体元数据,其中所述媒体元数据包括所述媒体内容的一或多个部分的一或多个预期生理上可观察状态;

[0316] 其中与所述媒体内容的所述一或多个部分相关的所述一或多个预期生理上可观察状态是至少部分基于描述与媒体内容的一或多个部分相关的情绪预期及叙事信息的创作意图输入产生;

[0317] 使用一或多个生理监测信号关于观看者确定与所述媒体内容的所述一或多个部分相关的一或多个经评估生理上可观察状态;

[0318] 至少部分基于所述一或多个预期生理上可观察状态及所述一或多个经评估生理上可观察状态从所述媒体内容产生并呈现经修改媒体内容同时调整且向所述观看者呈现从所述媒体内容产生的所述经修改媒体内容。

[0319] EEE5.根据EEE 4所述的方法,其中所述一或多个经评估生理上可观察状态包括所述观看者的经评估情绪状态,其中所述一或多个预期生理上可观察状态包括所述观看者的预期情绪状态,其是与所述观看者的所述经评估情绪状态相同类型的情绪状态。

[0320] EEE6.根据EEE 4或EEE 5所述的方法,其中所述一或多个经评估生理上可观察状态包括所述观看者的经评估叙事状态,其中所述一或多个预期生理上可观察状态包括所述观看者的预期叙事状态,其是与所述观看者的所述经评估叙事状态相同类型的叙事状态。

[0321] EEE7.根据EEE 4到6中任一EEE所述的方法,其中所述一或多个经评估生理上可观察状态包括所述观看者的经评估注意点,其中所述一或多个预期生理上可观察状态包括所述观看者的预期注意点。

[0322] EEE8.根据EEE 4到7中任一EEE所述的方法,其中所述媒体元数据包括用于响应于检测到所述一或多个经评估生理上可观察状态与所述一或多个预期生理上可观察状态之间的差异修改所述媒体内容的所述一或多个部分的一或多个信号修改选项。

[0323] EEE9. 根据EEE 8所述的方法, 其中所述一或多个信号修改选项的至少一个信号修改包括用于对以下各者中的一或多个者实施媒体内容修改: 亮度、空间分辨率、锐化、对比度、色彩饱和度、色相、色调映射、视场、色域、亮度动态范围、位深度、空间滤波、图像刷新率、缩放因子、图像转向、非视觉特性、运动呈现特性、枢轴、亮度映射的斜率及偏移、特定图像区域、特定对象、特定人物、背景中的亮度分布、亮度、音频对象的位置、频率均衡、混响、音色、相位、扬声器数目、扬声器配置、扬声器的频率范围、扬声器的相位失真、喇叭选择、音量、实际音频通道配置、用于选择单个扬声器呈现且用于选择多扬声器内插的捕捉容差选项、音频对象位置、音频对象大小、音频对象半径、音频对象方向、音频对象轨迹、对话音量、非对话音量、对话增强、音频动态范围、特定喇叭选择、特定喇叭配置、回声特性、延迟、信号攻击时间、或信号释放时间。

[0324] EEE10. 根据EEE 8或EEE 9所述的方法, 其中所述一或多个信号修改选项用于在所述媒体内容的内容回放中最小化关于所述观看者的所述一或多个经评估生理上可观察状态与所述一或多个预期生理上可观察状态之间的所述差异。

[0325] EEE11. 根据EEE 8到10中任一EEE所述的方法, 其中所述一或多个生理监测信号由以下各者的一或多个者产生: 基于显示器的传感器、可见光波长相机传感器、同步定位及映射传感器、热成像器、头戴式显示器传感器、入耳式传感器、手腕传感器、注视位置传感器、瞳孔直径传感器、面部表情传感器、头位置传感器、观看距离传感器、面部表情传感器、效价传感器、唤醒传感器、脑电图传感器、明确定位的电极、热传感器、光学传感器、眼电图传感器、呼吸传感器、基于体积描记法心率的传感器、皮肤电响应传感器、气体传感器、CO<sub>2</sub>含量传感器、R3COH含量传感器、或基于座椅的传感器。

