



(12) 发明专利

(10) 授权公告号 CN 111787986 B

(45) 授权公告日 2024. 08. 13

(21) 申请号 201980016107.6

(22) 申请日 2019.02.26

(65) 同一申请的已公布的文献号  
申请公布号 CN 111787986 A

(43) 申请公布日 2020.10.16

(30) 优先权数据  
15/908,603 2018.02.28 US  
16/033,111 2018.07.11 US

(85) PCT国际申请进入国家阶段日  
2020.08.28

(86) PCT国际申请的申请数据  
PCT/US2019/019554 2019.02.26

(87) PCT国际申请的公布数据  
W02019/168834 EN 2019.09.06

(73) 专利权人 苹果公司  
地址 美国加利福尼亚

(72) 发明人 S·A·拉普拉沙德

C·M·阿文达诺 A·M·林达尔

(74) 专利代理机构 中国贸促会专利商标事务所  
有限公司 11038

专利代理师 邹丹

(51) Int.Cl.  
A63F 13/213 (2006.01)  
A63F 13/60 (2006.01)  
A63F 13/215 (2006.01)  
A63F 13/424 (2006.01)

(56) 对比文件  
CN 104170318 A, 2014.11.26  
CN 105797374 A, 2016.07.27  
CN 106575500 A, 2017.04.19  
CN 107742515 A, 2018.02.27

审查员 杨翰

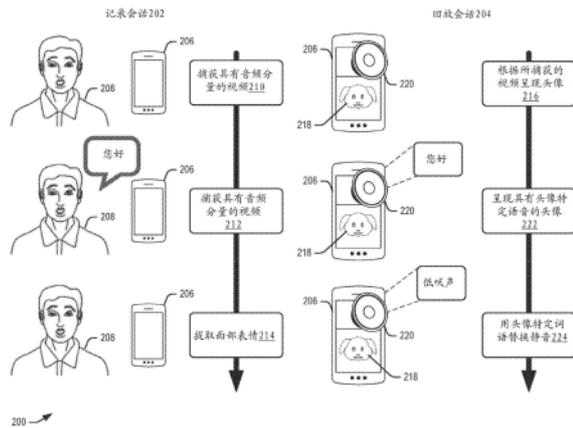
权利要求书3页 说明书19页 附图9页

(54) 发明名称

基于面部表情的语音效果

(57) 摘要

本公开的实施方案可提供用于至少部分地基于从硬件部件提取的面部特征和/或语音特征特性来调整视频剪辑的音频和/或视频信息的技术。例如,响应于检测到生成虚拟头像的头像视频剪辑的请求,可捕获与相机的视场中的面部相关联的视频信号,以及音频信号。语音特征特性和面部特征特性可分别从所述音频信号和所述视频信号中提取。在一些示例中,响应于检测到预览所述头像视频剪辑的请求,可至少部分地基于所述面部特征特性和所述语音特征特性来生成经调整的音频信号,并且可使用所述经调整的音频信号显示所述虚拟头像的所述视频剪辑的预览。



1. 一种用于实现基于面部表情的语音效果的方法,包括:
  - 在至少具有相机和麦克风的电子设备处:
  - 显示虚拟头像生成界面;
  - 接收对对应于多个头像类型中的头像类型的虚拟头像的选择;
  - 在所述虚拟头像生成界面中显示对应于所述头像类型的所述虚拟头像的第一预览内容,所述虚拟头像的所述第一预览内容对应于所述相机的视场中的用户头部特写的实时预览视频帧以及相关联的头部特写外观变化;
  - 在显示所述虚拟头像的所述第一预览内容时,检测所述虚拟头像生成界面中的输入;
  - 响应于检测到所述虚拟头像生成界面中的所述输入:
  - 在记录会话期间,经由所述相机捕获与所述用户头部特写相关联的视频信号;
  - 在所述记录会话期间,经由所述麦克风捕获用户音频信号;以及
  - 从所捕获的视频信号中提取与面部相关联的面部特征特性;并且
  - 响应于检测到所述记录会话到期:
  - 至少部分地基于所述面部特征特性来识别面部表情触发器;
  - 确定所识别的面部表情触发器是否具有用于对应于所选的虚拟头像的所述头像类型的对应的头像特定词语或声音;以及
  - 根据确定所识别的面部表情触发器具有用于对应于所选的虚拟头像的所述头像类型的对应的音频效果:
  - 检索用于所述头像类型的所述对应的头像特定词语或声音;
  - 确定所识别的面部表情触发器的持续时间;
  - 至少部分地基于所述持续时间,确定所述头像类型的对应的头像特定词语或声音的数量或者转换头像特定词语或声音用于插入经调整的音频信号中;
  - 通过调整所捕获的音频信号以及在所捕获的音频信号的一部分上插入对应于所述面部表情触发器的所述数量的对应的头像特定词语或声音或者转换的头像特定词语或声音,生成所述经调整的音频信号;
  - 根据所述面部特征特性和所述经调整的音频信号,在所述虚拟头像生成界面中生成所述虚拟头像的第二预览内容;以及
  - 在所述虚拟头像生成界面中呈现所述第二预览内容。
2. 根据权利要求1所述的方法,还包括存储与从所述视频信号中提取的所述面部特征特性相关联的面部特征元数据并且存储与从所述音频信号中提取的音频特征特性相关联的音频元数据。
3. 根据权利要求2所述的方法,还包括至少部分地基于所述面部特征特性和所述音频特征特性,从所述面部特征元数据生成经调整的面部特征元数据。
4. 根据权利要求3所述的方法,其中进一步根据所述经调整的面部特征元数据来显示所述虚拟头像的所述第二预览内容。
5. 一种用于实现基于面部表情的语音效果的电子设备,包括:
  - 相机;
  - 麦克风;以及
  - 一个或多个处理器,所述一个或多个处理器与所述相机和所述麦克风通信,所述一个

或多个处理器被配置为：

在显示虚拟头像的第一预览时，检测虚拟头像生成界面中的输入，并且接收对对应于多个头像类型中的头像类型的虚拟头像的选择，显示的所述虚拟头像对应于所述头像类型；

响应于检测到所述虚拟头像生成界面中的所述输入，发起包括以下操作的捕获会话：

经由所述相机捕获与所述相机的视场中的面部相关联的视频信号；

经由所述麦克风捕获与所捕获的视频信号相关联的音频信号；以及

从所捕获的视频信号中提取与面部相关联的面部特征特性；并且

响应于检测到所述捕获会话到期：

至少部分地基于所述面部特征特性来识别面部表情触发器；

确定所识别的面部表情触发器是否具有用于对应于所选的虚拟头像的所述头像类型的对应的头像特定词语或声音；以及

根据确定所识别的面部表情触发器具有用于对应于所选的虚拟头像的所述头像类型的对应的音频效果：

检索用于所述头像类型的所述对应的头像特定词语或声音；

确定所识别的面部表情触发器的持续时间；

至少部分地基于所述持续时间，确定所述头像类型的对应的头像特定词语或声音的数量或者转换头像特定词语或声音用于插入经调整的音频信号中；

通过调整所捕获的音频信号以及在所捕获的音频信号的一部分上插入对应于所述面部表情触发器的所述数量的对应的头像特定词语或声音或者转换的头像特定词语或声音，生成所述经调整的音频信号；以及

根据所述面部特征特性和所述经调整的音频信号，在所述虚拟头像生成界面中显示所述虚拟头像的第二预览。

6. 根据权利要求5所述的电子设备，其中至少部分地基于所述虚拟头像的头像类型来进一步调整所述音频信号。

7. 根据权利要求6所述的电子设备，其中至少部分地基于在所述虚拟头像生成界面中呈现的头像类型选择示能表示来接收所述虚拟头像的所述头像类型。

8. 根据权利要求6所述的电子设备，其中所述虚拟头像的所述头像类型包括动物类型，并且其中至少部分地基于与所述动物类型相关联的预先确定的声音来生成所述经调整的音频信号。

9. 根据权利要求5所述的电子设备，其中所述一个或多个处理器被进一步配置为，确定所述音频信号的一部分是否对应于所述视场中的所述面部。

10. 根据权利要求9所述的电子设备，其中所述一个或多个处理器被进一步配置为，根据确定所述音频信号的所述部分对应于所述面部，存储所述音频信号的所述部分以用于生成所述经调整的音频信号。

11. 根据权利要求9所述的电子设备，其中所述一个或多个处理器被进一步配置为，根据确定所述音频信号的所述部分不对应于所述面部，至少移除所述音频信号的所述部分。

12. 根据权利要求5所述的电子设备，其中音频特征特性包括与所述视场中的所述面部相关联的语音的特征。

13. 根据权利要求12所述的电子设备,其中所述一个或多个处理器被进一步配置为,存储与从所述视频信号中提取的所述面部特征特性相关联的面部特征元数据。

14. 根据权利要求13所述的电子设备,其中所述一个或多个处理器被进一步配置为,至少部分地基于所述面部特征特性和所述音频特征特性来生成经调整的面部特征元数据。

15. 根据权利要求14所述的电子设备,其中根据所述经调整的面部特征元数据和所述经调整的音频信号来生成所述虚拟头像的所述第二预览。

16. 一种存储计算机可执行指令的计算机可读存储介质,所述计算机可执行指令当由一个或多个处理器执行时,将所述一个或多个处理器配置为执行包括以下各项的操作:

响应于检测到生成虚拟头像的头像视频剪辑的请求:

接收对对应于多个头像类型中的头像类型的所述虚拟头像的选择;

经由电子设备的相机来捕获与所述相机的视场中的面部相关联的视频信号;

经由所述电子设备的麦克风来捕获音频信号;以及

从所捕获的视频信号中提取与面部相关联的面部特征特性;并且

响应于检测到预览所述头像视频剪辑的请求:

至少部分地基于所述面部特征特性来识别面部表情触发器;

确定所识别的面部表情触发器是否具有用于对应于所选的虚拟头像的所述头像类型的对应的头像特定词语或声音;以及

根据确定所识别的面部表情触发器具有用于对应于所选的虚拟头像的所述头像类型的对应的音频效果:

检索用于所述头像类型的所述对应的头像特定词语或声音;

确定所识别的面部表情触发器的持续时间;

至少部分地基于所述持续时间,确定所述头像类型的对应的头像特定词语或声音的数量或者转换头像特定词语或声音用于插入经调整的音频信号中;

通过调整所捕获的音频信号以及在所捕获的音频信号的一部分上插入对应于所述面部表情触发器的所述数量的对应的头像特定词语或声音或者转换的头像特定词语或声音,生成所述经调整的音频信号;以及

使用所述经调整的音频信号显示所述虚拟头像的所述视频剪辑的预览。

17. 根据权利要求16所述的计算机可读存储介质,其中至少部分地基于在与所述面部相关联的所述面部特征特性中识别的面部表情来调整所述音频信号。

18. 根据权利要求16所述的计算机可读存储介质,其中至少部分地基于与所述面部相关联的音量、音高、持续时间、可变回放速度、言语频谱格式位置、言语频谱格式级别、瞬时回放速度或语音变化来调整所述音频信号。

19. 根据权利要求16-18中任一项所述的计算机可读存储介质,其中通过插入一个或多个预先存储的音频样本来进一步调整所述经调整的音频信号。

20. 根据权利要求16所述的计算机可读存储介质,其中所述一个或多个处理器被进一步配置为,执行包括将所述虚拟头像的所述视频剪辑传输到另一个电子设备的操作。

21. 一种计算机程序产品,包括计算机指令,所述计算机指令在由电子设备的处理器执行时,使得所述电子设备执行根据权利要求1-4中任一项所述的方法。

## 基于面部表情的语音效果

[0001] 相关申请的交叉引用

[0002] 本申请要求于2018年2月28日提交的标题为“Voice Effects Based on Facial Expressions”的美国非临时专利申请15/908,603和于2018年7月11日提交的标题为“Techniques for Providing Audio and Video Effects”的美国部分继续专利申请16/033,111的权益,这些公开内容据此全文以引用方式并入以用于所有目。

### 背景技术

[0003] 多媒体内容(诸如,表情符号)可作为消息传送通信的一部分来发送。表情符号可表示多种预定义的人、物体、动作和/或其他事物。一些消息传送应用程序允许用户从预定义的表情符号库中进行选择,这些表情符号可作为可包含其他内容(例如,其他多媒体和/或文本内容)的消息的一部分来发送。动画表情是这种其他多媒体内容中的一种类型,在动画表情中用户可选择头像(例如,玩偶)来表示其自身。动画表情可以移动,可以说话,就像用户的视频一样。动画表情使得用户能够以有趣、富有创意的方式创建个性化的表情符号版本。

### 发明内容

[0004] 本公开的实施方案可提供用于实现头像视频剪辑修改和回放技术的系统、方法和计算机可读介质。在一些示例中,计算设备可呈现用于跟踪用户面部并且呈现虚拟头像表示(例如,用户面部的玩偶或视频角色版本)的用户界面(UI)。在识别要记录的请求时,计算设备可捕获音频和视频信息,提取和检测语境以及面部特征特性和语音特征特性,至少部分地基于提取/识别的特征修改音频和/或视频信息,并且使用已修改的音频和/或视频信息来呈现头像的视频剪辑。

[0005] 在一些实施方案中,可提供用于实现各种音频和视频效果技术的计算机实现的方法。该方法包括:显示虚拟头像生成界面。该方法还可包括:在虚拟头像生成界面中显示虚拟头像的第一预览内容,该虚拟头像的第一预览内容对应于相机的视场中的用户头部特写的实时预览视频帧以及相关联的头部特写外观变化。该方法还可包括:在显示该虚拟头像的第一预览内容时,检测虚拟头像生成界面中的输入。在一些示例中,响应于检测到虚拟头像生成界面中的输入,该方法还可包括:在记录会话期间,经由相机捕获与用户头部特写相关联的视频信号;在记录会话期间,经由麦克风捕获用户音频信号;从所捕获的用户音频信号中提取音频特征特性;以及从所捕获的视频信号中提取与面部相关联的面部特征特性。另外,响应于检测到记录会话到期:该方法还可包括:至少部分地基于面部特征特性和音频特征特性,从所捕获的音频信号生成经调整的音频信号;根据面部特征特性和经调整的音频信号,在虚拟头像生成界面中生成虚拟头像的第二预览内容;以及在虚拟头像生成界面中呈现第二预览内容。

[0006] 在一些实施方案中,该方法还可包括:存储与从视频信号中提取的面部特征特性相关联的面部特征元数据,以及至少部分地基于面部特征特性和音频特征特性,根据面部

