

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 8 月 6 日 (06.08.2020)



(10) 国际公布号
WO 2020/155915 A1

- (51) 国际专利分类号:
G06F 3/16 (2006.01) **G06K 9/00** (2006.01)
G06F 3/01 (2006.01)
- (21) 国际申请号: PCT/CN2019/126772
- (22) 国际申请日: 2019 年 12 月 19 日 (19.12.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910086010.4 2019 年 1 月 29 日 (29.01.2019) CN
- (71) 申请人: 北京字节跳动网络技术有限公司 (**BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市石景山区实兴大街 30 号院 3 号楼 2 层 B-0035 房间, Beijing 100041 (CN)。
- (72) 发明人: 黄佳斌 (**HUANG, Jiabin**); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。
- (74) 代理人: 北京天达共和律师事务所等 (**EAST & CONCORD PARTNERS et al.**); 中国北京市朝阳区东三环北路 8 号亮马河大厦写字楼 2 座 19 层, Beijing 100004 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** METHOD AND APPARATUS FOR PLAYING BACK AUDIO

(54) 发明名称: 用于播放音频的方法和装置

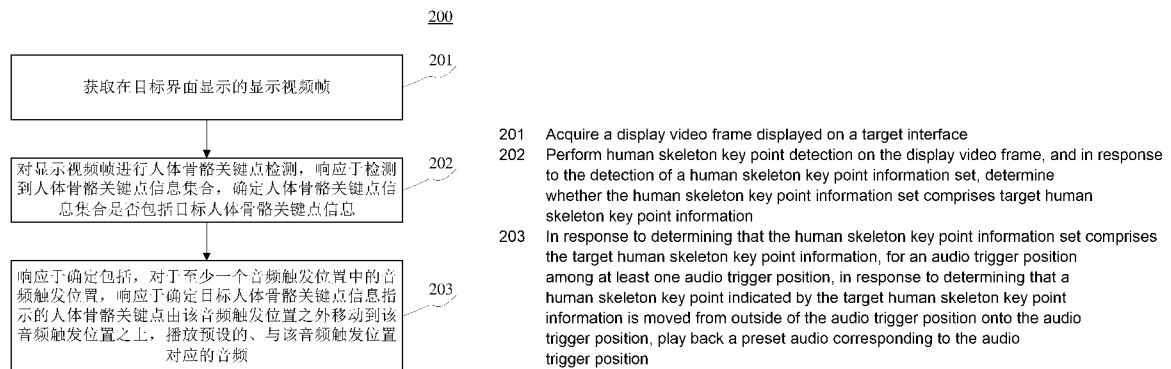


图 2

(57) **Abstract:** A method and an apparatus for playing back an audio. A specific embodiment of the method comprises: acquiring a display video frame displayed on a target interface (401); performing human skeleton key point detection on the display video frame, and in response to the detection of a human skeleton key point information set, determining whether the human skeleton key point information set comprises target human skeleton key point information (402); in response to determining that the human skeleton key point information set comprises the target human skeleton key point information, for an audio trigger position among at least one audio trigger position, in response to determining that a human skeleton key point indicated by the target human skeleton key point information is moved from outside of the audio trigger position onto the audio trigger position, playing back a preset audio corresponding to the audio trigger position (4031). The embodiment implements that a photographed person is able to trigger audio playback by means of a body movement, thereby improving the flexibility of triggering audio playback.

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(57) 摘要: 用于播放音频的方法和装置。该方法的一具体实施方式包括: 获取在目标界面显示的显示视频帧(401); 对显示视频帧进行人体骨骼关键点检测, 响应于检测到人体骨骼关键点信息集合, 确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息(402); 响应于确定包括, 对于至少一个音频触发位置中的音频触发位置, 响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上, 播放预设的、与该音频触发位置对应的音频(4031)。该实施方式实现了使被拍摄人员通过肢体动作即可触发播放音频, 提高了触发播放音频的灵活性。

用于播放音频的方法和装置

5 相关申请的交叉引用

本申请基于申请号为 201910086010.4、申请日为 2019 年 01 月 29 日的中国专利申请提出，并要求该中国专利申请的优先权，该中国专利申请的全部内容在此引入本申请作为参考。

10 技术领域

本公开的实施例涉及计算机技术领域，具体涉及用于播放音频的方法和装置。

背景技术

15 随着计算机技术的发展，人们可以使用手机、平板电脑等设备拍摄小视频、进行视频聊天等。当人们在拍摄演奏音乐的视频时，通常需要使用实际的乐器，或者通过触控电子设备上显示的虚拟乐器演奏音乐。

20 发明内容

本公开的实施例提出了用于播放音频的方法和装置。

第一方面，本公开的实施例提供了一种用于播放音频的方法，该方法包括：获取在目标界面显示的显示视频帧，其中，显示视频帧是当前拍摄的视频包括的视频帧，目标界面上预先设置有至少一个音频
25 触发位置；对显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息；响应于确定包括，对于至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放
30 预设的、与该音频触发位置对应的音频。

在一些实施例中，目标人体骨骼关键点信息为用于表征手部的人体骨骼关键点信息。

5 在一些实施例中，视频是对目标用户实时拍摄的视频；以及在对显示视频帧进行人体骨骼关键点检测之后，该方法还包括：响应于没有检测到人体骨骼关键点信息集合，在目标界面上显示用于提示目标用户站位错误的提示信息。

10 在一些实施例中，视频是对目标用户实时拍摄的视频；以及在确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息之后，该方法还包括：响应于确定人体骨骼关键点信息集合不包括目标人体骨骼关键点信息，在目标界面上显示用于提示目标用户站位错误的提示信息。

15 在一些实施例中，在响应于确定包括，对于至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频之前，该方法还包括：基于人体骨骼关键点信息集合确定人体图像；响应于确定人体图像的尺寸小于预设尺寸，对显示视频帧进行放大，以使人体图像的尺寸达到预设尺寸。

20 在一些实施例中，音频触发位置由以下至少一种表征：预设尺寸及形状的区域、预设长度的线。

25 在一些实施例中，在播放预设的、与该音频触发位置对应的音频之后，该方法还包括：响应于确定该音频触发位置由预设尺寸及形状的区域表征，且确定目标人体骨骼关键点信息指示的人体骨骼关键点在该音频触发位置上的停留时间达到预设时长，停止播放该音频触发位置对应的音频。

在一些实施例中，播放预设的、与该音频触发位置对应的音频，包括：确定目标人体骨骼关键点信息指示的人体骨骼关键点在目标界面上的移动速度；根据预设的与所确定的移动速度对应的音量，播放预设的、与该音频触发位置对应的音频。

30 第二方面，本公开的实施例提供了一种用于播放音频的装置，该

装置包括：获取单元，被配置成获取在目标界面显示的显示视频帧，其中，显示视频帧是当前拍摄的视频包括的视频帧，目标界面上预先设置有至少一个音频触发位置；检测单元，被配置成对显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息；播放单元，被配置成响应于确定包括，对于至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频。

10 在一些实施例中，目标人体骨骼关键点信息为用于表征手部的人体骨骼关键点信息。

在一些实施例中，视频是对目标用户实时拍摄的视频；以及检测单元进一步被配置成：响应于没有检测到人体骨骼关键点信息集合，在目标界面上显示用于提示目标用户站位错误的提示信息。

15 在一些实施例中，视频是对目标用户实时拍摄的视频；以及该装置还包括：显示单元，被配置成响应于确定人体骨骼关键点信息集合不包括目标人体骨骼关键点信息，在目标界面上显示用于提示目标用户站位错误的提示信息。