[0326] EEE12. 根据EEE 8到11中任一EEE所述的方法, 其中所述一或多个信号修改选项是至少部分基于回放装置特性化数据及呈现环境特性化数据产生。

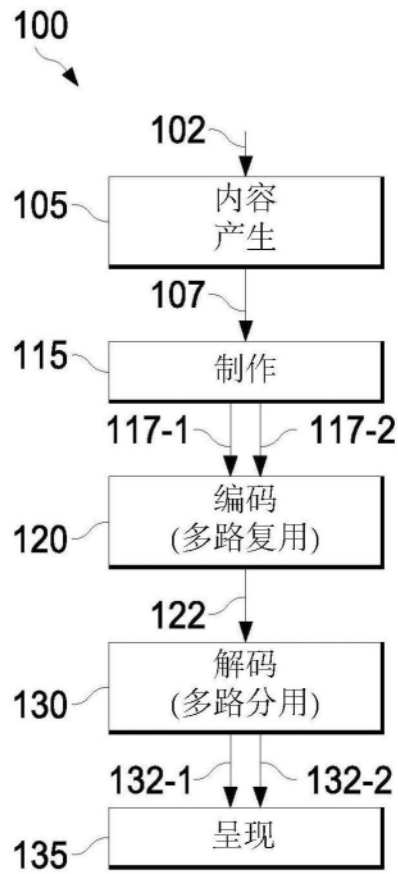


图1A



图1B