特征元数据生成经调整的面部特征元数据。另外,可进一步根据经调整的面部元数据来显示虚拟头像的第二预览。在一些示例中,可根据依照在预览会话期间的面部外观变化识别的预览面部特征特性来显示虚拟头像的第一预览。

[0007] 在一些实施方案中,可提供用于实现各种音频和视频效果技术的电子设备。该系统可包括相机、麦克风、预先记录/预先确定的音频库,以及与相机和麦克风通信的一个或多个处理器。在一些示例中,处理器可被配置为执行计算机可执行指令以执行操作。操作可包括:在显示虚拟头像的第一预览时,检测虚拟头像生成界面中的输入。操作还可包括:发起捕获会话,该捕获会话包括响应于检测到虚拟头像生成界面中的输入。捕获会话可包括:经由相机捕获与相机的视场中的面部相关联的视频信号;经由麦克风捕获与所捕获的视频信号相关联的音频信号;从所捕获的音频信号中提取音频特征特性;以及从所捕获的视频信号中提取与面部相关联的面部特征特性。在一些示例中,操作还可包括:至少部分地基于音频特征特性和面部特征特性来生成经调整的音频信号;以及至少响应于检测到捕获会话的到期,在虚拟头像生成界面中呈现第二预览内容。

[0008] 在一些情况下,可至少部分地基于虚拟头像的类型来进一步调整音频信号。另外,可至少部分地基于在虚拟头像生成界面中呈现的头像类型选择示能表示来接收虚拟头像的类型。在一些情况下,虚拟头像的类型可包括动物类型,并且可至少部分地基于与动物类型相关联的预先确定的声音来生成经调整的音频信号。预先确定的声音的使用和定时可基于来自所捕获的音频的音频特征和/或来自所捕获的视频的面部特征。该预先确定的声音本身也可基于来自所捕获的音频的音频特征和来自所捕获的视频的面部特征来进行修改。在一些示例中,该一个或多个处理器可被进一步配置为确定音频信号的一部分是否对应于视场中的面部。另外,根据确定音频信号的该部分对应于面部,可存储音频信号的该部分以用于生成经调整的音频信号,并且/或者根据确定音频信号的该部分不对应于面部,音频信号的至少该部分可被移除并且不考虑修改和/或回放该部分。另外,音频特征特性可包括与视场中的面部相关联的语音的特征。在一些示例中,该一个或多个处理器可被进一步配置为存储与从视频信号中提取的面部特征特性相关联的面部特征元数据。在一些示例中,该一个或多个处理器可被进一步配置为存储与从音频信号中提取的音频特征特性相关联的音频特征元数据。此外,该一个或多个处理器可被进一步配置为至少部分地基于面部特征特性和音频特征特性来生成经调整的面部元数据,并且虚拟头像的第二预览可根据经调整的面部元数据和经调整的音频信号来生成。

[0009] 在一些实施方案中,可提供计算机可读介质。该计算机可读介质可包括计算机可执行指令,该计算机可执行指令当由一个或多个处理器执行时,使得该一个或多个处理器执行操作。这些操作可包括响应于检测到生成虚拟头像的头像视频剪辑的请求而执行以下动作:经由电子设备的相机捕获与相机的视场中的面部相关联的视频信号;经由电子设备的麦克风捕获音频信号;从所捕获的音频信号中提取语音特征特性;以及从所捕获的视频信号中提取与面部相关联的面部特征特性。这些操作还可包括响应于检测到预览头像视频剪辑的请求而执行以下动作:至少部分地基于面部特征特性和语音特征特性来生成经调整的音频信号,并且使用经调整的音频信号显示虚拟头像的视频剪辑的预览。

[0010] 在一些实施方案中,可至少部分地基于在与面部相关联的面部特征特性中识别的面部表情来调整音频信号。在一些情况下,可至少部分地基于与面部相关联的音量、音高、

持续时间、格式或语音特性的变化来调整音频信号。此外,在一些实施方案中,该一个或多个处理器可被进一步配置为执行包括将虚拟头像的视频剪辑传输到另一个电子设备的操作。

[0011] 以下具体实施方式连同附图将提供对本公开的实质和优点的更好的理解。

### 附图说明

[0012] 图1是根据至少一个示例的简化框图,该框图示出了用于提供如本文所述的音频和/或视频效果技术的示例性流程。

[0013] 图2是根据至少一个示例的另一个简化框图,该框图示出了用于提供如本文所述的音频和/或视频效果技术的示例性流程。

[0014] 图3是根据至少一个示例的另一个简化框图,该框图示出了用于提供如本文所述的音频和/或视频效果技术的硬件和软件组件。

[0015] 图4是根据至少一个示例的示出了提供如本文所述的音频和/或视频效果技术的流程图。

[0016] 图5是根据至少一个示例的示出了提供如本文所述的音频和/或视频效果技术的另一个流程图。

[0017] 图6是根据至少一个示例的简化框图,该框图示出了用于提供如本文所述的音频和/或视频效果技术的用户界面。

[0018] 图7是根据至少一个示例的示出了提供如本文所述的音频和/或视频效果技术的另一个流程图。

[0019] 图8是根据至少一个示例的示出了提供如本文所述的音频和/或视频效果技术的另一个流程图。

[0020] 图9是根据至少一个示例的简化框图,该框图示出了用于提供如本文所述的音频和/或视频效果技术的计算机架构。

### 具体实施方式

[0021] 本公开的某些实施方案涉及用于实现至少部分地基于面部表情来提供语音效果(例如,已修改的音频)的各种技术的设备、计算机可读介质和方法。另外,在一些情况下,该各种技术还可至少部分地基于记录的音频特性来提供视频效果。更进一步地,该各种技术还可至少部分地基于记录的面部表情和音频特性中的一者或两者来提供语音效果和视频效果(例如,两者一起提供)。在一些示例中,语音效果和/或视频效果可呈现在用户界面(UI)中,该UI被配置为显示用户的卡通表示(例如,头像或数字玩偶)。表示用户的这种头像可被认为是动画表情,因为这种头像可能看起来像大多数智能电话用户熟悉的表情符号角色;不过,这种头像可以动画方式模拟用户的实际运动。

[0022] 例如,可向计算设备的用户呈现用于生成动画表情视频(例如,视频剪辑)的UI。视频剪辑可能受限于预先确定的时间量(例如,10秒、30秒等),或者视频剪辑可为不受限的。在UI中,预览区域可使用头像角色来向用户呈现其面部的实时表示。可提供各种头像角色,并且用户甚至可以能够生成或导入他们自己的头像。预览区域可被配置为提供头像的初始预览和记录的视频剪辑的预览。另外,记录的视频剪辑可(例如,在没有任何视频或音频效

果的情况下)以其原始形式预览,或者可在具有音频和/或视频效果的情况下预览。在一些情况下,用户可在已记录初始视频剪辑之后选择头像。然后,根据需要,可将视频剪辑预览的头像从一个改成另一个,并且可对其应用相同或不同的视频效果。例如,如果用户正在查看未加工的预览(例如,没有效果的原始形式),但是用户切换了头像角色,则UI可被更新以显示视频剪辑相同但采用了新选择的头像的渲染。换句话说讲,在记录期间捕获的面部特征和音频(例如,用户的语音)可由任何头像(例如,在没有任何效果的情况下)呈现。在预览中,看起来将像是头像角色在以用户在记录期间相同的移动方式在移动,并且在说用户在记录期间所说的内容。

[0023] 举例而言,用户可经由UI选择第一头像(例如,独角兽头部)或者可使用最初提供的默认头像。UI将在预览区域中呈现头像(在该示例中,如果用户选择了卡通独角兽的头部,则为卡通独角兽的头部;如果用户未选择,则为默认的任何其他可用玩偶),并且设备将(例如,使用一个或多个麦克风和/或一个或多个相机)开始捕获音频和/或视频信息。在一些情况下,初始预览屏幕仅需要视频信息。可分析视频信息,并且可提取面部特征。然后可将这些提取的面部特征实时映射到独角兽面部,使得独角兽头部的初始预览看起来像是用户头部的镜像。在一些情况下,术语“实时”用于指示响应于用户的每个运动而执行提取、映射、渲染和呈现的结果并且这些结果可基本上立即呈现。对于用户而言,看起来将像是他们正在照镜子,不同的是他们面部的图像被替换成了头像。

[0024] 当用户面部在设备的相机的视线(例如,视野)中时,UI将继续呈现初始预览。在选择UI上的记录示能表示(例如,虚拟按钮)后,设备可开始捕获具有音频分量的视频。在一些示例中,这包括相机捕获帧以及麦克风捕获音频信息。可利用也能够捕获三维(3D)信息的特殊相机。另外,在一些示例中,可利用能够捕获视频的任何相机。视频可以其原始形式存储并且/或者可存储与该视频相关联的元数据。因此,捕获视频和/或音频信息可不同于存储信息。例如,捕获信息可包括感测信息并且至少将其高速缓存,使得其可用于处理。经处理的数据也可被高速缓存,直到确定是否要存储该数据还是简单地利用该数据。例如,在初始预览期间,当用户的面部被实时呈现为玩偶时,视频数据(例如,与数据相关联的元数据)可被高速缓存,同时将该视频数据映射到该玩偶并且进行呈现。然而,该数据可能根本无法永久存储,使得初始预览不可重复使用或不可恢复。

[0025] 另选地,在一些示例中,一旦用户选择UI的记录示能表示,视频数据和音频数据就可被更永久地存储。这样,音频和视频(A/V)数据可被分析、处理等,以便提供本文所述的音频和视频效果。在一些示例中,可处理视频数据以提取面部特征(例如,面部特征特性),并且可将那些面部特征存储为元数据以用于动画表情视频剪辑。元数据集可用指示与视频剪辑相关联的时间、日期和用户的标识符(ID)来存储。另外,音频数据可用相同或其他ID来存储。一旦被存储,或在一些示例中,在存储之前,系统(例如,设备的处理器)可从音频数据中提取音频特征特性并且从视频文件中提取面部特征特性。该信息可用于识别用户的语境、关键词、意图和/或情绪,并且视频和音频效果可在渲染玩偶之前被引入到音频和视频数据中。在一些示例中,可至少部分地基于提取的特征来将音频信号调整为包括不同的词语、声音、音调、音高、定时等。另外,在一些示例中,还可以调整视频数据(例如,元数据)。在一些示例中,在预览本身期间实时提取音频特征。这些音频特征可以是头像特定的,仅在正在预览相关联的头像时才生成。音频特征可以是头像不可知的,针对所有头像生成。音频信号还

可基于这些实时音频特征提取以及在记录过程期间或之后但在预览之前创建的预先存储的提取视频特征来部分地调整。

[0026] 一旦至少部分地基于提取的特性来调整了视频和音频数据,就可渲染玩偶的第二预览。可针对每个可能的玩偶执行该渲染,诸如用户滚动经过并且选择不同的玩偶,这些玩偶已根据经调整的数据渲染了。或者可在选择每个玩偶之后执行渲染。在任何情况下,一旦用户选择玩偶,就可呈现第二预览。第二预览将回放由用户记录的视频剪辑,但该视频剪辑具有经调整的音频和/或视频。使用上面的示例,如果用户以愤怒的音调(例如,以粗哑的语音和皱起的眉头)来记录自己,则可检测到愤怒的语境或意图,并且音频文件可被调整为包括咆哮的声音。因此,第二预览将看起来像是独角兽在说用户所说的词语;然而,用户的语音可被调整为听起来像是咆哮的声音,或使音调更中音(例如,更低)。然后,用户可保存第二预览,或者选择第二预览以用于(例如,通过消息传送应用程序等)传输到另一个用户。在一些示例中,下方和上方动画表情视频剪辑可被共享为.mov文件。然而,在其他示例中,所述技术可实时使用(例如,与视频消息传送等一起使用)。

[0027] 图1是示出用于至少部分地基于在用户的记录中检测到的音频和/或视频特征来提供音频和/或视频效果的示例性流程100的简化框图。在示例性流程100中,存在两个单独的会话:记录会话102和回放会话104。在框110处,在记录会话102中,设备106可捕获具有用户108的音频分量的视频。在一些示例中,可使用两个不同的设备(例如,麦克风和相机)来分别捕获(例如,收集)视频和音频。可至少部分地基于用户108对记录示能表示的选择来触发视频和音频的捕获。在一些示例中,在框112处,用户108可说出词语“您好”。另外,在框112处,设备106可继续捕获用户动作的视频和/或音频分量。在框114处,设备106可继续捕获视频和音频分量,并且在该示例中,用户108可说出词语“狗叫”。在框114处,设备106还可从音频信息中提取说出的词语。然而,在其他示例中,说出的词语提取(或任何音频特征提取)实际上可在记录会话102完成之后进行。在其他示例中,说出的词语提取(或任何音频特征提取)实际上可在预览块124期间实时进行。也可在记录会话102仍在进行中时完成实时提取(例如,音频的分析)。在任一种情况下,由设备106执行的头像过程可通过提取用户说出的词语“狗叫”来识别,并且可采用一些逻辑来确定要实现什么音频效果。

[0028] 以举例的方式,记录会话102可在用户108再次选择记录示能表示(例如,指示希望结束记录)时、选择结束记录示能表示(例如,记录示能表示可在记录时充当结束记录示能表示)时或者至少部分地基于时间段(例如,10秒、30秒等)到期而结束。在一些情况下,该时间段可以是自动预先确定的,而在其他情况下,其可以是用户选择的(例如,从选项列表中选择或通过文本输入界面以自由形式输入)。一旦记录已完成,用户108就可选择预览示能表示,指示用户108希望观看记录的预览。一个选项可以是在没有任何视觉或音频效果的情况下播放原始记录。然而,另一个选项可以是播放视频剪辑的已修改版本。至少部分地基于检测到说出词语“狗叫”,头像过程可能已修改了视频剪辑的音频和/或视频。

[0029] 在框116处,设备106可在屏幕上呈现头像118(也被称为玩偶和/或动画表情)。设备106还可配置有扬声器120,该扬声器可播放与视频剪辑相关联的音频。在该示例中,框116对应于与框110相同的时间点,其中用户108可能已将其嘴部张开,但尚未说话。因此,头像118可呈现为其嘴部张开;然而,还没有呈现来自扬声器120的音频。在框122处,对应于用户108说出“您好”的框112,头像过程可为头像118呈现头像特定的语音。换句话讲,在框122

