



(12)发明专利

(10)授权公告号 CN 103442201 B

(45)授权公告日 2018.01.02

(21)申请号 201310364747.0

(22)申请日 2008.09.23

(65)同一申请的已公布的文献号
申请公布号 CN 103442201 A

(43)申请公布日 2013.12.11

(30)优先权数据
60/974,774 2007.09.24 US

(62)分案原申请数据
200880117393.7 2008.09.23

(73)专利权人 高通股份有限公司
地址 美国加利福尼亚州

(72)发明人 埃文·希尔德雷思

(74)专利代理机构 北京律盟知识产权代理有限公司 11287

代理人 宋献涛

(51)Int.Cl.
H04N 7/14(2006.01)
H04M 3/42(2006.01)
G06F 3/01(2006.01)
G06F 3/03(2006.01)

(56)对比文件
CN 1684485 A,2005.10.19,
CN 1684485 A,2005.10.19,
WO 2007093984 A2,2007.08.23,
CN 1981325 A,2007.06.13,
US 2004047461 A1,2004.03.11,

审查员 徐惠惠

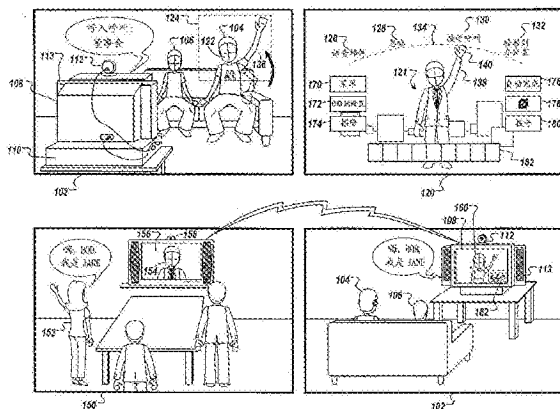
权利要求书3页 说明书28页 附图27页

(54)发明名称

用于语音和视频通信的增强接口

(57)摘要

本发明涉及用于语音和视频通信的增强接口。一种用于语音和视频通信的增强的接口,其中,从一系列相机图像中识别用户的姿态,并且提供包含控件和所述用户的表示的用户接口。所述过程还包括基于所识别的姿态使所述表示与所述控件交互作用,并基于所述交互作用控制电信会话。



1. 一种用于控制电信会话的方法,其包含:

当接收到呼入呼叫时通过移动计算装置将所述电信会话从待机状态改变到通知状态,其中所述待机状态是所述移动计算装置不响应于偶发的用户移动的状态;

通过所述移动计算装置从一系列相机图像中识别出第一用户的第一接合姿态,其中在所述一系列相机图像中的至少一个图像中呈现多个人,所述多个人包括所述第一用户;

使用基于所识别的第一接合姿态和所述第一用户的面部而识别的用户参考位置,将一个或多个相机图像聚焦在所述第一用户上;和

基于识别出所述第一接合姿态而通过所述移动计算装置将所述电信会话从所述通知状态改变到呼叫状态。

2. 如权利要求1所述的用于控制所述电信会话的方法,所述方法进一步包含:

当处于所述通知状态时,将呼入呼叫通知所述第一用户。

3. 如权利要求1所述的用于控制所述电信会话的方法,所述方法进一步包含:

提供包含控件的用户接口;

基于所识别的姿态致使与所述控件的交互作用;和

基于与所述控件的所述交互作用控制所述电信会话。

4. 如权利要求1所述的用于控制所述电信会话的方法,所述方法进一步包含:

检测一个或多个相机图像内的面部;和

确定所检测的脸部是否属于所述移动计算装置的已知用户。

5. 如权利要求4所述的用于控制所述电信会话的方法,其中将所述电信会话从所述通知状态改变到所述呼叫状态进一步基于属于所述已知用户的所检测的脸部。

6. 如权利要求1所述的用于控制所述电信会话的方法,所述方法进一步包含:

在第二系列相机图像中识别出解除接合姿态;和

基于识别出所述解除接合姿态而将所述电信会话改变到所述待机状态。

7. 如权利要求1所述的用于控制所述电信会话的方法,其进一步包含:

识别在所述第一用户已经执行了所述第一接合姿态以后第二用户已经执行了第二接合姿态;

基于识别出所述第二用户已经执行了所述第二接合姿态而把焦点改变到所述第二用户。

8. 如权利要求7所述的用于控制所述电信会话的方法,所述方法进一步包含:

确定所述第一用户是否已经放弃焦点,其中在所述第一用户已经放弃所述焦点之后,所述焦点被改变到所述第二用户。

9. 如权利要求1所述的用于控制所述电信会话的方法,所述方法进一步包含:

提供包含控件的用户接口;

基于所识别的姿态与所述控件交互作用,其包含:

基于所述所识别的姿态计算位置;和

基于所述位置与所述控件交互作用;和

基于所述交互作用控制所述电信会话。

10. 一种用于控制电信会话的系统,其包含:

一个或多个处理器;和

存储器,其通信地耦合到所述一个或多个处理器且被所述一个或多个处理器读取,所述存储器具有存储在其上的处理器可读指令,当由所述一个或多个处理器执行时,所述处理器可读指令致使所述一个或多个处理器:

当接收到呼入呼叫时将所述电信会话从待机状态改变到通知状态,其中所述待机状态是所述系统不响应于偶发的用户移动的状态;

从一系列相机图像中识别出第一用户的第一接合姿态,其中在所述一系列相机图像中的至少一个图像中呈现多个人,所述多个人包括所述第一用户;

使用基于所识别的第一接合姿态和所述第一用户的面部而识别的用户参考位置,将一个或多个相机图像聚焦在所述第一用户上;和

基于识别出所述第一接合姿态而将所述电信会话从所述通知状态改变到呼叫状态。

11. 如权利要求10所述的用于控制所述电信会话的系统,其中所述处理器可读指令在执行时进一步致使所述一个或多个处理器:

当处于所述通知状态时,将呼入呼叫通知所述第一用户。

12. 如权利要求10所述的用于控制所述电信会话的系统,其中所述处理器可读指令在执行时进一步致使所述一个或多个处理器:

提供包含控件的用户接口;

基于所识别的姿态致使与所述控件的交互作用;和

基于与所述控件的所述交互作用控制所述电信会话。

13. 如权利要求10所述的用于控制所述电信会话的系统,其中所述处理器可读指令在执行时进一步致使所述一个或多个处理器:

检测一个或多个相机图像内的面部;和

确定所检测的脸部是否属于所述系统的已知用户。

14. 如权利要求13所述的用于控制所述电信会话的系统,其中致使所述一个或多个处理器将所述电信会话从所述通知状态改变到呼叫状态的所述处理器可读指令包含下述处理器可读指令,其在执行时致使所述一个或多个处理器:

基于属于所述已知用户的所检测的脸部而将所述电信会话从所述通知状态改变到所述呼叫状态。

15. 如权利要求10所述的用于控制所述电信会话的系统,其中所述处理器可读指令在执行时进一步致使所述一个或多个处理器:

在第二系列相机图像中识别出解除接合姿态;和

基于识别出所述解除接合姿态而将所述电信会话改变到所述待机状态。

16. 如权利要求10所述的用于控制所述电信会话的系统,其中所述处理器可读指令在执行时进一步致使所述一个或多个处理器:

识别在所述第一用户已经执行了所述第一接合姿态以后第二用户已经执行了第二接合姿态;

基于识别出所述第二用户已经执行了所述第二接合姿态而把焦点改变到所述第二用户。

17. 如权利要求16所述的用于控制所述电信会话的系统,其中所述处理器可读指令在执行时进一步致使所述一个或多个处理器:

确定所述第一用户是否已经放弃焦点,其中在所述第一用户已经放弃所述焦点之后,所述焦点被改变到所述第二用户。

18. 如权利要求10所述的用于控制所述电信会话的系统,其中所述处理器可读指令在执行时进一步致使所述一个或多个处理器:

提供包含控件的用户接口;

基于所识别的姿态与所述控件交互作用,其中基于所识别的姿态与所述控件交互作用的所述处理器可读指令在执行时致使所述一个或多个处理器:

基于所述所识别的姿态计算位置;和

基于所述位置与所述控件交互作用;和

基于所述交互作用控制所述电信会话。

19. 一种用于控制电信会话的设备,其包含:

用于当接收到呼入呼叫时将所述电信会话从待机状态改变到通知状态的装置,其中所述待机状态是所述设备不响应于偶发的用户移动的状态;

用于从一系列相机图像中识别出第一用户的第一接合姿态的装置,其中在所述一系列相机图像中的至少一个图像中呈现多个人,所述多个人包括所述第一用户;

用于使用基于所识别的第一接合姿态和所述第一用户的面部而识别的用户参考位置将一个或多个相机图像聚焦在所述第一用户上的装置;和

用于基于识别出所述第一接合姿态而将所述电信会话从所述通知状态改变到呼叫状态的装置。

20. 如权利要求19所述的用于控制所述电信会话的设备,所述设备进一步包含:

用于提供包含控件的用户接口的装置;

用于基于所识别的姿态与所述控件交互作用的装置,其包含:

用于基于所述所识别的姿态计算位置的装置;和

用于基于所述位置与所述控件交互作用的装置;和

用于基于所述交互作用控制所述电信会话的装置。

用于语音和视频通信的增强接口

[0001] 分案申请的相关信息

[0002] 本申请为发明名称为“用于语音和视频通信的增强接口”的原中国发明专利申请的分案申请。原申请的申请号为200880117393.7；原申请的申请日为2008年9月23日；发明专利申请案的优先权日为2007年9月24日。

[0003] 相关申请的交叉引用

[0004] 本申请要求2007年9月24日提交的No.60/974,774号美国临时专利申请的优先权，并通过引用将其包含于此。

技术领域

[0005] 本公开涉及用户输入机制，并且至少一种具体配置涉及用户输入机制，该用户输入机制被配置成使用户能够控制语音和视频通信。

背景技术

[0006] 各种基于语音的电子控制系统可以允许用户和计算机应用程序或者其他的语音或者视频应用程序交互作用(interact)。但是，这些系统仅勉强适用于输入文本、数字，或者适用于从大量选项中进行选择。

发明内容

[0007] 根据一个一般性实施方式，用户可以通过执行各种规定的姿态控制通信会话，例如电话会议会话。通过在各种时间访问用户的图像，姿态被自动识别并映射到控件输入(control input)，例如用于开始或者终止电信会话的输入，或者用于将会话置于各种状态的输入。在其他的特征中，接口(interface)可以显示最初环绕所识别的用户的表示(representation) (例如，化身)的定制图标，通过用户身体的直率、直观和自然的运动来实现对人物(characters)命令的轻松选择。

[0008] 根据另一个一般性实施方式，一种计算机实施的过程包括从一系列相机图像中识别用户的姿态，并提供包括控件(control)和用户的表示的用户接口。所述过程还包括基于所识别的姿态使所述表示与控件交互作用，并基于所述交互作用控制电信会话。

[0009] 实施方式可以包括下列特征中的一个或多个。例如，控制电信会话还可以包括把电信会话从第一状态改变到不同的第二状态(a second different state)，并且第一或第二状态可以包括待机状态、通知状态、呼出(outgoing)呼叫状态、连接状态、静音状态、菜单状态、连接选项状态、直接拨号状态、重新拨号状态、搜索状态、浏览状态、消息菜单状态、消息回放状态，或者配置状态。电信会话还可以包括语音和视频电信会话。

[0010] 在额外的示例性实施方式中，控制电信会话还可以包括基于识别出接合姿态(engagement gesture)而把电信会话从待机状态改变到呼叫或者菜单状态。所述过程还可以包括确定用户的身份，并且基于所确定的身份使能(enabling)控件。所述过程可以包括只有所确定的用户身份和预期的(intended)呼叫接收者匹配才把电信会话从连接状态改

变到呼叫状态。所述过程可以包括确定图像内的用户的身份,并基于所确定身份中的至少一个使能或者禁止(disabling)控件。

[0011] 在进一步的示例性实施方式中,控制电信会话可以包括基于识别出解除接合姿态(disengagement gesture)而把电信会话改变到待机状态。所述过程可以包括把相机图像之一聚焦在用户上,其中聚焦相机图像之一可以包括使相机焦点保持在正在移动的用户上,或者使相机图像之一以用户为中心。所述过程可以包括定位(localizing)音频以聚焦在用户上。识别用户的姿态可以包括识别在第一用户已经执行了第一接合姿态以后第二用户已经执行了第二接合姿态,并且控制电信会话可以包括基于识别出第二用户已经执行了第二接合姿态而把焦点改变到第二用户。

[0012] 在其他的示例性实施方式中,所述过程可以包括确定第一用户是否已经放弃(relinquished)焦点,只有第一用户已经放弃焦点,所述焦点才可以被改变。确定第一用户是否已经放弃焦点可以包括确定第一用户是否已经结束讲话,或者确定第一用户是否已经执行了放弃姿态。所述过程可以包括确定用户的身份,其中,控制电信会话还包括在浏览状态中显示与所确定的身份相关联的联系人集合(a set of contacts),或者把所确定的用户身份传送到由用户设定的(placed)呼叫接收者。

[0013] 根据另一个一般性实施方式,一种设备包括被配置成从一系列相机图像中识别用户的姿态、并提供包括控件和用户的表示的用户接口的处理器。所述处理器还被配置成基于所识别的姿态使所述表示与控件交互作用,并基于所述交互作用控制电信会话。

[0014] 根据另一个一般性实施方式,一种计算机可读介质用计算机程序编码。所述计算机程序包括当被执行时,操作以使计算机执行操作的指令,所述操作包括从一系列相机图像中识别用户的姿态,并提供包括控件和用户的表示的用户接口。所述操作还包括基于所识别的姿态使所述表示与控件交互作用,并基于所述交互作用控制电信会话。

[0015] 根据另一个一般性实施方式,一种系统包括一个或多个计算机和耦合到所述一个或多个计算机的计算机可读介质。所述计算机可读介质具有存储于其上的指令,所述指令在被所述一个或多个计算机执行时使所述一个或多个计算机执行操作,所述操作包括从一系列相机图像中识别用户的姿态,并提供包括控件和用户的表示的用户接口。所述操作还包括基于所识别的姿态使所述表示与控件交互作用,并基于所述交互作用控制电信会话。

[0016] 下面在附图和描述中阐述了一个或多个实施方式的细节。从所述描述、附图和权利要求,本公开的其他潜在特征和优点将会清楚。

附图说明

- [0017] 图1是示出了增强的基于相机的输入的概念图。
- [0018] 图2是用来实施增强的基于相机的输入的设备框图。
- [0019] 图3示出了设备的部件的配置。
- [0020] 图4示出了使用增强的基于相机的输入的示范性过程。
- [0021] 图5到图7示出了包括用户的表示的接口的例子。
- [0022] 图8和图9示出了触摸图形的用户的表示的例子。
- [0023] 图10示出了用于选择项目的接口的例子。
- [0024] 图11到图29示出了沿着引导线对齐的项目的示范性选择。

- [0025] 图30是示出了系统的状态的流程图。
- [0026] 图31描绘了通知用户接口的例子。
- [0027] 图32示出了包括执行接合手姿态的用户的相机图像的例子。
- [0028] 图33示出了包括执行解除接合手姿态的用户的相机图像的例子。
- [0029] 图34示出了包括多个坐在会议桌旁的人的相机图像的例子。
- [0030] 图35到图36示出了通过剪裁和缩放相机图像聚焦在用户上的例子。
- [0031] 图37示出了把焦点从第一用户改变到第二用户的相机图像的例子。
- [0032] 图38示出了菜单状态用户接口的例子。
- [0033] 图39示出了直接拨号状态用户接口的例子。
- [0034] 图40到图41示出了重新拨号状态用户接口的例子。
- [0035] 图42示出了连接选项和信息状态用户接口的例子。
- [0036] 图43示出了消息回放状态中的用户接口的例子。
- [0037] 图44示出了根据另一个一般性实施方式的示范性计算机处理系统的部件。

具体实施方式

[0038] 使用这里描述的增强的方法,用户可以通过执行各种规定的姿态来控制通信会话,例如电话会议会话。通过在各种时间访问用户的图像,姿态被自动识别并映射到控件输入,例如用于开始或者终止电信会话的输入,或者用于将会话置于各种状态的输入。在其他的特征中,接口可以显示最初环绕所识别的用户的表示(例如,化身)的定制图标,通过用户身体的直率、直观和自然的运动来实现对人物命令的轻松选择。

[0039] 因此,描述了允许用户使用相机追踪(tracking)来控制语音和视频通信的方法。例如,这种方法可以在会议室环境(setting)中使用,其中,通信系统在物理上可以在用户所及之外,或者在用户可能希望显得如同在商业环境中那样的另一位置。例如,这种方法也可以在起居室环境中使用,其中,坐在沙发上的用户可以和包括在电视或者其他显示设备上显示的接口的通信系统交互作用。

[0040] 所述方法可以从一系列相机图像中识别用户的姿态,并提供包括控件和用户的表示的用户接口。基于所识别的姿态,系统可以使所述表示与控件交互作用,然后基于所述交互作用控制电信会话。

[0041] 例如,系统可以包括手姿态(hand gesture)识别过程,该手姿态识别过程被配置成检测用户的手姿态,并基于所检测的手姿态执行功能。例如,用户可能通过执行手、上臂或者身体“接合”姿态来应答呼入(incoming)呼叫。在多个用户参与呼叫的实施方式中,用户可以通过执行相同或者不同的姿态来获取或者改变焦点,例如“解除接合”姿态、“改变焦点”姿态或者“释放控制”姿态。

[0042] 图1是示出增强的基于相机的输入的上下文说明图。图1包括起居室102的视图。用户104(即,“BOB”)不正式地着装,并且正坐在起居室102中的沙发上,和孩子106相邻。用户104和孩子106正坐在显示器108(例如,电视屏幕)的前面,并且在接到呼叫之前可能正在观看节目或者在玩视频游戏。

[0043] 显示器108连接到媒体集线器110。媒体集线器110被配置成接受呼入电话或者视频会议呼叫。媒体集线器110连接到相机112(例如,网络摄像头),或者另外包括相机112。在

这个实施方式中,显示器108包括扬声器113。在其他的实施方式中,扬声器113被包括在媒体集线器中,或者在媒体集线器和显示器108外部(例如,电视扬声器)。媒体集线器110还包括或者另外连接到用于接收和数字化环境声音的麦克风。

[0044] 在这个例子中,媒体集线器110检测呼入呼叫并通过音频和/或视频消息警告用户104。在这个例子中,当扬声器113输出指示“呼入呼叫:董事会”的音频时,用户104被警告呼入呼叫来自该用户公司的董事会。例如,通过将文本转语音(text-to-speech)能力施加于呼叫者ID系统,或者通过其他机制,可以产生这样的警告。当进行正式的电话会议时,虽然BOB的非正式着装和周边环境本可能让他担心,但是这里描述的增强的方法通过产生隐藏他的非正式着装和周边环境的视觉化身(visual avatar)解决了这些问题,或者以更适当的可视化来代替。

[0045] 响应于检测到呼入呼叫,用户接口120可以被示出在显示器108上,可以替换、或者部分或完全地覆盖用户104在接收到呼叫之前正在显示器108上观看的视频内容。用户接口120包括用户表示121。在这个例子中,用户表示121是一种化身,其显示了被相机112在相机视场的区(region)124中捕获的用户104的面部122。

[0046] 尽管用户104未穿着工作装,但是在用户接口120中用户表示121被示出为穿着正装和领带。用户表示121的外观(appearance)(例如,化身的外观)可以被基于呼叫者的身份、一天中的时间、用户喜好、或者其他因素而改变。例如,根据用户设置,用户表示可以被配置成看起来更像用户104或者不像用户104。在更极端的例子中,用户表示121可以穿着燕尾服,或者比用户104实际上穿着的更不正式的服装(即,游泳衣)。此外,用户的面部122或者其他身体部分也可以被改变,例如显得更瘦,具有不同的发型和风格,去除附属物(即,手饰),或者具有完全不同的样子(即,怪兽、外星人、卡通人物、另一个人的样子,或者任何其他外表)。

[0047] 用户接口120包括项目126-132的表示。项目126-132可以沿着引导线134对齐(aligned)。引导线134可以显示或者不显示。

[0048] 响应于用户104的运动或者姿态而控制用户表示121。例如,当用户104抬起其左上臂136时,用户表示121的左上臂138被以对应的方式抬起。在其他的例子中,用户的运动 and 用户表示的运动可以被镜像(mirrored)。用户104可以移动其上臂136以使用户表示121的手140移动靠近或者覆盖项目126-132之一。例如,手140可以覆盖“接受呼叫”项目130,导致与该项目130的交互作用。作为与“接受呼叫”项目130交互作用的结果,呼入呼叫被接受。

[0049] 图1包括会议室150的视图。呼叫者152(即,“JANE”)正和其他人一起站在会议室150中,其中,呼叫者152正使用类似的系统执行类似的姿态以调用呼叫。作为用户104接受来自呼叫者152的呼入呼叫的结果,用户表示121的图像154现实在安装在会议室150中的显示器156上。图像154以用户104为中心(centered around),而不显示起居室102中的其他人(即,孩子106)。显示器156以及相机158和麦克风可以连接到位于会议室150中的媒体集线器。

[0050] 在显示器108上示出被相机158捕获的呼叫者152的表示160。在扬声器113上播放由会议室中的麦克风捕获的呼叫者152的语音的音频(即,“嗨,BOB,我是JANE”)。显示器108的插页(inset)162显示用户接口120,允许用户进一步控制通信,例如切断呼叫或者使呼叫静音。

[0051] 用户接口120还包括菜单按钮170、用户设置按钮172、擦除 (erase) 按钮174、自动完成按钮176、处于禁止状态的按钮178、拨号按钮180和文本区域 (area) 182。可以选择菜单按钮170以显示额外的特征。可以选择用户设置按钮172来配置特定于用户104的设置。例如,面部检测和识别过程可能已经识别了用户104。

[0052] 用户104可以把电话号码或者联系人姓名或者文本消息输入文本区域182。例如,沿着引导线134可以显示字母的表示,并且被选择的字母可以显示在文本区域182中。用户可以选择拨号按钮180来把呼叫连接到在文本区域182中显示的电话号码或者联系人。可以选择自动完成按钮176来接受先前拨号的号码或者与文本区域182中显示的文本部分地匹配的联系人姓名。用户可以选择擦除按钮174以便从文本区域182擦除字符。

[0053] 用户104可以选择语音邮件项目126以便导航到语音邮件特征。可以选择忽略项目128以忽略呼入呼叫。例如,可以选择转移到办公室项目132以便把呼入呼叫转移到办公室分机。

[0054] 按钮178可能因检测到存在孩子106或者出于另外的原因而处于禁止状态。例如,如果在相机112的视场内检测到孩子,则快速拨号或者长途拨号功能或者对面向成人的内容的访问可以被禁止。在另一个例子中,和按钮178相关联的功能可以被通信提供商禁止,例如,在文本消息服务正在维修的情况下或者服务被暂停的情况下。

[0055] 图2是用来实施增强的基于相机的输入的设备200的框图。简单地说,除了其他事物,设备200至少包括用户接口201、存储介质202、相机204、处理器205、麦克风206、一个或多个扬声器207和网络接口208。某些元件或者它们的相关功能可以被组合。

[0056] 在图3中示出了设备200的部件的示例性物理配置,其中,相机301和麦克风302被连接到处理器303并放在显示器104上方。扬声器305附着于显示器304,并且处理器303连接到网络306。

[0057] 往回参考图2,处理器205可以被包括在PC (个人计算机)、有线电视接收机、视频游戏控制台、电话或者其他类型的设备中。处理器205可以被包括在任何类型的电子设备中,并且可以和多个处理器结合操作。相机204捕获或者另外产生可被传送到处理器205的相机图像。处理器205可以使用贯穿本公开描述的技术来处理相机图像。

[0058] 麦克风206可以是全方向的或者单方向的,并接受过滤和未过滤的可听 (audible) 输入,并且可以产生被传送到处理器205的声音数据。处理器205可以产生用于在用户接口201上呈现 (rendering) 的用户界面,并且处理器205也可以产生被传送到扬声器207并由其转换为声音的音频数据。

[0059] 使用设备200,增强的基于相机的输入可以用来帮助与用户接口的控件进行交互作用,例如以允许由用户选择项目。和控件的交互作用不要求用户在手中持有例如逆向反射器、陀螺设备的任何专用 (specialized) 对象或者遥控器,而是直观地把上臂或者手运动映射到光标或者其他输入。基于系统的具体的期望配置,可以使用专用对象。

