



(12) 发明专利申请

(10) 申请公布号 CN 113593567 A

(43) 申请公布日 2021. 11. 02

(21) 申请号 202110700318.0

(22) 申请日 2021.06.23

(71) 申请人 荣耀终端有限公司

地址 518040 广东省深圳市福田区香蜜湖
街道东海社区红荔西路8089号深业中
城6号楼A单元3401

(72) 发明人 谭泳发

(74) 专利代理机构 深圳市赛恩倍吉知识产权代
理有限公司 44334

代理人 刘丽华 习冬梅

(51) Int. Cl.

G10L 15/26 (2006.01)

H04N 21/43 (2011.01)

H04N 21/488 (2011.01)

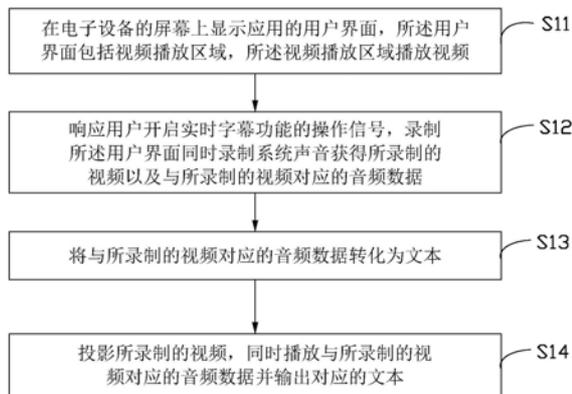
权利要求书2页 说明书19页 附图32页

(54) 发明名称

视频声音转文本的方法及相关设备

(57) 摘要

本申请的实施例提供了一种视频声音转文本的方法及相关设备。该方法在屏幕上显示用于播放视频的应用的用户界面。该方法响应用户的操作录制所述用户界面同时录制系统声音获得所录制的视频及与所录制的视频对应的音频数据。通过将所录制的视频对应的音频数据转化为文本,并投影所录制的视频并输出对应的文本,同时播放与所录制的视频对应的音频数据,实现视频声音转文本的功能。利用本实施例可使得声音和视频画面以及字幕文本能够同步,提升用户视频观看体验。



1. 一种视频声音转文本的方法,应用于电子设备,其特征在于,所述方法包括:
在电子设备的第一屏幕上显示应用的用户界面,所述用户界面包括视频播放区域,所述视频播放区域播放视频;
响应用户开启实时字幕功能的操作信号,录制所述用户界面同时录制系统声音获得所录制的视频以及与所录制的视频对应的音频数据;
将与所录制的视频对应的音频数据转化为文本;及
投影所录制的视频并输出对应的文本,同时播放与所录制的视频对应的音频数据。
2. 如权利要求1所述的视频声音转文本的方法,其特征在于,所述响应用户开启实时字幕功能的操作信号,录制所述用户界面包括:
响应所述操作信号,将所述用户界面缩小为一个小窗,并将所述小窗移动到所述第一屏幕的第一指定位置;
对所述第一屏幕的全屏画面进行录制获得所录制的全屏画面;及
根据所述小窗的大小和位置对所录制的全屏画面进行裁剪,从而获得所述小窗的画面内容,将所获得的所述小窗的画面内容作为所录制的视频。
3. 如权利要求1所述的视频声音转文本的方法,其特征在于,所述响应用户开启实时字幕功能的操作信号,录制所述用户界面包括:
响应所述操作信号,将所述用户界面缩小为一个小窗,并将所述小窗移动到所述第一屏幕的第一指定位置;及
对所述小窗的画面内容进行录制,获得所录制的视频。
4. 如权利要求2或3所述的视频声音转文本的方法,其特征在于,该方法于第一时间点 t_1 开始录制所述用户界面同时录制系统声音。
5. 如权利要求4所述的视频声音转文本的方法,其特征在于,所述第一时间点 t_1 为用户开启所述实时字幕功能的时间。
6. 如权利要求4所述的视频声音转文本的方法,其特征在于,该方法还包括:
于所述第一时间点 t_1 关闭电子设备的扬声器。
7. 如权利要求1至6中任意一项所述的视频声音转文本的方法,其特征在于,当所述小窗移动所述第一指定位置时,所述小窗展示所述用户界面的完整内容;或者仅展示所述视频播放区域的内容。
8. 如权利要求1至7中任意一项所述的视频声音转文本的方法,其特征在于,所述将与所录制的视频对应的音频数据转化为文本包括:
将与所录制的视频对应的音频数据上传到服务器,其中,所述服务器将与所录制的视频对应的音频数据转化为文本,并将转化获得的文本发送给所述电子设备;及
接收所述服务器发送过来的文本。
9. 如权利要求1至8中任意一项所述的视频声音转文本的方法,其特征在于,该方法还包括:
于将与所录制的视频对应的音频数据进行文本转化时,在所述第一屏幕上显示相应的文字提示。
10. 如权利要求4所述的视频声音转文本的方法,其特征在于,该方法于第二时间点 t_2 投影所录制的视频,其中, $t_2 = t_1 + \delta$, δ 的值的大小与将所录制的语音数据转化为文本所需

的时间相关。

11. 如权利要求10所述的视频声音转文本的方法,其特征在于,该方法将所录制的视频投影到所述第一屏幕的第二指定位置,所述第二指定位置与所述第一指定位置重叠或者不重叠。

12. 如权利要求10所述的视频声音转文本的方法,其特征在于,该方法将所录制的视频投影到所述电子设备的第二屏幕或者将所录制的视频投影到与所述电子设备通讯连接其他设备的屏幕上。

13. 如权利要求10至12中任意一项所述的视频声音转文本的方法,其特征在于,该方法将所录制的视频放大后再投影。

14. 如权利要求1至13中任意一项所述的视频声音转文本的方法,其特征在于,该方法将所转化得到的文本在系统最上层以浮窗形式显示。

15. 如权利要求1至13中任意一项所述的视频声音转文本的方法,其特征在于,该方法将所转化得到的文本内嵌到所录制的视频。

16. 一种视频声音转文本的方法,应用于电子设备,其特征在于,所述方法包括:

在电子设备的第一屏幕上显示应用的用户界面,所述用户界面包括视频播放区域,所述视频播放区域播放视频;

响应用户开启实时字幕功能的操作信号,录制所述用户界面同时录制系统声音,获得所录制的多个视频,以及获得与每个视频对应的音频数据;

按照所述多个视频的录制顺序将与每个视频对应的音频数据转化为文本;及

按照所述录制顺序依次投影所述多个视频并输出对应的文本,同时播放对应的音频数据。

17. 如权利要求16所述的视频声音转文本的方法,其特征在于,所述多个视频中的每个视频的时长相同。

18. 一种电子设备,其特征在于,包括处理器和存储器;所述存储器,用于存储指令;所述处理器,用于调用所述存储器中的指令,使得所述电子设备实现如权利要求1至15中任一项所述的视频声音转文本的方法或者实现如权利要求16至17中任一项所述的视频声音转文本的方法。

19. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有至少一个指令,所述至少一个指令被处理器执行时实现如权利要求1至15中任一项所述的视频声音转文本的方法或者实现如权利要求16至17中任一项所述的视频声音转文本的方法。

视频声音转文本的方法及相关设备

技术领域

[0001] 本申请实施例涉及终端技术领域,尤其涉及一种视频声音转文本的方法及相关设备。

背景技术

[0002] 现在很多设备都支持录制设备内声音,通过语音识别技术,把音频数据转写成文本,并立即显示在屏幕上,达到实时生成并显示字幕的效果。然而,由于模型推理有不可忽略的时延,通常显示的字幕和原音轨会有比较明显的时差,也即字幕领域常说的时间轴没有对齐。因此,用户实际观看的实时字幕会有明显滞后,导致用户体验较差。

发明内容

[0003] 本申请实施例公开了一种视频声音转文本的方法及相关设备,可使得声音和视频画面以及字幕文本能够同步,提升用户视频观看体验。

[0004] 本申请第一方面公开了一种视频声音转文本的方法,包括:在电子设备的第一屏幕上显示应用的用户界面,所述用户界面包括视频播放区域,所述视频播放区域播放视频;响应用户开启实时字幕功能的操作信号,录制所述用户界面同时录制系统声音获得所录制的视频以及与所录制的视频对应的音频数据;将与所录制的视频对应的音频数据转化为文本;及投影所录制的视频并输出对应的文本,同时播放与所录制的视频对应的音频数据。

[0005] 采用所述技术方案,通过同时录制用户界面和系统声音,并在投影所录制的画面的同时播放所录制的声音及输出对应的字幕,使得所播放的画面和声音及所输出的字幕同步,提升用户视频观看体验。

[0006] 在一些可选的实施方式中,所述响应用户开启实时字幕功能的操作信号,录制所述用户界面包括:响应所述操作信号,将所述用户界面缩小为一个小窗,并将所述小窗移动到所述第一屏幕的第一指定位置;对所述第一屏幕的全屏画面进行录制获得所录制的全屏画面;及根据所述小窗的大小和位置对所录制的全屏画面进行裁剪,从而获得所述小窗的画面内容,将所获得的所述小窗的画面内容作为所录制的视频。采用所述技术方案,可以实现对视频画面准确的录制。

[0007] 在一些可选的实施方式中,所述响应用户开启实时字幕功能的操作信号,录制所述用户界面包括:响应所述操作信号,将所述用户界面缩小为一个小窗,并将所述小窗移动到所述第一屏幕的第一指定位置;及对所述小窗的画面内容进行录制,获得所录制的视频。采用所述技术方案,无需对所录制的视频画面进行裁剪,缩短了视频录制时间。

[0008] 在一些可选的实施方式中,该方法于第一时间点 t_1 开始录制所述用户界面同时录制系统声音。

[0009] 在一些可选的实施方式中,所述第一时间点 t_1 为用户开启所述实时字幕功能的时间。采用所述技术方案,于用户开启实时字幕功能时即开始录制用户界面,及时响应用户的需求,进一步提升用户的观看体验。

[0010] 在一些可选的实施方式中,该方法还包括:于所述第一时间点t1关闭电子设备的扬声器。采用所述技术方案,使得所述视频播放区域所播放的视频的声音不能外放,避免给用户造成观看干扰。

[0011] 在一些可选的实施方式中,当所述小窗移动所述第一指定位置时,所述小窗展示所述用户界面的完整内容;或者仅展示所述视频播放区域的内容。

[0012] 在一些可选的实施方式中,所述将与所录制的视频对应的音频数据转化为文本包括:将与所录制的视频对应的音频数据上传到服务器,其中,所述服务器将与所录制的视频对应的音频数据转化为文本,并将转化获得的文本发送给所述电子设备;及接收所述服务器发送过来的文本。采用所述技术方案,由服务器完成音频转译字幕文本的步骤,由于服务器的算力和文本转化模型都比电子设备要大要强,因此可以进一步提升音频转译字幕文本的质量。

[0013] 在一些可选的实施方式中,该方法还包括:于将与所录制的视频对应的音频数据进行文本转化时,在所述第一屏幕上显示相应的文字提示。采用所述技术方案,优化了电子设备与用户之间的交互。

[0014] 在一些可选的实施方式中,该方法于第二时间点t2投影所录制的视频,其中, $t2=t1+\delta$, δ 的值的大小与将所录制的语音数据转化为文本所需的时间相关。采用所述技术方案,可以于字幕转好后即时播放所录制的视频,进一步提升用户视频观看体验。

[0015] 在一些可选的实施方式中,该方法将所录制的视频投影到所述第一屏幕的第二指定位置,所述第二指定位置与所述第一指定位置重叠或者不重叠。

[0016] 在一些可选的实施方式中,该方法将所录制的视频投影到所述电子设备的第二屏幕或者将所录制的视频投影到与所述电子设备通讯连接其他设备的屏幕上。采用所述技术方案,便利用户使用其他屏幕来观看视频,进一步提升用户的视频观看体验。

[0017] 在一些可选的实施方式中,该方法将所录制的视频放大后再投影。

[0018] 在一些可选的实施方式中,该方法将所转化得到的文本在系统最上层以浮窗形式显示。

[0019] 在一些可选的实施方式中,该方法将所转化得到的文本内嵌到所录制的视频。

[0020] 本申请第二方面公开了一种视频声音转文本的方法,应用于电子设备,所述方法包括:在电子设备的第一屏幕上显示应用的用户界面,所述用户界面包括视频播放区域,所述视频播放区域播放视频;响应用户开启实时字幕功能的操作信号,录制所述用户界面同时录制系统声音,获得所录制的多个视频,以及获得与每个视频对应的音频数据;按照所述多个视频的录制顺序将与每个视频对应的音频数据转化为文本;及按照所述录制顺序依次投影所述多个视频并输出对应的文本,同时播放对应的音频数据。

[0021] 在一些可选的实施方式中,所述多个视频中的每个视频的时长相同。

[0022] 采用所述技术方案,通过同时录制用户界面和系统声音,并在投影所录制的画面的同时播放所录制的声音及输出对应的字幕,使得所播放的画面和声音及所输出的字幕同步,提升用户视频观看体验。

[0023] 本申请第三方面公开了一种电子设备,包括处理器和存储器;所述存储器,用于存储指令;所述处理器,用于调用所述存储器中的指令,使得所述电子设备实现所述的视频声音转文本的方法。

[0024] 本申请第四方面公开了一种计算机可读存储介质,所述计算机可读存储介质存储有至少一个指令,所述至少一个指令被处理器执行时实现如所述的视频声音转文本的方法。

[0025] 第三方面至第四方面所带来的技术效果可参见上述方法部分各涉及的方法相关的描述,此处不再赘述。

附图说明

[0026] 图1A-图1B是一种视频声音转文本的方法的应用场景示意图。

[0027] 图2是本申请实施例提供的一种视频声音转文本的方法的流程示意图。

[0028] 图3A-图3D是本申请实施例提供的一种视频声音转文本的方法的应用场景示意图。

[0029] 图4A-图4C是本申请实施例提供的一种视频声音转文本的方法的应用场景示意图。

[0030] 图5A-图5C是本申请实施例提供的一种视频声音转文本的方法的应用场景示意图。

[0031] 图6A-图6D分别是本申请实施例提供的一种视频声音转文本的方法的应用场景示意图。

[0032] 图7是本申请实施例提供的一种视频声音转文本的方法的应用场景示意图。

[0033] 图8A-图8E分别是本申请实施例提供的一种视频声音转文本的方法的应用场景示意图。

[0034] 图9是本申请实施例提供的一种视频录制方法的流程示意图。

[0035] 图10A-图10B是本申请实施例提供的一种视频录制方法的应用场景示意图。

[0036] 图11是本申请实施例提供的另一种视频录制方法的流程示意图。

[0037] 图12是本申请实施例提供的另一种视频录制方法的应用场景示意图。

[0038] 图13是本申请实施例提供的另一种视频声音转文本的方法的流程示意图。

[0039] 图14A-图14D是本申请实施例提供的一种视频声音转文本的方法的应用场景示意图。

[0040] 图15是本申请实施例提供的一种利用Android的分屏、自由窗口 (freeform) 技术和虚拟屏幕显示技术来达成视频声音转文本的方法的示意图。

[0041] 图16A-图16D是本申请实施例提供的一种视频声音转文本的方法的应用场景示意图。

[0042] 图17是本申请实施例提供的电子设备的结构示意图。

具体实施方式

[0043] 需要说明的是,本申请实施例中“至少一个”是指一个或者多个,“多个”是指两个或多于两个。“和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B可以表示:单独存在A,同时存在A和B,单独存在B的情况,其中A,B可以是单数或者复数。本申请的说明书和权利要求书及附图中的术语“第一”、“第二”、“第三”、“第四”等(如果存在)是用于区别类似的对象,而不是用于描述特定的顺序或先后次序。

[0044] 图1A-图1B是一种视频声音转文本的方法的应用场景示意图。

[0045] 如图1A所示,视频中的发言人在时间点00:17的发言内容实际为“人工智能不是科学幻想”。但是,如图1B所示,发言人在时间点00:17的发言内容延迟到了时间点00:19才输出。即用户实际观看的实时字幕会有明显滞后,导致用户体验差。

[0046] 为解决上述问题,本申请提供的视频声音转文本的方法,可以避免实时字幕滞后显示的问题。具体细节下面介绍。

[0047] 图2是本申请实施例提供的第一种视频声音转文本的方法的流程图。本申请实施例提供的视频声音转文本的方法可以应用于电子设备。所述电子设备可以为手机、平板电脑等终端设备。所述电子设备可以具有一个屏幕或者两个屏幕。该电子设备还可以通过有线方式(例如USB)或者无线方式(例如WIFI)与其他设备(例如手机、平板电脑、个人电脑、服务器、电视或显示器)通讯连接。