处,可使用预定义的狗语音来说出词语“您好”。狗语音词语“您好”可由扬声器120呈现。如将在下文进一步详细描述,有多种不同的动物(和其他角色)头像可供用户108选择。在一些示例中,每个头像可与最适合该头像的特定预定义语音相关联。例如,狗可具有狗语音,猫可具有猫语音,猪可具有猪语音,并且机器人可具有机器人语音。这些头像特定的语音可以是预先记录的,或者可与可通过对原始声音执行数学运算而发生的特定频率或音频转换相关联,使得任何用户的语音可转换为听起来像狗语音的声音。然而,每个用户的狗语音至少部分地基于执行的特定音频转换而可能听起来不同。

[0030] 在框124处,头像过程可用头像特定词语替换说出的词语(例如,“狗叫”)。在该示例中,可将狗叫(例如,记录的或模拟的狗叫)的声音插入音频数据中(例如,代替词语“狗叫”),使得当在视频剪辑的呈现期间回放该声音时,扬声器120呈现“低吠声”。在一些示例中,将至少部分地基于不同的头像选择在124处呈现不同的头像特定词语,并且在其他示例中,无论头像选择如何,都可呈现相同的头像特定词语。例如,如果用户108说出“狗叫”,则可当选择狗头像时呈现“低吠声”。然而,在这种相同情况下,如果用户108稍后按相同的流程选择猫头像,则存在用于修改音频的几个选项。在一个示例中,该过程可将“狗叫”转换为“低吠声”,即使“低吠声”对于猫来说不适合。在不同的示例中,该过程可至少部分地基于选择猫头像来将“狗叫”转换为记录或模拟的“猫叫”。并且,在又一示例中,对于不是狗头像的头像,该过程可忽略“狗叫”。因此,即使在114处提取之后,也可执行第二级音频特征分析。视频和音频特征还可影响对头像特定话语的处理。例如,用户说出“狗叫”的音量和音高以及语调可作为音频特征提取的一部分被检测到,并且这可指引系统在预览过程之前和/或期间选择特定的“低吠声”样本或变换这种样本。

[0031] 图2是示出用于至少部分地基于在用户的记录中检测到的音频和/或视频特征来提供音频和/或视频效果的示例性流程200的另一个简化框图。在非常类似于图1的示例性流程100的示例性流程200中,存在两个单独的会话:记录会话202和回放会话204。在框210处,在记录会话202中,设备206可捕获具有用户208的音频分量的视频。可至少部分地基于用户208对记录示能表示的选择来触发视频和音频的捕获。在一些示例中,在框212处,用户208可说出词语“您好”。另外,在框212处,设备206可继续捕获用户动作的视频和/或音频分量。在框214处,设备206可继续捕获视频和音频分量,并且在该示例中,用户208可保持其嘴部张开,但不说任何内容。在框214处,设备206还可从视频中提取面部表情。然而,在其他示例中,面部特征提取(或任何视频特征提取)可实际上在记录会话202完成之后进行。仍然可在记录会话202仍在进行中时完成实时提取(例如,视频的分析)。在任一种情况下,由设备206执行的头像过程可通过提取用户(例如,在不说任何内容的情况下)短暂地张开其嘴部来识别,并且可采用一些逻辑来确定要实现什么音频和/或视频效果。在一些示例中,确定用户在不说任何内容的情况下保持其嘴部张开可能需要提取和分析音频和视频两者。例如,面部特征特性(例如,嘴部张开)的提取可能不够,并且该过程还可能检测用户208在记录的相同时间段期间没有说任何内容。视频和音频特征还可影响对头像特定话语的处理。例如,张开嘴部、张开眼睛等的持续时间可指引系统在预览过程之前和/或期间选择特定的“低吠声”样本或变换这种样本。一种这样的变换为改变低吠声的音量和/或持续时间,以匹配检测到的用户嘴部的张开和闭合。

[0032] 以举例的方式,记录会话202可在用户208再次选择记录示能表示(例如,指示希望

结束记录)时、选择结束记录示能表示(例如,记录示能表示可在记录时充当结束记录示能表示)时或者至少部分地基于时间段(例如,20秒、30秒等)到期而结束。一旦记录已完成,用户208就可选择预览示能表示,指示用户208希望观看记录的预览。一个选项可以是在没有任何视觉或音频效果的情况下播放原始记录。然而,另一个选项可以是播放记录的已修改版本。至少部分地基于检测到面部表情(例如,张开的嘴部),头像过程可能已修改了视频剪辑的音频和/或视频。

[0033] 在框216处,设备206可在设备206的屏幕上呈现头像218(也被称为玩偶和/或动画表情)。设备206还可配置有扬声器220,该扬声器可播放与视频剪辑相关联的音频。在该示例中,框216对应于与框210相同的时间点,其中用户208可能尚未讲话。因此,头像218可呈现为其嘴部张开;然而,还没有呈现来自扬声器220的音频。在框222处,对应于用户208说出“您好”的框212,头像过程可为头像218呈现头像特定的语音(如上所述)。

[0034] 在框224处,头像过程可用头像特定词语替换在框214处识别的静音。在该示例中,可将狗叫(例如,记录的或模拟的狗叫)的声音插入音频数据中(例如,代替静音),使得当在视频剪辑的呈现期间回放该声音时,扬声器220呈现“低吠声”。在一些示例中,将至少部分地基于不同的头像选择在224处呈现不同的头像特定词语,并且在其他示例中,无论头像选择如何,都可呈现相同的头像特定词语。例如,如果用户208保持其嘴部张开,则当选择了狗头像时可呈现“低吠声”,当选择了猫头像时可呈现“猫叫声”声音等。在一些情况下,当检测到用户208已保持其嘴部张开一时间量(例如,半秒、整秒等)而不讲话时,每个头像均可具有要播放的预定义声音。然而,在一些示例中,对于对该面部特征不具有预定义效果的头像,该过程可忽略检测到张开的嘴部。另外,即使在214处提取之后,也可执行第二级音频特征分析。例如,如果该过程(例如,基于检测到张开的嘴部)确定要针对狗头像插入“低吠声”,则该过程还可检测要插入多少“低吠声”声音(例如,如果用户保持其嘴部张开两倍用于指示狗叫的时间长度)或者插入请求的狗叫声(例如,在图1的场景中,其中用户将说出“狗叫”以指示应当插入“低吠声”声音)的数量是否是不可能的。因此,基于上述两个示例,应当显而易见的是,用户208可根据其面部表情和语音(例如,记录的头像消息)来控制回放的效果。此外,虽然在图1或图2中未明确示出,但是用户设备可配置有用于执行头像过程(例如,捕获A/V信息、提取特征、分析数据、实现逻辑、修改音频和/或视频文件以及渲染预览)的软件以及用于执行应用程序(例如,具有其自身UI的头像应用程序)的软件,该应用程序使得用户能够构建头像消息并且随后将这些头像消息发送到其他用户设备。

[0035] 图3是示出由上文和下文所述的头像过程利用的部件(例如,软件模块)的简化框图300。在一些示例中,可利用更多或更少的模块来至少部分地基于在用户的记录中检测到的音频和/或视频特征来实现音频和/或视频效果的提供。在一些示例中,设备302可配置有相机304、麦克风306和用于呈现UI和头像预览(例如,记录之前的初始预览以及发送之前记录的预览)的显示屏。在一些示例中,头像过程配置有头像引擎308和语音引擎310。头像引擎308可管理头像的列表、处理视频特征(例如,面部特征特性)、修改视频信息、在适当时与语音引擎301通信并且在所有处理完成并且效果已实现(或移除)时渲染头像的视频312。视频信息的修改可包括调整或以其他方式编辑与视频文件相关联的元数据。这样,当(经调整或未经调整的)视频元数据用于渲染玩偶时,面部特征可被映射到该玩偶。在一些示例中,语音引擎310可存储音频信息、执行用于确定要实现何种效果的逻辑、修改音频信息并且在

所有处理完成并且效果已实现(或移除)时提供已修改的音频314。

[0036] 在一些示例中,一旦用户选择记录新的头像视频剪辑,视频特征316就可由相机304捕获并且音频特征318可以由麦克风306捕获。在一些情况下,在视频特征316内可检测到多达(或多于)五十个面部特征。示例性视频特征包括但不限于表情持续时间、张开的嘴部、皱眉、微笑、眉毛向上或皱起等。另外,视频特征316可仅包括识别面部特征中的每一个面部特征的元数据(例如,指示用户面部上哪些位置移动了或在什么位置的数据点)。此外,视频特征316可被传递到头像引擎308和语音引擎310。在头像引擎308处,可存储和分析与视频特征316相关联的元数据。在一些示例中,头像引擎308可在存储元数据之前执行从视频文件的特征提取。然而,在其他示例中,可在将视频特征316发送到头像引擎之前执行特征提取(在这种情况下,视频特征316将是元数据本身)。在语音引擎310处,可将视频特征316与音频特征318进行比较,以帮助匹配哪些音频特征去对应于哪些视频特征(例如,查看某个音频特征和视频特征是否同时发生)。

[0037] 在一些情况下,音频特征也被传递到语音引擎310以供存储。示例性音频特征包括但不限于音量、音高、动态(例如,音量、音高、语音、共振峰、持续时间等的变化)。未加工的音频320在其被捕获时包括未处理的音频文件。未加工的音频320可被传递到语音引擎310以用于进一步处理和潜在的(例如,最终的)修改,并且该未加工的音频也可被单独地存储,使得如果需要可使用原始音频。未加工的音频320也可被传递到语音识别模块322。语音识别模块322可用于认出词语并且从用户的语音中识别他们的意图。例如,语音识别模块322可确定用户何时愤怒、悲伤、快乐等。另外,当用户说出关键词(例如,如上所述的“狗叫”)时,语音识别模块322将检测到该关键词。然后,由语音识别模块322检测和/或收集的信息可被传递到语音引擎310以用于进一步的逻辑和/或处理。如上所述,在一些示例中,在预览本身期间实时提取音频特征。这些音频特征可以是头像特定的,仅在正在预览相关联的头像时才生成。音频特征可以是头像不可知的,针对所有头像生成。音频信号还可基于这些实时音频特征提取以及在记录过程期间或之后但在预览之前创建的预先存储的提取视频特征来部分地调整。另外,一些特征提取可在336处由语音引擎310渲染期间执行。语音引擎310可根据需要使用一些预先存储的声音338来填充空白或替换提取的其他声音。

[0038] 在一些示例中,语音引擎310将做出关于如何处理从语音识别模块322中提取的信息的确定。在一些示例中,语音引擎310可将信息从语音识别模块322传递到特征模块324,以用于确定哪些特征对应于由语音识别模块322提取的数据。例如,特征模块324可(例如,基于一组规则和/或逻辑)指示由语音识别模块322检测到的悲伤语音对应于语音音高的升高,或语音速度或节奏的减慢。换句话讲,特征模块324可将提取的音频特征映射到特定语音特征。然后,效果类型模块326可将特定语音特征映射到期望的效果。语音引擎310还可负责为每个可能的头像存储每个特定语音。例如,对于每个头像可存在标准或硬编码的语音。在不进行任何其他改变的情况下,如果用户选择特定头像,则语音引擎310可选择适当的标准语音以与回放一起使用。在这种情况下,已修改的音频314可仅为基于所选择的头像转换为适当的头像语音的未加工的音频320。当用户滚动经过头像并且选择不同头像时,语音引擎310可在运行中修改未加工的音频320以使其听起来像新选择的头像的声音。因此,需要将头像类型328提供给语音引擎310以进行这种改变。然而,如果要提供效果(例如,要在音频文件内改变音高、音调或实际词语),则语音引擎310可修改未加工的音频文件320并且提

供已修改的音频314。在一些示例中,将向用户提供在打开/关闭330处使用原始音频文件的选项。如果用户选择“关闭”(例如,关闭效果),则未加工的音频320可与头像的视频312(例如,对应于未改变的视频)组合来进行A/V输出332。A/V输出332可被提供给在设备302的UI上呈现的头像应用程序。

[0039] 头像引擎308可负责至少部分地基于头像类型328的选择来提供初始头像图像。另外,头像引擎308负责将视频特征316映射到每个头像的适当面部标记。例如,如果视频特征316指示用户正在微笑,则指示微笑的元数据可被映射到所选择的头像的嘴部区域,使得该头像看起来正在头像的视频312中微笑。另外,头像引擎308可根据需要从语音引擎接收定时变化334。例如,如果语音引擎310(例如,基于特征模块324和/或效果类型326和/或头像类型)确定语音效果是使音频更多地为低声语音,并且将语音修改为更多地为低声语音,则该效果改变可除了降低音量和其他共振峰以及音高的改变之外,还包括减慢语音本身。因此,语音引擎可产生已修改的音频,该已修改的音频的回放速度相对于音频剪辑的原始音频文件较慢。在这种场景中,语音引擎310将根据定时变化334指示头像引擎308,使得可适当地减慢视频文件;否则,视频和音频将不会同步。

[0040] 如上所述,用户可使用设备302的头像应用程序来选择不同的头像。在一些示例中,语音效果可至少部分地基于该选择而改变。然而,在其他示例中,可为用户提供为给定头像选择不同语音(例如,为狗头像选择猫语音等)的机会。这种类型的自由形式的语音效果改变可由用户通过在UI上进行选择来执行,或者在一些情况下通过语音激活或面部运动来执行。例如,某个面部表情可触发语音引擎310改变给定头像的语音效果。此外,在一些示例中,语音引擎310可被配置为使儿童的语音听起来音高更高,或者另选地,确定不使儿童的语音听起来音高更高,因为在儿童语音的未加工的音频320可能已经是音高较高的情况下,该语音听起来将是不适当的。进行这种用户特定的效果确定可部分地由提取的音频特征驱动,并且在这种情况下,此类特征可包括整个记录中的音高值和范围。

[0041] 在一些示例中,语音识别模块322可包括识别引擎、词语检测仪、音高分析仪和/或共振峰分析仪。语音识别模块322执行的分析将能够识别用户是否沮丧、愤怒、快乐等。另外,语音识别模块322可能够识别用户语音的语境和/或语调,以及改变措辞的意图和/或确定用户的资料(例如,虚拟身份)。