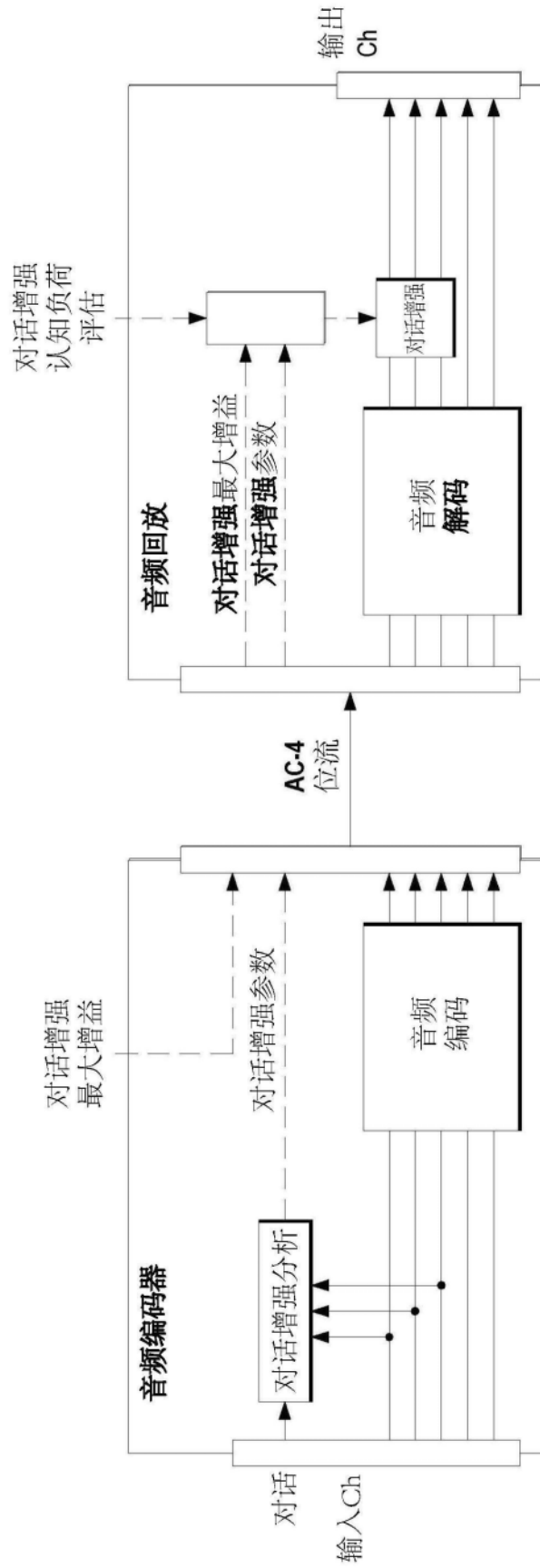


图1C

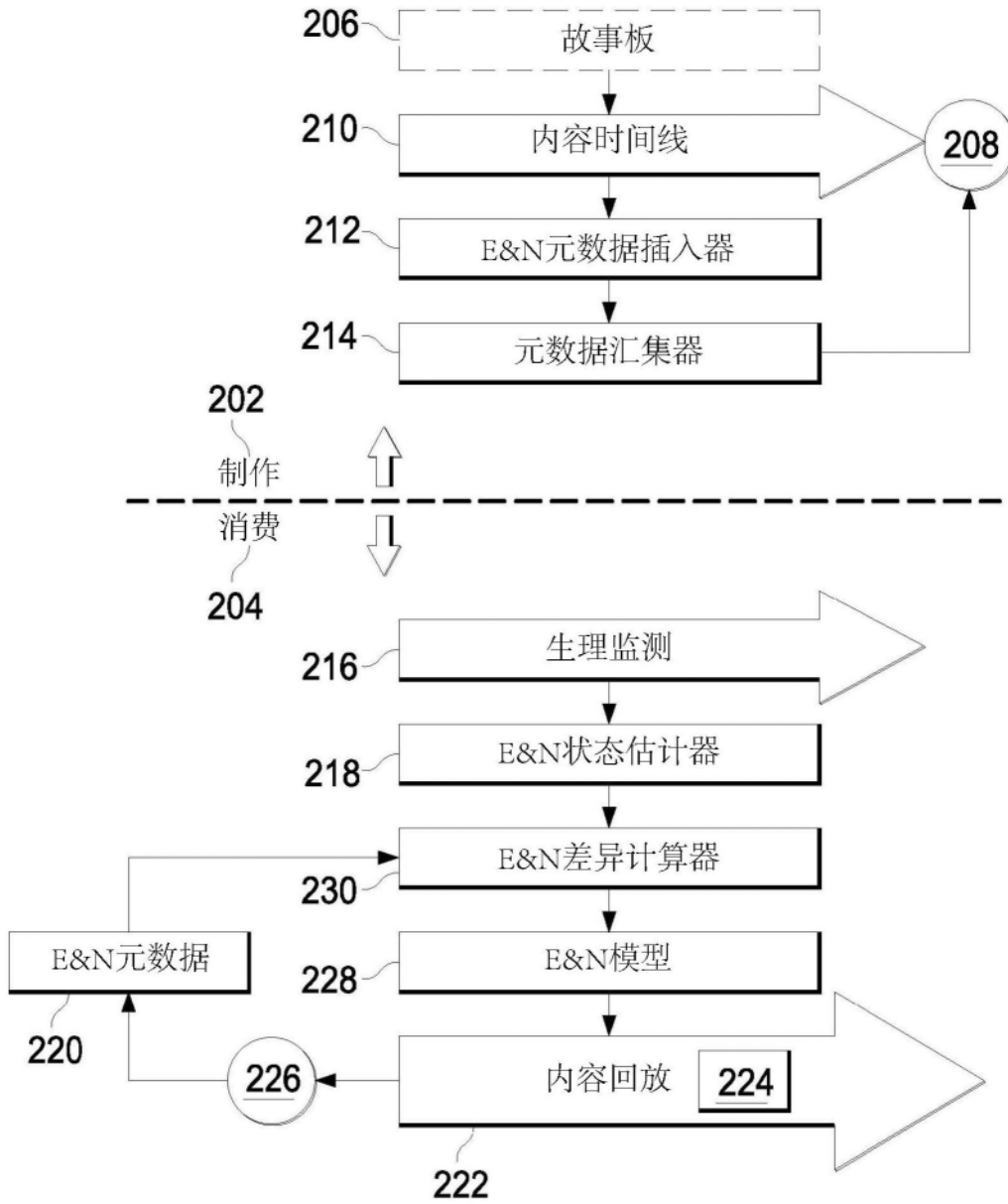


图2A

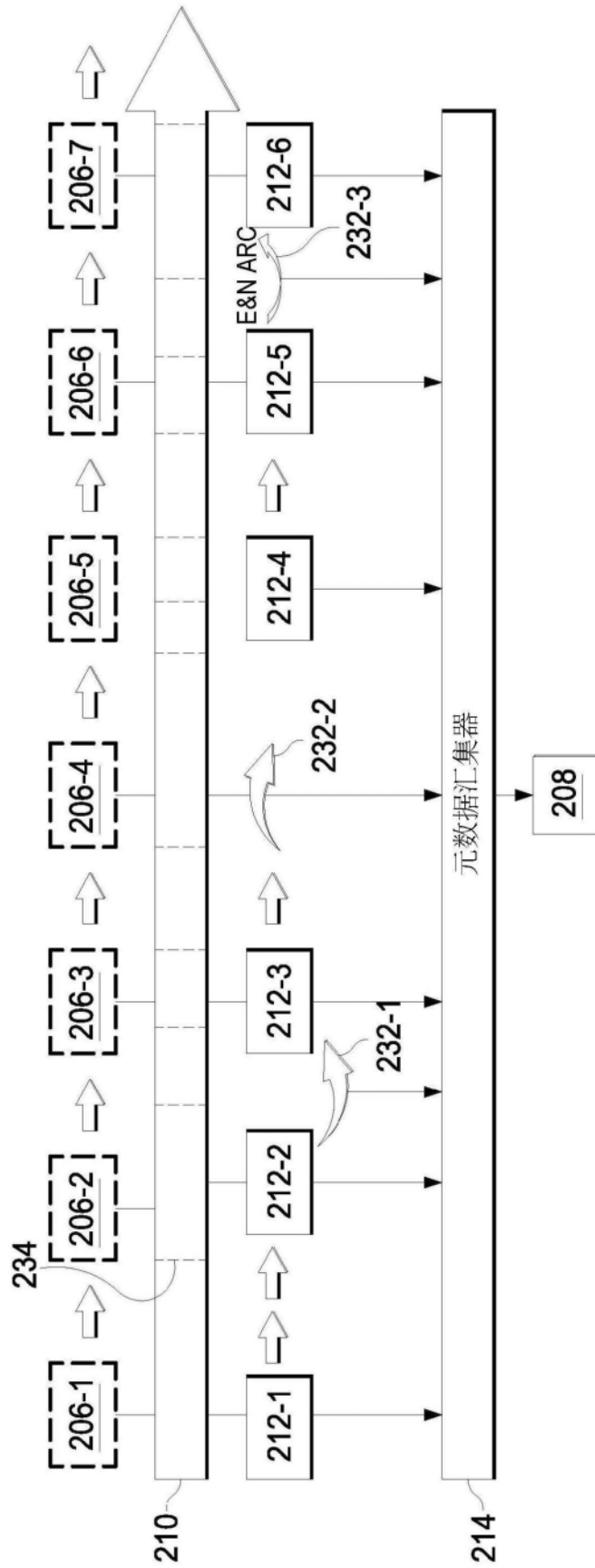


图2B