[0060] 用户接口201是用于允许用户和设备或者该设备调用的应用程序交互作用的机制。用户接口201可以实现输入和输出,允许用户操纵设备或者使设备产生用户操纵的效果。设备200可以利用任何类型的用户接口201,例如图形用户接口 (GUI)、语音用户接口或者触摸或触觉用户接口。

[0061] 用户接口201可以被配置成呈现视觉显示图像。例如,用户接口201可以是监视器、

电视、液晶显示器 (LCD)、等离子体显示设备、具有投影仪屏幕的投影仪、自动立体显示器、阴极射线管 (CRT) 显示器、数字光处理 (digital light processing, DLP) 显示器, 或者任何其他类型的被配置成呈现显示图像的显示设备。用户接口 201 可以包括一个或多个显示设备。在某些配置中, 用户接口 201 可以被配置成显示和应用程序相关联的图像, 例如由应用程序产生的显示图像, 包括控件和例如化身的对象。存储介质 202 存储和记录信息或者数据, 并且可以是光学存储介质、磁性存储介质、快闪存储器, 或者任何其他存储介质类型。

[0062] 相机 204 是用来捕获图像的设备, 所述图像或者作为静态照片, 或者作为一系列运动图像。相机 204 可以使用可见频谱的光或者具有电磁频谱的其他部分, 例如红外线。例如, 相机 204 可以是数码相机、数码摄像机, 或者被配置成捕获图像的任何其他类型的设备。相机 204 可以包括一个或多个相机。在某些例子中, 相机 204 可以被配置成捕获和应用程序交互作用的用户或者对象的图像。例如, 相机 204 可以被配置成在相机 204 的视场内身体上与应用程序交互作用的人或者用户的图像。

[0063] 相机 204 可以是立体相机、飞行时间 (time-of-flight) 相机, 或者任何其他相机。例如, 相机 204 可以是能够采样背景图像以便检测用户的运动和类似地检测用户的姿态的图像检测器。相机 204 可以产生灰度图像、彩色图像, 或者距离图像, 例如能够产生距离图像的立体相机或者飞行时间相机。立体相机可以包括两个在略微不同的视点获取图像的图像传感器, 其中, 处理器比较从不同的视点获取的图像以计算图像的部分的距离。飞行时间相机可以包括产生光脉冲的发射器, 所述光可以是红外光, 其中, 测量光脉冲从发射器传播到目标并回到传感器的时间以计算图像的部分的距离。

[0064] 设备 200 通过有线或者无线路径被电连接到相机 204 和用户接口 201, 并被配置成控制处理器 205 的操作, 以提供基于相机的输入。在一种配置中, 设备 200 使用处理器 205 或者其他的控制电路来执行提供给增强的基于相机的输入的应用程序。具体来说, 设备从相机 204 接收输入, 并处理所接收的输入以计算用户接口 201 中的用户的表示的位置和移动, 并基于这些移动与控件交互作用。

[0065] 在一个示例实施方式中, 通过使用相机检测执行姿态的用户的图像而产生输入。例如, 移动电话可以被放在桌子上, 并且可操作以产生使用面朝前方相机 (face-forward camera) 的用户的图像。或者, 使用相机 204 可以检测或者识别姿态, 例如通过使用光流 (optical flow) 或者某种其他方法检测“向左倾斜”姿态, 并使用这个检测的姿态把用户的表示向左移动并选择被布置在控件左侧的项目, 或者通过检测“向右前倾斜”姿态以便把用户的表示向中间 (neutral) 位置的上和右移动, 以选择被布置在控件右上侧的项目。

[0066] 因此, 相机 204 还可以包括某个其他类型的能够检测设备 200 的角位置的输入设备或者模块, 例如陀螺仪、加速度计或者光流追踪器, 或者被它们代替。在这个方面, 相机可以用倾斜传感器输入补充或者代替, 以执行用户期望的功能或者命令。因此, 用户姿态的检测可以不使用相机来进行。通过例如在用户接口上对控制可视化的、以相同种类的比划图案 (stroke pattern) 移动所述设备, 使用户能够以直率、直观和视觉愉悦和刺激的方式控制相同的接口或者应用程序。

[0067] 麦克风 206 可以包括多个可操作以便在空间上定位声音的传感器。麦克风 206 可以包括过滤过程, 所述过程可操作以便抑制背景噪声并取消回声。麦克风 206 可以是用户接口 201 的一部分, 例如计算机监视器包括搭载的 (on-board) 麦克风的情况, 或者, 可以与用户

接口201分离,例如具有内置(built-in)麦克风的网络摄像头被连接到计算机监视器上的USB端口的情况。音频扬声器207可操作以产生声音。和麦克风206类似,音频扬声器207可以是用户接口201的一部分,或者可以与用户接口201分离。

[0068] 处理器205可以在网络接口208上连接一个或多个网络。处理器205可以连接到能够把音频或视频通信传送到其他系统的计算机、电话或者视频网络(例如,因特网、网际协议语音(Voice over Internet Protocol,VOIP)、公共交换电话网络(public switched telephone network,PSTN)、电话会议服务,或者有线电视服务)。网络可以包括例如以下中的一个或多个:因特网、广域网(WAN)、局域网(LAN)、模拟或者数字的有线和无线电话网络(例如,PSTN、综合业务数字网络(Integrated Services Digital Network,ISDN)和数字订户线(Digital Subscriber Line,xDSL))、广播、电视、有线电视、卫星,和/或任何其他的用于携带数据或者VOIP服务的传递或者隧道机制。

[0069] 网络还可以包括电路交换语音网络、分组交换数据网络,或者任何其他能够携带语音通信的网络。例如,电路交换语音网络可以包括PSTN,并且分组交换数据网络可以包括基于网际协议(IP)、X.25或者帧中继或者其他相当的(comparable)技术的网络,并且可以使用例如VOIP或者其他相当的用于语音数据通信的协议支持语音。网络可以包括多个网络或者子网络,它们中的每一个均包括例如有线或者无线数据路径(pathway)。

[0070] 数据200可以实施协助语音和视频通信的过程。例如,处理器205可以执行与语音和视频通信相关的应用程序,并在用户接口201上呈现允许用户与和语音和视频通信相关的应用程序交互作用的界面。

[0071] 设备200可以实施音频数据解压缩过程。来自远程用户的语音可以在网络上作为压缩数据传送,所述压缩数据在被扬声器207转换为声音之前被解压缩。音频数据解压缩过程可以利用音频压缩方法,例如码激励线性预测(Code Excited Linear Prediction,CELP)、先进音频编码(Advanced Audio Coding,AAC),或者其他方法。

[0072] 设备200可以实施视频数据解压缩过程。远程用户的视频可以在网络上作为压缩数据传送,所述压缩数据在被用户接口201显示之前被解压缩。视频解压缩过程可以利用视频压缩方法,例如国际电信联盟(International Telecommunications Union,ITU)H.323、H.264、运动画面专家组(Moving Picture Experts Group)MPEG-2、MPEG-4,或者其他方法。

[0073] 设备200可以实施音频回声取消过程。回声取消(echo cancellation)可以包括声学回声取消。当扬声器207产生的声音被麦克风206感测到(sensed)时,可能出现声学回声。因此,来自远程用户的在网络上传送并通过扬声器207播放的语音可能被麦克风206感测,并在网络上传送回远程用户。这是不期望的。声学回声可能受到扬声器207、麦克风206和环境(例如房间的混响或声学性质)的特性的影响。

[0074] 声学回声取消过程可以包括:采样参考音频信号、使用扬声器从参考音频信号中产生声音、使用麦克风感测由扬声器产生的声音,并采样由麦克风感测的声音作为第二音频信号。补偿扬声器、麦克风和环境的特性的影响的预测参考音频信号可以被产生并与第二音频信号进行比较。可以从第二音频信号去除(remove)预测参考音频信号以产生回声被抑制或者消除(eliminate)的第三音频信号。本公开并不局限于特定的回声取消过程,也可以使用其他回声取消技术。

[0075] 设备200可以实施音频数据压缩过程。来自本地用户的由麦克风感测并被回声取

消过程处理过的语音可以被压缩,并在网络上作为压缩数据传送到远程用户。音频数据压缩过程可以利用音频压缩方法,例如码激励线性预测(Code Excited Linear Prediction, CELP)、先进音频编码(Advanced Audio Coding, AAC),或者其他方法。

[0076] 系统可以实施视频数据压缩过程。本地用户的由相机捕获的视频可以被压缩,并在网络上作为压缩数据传送到远程用户。视频压缩过程可以利用视频压缩方法,例如国际电信联盟(International Telecommunications Union, ITU) H.323、H.264、运动画面专家组(Moving Picture Experts Group) MPEG-2、MPEG-4,或者其他方法。

[0077] 处理器205可操作以执行几个相机追踪过程,包括手姿态识别过程、面部检测过程、面部识别过程、用户位置过程、产生用户的表示的过程,以及把用户位置与应用程序产生的项目比较的过程。这些过程中的一些是可选的,或者可以被组合。系统不局限于这些过程,并且可以执行其他过程来实施贯穿本公开描述的技术。

[0078] 在某些实施方式中,处理器205可操作以执行几个语音和视频通信过程,包括音频解压缩过程、视频解压缩过程、音频回声取消过程、音频压缩过程和视频压缩过程。用于协助语音和视频通信的一些或者全部过程可以由处理器205执行。或者,处理器205可以和执行通信操作的外部设备进行通信。

[0079] 尽管设备200已经被描述为个人计算机(PC)或者机顶盒,但是仅仅是为了简洁的目的才做出这样的描述,并且其他的实施方式或者表现形式也是可以预期的。例如,设备200可以被实施为电视、音乐播放器、数码相框、超移动个人计算机(ultra-mobile personal computer, UMPC)、移动因特网设备(mobile internet device, MID)、数码相框(digital picture frame, DPF)、便携式媒体播放器(portable media player, PMP)、通用或者专用计算机(例如台式计算机、工作站或者膝上型计算机)、服务器、游戏设备或者控制台,或者任何其他类型的包括处理器或者其他被配置成执行指令的控制电路的电子设备,或者任何其他的包括用户接口的装置。

[0080] 图4示出了用于控制电信会话的示范性过程400。简单地说,过程400包括:从一系列相机图像中识别用户的姿态、提供包括控件和用户的表示的用户接口、基于所识别的姿态使所述表示与控件交互作用,并基于所述交互作用控制电信会话。也可以使用其他的过程,并且过程400可以按不同于所示的顺序发生,或者可以包括更少或者更多的操作。

[0081] 更详细地,当过程400开始时(S401),从一系列相机图像中识别用户的姿态(S402)。例如,可以监控一个或多个相机图像以识别手姿态或者利用其他对象或者身体部分做出的姿态。例如,相机可以捕获与在显示设备上显示的接口交互作用的用户的一个或多个图像。处理器可以访问被相机捕获的图像,并通过处理图像识别手姿态。

[0082] 如这里通篇所使用的,“姿态”(gesture)意指利用整个或者部分人体或者多个人体做出的非语言通信的形式,并且和例如讲话的语言通信相反。例如,姿态可以由第一位置、姿势(pose)或者表情和第二姿势、位置或者表情之间的移动、变化或者变形来定义。示例的姿态包括例如“空中引号”姿态、鞠躬姿态、屈膝礼、吻脸颊、手指或者手运动、屈膝、摇头或者移动、庆祝成功的击掌、点头、悲伤面孔、举起的拳头、敬礼、猛击或者挥舞运动、竖大拇指运动、手画圆圈或者挥手姿态,或者手指指向姿态。

[0083] 因此,从一系列图像,可以导出定义用户的主意、见解、情绪、沟通、命令、演示或者表情的姿态。例如,用户的姿态可以是单个或者多个手指姿态、单个手姿态、单个手和上臂

姿态、单个手和上臂和身体姿态、双手姿态、头姿势或者体位 (posture) 的变化、眼睛位置的变化、面部表情的变化、身体姿势或者体位的变化, 或者任何其他有表现力的身体状态的变形。

[0084] 为了简洁, 用来执行相关姿态的身体部分一般被称为“控制对象”。例如, 用户可以使用其整个身体或者利用其他物理对象表达命令, 在这种情况下, 其他的整个身体或者其他的物理对象可以是控制对象。用户可以通过眨他们的眼睛、通过张开其鼻孔, 或者通过摆动手指来更巧妙地表达命令, 在这种情况下, 眼皮、鼻子或者手指可以是控制对象。单个图像或者两个图像之间的用户姿态可以表达使能或者“接合”姿态。控制对象也可以是物理设备, 例如红外手指灯、逆向反射器、或者遥控器。

[0085] 识别手姿态的过程可以包括分析运动。分析运动可以包括: 从一系列相机图像中计算光流图、聚类 (clustering) 具有类似方向和大小的运动的流图 (flow map) 的区, 和在一个时间段内聚类的 (clustered) 运动和统计模型 (例如隐式马可夫模型) 进行比较, 以便把运动分类为姿态。运动是外观和位置不变的, 因此, 可以在红外相机图像序列上被检测到 (例如其不取决于皮肤颜色)。在其中在红外相机图像序列中检测运动的实施方式中, 可以在低环境光中检测姿态。当用于能够测量距离的相机时 (例如立体相机或者飞行时间相机), 运动分析可以使用二维运动或三维运动。

[0086] 识别手姿态的过程可以包括分析颜色。分析颜色可以包括把相机图像和颜色模型进行比较, 识别相机图像的具有符合人类皮肤的颜色的一部分, 聚类相机图像的具有符合人类皮肤的颜色的那些部分, 并且如果聚类满足一组尺寸和形状标准, 则把其分类为手。分析颜色还可以包括把聚类形状和一组手姿势 (hand pose) 标准进行比较, 以便把手姿势分类为手姿态。另外, 分析颜色可以包括在一个时间段内产生手聚类位置的路径 (path), 并把所述路径和统计模型进行比较以便把运动分类为姿态。

[0087] 识别手姿态的过程可以包括分析特征。分析特征可以包括识别相机图像内的基本形状特征 (例如边缘), 并识别可能对应于手的特征集合。特征的空间结构可以被和统计模型进行比较, 以便把手姿势或者手移动分类为姿态。

[0088] 识别手姿态的过程还可以包括把手位置和面部 (或者身体) 位置进行比较, 并使用相对手位置。面部位置可以通过面部检测过程来确定, 并且可以使用相对于所检测的面部的手位置来检测姿态或者其他输入命令。

[0089] 识别手姿态的过程可以包括运动、颜色和形状技术的组合。本公开并不局限于特定的手姿态检测过程, 也可以使用其他用于检测手姿态或者其他的输入命令的技术。

[0090] 系统可以实施检测一个或多个相机图像内的面部的过程。面部检测过程可以确定一个或多个相机图像内的人面部的的位置、尺寸, 或者其他的物理特性。

[0091] 检测相机图像内的面部的过程可以包括分析颜色。分析颜色可以包括把相机图像和颜色模型进行比较, 识别相机图像的具有符合人类皮肤的颜色和面部特征的部分, 聚类相机图像的具有符合人类皮肤的颜色和面部特征的那些部分, 并且如果聚类满足一组尺寸和形状标准, 则把其分类为面部。

[0092] 检测相机图像内的面部的过程可以包括分析特征。分析特征可以包括识别一个或多个相机图像内的特征 (例如边缘或者梯度), 并识别可以对应于面部的特征集合。本公开并不局限于特定的面部检测过程, 也可以使用用于检测面部的其他技术。

[0093] 所述系统可以实施在一个或多个相机图像内识别面部以识别特定用户的过程。例如,处理器103可以分析由相机101捕获的图像以检测位于相机101前面的用户的面部。在检测面部之后,处理器103可以分析所检测的面部来确定所检测的面部是否属于可识别或者已知的用户。

[0094] 在相机图像内识别面部的过程可以包括选择图像的检测到面部的部分,并且把图像的该部分和模板进行比较。所述系统可以包括用户的模板。可以在配置状态中使用下面更详细地描述的技术产生所述模板。

[0095] 在使用了测程相机(例如立体相机或者飞行时间相机)的情况下,识别相机图像内的面部的过程可以包括使用三维面部识别过程。三维面部识别过程可以包括选择图像的检测到面部的部分(例如使用面部检测过程的结果),产生所检测的面部的三维模型(例如网格数据)和纹理模型,并把所述三维模型和纹理模型和用户的模板进行比较。在配置状态中可以使用在下面更详细地描述的技术产生所述模板。

[0096] 在相机图像内识别面部的过程可以包括使用本征图像(eigenimage)过程。本征图像面部识别过程可以包括选择图像的检测到面部的部分(例如使用面部检测过程的结果),计算协方差矩阵,计算所述协方差矩阵的本征向量(eigenvector)和本征值(eigenvalue),执行主(principal)成分分析以选择主成分并减少本征向量和本征值的维度,把作为结果的本征图像数据和预先定义的本征图像数据集合进行比较以产生在所检测的面部和预先定义的本征图像数据集合中的面部之间的相似性测量集合,并执行统计分析来把本征图像分类为特定用户的面部。统计分析可以利用在训练过程期间产生的数据,在配置状态中可以使用在下面更详细地描述的技术执行所述训练过程。本公开并不局限于特定的面部识别过程,也可以使用用于识别面部的其他技术。

[0097] 所述系统可以实施识别相机图像的与用户相关联的部分的过程。用户位置可以包括一个或多个用户的手的位置。或者,用户位置可以包括用户的身体或者用户的面部的位置。用户位置可以包括用户的形状的描述(例如剪影或者轮廓)。用户位置可以包括分割数据(segmentation data),所述分割数据对于相机图像的每一个像素,把像素分类为前景(例如用户的一部分)或者背景,并且用户位置过程可以包括产生被分割的相机图像的过程。

[0098] 被分割的相机图像可以是色度键控(chroma keyed)相机图像。可以获取正站在彩色背景前面的用户的色度键控相机图像。在某些实施方式中,色度键控过程识别相机图像的颜色与背景的颜色匹配的部分,并把那些部分分类为背景。相机图像的颜色偏离背景的颜色的一部分有可能是用户的部分,并且被分类为前景。

[0099] 色度键控过程可以包括产生背景的颜色模型。颜色模型可以定义代表背景的颜色范围。色度键控过程可以包括把图像和颜色模型进行比较,并且如果每一个像素的颜色值落入由颜色模型指示的颜色范围内,则把所述像素分类为背景,否则分类为前景。

[0100] 色度键控过程还可以包括过滤过程,以减少噪声并改变较小的孤立聚类(small isolated cluster)的分类(例如,去除背景中可能被分类为前景的孤立部分,并填充前景中的孔洞)。过滤过程可以包括卷积核(convolution kernel)。

[0101] 被分割的相机图像可以是深度键控相机图像。深度键控相机图像通常由能够确定距离的相机(例如立体相机或飞行时间相机)获取。飞行时间相机的例子包括由Canesta生

产的飞行时间相机,其中红外发射器发射红外光脉冲,并且光学传感器测量光传播到目标并回到相机所花费的时间。飞行时间相机基于测得的时间来计算目标的距离。

[0102] 立体相机的例子包括由Tyzx生产的立体相机,其中,立体相机包括从多个视点捕获场景的图像的多个光学传感器,并且比较所述多个图像以确定在多个视点之间目标位置上的差距。立体相机基于所确定的差距来计算目标的距离。

[0103] 基于在图像的部分中包括的目标的距离,可以把图像的部分分类为前景和背景。分类过程可以包括把距离和阈值进行比较,藉此,如果距离比阈值短,则把该部分分类为前景,并且如果距离比阈值长,则把该部分分类为背景。例如,分类过程可以包括把距离和背景模型进行比较,其中背景模型代表没有用户的场景部分(例如地板和家具)的深度。在这个例子中,如果距离比模型的对应部分短,则分类过程可以把部分分类为前景,并且如果距离等于模型的对应部分或者比模型的对应部分长,则把部分分类为背景。

[0104] 深度键控过程还可以包括过滤过程,以减少噪声和改变较小的孤立聚类的分类(例如,去除背景的可能被分类为前景的孤立部分,并填充前景中的孔洞)。过滤过程可以包括卷积核。

[0105] 被分割的相机图像可以使用背景模型来分割。可以获取正站在任意静态背景前面的用户的色度键控相机图像。相机图像可以和代表背景的预期外观的背景模型进行比较。对于背景的每一个部分(例如像素),背景模型可以包括代表背景的颜色范围。可以通过存储没有用户的场景的图像来获取背景颜色模型。如果相机图像的的部分的颜色和背景颜色模型的对应部分的颜色类似,则该部分可以被分类为背景。如果相机图像的的部分的颜色和背景颜色模型的对应部分的颜色不类似,则该部分可以被分类为前景。

[0106] 背景分割过程还可以包括过滤过程,以减小噪声和改变较小的孤立聚类的分类(例如,去除背景的可能被分类为前景的孤立部分,并填充前景中的孔洞)。过滤过程可以包括卷积核。背景分割过程还可以包括学习过程,用于更新背景模型以补偿背景中的变化或者环境光变化。

[0107] 被分割的相机图像可以被使用皮肤颜色模型来分割。相机图像可以被和代表用户的皮肤的预期外观的皮肤颜色模型进行比较。皮肤颜色模型可以被预先定义,并基于在大量人的图像中被采样的皮肤颜色。在某些实施方式中,如果相机图像的部分和皮肤颜色模型内的颜色类似,则该部分被分类为前景,并且,如果相机图像的部分的颜色和皮肤颜色模型中出现的任何颜色都不类似,则该部分被分类为背景。

[0108] 被分割的相机图像可以是运动图像,藉此,图像的包括运动的部分被分类为前景,并且图像的表现静态的部分被分类为背景。按运动分割相机图像的过程可以包括:获取一系列相机图像,计算连续图像之间的差异,并把所述差异和阈值进行比较。运动分割过程还可以包括过滤过程,以减小噪声和改变较小的孤立聚类的分类(例如,去除背景的可能被分类为前景的孤立部分,并填充前景中的孔洞)。过滤过程可以包括卷积核。

[0109] 运动分割过程还可以包括运动历史过程。在某个例子中,运动历史过程存储对应于每一个像素的最近发生运动的时间的表示。如果运动历史揭示对于每一个像素在一个时间阈值内运动已经发生,则运动历史过程可以把像素分类为前景,否则把像素分类为背景。

[0110] 把代表分割数据的用户位置细化(refine)为对应于用户的手的用户位置的过程可以包括分析被分割的相机图像的前景部分的形状。例如,过程可以识别最上面的(top-

most) 前景部分 (例如像素), 并计算用户位置为最上面的前景部分的位置。或者, 分析被分割的相机图像的前景部分的形状可以包括产生前景部分的轮廓的过程, 识别所述轮廓的形状特征的过程, 以及将形状特征识别为手的过程。用户位置可以被计算为手位置。

[0111] 确定对应于用户的手的用户位置的过程可以包括分析相机图像以识别用户的手, 并确定手的位置。识别相机图像内的手位置的示范性方法把相机图像的部分和代表用户皮肤的预期外观的皮肤颜色模型进行比较。如果相机图像的部分的颜色和皮肤颜色模型内的颜色类似, 则该部分可以被分类为皮肤。被分类为皮肤的部分可以被聚类到部分的聚类中, 并且总体位置和尺寸满足一个或多个标准的部分的聚类被分类为手。本公开并不局限于特定的识别相机图像内的手位置的方法, 也可以采用其他的手追踪方法。

[0112] 参考图4, 当姿态被识别时, 提供包括控件和用户的表示的用户接口 (S404)。可以显示用户的表示以最小化与接口交互作用所要求的训练和技能。用户的表示可以协助用户轻松地识别相对于屏幕上项目位置的其手位置, 并轻松地移动其手位置以便和项目位置重合 (coincident)。用户的表示可以是例如化身或者鼠标光标。

[0113] 当提供了用户接口时, 基于所识别的姿态使所述表示与控件交互作用 (S406)。例如, 可以选择邻近用户表示 (例如邻近鼠标光标, 或者邻近化身的手) 的按钮或者项目。

[0114] 基于所述表示和用户接口控件的交互作用, 电信会话被控制 (S408), 从而结束过程400 (S410)。例如, 仅举几例, 可以接受或者拒绝呼入呼叫, 可以播放语音邮件消息, 或者可以拨号电话号码或者联系人。

[0115] 图5示出了示例用户接口500。用户接口500包括用户表示501。用户接口500可以包括一组前景图形, 例如项目502、文本503和按钮504、505、506和507的表示, 它们中的一些可以被呈现在用户表示501的上面 (over), 部分地挡住 (occluding) 用户表示501。

