

(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
2023 年 3 月 2 日 (02.03.2023)



(10) 国际公布号
WO 2023/025005 A1

- (51) 国际专利分类号:
G06F 3/16 (2006.01) **G06K 9/00** (2022.01)
- (21) 国际申请号: PCT/CN2022/113074
- (22) 国际申请日: 2022 年 8 月 17 日 (17.08.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110971052.3 2021年8月23日 (23.08.2021) CN
- (71) 申请人: 维沃移动通信有限公司 (VIVO MOBILE COMMUNICATION CO., LTD.) [CN/CN];
中国广东省东莞市长安镇维沃路 1 号,
Guangdong 523863 (CN)。
- (72) 发明人: 喻超宁(YU, Chaoning); 中国广东省东莞市长安镇维沃路1号, Guangdong 523863 (CN)。
- (74) 代理人: 北京东方亿思知识产权代理有限公司 (BEIJING EAST IP LTD.); 中国北京市东城区东长安街1号东方广场东方经贸城东2座1601室, Beijing 100738 (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,

CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

(54) Title: AUDIO DATA PLAYING METHOD AND APPARATUS

(54) 发明名称: 音频数据播放方法与装置

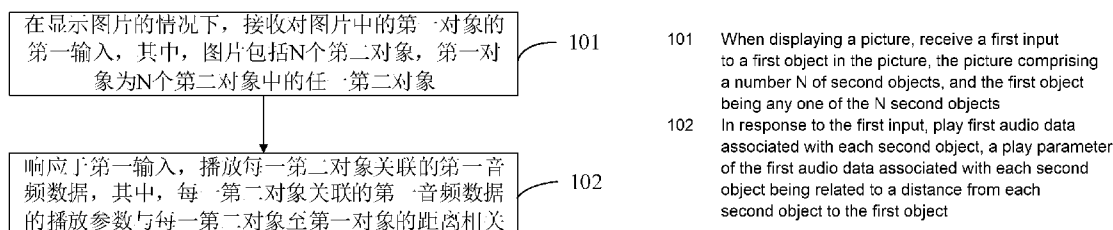


图 1

(57) Abstract: An audio data playing method and apparatus, belonging to the technical field of communications. The audio data playing method comprises: when displaying a picture, receiving a first input to a first object in the picture, the picture comprising a number N of second objects, and the first object being any one of the N second objects (101), N being an integer greater than 1; in response to the first input, playing first audio data associated with each second object, a play parameter of the first audio data associated with each second object being related to a distance from each second object to the first object (102).

(57) 摘要: 一种音频数据播放方法与装置, 属于通信技术领域。其中, 音频数据播放方法, 包括: 在显示图片的情况下, 接收对图片中的第一对象的第一输入, 图片包括N个第二对象, 第一对象为N个第二对象中的任一第二对象(101), N为大于1的整数; 响应于第一输入, 播放每一第二对象关联的第一音频数据, 每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关(102)。



WO 2023/025005 A1

音频数据播放方法与装置

相关申请的交叉引用

本申请主张在 2021 年 08 月 23 日在中国提交的中国专利申请号 202110971052.3 的优先权，其全部内容通过引用包含于此。

技术领域

本申请属于通信技术领域，具体涉及一种音频数据播放方法与装置。

背景技术

目前，随着对残障人士的关注度的提高，越来越多的电子设备支持无障碍模式，以便残障用户能够便捷使用电子设备。通常来说，在无障碍模式下，电子设备可以对用户点击区域的内容进行识别，并将识别结果以音频的形式进行播放，以告知用户点击区域的内容。

然而，现有技术中，电子设备在显示图片时，往往是对图片局部的内容进行识别与音频播放，难以向用户传达图片的整体内容。

发明内容

本申请实施例的目的是提供一种音频数据播放方法与装置，能够解决现有技术对图片局部的内容进行识别与音频播放，难以向用户传达图片的整体内容的问题。

第一方面，本申请实施例提供了一种音频数据播放方法，该方法包括：

在显示图片的情况下，接收对图片中的第一对象的第一输入，其中，图片包括 N 个第二对象，第一对象为 N 个第二对象中的任一第二对象，N 为大于 1 的整数；

响应于第一输入，播放每一第二对象关联的第一音频数据，其中，每

一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关。

第二方面，本申请实施例提供了一种音频数据播放装置，该装置包括：

第一接收模块，用于在显示图片的情况下，接收对图片中的第一对象的第一输入，其中，图片包括 N 个第二对象，第一对象为 N 个第二对象中的任一第二对象，N 为大于 1 的整数；