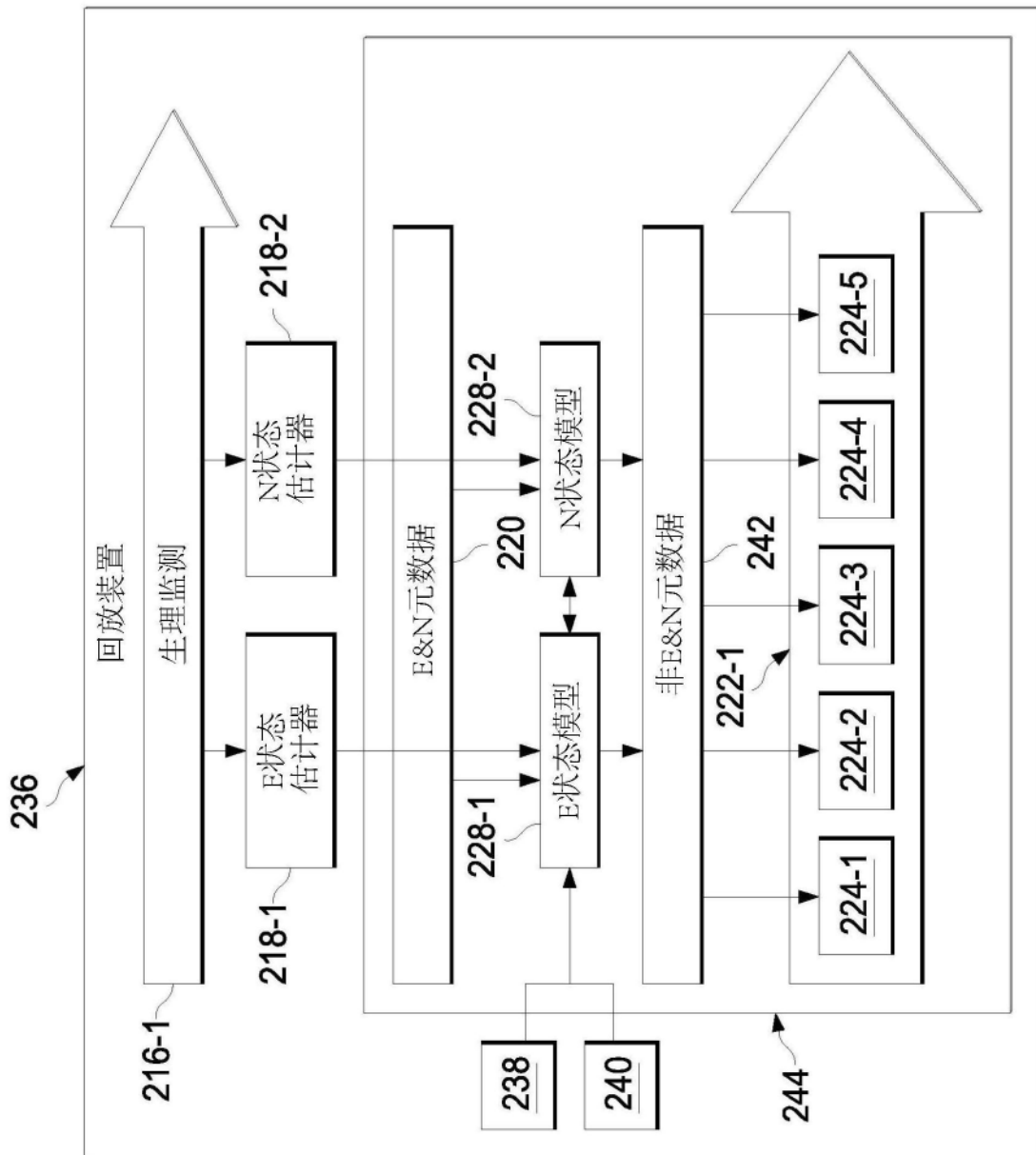


图2C

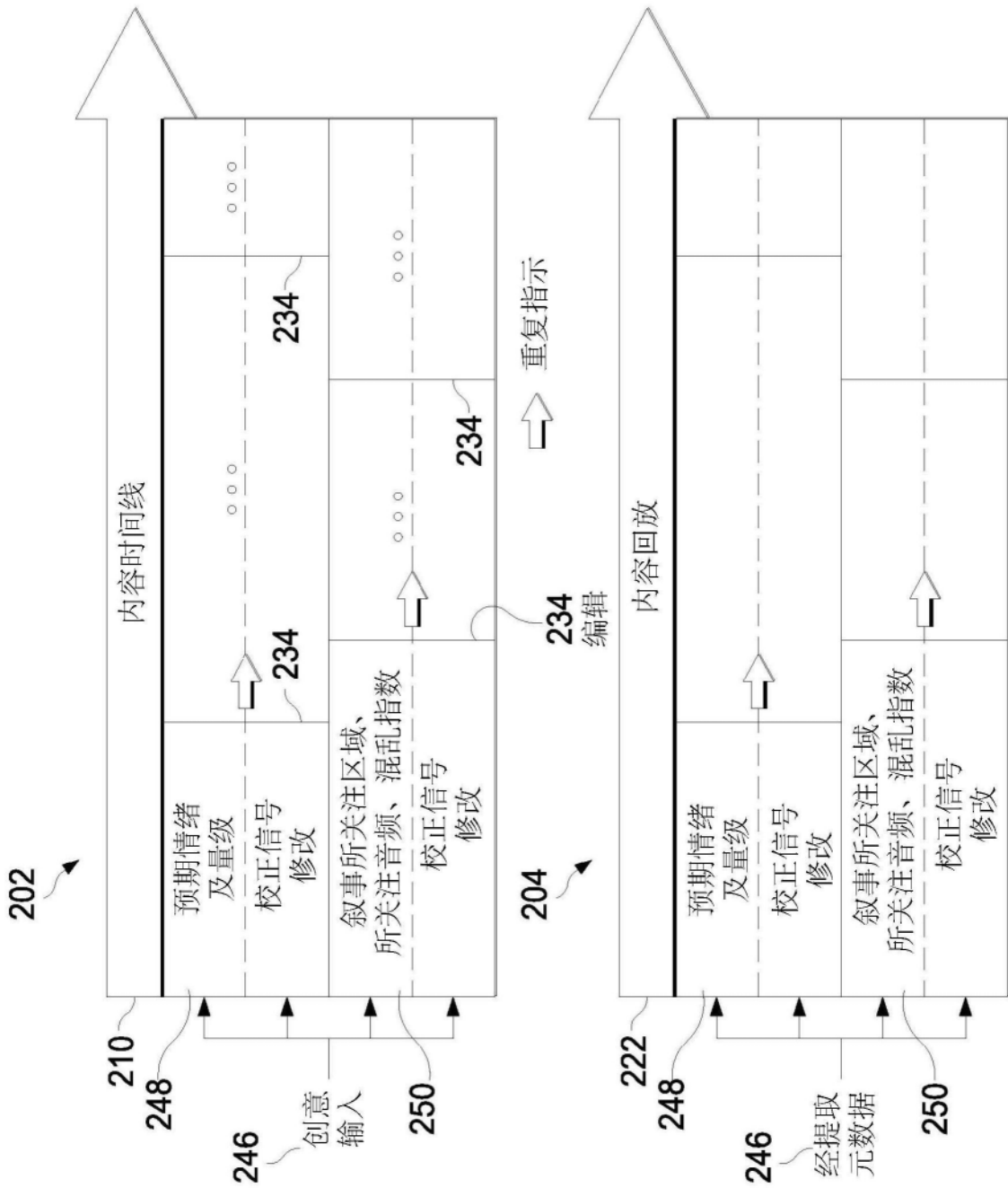


图2D

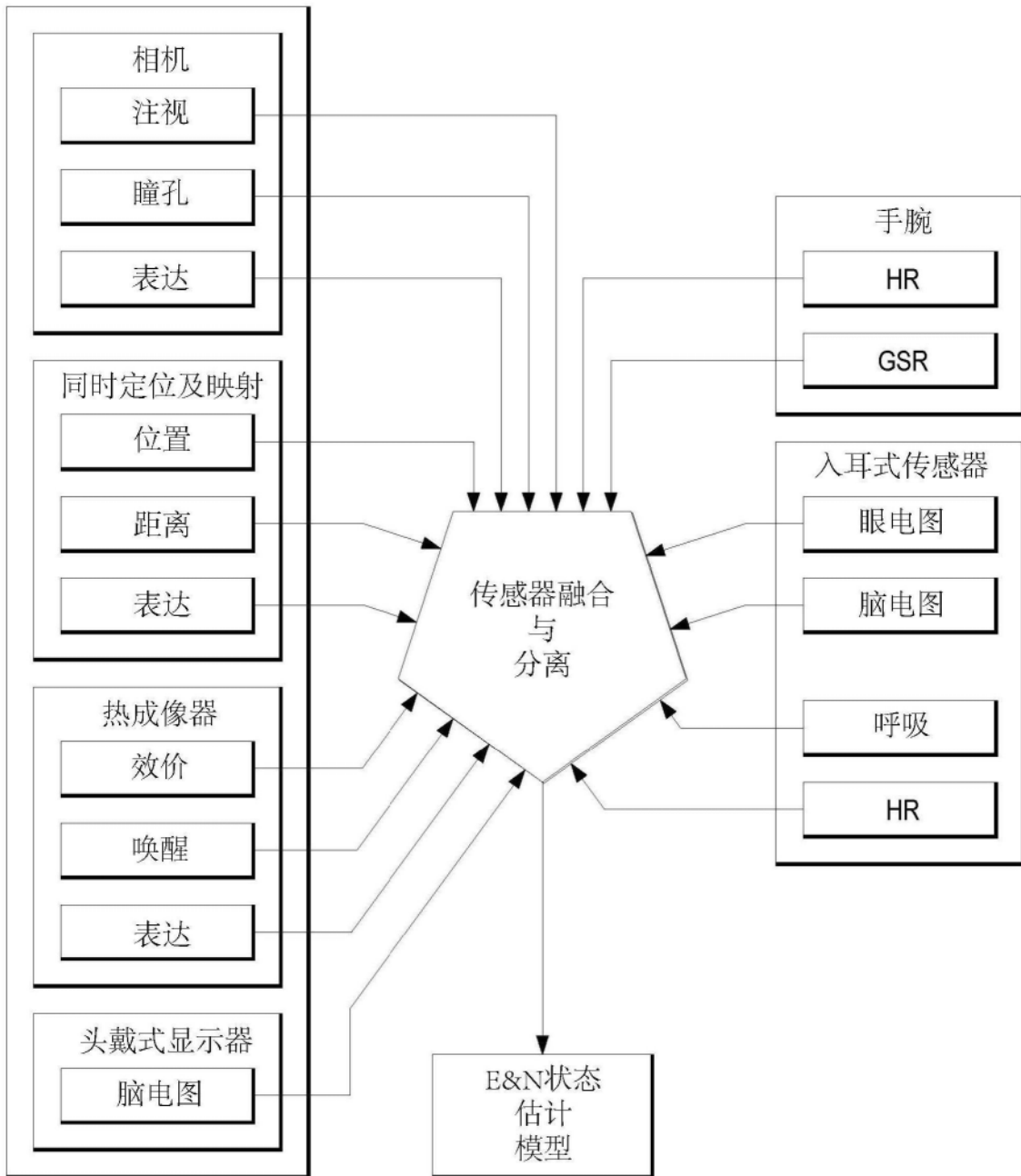


图2E

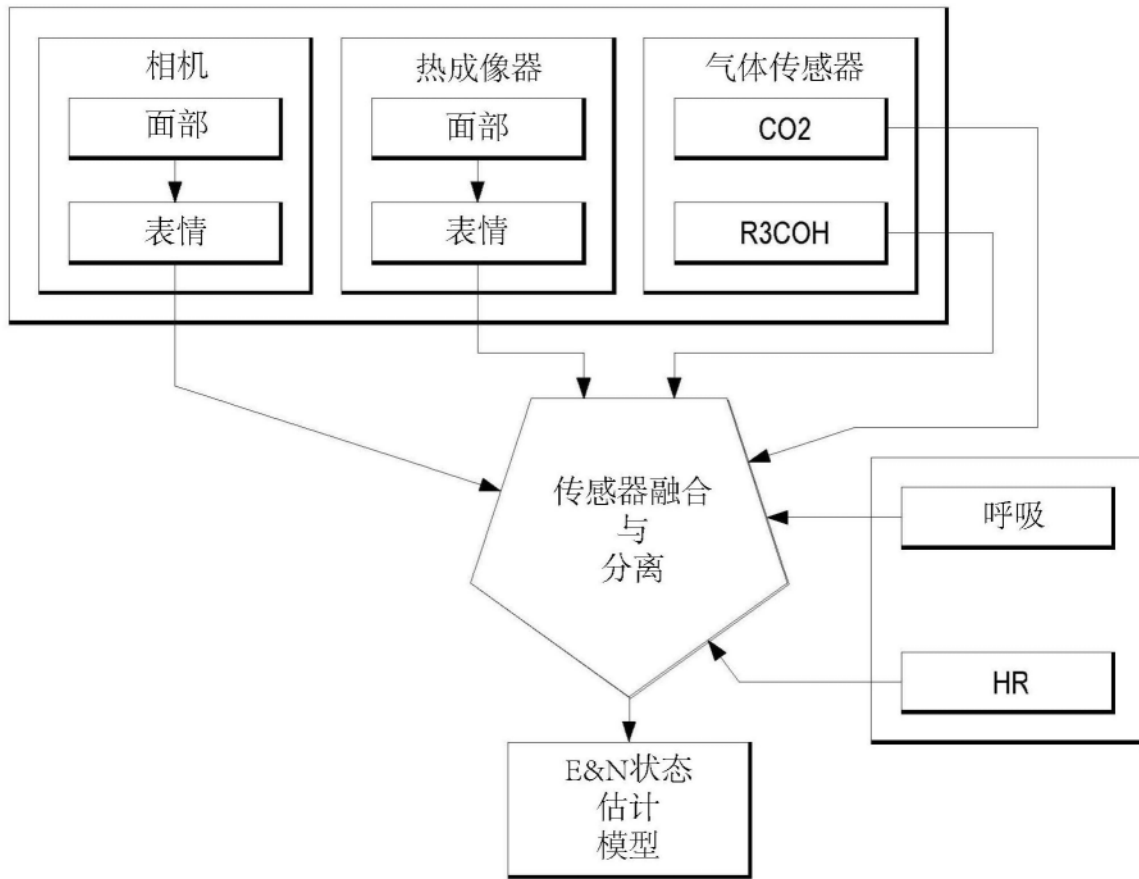