[0042] 在一些示例中,头像过程300可被配置为通过将头像的视频312和已修改的音频314或未加工的音频320组合成A/V输出332来打包/渲染视频剪辑。为了打包这两者,语音引擎310仅需要知道与头像的视频312相关联的元数据的ID(例如,该语音引擎实际上并不需要头像的视频312,而仅需要元数据的ID)。消息传送应用程序(例如,头像应用程序)内的消息可被传输到其他计算设备,其中该消息包括A/V输出332。当用户选择UI中的“发送”示能表示时,可发送要预览的最后一个视频剪辑。例如,如果用户用狗头像来预览他们的视频剪辑,然后切换到猫头像以进行预览,则当用户选择“发送”时将发送猫头像视频。另外,最后预览的状态可存储并且在稍后使用。例如,如果发送的最后一条消息(例如,头像视频剪辑)使用了特定效果,则正在生成的下一条消息的第一预览可利用该特定效果。

[0043] 由语音引擎310和/或头像引擎308实现的逻辑可检查某些提示和/或特征,然后修改音频和/或视频文件以实现期望的效果。一些示例性特征/效果对包括:检测到用户已张开嘴部并且暂停了一会。在该示例中,需要同时发生面部特征特性(例如,嘴部张开)和音频

特征特性(例如,静音)两者,以便实现期望的效果。对于该特征/效果对,修改音频和视频以使得头像呈现发出头像/动物特定声音的期望效果。例如,对于该角色/动物,狗将发出狗叫的声音,猫将发出猫叫的声音,猴子、马、独角兽等将发出适当的声音。其他示例性特征/效果对包括当检测到皱眉时降低音频音高和/或音调。在该示例中,仅需要检测视频特征特性。然而,在一些示例中,该效果可至少部分地基于语音识别模块322检测到用户语音中的悲伤来实现。在这种情况下,根本不需要视频特征316。其他示例性特征/效果对包括低声以使得音频和视频速度减慢、语调降低和/或改变减少。在一些情况下,视频改变可导致音频的修改,而在其他情况下,音频改变可导致视频的修改。

[0044] 如上所述,在一些示例中,头像引擎308可充当特征提取器,在这种情况下,视频特征316和音频特征318在被发送到头像引擎308之前可能不存在。相反,未加工的音频320和与该未加工的视频相关联的元数据可被传递到头像引擎308中,其中头像引擎308可提取音频特征特性和视频(例如,面部)特征特性。换句话讲,虽然在图3中没有以这种方式绘制,但是头像引擎308的部分实际上可存在于相机304内。另外,在一些示例中,与视频特征316相关联的元数据可存储在安全容器中,并且当语音引擎310正在运行时,该语音引擎可从该容器读取元数据。

[0045] 在一些情况下,因为头像的预览视频剪辑未实时显示(例如,该预览视频剪辑在记录了视频之后并且有时仅响应于播放示能表示的选择而渲染和显示),所以音频和视频信息可离线处理(例如,不是实时处理)。这样,头像引擎308和语音引擎310可提前读取音频和视频信息中,并且预先做出语境决定。然后,语音引擎310可相应地修改音频文件。这种离线地预先读取和做出决定的能力将极大地提高系统的效率,特别是对于更长的记录。此外,这使得能够进行第二阶段的分析,其中可处理附加逻辑。因此,可在做出任何最终决定之前分析整个音频文件。例如,如果用户连续说出“狗叫”两次,但所说的词语“狗叫”太靠近在一起,则预先记录的实际“低吠声”声音可能无法拟合用户说出“狗叫,狗叫”所花费的时间。在这种情况下,语音引擎310可从语音识别322获取信息并且确定忽略第二个“狗叫”,因为将不可能在音频文件中包括两个“低吠声”声音。

[0046] 如上所述,当音频文件和视频被打包在一起以进行A/V输出332时,语音引擎实际上并不需要访问头像的视频312。相反,在播放视频时,通过访问写入元数据文件的特征阵列(例如,浮点值)来创建视频文件(例如,.mov格式文件等)。然而,对音频和视频文件的所有排列/调整可预先进行,并且一些排列/调整甚至可在提取音频和视频时实时完成。另外,在一些示例中,可临时保存(例如,高速缓存)每个已修改的视频剪辑,使得如果用户重新选择已预览过的头像,则不需要重复进行生成/渲染该特定预览的处理。与每次在预览部分期间选择相同头像时重新渲染的已修改视频剪辑相反,已渲染的视频剪辑的上述高速缓存将使得能够实现大幅节省处理器功率和每秒的指令(IPS),特别是对于较长的记录和/或具有大量效果的记录。

[0047] 另外,在一些示例中,可采用噪声抑制算法以用于处理由麦克风306捕获的声音包括除用户语音之外的声音的情况。例如,当用户在多风地区或吵闹的房间(例如,餐厅或酒吧)中时。在这些示例中,噪声抑制算法可降低音频记录的某些部分的分贝输出。另选地或除此之外,可分离不同的语音并且/或者可仅收集来自某些视角(例如,用户面部的角度)的音频,并且可忽略或抑制其他语音。在其他情况下,如果头像过程300确定噪声水平太高或

将难以处理,则过程300可以禁用记录选项。

[0048] 图4示出了根据至少一些实施方案的示例性流程图,其示出了用于至少部分地基于音频和/或视频特征来实现各种音频和/或视频效果的过程400。在一些示例中,图1的计算设备106或其他类似用户设备(例如,利用至少图3的头像过程300)可执行图4的过程400。

[0049] 在框402处,计算设备106可捕获具有音频分量的视频。在一些示例中,视频和音频可由两个不同的硬件部件捕获(例如,相机可捕获视频信息,而麦克风可捕获音频信息)。然而,在一些情况下,单个硬件部件可被配置为捕获音频和视频两者。在任何情况下,视频和音频信息可(例如,通过共享ID、时间戳等)与彼此相关联。因此,视频可具有音频分量(例如,这两者是同一文件的一部分),或者视频可与音频分量有关(例如,相关联在一起的两个文件)。

[0050] 在框404处,计算设备106可分别从所捕获的视频信息和音频信息中提取面部特征和音频特征。在一些情况下,面部特征信息可经由头像引擎308提取并且存储为元数据。元数据可用于将每个面部特征映射到特定玩偶或映射到任何动画或虚拟面部。因此,不需要存储实际视频文件,从而提高了存储器存储效率并且节省显著。关于音频特征提取,可利用语音识别算法来提取不同的语音特征;例如,词语、短语、音高、速度等。

[0051] 在框406处,计算设备106可从提取的特征检测语境。例如,语境可包括用户的意图、情绪、设定、位置、背景物品、想法等。当采用逻辑来确定要应用什么效果时,语境可能是重要的。在一些情况下,语境可与检测到的说出的词语组合以确定是否和/或如何调整音频文件和/或视频文件。在一个示例中,用户可皱起他的眉毛并且缓慢讲话。皱起的眉毛为可能在框404处已提取的视频特征,并且缓慢的言语为可能在框404处已提取的音频特征。单独地,这两个特征可能意味着不一样的东西;然而,当组合在一起时,头像过程可确定用户所关心的东西。在这种情况下,消息的语境可能是父母正和孩子讲话,或者朋友正和另一个朋友讲严重的或令人担忧的问题。

[0052] 在框408处,计算设备106可至少部分地基于语境来确定用于渲染音频和/或视频文件的效果。如上所述,可能涉及到一个语境。因此,可采用特定的视频和/或音频特征以用于该效果。例如,可将语音文件调整为听起来更忧郁或减慢。在其他示例中,可用原始(例如,未加工的)音频的版本来替换头像特定语音以传达消息的严重性。可采用各种其他效果以用于各种其他语境。在其他示例中,语境可以是动物噪声(例如,基于用户说出“狗叫”或“猫叫”等)。在这种情况下,所确定的效果将用狗叫的声音来替换说出的词语“狗叫”。

[0053] 在框410处,计算设备106可执行附加效果的附加逻辑。例如,如果用户试图通过连续说出狗叫两次来使狗叫效果起作用,则可能需要利用附加逻辑来确定附加的狗叫在技术上是是否可行。例如,如果用于替换未加工的音频信息中说出的词语的狗叫的音频剪辑为0.5秒长,但是用户在0.7秒的跨度内说出“狗叫”两次,则附加逻辑可确定两个狗叫的声音无法在0.7秒可用的时间内拟合。因此,可能需要扩展音频文件和视频文件以便拟合两个狗叫的声音,可能需要(例如,通过处理存储的狗叫的声音)缩短狗叫的声音,或者可能需要忽略第二次说出的词语狗叫。

[0054] 在框412处,计算设备106可至少部分地基于所确定的效果和/或附加效果来修改音频和/或视频信息。在一些示例中,可仅使用一组效果。然而,在任一种情况下,可调整(例如,修改)未加工的音频文件以形成具有添加和/或减去的附加声音的新音频文件。例如,在

“狗叫”使用案例中,说出的词语“狗叫”将从音频文件中移除,并且将插入表示实际狗叫的新声音。新文件可用不同的ID或附加的ID(例如,未加工的音频ID,其具有用于指示其不是原始的.v2标识符)来保存。此外,未加工的音频文件将被单独地保存,使得其可被重复用于附加头像并且/或者用于如果用户决定不使用所确定的效果的情况。

[0055] 在框414处,计算设备106可接收来自用户的头像选择。用户可通过由计算设备106执行的头像应用程序的UI来选择多个不同头像中的一个头像。头像可经由滚轮、下拉菜单或图标菜单(例如,其中每个头像在屏幕上其自身位置中是可见的)来选择。

[0056] 在框416处,计算设备106可至少部分地基于所选择的头像来呈现已修改的音频和已修改的视频。在该示例中,可在用户选择头像之前,为每个相应头像生成每个经调整的视频剪辑(例如,具有经调整的音频和/或经调整的视频的头像的最终剪辑)。这样,处理已完成,并且准备好在选择头像时立即呈现经调整的视频剪辑。虽然这在头像选择之前可能需要附加的IPS,但是该附加的IPS将使呈现速度加快。另外,可在用户正在查看第一预览(例如,对应于UI中呈现的第一/默认头像的预览)时执行对每个经调整的视频剪辑的处理。

[0057] 图5示出了根据至少一些实施方案的示例性流程图,其示出了用于至少部分地基于音频和/或视频特征来实现各种音频和/或视频效果的过程500。在一些示例中,图1的计算设备106或其他类似用户设备(例如,利用至少图3的头像过程300)可执行图5的过程500。

[0058] 在框502处,计算设备106可捕获具有音频分量的视频。正如图4的框402中一样,视频和音频可由两个不同的硬件部件捕获(例如,相机可捕获视频信息,而麦克风可捕获音频信息)。如上所述,视频可具有音频分量(例如,这两者是同一文件的一部分),或者视频可与音频分量有关(例如,相关联在一起的两个文件)。

[0059] 在框504处,计算设备106可分别从所捕获的视频信息和音频信息中提取面部特征和音频特征。正如上文所述,面部特征信息可经由头像引擎308提取并且存储为元数据。元数据可用于将每个面部特征映射到特定玩偶或映射到任何动画或虚拟面部。因此,不需要存储实际视频文件,从而提高了存储器存储效率并且节省显著。关于音频特征提取,可利用语音识别算法来提取不同的语音特征;例如,词语、短语、音高、速度等。另外,在一些示例中,头像引擎308和/或语音引擎310可以执行音频特征提取。

[0060] 在框506处,计算设备106可从提取的特征检测语境。例如,语境可包括用户的意图、情绪、设定、位置、想法、身份等。当采用逻辑来确定要应用什么效果时,语境可能是重要的。在一些情况下,语境可与说出的词语组合以确定是否和/或如何调整音频文件和/或视频文件。在一个示例中,用户的年龄可至少部分地基于面部和/或语音特征被检测为语境(例如,儿童、成人等)。例如,儿童的面部可具有可被识别的特定特征(例如,大眼睛、较小的鼻部和相对较小的头部等)。因此,可检测到儿童语境。

[0061] 在框508处,计算设备106可接收来自用户的头像选择。用户可通过由计算设备106执行的头像应用程序的UI来选择多个不同头像中的一个头像。头像可经由滚轮、下拉菜单或图标菜单(例如,其中每个头像在屏幕上其自身位置中是可见的)来选择。

[0062] 在框510处,计算设备106可至少部分地基于语境和所选择的头像来确定用于渲染音频和/或视频文件的效果。在该示例中,每个头像的效果可在选择每个头像时生成,而不是同时生成所有头像。在一些情况下,这将使得能够实现处理器和存储器的显著节省,因为一次将仅执行一组效果和头像渲染。特别是当用户未选择要预览的多个头像时,可实现这

些节省。

[0063] 在框512处,计算设备106可执行附加效果的附加逻辑,类似于上文相对于图4的框410所述。在框514处,计算设备106可至少部分地基于所选择的头像的所确定的效果和/或附加效果来修改音频和/或视频信息,类似于上文相对于图4的框412所述。在框516处,计算设备106可至少部分地基于所选择的头像来呈现已修改的音频和已修改的视频,类似于上文相对于图4的框416所述。

[0064] 在一些示例中,头像过程300可至少部分地基于历史信息来确定是执行流程400还是执行流程500。例如,如果用户每次通常使用相同的头像,则流程500将更有效。然而,如果用户定期在头像之间切换,并且每个视频剪辑预览多个不同的头像,则遵循流程400可能更有效。

[0065] 图6示出了用于使得用户能够利用头像应用程序(例如,对应于头像应用程序示能表示602)的示例性UI 600。在一些示例中,在选择头像应用程序示能表示602之前,UI 600可看起来不同(例如,它可表现为标准文本(例如,短消息服务(SMS))消息传送应用程序)。如上所述,头像应用程序可与头像过程(例如,图3的头像过程300)通信,以提出捕获、处理(例如,提取特征、运行逻辑等)以及调整音频和/或视频的请求。例如,当用户选择记录示能表示(例如,记录/发送视频剪辑示能表示604)时,头像应用程序可对头像过程进行应用编程接口(API)调用以开始使用适当的硬件部件捕获视频和音频信息。在一些示例中,记录/发送视频剪辑示能表示604可在记录会话开始之前表示为红色圆圈(或没有图6所示的线的普通圆圈)。这样,示能表示将看起来更像标准记录按钮。在记录会话期间,记录/发送视频剪辑示能表示604的外观可改变以看起来像时钟倒计时或定时器的其他表示(例如,如果视频剪辑记录的长度有限)。然而,在其他示例中,记录/发送视频剪辑示能表示604可仅改变颜色以指示头像应用程序正在记录。如果不存在定时器或对记录长度的限制,则用户可能需要再次选择记录/发送视频剪辑示能表示604以终止记录。