20 在一些实施例中，该装置还包括：确定单元，被配置成基于人体骨骼关键点信息集合确定人体图像；放大单元，被配置成响应于确定人体图像的尺寸小于预设尺寸，对显示视频帧进行放大，以使人体图像的尺寸达到预设尺寸。

在一些实施例中，音频触发位置由以下至少一种表征：预设尺寸及形状的区域、预设长度的线。

25 在一些实施例中，播放单元进一步被配置成：响应于确定该音频触发位置由预设尺寸及形状的区域表征，且确定目标人体骨骼关键点信息指示的人体骨骼关键点在该音频触发位置上的停留时间达到预设时长，停止播放该音频触发位置对应的音频。

30 在一些实施例中，播放单元包括：确定模块，被配置成确定目标人体骨骼关键点信息指示的人体骨骼关键点在目标界面上的移动速

度；播放模块，被配置成根据预设的与所确定的移动速度对应的音量，播放预设的、与该音频触发位置对应的音频。

第三方面，本公开的实施例提供了一种终端设备，该终端设备包括：一个或多个处理器；存储装置，其上存储有一个或多个程序；当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

第四方面，本公开的实施例提供了一种计算机可读介质，其上存储有计算机程序，该计算机程序被处理器执行时实现如第一方面中任一实现方式描述的方法。

本公开的实施例提供的用于播放音频的方法和装置，通过获取在目标界面显示的显示视频帧，对显示视频帧进行人体骨骼关键点检测，如果检测到人体骨骼关键点信息集合，且人体骨骼关键点信息集合包括目标人体骨骼关键点信息，对于目标界面上的至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点处于该音频触发位置，播放预设的、与该音频触发位置对应的音频，从而通过被拍摄人员的肢体动作即可触发播放音频，提高了触发播放音频的灵活性，有助于使得被拍摄人员在不使用乐器的情况下，仅通过肢体动作即可演奏音乐。

附图说明

通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本公开的其它特征、目的和优点将会变得更明显：

图 1 是本公开的一个实施例可以应用于其中的示例性系统架构图；

图 2 是根据本公开的实施例的用于播放音频的方法的一个实施例的流程图；

图 3 是根据本公开的实施例的用于播放音频的方法的一个应用场景的示意图；

图 4 是根据本公开的实施例的用于播放音频的方法的又一个实施例的流程图；

图 5 是根据本公开的实施例的用于播放音频的装置的一个实施例的结构示意图；

图 6 是适于用来实现本公开的实施例的终端设备的结构示意图。

5 具体实施方式

下面结合附图和实施例对本公开作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关公开，而非对该公开的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关公开相关的部分。

10 需要说明的是，在不冲突的情况下，本公开中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本公开。

图 1 示出了可以应用本公开的实施例的用于播放音频的方法或用于播放音频的装置的示例性系统架构 100。

15 如图 1 所示，系统架构 100 可以包括终端设备 101、102、103，网络 104 和服务器 105。网络 104 用以在终端设备 101、102、103 和服务器 105 之间提供通信链路的介质。网络 104 可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

20 用户可以使用终端设备 101、102、103 通过网络 104 与服务器 105 交互，以接收或发送消息等。终端设备 101、102、103 上可以安装有各种通讯客户端应用，例如视频拍摄应用、视频播放应用、社交平台软件等。

25 终端设备 101、102、103 可以是硬件，也可以是软件。当终端设备 101、102、103 为硬件时，可以是各种电子设备。当终端设备 101、102、103 为软件时，可以安装在上述电子设备中。其可以实现成多个软件或软件模块（例如用来提供分布式服务的软件或软件模块），也可以实现成单个软件或软件模块。在此不做具体限定。

30 服务器 105 可以是提供各种服务的服务器，例如对终端设备 101、102、103 拍摄的视频提供支持的后台服务器。后台服务器可以用于设置目标界面上的音频触发位置，以及设置于音频触发位置对应的音频

等。

需要说明的是，本公开的实施例所提供的用于显示图像的方法一般由终端设备 101、102、103 执行，相应地，用于显示图像的装置一般设置于终端设备 101、102、103 中。

5 需要说明的是，服务器可以是硬件，也可以是软件。当服务器为硬件时，可以实现成多个服务器组成的分布式服务器集群，也可以实现成单个服务器。当服务器为软件时，可以实现成多个软件或软件模块（例如用来提供分布式服务的软件或软件模块），也可以实现成单个软件或软件模块。在此不做具体限定。

10 应该理解，图 1 中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务。在显示视频帧不需要从远程获取的情况下，上述系统架构可以不包括服务器和网络。

15 继续参考图 2，示出了根据本公开的用于播放音频的方法的一个实施例的流程 200。该用于播放音频的方法，包括以下步骤：

步骤 201，获取在目标界面显示的显示视频帧。

在本实施例中，用于播放音频的方法的执行主体（例如图 1 所示的终端设备）可以从本地获取在目标界面显示的显示视频帧。其中，
20 显示视频帧是当前拍摄的视频包括的视频帧。目标界面可以是用于显示上述视频的视频帧的界面。例如，目标界面可以是在上述执行主体上安装的视频播放应用的播放界面。目标界面上预先设置有至少一个音频触发位置。音频触发位置用于触发播放音频。

在本实施例的一些可选的实现方式中，音频触发位置可以由以下
25 至少一种表征：预设尺寸及形状的区域、预设长度的线。作为示例，音频触发位置可以由预设尺寸的矩形区域表征。上述预设长度的线可以是直线段或曲线段。

步骤 202，对显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定人体骨骼关键点信息集合是否包括
30 目标人体骨骼关键点信息。

在本实施例中，上述执行主体可以对显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息。其中，人体骨骼关键点信息用于指示人体骨骼关键点。人体骨骼关键点是用于表征人体的特定部位的点，例如用于表征头顶、肘关节、肩关节等部位的点。人体骨骼关键点信息可以包括在显示视频帧上建立的坐标系中的坐标，该坐标可以用于表征人体骨骼关键点在显示视频帧中的位置。

上述执行主体可以按照现有的各种确定人体骨骼关键点的方法对显示视频帧进行人体骨骼关键点检测。例如，上述执行主体可以将显示视频帧输入预先训练的用于进行人体骨骼关键点检测的卷积神经网络（Convolutional Neural Networks, CNN），得到人体骨骼关键点信息集合。上述卷积神经网络可以是现有的各种结构的卷积神经网络，例如 R-CNN（Region-CNN）、STN（Spatial Transform Networks, 空间变换网络）等。需要说明的是，上述人体骨骼关键点检测方法是目前广泛研究和应用的公知技术，在此不再赘述。

上述目标人体骨骼关键点信息可以是检测得到的人体骨骼关键点信息集合中的、用于表征人体的特定部位（例如手部、脚部等）的人体骨骼关键点信息。通常，人体骨骼关键点信息可以具有对应的序号，序号可以由上述执行主体在检测到人体骨骼关键点信息集合时，根据每个人体骨骼关键点信息指示的人体骨骼关键点对应的人体部位确定的。上述执行主体可以按照预设的、目标人体骨骼关键点信息对应的序号，从人体骨骼关键点信息集合中，确定目标人体骨骼关键点信息。