图2F

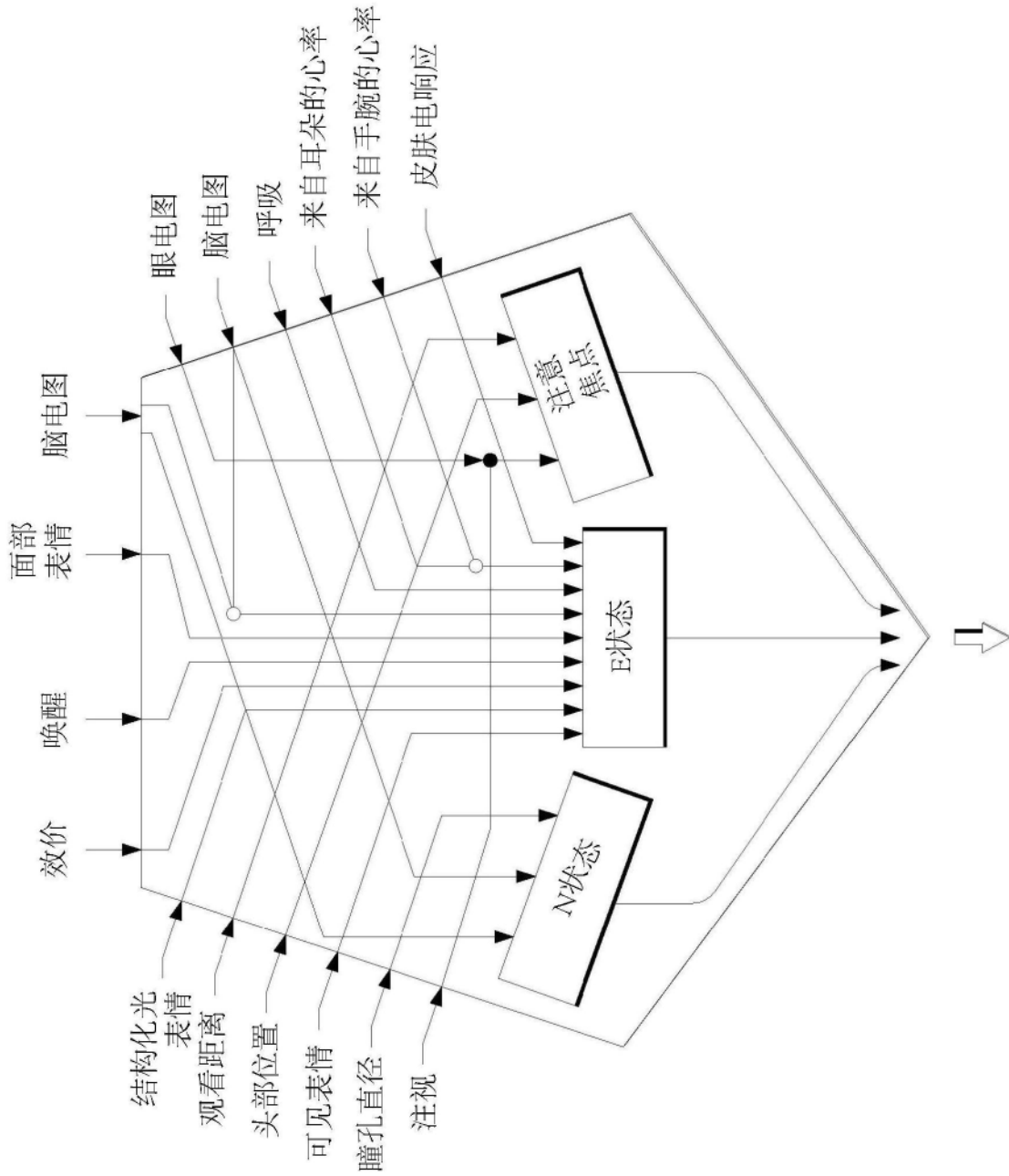


图2G

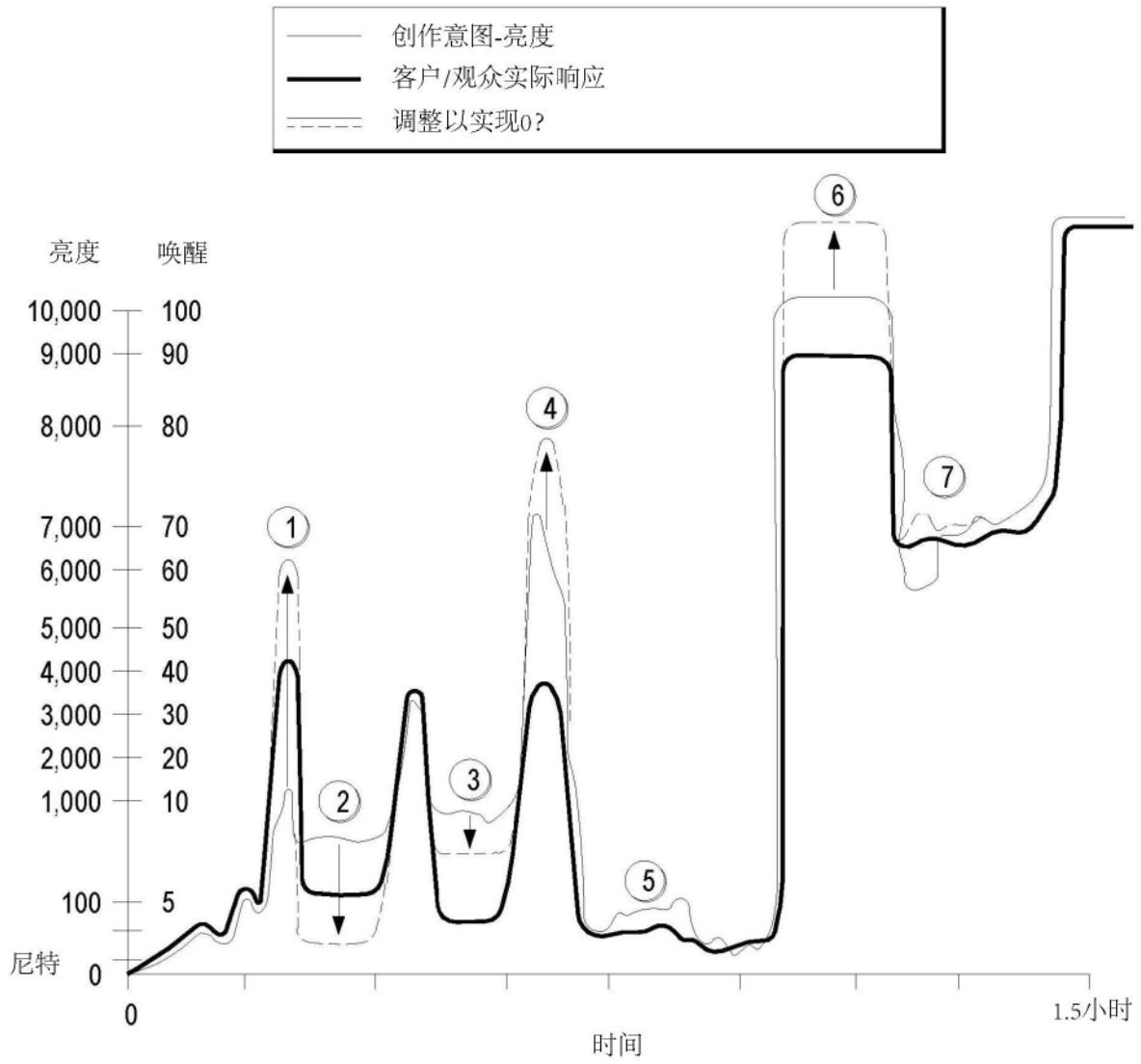


图2H

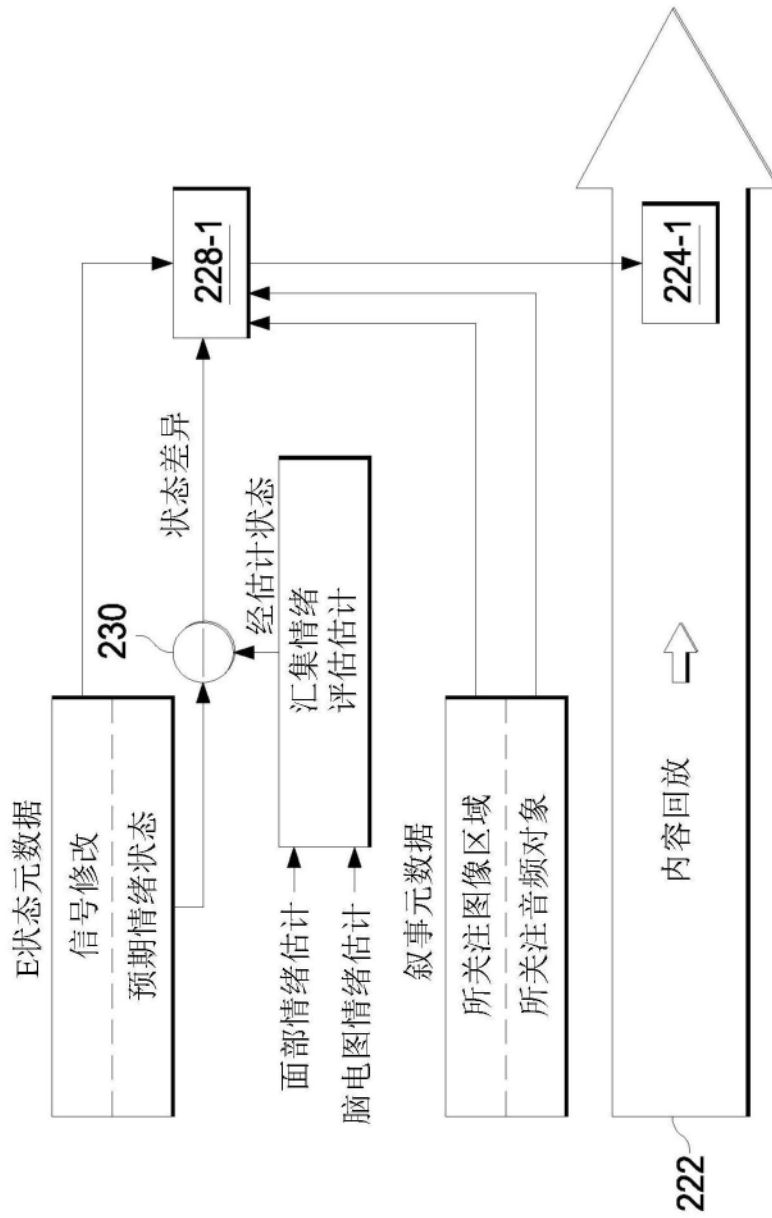