第一播放模块，用于响应于第一输入，播放每一第二对象关联的第一音频数据，其中，每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关。

第三方面，本申请实施例提供了一种电子设备，该电子设备包括处理器、存储器及存储在存储器上并可在处理器上运行的程序或指令，程序或指令被处理器执行时实现如第一方面的方法的步骤。

第四方面，本申请实施例提供了一种可读存储介质，可读存储介质上存储程序或指令，程序或指令被处理器执行时实现如第一方面的方法的步骤。

第五方面，本申请实施例提供了一种芯片，芯片包括处理器和通信接口，通信接口和处理器耦合，处理器用于运行程序或指令，实现如第一方面的方法。

第六方面，本申请实施例提供了一种计算机程序产品，包括有形地包含在计算机可读介质上的计算机程序，所述计算机程序包含用于执行如第一方面所述的方法的程序代码。

第七方面，本申请实施例提供了一种电子设备，被配置成用于执行如第一方面所述的方法的步骤。

本申请实施例提供的音频数据播放方法，在显示图片的情况下，接收对图片中的第一对象的第一输入，该图片包括多个第二对象，第一对象为这多个第二对象中的任一第二对象；响应于第一输入，播放每一第二对象关联的第一音频数据，每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关。本申请实施例有助于向用户传达图片

的整体内容，提高用户体验。

附图说明

图 1 是本申请实施例提供的音频数据播放方法的流程示意图；
图 2 是本申请实施例中图片的一个示例图；
图 3 是在一个具体应用例中，音频数据播放方法的流程示意图；
图 4 是本申请实施例提供的音频数据播放装置的结构示意图；
图 5 是本申请实施例提供的一种电子设备的结构示意图；
图 6 是本申请实施例提供的一种电子设备的硬件结构示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员获得的所有其他实施例，都属于本申请保护的范围。

本申请的说明书和权利要求书中的术语“第一”、“第二”等是用于区别类似的对象，而不适用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施，且“第一”、“第二”等所区分的对象通常为一类，并不限定对象的个数，例如第一对象可以是一个，也可以是多个。此外，说明书以及权利要求中“和/或”表示所连接对象的至少其中之一，字符“/”，一般表示前后关联对象是一种“或”的关系。

下面结合附图，通过具体的实施例及其应用场景对本申请实施例提供的音频数据播放方法与装置进行详细地说明。

如图 1 所示，本申请实施例提供的音频数据播放方法，包括：

步骤 101，在显示图片的情况下，接收对图片中的第一对象的第一输入，其中，图片包括 N 个第二对象，第一对象为 N 个第二对象中的任一第二对象，N 为大于 1 的整数；

步骤 102，响应于第一输入，播放每一第二对象关联的第一音频数

据，其中，每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关。

本申请实施例提供的音频数据播放方法，可以应用于电子设备。该电子设备可以是移动终端或者个人电脑等，此处不做具体限定。

电子设备可以显示图片，该图片中可以包括多个第二对象。

举例来说，上述的图片可以是在海边拍摄得到风景图片，相应地，图片可能包括大海、树以及鸟等第二对象。或者，图片可以是在车站拍摄得到图片，相应地，图片可能包括人物 A、人物 B 以及车辆等第二对象。

容易理解的是，图片中的第二对象，可以通过对图片的识别得到。

在一些举例中，可以预先基于深度学习模型，从图片中识别出各个第二对象。而在另一些举例中，也可以预先通过人工识别的方式，从图片中识别出各第二对象。

为简化描述，以下将主要以图片中的各个第二对象为通过深度学习模型进行识别得到为例进行说明。

在一个举例中，可以预先将图片发送至服务器，服务器使用深度学习模型对图片进行识别，并可以将相关的识别结果发送至电子设备。这些识别结果中可以包括了上述的 N 个第二对象。