[0048] 利用本申请提供的第一种视频声音转文本的方法,可以解决实时字幕滞后显示的技术问题,提升用户体验。

[0049] 具体地,所述第一种视频声音转文本的方法包括:

[0050] 步骤S11,在电子设备的第一屏幕上显示应用的用户界面,所述用户界面包括视频播放区域,所述视频播放区域播放视频。

[0051] 在一个实施例中,电子设备以全屏模式在所述第一屏幕上显示所述用户界面。

[0052] 在一个实施例中,在所述用户界面上,位于所述视频播放区域的下方位置还可以显示所述视频播放区域所播放的视频的相关资讯例如标题、用户对该视频的评论等。在一个实施例中,还可以在所述用户界面为用户提供针对所述视频的相关功能例如分享、下载、保存等。

[0053] 在一个实施例中,所述用户界面还包括其他资讯展示区域,该其他资讯展示区域用于展示其他的视频、文字,及/或图片等资讯。举例而言,参阅图3A所示,电子设备在第一屏幕上显示应用的用户界面10,所述用户界面10包括视频播放区域101,所述视频播放区域101正在播放视频。

[0054] 又如,参阅图4A所示,电子设备在第一屏幕上显示应用的用户界面20,所述用户界面20包括视频播放区域201,所述视频播放区域201正在播放视频。

[0055] 再如,参阅图5A所示,电子设备在第一屏幕上显示应用的用户界面30,所述用户界面30包括视频播放区域301,所述视频播放区域301正在播放视频。

[0056] 步骤S12,响应用户开启实时字幕功能的操作信号,录制所述用户界面同时录制系统声音获得所录制的视频以及与所录制的视频对应的音频数据。

[0057] 在一个实施例中,所述操作信号可以为用户按压了电子设备的指定的物理按键所产生的信号,例如该操作信号为用户同时按压了电源键和音量上调键所产生的信号。在其他实施例中,所述操作信号也可以为用户按压了显示在所述第一屏幕上的指定的虚拟按键所产生的信号。在其他实施例中,所述操作信号也可以是用户在所述第一屏幕上的手势轨迹所产生的信号,例如,所述操作信号可以为用户的两个手指同时从所述第一屏幕的上端下滑到所述第一屏幕的下端所产生的信号。

[0058] 在一个实施例中,可以于第一时间点t1开始录制所述用户界面同时录制系统声音。

[0059] 需要说明的是,与所录制的视频对应的音频数据也即是对系统声音进行录制所获得的音频数据。本实施例中,所述系统声音也即是所述用户界面的视频播放区域所播放的视频所对应的声音。

[0060] 在一个实施例中,所述第一时间点t1为用户开启所述实时字幕功能的时间,也即是接收到开启所述实时字幕功能的操作信号的时间。

[0061] 在一个实施例中,可以通过电子设备的音频编解码模块控制所述视频播放区域所播放的视频的声音不输出到所述扬声器。

[0062] 在一个实施例中,也可以于所述第一时间点t1关闭电子设备的扬声器,即将所述电子设备的扬声器静音,从而使得所述视频播放区域所播放的视频的声音不能外放。

[0063] 以下以录制视频和投影视频均在电子设备的同一屏幕上执行为例说明(简称为“第一应用场景”)。

[0064] 在第一实施例中,所述响应用户开启实时字幕功能的操作信号,录制所述用户界面包括:响应所述操作信号,将所述用户界面缩小为一个小窗,并将所述小窗移动到所述第一屏幕的第一指定位置;对所述第一屏幕的全屏画面进行录制获得所录制的全屏画面;根据所述小窗的大小和位置对所录制的全屏画面进行裁剪,从而获得所述小窗的画面内容,将所获得的所述小窗的画面内容作为所录制的视频。

[0065] 需要说明的是,所述根据所述小窗的大小和位置对所录制的全屏画面进行裁剪也即是从所录制的全屏画面中,裁剪所述小窗所在的位置区域的画面内容,从而获得所述小窗的画面内容。

[0066] 本实施例中,可以使用Android的MediaProjection相关接口对所述第一屏幕的全屏画面进行录制,获得所录制的全屏画面;以及使用FFmpeg多媒体处理库从所录制的全屏画面中,裁剪所述小窗所在的位置区域的画面内容,从而获得所述小窗的画面内容,并将所获得的所述小窗的画面内容作为所录制的视频。具体可参阅后面对图9的介绍。

[0067] 在第二实施例中,所述响应用户开启实时字幕功能的操作信号,录制所述用户界面包括:响应所述操作信号,将所述用户界面缩小为一个小窗,并将所述小窗移动到所述第一屏幕的第一指定位置;对所述小窗的画面内容进行录制,获得所录制的视频。

[0068] 具体地,可通过在Android的media framework中的MediaProjection增加一个接口,利用所增加的接口指定录制的位置区域;及对所指定的位置区域进行录制。在这里,为仅对所述小窗的画面内容进行录制,则利用所增加的接口将所述小窗的位置区域指定为录制的位置区域即可。具体地可参后面对图11的介绍。

[0069] 本实施例中,所述小窗的尺寸小于所述第一屏幕的尺寸。

[0070] 在一个实施例中,当接收到所述操作信号时,可以调用Android(安卓)的小窗模式将所述用户界面缩小为小窗。

[0071] 在一个实施例中,所述第一指定位置可以为所述第一屏幕上的任意位置。例如,所述第一指定位置可以为所述第一屏幕的右下角的位置。

[0072] 在一个实施例中,当所述小窗移动所述第一指定位置时,所述小窗展示所述用户界面的完整内容。

[0073] 在其他实施例中,当所述小窗移动所述第一指定位置时,所述小窗仅展示所述视频播放区域的内容,也即仅展示所述视频播放区域所播放的视频,即小窗的画面为画中画

的视频。

[0074] 举例而言,参阅图3A和图3B所示,电子设备响应用户开启实时字幕功能的操作信号,将用户界面10缩小成小窗11,并将所述小窗11移动到第一屏幕的右下角。所述小窗11展示的内容包括所述用户界面10的完整内容。

[0075] 再如,参阅图4A和图4B所示,电子设备响应用户开启实时字幕功能的操作信号,将用户界面20缩小成小窗21,并将所述小窗21移动到第一屏幕的右下角。所述小窗21展示的内容包括所述用户界面20的完整内容。

[0076] 又如,参阅图5A和图5B所示,电子设备响应用户开启实时字幕功能的操作信号,将用户界面30缩小成小窗31,并将所述小窗31移动到第一屏幕的右下角。所述小窗31仅展示所述用户界面30的视频播放区域的内容。

[0077] 以下以录制视频和投影视频不在电子设备的同一个屏幕为例说明(简称为“第二应用场景”)。需要说明的是,录制视频和投影视频不在电子设备的同一个屏幕可以包括两种情况。其中,第一种情况为:电子设备包括两个屏幕,在电子设备的一个屏幕上录制视频,在电子设备的另一个屏幕上投影视频。第二种情况为:在电子设备的一个屏幕上录制视频,在与电子设备通讯连接的另一个设备的屏幕上投影视频,该电子设备可以包括一个屏幕或者两个屏幕。

[0078] 在一个实施例中,所述响应用户开启实时字幕功能的操作信号,录制所述用户界面包括:响应所述操作信号,对电子设备的所述第一屏幕的全屏画面进行录制获得所录制的视频。

[0079] 同样地,可以使用Android的MediaProjection相关接口对电子设备的所述第一屏幕的全屏画面进行录制。

[0080] 举例而言,参阅图6A所示,对于具有大小不一的外屏和内屏的折叠屏(内折款)手机而言,可以在其中一个屏幕例如外屏61显示所述用户界面时,响应操作信号对外屏61的全屏画面进行录制获得所录制的视频。

[0081] 又如,参阅图6B所示,对于具有大小相同的两半屏幕的折叠屏的手机(图中示意为半展开状态)而言,可以在其中一半屏幕63显示所述用户界面时,响应操作信号对屏幕63的全屏画面进行录制获得所录制的视频。

[0082] 再如,参阅图6C所示,对于具有双屏的手机(折叠双屏)而言,可以在其中一个屏幕65显示所述用户界面时,响应操作信号对屏幕65的全屏画面进行录制获得所录制的视频。

[0083] 再如,参阅图6D所示,对于具有双屏的手机(正面和背面分别有一个屏幕)而言,可以在其中一个屏幕例如正面屏幕67显示所述用户界面时,响应操作信号对屏幕67的全屏画面进行录制获得所录制的视频。

[0084] 步骤S13,将与所录制的视频对应的音频数据转化为文本。

[0085] 本实施例中,可以利用语音识别技术(Automatic Speech Recognition,ASR)将与所录制的视频对应的音频数据转化为文本。

[0086] 在一个实施例中,可以由所述电子设备来将与所录制的视频对应的音频数据转化为文本。

[0087] 在其他实施例中,电子设备也可以将与所录制的视频对应的音频数据上传到服务器如图7所示,由所述服务器将与所录制的视频对应的音频数据转化为文本,并将转化获得

的文本发送给所述电子设备,所述电子设备接收所述服务器发送过来的文本。需要说明的是,这里电子设备将与所录制的视频对应的音频数据上传到服务器,在服务器上完成音频转译字幕文本的步骤,由于服务器的算力和文本转化模型都比电子设备要大要强,因此可以进一步提升音频转译字幕文本的质量。

[0088] 在一个实施例中,在所述第一应用场景中(也即录制视频和投影视频均在电子设备的同一屏幕上执行),若电子设备正在将与所录制的视频对应的音频数据转化为文本,则可以在所述第一屏幕上展示一张背景图。该背景图可以是如图3B或图5B所示的黑屏背景图,或者如图4B所示的截图,该截图可以是针对所述用户界面在第一时间点 t_1 的画面内容所截取的图片。在一个实施例中,还可以在所述背景图上显示相应的文字提示,例如图3B或图4B或图5B所示的文字提示“字幕处理中”以提示用户等待,由此可以更好地与用户进行交互。

[0089] 步骤S14,投影所录制的视频并输出对应的文本,同时播放与所录制的视频对应的音频数据。

[0090] 以下以应用在所述第一应用场景(也即录制视频和投影视频均在电子设备的同一屏幕上执行)为例说明:

[0091] 在一个实施例中,可以于第二时间点 t_2 将所录制的视频投影到所述第一屏幕的第二指定位置并输出对应的文本,同时播放与所录制的视频对应的音频数据。

[0092] 在一个实施例中,若由所述电子设备将所录制的视频对应的音频数据执行文本转化,则所述第二时间点 t_2 可以为所述电子设备将与所录制的视频对应的音频数据转化为了文本的时间。可选地,所述第二时间点 t_2 为电子设备从所录制的视频对应的音频数据识别到第一个文字的时间。

[0093] 在其他实施例中,若由与所述电子设备通讯连接的服务器来将所录制的视频对应的音频数据执行文本转化,则所述第二时间点 t_2 可以为所述电子设备接收到所述服务器发送过来的文本的时间。可选地,所述第二时间点 t_2 为电子设备接收到所述服务器发送过来的第一个文字的时间。

[0094] 在一个实施例中,还于所述第二时间点 t_2 开启所述扬声器,从而使得可以在所述第二时间点 t_2 外放所录制的视频对应的音频数据。

[0095] 在一个实施例中 $t_2 = t_1 + \delta$ 。其中, δ 的值可以根据所统计的有关语音转文本所需的时间数据来获得。即 δ 的值可以根据统计数据来调整。可选地, δ 等于1300ms。

[0096] 在一个实施例中,所述第二指定位置与所述第一指定位置不重叠。例如,所述第二指定位置可以为所述第一屏幕上,除所述第一指定位置之外的其他位置区域。

[0097] 在其他实施例中,所述第二指定位置可以与所述第一指定位置重叠。例如,所述第二指定位置为所述第一屏幕的整个位置区域。

[0098] 在一个实施例中,可以将所录制的视频的画面放大后投影到所述第二指定位置。本实施例中,可以通过调用Android中的VirtualDisplay、Surface或MediaCodec接口将所录制的视频的画面放大后投影到所述第二指定位置。

[0099] 在一个实施例中,当所录制的视频的内容包括所述用户界面的完整内容时,可以将所录制的视频的画面放大至全屏大小,并将放大后的所录制的视频的画面投影至所述第二指定位置如图3C或图4C所示。在这里,所述第二指定位置是指所述第一屏幕所在的整个

位置区域。

[0100] 在一个实施例中,当所录制的视频的内容包括所述用户界面的完整内容时,可以放大所录制的视频的画面,使得放大后的画面与所述小窗能够分别在所述第二指定位置和所述第一指定位置实现分屏展示如图3D所示。在这里,所述第二指定位置是指在所述第一屏幕上,除所述第一指定位置之外的其他位置区域。

[0101] 在一个实施例中,当所录制的视频的内容仅包括所述用户界面的视频播放区域所播放的视频时,可以放大所录制的视频的画面,使得放大后的画面与所述小窗能够分别在所述第二指定位置和所述第一指定位置实现分屏展示如图5C所示。在这里,所述第二指定位置是指在所述第一屏幕上,除所述第一指定位置之外的其他位置。

[0102] 在其他实施例中,所述将所录制的视频投影到所述第一屏幕的第二指定位置也可以是指:将所录制的视频的画面在系统显示界面的最上层全屏或局部显示,并将所述小窗置于系统显示界面的其他层级(即非最上层),从而使得用户无法再看见所述小窗所显示的界面画面。

[0103] 在一个实施例中,所述输出对应的文本包括:将所转化得到的文本在系统最上层以浮窗形式进行显示。

[0104] 例如,参阅图3C所示,将所转化得到的文本在系统最上层以浮窗121形式显示;参阅图图4C所示,将所转化得到的文本在系统最上层以浮窗221形式显示;参阅图图5C所示,将所转化得到的文本在系统最上层以浮窗321形式显示。

[0105] 在其他实施例中,所述输出对应的文本包括:将所转化得到的文本内嵌到所录制的视频。

[0106] 以下以应用在所述第二应用场景(也即录制视频和投影视频均不在电子设备的同一屏幕上执行)为例说明:

[0107] 在一个实施例中,可以于所述第二时间点 t_2 将所录制的视频投影到第二屏幕并在所述第二屏幕输出对应的文本,同时播放与所录制的视频对应的音频数据。所述第二屏幕可以为所述电子设备的另一个屏幕,或者可以为与所述电子设备通讯连接的一个设备的屏幕。

[0108] 以所述第二屏幕为所述电子设备的另一个屏幕为例,参阅图6A所示,可以将所录制的视频投影到屏幕62并在所述屏幕62输出对应的文本;参阅图6B所示,可以将所录制的视频投影到屏幕64并在所述屏幕64输出对应的文本;参阅图6C所示,可以将所录制的视频投影到屏幕66并在所述屏幕66输出对应的文本;参阅图6D所示,可以将所录制的视频投影到屏幕68并在所述屏幕68输出对应的文本。

[0109] 以所述电子设备为手机,所述第二屏幕为与手机通讯连接的一个设备的屏幕为例,参阅图8A所示,可以将所录制的视频投影到电视的屏幕上,并在电视的屏幕上输出对应的文本;参阅图8B所示,可以将所录制的视频投影到另一个手机的屏幕上,并在该另一个手机的屏幕上输出对应的文本;参阅图8C所示,可以将所录制的视频投影到一个显示器上,并在显示器上输出对应的文本;参阅图8D所示,可以将所录制的视频投影到一个笔记本的屏幕上,并在笔记本的屏幕上输出对应的文本;参阅图8E所示,可以将所录制的视频投影到一个平板的屏幕上,并在平板的屏幕上输出对应的文本。

[0110] 根据本申请的上述流程可知,由于本案将所录制的画面进行投影的时间相比原视

频播放的时间进行了延迟,即延迟后进行播放并输出语音对应的文本,从而使得声音和视频画面以及文本能够对齐同步。

[0111] 参阅图9所示,是本申请提供的第一种视频录制方法的流程图。

[0112] 步骤S61,使用Android中的MediaProjection相关接口录制整个屏幕。

[0113] 步骤S62,使用FFmpeg多媒体处理库对所录制的视频,指定小窗所在位置和区域进行裁剪。

[0114] 举例而言,参阅图10A所示,整个屏幕作为录制区域进行录制,因此所录制的视频的画面既包括了小窗所在位置区域的画面内容,还包括了屏幕上除小窗所在位置区域之外的其他内容。参阅图10B所示,从所录制的视频中,将小窗所在的位置区域作为裁剪区域,那么裁剪下来的区域即为所录制的画面。