[0116] 显示用户接口500可以包括把相机图像和第二图像混和 (blending)。第二图像可以包括由应用程序产生的图像、视频游戏图像、电视图像, 等等。对于用户接口的每一个像素, 混和相机图像可以包括把相机图像像素值和第二图像像素值组合 (combining)。如下所示, 等式 (1) 可以用于把相机图像与第二图像混和, 其中D表示结果, C表示相机图像, S表示第二图像, x和y参照图像内的像素位置, 并且 α 表示范围0到1的混和比例:

$$D(x, y) = \alpha \cdot C(x, y) + (1 - \alpha) \cdot S(x, y) \quad (1)$$

[0118] 图6示出了包括用户表示601的用户接口600的另一个例子。显示用户的表示的实施方式可以包括显示分割的图像, 其中所述图像被分割为前景和背景部分。在某些实施方式中, 有可能是用户表示的一部分的图像的部分 (例如像素、区域、块、点或者区) 被分类为前景, 并且不可能是用户表示的一部分的部分被分类为背景。

[0119] 显示被分割的图像可以包括只呈现被分类为前景的部分。只呈现被分类为前景的部分的实施方式可以包括产生表示不透明度的阿尔法通道 (alpha channel), 其中前景部分被定义为不透明的, 并且背景部分被定义为透明的。在这个实施方式中, 用户表示601可以被使用阿尔法通道和一组背景图形组合, 以使用户表示601挡住 (例如在区域602中) 背景图形603。包括项目604的表示、文本606和按钮608、610、612和614的一组前景图形可以被呈现在组合的用户表示601和背景图形的上面, 部分地挡住用户表示601和背景图像。额外的前景和背景图形可以由应用程序产生。

[0120] 图7示出了包括用户表示701的用户接口700的又一个例子。显示用户的表示的实

施方式可以包括显示计算机产生的图像,例如计算机鼠标光标或者手的图形表示。如图7中所示,计算机产生的图像701的位置可以对应于用户位置702,并且可以是用户的表示。

[0121] 显示用户的表示的实施方式可以包括显示呼出的视频的预览(例如在视频会议呼叫中)。可以显示其他的用户表示以帮助用户和接口交互作用。可以显示用户的运动的表示(例如过去的运动或者运动历史)来帮助用户和接口交互作用。

[0122] 所述系统可以实施检测用户何时触摸例如按钮的前景图形的过程。当用户的表示、或者所述表示的一部分占用、指向或者邻近用户接口内和被前景图形占用的位置重合的位置时,用户可以触摸前景图形。

[0123] 图8示出了触摸图形的用户的表示的例子。确定用户触摸的过程可以包括分析被分割的相机图形,其中相机图形被分割为前景和背景部分。参考图8,图像的可能表示用户的部分(例如像素)可以被分类为前景803,并且不可能表示用户的部分(例如像素)可以被分类为背景。可以定义区801以使区801表示例如按钮802的交互作用前景图形的尺寸和位置。所述过程可以确定被分割的相机图像在区内(例如在区801内)的部分(例如像素),如阴影区域804所示。使用被分割的相机图像确定触摸的过程可以包括对区801内的前景部分(例如像素)的数量计数,把计数和阈值进行比较,并且如果所述计数大于阈值,则把该区分类为被触摸。

[0124] 在某些实施方式中,使用被分割的相机图像确定用户位置的过程可以包括在一个时间段内累积对区801内前景部分(例如像素)的数量的计数。对于每一个视频帧,可以从累积的计数去除预先定义的数量,到最小值零,并且当前计数可以被添加到累积计数器,到最大值阈值。当累积计数达到阈值时,所述区可以被分类为被触摸。前景图形可以包括相对于阈值的累积计数的表示,以便向用户提供视觉反馈。例如,图形可以包括条形图或者饼形图805,或者可以用取决于累积计数的颜色或者尺寸来呈现。

[0125] 图9示出了触摸图形的用户的表示的另一个例子。确定用户触摸的过程可以包括把用户位置和区进行比较。参考图9,可以定义区901以使区901代表例如按钮902的交互作用前景图形的尺寸和位置。使用用户位置确定触摸的过程可以包括如果用户位置903在区901内,则把区901分类为被触摸。

[0126] 在某些实施方式中,使用被分割的相机图像确定用户位置的过程可以包括对视频帧的数量或者用户位置处于区内的时间量进行计数。对于每一个视频帧,如果确定用户位置处于区内,则计数可以被增大(到最大值阈值),并且如果用户位置不在区内,则计数可以被减小(到最小值零)。当累积计数达到阈值时,所述区可以被分类为被触摸。前景图形可以包括相对于阈值累积计数的表示,以便向用户提供视觉反馈。例如,图形可以包括条形图或者饼形图905,或者可以用取决于累积计数的颜色或者尺寸来呈现。

[0127] 所述系统可以实施协助用户选择项目的增强的用户接口过程。项目可以包括文本和数字,藉此,增强的用户接口过程协助输入文本和数字。

[0128] 图10示出了包括对象1002和控件1005的用户接口1001,对象1002例如图标或者化身(描绘为虚线圆),控件1005还可以包括引导线1004。尽管在图10中引导线1004示出为虚线,但是在其他的示例实施方式中,引导线1004被描绘为实线,或者根本不被描绘。

[0129] 控件1005允许直观选择或者输出例如字母1006的项目,同时允许在邻近、靠近、部分或者完全被引导线1004限制或者包围或者限定的区中显示对象1002的一部分或者整体。

当确定项目要被输出或者选择时,例如当用户选择调用文本数据的录入的输入字段时,控件1005被定位在用户接口1001中允许显示项目1006和对象1002的位置,因而被相对于、关于、基于或者针对于对象1002定义。换句话说,对象1002的位置可以表示锚定位置,引导线1004和控件1005的其他元素可以关于所述锚定位置来取向或者对准,或者,其他元素可以对于该锚定位置来定位。

[0130] 为了确定这个锚定位置,确定定义引导线1004的对象和该对象的位置。在某些情况下,可以预先确定或者预先定义对象,例如控件1005被结合化身、用户图形、把手、图标、列表、数据表、数据图、文本录入字段、另一个控件或者部件,或者用户接口1001的已知空白区输出的情况,或者控件1005包括它们的情况。在其他情况下,对象在控件1005要被输出时被动态地确定,例如用户跳转到 (tabs to) 文本字段并且该文本字段被用作对象的情况,用户接口1001的空白区被基于用户1001的当前状态定位的情况,屏幕上最大的、突出的、最绚丽的 (colorful) 或最不绚丽的对象被动态地确定为所述对象的情况,或者聚焦的 (in focus) 元素、区或者窗口被动态确定为所述对象的情况。

[0131] 在这些和其他情况下,检测对象1002的尺寸、形状、位置、边界或者其他上下文,并且引导线1004被以与这些检测的上下文的关系来定义。例如,引导线1004可以被定义成和对象或者对象的一部分具有覆盖 (overlapping) 或者非覆盖关系、二等分关系、分开关系、空间约束或者受限关系,或者任何其他关系,例如基于对象1002的尺寸、形状、比例或者解剖模型的关系。

[0132] 总之,并且除此之外,用户接口1001包括可以是用户的表示的对象1002和构成项目集合的项目1006。可以以动态方式显示项目1006,以使项目1006被用允许用户方便并且可靠地选择每一个项目的尺寸和位置显示。由于本不可能把项目1006中的每一个均以大尺寸安排在用户接口1001内或者和引导线1004对准,可以以该大尺寸呈现项目1006的子集。

[0133] 图11到图15示出了示范性的从沿着引导线对齐的项目集合中选择特定项目。在图11中,项目集合1101被沿着引导线1102对齐。被映射到并描绘为化身的手的光标1104向着关注的项目1101r移动。

[0134] 在图12中,光标1204的位置和引导线1202交叉。引导线1202的交叉可以启动进一步的检测过程,该检测过程基于每一个项目和光标1204之间的距离选择或者识别项目1201的子集1201a。在相机图形中检测到光标1204时,基于用户的手在用户身体周围空间中的位置来确定光标1204的位置。

[0135] 项目1201的子集1201a被以更大的字体尺寸或者比例显示,以协助用户更容易或者更直观的选择。选择项目1201的子集1201a和放大项目1201的子集1201a的显示可以响应于检测到光标1204已经和引导线1202交叉而发生,或者可以和光标1204相对于引导线1202的位置无关地发生。在检测到光标1204在基位置 (base position) 1206和引导线1202交叉或者检测到项目已被选择时,可以产生用户反馈,包括例如声音、影像、闪光,和/或例如振动的触觉输出。

[0136] 在图13中,用户选择关注的项目1301r。突出显示 (highlight) 项目的子集1301允许用户选择包含关注的项目1301r的一般区域,在该区域内对项目“放大”,并且可靠和方便地选择关注的表示字母“R”的项目1301r。选择可以使用鼠标事件、键盘或者小键盘敲击、姿态识别、基于相机的输入或者通过很多其他的方法发生。

[0137] 可以以很多方式检测代表光标1304已经和引导线1302交叉的位置的基位置1306的位置。例如,如图14中所示,基位置1406的位置可以被确定为在引导线1401上方 (above) 检测到光标1404后的一个时间所观察到的、引导线1401上最靠近光标1404,或者最靠近被突出显示的例如项目1401r的项目的位置。

[0138] 也可以使用其他方法检测基位置。例如,可以在光标1404和引导线1401交叉时检测基位置1406的位置,或者使用在和引导线1401交叉之前或者之后时的光标1404的位置。例如,图15示出了表示在和引导线1502交叉之前观察到的光标位置的端点(endpoint) 1508和表示在和引导线1502交叉以后时观察到的光标位置的端点1510。基位置1506可以被确定为由端点1508和端点1510限定的线段1504和引导线1501的交叉点。

[0139] 回到图12,突出显示项目1201的子集1201a可以包括沿着引导线1202确定对应于构成子集1201a的项目的位置。在某些实施方式中,位于基位置1206附近的项目被选择为要突出显示的项目1201的子集1201a的一部分(例如被以大尺寸显示),以使靠近基位置1206的项目保持在或者靠近其原始的、非突出显示的位置,并且更远离基位置1206的项目朝外移动以容纳项目1201的子集1201a的尺寸上的增加。

[0140] 下面的等式(2)可以用来确定在子集1201a已被突出显示以后不在子集1201a内的项目的位置。

$$[0141] \quad X_i' = X_b + (X_i - X_b) \cdot \left(\frac{S_i'}{S_i} \right) \quad (2)$$

[0142] 在等式(2)中, X_i 表示在原始状态中沿着引导线的项目*i*的位置, X_i' 表示在放大状态中沿着引导线的项目*i*的位置, X_b 表示沿着引导线的基位置, S_i 表示在原始状态中项目*i*的基本尺寸,并且 S_i' 表示在放大状态中项目*i*的尺寸。

[0143] 图16示出了在突出显示项目的子集之前的第一状态1601中和在突出显示项目的子集之后的第二状态1602中的项目的子集。例如,如果光标1604起初在项目“S”下和引导线1605交叉(以使基位置1606和项目“S”重合),则项目“S”保持在其原始位置,并且项目“R”相对于其原始位置朝左位移了距离1607。因此第二状态1602演示了突出显示后项目的按比例缩放尺寸和位置。

[0144] 图17示出了在突出显示项目的子集之前的第一状态1701中和在突出显示项目的子集之后的第二状态1702中的项目的子集。例如,如果光标1704起初在项目“Q”下和引导线1705交叉(以使基位置1706和项目“Q”重合),则项目“Q”将保持在其原始位置,并且项目“R”相对于其原始位置朝右位移了距离1707。因此第二状态1702演示了突出显示后按比例缩放的项目。因此,用来选择特定项目的光标1704的位置可以取决于起初光标1704和引导线1705交叉的位置。

[0145] 图18示出了在与项目子集的突出显示相关联的状态1801到1804中的项目子集。具体来说,图18示出了分别选择和突出显示项目1801a-c的第一到第三子集1806到1808的概观。第二和第三子集1807和1808可以分别根据光标1811相对于第一和第二子集1806和1807的位置定位。

[0146] 在状态1801中,项目1801a反映其原始的非突出显示的尺寸和位置。在状态1802中,项目1801b的第一子集1806已被选择和突出显示。在状态1803中,项目1810c的第二子集1807已经被选择和突出显示。在状态1804中,项目1801d的第三子集1808已经被选择和突出

显示。

[0147] 在状态1802中,光标1811a起初在项目“S”下和引导线1812a交叉(以使基位置和项目“S”重合),项目“S”保持在其原始位置中,并且周围的字母被从其原始位置朝外位移。从状态1802过渡到状态1803,如果光标1811b移动到右边,则选择在光标1811c的距离内的项目1810c的第二子集1807。在第三状态1803中,如果光标1811c沿着引导线1812c移动到和被放大的项目“T”重合,则项目“T”保持在其放大的位置,并且项目“V”被突出显示,沿着引导线1812c朝右进一步移动距离1815。

[0148] 如果对于某些项目在引导线1812c上不存在足够的空间,例如项目“W”到项目“Z”,则项目被“推离”引导线的末端,并且不被显示。从状态1803过渡到状态1804,如果光标1811c继续进一步朝着引导线1812c的右端移动,可能没有足够的空间来显示作为第二子集1807的一部分的额外项目,并且第三子集1808可以被形成(作为第二子集1807的子集)。

[0149] 为了选择在第二子集1807右边位置上的项目,例如项目“U”,或者已经被“推离”引导线末端的项目,例如项目“W”到项目“Z”,用户可以把引导线1812c和光标1811c重新交叉,并且和引导线1812c第三次交叉以建立更靠近期望项目的新的基位置。此外,不是把项目“推离”到引导线1812c的右端,可以转而“推离”在引导线左端的项目以容纳否则将被“推离”右端的项目的显示。在某些实施方式中,项目可以根据引导线上可用的空间而在尺寸上被减小,以使所有项目都在引导线上显示,而不是“推离”项目。

[0150] 在其他的实施方式中,被“推离”引导线的右端的项目可以在引导线的左端上重新出现。在利用连续的(例如环形的)引导线的实施方式中,可以绕着引导线推动项目。因此,在用户可以绕着连续的引导线顺时针地移动光标的环形引导线的例子中,被从当前的项目子集排除的项目可以以小于光标的角速度顺时针流动(因为它们在增长的项目被添加到子集时要为增长的项目让位)。在这个例子中,对于项目绕着引导线的一圈,光标可以绕着引导线移动多圈。

[0151] 可以使用滚动来协助选择否则将被推离引导线末端的项目。例如,滚动可以包括检测光标1811c是否在引导线1812c末端的预先定义的距离内,并且将速度施加于项目位置。在相对于基位置计算项目位置(见上面等式(1))的情况下,速度可以被施加于基位置,并且项目可以被相应地移动。

[0152] 图19示出了示范性速度函数,其中,水平轴1901表示沿着引导线的位置,并且垂直轴1902表示速度。使用这个速度函数,通过在光标位置靠近引导线的末端时施加速度,项目被移动或者平移,并且当光标位置在引导线中心时,项目不被移动或者平移(因为速度是零)。

[0153] 图20示出了可用来基于光笔位置相对于引导线之间的距离(由水平轴2002反映)按比例缩放项目子集中的项目的尺寸(由垂直轴2001反映)的示范性按比例缩放函数。如曲线2004所表示的,项目的尺寸(S_i')因此可以是光标的当前位置的函数。和非突出显示项目相关联的或者如果光标还未和引导线交叉,则和所有项目相关联的基尺寸(S_i)由参考点2005确定,并且线2006定义了和被突出显示的项目相关联的最大尺寸。

[0154] 在一个示例实施方式中,靠近点2005曲线2004的斜率近似1:1,所以随着对光标的靠近减小,项目尺寸显得线性地并且成比例地增长。靠近最大尺寸,项目的增长可以开始逐渐减小,以便随着光标变得越来越多地靠近被突出显示的项目,产生审美愉悦的过渡。

[0155] 图21和图22示出了例如当光标2101和引导线2102交叉时由增强的控件使用按比例缩放特征。高度2104表示光标2101的位置和引导线2102之间的距离,并且高度2105表示子集2107中的项目的尺寸,例如项目2106r。子集2107中的项目的高度2105被基于高度2104按比例缩放,使项目在光标2101已和引导线2102以小量交叉的图21中比在光标2201已和引导线2202以更小量交叉的图22显得更大。和上面一样,可以使用基于相机的输入来确定光标2101的位置。

[0156] 图23和图24示出了示范性引导线,其中,项目的子集被突出显示。具体来说,图23示出了示范性引导线2301,其中项目2304的子集2302被以放大方式显示。选择项目2304的子集2302可以包括选择预先定义数量的项目,或者基于项目的尺寸动态地选择许多项目包括在子集2302内。

[0157] 包括在子集2302中的项目的数量可以被动态地选择,以使子集2302的项目跨过引导线2301的整个长度2307,或者,它们可以跨过引导线2301的一部分。如图23中所示,被从子集2302排除的项目2305a和项目2305b也可以被沿着引导线2301显示。子集2302中的项目的数量也可以基于子集内的项目的尺寸变化,例如,子集2107(图21中)显示五个项目,而子集2207(图22中)显示三个项目,尽管子集2107的宽度和子集2207的宽度相同。

[0158] 通过以大尺寸显示子集的项目来突出显示项目可以包括以相同的大尺寸显示所有的被突出显示的项目,如图23中所示,或者,可以通过以取决于其沿着引导线相对于光标位置的位置的尺寸显示子集内的各个项目,如图24中所示。

[0159] 图24示出了示范性引导线2401,其中,项目2404的子集2402被以变化的项目尺寸显示。例如,可以把在子集2402的末端的项目2405p和项目2405t(分别表示字母“P”和“T”)的尺寸定得比在子集2402的中心的项目,例如项目2405r(表示字母“R”)小。在子集2402中以变化的尺寸显示项目可以产生愉悦的审美外观,并且可以使对增强的控件的使用更为直观。如图24中所示,被从子集2402排除的项目2405a和项目2405b也可以被沿着引导线2401显示。

[0160] 以大尺寸显示子集2302和2402的项目可以包括给项目赋予动画。动画可以包括放大子集的项目,并在较短的时间段上沿着引导线平移项目的位置(例如保持项目垂直于引导线)。被从子集2302和2402排除的项目可以被赋予动画以便在尺寸上收缩,并沿着引导线朝外移动,从而给子集2302和子集2402“让位”。

[0161] 被从引导线的末端“推离”的项目可以简单地消失,或者可以被赋予动画以便从引导线的边缘离开或者被以视觉上刺激或者滑稽的方式破坏,例如被燃烧、内爆、蒸发、爆炸、液化、破碎,或者其他的技术。类似地,因在引导线上空间被让出而再次出现的先前被“推离”的项目可以简单地再次出现,或者可以被赋予动画以便从用户接口的顶部落回到引导线上或者被以视觉上刺激或者滑稽的方式自然地产生。

[0162] 在引导线不连续的情况下,项目可以被赋予动画以跨过不连续性移动。项目可以被赋予动画以便以高速度跨过间隙移动,或者可以被赋予动画以便使用上面的视觉效果中的任何一个被“推离”或者“再次出现”。同样地,被“推离”引导线的一端以在相反端再次出现的项目可以被赋予动画以便在引导线端点之间高速移动,或者可以被赋予动画以便使用上面的视觉效果中的任何一个被“推离”或者“再次出现”。

[0163] 图25描绘了沿着引导线的项目的激活,其中“激活”或者“突出显示”一般涉及确定

要选择的项目。确定要选择的项目的系统可以使用迟滞。选择可以包括当光标最初和引导线交叉时确定初始被选择项目,其中所述初始被选择项目可以是其位置在距离上最靠近基位置的项目。此后,为了减小当光标被置于项目之间时在两个邻近项目之间的无意的摇曳,选择可以包括每当光标运动超过了相邻项目之间的中点预先定义的距离就确定新的被选择项目。

[0164] 例如,两个项目2502r和2502s沿着引导线2501的位置由线2504r和2504s指示,项目2502r和项目2502s之间的中点的位置由线2505指示,并且距离2506表示预先定义的距离。例如,如果项目2502s(即字母“S”)是初始被选择项目,用户将沿朝着项目2502r(即字母“R”)的方向把光标移动到线2507左面以选择项目2502r,线2507是超过中点2505的预先定义的距离。如果随后选择了项目2502r,则为了重新选择项目2502s,用户将把光标沿朝着项目2502s的方向移动到线2509的右面,线2509是超过中点2505的预先定义的距离2506。

[0165] 可以基于预期光标在用户试图保持静止时摇曳或者晃动的期望最大距离定义所述预先定义的距离,补偿由无意识的身体震颤或者输入单元的限制导致的用户移动的影响。当光标的位置在引导线以下时,或者当光标未被找到或者不在特定用户接口内或者由增强的控件占据的用户接口区内时,项目可以被解除选择。项目的选择还可以产生用户反馈,例如包括声音、影像,和/或例如震动的触觉输出。

[0166] 显示子集的项目可以包括显示项目以使其外观提供关于其选择状态的用户反馈。例如,被选择的项目可以被以独特的颜色显示,或者以例如鲜艳外观的视觉效果显示。当项目处于被选择状态比第一预先定义的持续时间阈值长时,当前选择的项目可以被激活或者突出显示。就此而言,当光标被保持在项目上方一段时间后,项目被激活。如果被选择的项目保持被选择比第二预先定义的持续时间阈值长,则激活可以被重复或者取消。

[0167] 而且,当光标的位置静止了一段时间后,项目可以被激活或者突出显示。当平行于引导线的位置分量中的变化小于预先定义的距离阈值超过了预先定义的时间阈值时,光标可以被分类为静止。例如,最靠近光标的位置的项目被识别和激活。如果光标保持被分类为静止比第二预先定义的持续时间阈值长,则激活可以被重复或者取消。此外,项目可以被基于光标位置和相对于引导线之间的距离激活。例如,当这样的距离超过了预先定义的距离阈值时,项目可以被激活。

[0168] 在其他的例子中,在选择项目后,可以由另一类型的用户输入导致项目的激活。例如,用户可以提供另一类型的用户输入来激活被选择的项目。在这些例子中,为了激活被选择的项目,用户可以触摸用户输入按钮(例如在控制器上),提供可听输入(例如说“激活”),执行另一类型的姿态(例如,朝显示器移动用来选择项目的手,或者把用户的另一只手移动到用户接口的另一位置来激活被选择的项目),或者提供任何其他类型的用户输入。

[0169] 图26到图29示出了示范性项目集合。在图26中,包括来自英文字母表的字符的项目2601包括项目2602,项目2602在被激活或被选择时,打开第二个项目集合。当被选择时,第二个项目集合可以沿着引导线2604和项目2601或者项目2601的一部分一起出现,或者第二个项目集合可以在引导线2604上替换项目2601。从用来表示项目2602的符号,用户可以直观地确定第二个项目集合包括数字。

[0170] 图27示出了与引导线2702对准的项目2701。当选择图26中的项目2602时,显示项目2701。一旦选择了第二个项目集合,在引导线2702下面去除光标或者从显示项目2701的

用户接口去除光标可以让项目2601被重新选择或者重新激活。项目2701包括项目2704,项目2704在被激活时重新打开、重新激活或者重新选择项目2601。从用来表示项目2704的符号,用户可以直观地确定项目2601包括来自英文字母表的字符。

[0171] 图28示出了与引导线2803对准的项目2801,其中,项目均表示被已知与标准电话小键盘上的键相关联的数字和字母的组合。图29示出了项目2901,当项目2801的项目2802被选择时(见图28)显示项目2901,并且项目2901包括与标准电话小键盘上的键数字“7”相关联的字符的组合。

[0172] 使用在图26到图29中示出的项目,可以提供基于相机的文本输入机制,例如通过顺序地把形成单词的字母填入文本字段中。在激活数字、字母、文本符号或者预先定义的文本后,所述数字、字母、文本符号或者预先定义的文本可以被附加到已经被输入的字符串。或者,项目的激活(例如来自图26中的项目2601的项目)可以显示进一步的项目,并且进一步的项目的激活可以把被激活的字符附加到字符串。所述项目之一可以是退格项目,用于协助从项目串去除最近的项目。