在本实施例的一些可选的实现方式中，目标人体骨骼关键点信息为用于表征手部的人体骨骼关键点信息。其中，目标人体骨骼关键点信息可以是表征人的任一只手的人体骨骼关键点信息。通过设置表征手部的人体骨骼关键点信息作为目标人体骨骼关键点信息，可以有助于使得被拍摄的人物通过手部动作灵活地控制音频的播放。

在本实施例的一些可选的实现方式中，上述视频是对目标用户实时拍摄的视频。其中，目标用户可以是上述执行主体包括的摄像头或

与上述执行主体通信连接的电子设备包括的摄像头拍摄到的用户。上述执行主体可以响应于确定未检测到人体骨骼关键点信息集合，在目标界面上显示用于提示用户站位错误的提示信息。具体地，上述执行主体未检测到人体骨骼关键点信息集合的原因，通常是由于目标用户的站位不准确，上述执行主体无法获得目标用户的人体图像。此时，上述执行主体可以在目标界面上显示用于提示用户站位错误的提示信息。提示信息可以包括但不限于以下至少一种：文字、图像等。作为示例，提示信息可以是用于表征人体的轮廓的图像，目标用户可以参考该图像，调整站位，使得自己的人体图像位于该图像所处的位置。实践中，当人体图像处于该图像所处的位置时，通常为触发播放音频的最佳位置。

在本实施例的一些可选的实现方式中，上述执行主体可以在步骤203之前，执行如下步骤：

首先，基于人体骨骼关键点信息集合确定人体图像。具体地，作为示例，上述执行主体可以从显示视频帧中，将包括所有的人体骨骼关键点信息分别指示的人体骨骼关键点的矩形区域（例如最小矩形包括的区域，或者在最小矩形的基础上放大预设倍数所得到的矩形包括的区域）确定为人体图像。或者，人体骨骼关键点信息可以具有对应的序号，上述执行主体可以按照预先指定的序号，将包括这些序号分别对应的人体骨骼关键点的矩形区域确定为人体图像。通常，按照上述指定的序号确定的人体图像可以为人体的上半身图像。

然后，响应于确定人体图像的尺寸小于预设尺寸，对显示视频帧进行放大，以使人体图像的尺寸达到预设尺寸。

其中，图像的尺寸通常用像素数表征，例如 $x \times y$ ，其中， x 为横向像素数， y 为纵向像素数。上述预设尺寸可以是预设的固定尺寸；或者，上述预设尺寸可以是根据预设的比例确定的尺寸，例如，假设用于显示上述显示视频帧的界面的尺寸为 $m \times n$ ，预设的比例为 0.8，则预设尺寸为 $0.8m \times 0.8n$ 。需要说明的是，当满足以下至少一项条件时，可以确定人体图像的尺寸小于预设尺寸：人体图像的横向像素数小于预设尺寸的横向像素数，人体图像的纵向像素数小于预设尺寸的

纵向像素数，人体图像的对角线包括的像素数小于预设尺寸表征的矩形的对角线包括的像素数。相应地，当满足以下至少一项条件时，确定人体图像达到预设尺寸：人体图像的横向像素数与预设尺寸的横向像素数相同，人体图像的纵向像素数与预设尺寸的纵向像素数相同，

5 人体图像的对角线包括的像素数与预设尺寸表征的矩形的对角线包括的像素数相同。应当理解，上述条件仅仅是示例性的，实践中，还可以包括其他条件。

通过执行本可选的实现方式，可以在被拍摄的用户的人体图像较小时，将人体图像放大至预设的尺寸，从而有助于使用户更准确地通过

10 肢体动作触发音频播放。

步骤 203，响应于确定包括，对于至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频。

15 在本实施例中，上述执行主体可以响应于确定人体骨骼关键点信息集合包括目标人体骨骼关键点信息，对于上述至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频。其中，与音频触发位置对

20 应的音频可以预先存储在上述执行主体中，音频触发位置与音频的对应关系可以通过列表、指针等形式预先建立。

具体地，作为示例，对于某个音频触发位置，如果该音频触发位置由预设尺寸及形状的区域表征，当检测到目标人体骨骼关键点信息指示的人体骨骼关键点由该区域外移动到该区域内时，确定目标人体

25 骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，此时，播放预设的、与该音频触发位置对应的音频。再例如，对于某个音频触发位置，如果该音频触发位置由预设长度的直线段表征，当检测到目标人体骨骼关键点信息指示的人体骨骼关键点与该直线段接触时，确定目标人体骨骼关键点信息指示的人

30 体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，此

时，播放预设的、与该音频触发位置对应的音频。

实践中，音频触发位置对应的音频为模拟某种乐器的某个音调，通过触发播放各个音频触发位置分别对应的音频，可以实现通过人的肢体动作模拟演奏乐器。此外，音频触发位置对应的音频还可以是其他类型的音频，例如一段音乐，一种音效等。

在本实施例的一些可选的实现方式中，视频是对目标用户实时拍摄的视频。其中，目标用户可以是上述执行主体包括的摄像头或与上述执行主体通信连接的电子设备包括的摄像头拍摄到的用户。上述执行主体可以响应于确定人体骨骼关键点信息集合不包括目标人体骨骼关键点信息，在目标界面上显示用于提示目标用户站位错误的提示信息。具体地，当确定人体骨骼关键点信息集合不包括目标人体骨骼关键点信息时，表示目标人物站位不准确，摄像头无法拍摄到完整的人体图像，上述执行主体无法检测到目标人体骨骼关键点信息。此时，上述执行主体可以在目标界面上显示提示信息，以提醒目标用户站位错误。本实现方式中的提示信息可以与上述可选的实现方式中的提示信息相同，这里不再赘述。

继续参见图 3，图 3 是根据本实施例的用于播放音频的方法的应用场景的一个示意图。在图 3 的应用场景中，终端设备 301 首先获取当前在目标界面 302（即终端设备上用于显示拍摄的视频包括的视频帧的页面）显示的显示视频帧，其中，显示视频帧是对使用终端设备自拍的用户拍摄的视频包括的视频帧。目标界面 302 上预先设置有 7 个音频触发位置（即图中的矩形区域 A-G），7 个音频触发位置用于模拟钢琴的琴键。然后，终端设备 301 对显示视频帧进行人体骨骼关键点检测，得到人体骨骼关键点信息集合，其中包括的人体骨骼关键点信息分别对应于图中黑色原点所示的人体骨骼关键点。其中，人体骨骼关键点 303 和 304 分别对应的人体骨骼关键点信息均为目标人体骨骼关键点信息，用于表征人手。最后，终端设备 301 响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点 303 由音频触发位置 G 之外移动到 G 之上，播放预设的、与该音频触发位置对应的音频。

本公开的上述实施例提供的方法，通过获取在目标界面显示的显

示视频帧，对显示视频帧进行人体骨骼关键点检测，如果检测到人体骨骼关键点信息集合，且人体骨骼关键点信息集合包括目标人体骨骼关键点信息，对于目标界面上的至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点处于该音频触发位置，播放预设的、与该音频触发位置对应的音频，从而实现了使被拍摄人员通过肢体动作即可触发播放音频，提高了触发播放音频的灵活性，有助于使得被拍摄人员在不使用乐器的情况下，仅通过肢体动作即可演奏音乐。