[0066] 在一些示例中,用户可使用头像选择示能表示606来选择头像。这可在记录头像视频剪辑之前和/或在记录头像视频剪辑之后进行。当在记录之前选择时,用户的运动和面部特性的初始预览将被呈现为所选择的头像。另外,将在呈现记录的实时预览时执行记录,其中用户的面部由所选择的头像来表示。一旦记录完成,将再次使用所选择的头像来呈现第二预览(例如,实际记录的回放)。然而,在该阶段,用户可滚动经过头像选择示能表示606以选择新的头像来查看记录预览。在一些情况下,在选择新头像时,UI将开始使用所选择的头像来预览记录。新预览可用音频/视频效果来呈现,或者按原始记录的来呈现。如上所述,关于呈现已有效果的版本还是原始版本的确定可至少部分地基于所使用的最后回放的方法。例如,如果最后回放使用了效果,则在新头像选择之后的第一回放可使用效果。然而,如果最后回放不使用效果,则在新头像选择之后的第一回放可不使用效果。在一些示例中,用户可通过选择效果预览示能表示608来重放具有效果的视频剪辑,或者通过选择原始预览示能表示610来重放不具有效果的视频剪辑。一旦满意视频剪辑(例如,消息),用户就可以使用记录/发送视频剪辑示能表示604来将消息中的头像视频发送到另一个计算设备。将使用对应于(例如,具有或不具有效果的)最后预览的格式来发送视频剪辑。在任何时间,如果用户需要,可选择删除视频剪辑示能表示612来删除头像视频,并且开始或退出头像和/或消息传送应用程序。

[0067] 图7示出了根据至少一些实施方案的示例性流程图(例如,计算机实现的方法),其示出了用于至少部分地基于音频和/或视频特征来实现各种音频和/或视频效果的过程700。在一些示例中,图1的计算设备106或其他类似用户设备(例如,至少利用类似于图6所示的头像应用程序和图3的头像过程300)可执行图7的过程700。

[0068] 在框702处,计算设备106可显示虚拟头像生成界面。虚拟头像生成界面可看起来类似于图6所示的UI。然而,可使用被配置为实现本文所述的相同特征的任何UI。

[0069] 在框704处,计算设备106可显示虚拟头像的第一预览内容。在一些示例中,第一预览内容可以是用户面部的实时表示,该表示包括移动和面部表情。然而,第一预览将提供头像(例如,卡通角色、数字/虚拟玩偶)来表示用户的面部而不是用户面部的图像。该第一预览可仅为视频,或者至少是没有声音的头像的渲染。在一些示例中,该第一预览未被记录,并且只要用户需要就可使用,而不受除了计算设备106的电池电力或存储器空间之外的限制。

[0070] 在框706处,计算设备106可检测虚拟头像生成界面中的输入(例如,图6的记录/发送视频剪辑示能表示604)的选择。可在UI正在显示第一预览内容时做出该选择。

[0071] 在框708处,计算设备106可至少部分地基于在框706处检测到的输入来开始捕获视频和音频信号。如上所述,视频和音频信号可由适当的硬件部件捕获,并且可由一个此类部件或此类部件的组合捕获。

[0072] 在框710处,计算设备106可提取音频特征特性和面部特征特性,如上文详细所述。如上所述,提取可以由图3的头像过程300的特定模块或由头像应用程序和/或计算设备106的其他提取和/或分析部件执行。

[0073] 在框712处,计算设备106可至少部分地基于面部特征特性和音频特征特性来生成经调整的音频信号。例如,在框708处捕获到的音频文件可被永久地(或暂时地)修改(例如,调整)以包括新的声音、新的词语等,并且/或者以调整原始音高、音调、音量等。这些调整可至少部分地基于经由面部特征特性和音频特征特性的分析检测到的语境来进行。另外,可基于所选择的头像类型和/或基于用户在记录会话期间执行(例如,由用户的面部表示)的特定运动、面部表情、词语、短语或动作来进行调整。

[0074] 在框714处,计算设备106可根据经调整的音频信号来在UI中生成虚拟头像的第二预览内容。生成的第二预览内容可至少部分地基于当前选择的头像或某个默认头像。在框716处,一旦生成第二预览内容,计算设备106就可在UI中呈现第二预览内容。

[0075] 图8示出了根据至少一些实施方案的示例性流程图,其示出了用于至少部分地基于音频和/或视频特征来实现各种音频和/或视频效果的过程800(例如,存储在计算机可读存储器上可被执行的指令)。在一些示例中,图1的计算设备106或其他类似用户设备(例如,至少利用类似于图6所示的头像应用程序和图3的头像过程300)可执行图8的过程800。

[0076] 在框802处,计算设备106可检测生成虚拟头像的头像视频剪辑的请求。在一些示例中,请求可至少部分地基于用户对图6的发送/记录视频剪辑示能表示604的选择。

[0077] 在框804处,计算设备106可捕获与相机的视场中的面部相关联的视频信号。在框806处,计算设备106可捕获对应于视频信号的(例如,来自相机捕获的面部的)音频信号。

[0078] 在框808处,计算设备106可从音频信号中提取语音特征特性,并且在框810处,计算设备106可从视频信号中提取面部特征特性。

[0079] 在框812处,计算设备106可检测预览头像视频剪辑的请求。该请求可至少部分地基于用户经由图6的头像选择示能表示606对新头像的选择,或者至少部分地基于用户对图6的效果预览示能表示608的选择。

[0080] 在框814处,计算设备106可至少部分地基于面部特征特性和语音特征特性来生成经调整的音频信号。例如,可修改(例如,调整)在框806处捕获到的音频文件以包括新的声音、新的词语等,并且/或者以调整原始音高、音调、音量等。这些调整可至少部分地基于经由面部特征特性和语音特征特性的分析检测到的语境来进行。另外,可基于所选择的头像类型和/或基于用户在记录会话期间执行(例如,由用户的面部表示)的特定运动、面部表情、词语、短语或动作来进行调整。

[0081] 在框816处,计算设备106可根据经调整的音频信号来在UI中生成虚拟头像的预览。生成的预览可至少部分地基于当前选择的头像或某个默认头像。在框816处,一旦生成预览,计算设备106还可在UI中呈现第二预览内容。

[0082] 图9是示出根据至少一个实施方案的用于实现本文所述的特征的示例性架构900的简化框图。在一些示例中,具有示例性架构900的计算设备902(例如,图1的计算设备106)可被配置为呈现相关UI、捕获音频和视频信息、提取相关数据、执行逻辑、修改音频和视频信息并且呈现动画表情视频。

[0083] 计算设备902可被配置为执行或以其他方式管理用于执行所述技术的应用程序或指令,这些技术诸如但不限于提供用于记录、预览和/或发送虚拟头像视频剪辑的用户界面(例如,图6的用户界面600)。计算设备602可在用户界面处(例如,利用I/O设备904诸如触摸屏)从用户接收输入、捕获信息、处理信息,然后也利用I/O设备904(例如,计算设备902的扬声器)将视频剪辑呈现为预览。计算设备902可被配置为至少部分地基于从所捕获的视频中提取的面部特征和/或从所捕获的音频中提取的语音特征来修改音频和/或视频文件。

[0084] 计算设备902可为任何类型的计算设备,诸如但不限于移动电话(例如,智能电话)、平板电脑、个人数字助理(PDA)、膝上型计算机、台式计算机、瘦客户端设备、智能手表、无线耳机等。

[0085] 在一种示例性配置中,计算设备902可包括至少一个存储器914和一个或多个处理单元(或处理器)916。处理器916可视情况而在硬件、计算机可执行指令或它们的组合中实现。处理器916的计算机可执行指令或固件实施方式可包括以任何合适的编程语言编写的计算机可执行指令或机器可执行指令,以执行所描述的各种功能。

[0086] 存储器914可存储能够在处理器916上加载和执行的程序指令以及在执行这些程序过程期间所生成的数据。根据计算设备902的配置和类型不同,存储器914可为易失性存储器(诸如随机存取存储器(RAM))和/或非易失性存储器(诸如只读存储器(ROM)、闪存存储器等)。计算设备902还可包括附加的可移除存储装置和/或不可移除存储装置926,包括但不限于磁存储装置、光盘和/或磁带存储装置。磁盘驱动器及其相关联的非暂态计算机可读介质可为计算设备提供计算机可读指令、数据结构、程序模块及其他数据的非易失性存储装置。在一些实施方式中,存储器914可包括多种不同类型的存储器,诸如静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)或ROM。虽然本文所述的易失性存储器可以被称为RAM,但是从主机和/或电源拔下后将不会保留其中所存储的数据的任何易失性存储器都是适当的。

[0087] 存储器914和附加存储装置926为可移除或不可移除的,均为非暂态计算机可读存储介质的示例。例如,非暂态计算机可读存储介质可包括易失性或非易失性、可移除或不可移除介质,其通过任何方法或技术实现为用于存储信息诸如计算机可读指令、数据结构、程序模块或其他数据。存储器914和附加存储装置926两者均为非暂态计算机存储介质的示例。可存在于计算设备902中的附加类型的计算机存储介质可包括但不限于:相变RAM (PRAM)、SRAM、DRAM、RAM、ROM、电可擦除可编程只读存储器 (EEPROM)、闪存存储器或其他存储器技术、光盘只读存储器 (CD-ROM)、数字视频光盘 (DVD) 或其他光学存储装置、磁带盒、磁带、磁盘存储装置或其他磁性存储设备或任何其他可用于存储所需信息并可被计算设备902访问的介质。上述任意组合也应该包括在非暂态计算机可读存储介质的范围内。

[0088] 另选地,计算机可读通信介质可包括计算机可读指令、程序模块或在数据信号诸如载波或其他传输手段内传输的其他数据。然而,如本文所用,计算机可读存储介质不包括计算机可读通信介质。

[0089] 计算设备902还可包含通信连接件928,该通信连接件允许计算设备902经由一个或多个网络与数据存储库、另一种计算设备或服务器、用户终端和/或其他设备进行通信。此类网络可包括许多不同类型的网络中的任一种或组合,诸如有线网络、互联网、无线网络、蜂窝网络、卫星网络、其他专用网络和/或公共网络或者它们的任意组合。计算设备902还可以包括I/O设备904,诸如触摸输入设备、键盘、鼠标、笔、语音输入设备、显示器、扬声器、打印机等。

[0090] 更详细地转至存储器914的内容,存储器914可包括操作系统932和/或用于实现本文公开的特征的一个或多个应用程序或服务,该一个或多个应用程序或服务包括用户界面模块934、头像控制模块936、头像应用程序模块938以及消息传送模块940。存储器914还可被配置为存储要用于产生音频和视频输出的一个或多个音频文件和视频文件。这样,计算设备902可执行本文所述的所有操作。

[0091] 在一些示例中,用户界面模块934可被配置为管理计算设备902的用户界面。例如,用户界面模块934可呈现由计算设备902请求的任何数量的各种UI。具体地讲,用户界面模块934可被配置为呈现图6的UI 600,这使得能够实现本文所述的特征以及与图3的头像过程300的通信,该头像过程如上所述负责捕获视频和音频信息、提取适当的面部特征和语音特征信息,以及在呈现生成的头像视频剪辑之前修改视频和音频信息。

[0092] 在一些示例中,头像控制模块936被配置为实现头像过程300(例如,执行用于实现头像过程的指令),而头像应用程序模块938被配置为实现面向用户的应用程序。如上所述,头像应用程序模块938可利用一个或多个API来请求和/或向头像控制模块936提供信息。

[0093] 在一些实施方案中,消息传送模块940可实现可与头像控制模块936和/或头像应用程序模块938通信的任何独立或附加消息传送应用程序。在一些示例中,消息传送模块940可与头像应用程序模块938完全集成(例如,如图6的UI 600所示),其中头像应用程序看起来是消息传送应用程序的一部分。然而,在其他示例中,当用户请求生成头像视频剪辑时,消息传送应用程序940可调用头像应用程序模块938,并且头像应用程序模块938可完全打开与消息传送模块940集成的新应用程序。

[0094] 计算设备902还可配备有相机和麦克风,如至少图3所示,并且处理器916可被配置为执行显示虚拟头像的第一预览的指令。在一些示例中,在显示虚拟头像的第一预览时,可

经由用户界面模块934所呈现的虚拟头像生成界面来检测输入。在一些情况下,响应于检测到虚拟头像生成界面中的输入,头像控制模块936可发起包括以下操作的捕获会话:经由相机捕获与该相机的视场中的面部相关联的视频信号;经由麦克风捕获与所捕获的视频信号相关联的音频信号;从所捕获的音频信号中提取音频特征特性;并且从所捕获的视频信号中提取与面部相关联的面部特征特性。另外,响应于检测到捕获会话的到期,头像控制模块936可至少部分地基于音频特征特性和面部特征特性来生成经调整的音频信号,并且根据面部特征特性和经调整的音频信号在虚拟头像生成界面中显示虚拟头像的第二预览。

[0095] 上文描述了用于提供用于至少部分地基于语音和/或面部特征特性来调整音频和/或视频内容的各种技术的例示性方法、计算机可读介质和系统。这些系统、介质和方法中的一些或全部可以但不需要至少部分地由架构和流程(诸如至少在上面的图1至图9中示出的那些架构和流程)来实现。虽然上文参考消息传送应用程序描述了许多实施方案,但应当理解,任何上述技术都可用于任何类型的应用程序(包括实时视频回放或实时视频消息传送应用程序)中。为了解释的目的,阐述了很多具体配置和细节以便提供对示例的彻底理解。但是,对本领域的技术人员也将显而易见的是,一些示例可在没有这些具体细节的情况下被实施。此外,有时省略或简化熟知的特征部以防止对本文所述的示例造成混淆。

[0096] 还可在各种操作环境中实现各种实施方案,在一些情况下,操作环境可包括可用于操作许多应用程序中的任何应用程序的一个或多个用户计算机、计算设备或处理设备。用户设备或客户端设备可包括许多通用个人计算机中的任何一个,诸如运行标准操作系统的台式计算机或膝上型计算机,以及运行移动软件并能够支持多个联网协议和即时消息协议的蜂窝设备、无线设备和手持设备。此系统还可包括运行各种可商购获得的操作系统和用于目的诸如开发和数据库管理等其他已知应用程序中的任何一者的多个工作站。这些设备还可包括其他电子设备,诸如虚拟终端、瘦客户端、游戏系统以及能够经由网络进行通信的其他设备。

[0097] 大多数实施方案利用本领域技术人员熟悉的至少一个网络来支持使用各种商用协议诸如TCP/IP、OSI、FTP、UPnP、NFS、CIFS和AppleTalk中的任何协议的通信。网络可以是例如局域网、广域网、虚拟专用网络、互联网、内联网、外联网、公共交换电话网、红外网络、无线网络及其任何组合。