[0173] 基于相机的文本输入机制可以包括组合字符来形成复合字符。用于日文文本输入的文本输入机制可以包括组合假名字符来形成日文汉字字符(用日文个人计算机键盘的用户熟悉的方法)。第一个项目集合可以包括表示假名的项目。当假名项目被激活时,对应的假名字符被附加到字符串。项目集合可以包括表示操作的项目,所述项目在被激活时,激活把最近的假名串字符转换为日文汉字的过程。把最近的串的假名字符转换为日文汉字的过程可以包括显示第二个候选日文汉字项目集合。激活日文汉字项目激活了用所激活的日文汉字替换最近的串的假名字符的过程。这种基于相机的文本输入方法可以被扩展到其他语言。

[0174] 用于日文文本输入的替换方法可以包括显示表示假名的第一个项目集合。当假名项目被激活时,显示第二个项目集合。第二个项目集合可以包括日文汉字,所述被激活的假名形成了所述日文汉字的一部分。这个集合可以大于上面描述的第二集合。这种文本输入方法可以扩展到其他语言。

[0175] 一种文本输入机制可以在项目集合中包括确认项目。当被激活时,确认项目激活一个过程,藉此将字符串提供给应用程序。此外,文本输入机制可以包括推测文本完成过程。推测文本完成过程可以搜索字典以找到包含字符串的字符的最可能的文本。可以在输出字段中显示所述最可能的文本。可以基于文本输入机制被使用的上下文选择在推测文本完成过程中使用的字典。例如,字典可以包括当文本输入机制被用来输入名称时的名称。

[0176] 取决于应用程序的状态,应用程序过程可以确定用于用户接口的项目。激活项目可以向应用程序过程提供消息。可以基于所述消息控制应用程序过程。

[0177] 图30示出了系统状态的例子。系统可以包括在图30中所示的各种状态。系统可以包括下列状态:待机状态3001,其中系统等待呼入的呼叫或者用户来激活所述系统;通知状态3002,其中,系统向用户通知呼入的呼叫并等待用户来激活系统;呼叫状态3003,其中系统协助与另一系统的音频或者视频通信;菜单状态3004,其中系统显示选项和功能供用户选择;直接拨号状态3005,其中系统允许用户输入电话号码或者地址;重新拨号状态3006,其中系统允许用户选择和近来呼叫相关联的联系人;搜索状态3007,其中系统允许用户输入名称,并且匹配的联系人被在数据库中找到;浏览状态3008,其中系统允许用户浏览数据

库中的联系人;连接选项和确认状态3009,其中系统显示关于联系人的信息并确认用户呼叫该联系人的意图;连接状态3010,其中系统和远程系统建立连接;消息菜单状态3011,其中系统允许用户选择消息来查看;消息回放状态3012,其中系统播放消息;消息静音状态3013,其中呼叫被静音;和,配置状态3014,其中用户可以配置所述系统,例如注册其面部。所述系统可以包括更少的状态,或者未被列出的额外的状态。系统可以以不同于图30中所示的方式或者顺序输入各种状态。

[0178] 系统包括待机状态3001。系统在处于待机状态3001中时可以是不显眼的(inconspicuous),以使系统不响应于偶发的用户移动或者在显示设备104上不显示任何不必要的东西。当系统处于待机时,这可以允许显示器被其他应用程序使用,或者用于其他目的(例如看电视或者玩儿视频游戏)。

[0179] 当处于待机状态3001中时,系统可以针对呼入的呼叫监控网络。当接收到呼入的呼叫或者其他的消息时,系统可以进入通知状态3002。

[0180] 系统可以被配置成在接收到呼入的呼叫时把控制信号发送到显示设备。控制信号可以打开显示设备,并选择输入设置,以使所述显示设备显示由系统产生的图像。系统可以被配置成在接收到呼入的呼叫时把控制信号发送到其他的应用程序或者设备。例如,当接收到呼叫时,控制信号可以暂停应用程序或者程序执行(例如在接收到呼叫时暂停正玩儿的视频游戏或者正观看的电视节目)。举几个例子来说,控制信号可以是红外信号、直接的有线连接上的电子信号或者网络消息。

[0181] 系统可以被配置成如果面部检测过程检测到面部则进入通知状态3002。这可以允许系统避免在不存在用户时打开显示设备。

[0182] 系统可以被配置成接受预期用于多个用户的呼叫。系统可以以识别对应于呼叫所预期用户的用户面部的面部识别过程为条件进入通知状态3002。这可以允许系统避免在预期的接收者不在场时打断其他用户的活动。这也可以帮助确保预期的接收者的隐私受到保护。

[0183] 当检测到进入的通信时,系统可以进入通知状态3002。当处于通知状态3002中时,系统可以产生在显示设备上显示的视觉提示。

[0184] 图31描绘了通知用户接口3101的例子。视觉提示3102可以叠加在由其他的应用程序或者其他的设备(例如电视节目或者视频游戏,如图31中所示)产生的其他的视频内容3103上。当处于通知状态3002中时,系统也可以产生在音频扬声器上产生的音频提示。音频提示可以和由其他的应用程序或者其他的设备,例如电视节目或者视频游戏产生的其他音频内容混和。

[0185] 当处于待机状态3001中和通知状态3002中时,系统可以针对接合手姿态监控由相机捕获的图像。处理器可以检测接合手姿态以便用户可以通过执行接合手姿态与系统接合。接合手姿态可以是不可能被用户意外地执行的姿态,并且足够简单,以至于其可以被无需在显示设备上显示用户的表示地执行。

[0186] 图32示出了包括执行接合手姿态的用户的相机图形的例子。如图32中所示,示例接合手姿态包括按垂直运动向上抬起手,直到其在用户的头上方为止。这个姿态足够不寻常以至于其不可能被用户意外地或者无意识地执行。用户可以认为这个姿态是直观的并且容易记住,因为它可以是学校中学生抬起他的手来“吸引”(engage)老师的回忆。

[0187] 接合姿态可以是在显示器前相机的追踪区中做出的被保持预先确定的时间量的特定手姿势或者手运动序列。一个示例姿态是一种保持处于竖直位置，所有的手指和拇指分散得很开的手姿势。另一个例子是挥手运动，或者通过在用户的面部前伸出其上臂，并在其头部前按圆圈移动其上臂做出的环形手运动。本质上，接合姿态向产生用户接口的设备指明了用户为发生进一步的基于相机的输入做好了准备。为了减少错误，接合姿态可以是非典型姿态，例如在正常对话其间将不会利用身体语言下意识地做出的姿态，或者将不会在普通的正常人类活动的进行中做出的姿态。

[0188] 手姿态识别过程可以在系统处于待机状态3001和/或通知状态3002中时识别接合手姿态。参考图30，在系统在处于待机状态3001中时，当检测到接合手姿态时可以进入菜单状态3004。在系统处于通知状态3002中时，当检测到接合手姿态时可以呼入呼叫状态3003。

[0189] 系统可以被配置成使得只有呼叫的预期接收者或者系统的授权用户才可以回答呼叫。手姿态识别过程可以检测接合手姿态，并且辨别识别过程可以识别执行接合手姿态的用户。在这个例子中，如果，或者只有执行接合手姿态的用户被面部识别过程识别为呼叫的预期接收者或者系统的授权用户，系统才可以呼入呼叫状态3003。

[0190] 可以定义解除接合手姿态。解除接合手姿态可以是不可能被用户意外地执行的姿态。图33示出了包括执行解除接合手姿态的用户的相机图像的例子。如图33中所示，解除接合手姿态可以包括水平地移动手。

[0191] 系统可以被配置成当检测到解除接合手姿态时静音通知消息。当处于通知状态3002时，由处理器执行的手姿态识别过程可以针对解除接合手姿态监控由相机捕获的图像。如果处理器检测到解除接合手姿态，则处理器可以指挥系统返回待机状态3001。

[0192] 在某些实施方式中，系统可以被配置成当检测到解除接合手姿态时断开呼叫。当处于呼叫状态3003中时，由处理器执行的手姿态识别过程可以针对解除接合手姿态监控由相机捕获的图像。如果处理器检测到解除接合手姿态，则处理器可以指挥系统返回待机状态3001。

[0193] 也可以定义静音手姿态。手姿态识别过程可以针对静音手姿态监控由相机捕获的图像，并且如果检测到静音手姿态则进入静音状态3013。

[0194] 系统可以被配置成聚集在用户上。聚集在用户上可以包括对相机摇全景和变焦，以使用户的面部出现在相机图像的中心，并且处于规定的尺寸。聚集在用户上可以包括剪裁和缩放相机图像，以便在相机图像中用户的面部出现在中心。可以定义规定的尺寸以便贯穿运动范围，相机图像包括用户的上臂。也可以定义规定的尺寸，以使用户能够到达相机图像边缘的预先定义的距离内，从而保证交互作用项目和按钮（例如图5的项目502、504、505、506和507）在用户所及之内。

[0195] 系统也可以被配置成基于用户的参考位置聚集在用户上。用户参考位置可以由面部检测过程确定。例如，用户参考位置可以被确定为最靠近检测到接合手姿态的位置的面部位置。

[0196] 图34示出了包括坐在会议桌边的多个人的相机图像的例子。坐在会议桌边的人可能参与视频会议会话。参考图34，数字3401、3402、3403、3404和3405指示由面部检测过程确定的面部位置。接合手姿态的位置由数字3406指示。靠近面部位置（3404）被选择作为用户参考位置。靠近面部位置可以被作为具有面部位置和姿态位置之间的距离的水平分量

(3407)的最小绝对值的面部位置和姿态位置对计算。孤立距离的水平分量可能是有益的，因为可以预计在接合手姿态期间手位置在头上方。用户参考位置也可以由用户位置过程确定，在用户位置过程中，被分割的图像的前景部分的面心被确定是用户位置。

[0197] 系统可以追踪用户参考位置以使相机在用户移动时保持聚集在用户上(例如相机图像跟随用户)。相机摇全景和变焦可以帮助保证用户保持在被传送的图像内(例如在视频会议期间)。相机摇全景和变焦也可以帮助保证可以在显示设备上显示的按钮保持在用户的轻松可及的范围内。

[0198] 对相机摇全景和变焦可以包括把控制信息传送到被电动致动器。电动致动器可以摇相机的视点，并且操纵相机光学器件对相机图像变焦距。聚集在用户上可以包括使用图像处理或者其他的计算技术来模拟摇全景变焦功能。例如，聚集在用户上可以包括通过剪裁和缩放相机图像来模拟摇全景变焦功能。

[0199] 图35和图36示出了通过剪裁和缩放相机图像聚焦在用户上的例子。参考图35，剪裁相机图像3501可以包括选择包括用户参考位置3503的相机图像的一部分3502。类似的，参考图36，剪裁相机图像3601可以包括选择包括用户参考位置3603的相机图像的一部分3602。缩放相机图像可以包括通过放大或者抽选产生图像，其中所产生的图像的尺寸可以和该部分的尺寸不同。

[0200] 系统还可以包括基于用户参考位置定位音频以聚焦在用户上。麦克风可以包括多个传感器(例如多个麦克风)。声音定位过程可以增大起源于和用户参考位置对应的方向的声音的灵敏度，并减小起源于其他方向的声音的灵敏度。例如，执行声音定位过程的处理器可以关闭位于远离用户参考位置的麦克风或者传感器(例如，处理器可以保留最靠近的麦克风或者传感器开启)。声音定位过程可以利用波束形成过程，藉此比较麦克风阵列的每一个传感器接收的信号的相位和幅度。

[0201] 系统可以被配置成在多用户环境(例如会议室)把焦点从第一个用户改变到第二个用户。如上所述，第一个用户可以执行手姿态(例如接合手姿态)以获取焦点。稍后(但是在同一呼叫期间)，第二个用户可以执行手姿态(例如，接合手姿态，或者焦点转移姿态)来从第一个用户获取焦点。例如，响应于检测到由第二个用户执行的接合手姿态，相机图像可以从第一个用户摇到第二个用户，并且麦克风可以定位第二个用户的声音。

[0202] 图37示出了把焦点从第一个用户改变到第二个用户的相机图像的例子。参考图37的例子(其描绘了和图36的例子类似的图像)，相机图像3701的第一部分3702基于第一用户参考位置3703选择。系统可以包括当处于呼叫状态3003中时操作的手姿态识别过程。在对相机摇全局和变焦包括剪裁和缩放相机图像的实施方式中，手姿态识别过程可以分析整个相机图像3701(例如不仅是部分3702)以识别第二个用户。在检测到第二个用户的接合手姿态以后，相机图像3701的第二部分3704基于第二用户参考位置3705选择。相机图像的一部分(3702或者3704)可以被视频压缩过程压缩，并在网络上传送。这个例子允许多用户环境中的用户通过执行手姿态选择其自身作为主动的发言人。

[0203] 系统还可以包括仲裁过程以厉行对用户何时可以获得焦点的限制。例如，仲裁过程可以通过分析来自第一个用户的方向的定位声音，把该声音分类为语音或非语音，并只有在未检测到第一个用户的语音时才允许第二个用户获得焦点来防止第二个用户打断第一个用户。仲裁过程可以排队用户获得焦点的请求，并在不再检测到第一个用户的语音时

聚焦在下一个排队的用户上。或者,仲裁过程可以排队获得焦点的用户请求,并在姿态识别过程检测到由第一个用户执行的手姿态,从而允许第一个用户指示他或她何时完成并打算放弃焦点时聚焦在下一个排队的用户上。

[0204] 当处于呼叫状态3003中时,系统可以显示被聚焦用户的预览图像。预览图像可以包括被摇全局或变焦的相机图像,或者相机图像的被剪裁和缩放的部分。预览图像可以占据用户接口的一部分。一个或多个远程用户的视频图像可以占据用户接口的其他部分。

[0205] 一些操作,例如接合系统和回答呼入的呼叫,可以使用无需直接的用户反馈的手姿态被执行。但是,用户可以无需直接的用户反馈执行的手姿态的数量和复杂度可能受用户训练限制。某些系统状态中的某些操作可能要求复杂的用户交互作用,例如选择联系人或者拨号号码。当在这些状态中时,系统可以显示用户的表示。显示用户的表示可以帮助用户触摸屏幕上项目,从而减少用户训练要求并产生直观的接口。上面描述了产生用户的表示的过程、把用户位置和应用程序产生的项目比较的过程,以及协助输入文本和数字的增强的用户接口过程。

[0206] 系统可以奥克菜单状态3004。菜单状态3004可以使用户能够选择进入不同的状态或者可以使用户能够改变系统选项或者设置。图38示出了菜单状态用户接口3801的例子。菜单状态用户接口3801包括一组按钮(3802到3807),所述一组按钮在被选择时事的系统进入另一状态。利用每一个按钮上的图标图像或者文本,按钮可以表示每一个状态。

[0207] 菜单状态3004可以包括确定用户的身份的面部识别过程。如果用户未被识别,或者对某些功能的未授权的访问,则状态可以被禁止。例如,可以不允许孩子进入直接拨号状态3005,因为直接拨号状态允许任意联系人被拨号。在菜单状态中可以显示状况(例如,未查看的消息的数量)。

[0208] 系统可以包括直接拨号状态3005。在直接拨号状态3005中,用户可以输入联系人号码或者地址(例如电话号码、电子邮件地址、Skype用户手柄、Xbox实况玩儿家标签,等等)。直接拨号状态3005可以使用协助用户选择项目的增强的用户接口。协助用户选择项目的增强的用户接口过程可以包括表示数字、字母、常用短语(例如“.com"或者本地电话区号)的项目集合和功能(例如退格)。可以基于通信网络(例如,如果到电话网络的连接可用时,项目集合可以包括数字)确定所述项目集合。被输入的字母或者数字可以显示为文本。

[0209] 图39示出了直接拨号状态用户接口3900的例子。表示字母的项目集合被显示为项目的表示3902。通过增强的用户接口过程输入的文本被示出为文本3903。在直接拨号状态3005中,针对退格功能,可以显示按钮(例如按钮3904),或者,可以识别手姿态,以去除近来的数字或者字母。

[0210] 在直接拨号状态3005中,针对拨号功能可以显示按钮(例如按钮3905),或者,可以识别手姿态,这启动了初始化与当前输入的联系人或者电话号码的通信的过程。用于拨号功能的按钮或者手姿态可以被缺省地禁止,并且在确定联系人信息是完整的时(例如正确的电话号码位数)被使能。多个拨号功能可能可用,表示直接拨号状态3005中或者连接选项状态3009中的不同的连接选项(例如只有语音,或者语音和视频)。在选择拨号功能以后,系统可以进入连续选项状态3009。或者,系统可以绕过连接选项状态并直接到达连接状态3010。

[0211] 直接拨号状态3005可以包括自动完成过程。自动完成过程基于部分输入的联系人

号码或者地址针对潜在匹配搜索联系人集合。联系人集合可以包括近来被拨号的联系人集合、错过的呼叫的集合,以及地址簿中的联系人集合。自动完成过程可以包括基于由面部识别过程确定的用户的身份过滤联系人集合。例如,只有由所识别的联系人拨号的联系人,或者所识别的用户的个人地址簿中的联系人可以被包括供考虑作为潜在匹配。这可以增强用户的隐私。在直接拨号状态3005中,自动完成过程潜在匹配可以用将潜在匹配与由用户输入的文本(例如图39中所示的文本3903的灰色字母)区别的视觉样式(例如,彩色)显示。针对接受功能,可以显示按钮(例如按钮3906),或者,手姿态被识别,接受功能接受由自动完成过程指示的联系人,并开始初始化与该联系人的通信的过程。

[0212] 直接拨号状态3005可以包括确定用户的身份的面部识别过程。可以基于用户的身份来限制用户可以输入和拨号的联系人。例如,可以防止未被识别用户拨长途电话号码。

[0213] 直接拨号状态3005可以显示针对退出功能的按钮(例如按钮3907)或识别手姿态,退出功能使系统返回菜单状态3004。或者,退出功能可以返回到待机状态3001。

[0214] 系统可以包括搜索状态3007。在搜索状态中,用户可以输入名称。对于输入的名称,系统可以针对潜在匹配搜索联系人集合。搜索状态可以使用协助用户选择项目的增强的用户接口。增强的用户接口可以包括表示字母、数字和功能(例如退格)的项目集合。被输入的字母或者数字可以被显示为文本。

[0215] 例如,图39中的用户接口3900示出了搜索状态用户接口的例子,其中,用户从项目3902选择字母,并且输入的字母被显示为文本3903。和直接拨号状态3005类似,搜索状态3007可以包括自动完成过程。基于部分输入的联系人名称,或者完整的联系人名称,搜索过程针对潜在匹配搜索联系人集合。联系人集合可以包括近来被拨号的联系人集合、错过的呼叫的集合,以及地址簿中的联系人集合。自动完成过程可以包括基于由面部识别过程确定的用户的身份过滤联系人集合。例如,只有被该用户拨号的联系人,或者在该用户的个人地址簿中的联系人可以被包括供考虑作为潜在匹配。这可以增强用户的隐私。搜索状态3007可以用将潜在匹配与由用户输入的文本(例如图39中所示的文本3903的灰色字母)区别的视觉样式(例如,彩色)显示自动完成过程潜在匹配。针对接受功能,搜索状态3007可以显示按钮(例如按钮3906),或者,识别手姿态,接受功能接受由自动完成过程指示的联系人,并开始初始化与该联系人的通信的过程。

[0216] 针对退出功能,搜索状态3007可以显示按钮(例如按钮3907),或者,识别手姿态,退出功能使系统返回到菜单状态3004。或者,退出功能可以返回待机状态3001。系统可以包括重新拨号状态3006。在重新拨号状态3006中,用户可以选择先前被拨号的联系人。重新拨号状态3006可以列出联系人集合。联系人集合可以包括近来被拨号的联系人、呼出的呼叫、呼入的呼叫和错过的呼叫。

[0217] 重新拨号状态3006可以包括确定用户的身份的面部识别过程。可以基于用户的身份过滤联系人集合。例如,联系人集合可以只包括由所识别的用户先前拨号过的联系人。

[0218] 图40示出了重新拨号用户接口4000的例子。对应于联系人集合的子集,重新拨号状态可以显示一组按钮(例如按钮4001)或者识别手姿态集合。当联系人被选择时,系统可以开始初始化与被选择的联系人的通信的过程。或者,系统可以显示表示与联系人相关联的功能的额外的按钮或者识别额外的手姿态。与联系人相关联的功能可以包括显示关于联系人的信息、选择与联系人通信的方法(例如拨联系人的手机、办公室或者家庭电话号码)、

给联系人发送消息,或者从集合擦除该联系人。这些选项可以在重新拨号状态3006或者连接选项状态3009中展示。在选择了拨号功能以后,系统可以进入连接选项状态3009。或者,系统可以绕过连接选项状态3009并直接到达连接状态3010。

[0219] 重新拨号状态还可以显示滚动联系人子集的按钮(例如按钮4002和按钮4003)或者识别手姿态。当选择了按钮4002和按钮4003其中之一时,可以选择新的联系人子集。滚动允许用户从大量联系人中选择。

[0220] 重新拨号状态3006可以使用协助用户选择项目的增强的用户接口。所述增强的用户接口可以包括表示联系人的集合。图41示出了重新拨号用户接口4100的例子。项目集合4102被以小尺寸显示,所以可以显示大的列表。联系人的大小和位置被动态地调整以便协助选择联系人。项目集合4102可以被相对于引导线4101对齐,其中,引导线4101如图41中所示那样垂直取向。增强的用户接口可以包括表示类别的第一集合。选择第一集合的项目可以显示第二个项目集合,第二集合包括归档在第一被选择类别下的联系人。类别可以包括字母表的字母,并且类别中的项目可以包括以该字母开始的联系人名称。

[0221] 显示联系人可以包括显示联系人的状态。状态可以表示联系人的有效性(例如,该联系人当前是否在线)。针对退出功能,重新拨号状态3006可以显示按钮(例如按钮4104)或者识别手姿态,退出功能可以使系统返回菜单状态3004。或者,退出功能可以返回待机状态3001。

[0222] 系统可以包括浏览状态3008。在浏览状态3008中,用户可以选择先前已知的联系人。浏览状态3008可以列出联系人集合。联系人集合可以包括个人联系人列表中的联系人,或者远程数据库中的联系人(例如,在线电话簿或者公司目录,并且还可以包括人名、企业名称、部门名称等等。)

[0223] 浏览状态3008可以包括确定用户的身份的面部识别过程。可以基于用户的身份过滤联系人集合。例如,联系人集合可以包括所识别的用户的个人地址簿中的联系人。在另一个例子中,如果授权用户未被识别,则联系人集合可以排除数据库中的联系人(例如,可以防止未成年的用户联系面向成人的业务)。

[0224] 浏览状态3008用户接口可以显得和图40中示出的用户接口4000类似。对于联系人集合的子集的每一个联系人,浏览状态3008可以显示按钮或者识别手姿态。当选择了联系人时,系统可以开始初始化与被选择的联系人的通信的过程。或者,系统可以显示表示与联系人相关联的功能的额外的按钮或者识别额外的手姿态。与联系人相关联的功能可以包括显示关于联系人的信息、选择与联系人通信的方法(例如拨联系人的手机、办公室或者家庭电话号码)、给联系人发送消息,或者显示关于该联系人的信息(例如地址或者广告)。这些选项可以在重新拨号状态3006或者连接选项状态3009中展示。在选择了拨号功能以后,系统可以进入连接选项状态3009。或者,系统可以绕过连接选项状态3009并直接到达连接状态3010。

[0225] 浏览状态3008还可以显示滚动联系人子集的一个或多个按钮或者识别手姿态。当滚动按钮被选择或者滚动手姿态被执行时,可以选择新的联系人子集。滚动允许用户从大量联系人中选择。

[0226] 浏览状态3008可以使用协助用户选择项目的增强的用户接口。联系人可以被以小尺寸显示。联系人的大小和位置被动态地调整以便协助选择联系人。增强的用户接口可以

包括表示类别的第一集合。选择第一集合的项目可以显示第二个项目集合，第二集合包括归档在第一被选择类别下的联系人。类别可以包括字母表的字母，并且类别中的项目可以包括以该字母开始的联系人名称。类别可以包括表示和用户的关系的关键词（例如朋友、家庭、业务伙伴，等等）。类别可以包括表示服务的关键词（例如销售、支持、运输，等等）。

[0227] 显示联系人可以包括显示联系人的状态。状态可以表示联系人的有效性（例如，该联系人当前是在线、离线还是请求不被打扰）。状态可以由图标（例如图标4005）表示。连接选项和信息状态3009可以展示多个连接选项（例如只有语音、语音和视频，以及拨联系人的手机、办公室或者家庭电话号码）。连接选项和信息状态3009也可以提供向联系人发送消息或者添加和从联系人集合擦除联系人的选项。连接选项和信息状态3009还可以显示关于联系人的信息。