[0115] 参阅图11所示,是本申请提供的第二种视频录制方法的流程图。

[0116] 步骤S81,在Android中的media framework中的MediaProjection增加一个接口,该接口用于指定录制区域。

[0117] 步骤S82,直接使用上述增加接口,指定小窗的位置和区域为录制区域并进行录制。

[0118] 举例而言,参阅图12所示,利用增加的接口指定小窗的位置和区域为录制区域并进行录制,从而所录制的视频的画面仅包括了小窗所在区域的画面内容。

[0119] 图13是本申请实施例提供的第二种视频声音转文本的方法的流程图。本申请实施例提供的第二种视频声音转文本的方法可以应用于电子设备。所述电子设备可以为手机、平板电脑等终端设备。所述电子设备可以具有一个屏幕或者两个屏幕。该电子设备还可以通过有线方式(例如USB)或者无线方式(例如WIFI)与其他设备(例如手机、平板电脑、个人电脑、服务器、电视或显示器)通讯连接。利用本申请提供的第二种视频声音转文本的方法,可以解决实时字幕滞后显示的技术问题,提升用户体验。

[0120] 具体地,所述第二种视频声音转文本的方法包括:

[0121] 步骤S101,在电子设备的第一屏幕上显示应用的用户界面,所述用户界面包括视频播放区域,所述视频播放区域播放视频。

[0122] 参阅图14A所示,电子设备在第一屏幕上显示应用的用户界面40,所述用户界面40包括视频播放区域401,所述视频播放区域401正在播放视频。

[0123] 需要说明的是,本步骤S101可以同于上述步骤S11,于此不再赘述。

[0124] 步骤S102,响应用户开启实时字幕功能的操作信号,录制所述用户界面同时录制系统声音,获得所录制的多个视频,以及获得与每个视频对应的音频数据。

[0125] 在一个实施例中,所录制的每个视频的时长为预设时长 δ_1 (例如1分钟)。对应地,与每个视频对应的音频数据的时长也为预设时长 δ_1 。

[0126] 需要说明的是,步骤S102与步骤S12的区别在于,在本步骤S102中,对所述用户界面持续录制,每录制了所述预设时长 δ_1 则生成一个视频文件,由此获得所述多个视频。同样地,对系统声音持续录制,每录制所述预设时长 δ_1 则生成一个音频文件,由此获得与每个视频对应的音频数据。同样地,本步骤S102中,每个视频的录制可以采用步骤S12中所记载的视频录制方法进行录制。也即可以采用图9所述第一种视频录制方法或者图11所示的所述第二种视频录制方法来录制,于此不再赘述。

[0127] 在本实施例中,以应用在所述第一场景(也即录制视频和投影视频均在电子设备的同一屏幕上执行)为例,可以在录制第一个视频的期间,在所述第一屏幕上展示一张背景图。该背景图可以是如图14B所示的黑屏背景图。当然该背景图也可以是针对所述用户界面在第一时间点 t_1 的画面内容所截取的图片。在一个实施例中,还可以在所述背景图上显示相应的第一文字提示,例如图14B所示的文字提示“开始录制第1分钟的视频内容,处理完字幕之后可连续播放不用等待”以提示用户等待视频录制,由此可以更好地与用户进行交互。在一个实施例中,在将所录制的第一个视频对应的音频数据转化为文本的期间,还可以在所述第一屏幕上继续展示所述背景图,并显示第二文字提示例如图14C所示的文字提示“字幕处理中,1分钟后开始播放”以提示用户等待文本转换,由此可以更好地与用户进行交互。

[0128] 步骤S103,按照所述多个视频的录制顺序将与每个视频对应的音频数据转化为文本;按照所述录制顺序依次投影所述多个视频并输出对应的文本,同时播放对应的音频数据。

[0129] 需要说明的是,在本步骤S103中,对与每个视频对应的音频数据转化为文本的操作可以同于步骤S13的记载。也即可以由电子设备将与所录制的每个视频对应的音频数据转化为文本;或者可以由电子设备将与所录制的每个视频对应的音频数据发送到服务器,由服务器来将与所录制的每个视频对应的音频数据转化为文本,并将所转化的文本发送到电子设备。

[0130] 需要说明的是,在本步骤S103中,投影每个视频的操作也可以同于步骤S14中所记载的投影所录制的视频的操作。

[0131] 还需要说明的是,本实施例中, δ_1 是每个视频以及每个视频对应的音频数据的采集时长(也即所述预设时长),若用 δ_2 表示为将每个视频对应的音频数据转化为文本字幕所需的时间,那么在本步骤S103中,按照所述录制顺序投影所录制的第一个视频的时间点 $t_2 = t_1 + \delta_1 + \delta_2$;其中, δ_1 为所述预设时长, δ_2 为将所录制的第一个视频对应的音频数据转化为文本所需的时间。需要说明的是,不同视频对应的音频数据转化为文本所需要的时长 δ_2 可能不同。

[0132] 在本步骤S103中,在投影完所述第一个视频之后,则按照所述录制顺序投影所录制的下一个视频并输出对应的文本,同时播放与所述下一个视频对应的音频数据。依次类推,实现非串流音频字幕对齐。

[0133] 还需要说明的是,在本实施例中,所述多个视频的录制是持续进行的,即每录制完一个所述预设时长 δ_1 的视频则进行下一个视频的录制。即该下一个视频的录制无需等到上一个视频对应的音频数据转化为了文本才进行。同样地,对所录制的上一个音频数据执行完文本转化即可开始对所录制的下一个音频数据执行文本转化。另外,一旦投影的前一个视频播放完成,只要下一个视频对应的音频数据完成了文本转化,即可开始投影该下一个视频,由此保证视频及音频数据的连续处理和播放。这里需要说明的是,由于不同视频对应的音频数据转化为文本所需要的时长 δ_2 可能不同,若投影的前一个视频播放完成,但是下一个视频对应的音频数据还没有完成文本转化,则可以在屏幕上再次显示“字幕处理中”的提示并暂停投影。

[0134] 仍然以应用在所述第一场景(也即录制视频和投影视频均在电子设备的同一屏幕上执行)为例,参阅图14D所示,将录制的视频的画面放大后投影至第二指定位置42,以及将

对应的文本以浮窗421的形式进行显示。

[0135] 根据本申请的上述流程可知,本申请通过对用户界面进行分段录制,只要录制完上一段,马上继续录制下一段视频,并按照录制顺序将每段视频对应的音频数据文本转化后即投影视频,由此保证了视频连续处理和播放,也即本申请利用非串流音频字幕对齐方法,对串流字幕对齐方法进行了改进。

[0136] 需要说明的是,本申请提供的上述方案是利用了Android的分屏、自由窗口(freeform)技术和虚拟屏幕显示技术来达成上述效果。具体实现方案可参阅对图15的介绍。

[0137] 预先在Android系统的Activity Manager中,构造一个virtualscreen_workspace_stack(stack_id=5),属性类似fullscreen_workspace_stack(stack_id=1),其中,virtualscreen_workspace_stack与fullscreen_workspace_stack的区别是显示位置为Android的一个虚拟屏幕(virtual display);在Android系统的Activity Manager中,构造一个新方法接口moveActivityToVirtualStack,类似AMS.startActivityFromRecents(),可把本来在fullscreen_workspace_stack(stack_id=1)的Activity移动到virtualscreen_workspace_stack(stack_id=5)。

[0138] 若用户启动AI字幕(例如图16A所示接收到用户启动开启实时字幕功能的操作信号);系统调用AMS.moveActivityToVirtualStack把原始Activity移动到虚拟屏幕栈中(如图16B和16C所示);在虚拟屏幕中渲染该Activity的界面;及录制该虚拟屏幕的画面并投影到设备的主屏上,并显示对齐后的字幕(如图16D所示)。

[0139] 图17是本申请实施例提供的电子设备100的结构示意图。电子设备100可以包括处理器110,外部存储器接口120,内部存储器121,通用串行总线(universal serial bus,USB)接口130,充电管理模块140,电源管理模块141,电池142,天线1,天线2,移动通信模块150,无线通信模块160,音频模块170,扬声器170A,受话器170B,麦克风170C,耳机接口170D,传感器模块180,按键190,马达191,指示器192,摄像头193,显示屏194,以及用户标识模块(subscriber identification module,SIM)卡接口195等。其中传感器模块180可以包括压力传感器180A,陀螺仪传感器180B,气压传感器180C,磁传感器180D,加速度传感器180E,距离传感器180F,接近光传感器180G,指纹传感器180H,温度传感器180J,触摸传感器180K,环境光传感器180L,骨传导传感器180M等。

[0140] 可以理解的是,本申请实施例示意的结构并不构成对电子设备100的具体限定。在本申请另一些实施例中,电子设备100可以包括比图示更多或更少的部件,或者组合某些部件,或者拆分某些部件,或者不同的部件布置。图示的部件可以以硬件,软件或软件和硬件的组合实现。

[0141] 处理器110可以包括一个或多个处理单元,例如:处理器110可以包括应用处理器(application processor,AP),调制解调处理器,图形处理器(graphics processing unit,GPU),图像信号处理器(image signal processor,ISP),控制器,存储器,视频编解码器,数字信号处理器(digital signal processor,DSP),基带处理器,和/或神经网络处理器(neural-network processing unit,NPU)等。其中,不同的处理单元可以是独立的器件,也可以集成在一个或多个处理器中。

[0142] 其中,控制器可以是电子设备100的神经中枢和指挥中心。控制器可以根据指令操

作码和时序信号,产生操作控制信号,完成取指令和执行指令的控制。

[0143] 处理器110中还可以设置存储器,用于存储指令和数据。在一些实施例中,处理器110中的存储器为高速缓冲存储器。所述存储器可以保存处理器110刚用过或循环使用的指令或数据。如果处理器110需要再次使用所述指令或数据,可从所述存储器中直接调用。避免了重复存取,减少了处理器110的等待时间,因而提高了系统的效率。

[0144] 在一些实施例中,处理器110可以包括一个或多个接口。接口可以包括集成电路(inter-integrated circuit,I2C)接口,集成电路内置音频(inter-integrated circuit sound,I2S)接口,脉冲编码调制(pulse code modulation,PCM)接口,通用异步收发传输器(universal asynchronous receiver/transmitter,UART)接口,移动产业处理器接口(mobile industry processor interface,MIPI),通用输入输出(general-purpose input/output,GPIO)接口,用户标识模块(subscriber identity module,SIM)接口,和/或通用串行总线(universal serial bus,USB)接口等。

[0145] I2C接口是一种双向同步串行总线,包括一根串行数据线(serial data line,SDA)和一根串行时钟线(derail clock line,SCL)。在一些实施例中,处理器110可以包含多组I2C总线。处理器110可以通过不同的I2C总线接口分别耦合触摸传感器180K,充电器,闪光灯,摄像头193等。例如:处理器110可以通过I2C接口耦合触摸传感器180K,使处理器110与触摸传感器180K通过I2C总线接口通信,实现电子设备100的触摸功能。

[0146] I2S接口可以用于音频通信。在一些实施例中,处理器110可以包含多组I2S总线。处理器110可以通过I2S总线与音频模块170耦合,实现处理器110与音频模块170之间的通信。在一些实施例中,音频模块170可以通过I2S接口向无线通信模块160传递音频信号,实现通过蓝牙耳机接听电话的功能。

[0147] PCM接口也可以用于音频通信,将模拟信号抽样,量化和编码。在一些实施例中,音频模块170与无线通信模块160可以通过PCM总线接口耦合。在一些实施例中,音频模块170也可以通过PCM接口向无线通信模块160传递音频信号,实现通过蓝牙耳机接听电话的功能。所述I2S接口和所述PCM接口都可以用于音频通信。

[0148] UART接口是一种通用串行数据总线,用于异步通信。所述总线可以为双向通信总线。它将要传输的数据在串行通信与并行通信之间转换。在一些实施例中,UART接口通常被用于连接处理器110与无线通信模块160。例如:处理器110通过UART接口与无线通信模块160中的蓝牙模块通信,实现蓝牙功能。在一些实施例中,音频模块170可以通过UART接口向无线通信模块160传递音频信号,实现通过蓝牙耳机播放音乐的功能。

[0149] MIPI接口可以被用于连接处理器110与显示屏194,摄像头193等外围器件。MIPI接口包括摄像头串行接口(camera serial interface,CSI),显示屏串行接口(display serial interface,DSI)等。在一些实施例中,处理器110和摄像头193通过CSI接口通信,实现电子设备100的拍摄功能。处理器110和显示屏194通过DSI接口通信,实现电子设备100的显示功能。

[0150] GPIO接口可以通过软件配置。GPIO接口可以被配置为控制信号,也可被配置为数据信号。在一些实施例中,GPIO接口可以用于连接处理器110与摄像头193,显示屏194,无线通信模块160,音频模块170,传感器模块180等。GPIO接口还可以被配置为I2C接口,I2S接口,UART接口,MIPI接口等。

[0151] USB接口130是符合USB标准规范的接口,具体可以是Mini USB接口, Micro USB接口, USB Type C接口等。USB接口130可以用于连接充电器为电子设备100充电,也可以用于电子设备100与外围设备之间传输数据。也可以用于连接耳机,通过耳机播放音频。所述接口还可以用于连接其他电子设备,例如AR设备等。

[0152] 可以理解的是,本申请实施例示意的各模块间的接口连接关系,只是示意性说明,并不构成对电子设备100的结构限定。在本申请另一些实施例中,电子设备100也可以采用上述实施例中不同的接口连接方式,或多种接口连接方式的组合。

[0153] 充电管理模块140用于从充电器接收充电输入。其中,充电器可以是无线充电器,也可以是有线充电器。在一些有线充电的实施例中,充电管理模块140可以通过USB接口130接收有线充电器的充电输入。在一些无线充电的实施例中,充电管理模块140可以通过电子设备100的无线充电线圈接收无线充电输入。充电管理模块140为电池142充电的同时,还可以通过电源管理模块141为电子设备供电。

[0154] 电源管理模块141用于连接电池142,充电管理模块140与处理器110。电源管理模块141接收电池142和/或充电管理模块140的输入,为处理器110,内部存储器121,外部存储器,显示屏194,摄像头193,和无线通信模块160等供电。电源管理模块141还可以用于监测电池容量,电池循环次数,电池健康状态(漏电,阻抗)等参数。在其他一些实施例中,电源管理模块141也可以设置于处理器110中。在另一些实施例中,电源管理模块141和充电管理模块140也可以设置于同一个器件中。

[0155] 电子设备100的无线通信功能可以通过天线1,天线2,移动通信模块150,无线通信模块160,调制解调处理器以及基带处理器等实现。

[0156] 天线1和天线2用于发射和接收电磁波信号。电子设备100中的每个天线可用于覆盖单个或多个通信频带。不同的天线还可以复用,以提高天线的利用率。例如:可以将天线1复用为无线局域网的分集天线。在另外一些实施例中,天线可以和调谐开关结合使用。

[0157] 移动通信模块150可以提供应用在电子设备100上的包括2G/3G/4G/5G等无线通信的解决方案。移动通信模块150可以包括至少一个滤波器,开关,功率放大器,低噪声放大器(low noise amplifier,LNA)等。移动通信模块150可以由天线1接收电磁波,并对接收的电磁波进行滤波,放大等处理,传送至调制解调处理器进行解调。移动通信模块150还可以对经调制解调处理器调制后的信号放大,经天线1转为电磁波辐射出去。在一些实施例中,移动通信模块150的至少部分功能模块可以被设置于处理器110中。在一些实施例中,移动通信模块150的至少部分功能模块可以与处理器110的至少部分模块被设置在同一个器件中。

[0158] 调制解调处理器可以包括调制器和解调器。其中,调制器用于将待发送的低频基带信号调制为中高频信号。解调器用于将接收的电磁波信号解调为低频基带信号。随后解调器将解调得到的低频基带信号传送至基带处理器处理。低频基带信号经基带处理器处理后,被传递给应用处理器。应用处理器通过音频设备(不限于扬声器170A,受话器170B等)输出声音信号,或通过显示屏194显示图像或视频。在一些实施例中,调制解调处理器可以是独立的器件。在另一些实施例中,调制解调处理器可以独立于处理器110,与移动通信模块150或其他功能模块设置在同一个器件中。

[0159] 无线通信模块160可以提供应用在电子设备100上的包括无线局域网(wireless local area networks,WLAN)(如无线保真(wireless fidelity,Wi-Fi)网络),蓝牙