进一步参考图 4，其示出了用于播放音频的方法的又一个实施例的流程 400。该用于播放音频的方法的流程 400，包括以下步骤：

步骤 401，获取在目标界面显示的显示视频帧。

在本实施例中，用于播放音频的方法的执行主体（例如如图 1 所示的终端设备）可以从本地获取在目标界面显示的显示视频帧。其中，显示视频帧是当前拍摄的视频包括的视频帧。目标界面可以是用于显示上述视频的视频帧的界面。例如，目标界面可以是在上述执行主体上安装的视频播放应用的播放界面。目标界面上预先设置有至少一个音频触发位置。音频触发位置用于触发播放音频。

在本实施例中，音频触发位置可以由以下至少一种表征：预设尺寸及形状的区域、预设长度的线。作为示例，音频触发位置可以由预设尺寸的矩形区域表征。上述预设长度的线可以是直线段或曲线段。

步骤 402，对显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息。

在本实施例中，步骤 402 与图 2 对应实施例中的步骤 202 基本一致，这里不再赘述。

步骤 403，响应于确定包括，对于至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频；响应于确定该音频触发位置由预设尺

寸及形状的区域表征，且确定目标人体骨骼关键点信息指示的人体骨骼关键点在该音频触发位置上的停留时间达到预设时长，停止播放音频。

在本实施例中，上述执行主体可以响应于确定人体骨骼关键点信息集合包括目标人体骨骼关键点信息，对于上述至少一个音频触发位置中的音频触发位置，执行如下子步骤（包括步骤 4031-步骤 4032）：

步骤 4031，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频。

具体地，步骤 4031 与图 2 对应实施例中的步骤 203 基本一致，这里不再赘述。

步骤 4032，响应于确定该音频触发位置由预设尺寸及形状的区域表征，且确定目标人体骨骼关键点信息指示的人体骨骼关键点在该音频触发位置上的停留时间达到预设时长，停止播放音频。

具体地，当该音频触发位置由预设尺寸及形状的区域（例如预设尺寸的矩形区域）表征时，上述执行主体可以确定目标人体骨骼关键点信息指示的人体骨骼关键点在该触发位置上的停留时间。通常，上述执行主体可以在播放该音频触发位置对应的音频的同时，开始计时，并实时地对目标界面上显示的视频帧进行人体骨骼关键点检测。当检测到目标人体骨骼关键点信息指示的人体骨骼关键点在该音频触发位置上的停留时间达到预设时长（例如 3 秒）时，停止播放该音频触发位置对应的音频。实践中，通常可以以音量逐渐减小的方式停止播放音频。

在本实施例的一些可选的实现方式中，上述执行主体可以按照如下步骤播放与该音频触发位置对应的音频：

首先，确定目标人体骨骼关键点信息指示的人体骨骼关键点在目标界面上的移动速度。具体地，上述执行主体可以实时地对目标界面显示的视频帧进行人体骨骼关键点检测，通过对目标人体骨骼关键点信息指示的人体骨骼关键点在相邻的两个视频帧（或者中间间隔预设数量个视频帧的两个视频帧）的所处的位置的变化，以及两个视频

帧的播放时间间隔，可以实时地确定出目标人体骨骼关键点信息指示的人体骨骼关键点在目标界面上的移动速度。当检测到目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上时，确定此时的目标人体骨骼关键点信息指示的人体骨骼关键点移动速度为下述确定音频的音量使用的移动速度。

然后，根据预设的与所确定的移动速度对应的音量，播放预设的、与该音频触发位置对应的音频。本实现方式可以根据目标人体骨骼关键点信息指示的人体骨骼关键点的移动速度，控制播放的音频的音量，从而有助于更准确地模拟乐器的演奏。例如，上述音频触发位置用于模拟钢琴时，可以由目标人体骨骼关键点信息指示的人体骨骼关键点的移动速度表征人的手指敲击琴键的力度，从而更真实的模拟钢琴演奏。

从图 4 中可以看出，与图 2 对应的实施例相比，本实施例中的用于播放音频的方法的流程 400 突出了根据目标人体骨骼关键点信息指示的人体骨骼关键点在音频触发位置的停留时间，停止播放音频的步骤。由此，本实施例描述的方案可以更加灵活地控制音频的播放，有助于更加准确地模拟乐器的演奏。

进一步参考图 5，作为对上述各图所示方法的实现，本公开提供了一种用于播放音频的装置的一个实施例，该装置实施例与图 2 所示的方法实施例相对应，该装置具体可以应用于各种电子设备中。

如图 5 所示，本实施例的用于播放音频的装置 500 包括：获取单元 501，被配置成获取在目标界面显示的显示视频帧，其中，显示视频帧是当前拍摄的视频包括的视频帧，目标界面上预先设置有至少一个音频触发位置；检测单元 502，被配置成对显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息；播放单元 503，被配置成响应于确定包括，对于至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该

音频触发位置对应的音频。

在本实施例中，获取单元 501 可以获取在目标界面显示的显示视频帧。其中，显示视频帧是当前拍摄的视频包括的视频帧。目标界面可以是用于显示上述视频的视频帧的界面。例如，目标界面可以是在
5 上述装置 500 上安装的视频播放应用的播放界面。目标界面上预先设置有至少一个音频触发位置。音频触发位置用于触发播放音频。

在本实施例中，检测单元 502 可以对显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息。其中，人体骨骼关键点信息用于指示人体骨骼关键点。人体骨骼关键点是用于表征人体的特定部位的点，例如用于表征头顶、肘关节、肩关节等部位的点。
10 人体骨骼关键点信息可以包括在显示视频帧上建立的坐标系中的坐标，该坐标可以用于表征人体骨骼关键点在显示视频帧中的位置。

上述检测单元 502 可以按照现有的各种确定人体骨骼关键点的方法对显示视频帧进行人体骨骼关键点检测。例如，上述检测单元 502
15 可以将显示视频帧输入预先训练的卷积神经网络（Convolutional Neural Networks, CNN），得到人体骨骼关键点信息集合。上述卷积神经网络可以是现有的各种结构的卷积神经网络，例如 R-CNN（Region-CNN）、STN (Spatial Transform Networks, 空间变换网络)等。
20 需要说明的是，上述人体骨骼关键点检测方法是目前广泛研究和应用的公知技术，在此不再赘述。

上述目标人体骨骼关键点信息可以是所检测得到的人体骨骼关键点信息集合中的、用于表征人体的特定部位（例如手部、脚部等）的人体骨骼关键点信息。通常，人体骨骼关键点信息可以具有对应的序号，序号可以由上述检测单元 502 在检测到人体骨骼关键点信息集
25 合时，根据每个人体骨骼关键点信息指示的人体骨骼关键点对应的人体部位确定的。上述检测单元 502 可以按照预设的、目标人体骨骼关键点信息对应的序号，从人体骨骼关键点信息集合中，确定目标人体骨骼关键点信息。

30 在本实施例中，播放单元 503 可以响应于确定人体骨骼关键点信

息集合包括目标人体骨骼关键点信息，对于上述至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频。

5 具体地，作为示例，对于某个音频触发位置，如果该音频触发位置由预设尺寸及形状的区域表征，当检测到目标人体骨骼关键点信息指示的人体骨骼关键点由该区域外移动到该区域内时，确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，此时，播放预设的、与该音频触发位置对应的
10 音频。再例如，对于某个音频触发位置，如果该音频触发位置由预设长度的直线段表征，当检测到目标人体骨骼关键点信息指示的人体骨骼关键点与该直线段接触时，确定目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，此时，播放预设的、与该音频触发位置对应的音频。

15 实践中，音频触发位置对应的音频为模拟某种乐器的某个音调，通过触发播放各个音频触发位置分别对应的音频，可以实现通过人的肢体动作模拟演奏乐器。此外，音频触发位置对应的音频还可以是其他类型的音频，例如一段音乐，一种音效等。