图3A

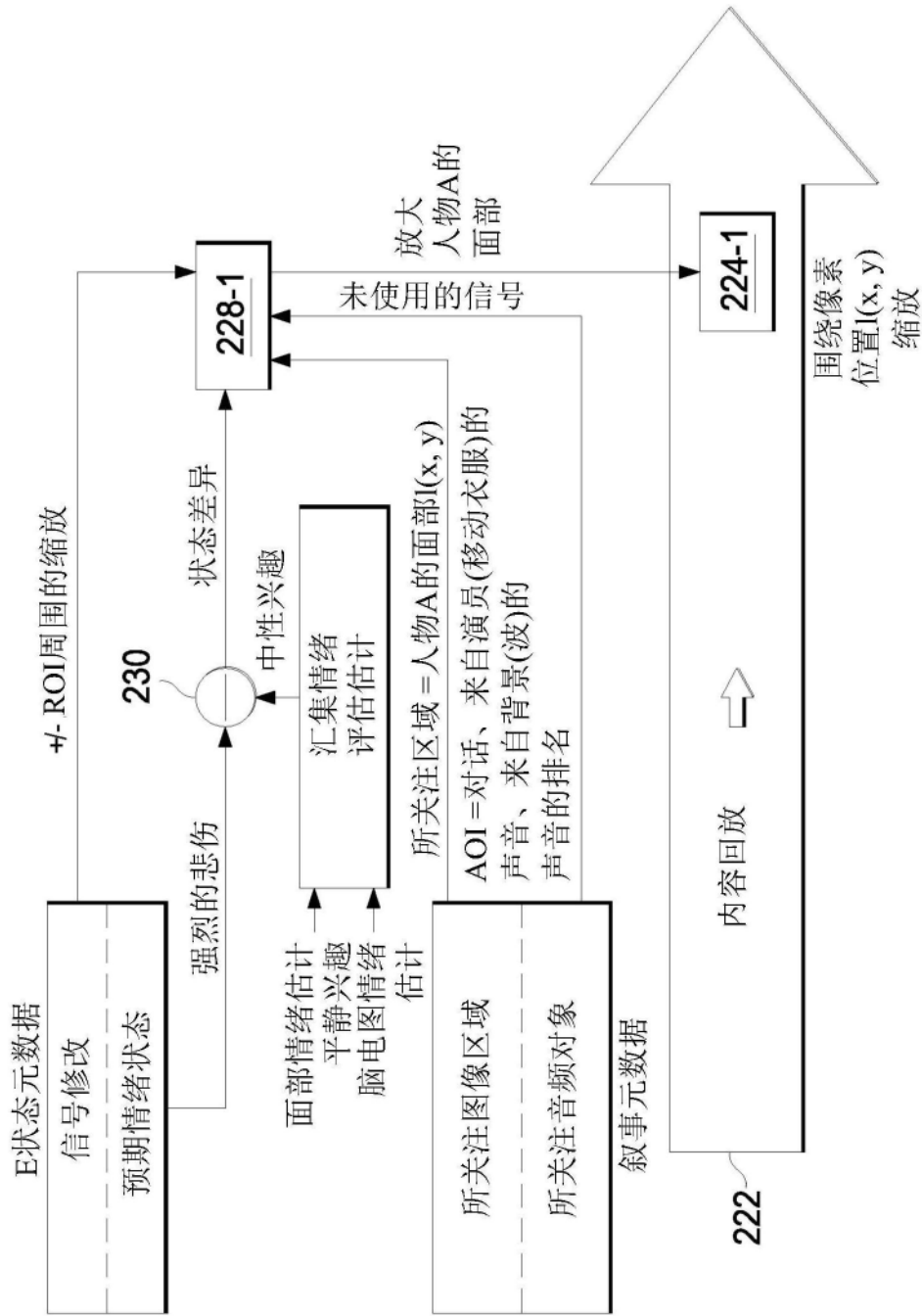


图3B

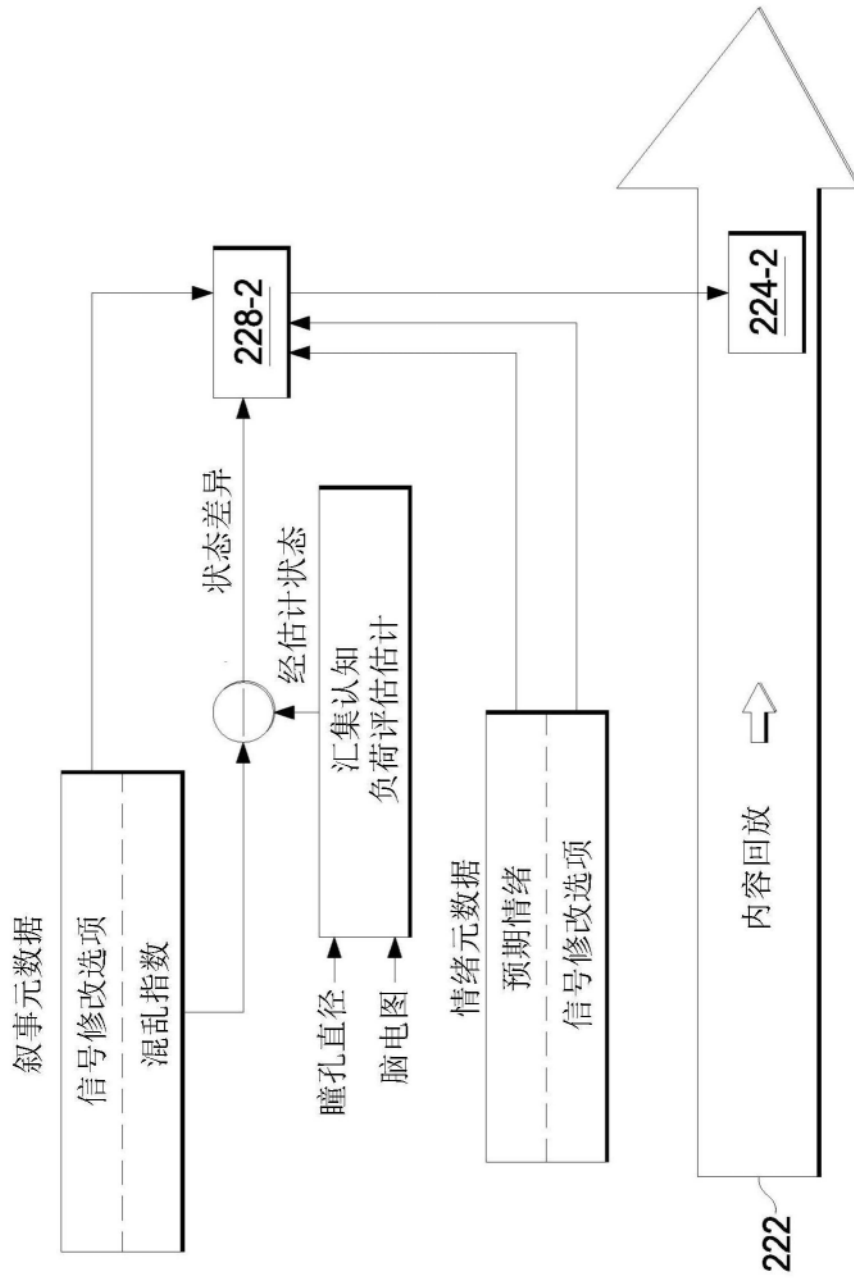


图3C

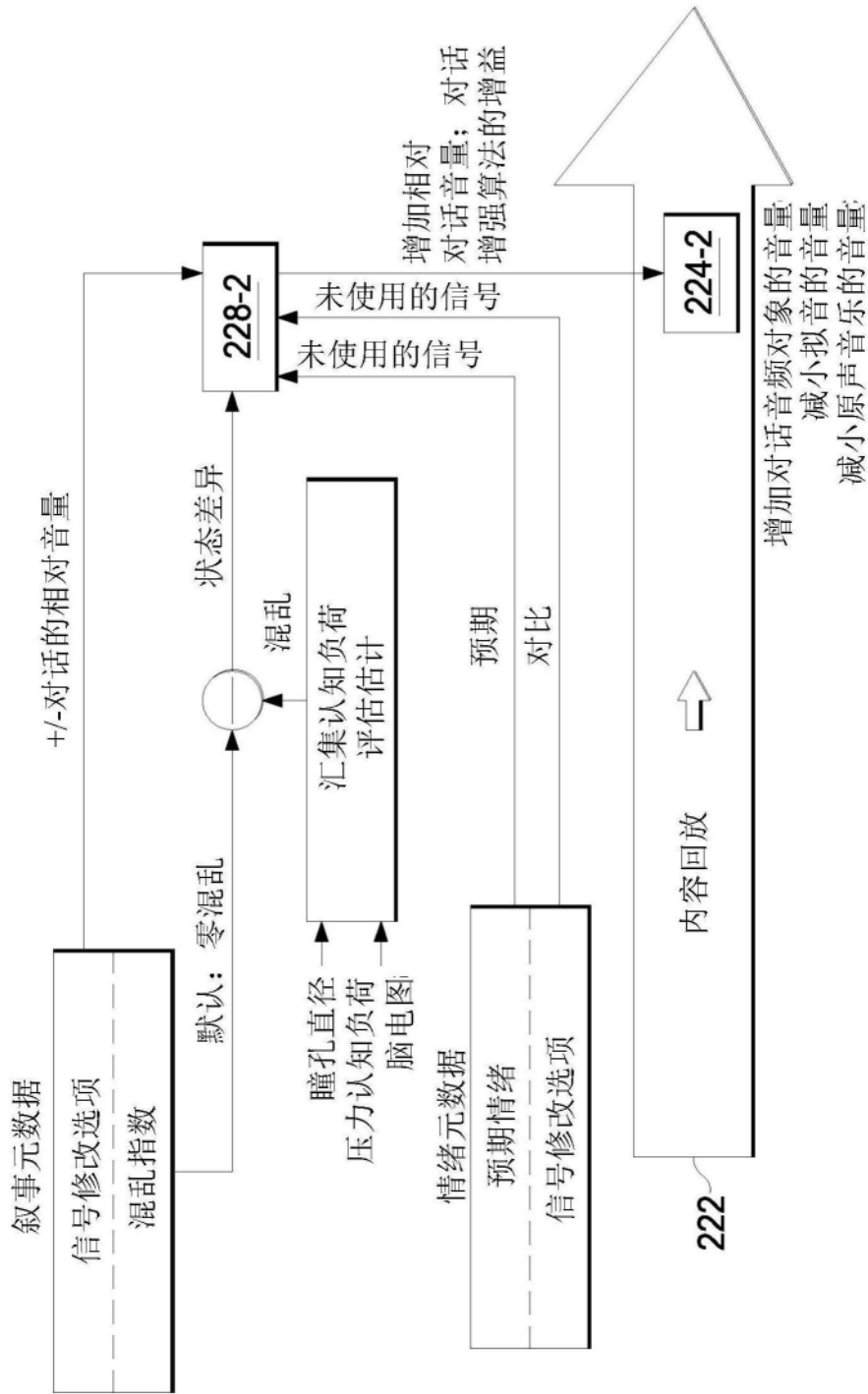