(bluetooth,BT),全球导航卫星系统(global navigation satellite system,GNSS),调频(frequency modulation,FM),近距离无线通信技术(near field communication,NFC),红外技术(infrared,IR)等无线通信的解决方案。无线通信模块160可以是集成至少一个通信处理模块的一个或多个器件。无线通信模块160经由天线2接收电磁波,将电磁波信号调频以及滤波处理,将处理后的信号发送到处理器110。无线通信模块160还可以从处理器110接收待发送的信号,对其进行调频,放大,经天线2转为电磁波辐射出去。

[0160] 在一些实施例中,电子设备100的天线1和移动通信模块150耦合,天线2和无线通信模块160耦合,使得电子设备100可以通过无线通信技术与网络以及其他设备通信。所述无线通信技术可以包括全球移动通讯系统(global system for mobile communications,GSM),通用分组无线服务(general packet radio service,GPRS),码分多址接入(code division multiple access,CDMA),宽带码分多址(wideband code division multiple access,WCDMA),时分码分多址(time-division code division multiple access,TD-SCDMA),长期演进(long term evolution,LTE),第五代无线通信系统(5G,the 5th Generation of wireless communication system),BT,GNSS,WLAN,NFC,FM,和/或IR技术等。所述GNSS可以包括全球卫星定位系统(global positioning system,GPS),全球导航卫星系统(global navigation satellite system,GLONASS),北斗卫星导航系统(beidou navigation satellite system,BDS),准天顶卫星系统(quasi-zenith satellite system,QZSS)和/或星基增强系统(satellite based augmentation systems,SBAS)。

[0161] 电子设备100通过GPU,显示屏194,以及应用处理器等实现显示功能。GPU为图像处理的微处理器,连接显示屏194和应用处理器。GPU用于执行数学和几何计算,用于图形渲染。处理器110可包括一个或多个GPU,其执行程序指令以生成或改变显示信息。

[0162] 显示屏194用于显示图像,视频等。显示屏194包括显示面板。显示面板可以采用液晶显示屏(liquid crystal display,LCD),有机发光二极管(organic light-emitting diode,OLED),有源矩阵有机发光二极体或主动矩阵有机发光二极体(active-matrix organic light emitting diode的,AMOLED),柔性发光二极管(flex light-emitting diode,FLED),Miniled,MicroLed,Micro-oLed,量子点发光二极管(quantum dot light emitting diodes,QLED)等。在一些实施例中,电子设备100可以包括1个或N个显示屏194,N为大于1的正整数。

[0163] 电子设备100可以通过ISP,摄像头193,视频编解码器,GPU,显示屏194以及应用处理器等实现拍摄功能。

[0164] ISP用于处理摄像头193反馈的数据。例如,拍照时,打开快门,光线通过镜头被传递到摄像头感光元件上,光信号转换为电信号,摄像头感光元件将所述电信号传递给ISP处理,转化为肉眼可见的图像。ISP还可以对图像的噪点,亮度,肤色进行算法优化。ISP还可以对拍摄场景的曝光,色温等参数优化。在一些实施例中,ISP可以设置在摄像头193中。

[0165] 摄像头193用于捕获静态图像或视频。物体通过镜头生成光学图像投射到感光元件。感光元件可以是电荷耦合器件(charge coupled device,CCD)或互补金属氧化物半导体(complementary metal-oxide-semiconductor,CMOS)光电晶体管。感光元件把光信号转换成电信号,之后将电信号传递给ISP转换成数字图像信号。ISP将数字图像信号输出到DSP加工处理。DSP将数字图像信号转换成标准的RGB,YUV等格式的图像信号。在一些实施例中,

电子设备100可以包括1个或N个摄像头193,N为大于1的正整数。

[0166] 数字信号处理器用于处理数字信号,除了可以处理数字图像信号,还可以处理其他数字信号。例如,当电子设备100在频点选择时,数字信号处理器用于对频点能量进行傅里叶变换等。

[0167] 视频编解码器用于对数字视频压缩或解压缩。电子设备100可以支持一种或多种视频编解码器。这样,电子设备100可以播放或录制多种编码格式的视频,例如:动态图像专家组(moving picture experts group,MPEG)1,MPEG2,MPEG3,MPEG4等。

[0168] NPU为神经网络(neural-network,NN)计算处理器,通过借鉴生物神经网络结构,例如借鉴人脑神经元之间传递模式,对输入信息快速处理,还可以不断的自学习。通过NPU可以实现电子设备100的智能认知等应用,例如:图像识别,人脸识别,语音识别,文本理解等。

[0169] 外部存储器接口120可以用于连接外部存储卡,例如Micro SD卡,实现扩展电子设备100的存储能力。外部存储卡通过外部存储器接口120与处理器110通信,实现数据存储功能。例如将音乐,视频等文件保存在外部存储卡中。

[0170] 内部存储器121可以用于存储计算机可执行程序代码,所述可执行程序代码包括指令。处理器110通过运行存储在内部存储器121的指令,从而执行电子设备100的各种功能应用以及数据处理。内部存储器121可以包括存储程序区和存储数据区。其中,存储程序区可存储操作系统,至少一个功能所需的应用程序(比如声音播放功能,图像播放功能等)等。存储数据区可存储电子设备100使用过程中所创建的数据(比如音频数据,电话本等)等。此外,内部存储器121可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件,闪存器件,通用闪存存储器(universal flash storage,UFS)等。

[0171] 电子设备100可以通过音频模块170,扬声器170A,受话器170B,麦克风170C,耳机接口170D,以及应用处理器等实现音频功能。例如音乐播放,录音等。

[0172] 音频模块170用于将数字音频信息转换成模拟音频信号输出,也用于将模拟音频输入转换为数字音频信号。音频模块170还可以用于对音频信号编码和解码。在一些实施例中,音频模块170可以设置于处理器110中,或将音频模块170的部分功能模块设置于处理器110中。

[0173] 扬声器170A,也称“喇叭”,用于将音频电信号转换为声音信号。电子设备100可以通过扬声器170A收听音乐,或收听免提通话。

[0174] 受话器170B,也称“听筒”,用于将音频电信号转换为声音信号。当电子设备100接听电话或语音信息时,可以通过将受话器170B靠近人耳接听语音。

[0175] 麦克风170C,也称“话筒”,“传声器”,用于将声音信号转换为电信号。当拨打电话或发送语音信息时,用户可以通过人嘴靠近麦克风170C发声,将声音信号输入到麦克风170C。电子设备100可以设置至少一个麦克风170C。在另一些实施例中,电子设备100可以设置两个麦克风170C,除了采集声音信号,还可以实现降噪功能。在另一些实施例中,电子设备100还可以设置三个,四个或更多麦克风170C,实现采集声音信号,降噪,还可以识别声音来源,实现定向录音功能等。

[0176] 耳机接口170D用于连接有线耳机。耳机接口170D可以是USB接口130,也可以是3.5mm的开放移动电子设备平台(open mobile terminal platform,OMTP)标准接口,美国

蜂窝电信工业协会 (cellular telecommunications industry association of the USA, CTIA) 标准接口。

[0177] 压力传感器180A用于感受压力信号,可以将压力信号转换成电信号。在一些实施例中,压力传感器180A可以设置于显示屏194。压力传感器180A的种类很多,如电阻式压力传感器,电感式压力传感器,电容式压力传感器等。电容式压力传感器可以是包括至少两个具有导电材料的平行板。当有力作用于压力传感器180A,电极之间的电容改变。电子设备100根据电容的变化确定压力的强度。当有触摸操作作用于显示屏194,电子设备100根据压力传感器180A检测所述触摸操作强度。电子设备100也可以根据压力传感器180A的检测信号计算触摸的位置。在一些实施例中,作用于相同触摸位置,但不同触摸操作强度的触摸操作,可以对应不同的操作指令。例如:当有触摸操作强度小于第一压力阈值的触摸操作作用于短消息应用图标时,执行查看短消息的指令。当有触摸操作强度大于或等于第一压力阈值的触摸操作作用于短消息应用图标时,执行新建短消息的指令。

[0178] 陀螺仪传感器180B可以用于确定电子设备100的运动姿态。在一些实施例中,可以通过陀螺仪传感器180B确定电子设备100围绕三个轴(即,x,y和z轴)的角速度。陀螺仪传感器180B可以用于拍摄防抖。示例性的,当按下快门,陀螺仪传感器180B检测电子设备100抖动的角度,根据角度计算出镜头模组需要补偿的距离,让镜头通过反向运动抵消电子设备100的抖动,实现防抖。陀螺仪传感器180B还可以用于导航,体感游戏场景。

[0179] 气压传感器180C用于测量气压。在一些实施例中,电子设备100通过气压传感器180C测得的气压值计算海拔高度,辅助定位和导航。

[0180] 磁传感器180D包括霍尔传感器。电子设备100可以利用磁传感器180D检测翻盖皮套的开合。在一些实施例中,当电子设备100是翻盖机时,电子设备100可以根据磁传感器180D检测翻盖的开合。进而根据检测到的皮套的开合状态或翻盖的开合状态,设置翻盖自动解锁等特性。

[0181] 加速度传感器180E可检测电子设备100在各个方向上(一般为三轴)加速度的大小。当电子设备100静止时可检测出重力的大小及方向。还可以用于识别电子设备姿态,应用于横竖屏切换,计步器等应用。

[0182] 距离传感器180F,用于测量距离。电子设备100可以通过红外或激光测量距离。在一些实施例中,拍摄场景,电子设备100可以利用距离传感器180F测距以实现快速对焦。

[0183] 接近光传感器180G可以包括例如发光二极管(LED)和光检测器,例如光电二极管。发光二极管可以是红外发光二极管。电子设备100通过发光二极管向外发射红外光。电子设备100使用光电二极管检测来自附近物体的红外反射光。当检测到充分的反射光时,可以确定电子设备100附近有物体。当检测到不充分的反射光时,电子设备100可以确定电子设备100附近没有物体。电子设备100可以利用接近光传感器180G检测用户手持电子设备100贴近耳朵通话,以便自动熄灭屏幕达到省电的目的。接近光传感器180G也可用于皮套模式,口袋模式自动解锁与锁屏。

[0184] 环境光传感器180L用于感知环境光亮度。电子设备100可以根据感知的环境光亮度自适应调节显示屏194亮度。环境光传感器180L也可用于拍照时自动调节白平衡。环境光传感器180L还可以与接近光传感器180G配合,检测电子设备100是否在口袋里,以防误触。

[0185] 指纹传感器180H用于采集指纹。电子设备100可以利用采集的指纹特性实现指纹

解锁,访问应用锁,指纹拍照,指纹接听来电等。

[0186] 温度传感器180J用于检测温度。在一些实施例中,电子设备100利用温度传感器180J检测的温度,执行温度处理策略。例如,当温度传感器180J上报的温度超过阈值,电子设备100执行降低位于温度传感器180J附近的处理器的性能,以便降低功耗实施热保护。在另一些实施例中,当温度低于另一阈值时,电子设备100对电池142加热,以避免低温导致电子设备100异常关机。在其他一些实施例中,当温度低于又一阈值时,电子设备100对电池142的输出电压执行升压,以避免低温导致的异常关机。

[0187] 触摸传感器180K,也称“触控面板”。触摸传感器180K可以设置于显示屏194,由触摸传感器180K与显示屏194组成触摸屏,也称“触控屏”。触摸传感器180K用于检测作用于其上或附近的触摸操作。触摸传感器可以将检测到的触摸操作传递给应用处理器,以确定触摸事件类型。可以通过显示屏194提供与触摸操作相关的视觉输出。在另一些实施例中,触摸传感器180K也可以设置于电子设备100的表面,与显示屏194所处的位置不同。

[0188] 骨传导传感器180M可以获取振动信号。在一些实施例中,骨传导传感器180M可以获取人体声部振动骨块的振动信号。骨传导传感器180M也可以接触人体脉搏,接收血压跳动信号。在一些实施例中,骨传导传感器180M也可以设置于耳机中,结合成骨传导耳机。音频模块170可以基于所述骨传导传感器180M获取的声部振动骨块的振动信号,解析出语音信号,实现语音功能。应用处理器可以基于所述骨传导传感器180M获取的血压跳动信号解析心率信息,实现心率检测功能。

[0189] 按键190包括开机键,音量键等。按键190可以是机械按键。也可以是触摸式按键。电子设备100可以接收按键输入,产生与电子设备100的用户设置以及功能控制有关的键信号输入。

[0190] 马达191可以产生振动提示。马达191可以用于来电振动提示,也可以用于触摸振动反馈。例如,作用于不同应用(例如拍照,音频播放等)的触摸操作,可以对应不同的振动反馈效果。作用于显示屏194不同区域的触摸操作,马达191也可对应不同的振动反馈效果。不同的应用场景(例如:时间提醒,接收信息,闹钟,游戏等)也可以对应不同的振动反馈效果。触摸振动反馈效果还可以支持自定义。

[0191] 指示器192可以是指示灯,可以用于指示充电状态,电量变化,也可以用于指示消息,未接来电,通知等。

[0192] SIM卡接口195用于连接SIM卡。SIM卡可以通过插入SIM卡接口195,或从SIM卡接口195拔出,实现和电子设备100的接触和分离。电子设备100可以支持1个或N个SIM卡接口,N为大于1的正整数。SIM卡接口195可以支持Nano SIM卡, Micro SIM卡, SIM卡等。同一个SIM卡接口195可以同时插入多张卡。所述多张卡的类型可以相同,也可以不同。SIM卡接口195也可以兼容不同类型的SIM卡。SIM卡接口195也可以兼容外部存储卡。电子设备100通过SIM卡和网络交互,实现通话以及数据通信等功能。在一些实施例中,电子设备100采用eSIM,即:嵌入式SIM卡。eSIM卡可以嵌在电子设备100中,不能和电子设备100分离。

[0193] 电子设备100的软件系统可以采用分层架构,事件驱动架构,微核架构,微服务架构,或云架构。本申请实施例以分层架构的Android系统为例,示例性说明电子设备100的软件结构。

[0194] 所述电子设备100集成的模块如果以软件功能模块的形式实现并作为独立的产品

销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请实现上述实施例方法中的全部或部分流程,也可以通过计算机可读指令来指令相关的硬件来完成,所述的计算机可读指令可存储于一计算机可读存储介质中,所述计算机可读指令在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机可读指令包括计算机可读指令代码,所述计算机可读指令代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机可读指令代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM)、随机存取存储器(RAM)等。

[0195] 本实施例还提供一种计算机存储介质,所述计算机存储介质中存储有计算机指令,当所述计算机指令在电子设备上运行时,使得电子设备执行上述相关方法步骤实现上述实施例中的视频声音转文本的方法。

[0196] 本实施例还提供了一种计算机程序产品,当所述计算机程序产品在电子设备上运行时,使得电子设备执行上述相关步骤,以实现上述实施例中的视频声音转文本的方法。

[0197] 另外,本申请的实施例还提供一种装置,这个装置具体可以是芯片,组件或模块,所述装置可包括相连的处理器和存储器;其中,存储器用于存储计算机执行指令,当装置运行时,处理器可执行存储器存储的计算机执行指令,以使芯片执行上述各方法实施例中的视频声音转文本的方法。

[0198] 其中,本实施例提供的电子设备、计算机存储介质、计算机程序产品或芯片均用于执行上文所提供的对应的方法,因此,其所能达到的有益效果可参考上文所提供的对应的方法中的有益效果,此处不再赘述。

[0199] 通过以上的实施方式的描述,所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将装置的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。

[0200] 在本申请所提供的几个实施例中,应所述理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个装置,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0201] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是一个物理单元或多个物理单元,即可以位于一个地方,或者也可以分布到多个不同地方。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0202] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0203] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个可读取存储介质中。基于这样的理解,本申请实施例的技术方案本质上

或者说对现有技术做出贡献的部分或者所述技术方案的全部或部分可以以软件产品的形式体现出来,所述软件产品存储在一个存储介质中,包括若干指令用以使得一个设备(可以是单片机,芯片等)或处理器(processor)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory, RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0204] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何在本申请揭露的技术范围内的变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以权利要求的保护范围为准。