[0228] 图42示出了连接选项和信息状态用户接口4200的例子。针对每一个连接方法，状态3009可以包括按钮（例如按钮4201、4202和4203）或者识别手姿态。选择连接方法可以使过程前进到连接状态3010。显示连接方法可以包括显示连接方法的状态。状态可以表示通过该方法连接到联系人的有效性（例如，该联系人当前是在线、离线还是请求不被打扰）。状态可以被表示为图标（例如图标4204）。

[0229] 针对消息功能，连接选项状态3009可以包括按钮（例如按钮4205）或者识别手姿态。消息功能可以允许用户记录消息。消息可以被传送到联系人。

[0230] 如果联系人不在用户的联系人列表中，连接选项状态3009可以包括用于把联系人添加到用户的联系人列表的按钮或者识别手姿态。如果联系人在用户的联系人列表中，则该状态可以包括用于把联系人从用户的联系人列表去除的按钮（例如按钮4206）或者识别手姿态。所述状态也可以显示用于退出功能的按钮（例如按钮4207）或者识别手姿态，退出功能使系统返回菜单状态3004。或者，退出功能可以返回待机状态3001。

[0231] 连接选项状态3009可以显示联系人名称4208和关于联系人的信息4209。显示关于联系人的信息可以包括搜索并从数据库检出信息。在连接状态3010中，系统可以包括初始化与被选择的联系人的通信的过程。响应于用户完成输入电话号码或者联系人、启动自动完成过程、选择拨号功能、接受呼入的呼叫等，过程可以前进到连接状态3010。

[0232] 系统可以包括面部识别过程。当建立与远程用户的通信时，呼叫者的身份可以被传送，以使远程用户的系统可以显示正在呼叫的人的身份。在连接状态3010中时，系统可以显示连接状况。连接状况可以指示连接的强度（例如强或者弱），或者可以指示连接的有效性（例如连接或者断开）。

[0233] 消息菜单状态3011允许用户检查消息。消息菜单状态3011可以展示消息列表。消息列表可以显得类似于如图40或者图41中所示的联系人列表。选择消息允许在消息回放状态3012中播放该消息。

[0234] 在消息回放状态3012中播放消息可以包括展示音频或者视频记录。消息回放状态3012可以显示对应于回放操作（例如暂停、快进、后退或者擦除）的一组按钮或者识别一组手姿态。

[0235] 在消息互访状态3012的实施方式中，用户的表示可以挡住正被播放的视频的一部分。用户的表示可以和记录的视频混和。或者，可以在消息状态用户接口的单独部分中显示用户的表示。

[0236] 图43示出了消息回放状态中的用户接口的例子。参考图43的例子,用户接口4300可以包括记录的视频被播放的部分4301。用户接口4300还可以包括用户的表示被显示的部分4302。在部分4302中可以包括按钮(例如按钮4303),提供对回放操作的访问。用户接口还可以包括关于联系人和消息的信息4304(例如日期和时间)。

[0237] 静音状态3013可以在呼叫期间给本地用户提供隐私。静音状态3013可以继续显示远程用户。静音状态3013可以包括继续呼叫的按钮或者识别手姿态,从而重新呼入呼叫状态3003。静音状态3013可以包括结束呼叫的按钮或者识别手姿态,从而到达待机状态3001或者菜单状态3004。静音状态3013可以包括进入直接拨号3005、重新拨号3006、搜索3007或者浏览状态3008的按钮或者识别手姿态,藉此被选择的联系人可以被添加到现有呼叫(例如三方呼叫)。

[0238] 系统可以包括配置状态3014。配置状态3014可以展示通信和网络选项(例如IP地址)。配置状态3014可以展示个人地址簿功能(例如添加、编辑或者删除联系人)。配置状态3014可以包括面部识别学习过程,藉此用户的面部被分析并以该用户的名称存储。所存储的图像可以用作面部识别过程的模板。

[0239] 图44是通用或专用计算机系统4400的例子的示意图。系统4400可以用于结合根据一个实施方式的过程400描述的操作。系统4400包括处理器4410、存储器4420、存储设备4430和输入/输出设备4440。使用系统总线4450互连部件4410、4420、4430和4440中的每一个。处理器4410能够处理用于在系统4400内执行的指令。在一个实施方式中,处理器4410是单线程处理器。在另一个实施方式中,处理器4410是多线程处理器。处理器4410能够处理存储在存储器4420中或者存储设备4430上的指令以便在输入/输出设备4440上显示用于用户接口的图形信息。

[0240] 存储器4420在系统4400内存储信息。在一个实施方式中,存储器4420是计算机可读介质。在一个实施方式中,存储器4420是易失性存储器单元。在另一个实施方式中,存储器4420是非易失性存储器单元。

[0241] 存储设备4430能够为系统4400提供海量存储。在一个实施方式中,存储设备4430是计算机可读介质。在各种不同的实施方式中,存储设备4430可以是软盘设备、硬盘设备、光盘设备,或者磁带设备。

[0242] 输入/输出设备4440为系统4400提供输入/输出操作。在一个实施方式中,输入/输出设备4440包括键盘和/或指点设备。在另一个实施方式中,输入/输出设备4440包括用于显示图形用户接口的显示单元。

[0243] 所描述的特征可以在数字电子线路中或者在计算机硬件、固件、软件或者其组合中实施。所述装置可以被在计算机程序产品中实施,所述计算机程序产品有形地体现在信息载体中,例如在机器可读存储设备中或者传播的信号中,用于由可编程处理器执行,并且方法步骤可以由执行指令程序的可编程处理器执行,以便通过对输入数据操作并产生输出来执行所描述的实施方式的功能。所描述的特征可以有利地在一个或多个计算机程序中实施,所述计算机程序可在包括至少一个可编程处理器、至少一个输入设备和至少一个输出设备的可编程系统上执行,所述可编程处理器被耦合到从数据存储系统接收数据和指令,并把数据和指令传送到所述数据存储系统。计算机程序是可在计算机中直接或者间接地使用以执行某个活动或者带来某个结果的指令集合。计算机程序可以用任何形式的编程语言

书写,包括汇编或者解释语言,并且其可以被以任何形式部署,包括作为独立程序或者作为模块、部件、子例程或者其他适于在计算环境中使用的单元。

[0244] 举例来说,用于执行指令程序的合适的处理器包括通用和专用微处理器,并且任何种类的计算机的唯一的处理器或者多个处理器中的一个。一般地,处理器将从只读存储器或者随机访问存储器或者这两者接收指令和数据。计算机的必要元件是用于执行指令的处理器和用于存储指令和数据的一个或多个存储器。一般地,计算机也将包括或者可操作地耦合与用于存储数据文件的一个或多个海量存储设备通信,这些设备包括磁盘、例如内部硬盘和可移动盘、磁光盘和光盘。适于具体实施计算机程序指令和数据的存储设备包括所有形式的非易失存储器,例如包括半导体存储器设备、例如EPROM、EEPROM和快闪存储器设备;磁盘,例如内部硬盘和可移动盘、磁光盘和CD-ROM和DVD-ROM盘。处理器和存储器可以被ASIC(专用集成电路)补充或者被包含在ASIC中。

[0245] 为了提供与用户的交互作用,所述特征可以被实施在具有例如CRT(阴极射线管)或者LCD(液晶显示器)监视器的用于向用户显示信息的显示设备和键盘以及例如鼠标或者轨迹球的指点设备的计算机上,通过键盘和指点设备用户可以向计算机提供输入。

[0246] 所述特征可以被实施在包括例如数据服务器的后端部件或者包括例如应用程序服务器或者因特网服务器的中间件部件或者包括例如具有图形用户接口或者因特网浏览器的客户端计算机的前端部件,或者其任意组合的计算机系统中。系统的部件可以通过任何形式或者介质的数字数据通信连接,例如通信网络。通信网络的例子包括例如LAN、WAN,以及形成因特网的计算机和网络。

[0247] 计算机系统可以包括客户端和服务器。客户端和服务器一般彼此远离,并且通常通过网络交互作用,例如所描述的那个。客户端和服务器的关系因运行于相应计算机上并且彼此具有客户端-服务器关系的计算机程序而出现。