在另一个举例中，也可以是电子设备直接使用深度学习模型对图片进行识别，以得到 N 个第二对象。

在步骤 101 中，电子设备可以接收对第一对象的第一输入，也就是任一第二对象的第一输入。

第一输入可以对应了单击、多击或者长按等形式的输入，此处不做具体限定。

容易理解的是，深度学习模型识别第二对象时，可以同时获取到各个第二对象在图片中的位置，具体来说，可以获取到各个第二对象对应的图像区域在图片中的位置。

在接收到第一输入的情况下，电子设备可以获取第一输入在图片中的输入位置，基于该输入位置，以及各个第二对象在图片中的位置，电子设备从图片中确定出上述的第一对象。

当然，在一些可行的实施方式中，电子设备也可以根据第一输入在图片中的输入位置，实时对输入位置及其预设距离范围内的图像区域进行识别，进而确定出第一对象。

步骤 102 中，电子设备可以响应于第一输入，播放每一第二对象关联的第一音频数据。

如上文所示的，通过深度学习模型可以识别出各第二对象。在一些举例中，每一类第二对象可以关联有一标识，该标识在一定程度上可以体现对第二对象的分类与识别结果。

比如，对于人物这一类第二对象，可以关联有标识“人物”，该标识可以通过文本的方式进行表达的。当然，在一些可行的实施方式中，上述的标识也可以是通过编号等方式进行表达的。

每一类第二对象可以关联有相应的第一音频数据，在一些可行的实施方式中，第二对象与第一音频数据的关联关系，可以体现在第二对象与第一音频数据的关联关系中。

结合一些应用场景的举例，电子设备可以根据各个第二对象的标识，在预设的音频数据库中查询各第二对象关联的第一音频数据，以便对各第二对象关联的第一音频数据进行播放。

当然，在另一些应用场景中，也可以是上述服务器在完成对第二对象的识别的情况下，从音频数据块中查询各第二对象关联的第一音频数据，并将各第一音频数据以及音频数据第二对象的关联关系发送至电子设备，以供电子设备进行播放。

为便于理解第一音频数据，以下主要以播放第一音频数据时发出的声音内容为例，来对第一音频数据进行代替说明。

在一些示例中，第一音频数据可以是对关联的第二对象简单的描述词。比如，对于人物这类第二对象，关联的第一音频数据可以是“人”；对于小狗这类第二对象，关联的第一音频数据可以是“小狗”；对于大海这类第二对象，关联的第一音频数据可以是“大海”。

而在另一示例中，第一音频数据也可以是关联的第二对象在实际环境下经常发出的声音。比如，对于人物这类第二对象，关联的第一音频数据

可以是“你好”；对于小狗这类第二对象，关联的第一音频数据可以是“汪汪汪”；对于大海这类第二对象，关联的第一音频数据可以是“哗哗哗”（海浪的声音）。

为了比较准确地向用户传达图片的整体内容，本实施例中，可以播放N个第二对象中每一第二对象关联的第一音频数据，同时，各个第二对象关联的第一音频数据的播放参数，可以与各第二对象与第一对象之间的距离相关。

比如，第一音频数据的播放参数可以包括播放音量，各个第二对象关联的第一音频数据的播放音量，可以与各第二对象与第一对象之间的距离负相关。换言之，对于一个第二对象，其与第一对象之间的距离越近，则其关联的第一音频数据的播放音量则可以越高。

再比如，第一音频数据的播放参数可以包括播放速度，各个第二对象关联的第一音频数据的播放频率，可以与各第二对象与第一对象之间的距离负相关。换言之，对于一个第二对象，其与第一对象之间的距离越近，则其关联的第一音频数据的播放速度则可以越快。

当然，以上是对播放参数的类型，以及播放参数与距离的相关关系的一些举例说明，在实际应用中，播放参数的类型与相关关系均可以根据需要进行设定。

为了简化说明，以下主要以播放参数包括播放音量为例进行说明。

对于第一对象，其与自身之间的距离为0。对于除第一对象以外的任一第二对象，其与第一对象之间的距离，可以在图像坐标系中的距离，也可以是在大地坐标系等坐标系中的距离。

比如，在图像坐标系中，任一第二对象的坐标，可以是该第二对象对应的图像区域的中点在图像坐标系中的坐标；第一对象与第二对象之间的距离，可以通过第一对象的坐标与第二对象的坐标进行计算。

再比如，可以从图片中识别出直线线条，并将直线线条作为参考线，来确定第一对象与第二对象之间的距离。结合一些应用场景，直线线条通常可以是道路边界、水面与天空的交界线等等。以直线线条作为参考线，可以得到第一对象与第二对象在大地坐标系中大致距离。