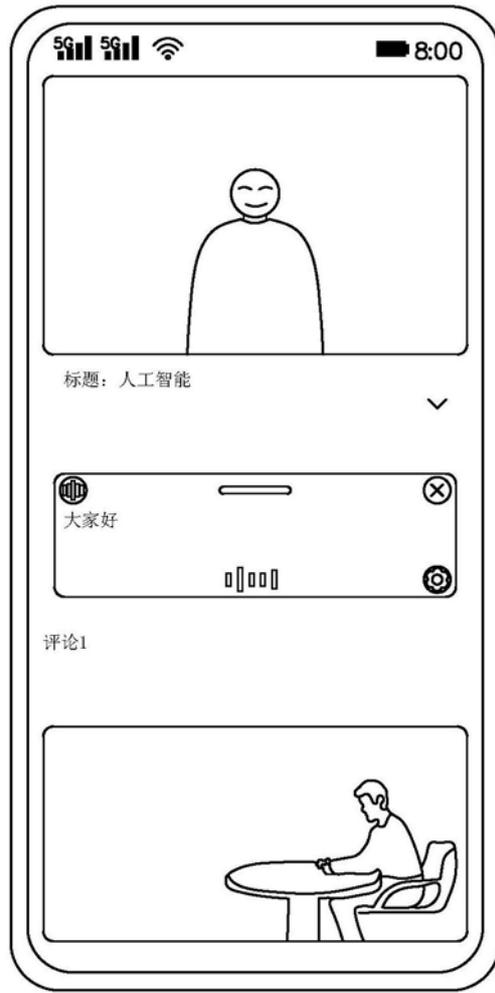


图1A

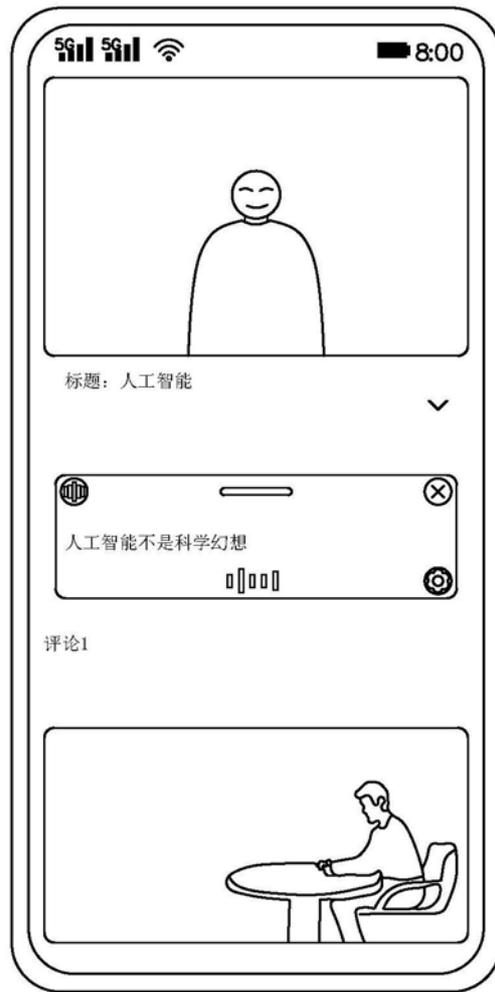


图1B

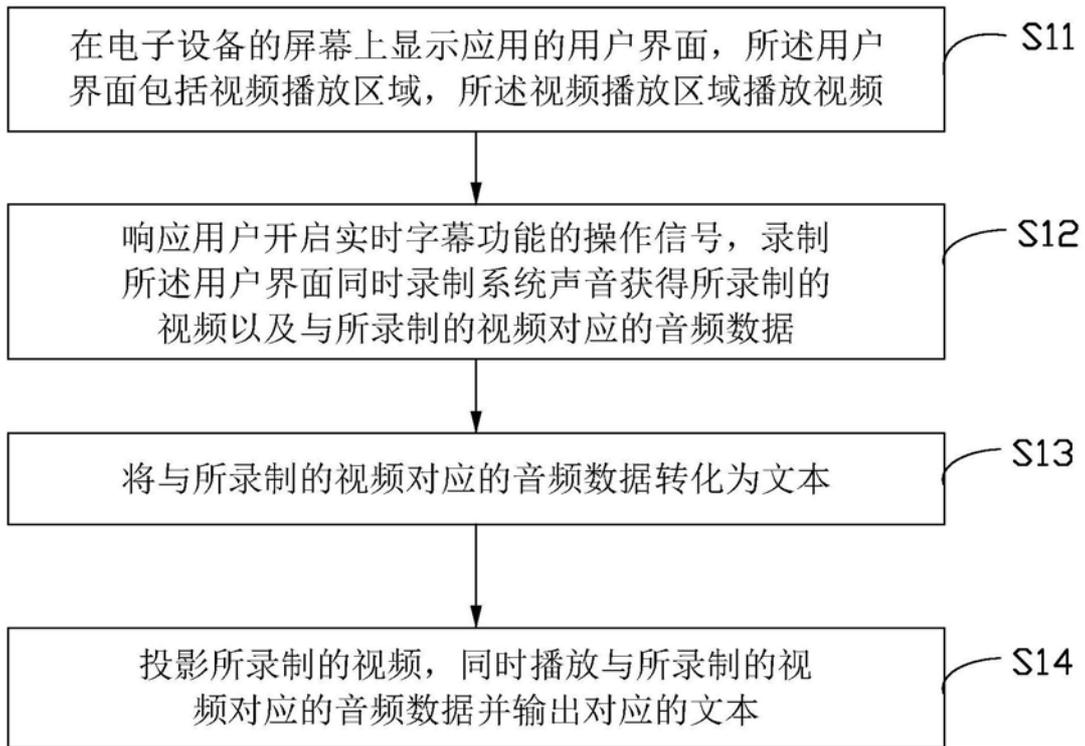


图2

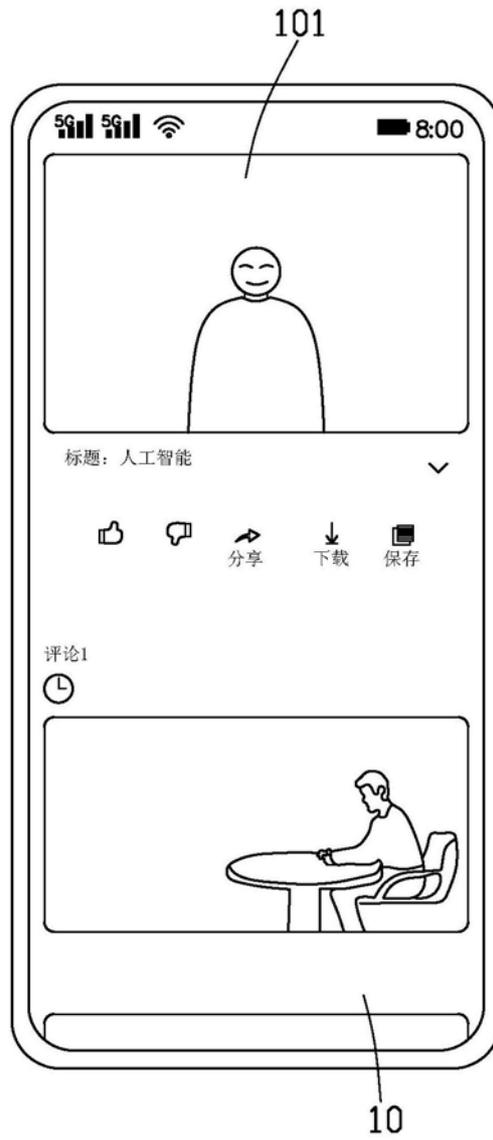


图3A

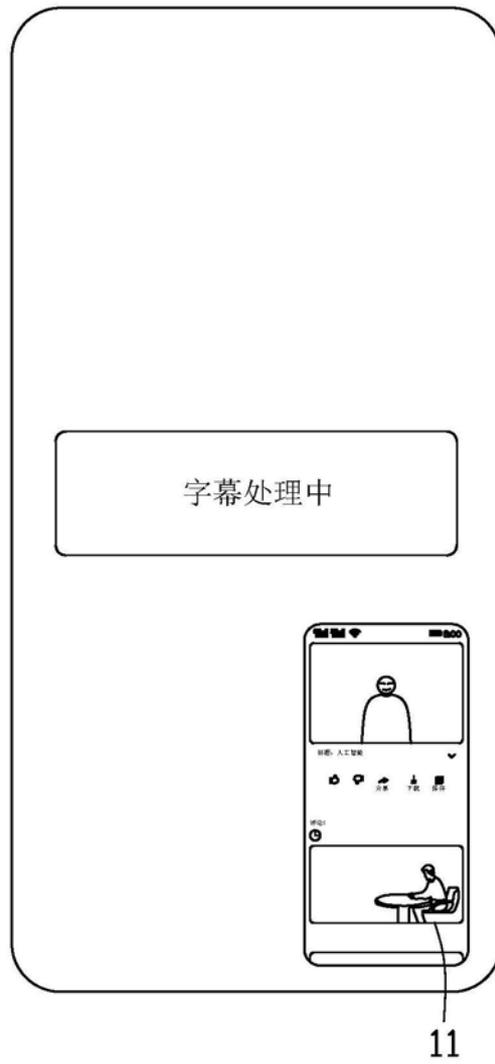


图3B

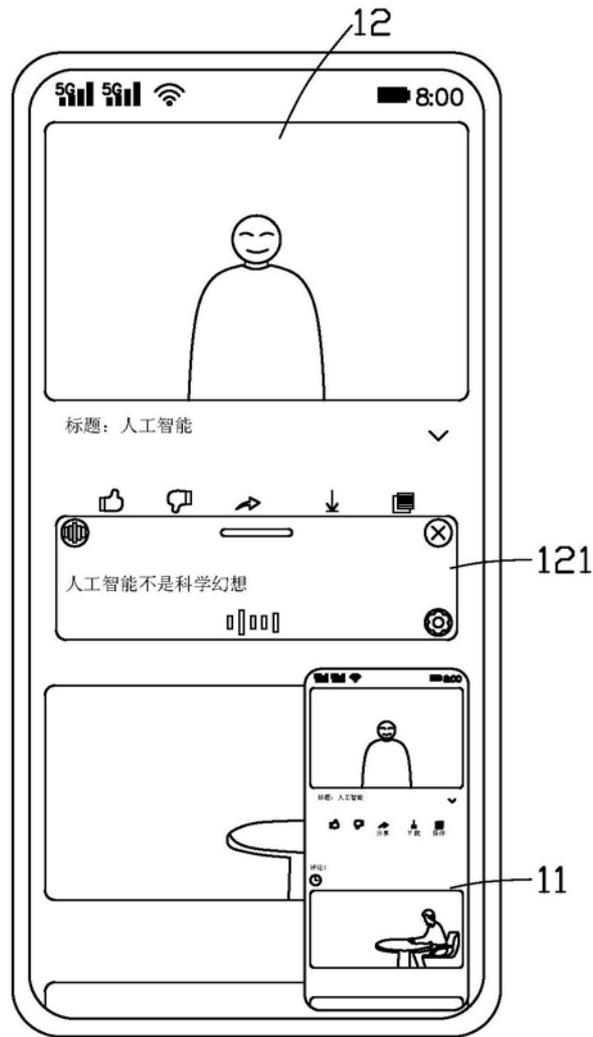


图3C

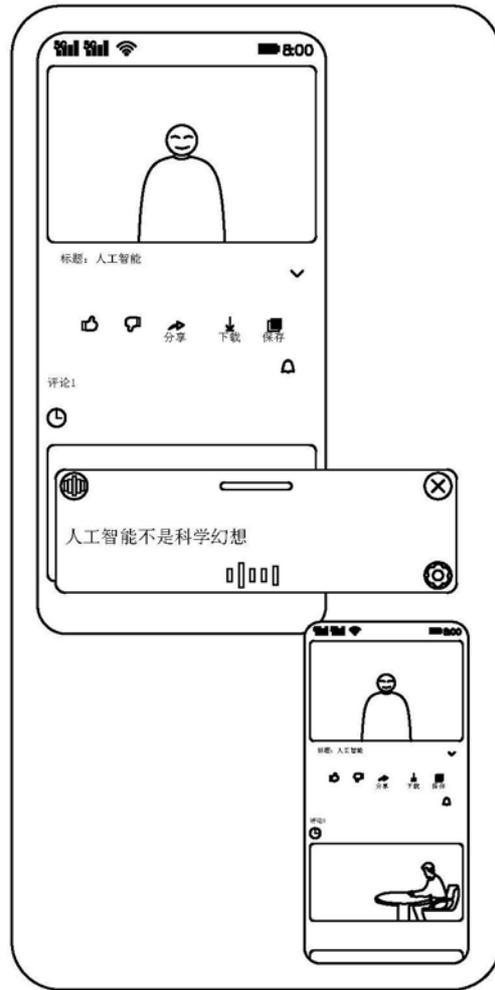


图3D

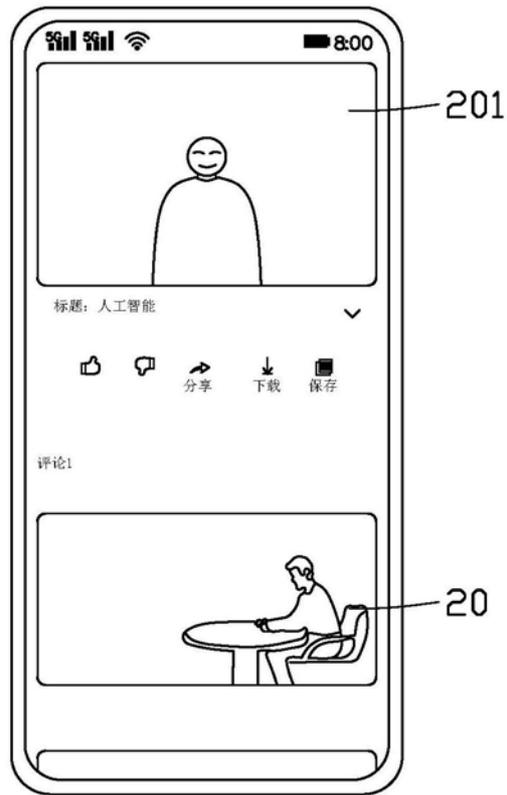


图4A

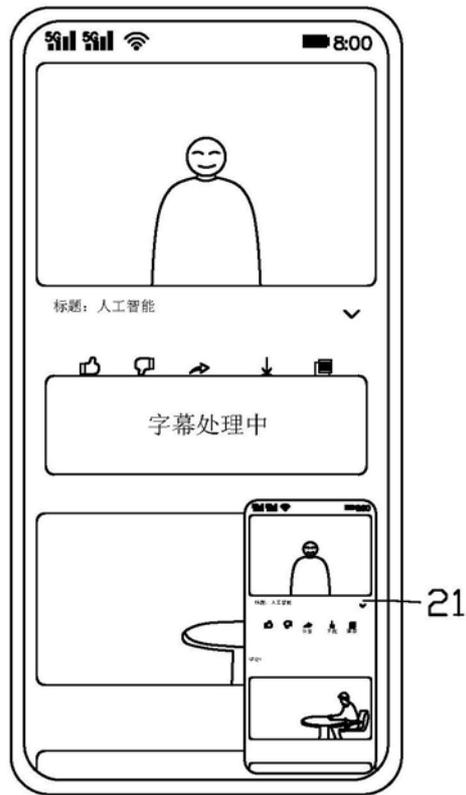