图3D

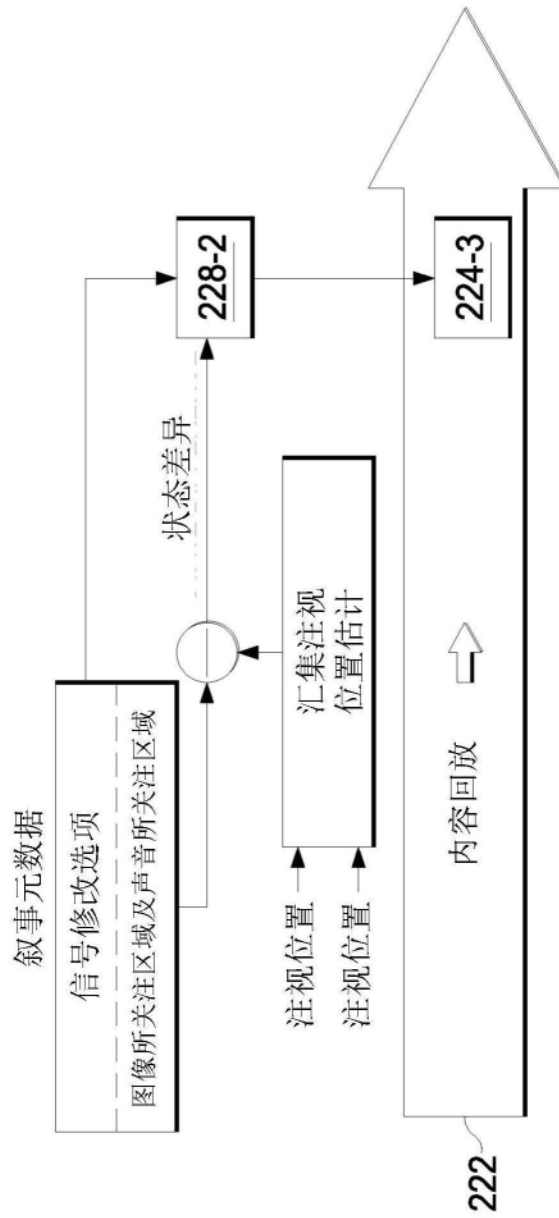


图3E

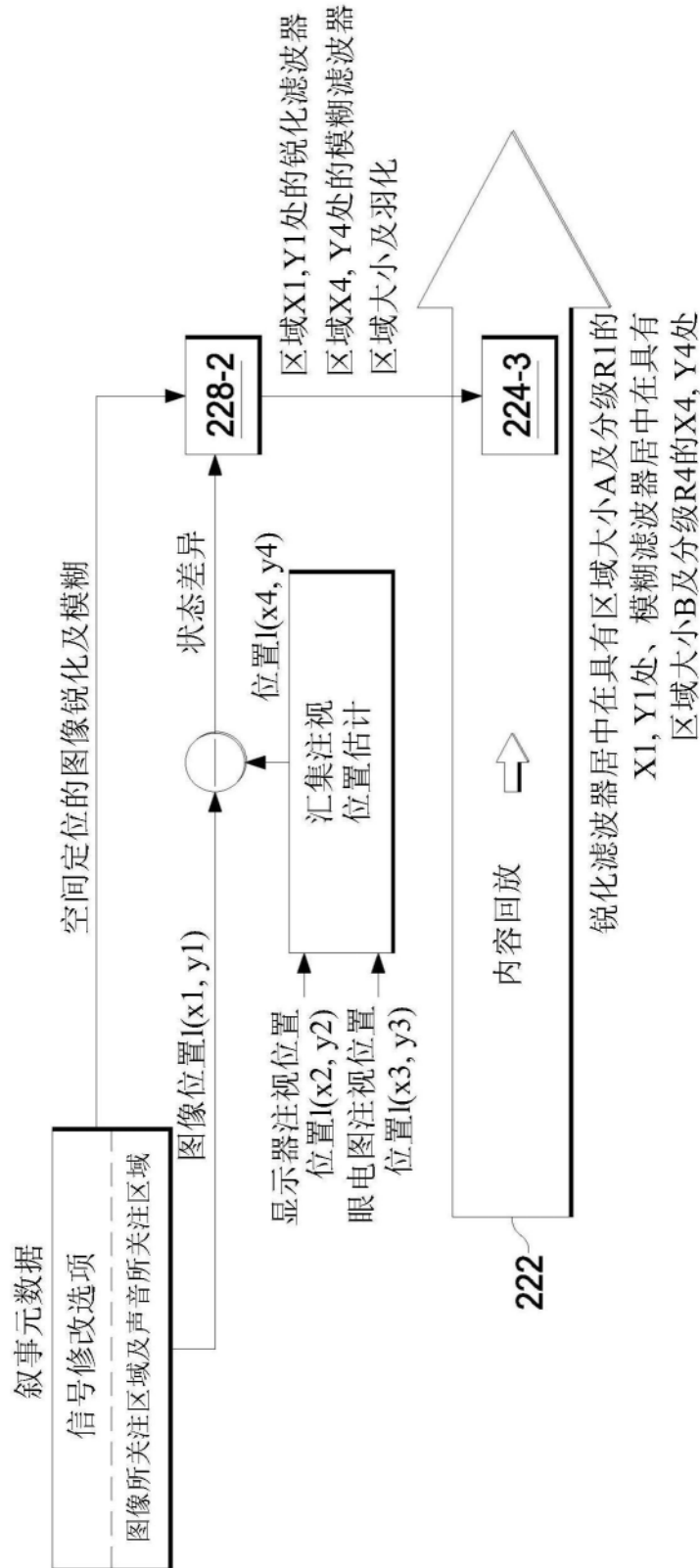


图3F

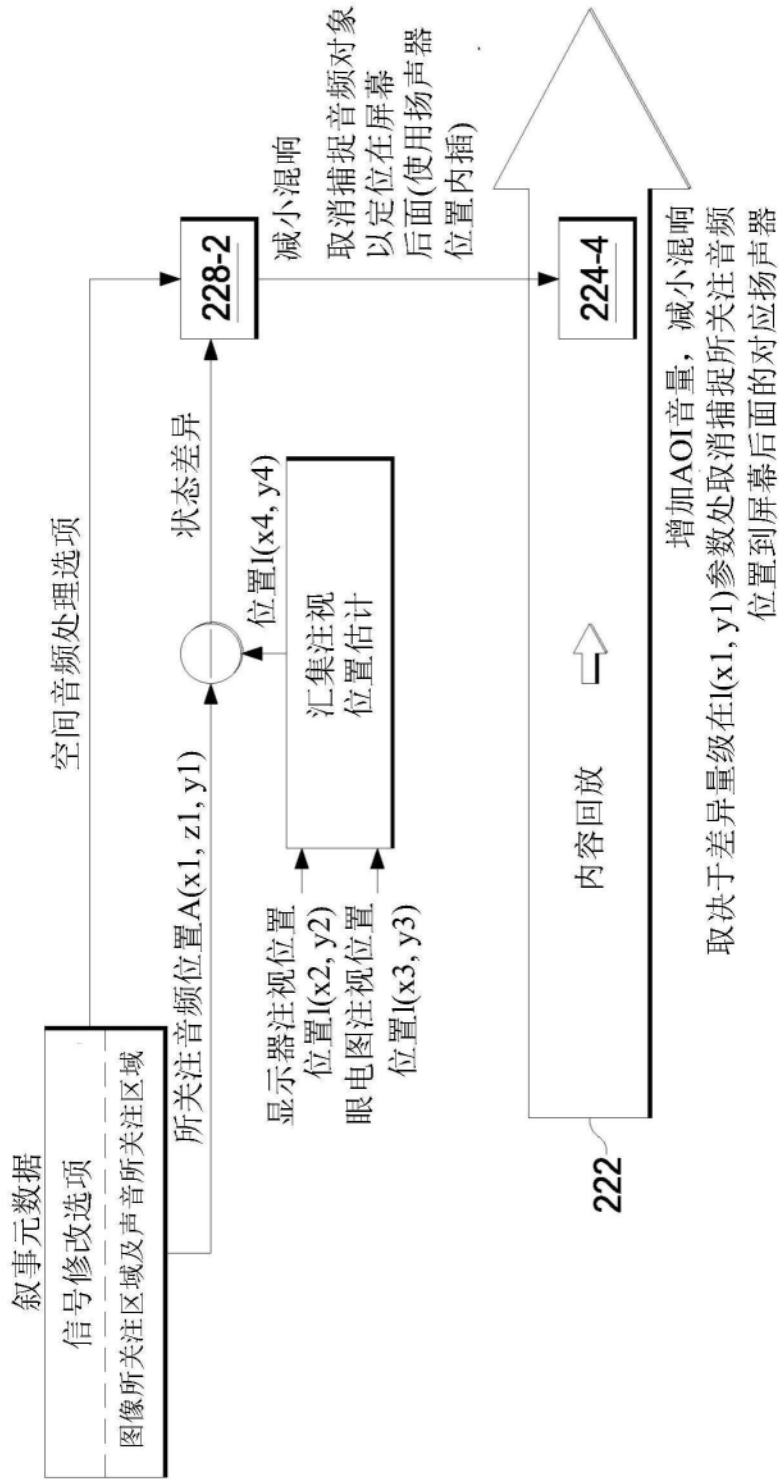