为便于更好地理解各个第二对象至第一对象的距离，与各个第二对象关联的第一音频数据的播放音量的关系，以下结合一个举例进行说明。

对于第一对象，其与自身之间的距离为 0，因此，第一对象关联的第一音频数据可以按照较高的音量进行播放，例如，以电子设备最大音量的 50% 进行播放。N 个第二对象中包括第一对象、第二对象 A 以及第二对象 B，其中，第二对象 A 与第一对象的距离，小于第二对象 B 与第一对象的距离。则可以以 40% 的最大音量播放第二对象 A 关联的第一音频数据，以 30% 的最大音量播放第二对象 B 关联的第一音频数据。

可见，通过播放图片中各个第二对象关联的第一音频数据，可以便于用户了解图片中所包括的对象；而各第二对象关联的第一音频数据的播放音量，与各第二对象与第一对象之间的距离负相关，有助于用户能够根据播放音量确定各第二对象的相对位置。综合各第二对象关联的第一音频数据，以及从播放音量中体现的各第二对象的相对位置，用户可以获知图片中各第二对象的类型与分布，从而比较准确地理解图片整体表达的内容。

类似地，当各第二对象关联的第一音频数据的播放参数包括播放速度时，也可以通过播放速度来体现各第二对象的相对位置。

本申请实施例提供的音频数据播放方法，在显示图片的情况下，接收对图片中的第一对象的第一输入，该图片包括多个第二对象，第一对象为这多个第二对象中的任一第二对象；响应于第一输入，播放每一第二对象关联的第一音频数据，每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关。本申请实施例有助于向用户传达图片的整体内容，提高用户体验。

与此同时，用户对图片进行第一输入，即可比较准确地理解图片的内容，因此，本申请实施例提供的音频数据播放方法，也能够有效提升操作便捷度。

在一个实施方式中，电子设备响应于第一输入，可以先播放第一对象关联的第一音频数据，在间隔预设时长后，再播放除第一对象以外的第二对象所关联的第一音频数据。

进一步地，除第一对象以外的第二对象所关联的第一音频数据的播放

顺序，也可以是与各第二对象至第一对象的距离相关的，第二对象距离第一对象越远，相关联的第一音频数据的播放顺序约靠后。

可选地，每一第二对象关联的第一音频数据的播放音色，与每一第二对象对应的图像区域的颜色相匹配。

如上文所示的，在完成对图片中的第二对象的识别的情况下，每一第二对象在图片中的图像区域也可以得到确定。而各个图像区域中像素的颜色通常是已知的，根据图像区域中各像素的颜色，可以确定图像区域的颜色。

通常情况下，像素的颜色可以通过 RGB 值或者灰度值进行表示，为简化描述，以下主要以像素的颜色通过 RGB 值进行表示为例进行说明。

结合一些举例，第二对象在图片中的图像区域中可能包括了多个像素，该图像区域的颜色，可以对应该图像区域中各像素的 RGB 值的众数或者平均数等。

播放音色可以与颜色存在预设的匹配关系，例如，播放音色可以与颜色的冷暖程度或者色系相匹配。相应地，在确定了各个第二对象对应的图像区域的颜色颜色的情况下，可以根据该预设的匹配关系，确定各个第二对象关联的第一音频数据的播放音色。

举例来说，颜色的冷暖程度可以通过如下公式进行计算：

$$CW=r*0.299+g*0.578+b*0.114$$

其中，CW 是用于衡量颜色冷暖程度的值，r 是 RGB 值中红色通道的数值，g 是 RGB 值中绿色通道的数值，b 是 RGB 值中蓝色通道的数值。r、g 以及 b 的取值范围通常均为 0~255。

在一个示例中，当 CW 大于或等于 192 时，可以认为是冷色，当 CW 小于 192 时，可以认为是暖色。对于冷色，可以匹配有比较沉闷的音色；对于暖和，可以匹配有比较轻快的音色。

或者，在另一个示例中，可以将 CW 的取值范围（0~253）按照预设步长划分为多个区间，按照 CW 从小到大的顺序，针对各个区间按轻快到沉闷的趋势匹配音色。

结合一个应用场景，白天拍摄的包括大海的图片中，大海这一第二对

象的颜色为蓝色，相应的 CW 较低，匹配的音色比较轻快，因此，在播放大海对应的第一音频数据时，可以以较轻快的音色进行播放。

而晚上拍摄的包括大海的图片中，大海这一第二对象的颜色近似黑色或墨绿色，相应的 CW 较高，匹配的音色比较沉闷，在在播放大海对应的第一音频数据时，可以以较沉闷的音色进行播放。