图4B

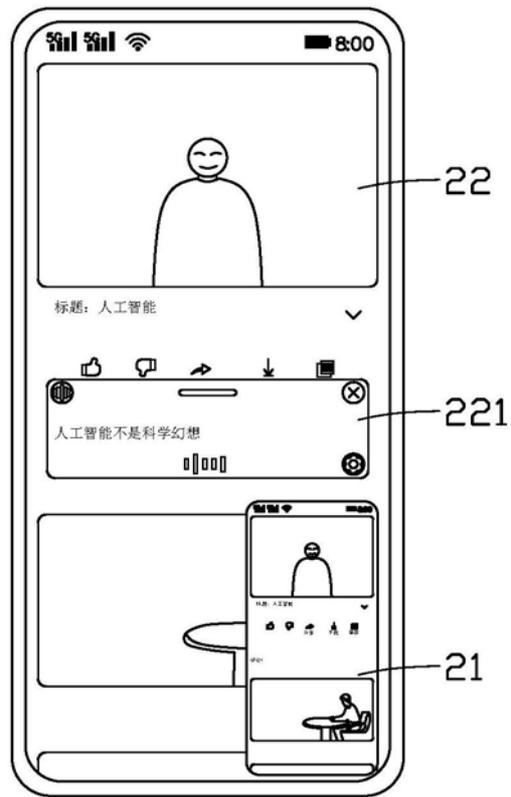


图4C



图5A



图5B

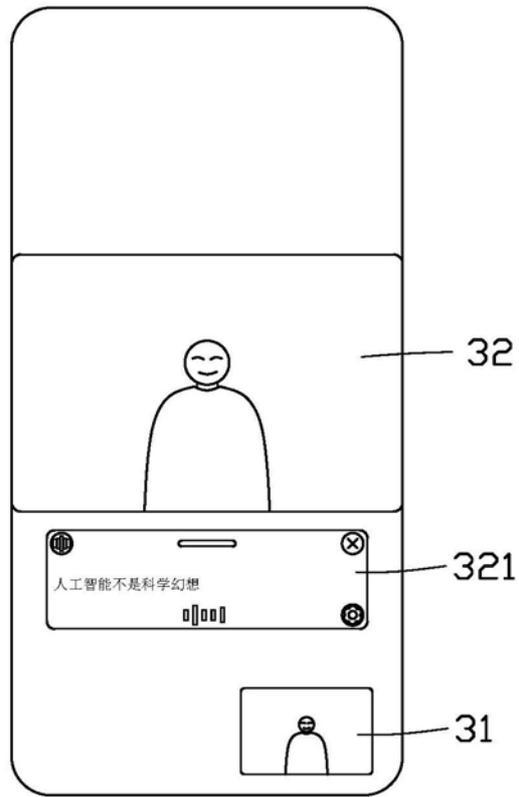


图5C

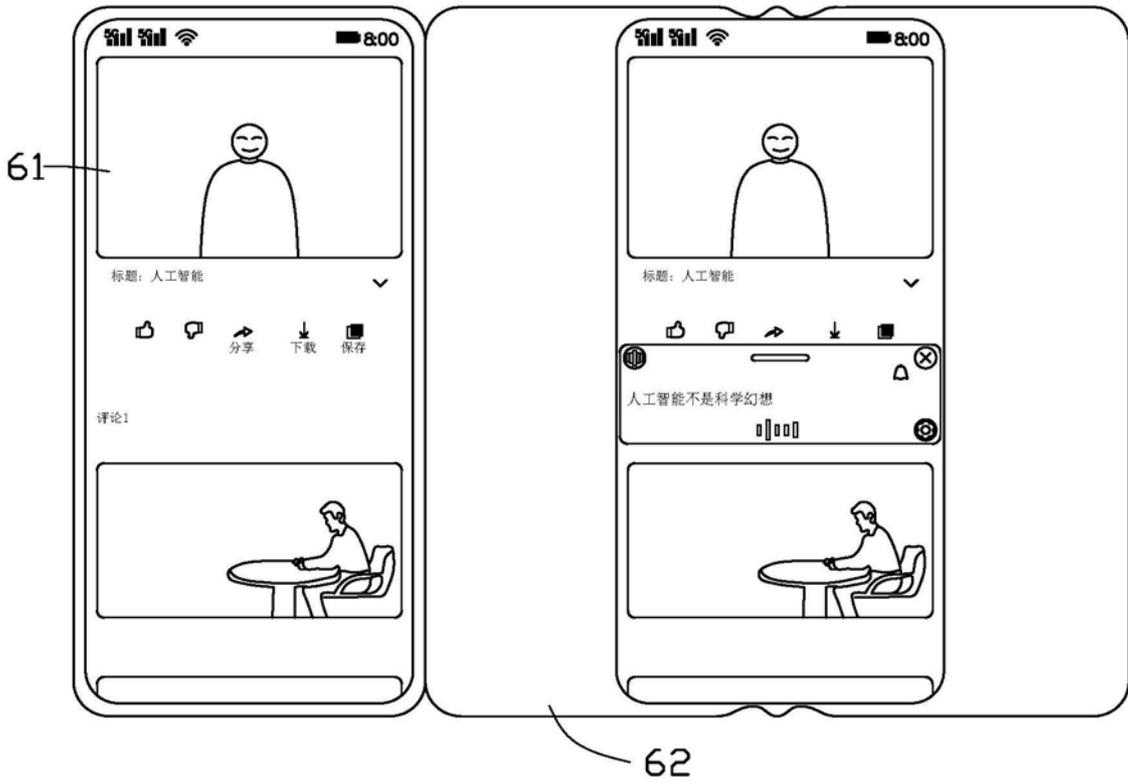


图6A

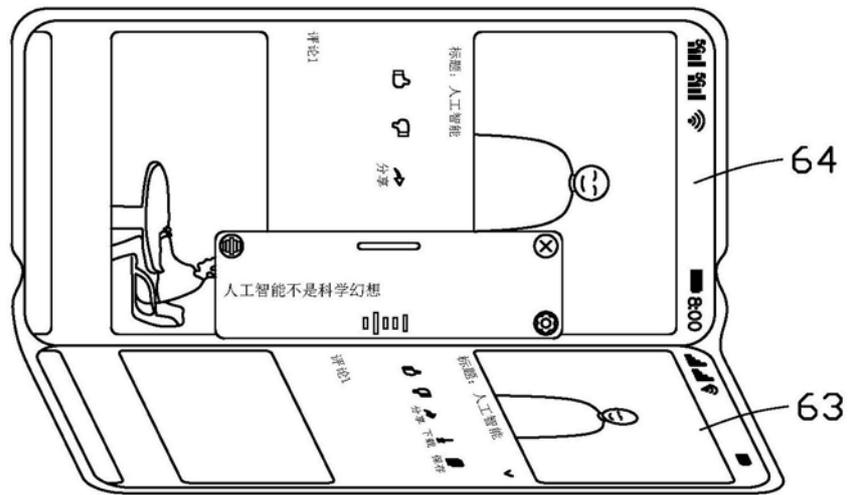


图6B

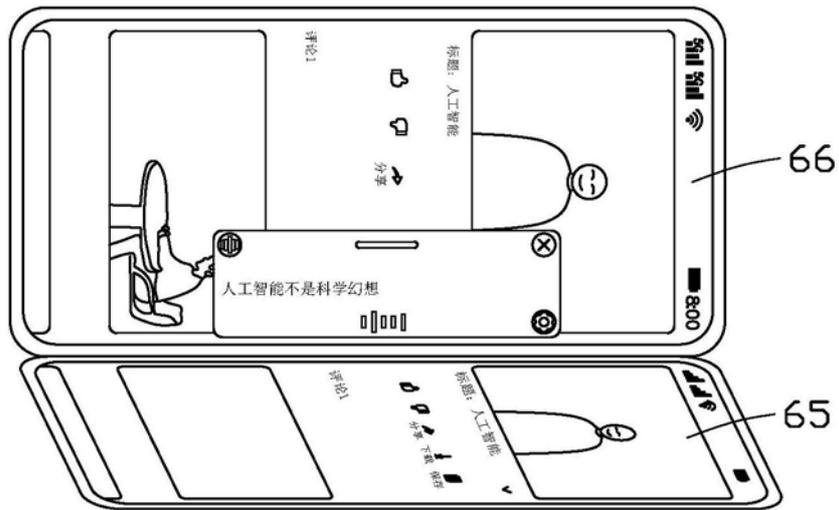


图6C

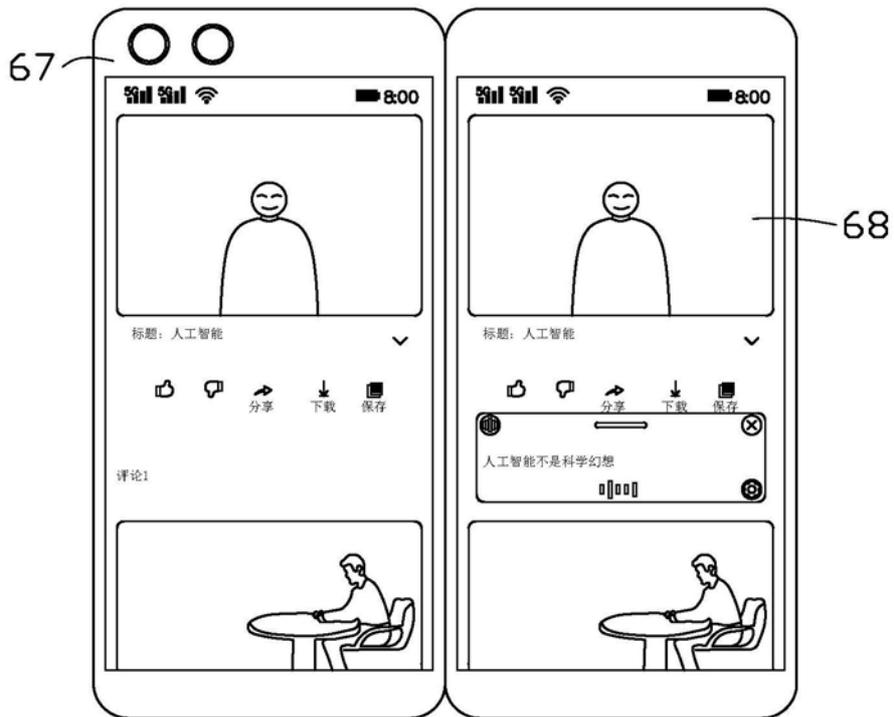


图6D

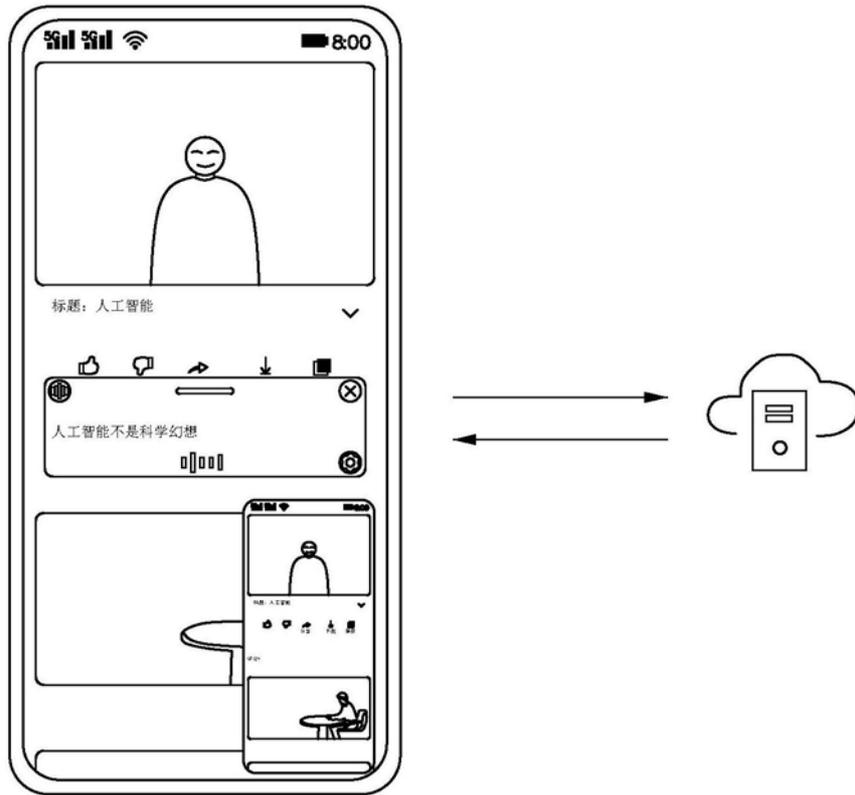


图7

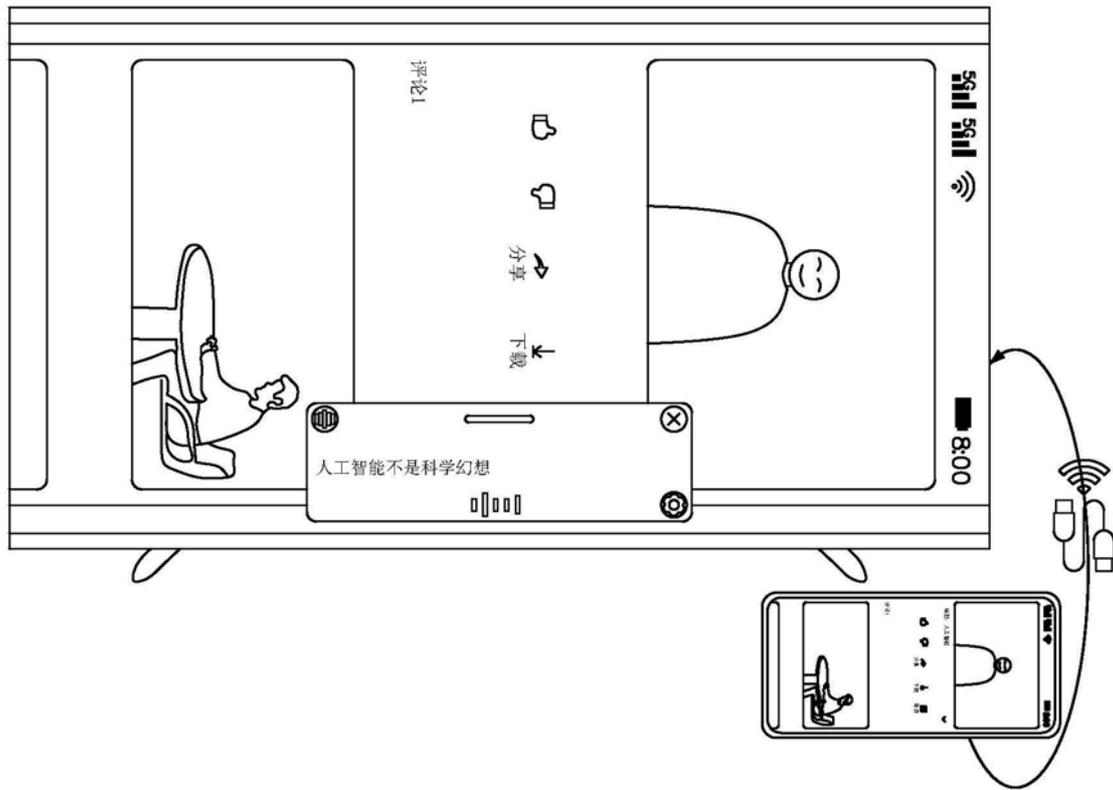


图8A