[0248] 已经描述了许多实施方式。尽管如此,将会理解,不偏离权利要求的精神和范围可以做出各种修改。因此,其他的实施例在下列权利要求的范围内。

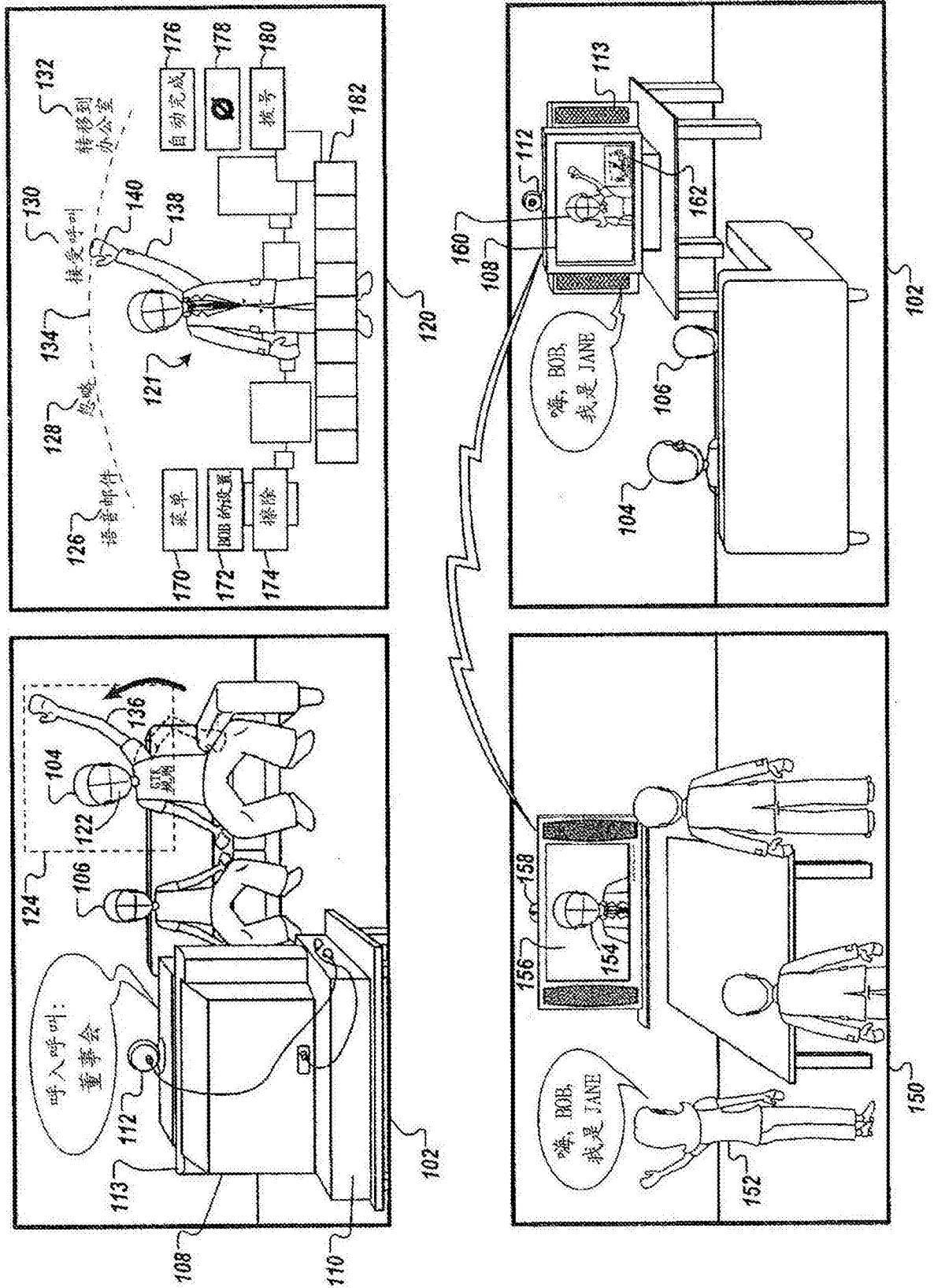


图1

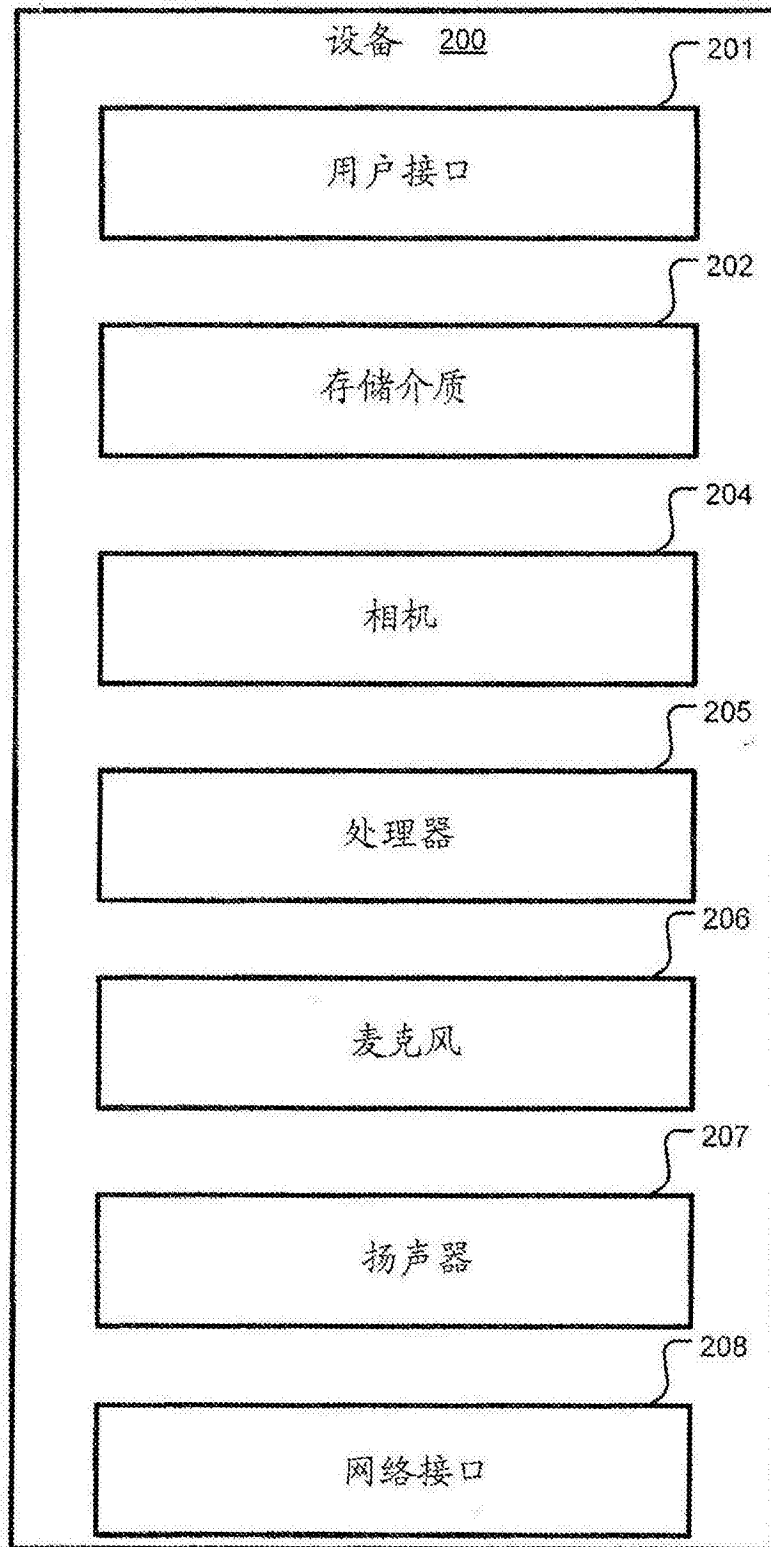


图2

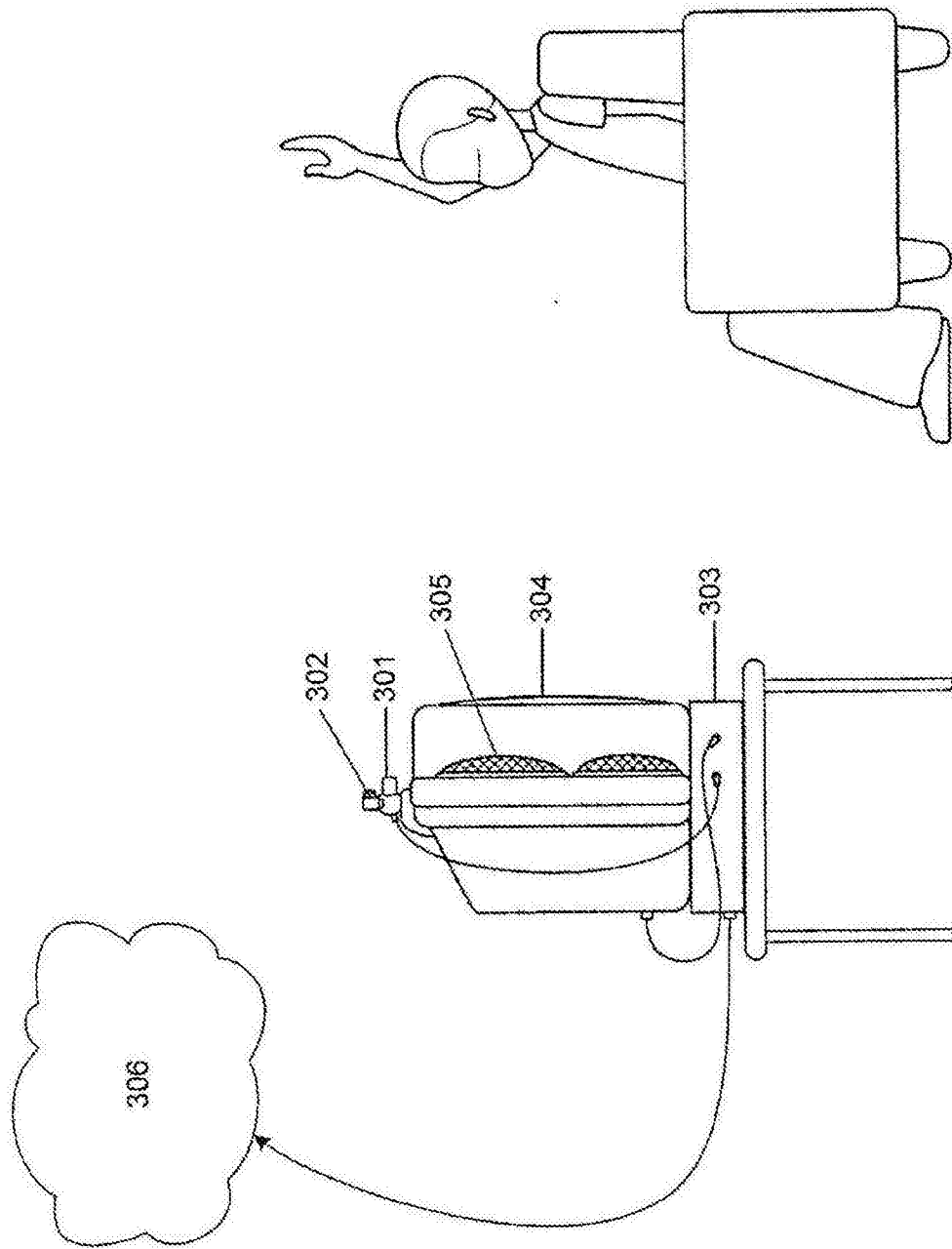


图3

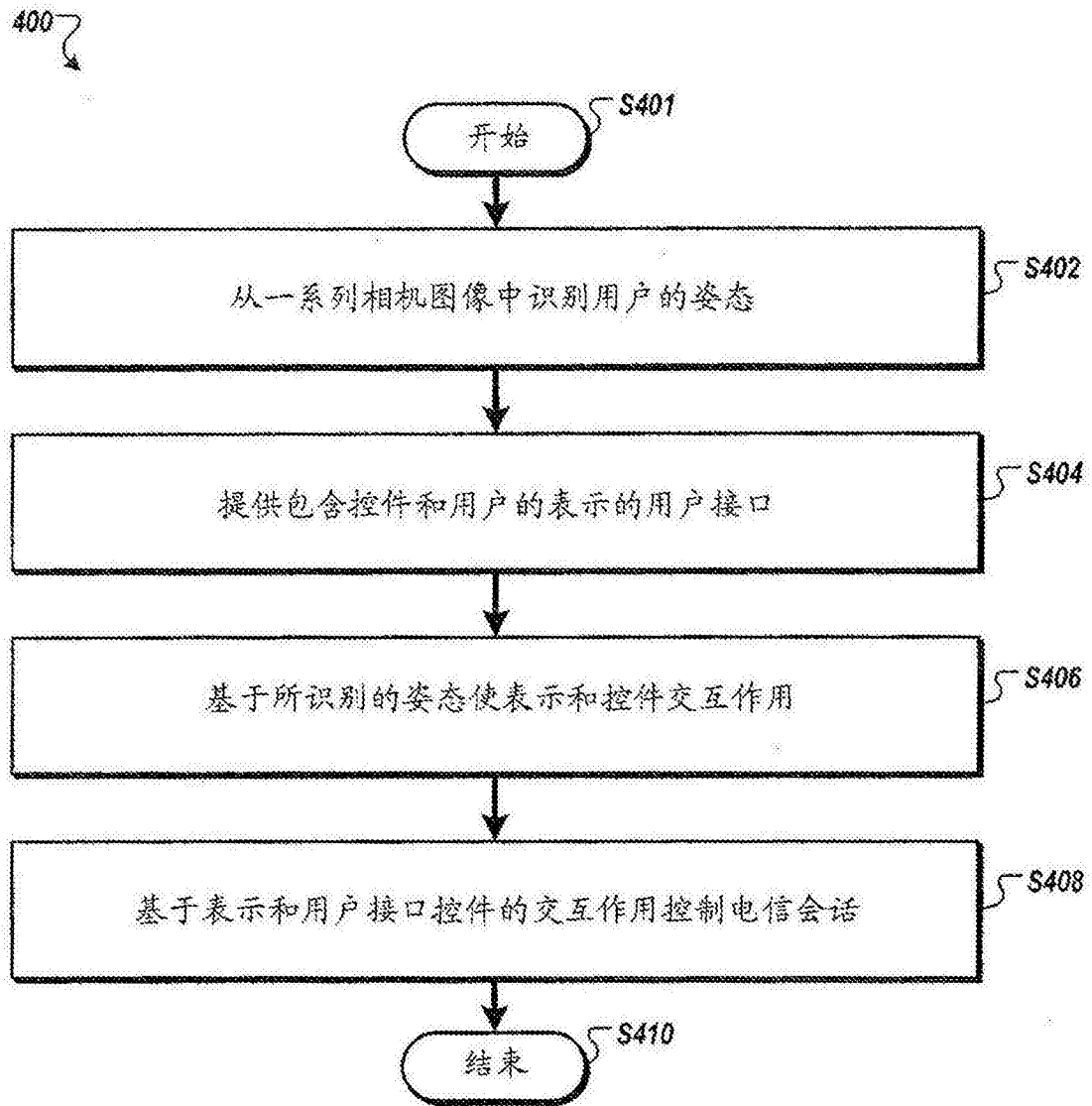


图4

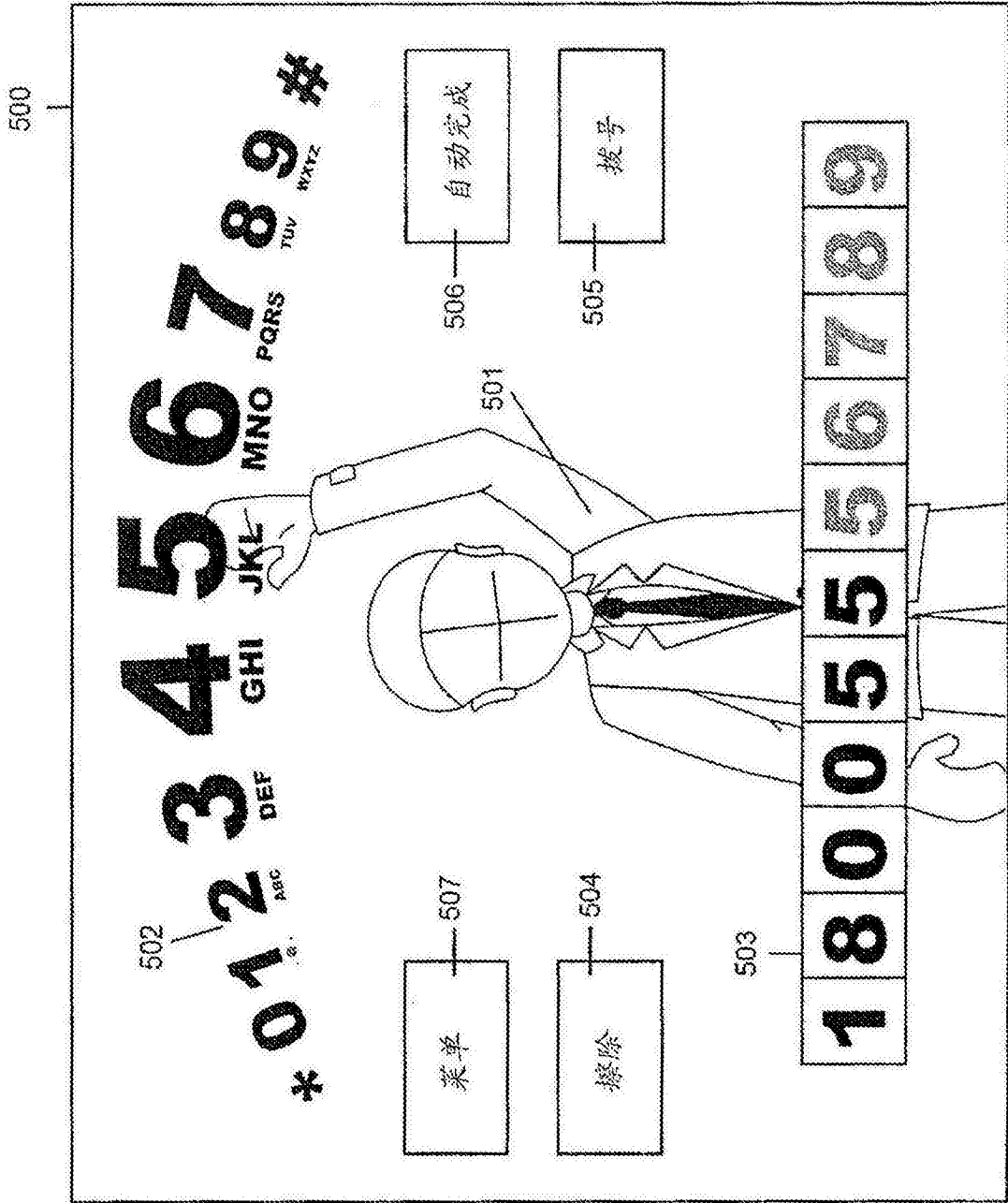


图5

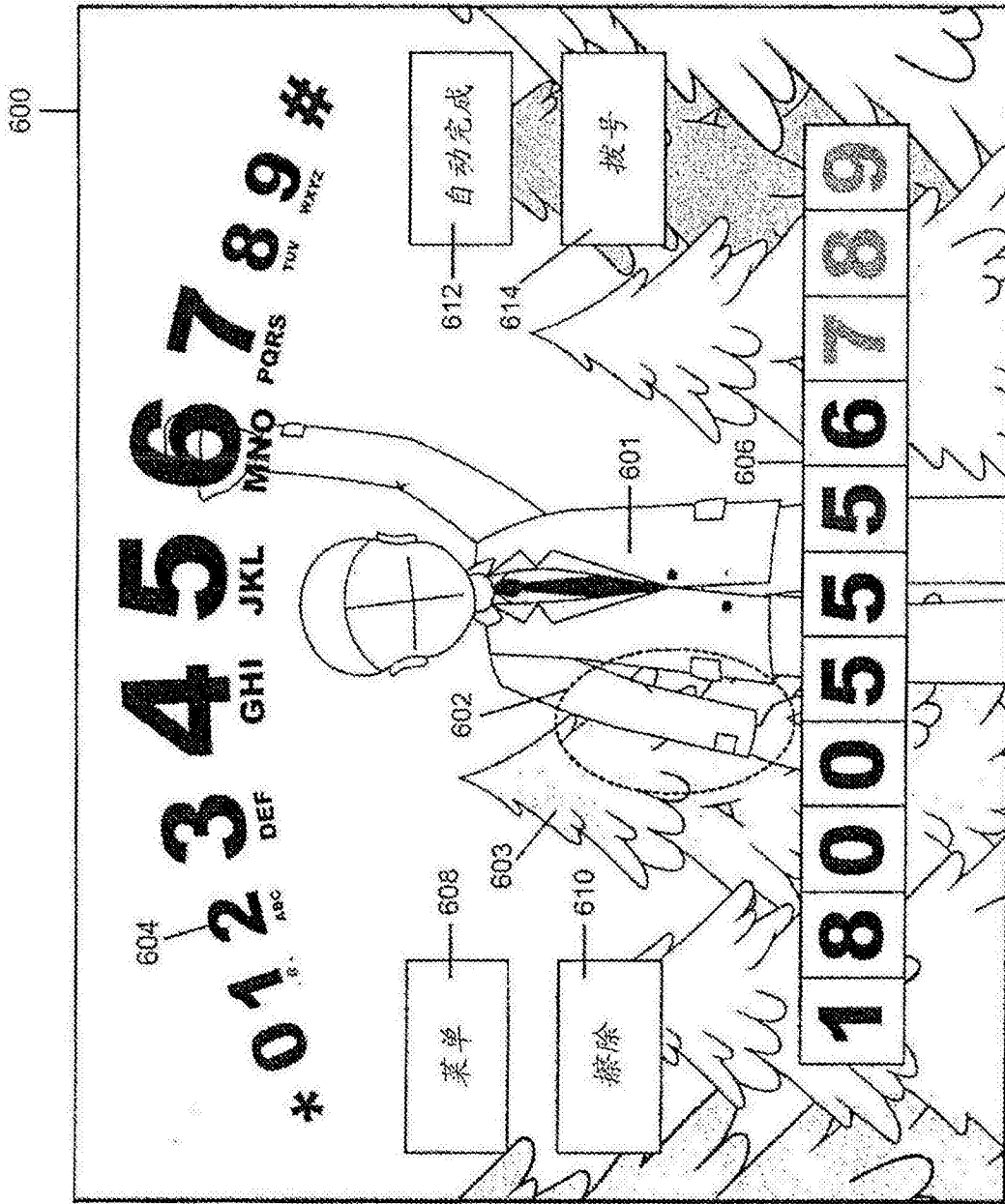


图6

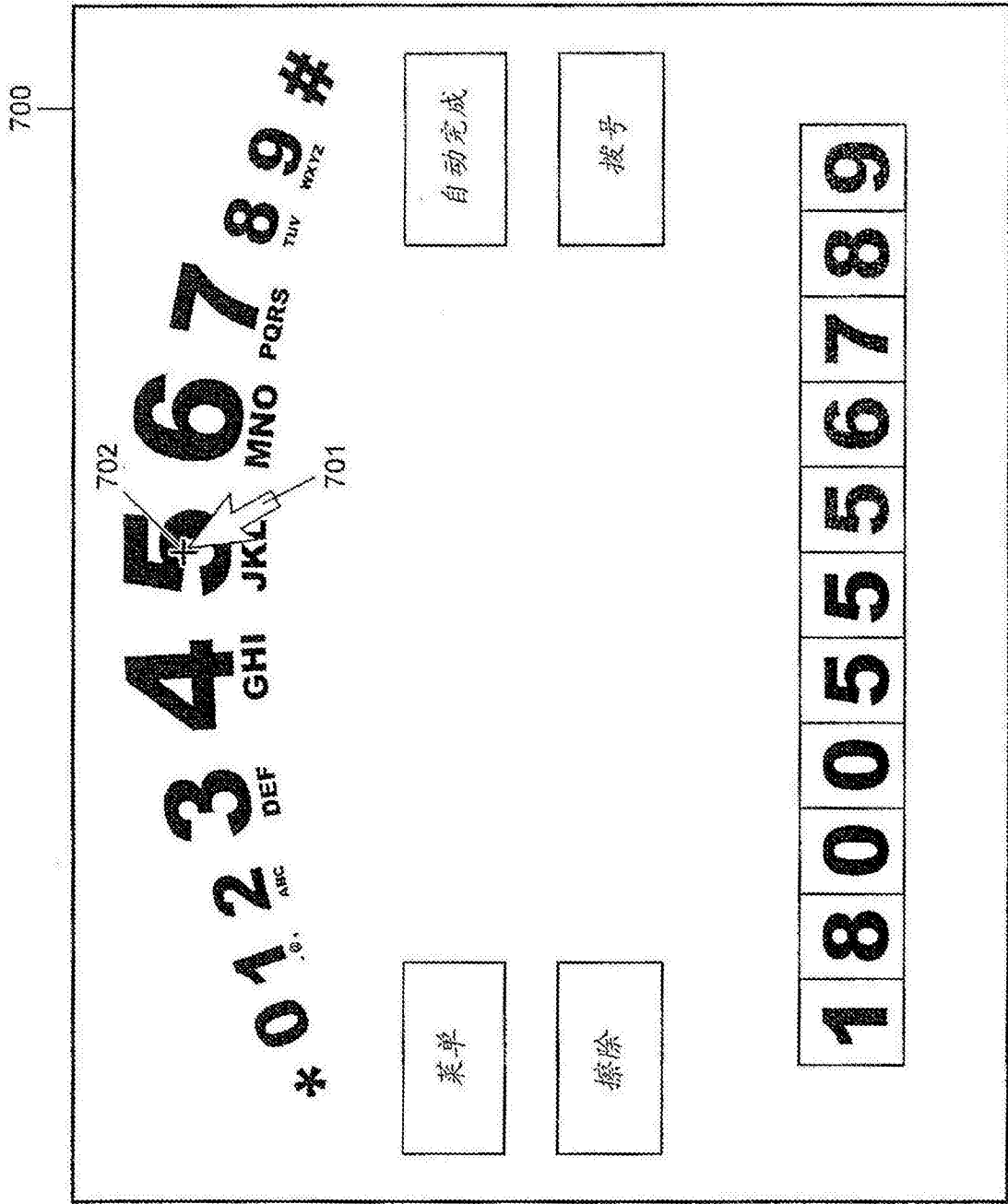


图7

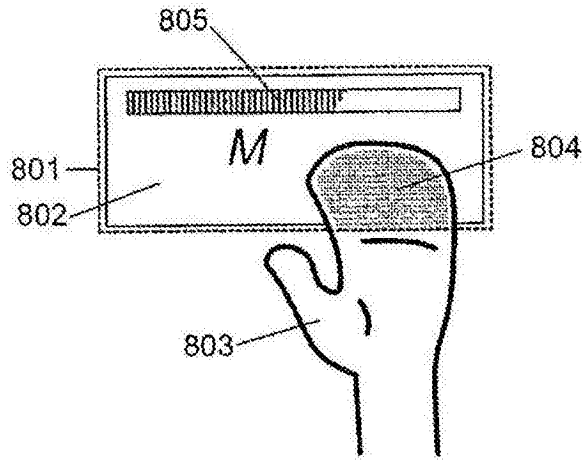


图8

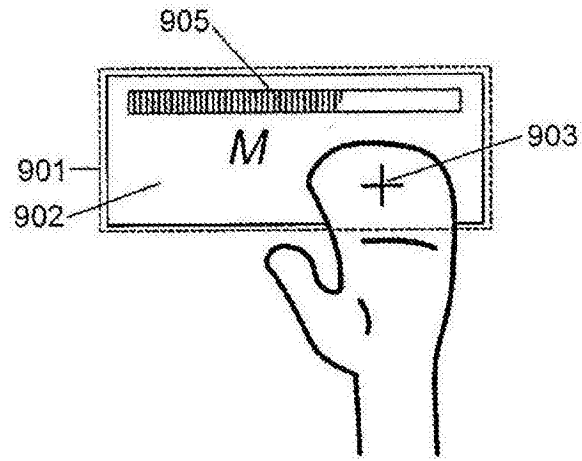


图9

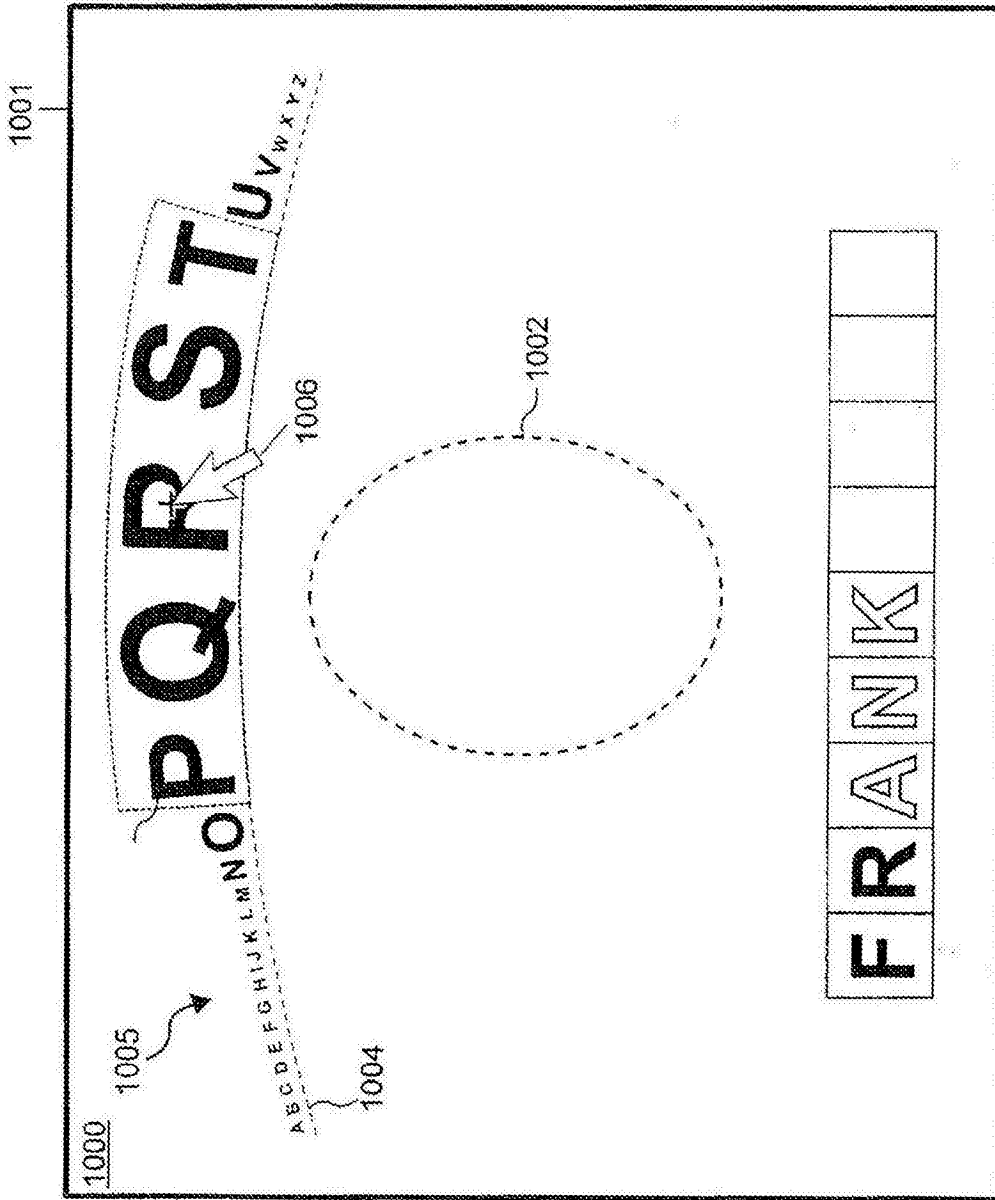


图10

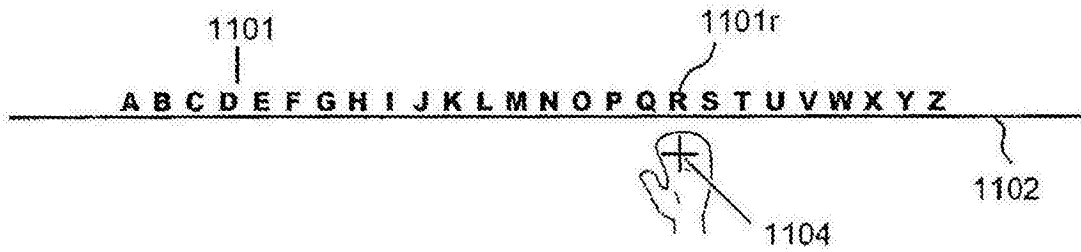


图11

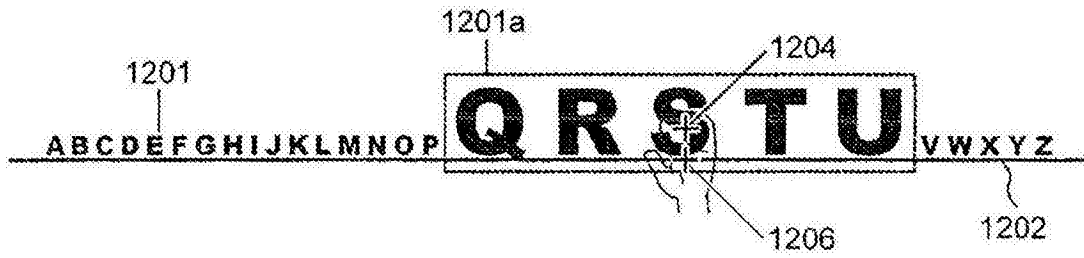


图12

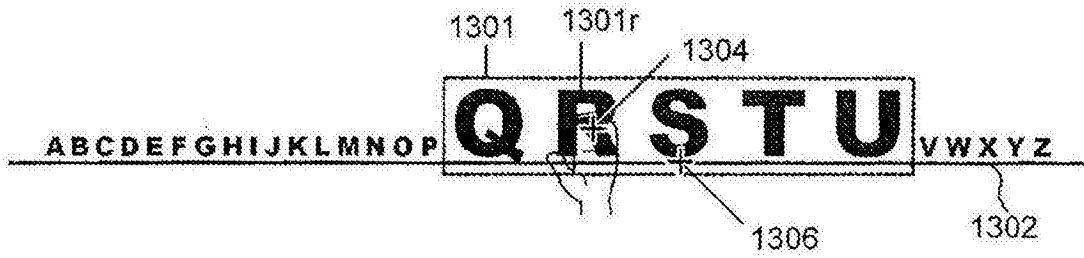


图13

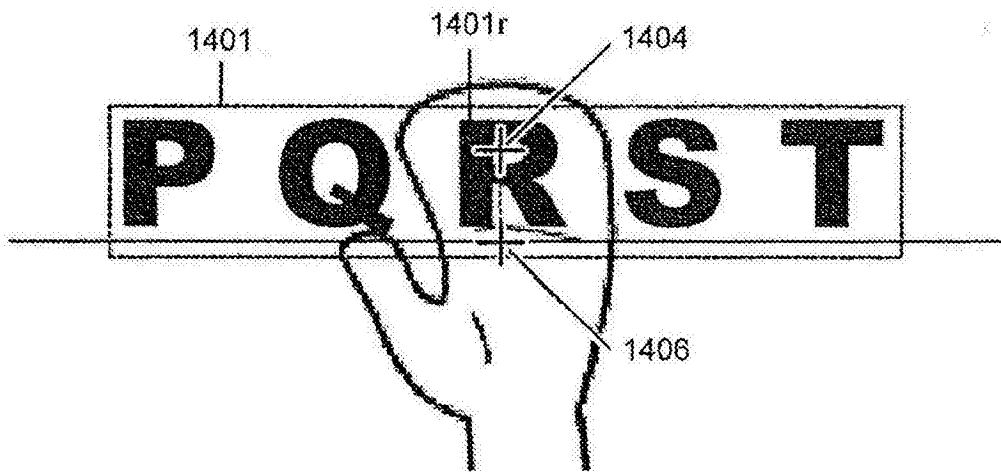