当然，在实际应用中，上述用于衡量颜色冷暖程度的值的计算方式，也可以根据实际需要进行调整。

如上文所示的，音色与颜色的匹配关系，还可以体现在音色与颜色所属的色系的匹配关系，也就是说，针对不同的色系，可以匹配不同的音色。为了简化说明，此处不再对根据各第二对象对应的图像区域的颜色的色系，确定各第二对象的第一音频数据的播放音色的过程进行详细说明。

结合上文中的举例，各个第二对象关联的第一音频数据的播放音色，可以是服务器完成对图片中的各第二对象识别后，根据各第二对象对应的图像区域的颜色确定的。后续服务器可以将播放音色与第二对象之间的关联关系发送至电子设备，以便电子设备选择合适的播放音色对各第二对象关联的第一音频数据进行播放。

当然，也可以是电子设备在接收到第一输入的情况下，再对图片进行识别得到 N 个第二对象，根据各第二对象对应的图像区域的颜色，以及预设的播放音色与颜色之间的匹配关系，确定出各第二对象对应的播放音色，并以确定的播放音色，播放第二对象关联的第一音频数据。

本实施例中，对于播放音色的确定时机可以不作具体限定，保证在播放各第二对象关联的第一音频数据时，播放音色与各第二对象对应的图像区域的颜色相匹配即可。本实施例中播放音色与第二对象对应图像区域的颜色匹配，有助于更有效地向用户传达图片的内容，提升用户对图片内容的收听体验。

实际应用中，例如大海、道路等类型的第二对象，在图片中对应的图像区域可能被其他的第二对象分割为多个子图像区域。相应地。在一些实施方式中，可以根据这些子图像区域的颜色，将颜色相同或相近的子图像区域归并为同一个第二对象对应的图像区域，再根据归并得到的图像区域

的颜色的冷暖程度或色系，确定匹配的播放音色。

可选地，上述步骤 102，响应于第一输入，播放每一第二对象关联的第一音频数据之后，音频数据播放方法还可以包括：

接收对第一对象的第二输入；

在第一对象的对象类型为预设对象类型的情况下，响应于第二输入，播放第一对象关联的第二音频数据；

其中，第二音频数据包括以下至少一项：

用于提示至少一个第二对象的状态的音频数据；

用于提示至少一个第三对象至第一对象的距离的音频数据，第三对象为 N 个第二对象中除第一对象之外的第二对象。

容易理解的是，第一对象可以是 N 个第二对象中的任一个第二对象，而第一对象的确定可以是与用户的输入相关的。比如，若用户的第一输入，是对大海这一第二对象的点击输入，则可以将点击的大海确定为第一对象；若用户的第二输入，是对人物这一第二对象的长按输入，则可以将第一对象更新为长按的人物。

当然，此时的第二输入的具体方式，可以并不限于长按输入，还可以是其他预设的手势输入等。比如，第二输入可以对应划问号或者划勾的手势输入。

本实施例中，在第一对象的对象类型为预设对象类型的情况下，电子设备可以响应于第二输入，播放第一对象关联的第二音频数据。

举例来说，预设对象类型可以是人物，当用户对对象类型为人物的第一对象进行第二输入后，电子设备可以以人物的口吻，来播放第一对象关联的第二音频数据。

而在另一些举例中，预设对象类型还可以是鹦鹉、喇叭或者其他对象类型，可以根据实际需要进行设置。

结合以上举例可见，作为一些可行的实施方式，预设对象类型可以对应在实际环境中能够发出提示性语言的对象。当然，在其他一些实施方式中，预设对象类型也可以是动物或者植物等对象类型，对应的对象可以通过拟人化的方式发出提示性的语言。

为了简化描述，以下将主要以预设对象类型为人物为例进行说明。

在一个示例中，第二音频数据可以包括用于提示至少一个第二对象的状态的音频数据。

结合一个应用场景，第一对象为人物时，可以以第一人称的口吻，对包括自身在内的各个第二对象进行状态的介绍。第二音频数据可以对应用于状态的介绍的音频数据。比如，第二音频数据可以是“我是人物 A，正在读报”、“我的身后是蓝色的大海”等。