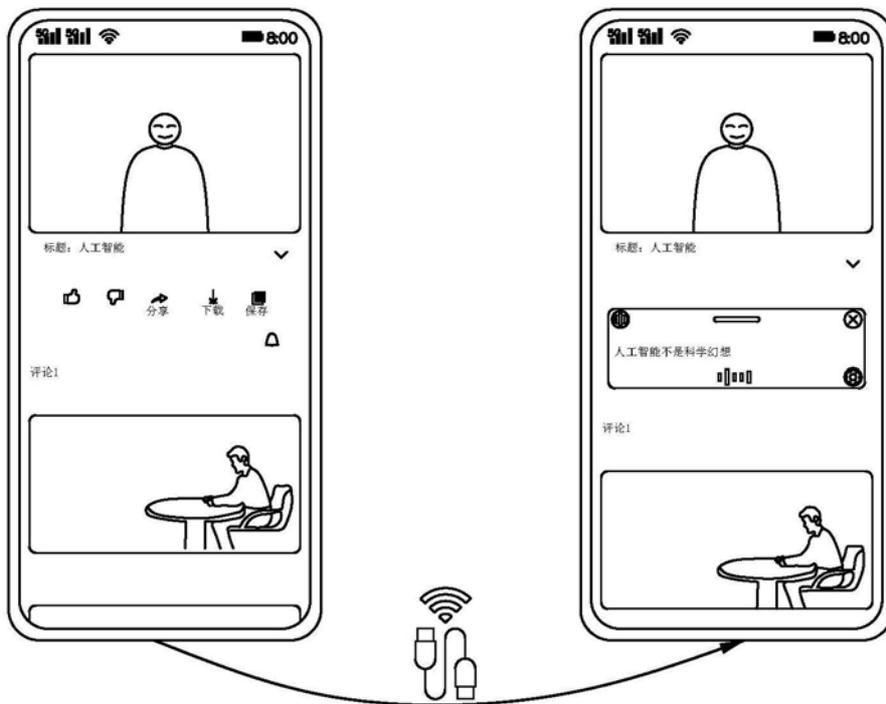


图8B

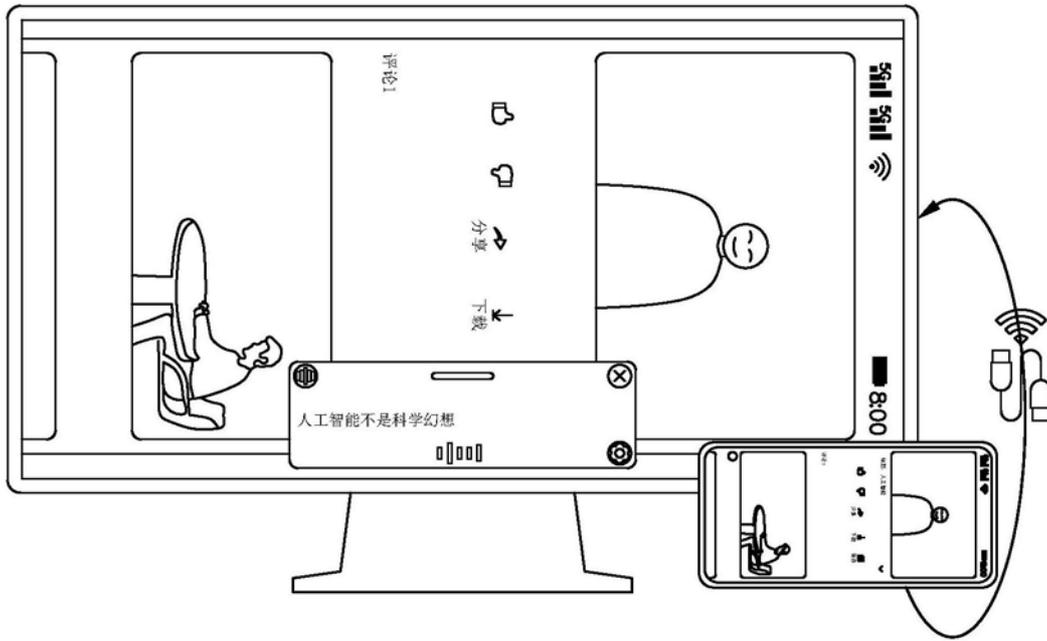


图8C

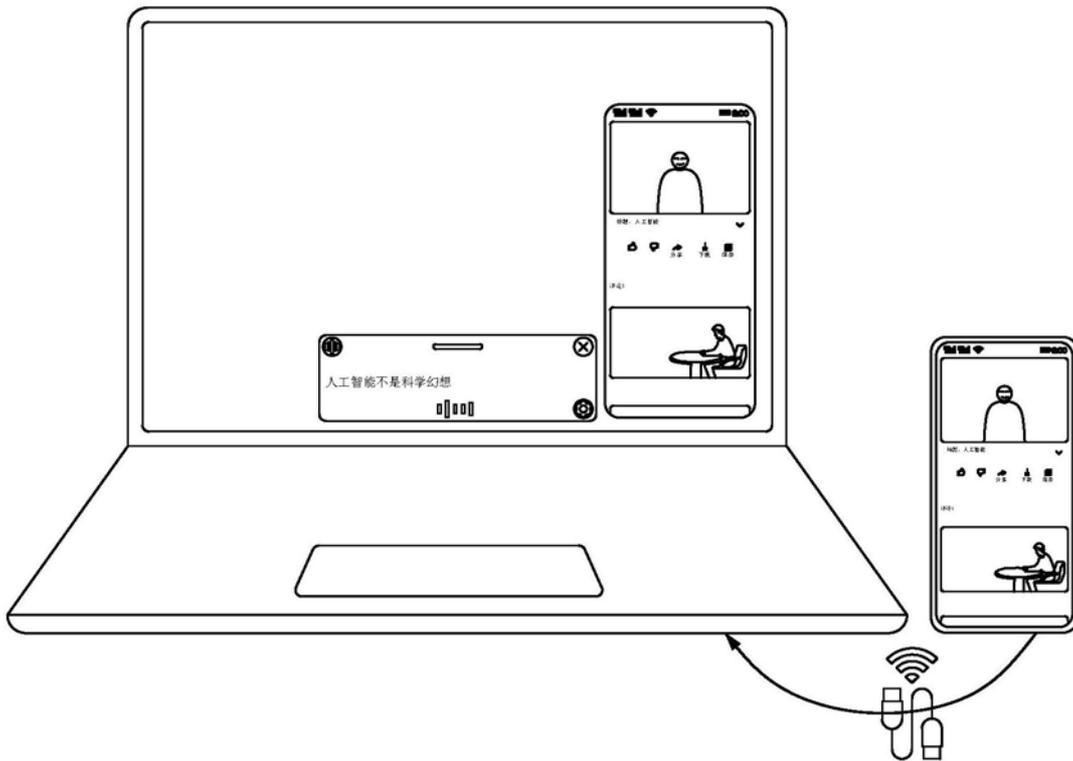


图8D

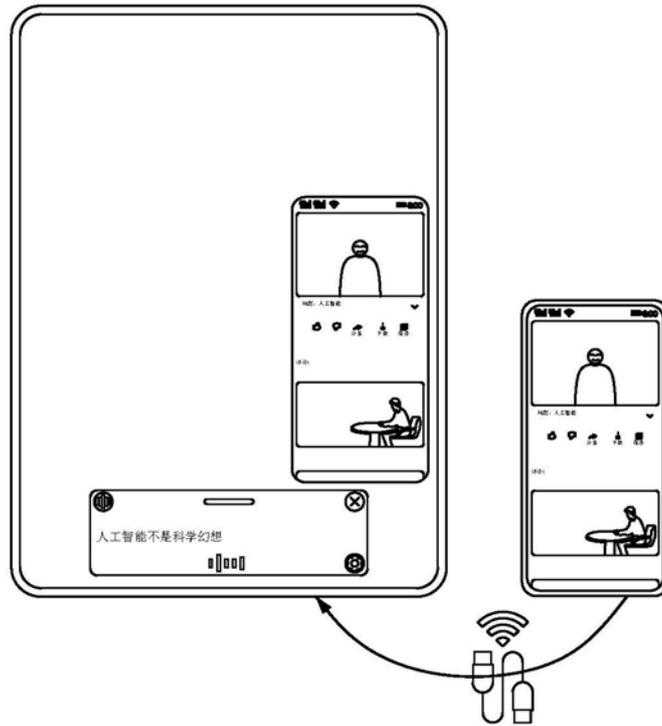


图8E

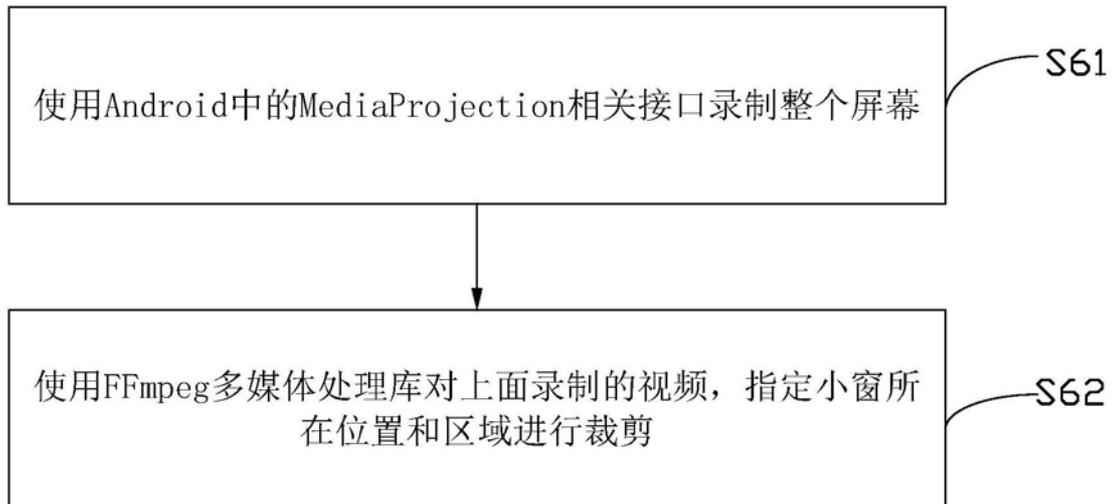


图9



图10A

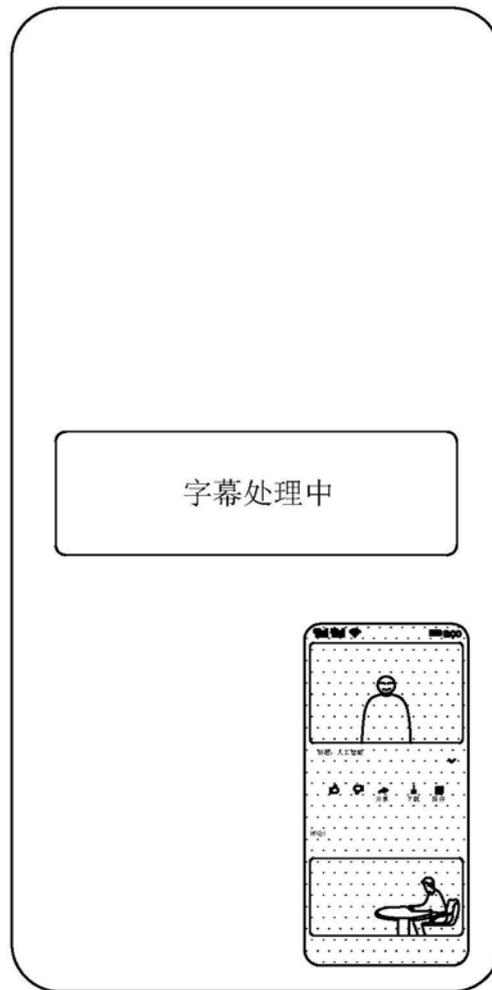


图10B

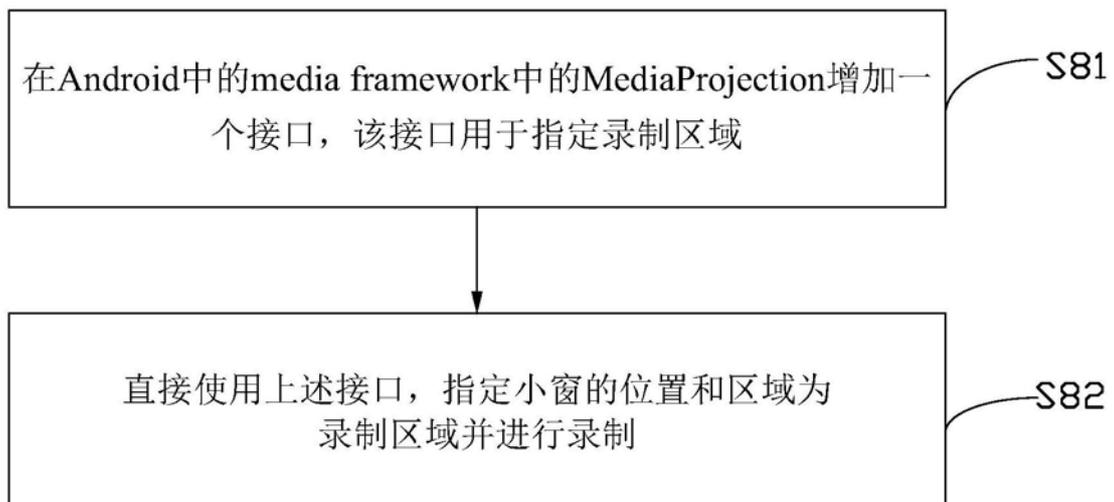


图11



图12

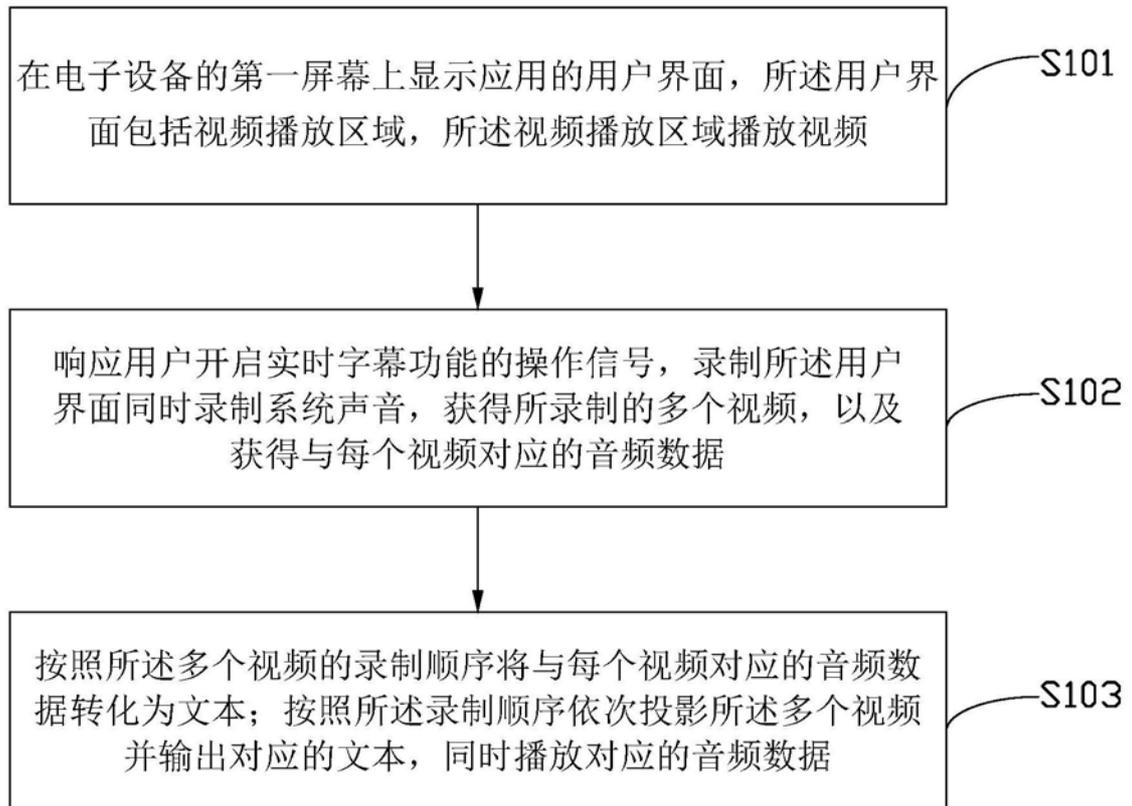


图13

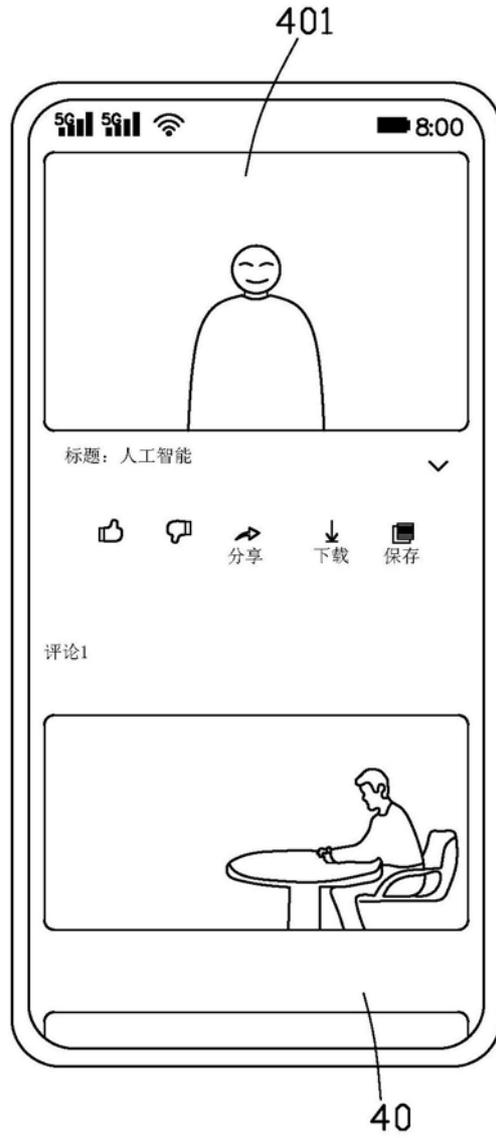