[0098] 在利用网络服务器的实施方案中,网络服务器可运行各种服务器或中间层应用程序中的任何一者,包括HTTP服务器、FTP服务器、CGI服务器、数据服务器、Java服务器和业务应用程序服务器。一个或多个服务器还可能响应于来自用户设备的请求而执行程序或脚本,诸如通过执行一个或多个应用程序,所述一个或多个应用程序可被实现为以任何编程语言诸如Java<sup>®</sup>、C、C#或C++或者任何脚本语言诸如Perl、Python或TCL以及它们的组合编写的一个或多个脚本或程序。一个或多个服务器还可包括数据库服务器,包括但不限于可从Oracle<sup>®</sup>、Microsoft<sup>®</sup>、Sybase<sup>®</sup>和IBM<sup>®</sup>商购获得的那些。

[0099] 环境可包括各种数据存储库和其他存储器和存储介质,如上所述。这些可驻留在各个位置,诸如在一个或多个计算机本地的存储介质上或者远离网络上的任何或全部计算机的存储介质上(和/或驻留在一个或多个计算机中)。在特定的一组实施方案中,信息可驻留在本领域技术人员熟悉的存储区域网络(SAN)中。类似地,用于执行归属于计算机、服务

器或其他网络设备的功能的任何必要文件可以根据需要本地存储以及/或者远程存储。当系统包括计算机化设备时,每个此类设备可包括可经由总线电耦接的硬件元件,所述元件包括例如至少一个中央处理单元(CPU)、至少一个输入设备(例如,鼠标、键盘、控制器、触摸屏或小键盘),以及至少一个输出设备(例如,显示设备、打印机或扬声器)。此类系统还可包括一个或多个存储设备,诸如磁盘驱动器、光存储设备和固态存储设备诸如RAM或ROM,以及可移除媒体设备、存储卡、闪存卡,等等。

[0100] 此类设备还可包括如上所述的计算机可读存储介质读取器、通信设备(例如,调制解调器、网卡(无线或有线)、红外通信设备等)和工作存储器。计算机可读存储介质读取器可连接至或配置为接收表示远程、本地、固定和/或可移除的存储设备的非暂态计算机可读存储介质,以及用于临时和/或更永久地包含、存储、传输和检索计算机可读信息的存储介质。系统和各种设备通常还将包括位于至少一个工作存储器设备内的多个软件应用程序、模块、服务或其他元件,包括操作系统和应用程序,诸如客户端应用程序或浏览器。应当理解的是,另选实施方案可具有根据上文所述的许多变型形式。例如,还可使用定制硬件,和/或可在硬件、软件(包括便携式软件,诸如小应用程序)或两者中实现特定元件。此外,可使用与其他计算设备诸如网络输入/输出设备的连接。

[0101] 用于包含代码或代码的部分的非暂态存储介质和计算机可读存储介质可包括本领域中已知或使用的任何适当的介质(载波等暂态介质除外),诸如但不限于在任何方法或技术中实现的用于存储信息诸如计算机可读指令、数据结构、程序模块或其他数据的易失性和非易失性、可移除和不可移除的介质,包括RAM、ROM、电可擦除可编程只读存储器(EEPROM)、闪存存储器或其他存储器技术、CD-ROM、DVD或其他光学存储装置、磁带盒、磁带、磁盘存储装置或其他磁性存储设备,或者可用于存储所需信息并且可由系统设备访问的任何其他介质。基于本文提供的公开和教导,本领域的普通技术人员将理解实现各种实施方案的其他方式和/或方法。然而,如上所述,计算机可读存储介质不包括暂态介质诸如载波等。

[0102] 相应地,说明书和附图应被视为具有例示性的而非限制性的意义。然而,显而易见的是,在不脱离权利要求中阐述的本公开的更广泛的实质和范围的情况下,可对其进行各种修改和改变。

[0103] 其他变型形式在本公开的实质内。因此,尽管所公开的技术容易受到各种修改和另选构造的影响,但是其某些例示的实施方案在附图中示出并且已经在上面详细描述。然而,应当理解,并不旨在将本公开限制于所公开的特定形式,相反,其目的在于覆盖落入由所附权利要求所限定的本公开的实质和范围内的所有修改、另选构造和等同形式。

[0104] 在描述所公开的实施方案的上下文中(特别是在下面的权利要求书的上下文中)使用术语“一”、“一个”和“该”以及类似的指示词将被解释为覆盖单数和复数,除非另有说明或与上下文明显矛盾。除非另有说明,否则术语“包含”、“具有”、“包括”和“含有”应被解释为开放式术语(即,意思为“包括但不限于”)。术语“连接”被解释为即使有干预的东西,也被部分或全部地包含在内、附接或接合在一起。短语“基于”应当被理解为开放式的,并且不以任何方式进行限制,并且旨在在适当的情况下被解释或以其他方式理解为“至少部分地基于”。除非本文另有说明,否则本文中对数值范围的叙述仅仅旨在用作单独提及落入该范围内的每个单独值的简单方法,并且每个单独的值被并入说明书中,如同在本文中单独引

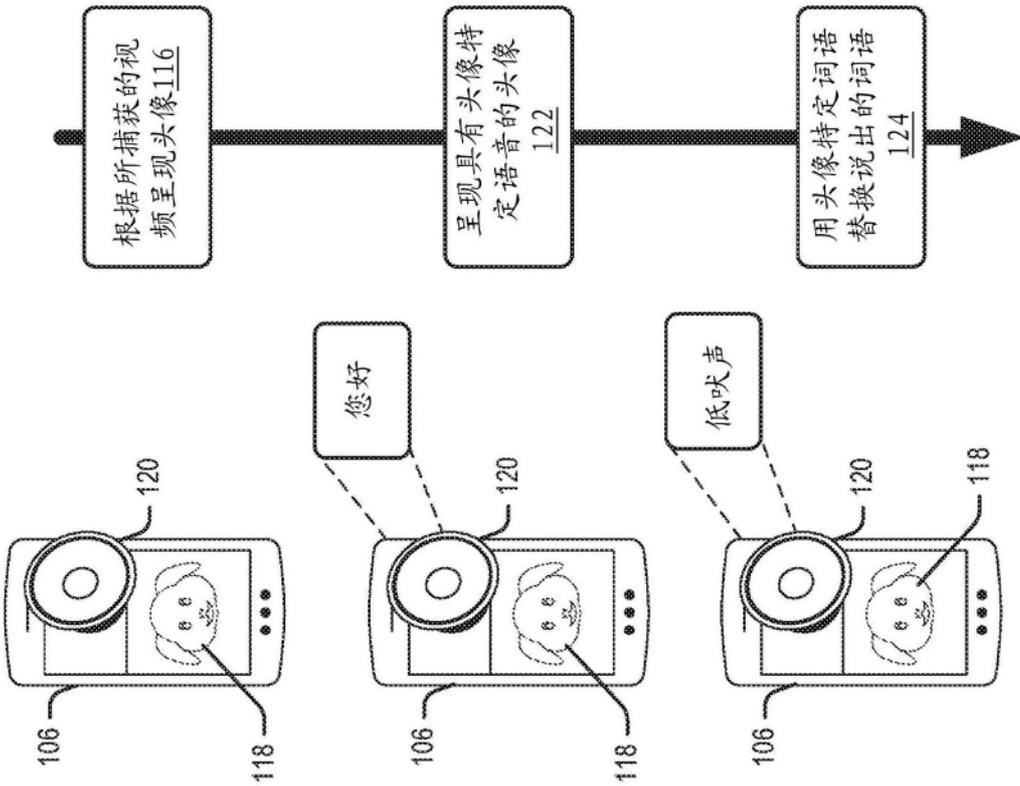
用。本文描述的所有方法能够以任何合适的顺序执行,除非本文另有说明或以其他方式与上下文明显矛盾。除非另有声明,否则本文提供的任何和所有示例或示例性语言(例如,“诸如”)的使用仅仅旨在更好地说明本公开的实施方案,并且不会限制本公开的范围。说明书中的任何语言都不应被解释为指示任何未声明的元素对于本公开的实践是必不可少的。

[0105] 除非另外特别说明,否则析取语言诸如短语“X、Y或Z中的至少一者”在上下文被理解为通常用于呈现项目、术语等,其可以是X、Y或Z,或它们的任何组合(例如,X、Y和/或Z)。因此,此类析取语言通常不旨在并且不应该暗示某些实施方案要求X中的至少一个、Y中的至少一个或者Z中的至少一个均各自存在。另外,除非另外特别说明,否则诸如短语“X、Y和Z中的至少一者”的联合语言也应理解为意指X、Y、Z或它们的任何组合,包括“X、Y和/或Z”。

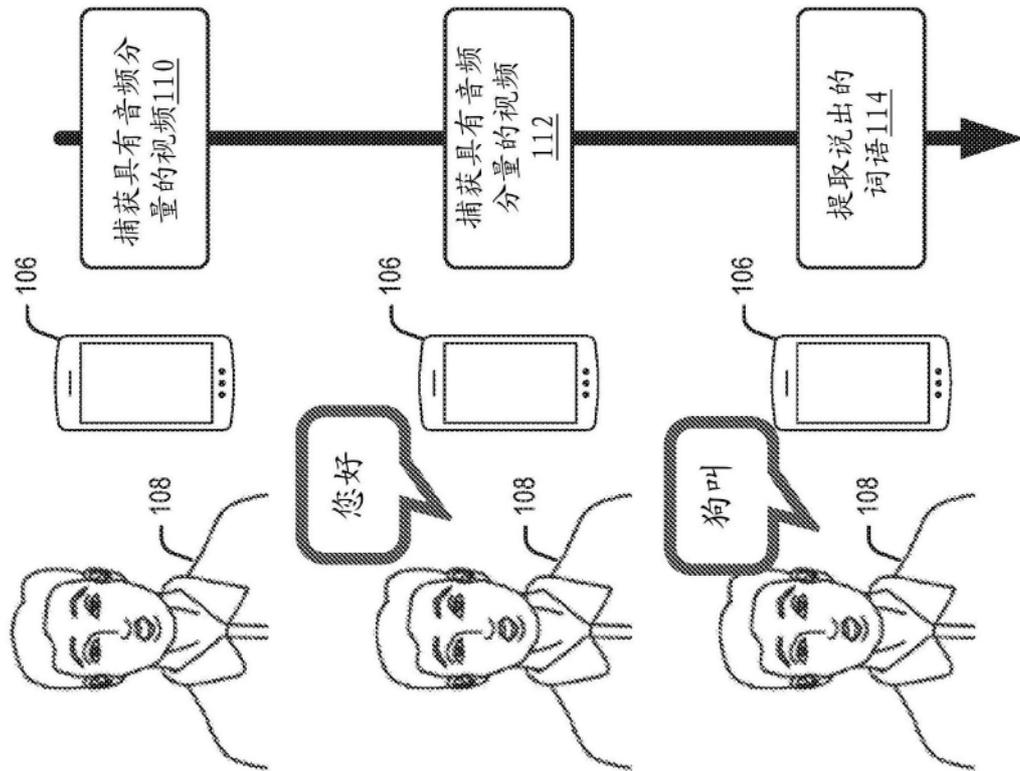
[0106] 本文描述了本公开的优选实施方案,包括发明人已知的用于执行本公开的最佳模式。在阅读前面的描述之后,那些优选实施方案的变型形式对于本领域的普通技术人员来说可变得显而易见。发明人期望技术人员适当地采用此类变型形式,并且发明人旨在以不同于本文具体描述的方式来实践本公开。因此,如适用法律所允许的,本公开包括所附权利要求中记载的主题的所有修改和等同形式。此外,除非在本文中另外指出或者明显与上下文矛盾,否则本公开包含上述元素的所有可能变型形式的任何组合。

[0107] 本文引用的所有参考文献,包括出版物、专利申请和专利,均据此以引用方式并入本文,正如每篇参考文献被单独且具体地指示为以引用方式并入并且在本文全文阐述。