也就是说，第二对象的状态，可以指第二对象的行为状态、颜色状态等等，可以根据实际需要进行设置，此处可以不做具体限定。

在另一个应用场景下，第一对象为人物时，可以以第一人称的口吻，对至少一个第三对象与自身的距离进行介绍。第二音频数据可以对应用于提示至少一个第三对象与第一对象之间的距离。比如，第二音频数据可以是“在我右侧约两指宽的距离处是人物 B”、“我的左侧是一棵树，它大约距离我一指宽”等。

当然，在又一应用场景下，第二音频数据也可以同时用于提示上述的状态与距离。比如，第二音频数据可以是“在我左侧约两指宽的距离处是人物 B，他正在打电话”。

可见，本实施例中，接收对第一对象的第二输入，在第一对象的对象类型为预设对象类型的情况下，响应于第二输入，播放第一对象关联的第二音频数据，有助于向用户传达图片中各个第二对象的状态或者距离等信息，以使用户更好地理解图片所表达的内容。

可选地，响应于第二输入，播放第一对象关联的第二音频数据之后，方法还包括：

在接收到第三输入的情况下，响应于第三输入，将第一对象更新为距离第三输入的终止位置最近，且对象类型为预设对象类型的第二对象；

播放更新后的第一对象所关联的第二音频数据。

以下结合一个应用例，来对本实施例的实施过程进行介绍。

该应用例中，图片中可以包括人物 A 和人物 B 这两个第二对象，用户可以点击（对应第一输入）人物 A 对应的图像区域，此时，电子设备可以

播放人物 A 关联的第一音频数据，例如“人物”或者“你好”等。

当用户在人物 A 对应的图像区域进行划问号的输入时（对应第二输入），此时，电子设备可以播放人物 A 关联的第二音频数据，例如“在我右侧约两指宽的距离处是人物 B”。

当用户从人物 A 所在的图像区域向右时（即滑动输入，对应第三输入），电子设备可以根据第三输入的终止位置，来重新确定第一对象。

比如，当第三输入的终止位置在人物 B 所在的图像区域时，可以将第一对象更新为人物 B，此时，可以播放与人物 B 关联的第二音频数据，例如“我是人物 B，我正在打电话，在我的右侧约一指宽的距离处是人物 C”。

当然，在实际应用中，第三输入的具体方式，也可以是长按输入等，此处可以不做具体限定。比如，用户可以根据人物 A 关联的第二音频数据的提示，确定出人物 B 对应的图像区域，进而可以对人物 B 对应的图像区域进行长按输入。

而至于预设对象类型，在上一实施例中进行了说明，此处不再赘述。

基于以上应用例可见，本实施例中，根据第三输入的终止位置，更新第一对象，并播放更新后的第一对象所关联的第二音频数据，有助于引导用户更为详细地获取图片中各个第二对象的状态或位置等信息，便于用户对图片的整体内容进行理解，提升用户体验。

在一个示例中，用户在进行第三输入时，可以是手指从人物 A 所在的图像区域划至人物 B 所在的图像区域。在滑动过程中，当用户手指距离人物 A 的图像区域较近时，可以保持播放人物 A 关联的第二音频数据，并且，随着用户手指至人物 A 的图像区域的距离的增加，第二音频数据的播放音量可以减小。

当用户的手指距离人物 B 所在的图像区域较近时，可以将第一对象从人物 A 更新为人物 B，进而播放人物 B 关联的第二音频数据，并且，随着用户手指至人物 B 的图像区域的距离的减小，第二音频数据的播放音量可以增大。

也就是说，该示例中，可以根据第三输入对应的实时输入位置，确定

第二音频数据的内容与播放参数（例如播放音量与播放速度），从而使得用户能够对第三输入的输入位置与各个第二对象之间的距离关系进行实时获取，便于更好地引导用户获取图片所表达的内容。

在一个示例中，电子设备可以统计在第三输入期间播放的全部第二音频数据所关联的 P 个第二对象，以及图片中对象类型为预设对象类型的第二对象的数量 Q 。其中， Q 为正整数， P 为小于或等于 Q 的正整数。

在第三输入结束后，电子设备可以输出 $Q-P$ 的值，结合以上应用例， $Q-P$ 的值，可以认为是未进行状态或距离等信息的介绍的第二对象。至于 $Q-P$ 的值，也可以是以音频播放的方式输出。如此，可以使得用户能够比较完整地了解图片中所展现的内容。