图14

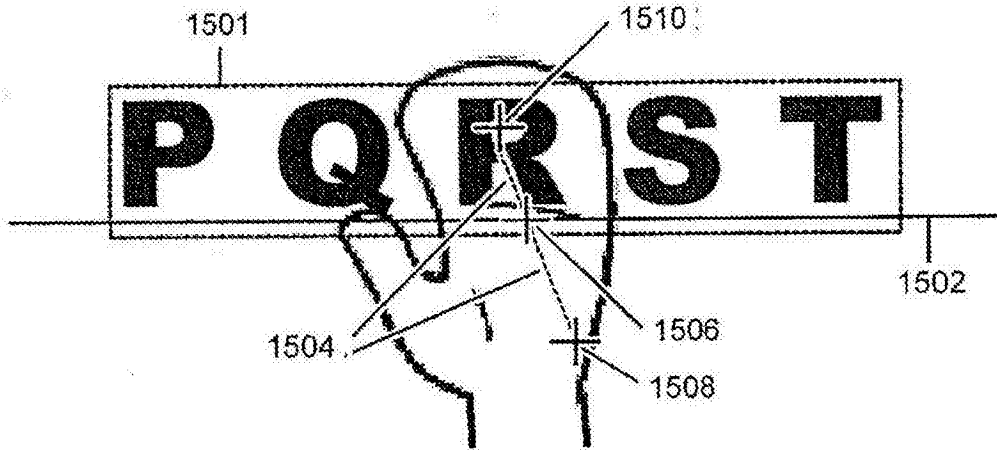


图15

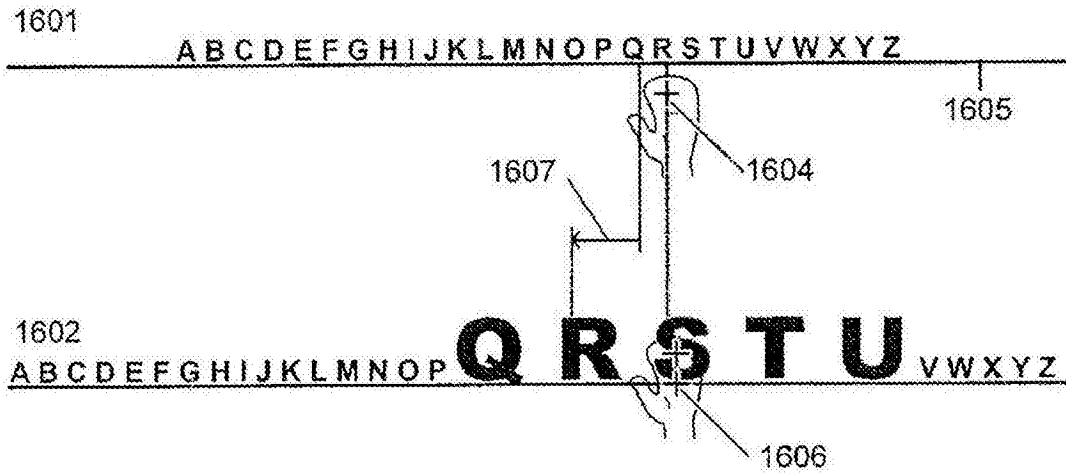


图16

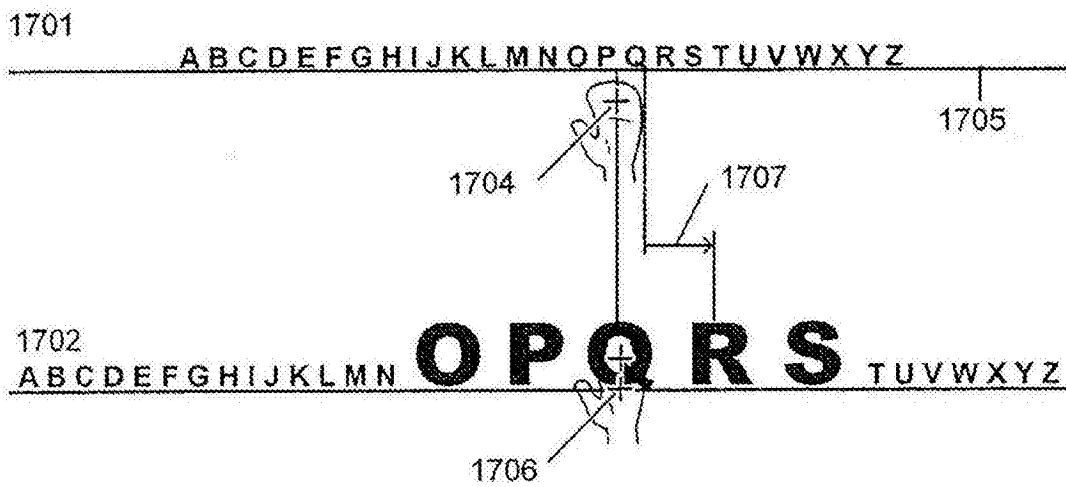


图17

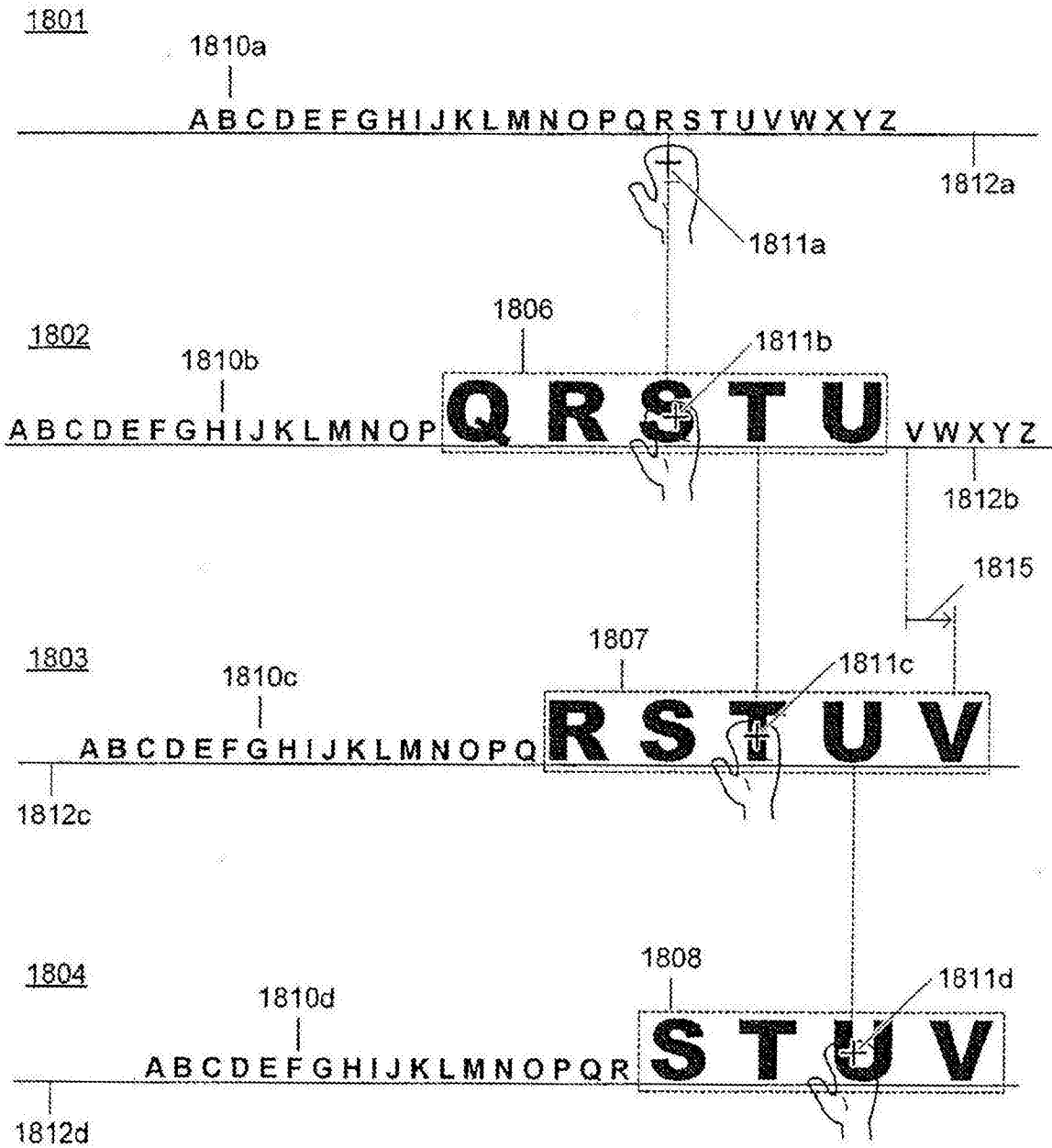


图18

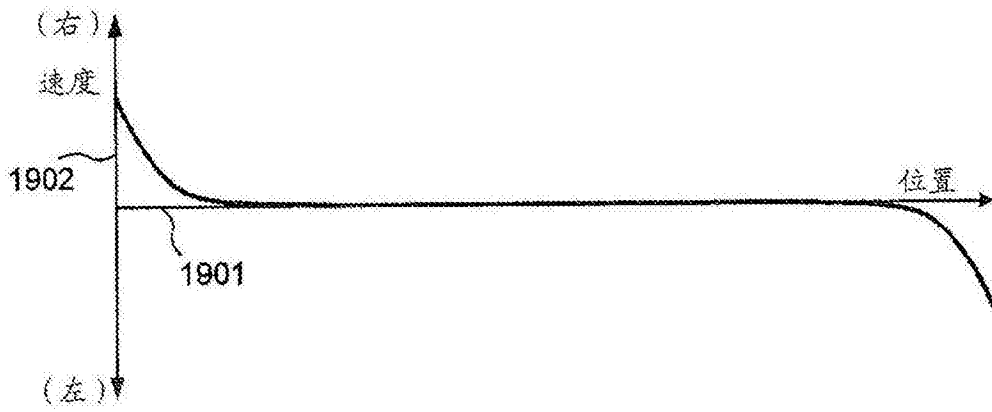


图19

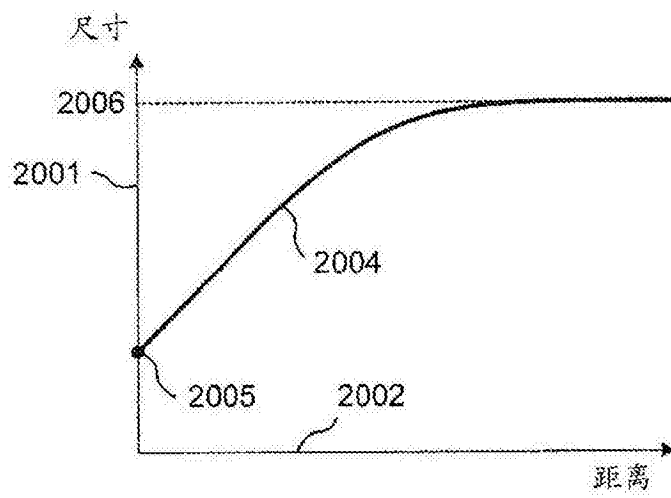


图20

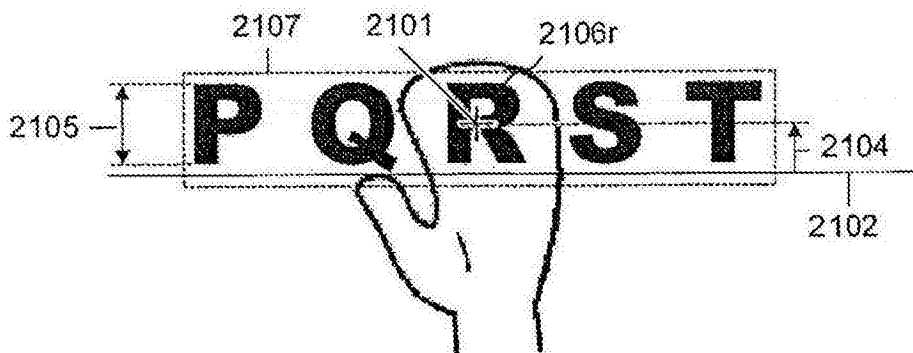


图21

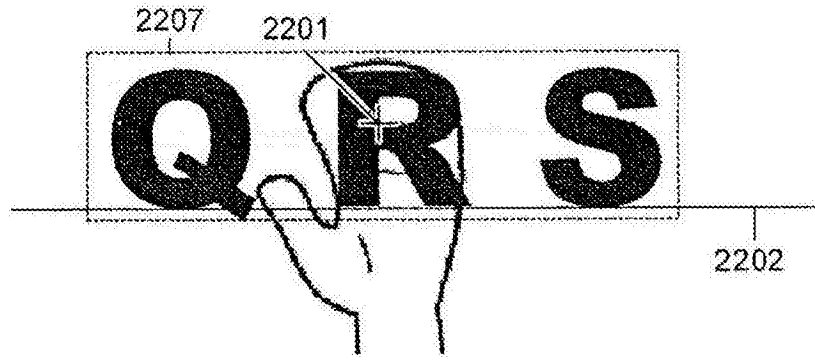


图22

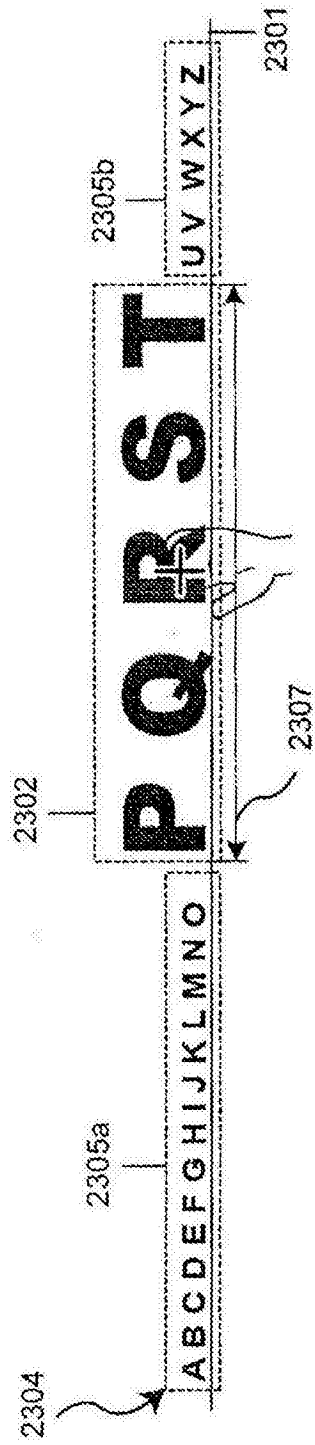


图23

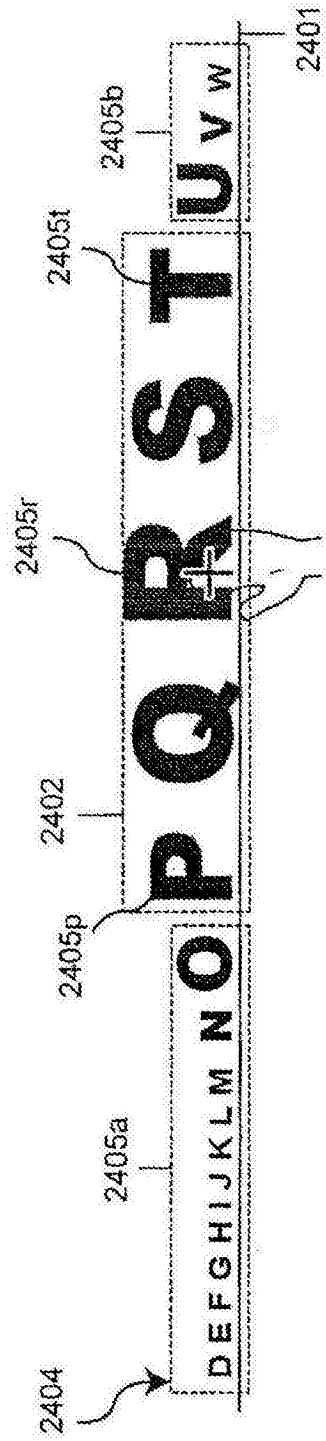


图24

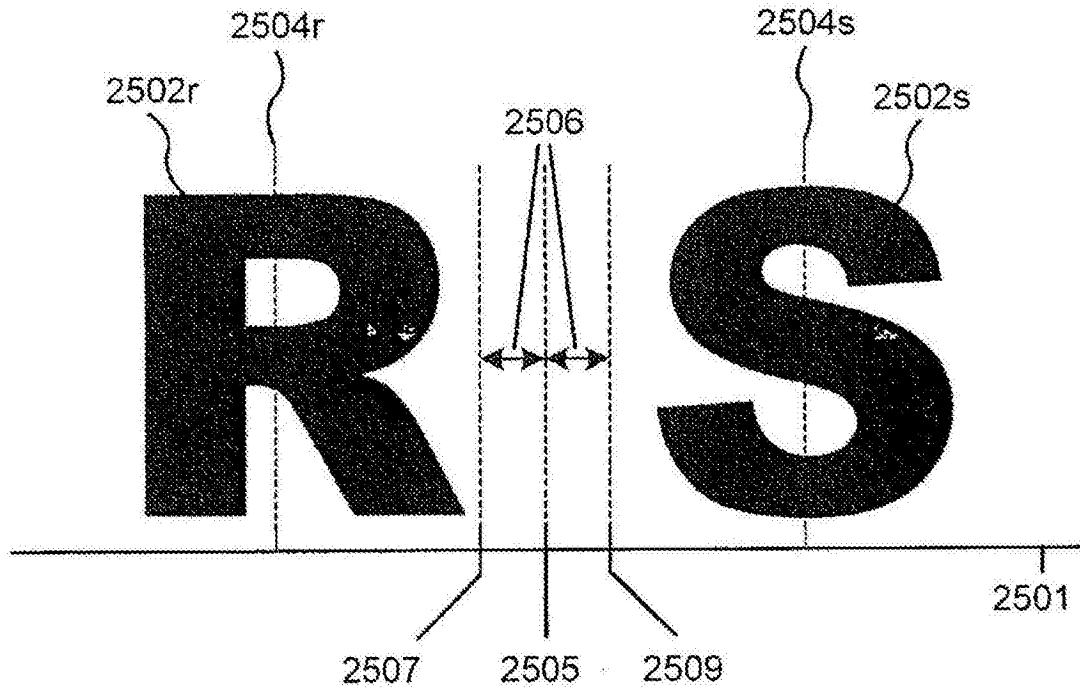


图25

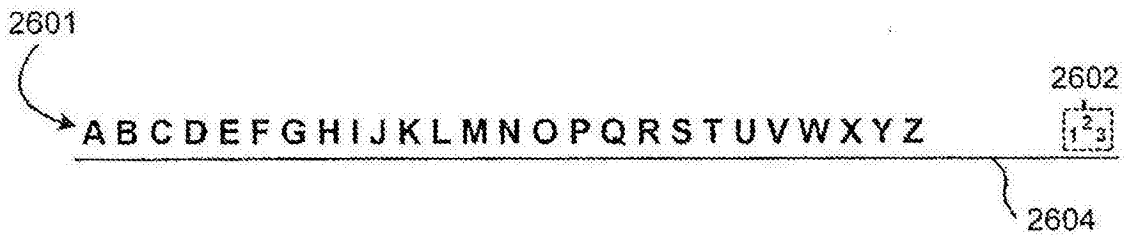


图26

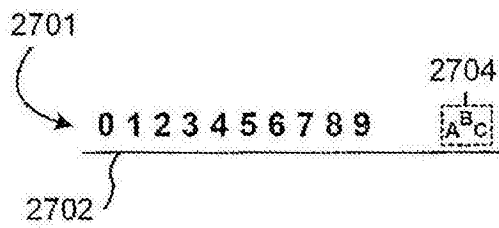


图27

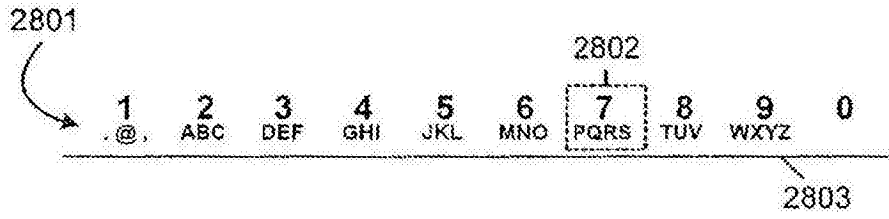


图28

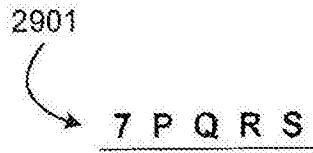