回放会话104



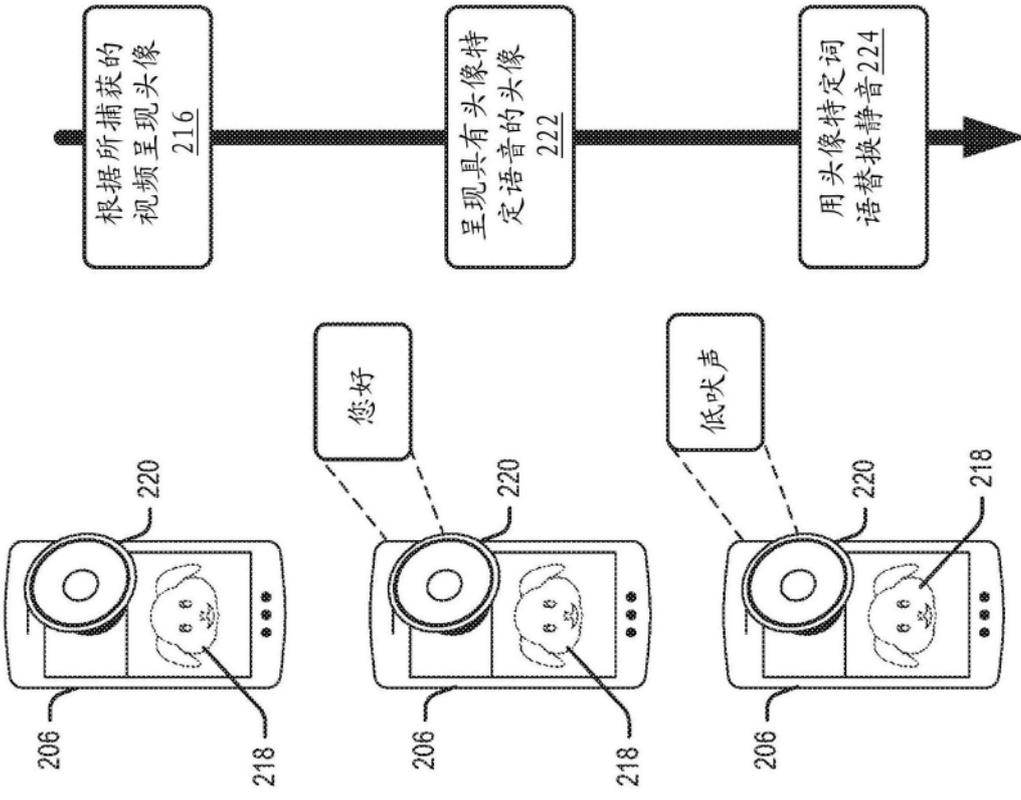
记录会话102



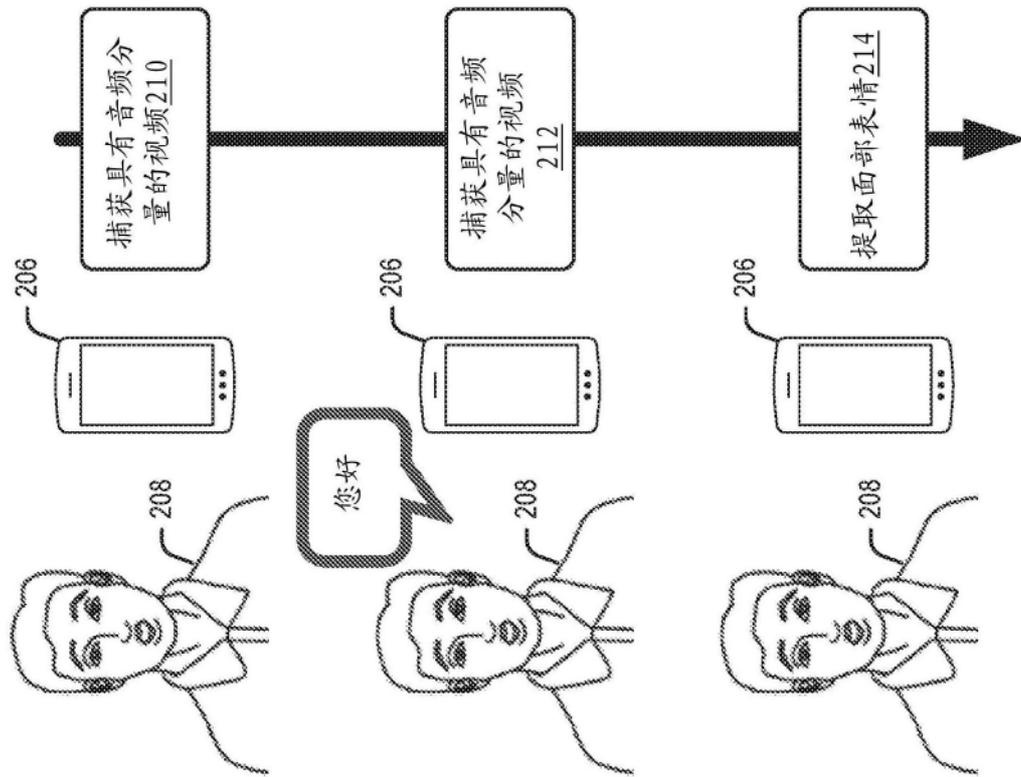
100

图1

回放会话 204



记录会话 202



200

图2

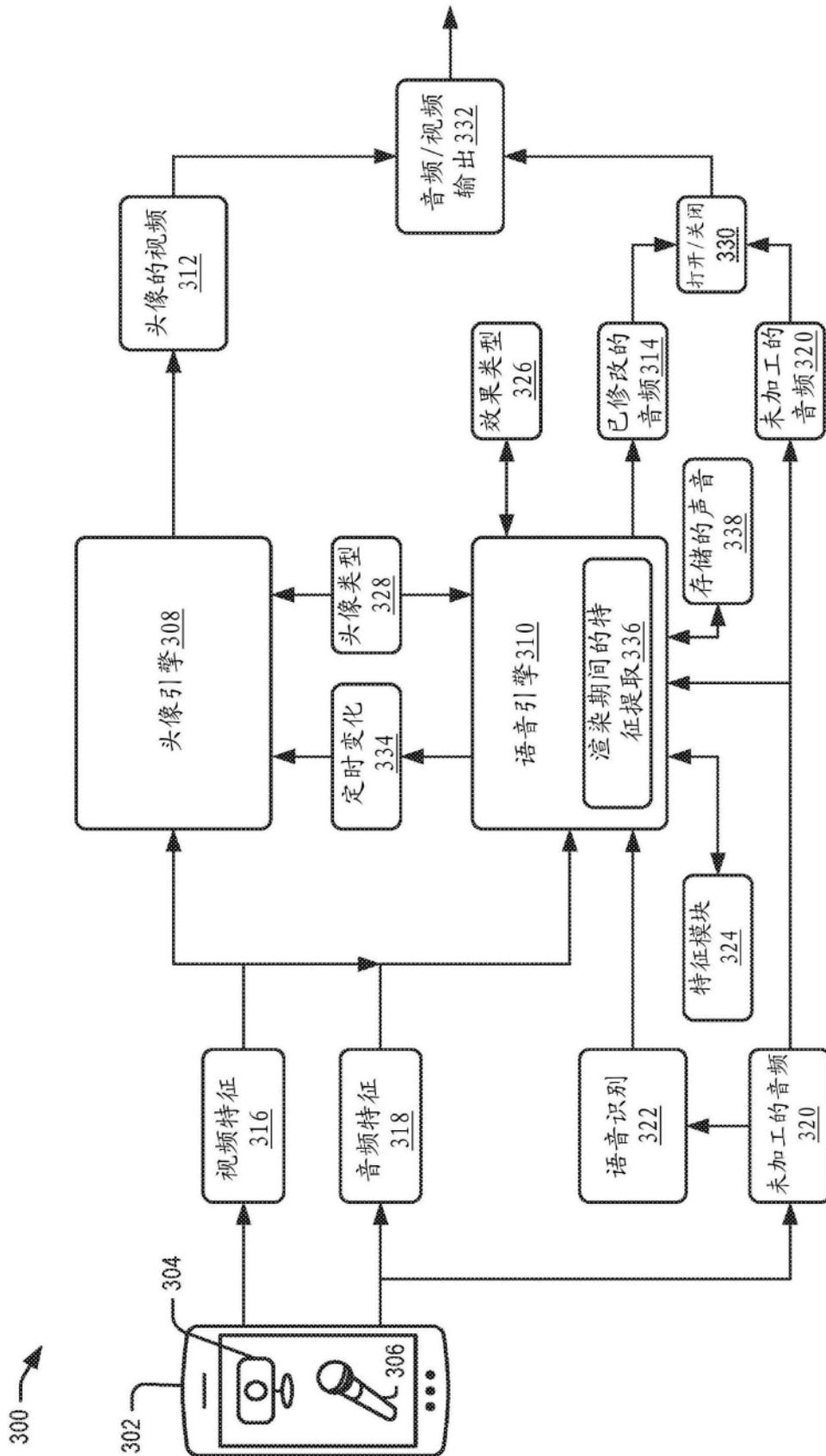


图3

400

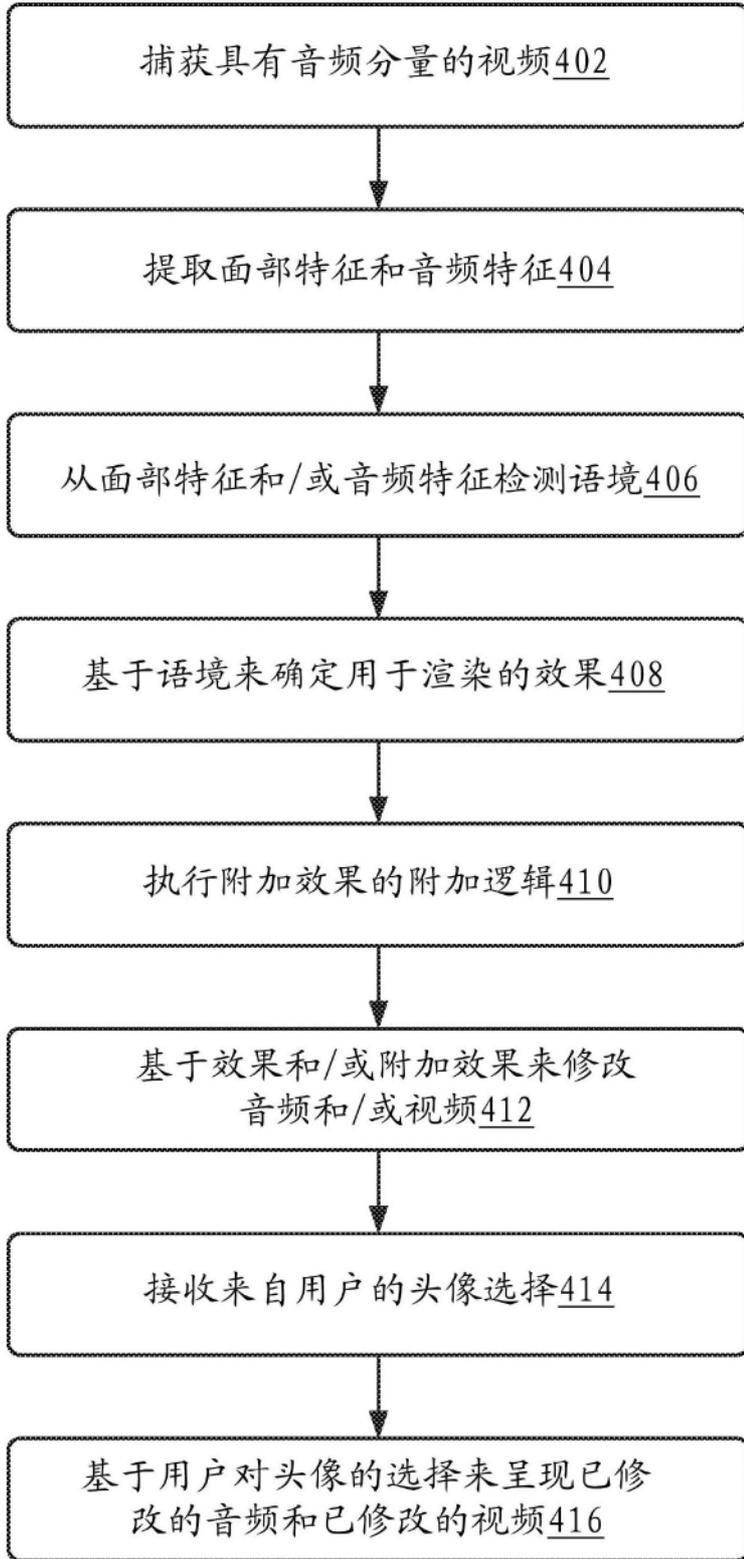


图4