在本实施例的一些可选的实现方式中，目标人体骨骼关键点信息
20 为用于表征手部的人体骨骼关键点信息。

在本实施例的一些可选的实现方式中，视频是对目标用户实时拍摄的视频；以及检测单元 502 可以进一步被配置成：响应于没有检测到人体骨骼关键点信息集合，在目标界面上显示用于提示目标用户站位错误的提示信息。

25 在本实施例的一些可选的实现方式中，视频是对目标用户实时拍摄的视频；以及该装置 500 还可以包括：显示单元（图中未示出），被配置成响应于确定人体骨骼关键点信息集合不包括目标人体骨骼关键点信息，在目标界面上显示用于提示目标用户站位错误的提示信息。

在本实施例的一些可选的实现方式中，该装置 500 还可以包括：
30 确定单元（图中未示出），被配置成基于人体骨骼关键点信息集合确定

人体图像；放大单元（图中未示出），被配置成响应于确定人体图像的尺寸小于预设尺寸，对显示视频帧进行放大，以使人体图像的尺寸达到预设尺寸。

在本实施例的一些可选的实现方式中，音频触发位置由以下至少一种表征：预设尺寸及形状的区域、预设长度的线。

在本实施例的一些可选的实现方式中，播放单元 503 可以进一步被配置成：响应于确定该音频触发位置由预设尺寸及形状的区域表征，且确定目标人体骨骼关键点信息指示的人体骨骼关键点在该音频触发位置上的停留时间达到预设时长，停止播放该音频触发位置对应的音频。

在本实施例的一些可选的实现方式中，播放单元 503 可以包括：确定模块（图中未示出），被配置成确定目标人体骨骼关键点信息指示的人体骨骼关键点在目标界面上的移动速度；播放模块（图中未示出），被配置成根据预设的与所确定的移动速度对应的音量，播放预设的、与该音频触发位置对应的音频。

本公开的上述实施例提供的装置，通过获取在目标界面显示的显示视频帧，对显示视频帧进行人体骨骼关键点检测，如果检测到人体骨骼关键点信息集合，且人体骨骼关键点信息集合包括目标人体骨骼关键点信息，对于目标界面上的至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的人体骨骼关键点处于该音频触发位置，播放预设的、与该音频触发位置对应的音频，从而实现了使被拍摄人员通过肢体动作即可触发播放音频，提高了触发播放音频的灵活性，有助于使得被拍摄人员在不使用乐器的情况下，仅通过肢体动作即可演奏音乐。

下面参考图 6，其示出了适于用来实现本公开的实施例的终端设备 600 的结构示意图。本公开的实施例中的终端设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、PDA（个人数字助理）、PAD（平板电脑）、PMP（便携式多媒体播放器）、车载终端（例如车载导航终端）等等的移动终端以及诸如数字 TV、台式计算机等等

的固定终端。图 6 示出的终端设备仅仅是一个示例，不应对本公开的实施例的功能和使用范围带来任何限制。

如图 6 所示，终端设备 600 可以包括处理装置（例如中央处理器、图形处理器等）601，其可以根据存储在只读存储器（ROM）602 中的程序或者从存储装置 608 加载到随机访问存储器（RAM）603 中的程序而执行各种适当的动作和处理。在 RAM 603 中，还存储有终端设备 600 操作所需的各种程序和数据。处理装置 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出（I/O）接口 605 也连接至总线 604。

通常，以下装置可以连接至 I/O 接口 605：包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置 606；包括例如液晶显示器（LCD）、扬声器、振动器等的输出装置 607；包括例如内存等的存储装置 608；以及通信装置 609。通信装置 609 可以允许终端设备 600 与其他设备进行无线或有线通信以交换数据。虽然图 6 示出了具有各种装置的终端设备 600，但是应理解的是，并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。图 6 中示出的每个方框可以代表一个装置，也可以根据需要代表多个装置。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信装置 609 从网络上被下载和安装，或者从存储装置 608 被安装，或者从 ROM 602 被安装。在该计算机程序被处理装置 601 执行时，执行本公开的实施例的方法中限定的上述功能。需要说明的是，本公开的实施例所述的计算机可读介质可以是计算机可读信号介质或者计算机可读介质或者是上述两者的任意组合。计算机可读介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、

便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。

5 在本公开的实施例中，计算机可读介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开的实施例中，计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介
10 质还可以是计算机可读介质以外的任何计算机可读介质，该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：电线、光缆、RF（射
15 频）等等，或者上述的任意合适的组合。

 上述计算机可读介质可以是上述终端设备中所包含的；也可以是单独存在，而未装配入该终端设备中。上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该终端设备执行时，使得该终端设备：获取在目标界面显示的显示视频帧，其中，显示视频
20 帧是当前拍摄的视频包括的视频帧，目标界面上预先设置有至少一个音频触发位置；对显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息；响应于确定包括，对于至少一个音频触发位置中的音频触发位置，响应于确定目标人体骨骼关键点信息指示的
25 人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频。

 可以以一种或多种程序设计语言或其组合来编写用于执行本公开的实施例的操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如 Java、Smalltalk、C++，还包括常规的过程式
30 程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可

以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本公开的实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器包括获取单元、检测单元、和播放单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，获取单元还可以被描述为“获取在目标界面显示的显示视频帧的单元”。

以上描述仅为本公开的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本公开的实施例中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开的实施例中公开的（但不限于）具有类似功能的技术特征进行互相替换而形成的技术方案。

权 利 要 求 书

1、一种用于播放音频的方法，包括：

5 获取在目标界面显示的显示视频帧，其中，所述显示视频帧是当
前拍摄的视频包括的视频帧，所述目标界面上预先设置有至少一个音频
触发位置；

对所述显示视频帧进行人体骨骼关键点检测，响应于检测到人体
骨骼关键点信息集合，确定所述人体骨骼关键点信息集合是否包括目
标人体骨骼关键点信息；

10 响应于确定包括，对于所述至少一个音频触发位置中的音频触发
位置，响应于确定所述目标人体骨骼关键点信息指示的人体骨骼关键
点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、
与该音频触发位置对应的音频。

15 2、根据权利要求 1 所述的方法，其中，所述目标人体骨骼关键点
信息为用于表征手部的人体骨骼关键点信息。

3、根据权利要求 1 所述的方法，其中，所述视频是对目标用户实
时拍摄的视频；以及

20 在所述对所述显示视频帧进行人体骨骼关键点检测之后，所述方
法还包括：

响应于没有检测到人体骨骼关键点信息集合，在所述目标界面上
显示用于提示所述目标用户站位错误的提示信息。

25 4、根据权利要求 1 所述的方法，其中，所述视频是对目标用户实
时拍摄的视频；以及

在所述确定所述人体骨骼关键点信息集合是否包括目标人体骨骼
关键点信息之后，所述方法还包括：

30 响应于确定所述人体骨骼关键点信息集合不包括目标人体骨骼关
键点信息，在所述目标界面上显示用于提示所述目标用户站位错误的

提示信息。

5 5、根据权利要求1所述的方法，其中，在所述响应于确定包括，对于所述至少一个音频触发位置中的音频触发位置，响应于确定所述目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频之前，所述方法还包括：