图29

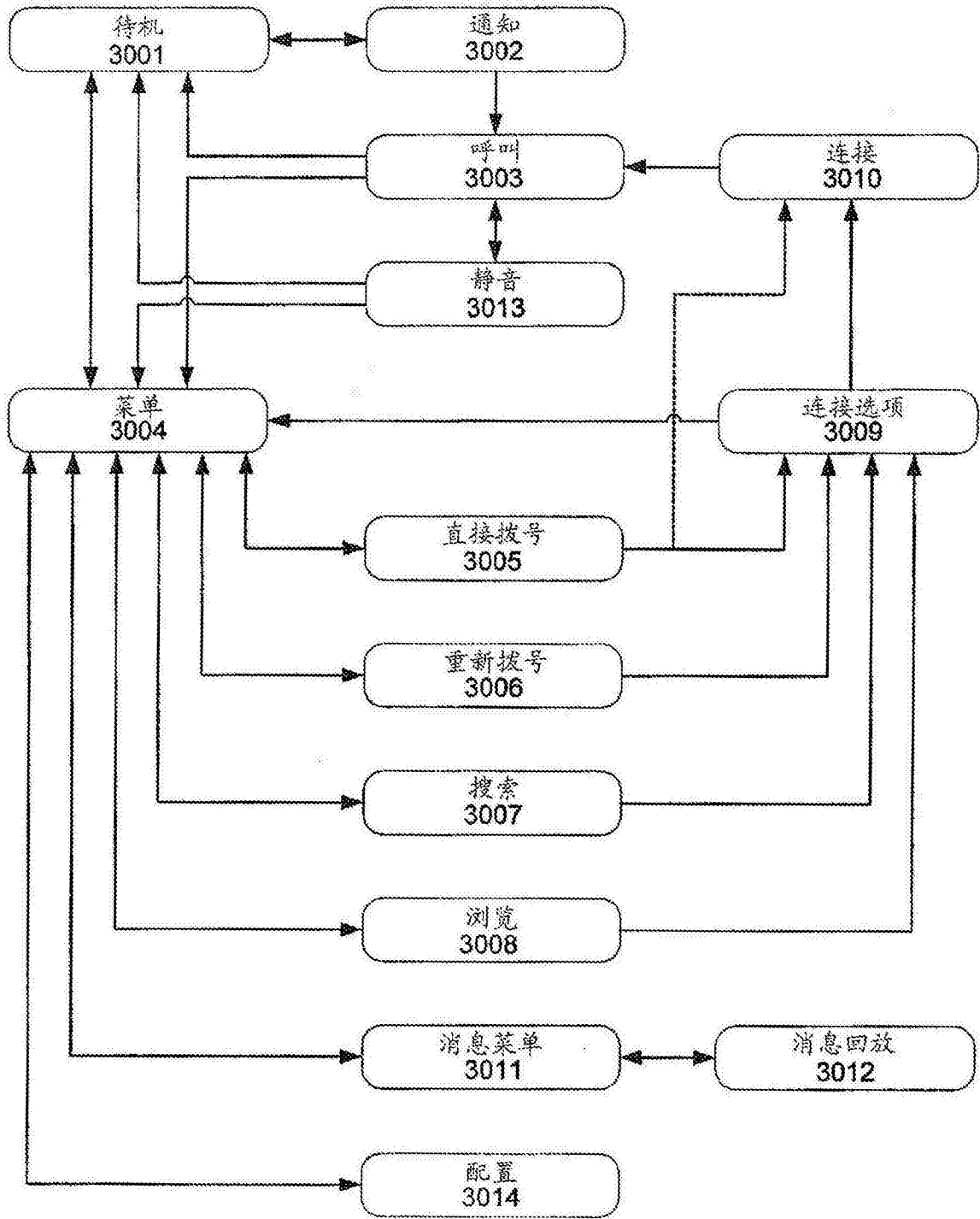


图30

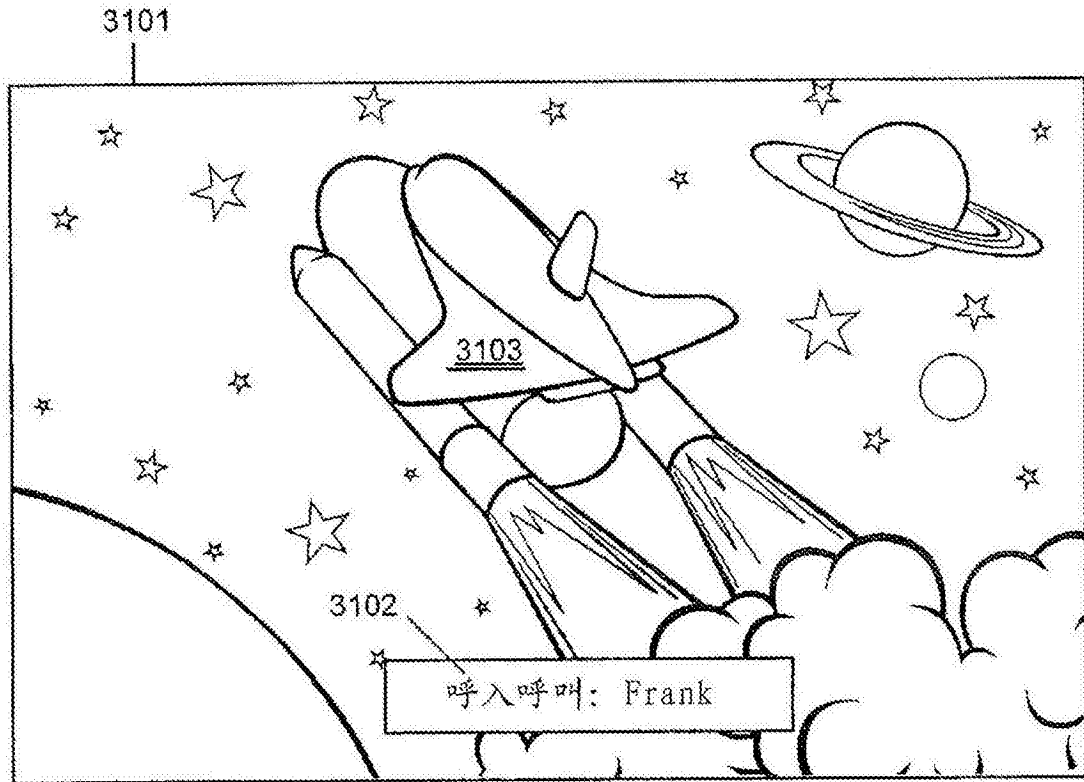


图31

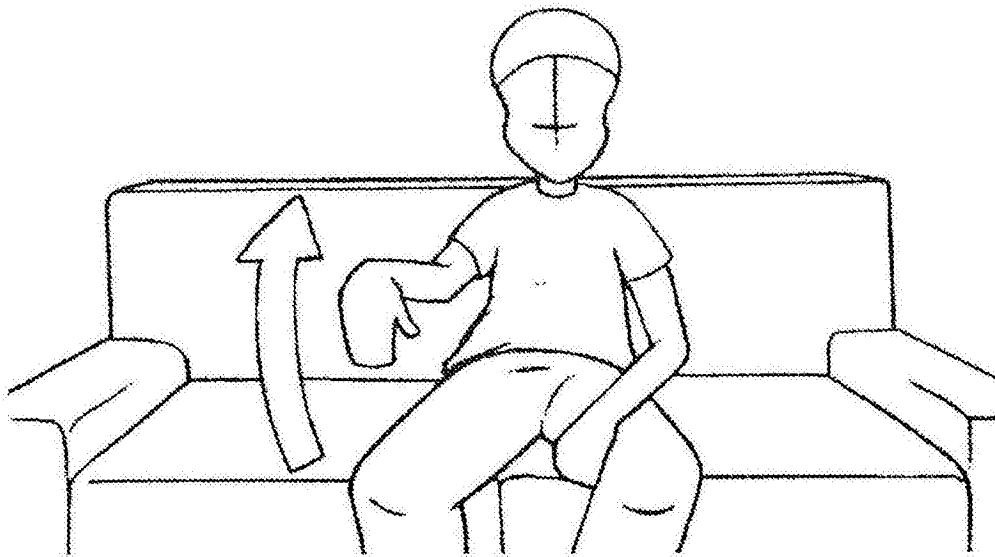


图32

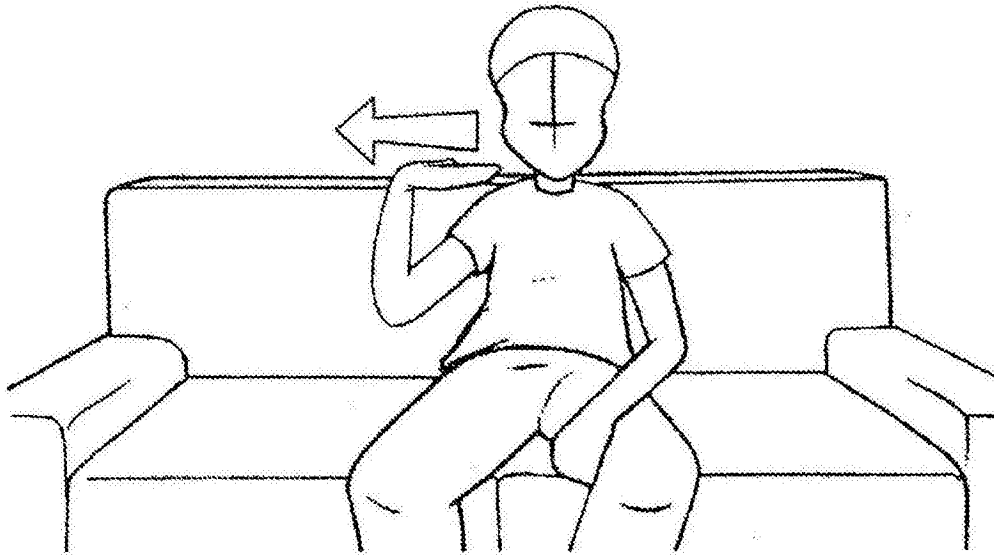


图33

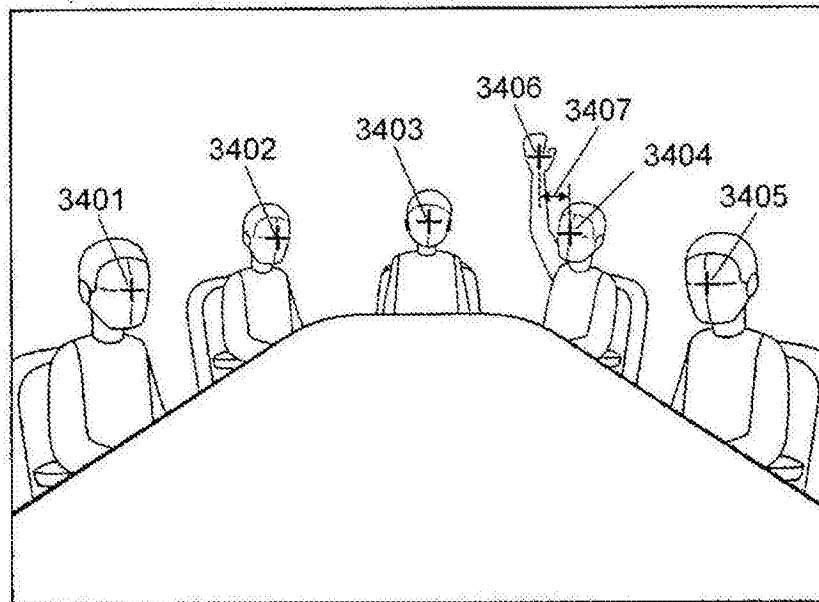


图34

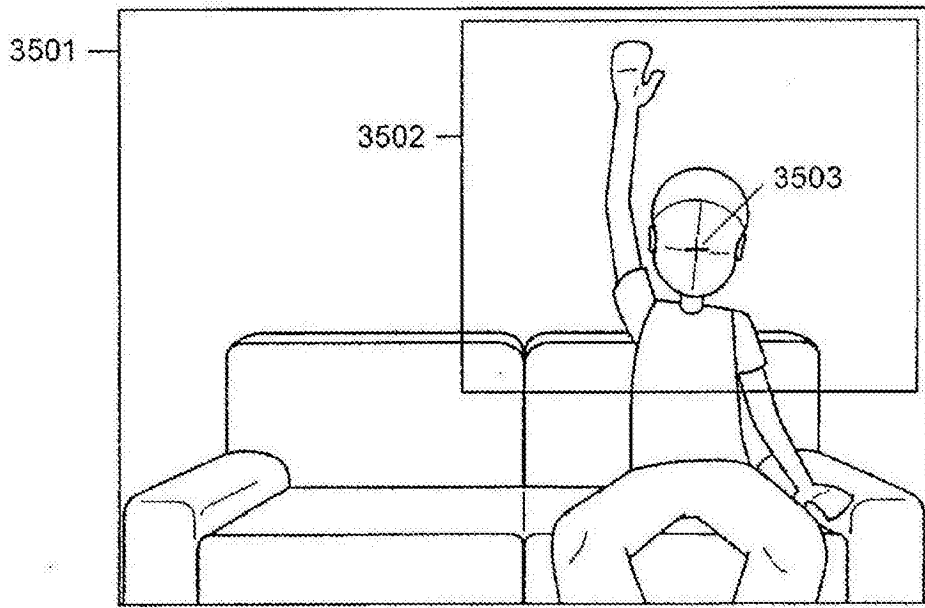


图35

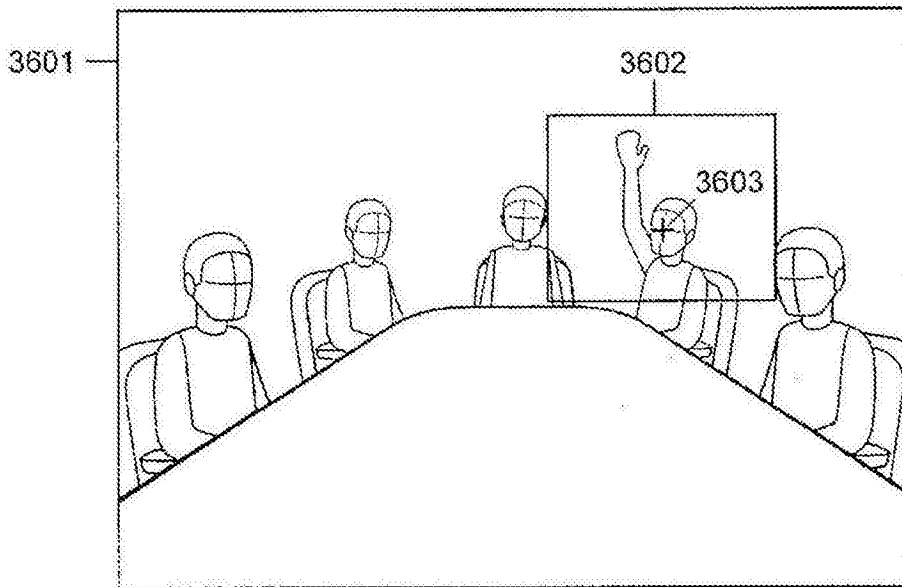


图36

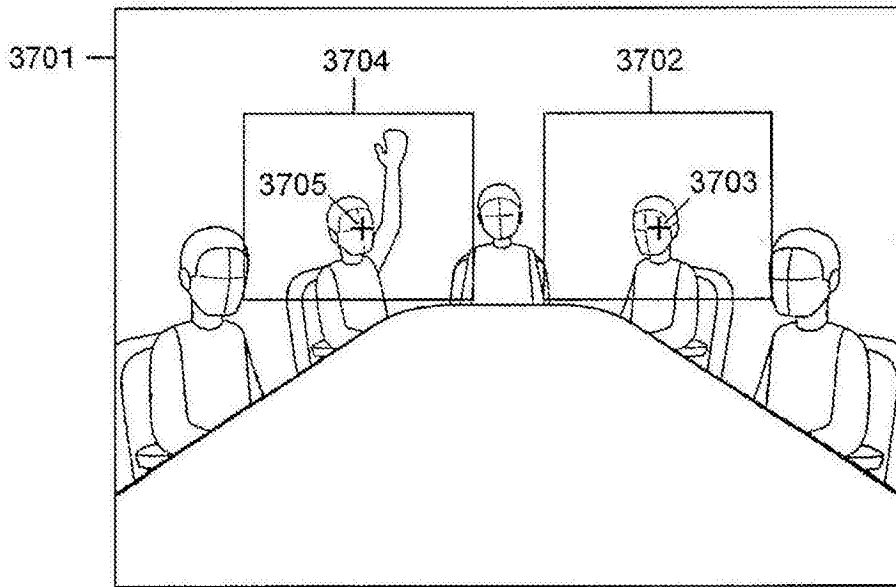


图37

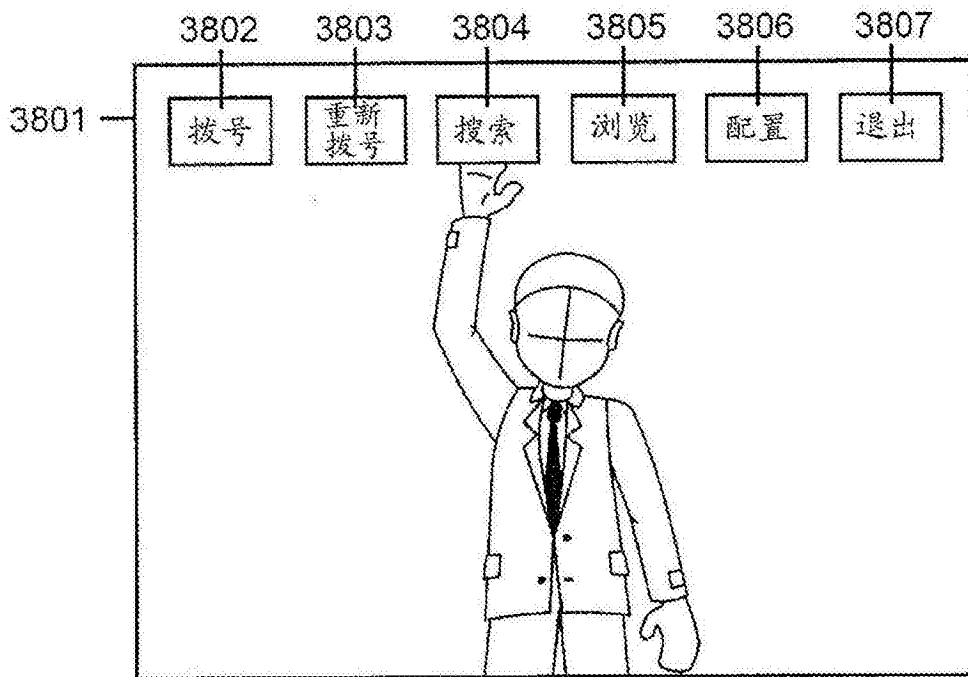


图38

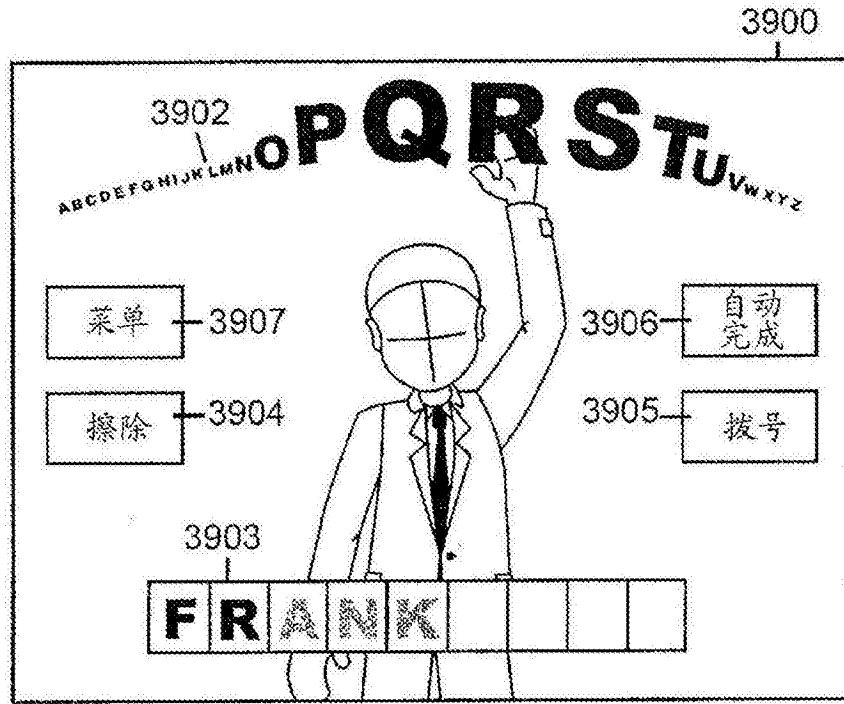


图39

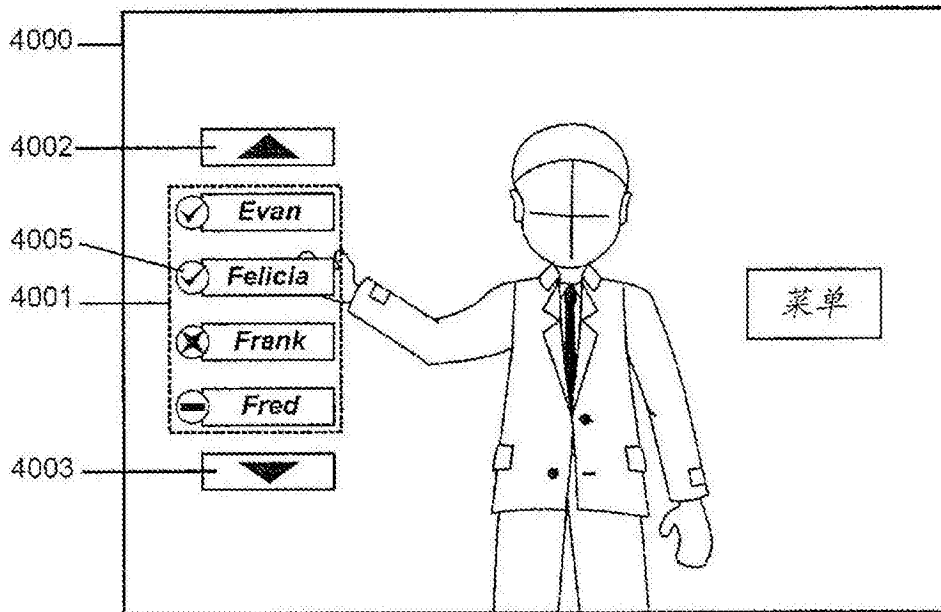


图40

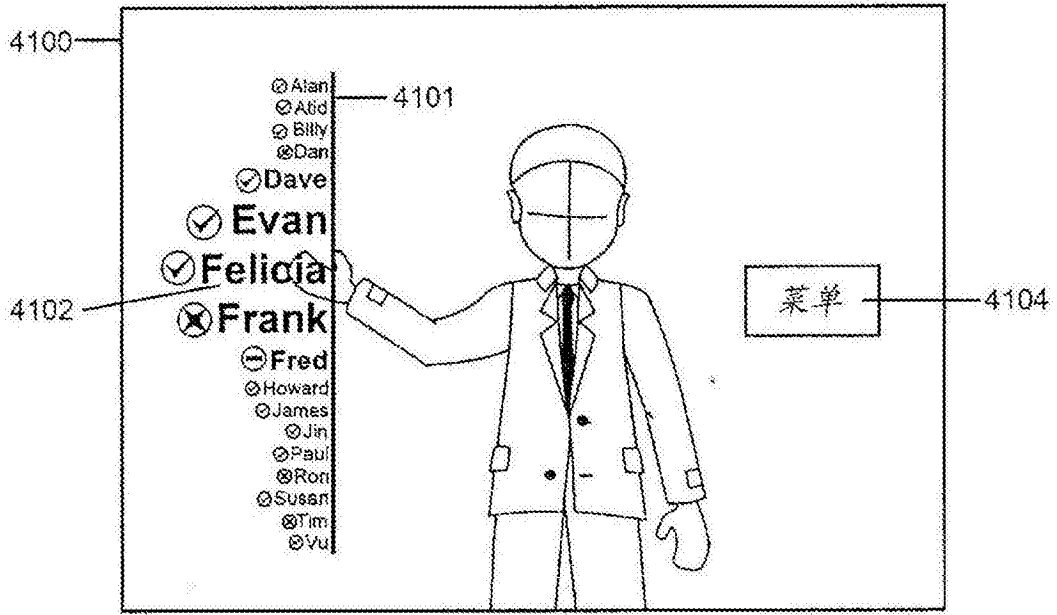


图41

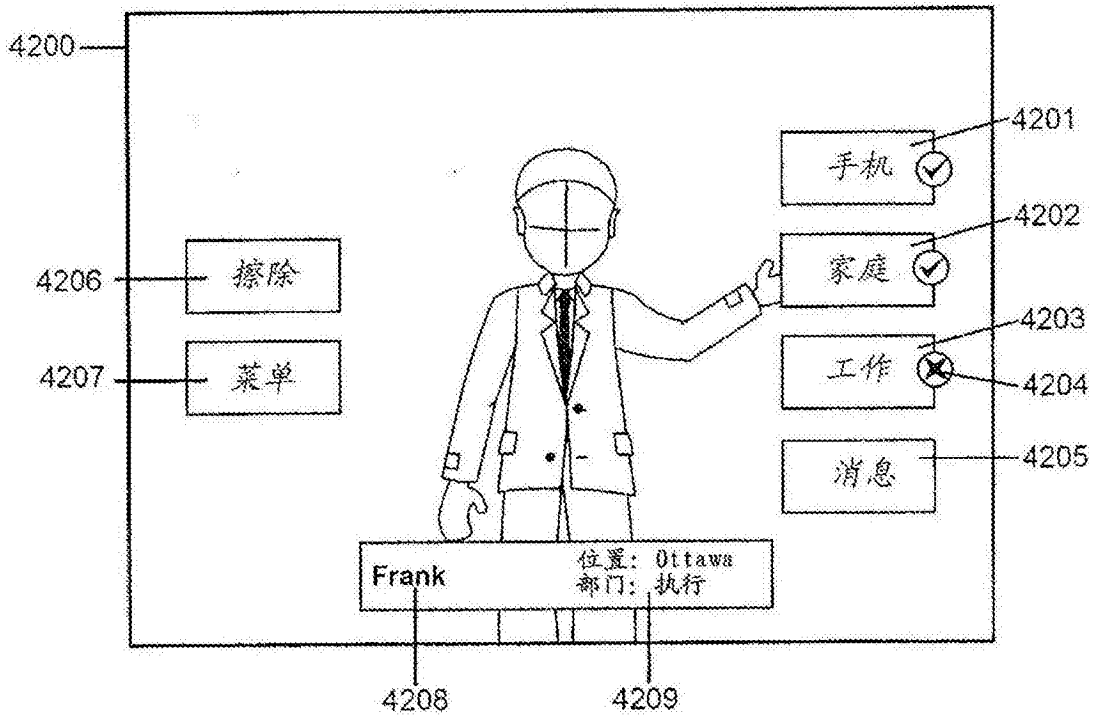


图42

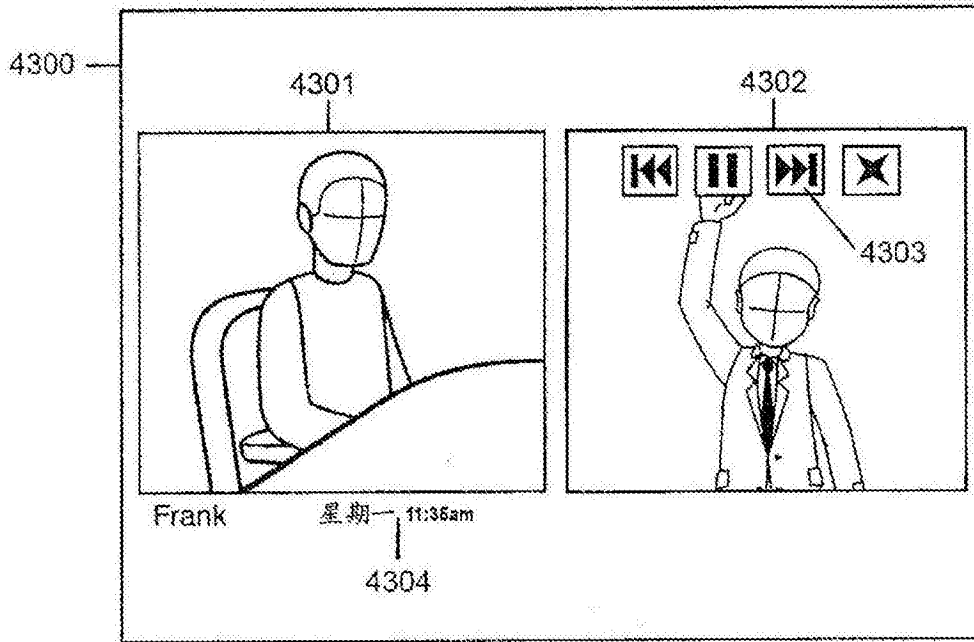


图43

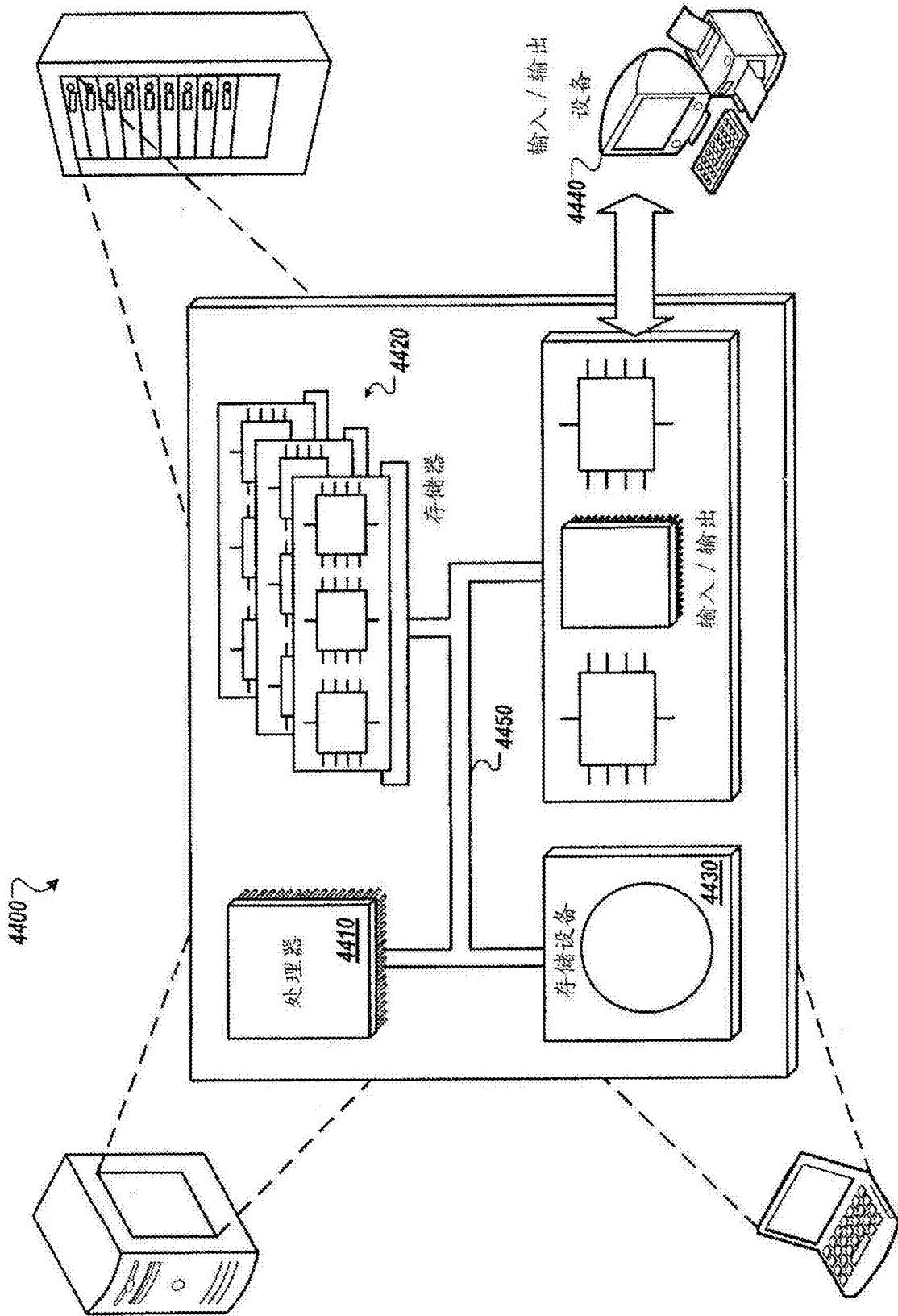


图44