可选地，响应于第一输入，播放每一第二对象关联的第一音频数据之后，音频数据播放方法还可以包括以下至少一项：

在接收到对第一对象的第四输入的情况下，响应于第四输入，停止播放对象类型与第一对象的对象类型不同的第二对象所关联的第一音频数据；

在接收到对图片的第五输入的情况下，响应于第五输入，从图片中确定与第五输入匹配的目标图像区域，停止播放第四对象关联的第一音频数据，第四对象为对应的图像区域位于目标图像区域之外的第二对象；

在接收到对第一对象的第六输入的情况下，响应于第六输入，播放于第一对象关联的第三音频数据；

在接收到第七输入的情况下，响应于第七输入，根据第七输入的输入参数，调整第一音频数据的播放音量。

如上文所示的，第一对象可以是 N 个第二对象中的任一个第二对象，而第一对象的确定可以是与用户的输入相关的。在不同的输入阶段，第一对象可以是 N 个第二对象中的同一第二对象或者不同的第二对象。

以下结合一些应用例来对本实施例进行说明。在这些应用例中，图片可以包括人物 A、人物 B、大海以及草地。

在第一个应用例中，第四输入可以是对第一对象的双击或多击输入。例如，用户可以对图片中海进行双击，电子设备响应于用户对大海这一

第一对象的双击输入，保持对大海关联的第一音频数据的播放。例如，可以保持“哗哗哗”的海浪声音的播放；或者，大海关联的第一音频数据，还可以包括海风“呼呼呼”的声音，以及海鸟的叫声，保持对这些第一音频数据的播放即可。

而相应地，电子设备可以停止播放人物 A、人物 B 以及草地分别关联的第一音频数据的播放。

例如，在接收到第一输入的情况下，电子设备可以响应于第一输入，播放各个第二对象关联的第一音频数据。其中，人物 A 关联的第一音频数据，可以包括“你好”、“我是人物”以及“有什么可以帮助你呢”，这些第一音频数据可以间隔预设时长轮流播放或随机播放。而在接收到对大海的第四输入后，电子设备不再播放人物 A 关联的第一音频数据。

当然，如果第四输入为对人物 A 或者人物 B 的双击输入，则响应于第四输入，可以保持对人物 A 关联的第一音频数据与人物 B 关联的第一音频数据的播放，而停止播放大海关联的第一音频数据与草地关联的第一音频数据。

在第二个应用例中，第五输入可以是捏合输入。比如，捏合输入可以具体是至少三根手指相互靠拢的手势输入。

电子设备检测到上述捏合输入时，可以根据至少三根手指的终止位置，从图片中确定出与第五输入匹配的目标图像区域。比如，目标图像区域可以是三根手指终止位置的触点的连线围合的图像区域。

而在判断各个第二对象对应的图像区域是否位于目标图像区域之内时，可以是判断各第二对象对应的图像区域的中点是否位于目标图像区域之内；或者，可以是判断各第二对象对应的图像区域是否整体或部分位于目标图像区域之内等等，可以根据实际需要进行设置。

为简化说明，可以认为当某一第二对象对应的图像区域全部位于目标图像区域之外时，判定该第二对象对应的图像区域位于目标图像区域之外，该第二对象即可以确定为上述第四对象。后续电子设备可以停止对第四对象关联的第一音频数据的播放。

如此，电子设备可以根据用户的输入情况，对用户更加关注的第二对

象进行关联音频数据的播放。

当然，第五输入也可以是其他类型的手势输入，例如，第五输入可以是沿封闭轨迹的输入，则目标图像区域可以是对应封闭轨迹围合的区域。

在第三个应用例中，第六输入可以是往返滑动输入。

比如，用户对图片中的草地进行往返滑动输入时，电子设备可以发出“沙沙沙”的声音，以模仿草地被拨动的声音。

同一个第二对象关联的第三音频数据与第一音频数据之间可以存在差异。比如，对于草地这一第二对象，关联的第一音频数据，可以是频率较慢的“沙沙沙”的声音；而关联的第三音频数据，则可以是频率较快的“沙沙沙”的声音。

再比如，对于人物 A 这一第二对象，关联的第一音频数据，可以是“你好”，而关联的第三音频数据，可以是“请问有什么事情么”。

换言之，用户对任一第二对象的第六输入，可以认为是与该第二对象进行的动作交互。相应地，从电子设备的角度来说，当接收到对第一对象的第六输入时，可以响应于第六输入，播放预设的体现第一对象被交互时发出的声音，即播放上述的第一对象关联的第三音频数据，从而使得用户能够获得较好的交互体验。