基于所述人体骨骼关键点信息集合确定人体图像；

10 响应于确定所述人体图像的尺寸小于预设尺寸，对所述显示视频帧进行放大，以使所述人体图像的尺寸达到所述预设尺寸。

6、根据权利要求1-5之一所述的方法，其中，音频触发位置由以下至少一种表征：预设尺寸及形状的区域、预设长度的线。

15 7、根据权利要求6所述的方法，其中，在所述播放预设的、与该音频触发位置对应的音频之后，所述方法还包括：

响应于确定该音频触发位置由预设尺寸及形状的区域表征，且确定所述目标人体骨骼关键点信息指示的人体骨骼关键点在该音频触发位置上的停留时间达到预设时长，停止播放该音频触发位置对应的音频。

20

8、根据权利要求1-5之一所述的方法，其中，所述播放预设的、与该音频触发位置对应的音频，包括：

25 确定所述目标人体骨骼关键点信息指示的人体骨骼关键点在所述目标界面上的移动速度；

根据预设的与所确定的移动速度对应的音量，播放预设的、与该音频触发位置对应的音频。

9、一种用于播放音频的装置，包括：

30 获取单元，被配置成获取在目标界面显示的显示视频帧，其中，

所述显示视频帧是当前拍摄的视频包括的视频帧，所述目标界面上预先设置有至少一个音频触发位置；

检测单元，被配置成对所述显示视频帧进行人体骨骼关键点检测，响应于检测到人体骨骼关键点信息集合，确定所述人体骨骼关键点信息集合是否包括目标人体骨骼关键点信息；

播放单元，被配置成响应于确定包括，对于所述至少一个音频触发位置中的音频触发位置，响应于确定所述目标人体骨骼关键点信息指示的人体骨骼关键点由该音频触发位置之外移动到该音频触发位置之上，播放预设的、与该音频触发位置对应的音频。

10

10、根据权利要求 9 所述的装置，其中，所述目标人体骨骼关键点信息为用于表征手部的人体骨骼关键点信息。

11、根据权利要求 9 所述的装置，其中，所述视频是对目标用户实时拍摄的视频；以及

所述检测单元进一步被配置成：

响应于没有检测到人体骨骼关键点信息集合，在所述目标界面上显示用于提示所述目标用户站位错误的提示信息。

12、根据权利要求 9 所述的装置，其中，所述视频是对目标用户实时拍摄的视频；以及

所述装置还包括：

显示单元，被配置成响应于确定所述人体骨骼关键点信息集合不包括目标人体骨骼关键点信息，在所述目标界面上显示用于提示所述目标用户站位错误的提示信息。

13、根据权利要求 9 所述的装置，其中，所述装置还包括：

确定单元，被配置成基于所述人体骨骼关键点信息集合确定人体图像；

放大单元，被配置成响应于确定所述人体图像的尺寸小于预设尺

寸，对所述显示视频帧进行放大，以使所述人体图像的尺寸达到所述预设尺寸。

14、根据权利要求 9-13 之一所述的装置，其中，音频触发位置由
5 以下至少一种表征：预设尺寸及形状的区域、预设长度的线。

15、根据权利要求 14 所述的装置，其中，所述播放单元进一步被配置成：

响应于确定该音频触发位置由预设尺寸及形状的区域表征，且确定
10 所述目标人体骨骼关键点信息指示的人体骨骼关键点在该音频触发位置上的停留时间达到预设时长，停止播放该音频触发位置对应的音频。