500

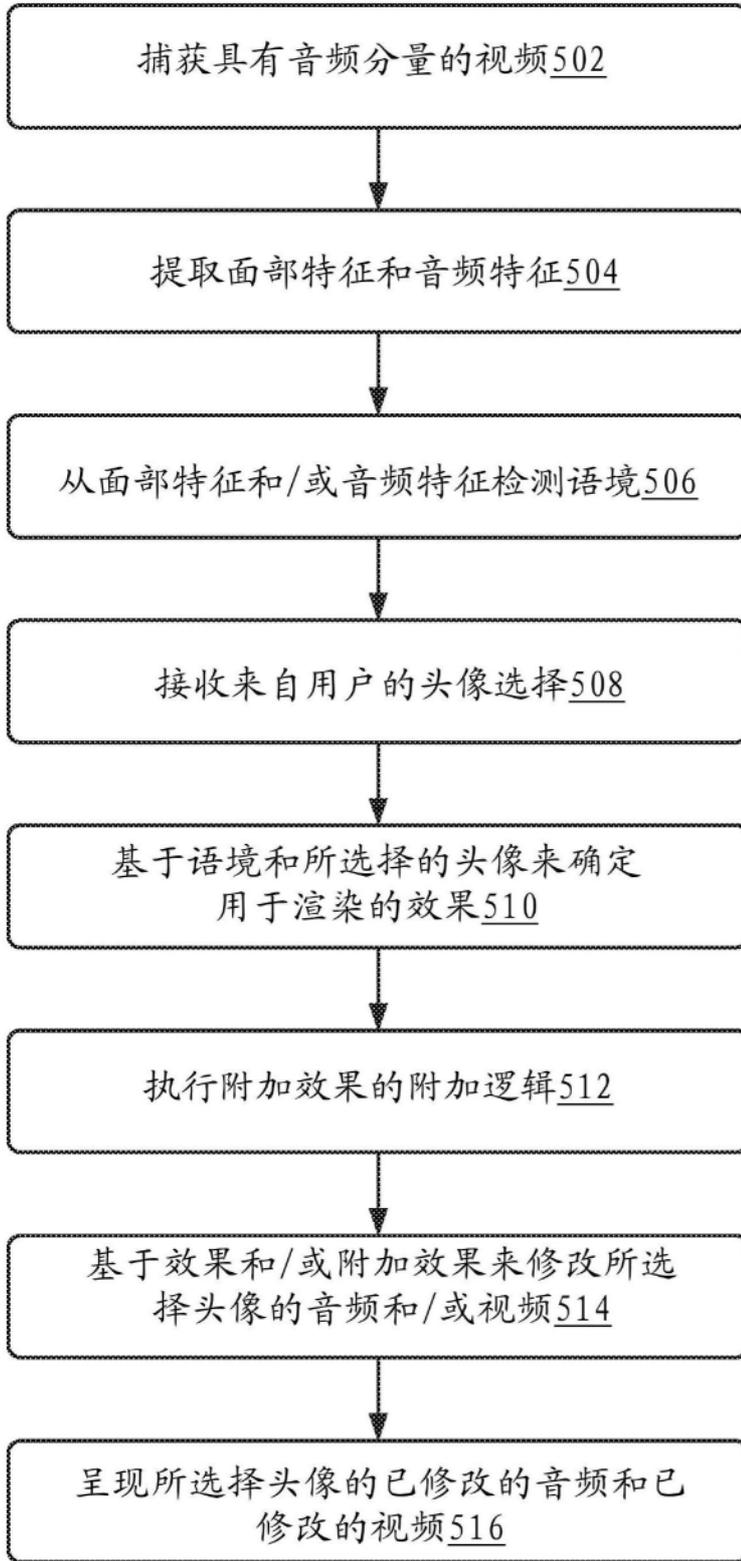


图5

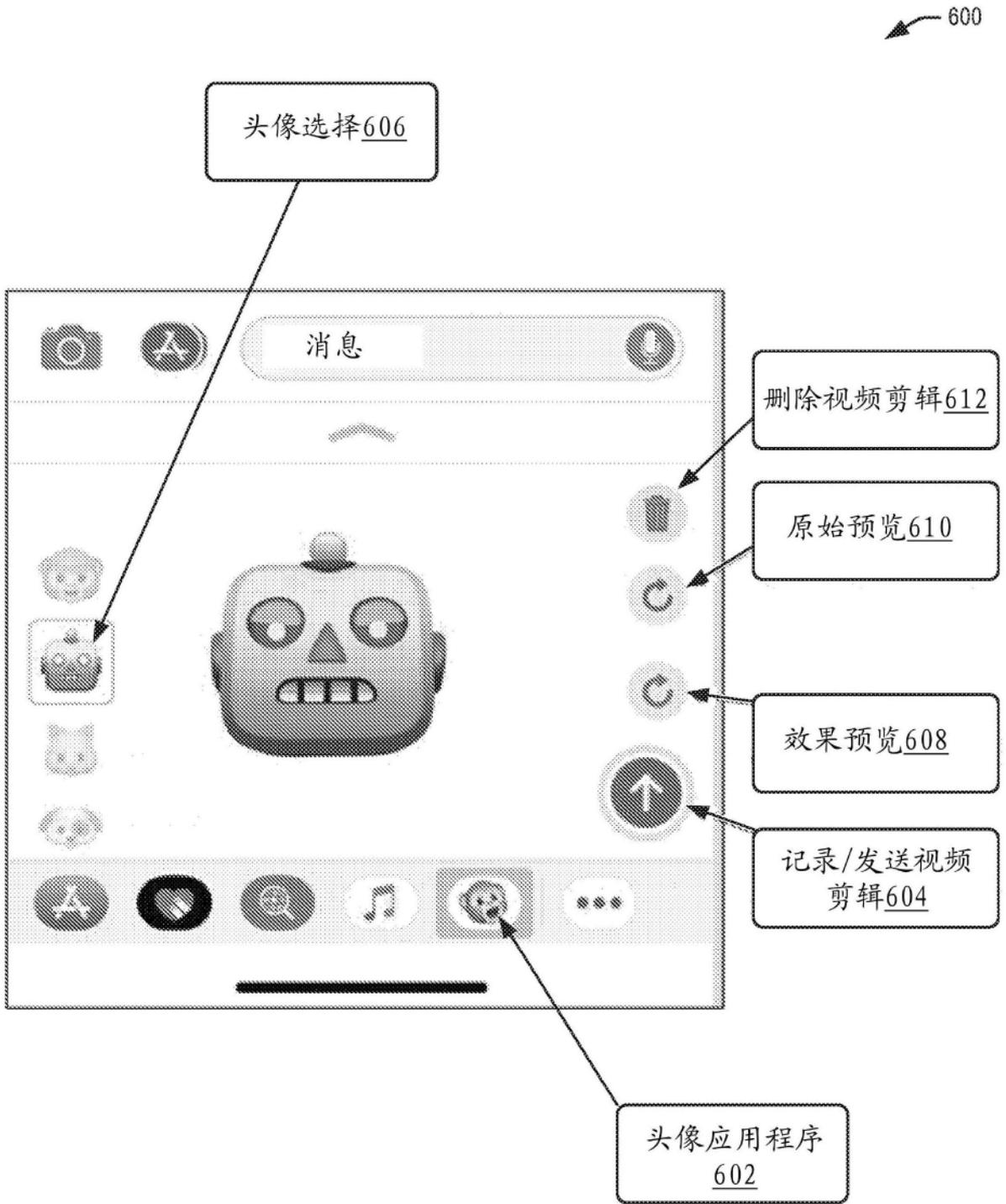


图6

700

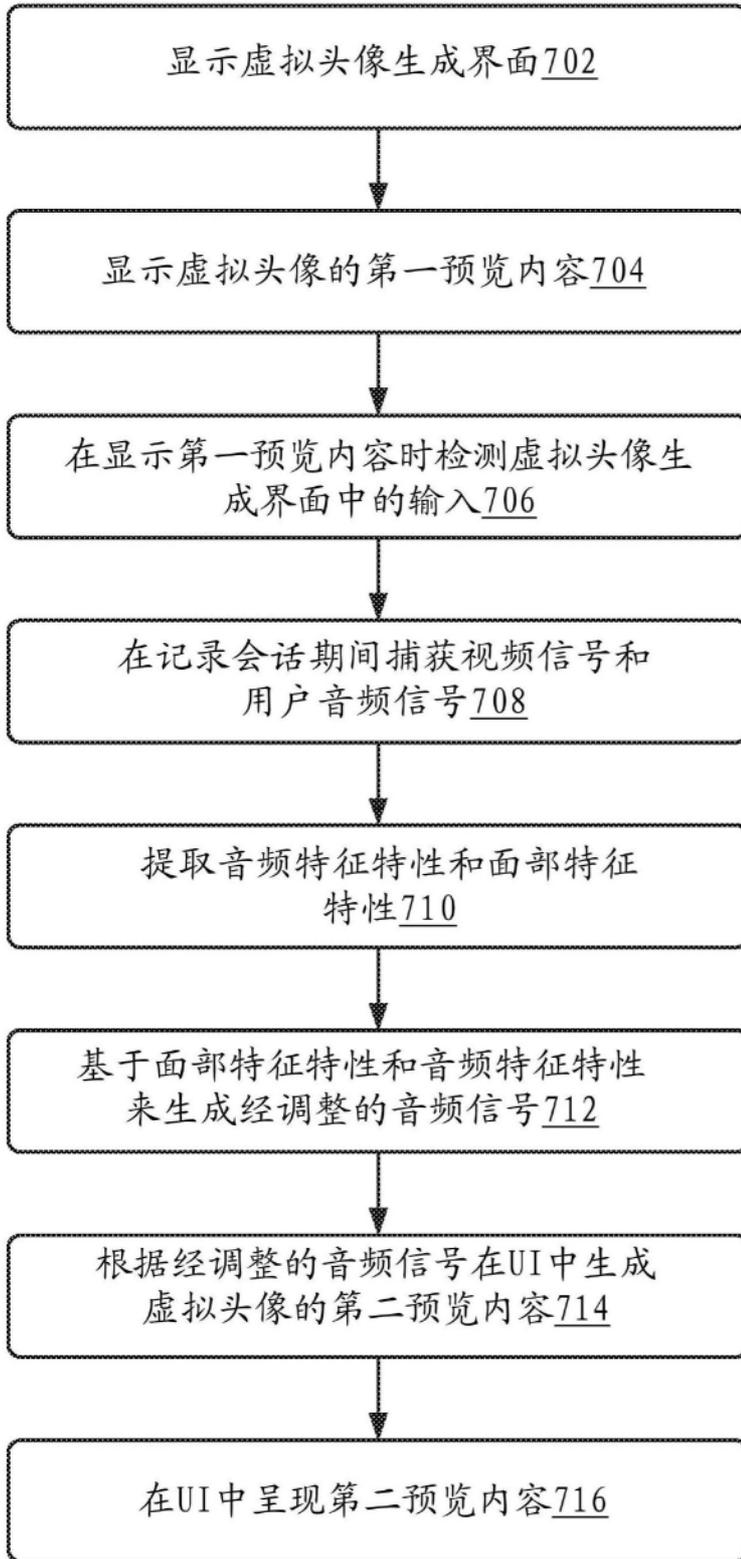


图7

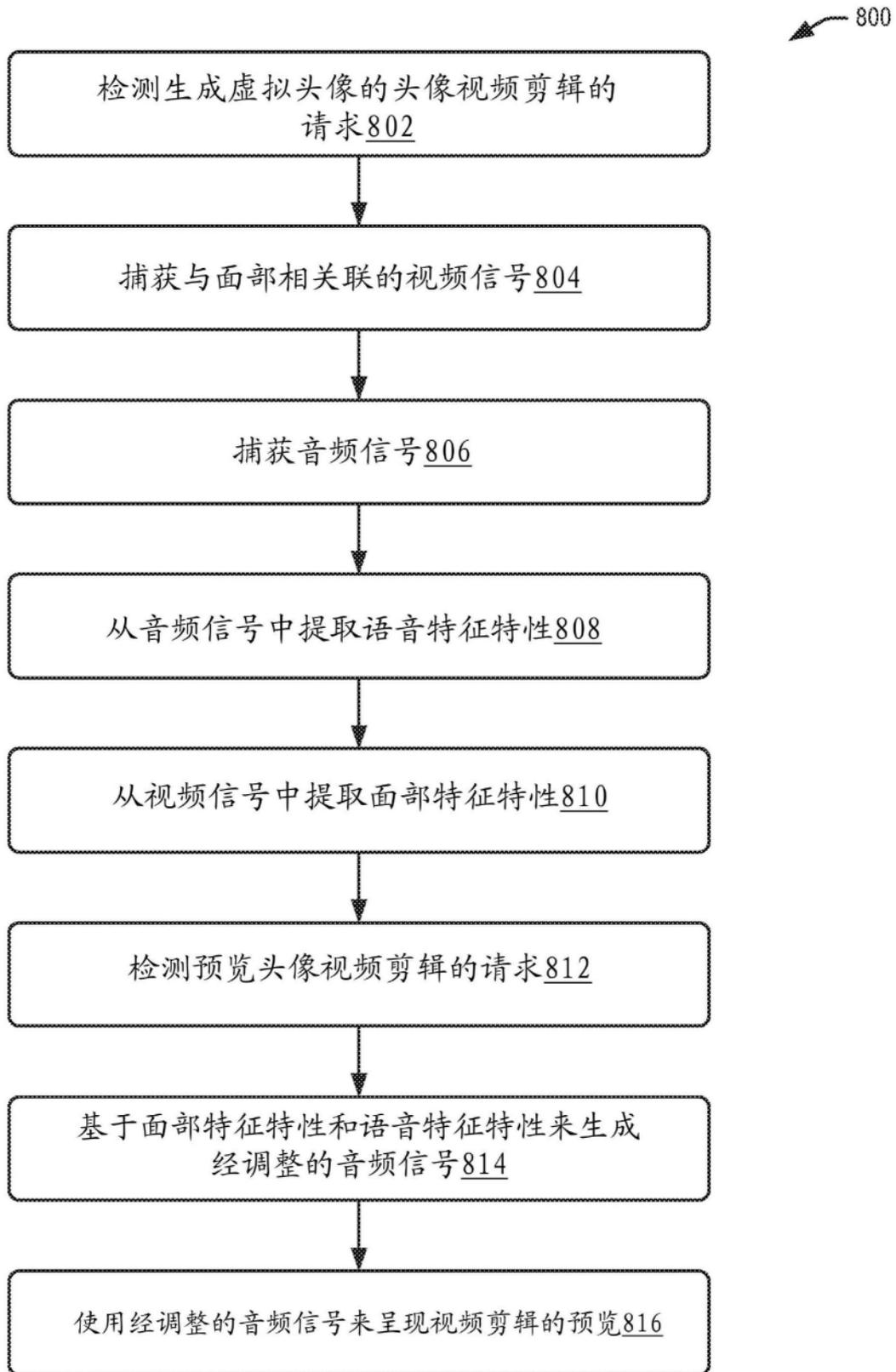


图8

900 →

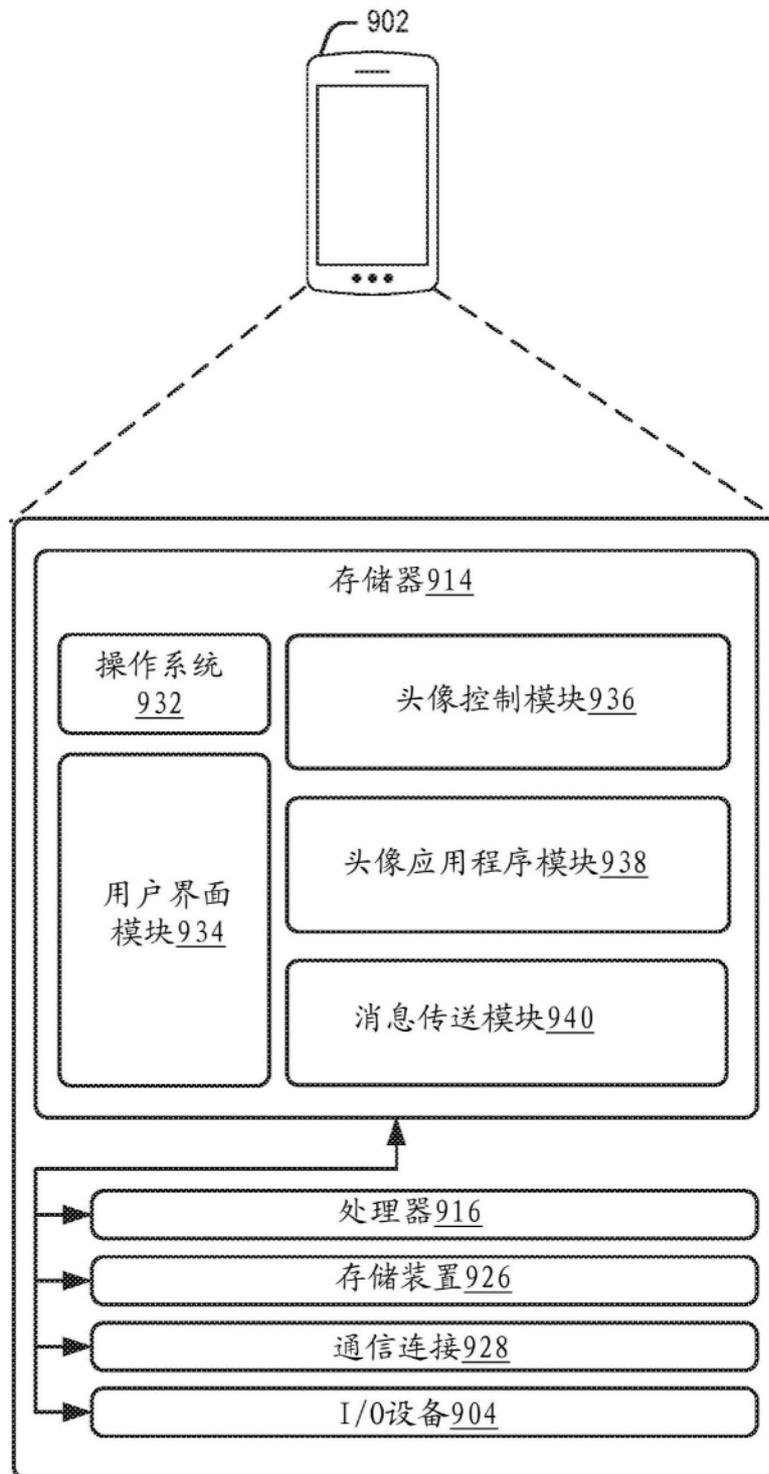


图9