图3G

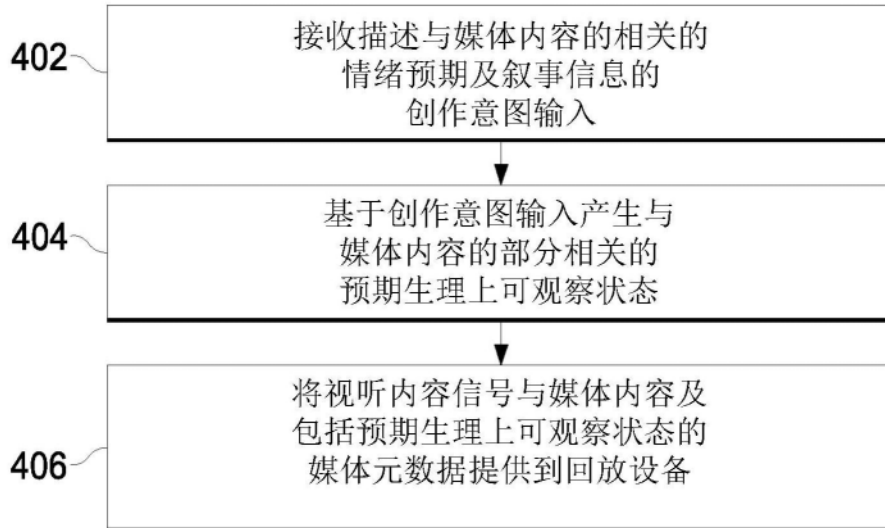


图4A

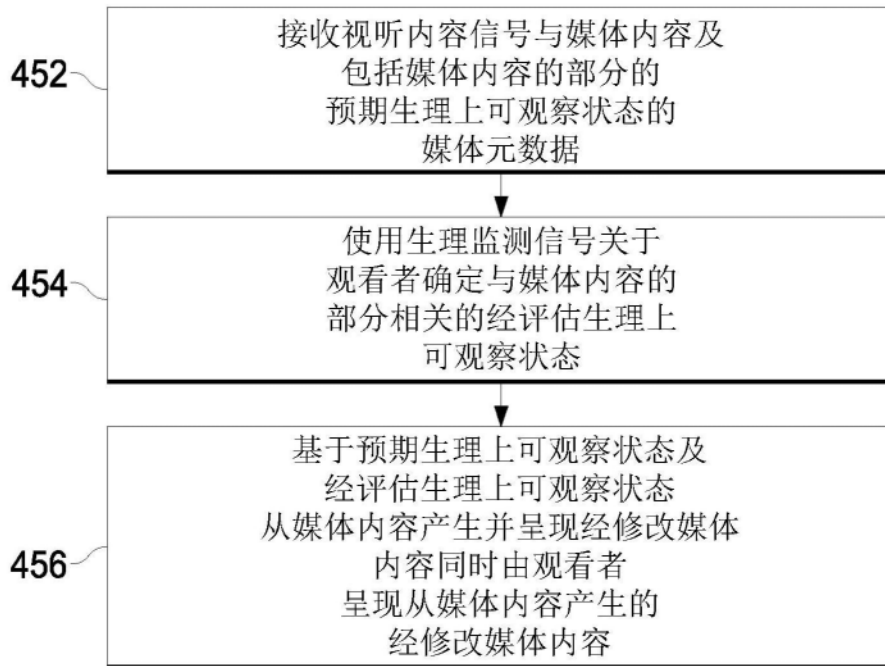


图4B

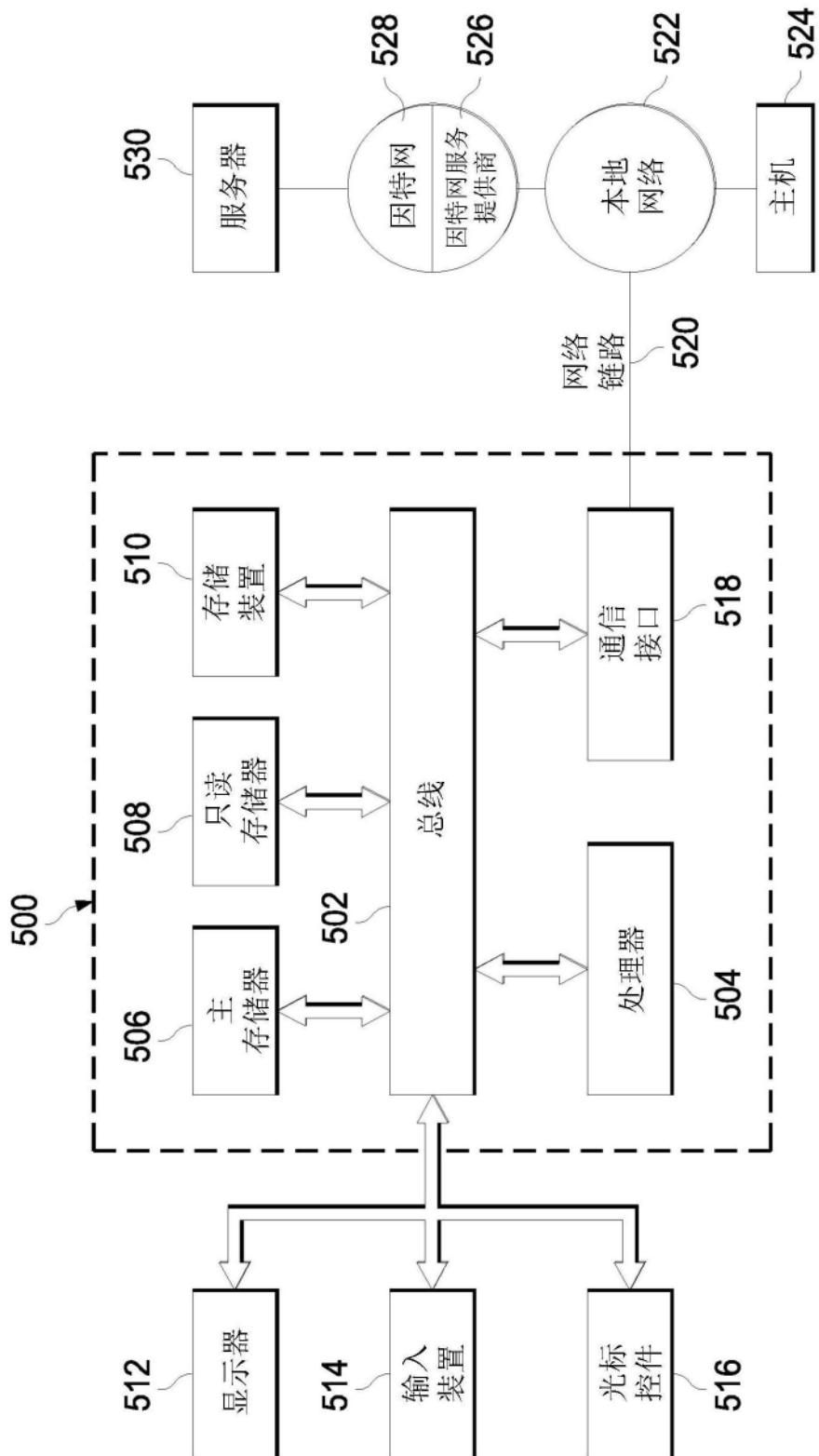


图5