16、根据权利要求 9-13 之一所述的装置，其中，所述播放单元包
15 括：

确定模块，被配置成确定所述目标人体骨骼关键点信息指示的人体骨骼关键点在所述目标界面上的移动速度；

播放模块，被配置成根据预设的与所确定的移动速度对应的音量，播放预设的、与该音频触发位置对应的音频。

20

17、一种终端设备，包括：

一个或多个处理器；

存储装置，其上存储有一个或多个程序，显示屏？

当所述一个或多个程序被所述一个或多个处理器执行，使得所述
25 一个或多个处理器实现如权利要求 1-8 中任一所述的方法。

18、一种计算机可读介质，其上存储有计算机程序，其中，该程序被处理器执行时实现如权利要求 1-8 中任一所述的方法。

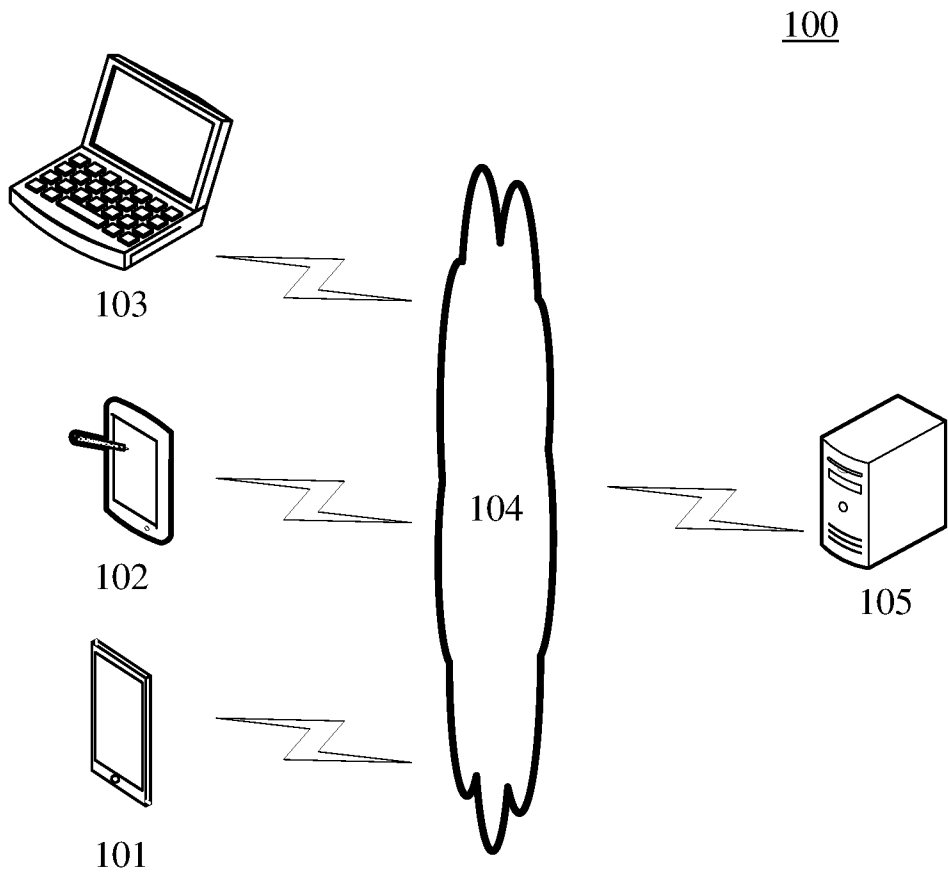


图 1

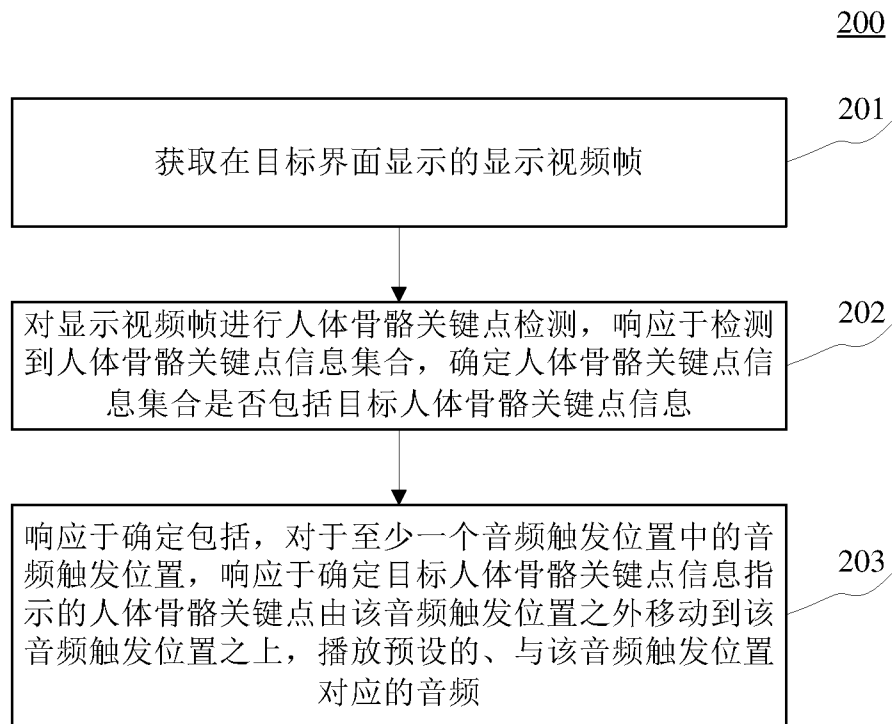


图 2

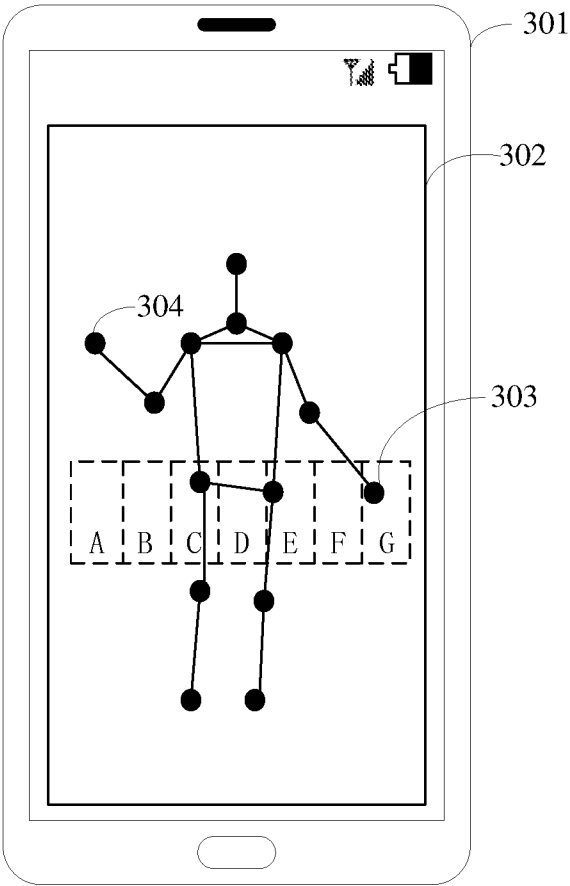


图 3

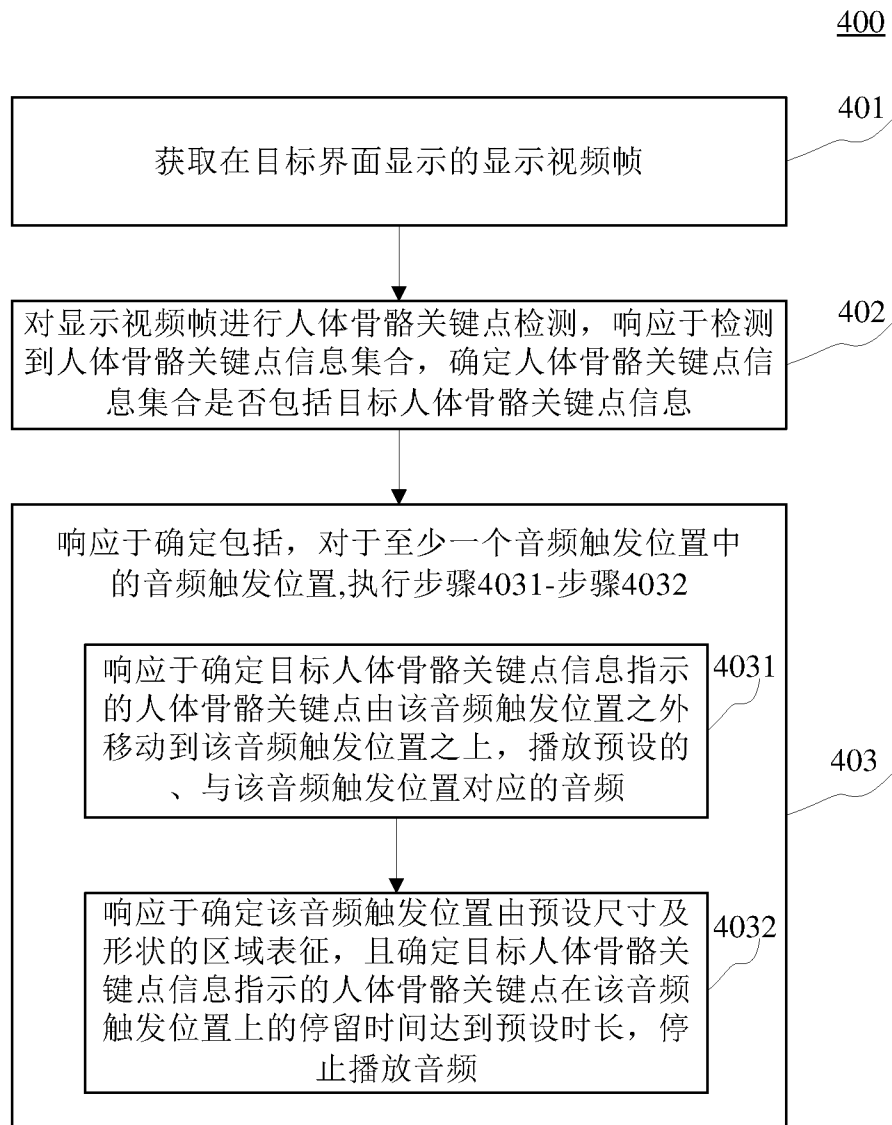


图 4

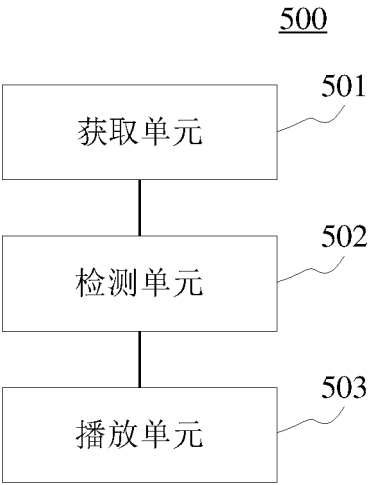


图 5

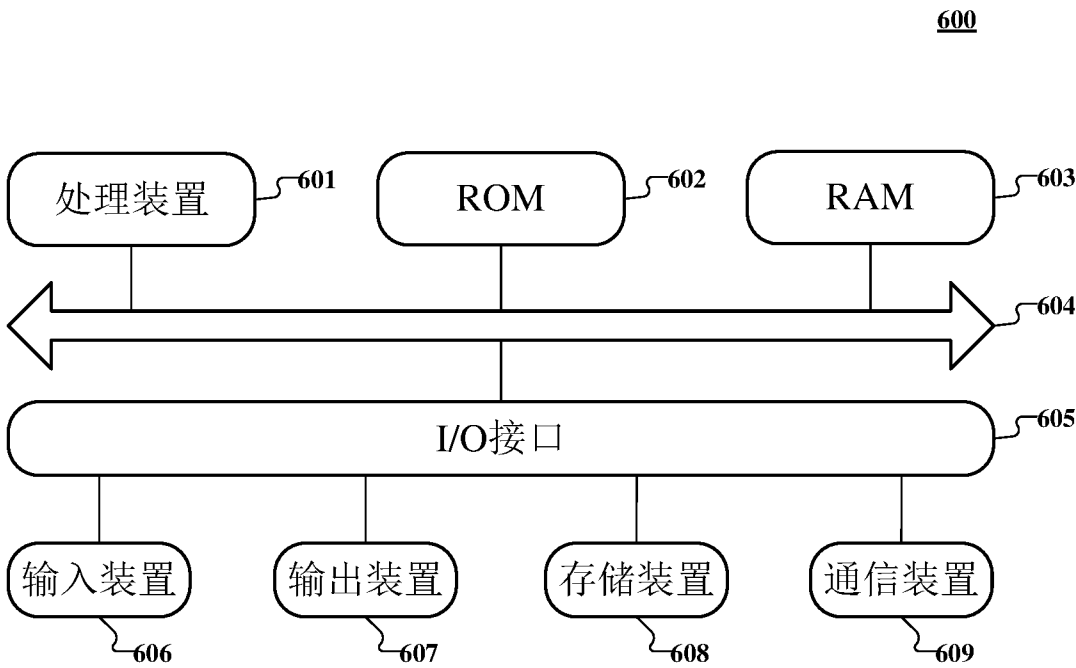


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/126772

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/16(2006.01)i; G06F 3/01(2006.01)i; G06K 9/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F,G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC: 播放, 音频, 音乐, 视频, 手指, 手势, 乐器, 虚拟, 人体, 触发, 站位, 按键, PLAY, VIDEO, AUDIO, MUSIC+, FINGER, GESTURE, BODY, KEY, VIRTUAL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109828741 A (BEIJING ZIJIE TIAODONG NETWORK TECHNOLOGY CO., LTD.) 31 May 2019 (2019-05-31) claims 1-18	1-18
X	CN 109166565 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 08 January 2019 (2019-01-08) description, paragraphs [0053]-[0128], and figures 2-6	1, 2, 5-10, 13-18
Y	CN 109166565 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 08 January 2019 (2019-01-08) description, paragraphs [0053]-[0128], and figures 2-6	3-4, 11-12
Y	CN 106934406 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 07 July 2017 (2017-07-07) description, paragraphs [0034]-[0071], and figures 1-4	3-4, 11-12
X	US 2012007884 A1 (SAMSUNG ELECTRONICS CO., LTD.) 12 January 2012 (2012-01-12) description, paragraphs [0027]-[0074], and figures 1-4	1, 2, 5-10, 13-18
A	CN 107302548 A (CHINA TELECOM CORPORATION LIMITED) 27 October 2017 (2017-10-27) entire document	1-18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

04 March 2020

Date of mailing of the international search report

18 March 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/126772**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2008010024 A1 (CHERRADI, I.) 24 January 2008 (2008-01-24) entire document	1-18
A	CN 108829253 A (UNIVERSITY OF SCIENCE AND TECHNOLOGY BEIJING) 16 November 2018 (2018-11-16) entire document	1-18
A	CN 105389013 A (SHENZHEN-HONG KONG PRODUCTION AND REAESEARCH BASE) 09 March 2016 (2016-03-09) entire document	1-18

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/126772

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109828741	A	31 May 2019	None			