图14A

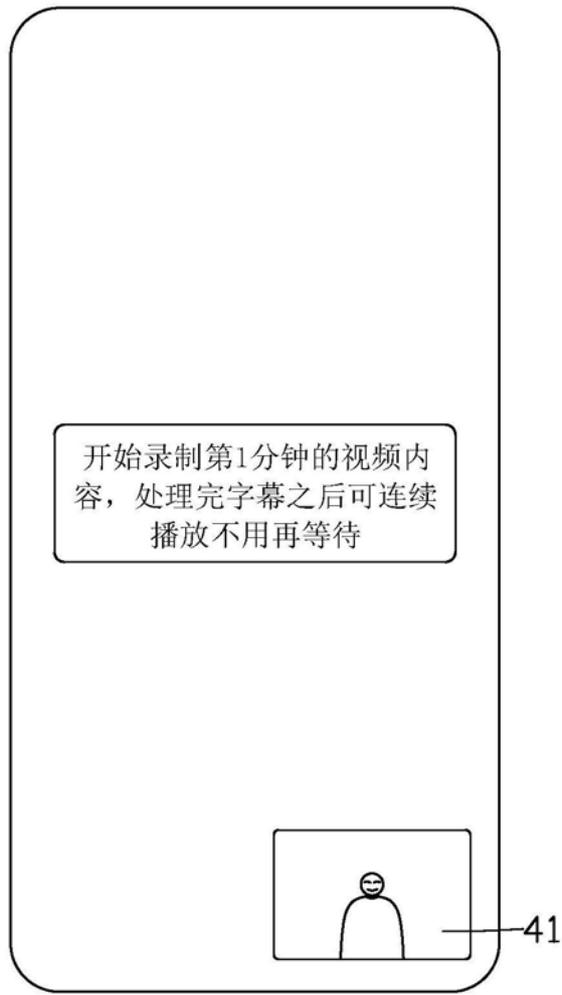


图14B

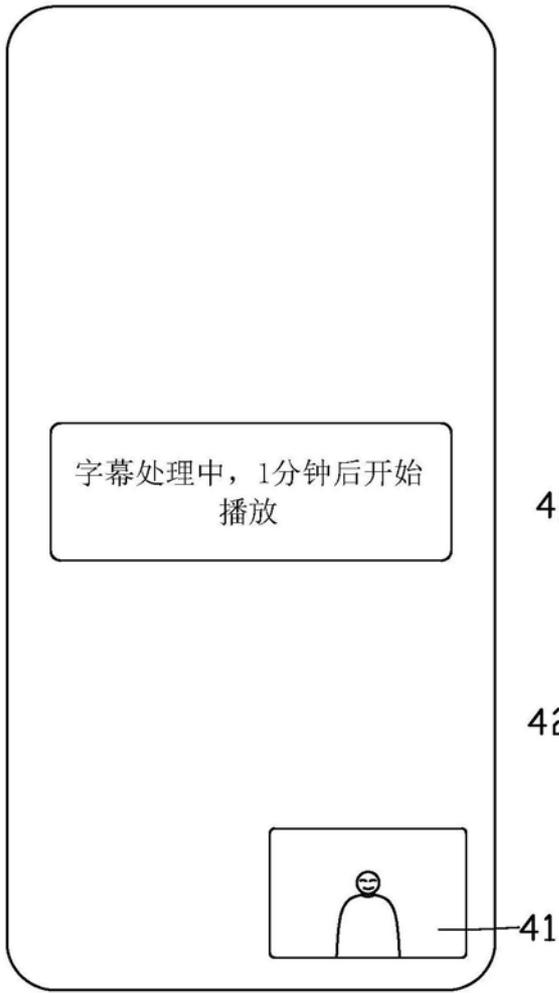


图 14C

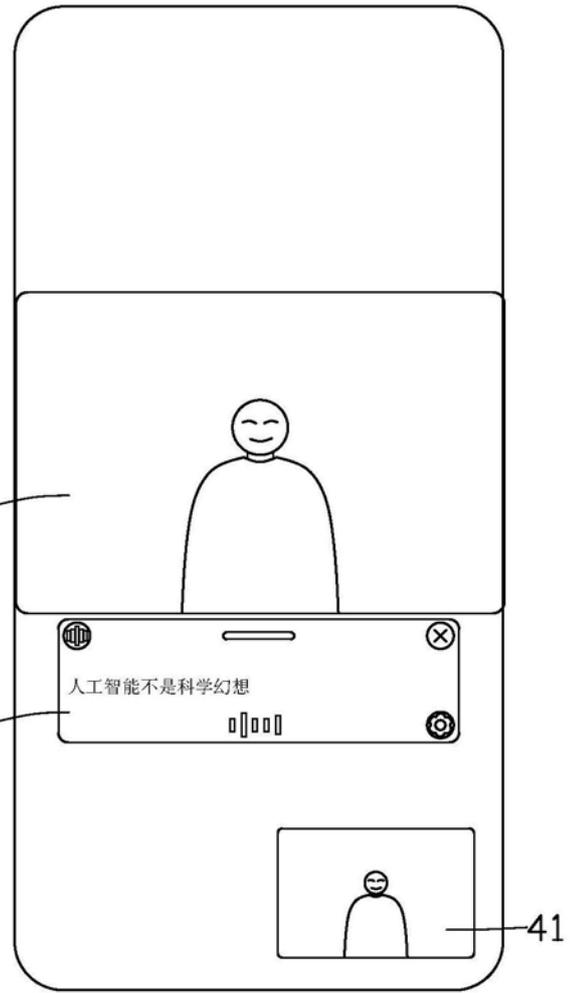


图 14D

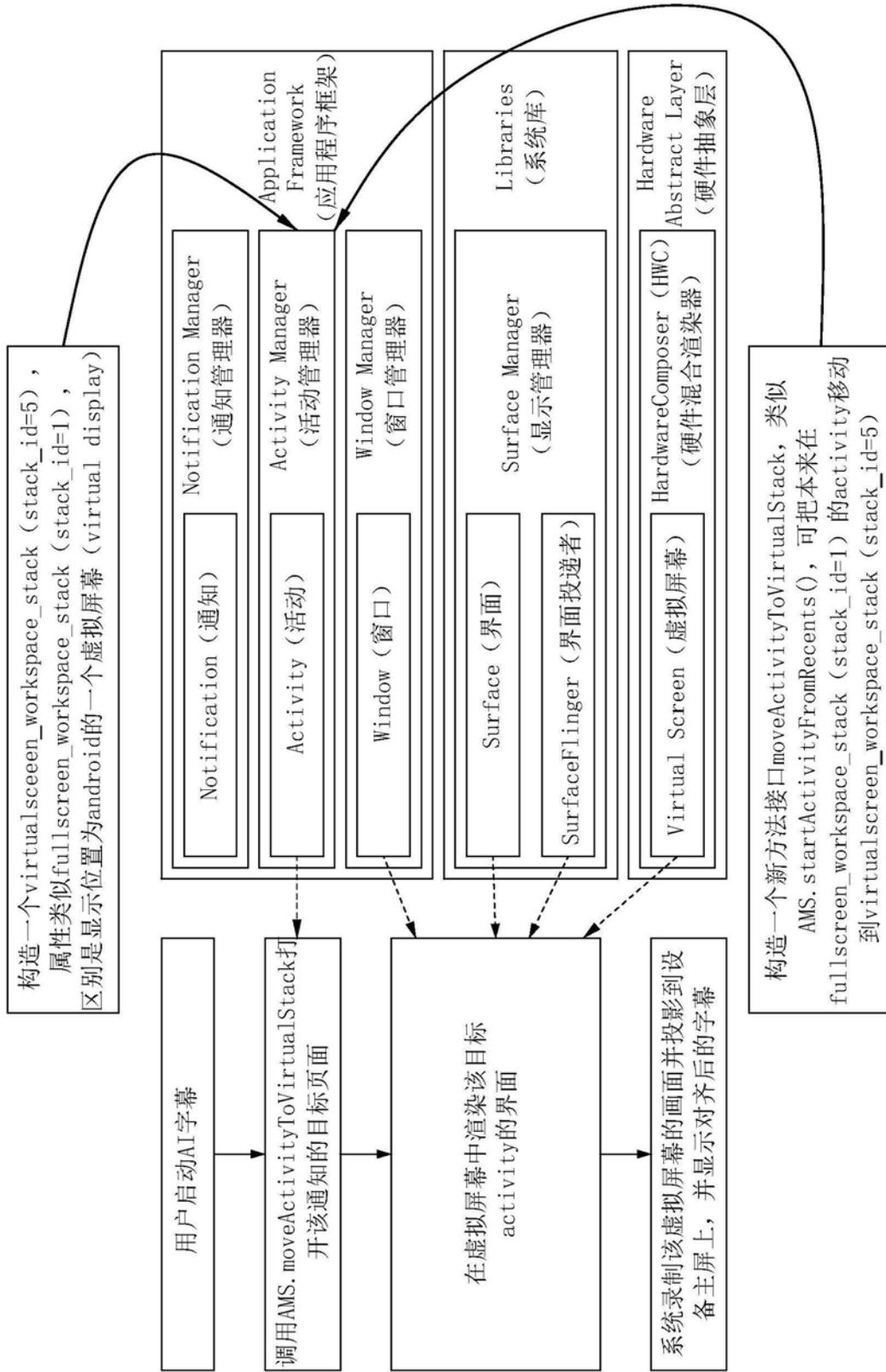


图15

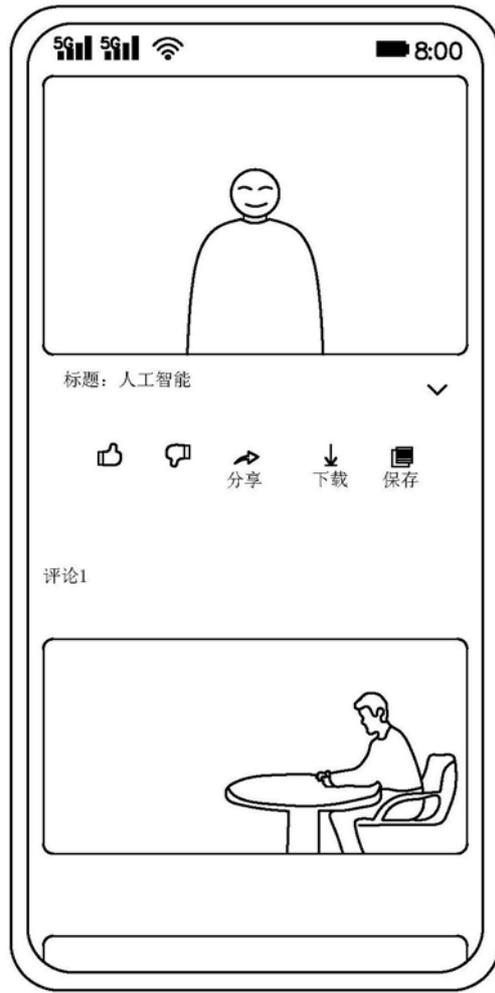


图16A



图16B

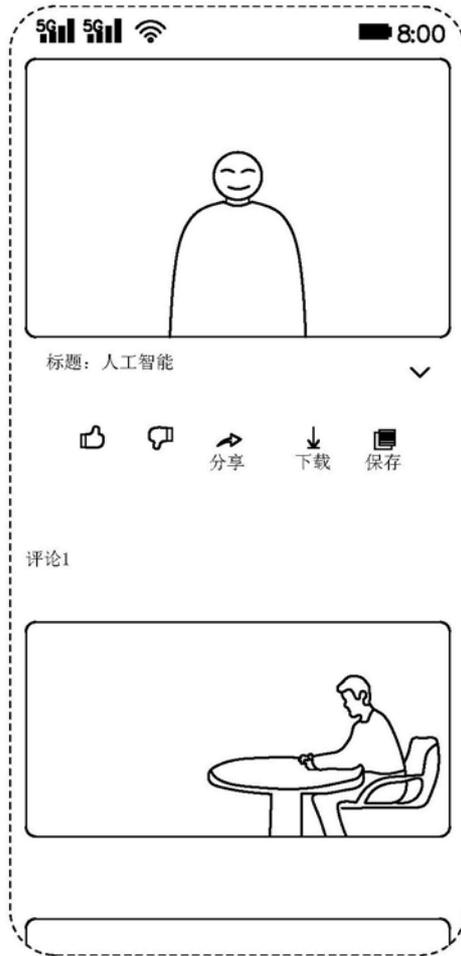


图16C

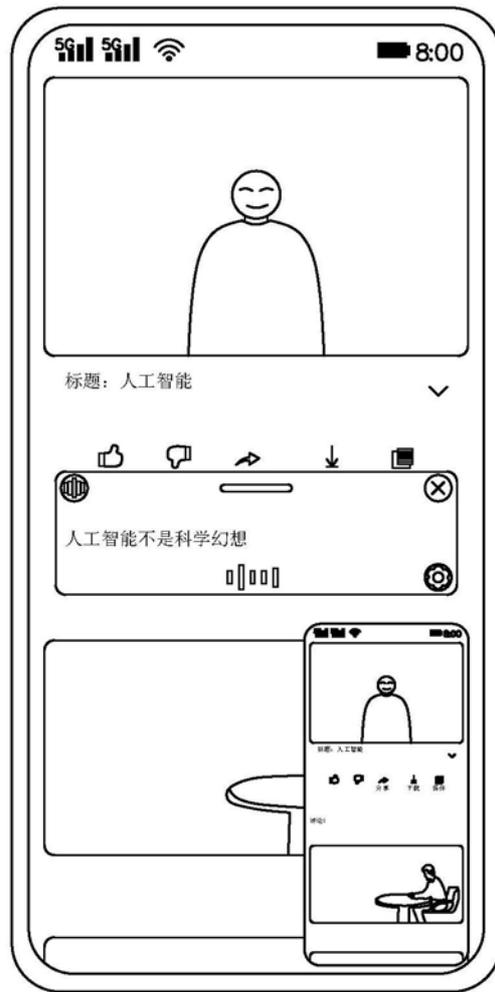


图16D

电子设备100

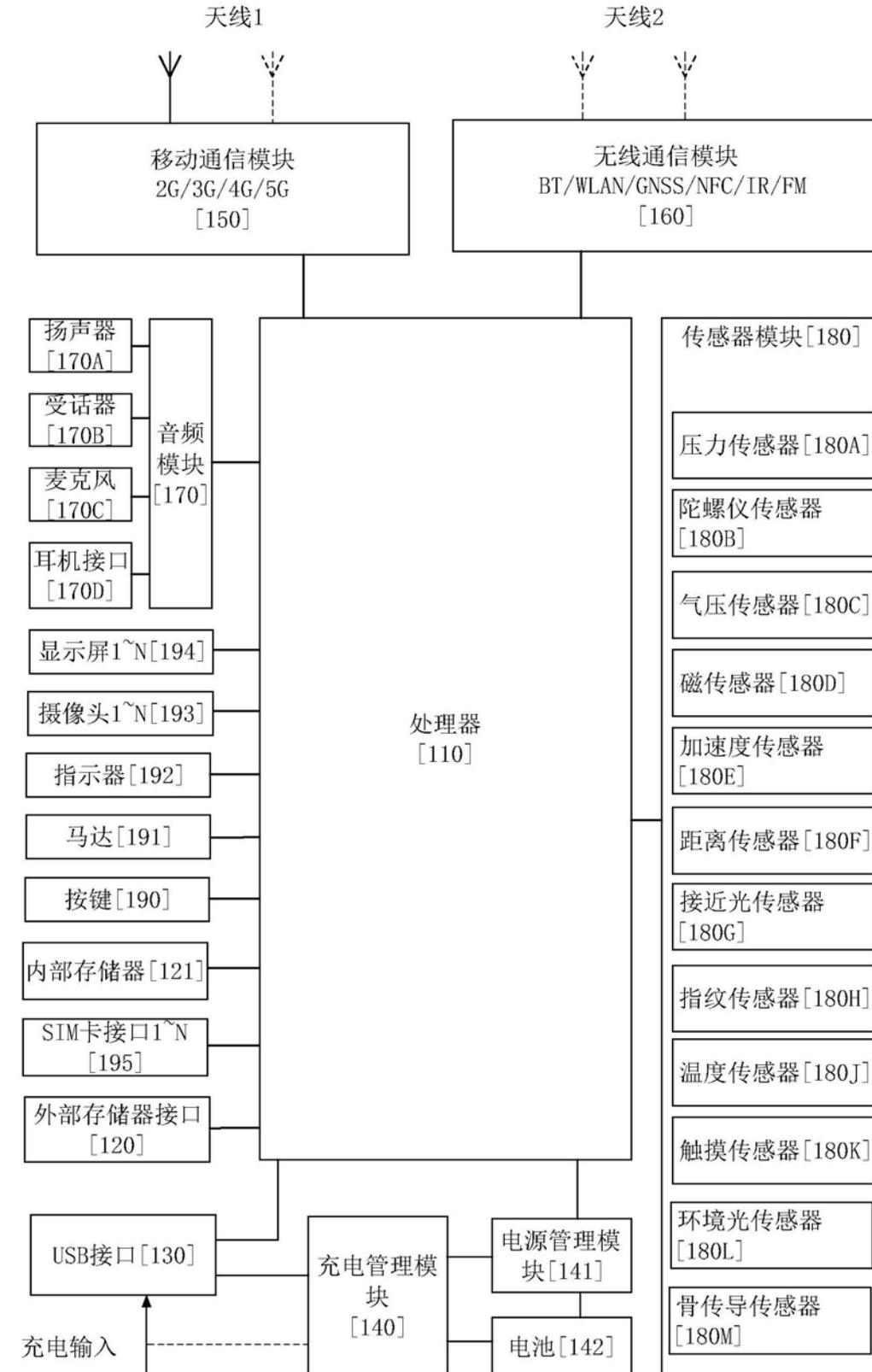


图17