CN	109166565	A	08 January 2019	None			
CN	106934406	A	07 July 2017	None			
US	2012007884	A1	12 January 2012	KR	20120004023	A	12 January 2012
				US	8866846	B2	21 October 2014
				KR	101679239	B1	24 November 2016
CN	107302548	A	27 October 2017	None			
WO	2008010024	A1	24 January 2008	US	2009322673	A1	31 December 2009
				EP	2041640	B1	25 January 2012
				US	9477310	B2	25 October 2016
				EP	2041640	A1	01 April 2009
				AT	543132	T	15 February 2012
CN	108829253	A	16 November 2018	None			
CN	105389013	A	09 March 2016	None			

国际检索报告

国际申请号

PCT/CN2019/126772

A. 主题的分类

G06F 3/16(2006.01)i; G06F 3/01(2006.01)i; G06K 9/00(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F, G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNKI, CNPAT, WPI, EPDOC: 播放, 音频, 音乐, 视频, 手指, 手势, 乐器, 虚拟, 人体, 触发, 站位, 按键, PLAY, VIDEO, AUDIO, MUSIC+, FINGER, GESTURE, BODY, KEY, VIRTUAL

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 109828741 A (北京字节跳动网络技术有限公司) 2019年 5月 31日 (2019 - 05 - 31) 权利要求1-18	1-18
X	CN 109166565 A (百度在线网络技术北京有限公司) 2019年 1月 8日 (2019 - 01 - 08) 说明书第[0053]-[0128]段, 图2-6	1, 2, 5-10, 13-18
Y	CN 109166565 A (百度在线网络技术北京有限公司) 2019年 1月 8日 (2019 - 01 - 08) 说明书第[0053]-[0128]段, 图2-6	3-4, 11-12
Y	CN 106934406 A (华南理工大学) 2017年 7月 7日 (2017 - 07 - 07) 说明书第[0034]-[0071]段, 图1-4	3-4, 11-12
X	US 2012007884 A1 (SAMSUNG ELECTRONICS CO., LTD.) 2012年 1月 12日 (2012 - 01 - 12) 说明书第[0027]-[0074]段, 附图1-4	1, 2, 5-10, 13-18
A	CN 107302548 A (中国电信股份有限公司) 2017年 10月 27日 (2017 - 10 - 27) 全文	1-18
A	WO 2008010024 A1 (CHERRADI, I.) 2008年 1月 24日 (2008 - 01 - 24) 全文	1-18

☒ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 3月 4日

国际检索报告邮寄日期

2020年 3月 18日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

李英

电话号码 86-(10)-53961445

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 108829253 A (北京科技大学) 2018年 11月 16日 (2018 - 11 - 16) 全文	1-18
A	CN 105389013 A (深港产学研基地) 2016年 3月 9日 (2016 - 03 - 09) 全文	1-18

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/126772

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	109828741	A	2019年 5月 31日	无			
CN	109166565	A	2019年 1月 8日	无			
CN	106934406	A	2017年 7月 7日	无			
US	2012007884	A1	2012年 1月 12日	KR	20120004023	A	2012年 1月 12日
				US	8866846	B2	2014年 10月 21日
				KR	101679239	B1	2016年 11月 24日
CN	107302548	A	2017年 10月 27日	无			
WO	2008010024	A1	2008年 1月 24日	US	2009322673	A1	2009年 12月 31日
				EP	2041640	B1	2012年 1月 25日
				US	9477310	B2	2016年 10月 25日
				EP	2041640	A1	2009年 4月 1日
				AT	543132	T	2012年 2月 15日
CN	108829253	A	2018年 11月 16日	无			
CN	105389013	A	2016年 3月 9日	无			