在第四个应用例中，用户可以在电子设备上以较小的幅度画圈，从而调整整个第一音频数据的播放音量。该较小的幅度画圈的输入，可以认为对应第七输入。而较小的幅度的判断，可以是根据用户画圈的区域的大小进行判断。比如，当画圈的区域小于预设区域面积时，可以认为是以较小的幅度画圈。

容易理解的是，画圈的输入，也就是上述的第七输入，可以存在相应的输入参数，比如画圈的方向与圈数。

在一个示例中，当画圈的方向是顺时针时，可以调高各第一音频数据的播放音量；当画圈的方向是逆时针时，可以调低各第一音频数据的播放音量。而上述播放音量的调高或调低的程度，可以由圈数进行确定。

在一个示例中，在对各第一音频数据的播放音量进行调整后，各第一音频数据的播放音量之间的相对大小关系可以保持不变，也就是说，依然

存在每一第二对象关联的第一音频数据的播放音量，与每一第二对象至第一对象的距离负相关的关系。

结合以上应用例可见，本实施例中，电子设备可以根据用户的不同的手势输入，实现不同的音频数据播放功能，极大提高用户的操作便捷性。

可选地，上述步骤 102，响应于第一输入，播放每一第二对象关联的第一音频数据之前，音频数据播放方法还可以包括：

确定图片中的背景图像区域以及每一第二对象对应的图像区域；

从背景图像区域中确定出直线线条；

根据每一第二对象对应的图像区域与直线线条的位置关系，确定任两个第二对象之间的距离。

一般情况下，图片中会存在背景图像区域，比如天空或者大地所在的图像区域等。在一个示例中，可以将识别为天空或大地的第二对象所在的图像区域，直接确定为背景图像区域。

在实际应用中，例如天空或大地所在的图像区域可能被分割为多个子图像区域，这些子图像区域的颜色可能相同或相近，因此可以根据子图像区域的颜色，将这些子图像区域归入到背景图像区域。

结合一些应用场景，图片可以是通过相机拍摄得到的，相应地，图片中的内容可以是呈透视图的形式进行呈现的。也就是说，图片中的第二对象，可以是呈近大远小的形式呈现的。

比如，如图2所示，图片中包括大地 D1、公路 D2、树 D3、天空 D4、人物 D5 以及车辆 D6 这些第二对象，公路 D2 在远端汇集成一个点 TP，点 TP 可以是直线线条 L1、直线线条 L2 以及直线线条 L3 之间的交点。其中，直线线条 L1 可以是大地 D1 与天空 D4 之间的分割线；直线线条 L2 与直线线条 L3 为大地 D1 与公路 D2 之间的分割线。

若将大地 D1 与天空 D4 作为背景图像区域，则直线线条 L1、直线线条 L2 以及直线线条 L3 均可以从背景图像区域中确定出来。

而关于背景图像区域中直线线条的确定方式，可以通过图像分割或特征提取等技术进行获取，具体可以通过现有技术实现，此处不做赘述。

在从背景图像区域确定出直线线条的情况下，可以根据直线线条与各

第二对象对应的图像区域之间的位置关系，来确定任两个第二对象之间的距离。

为了简化说明，可以将第二对象对应的图像区域的中点，作为第二对象在图片中的位置。而至于第二对象对应的图像区域的确定方式，已在上文实施例中进行了说明，此处不做赘述

同样结合图 2，对于人物 D5 与车辆 D6，两个第二对象均与直线线条 L2 的距离较近，可以以直线线条 L2 作为参考，确定人物 D5 与车辆 D6 之间的距离。人物 D5 与车辆 D6 之间的连接线，可以分解至平行于直线线条 L2 的子线段，以及垂直于直线线条 L2 的子线段，根据两条子线段的长度，可以大致确定人物 D5 与车辆 D6 之间在两个方向上的距离，进而可以得到人物 D5 与车辆 D6 之间的距离。

可见，本实施例中，通过从背景图像区域确定出直线线条，基于直线线条来确定任两个第二对象之间的距离，可以比较准确地获取任两个第二对象在大地坐标系中的距离。后续在根据距离播放各第二对象关联的第一音频数据时，可以比较准确地向用户传达各第二对象与第一对象之间的距离关系。

可选地，上述步骤 101，播放每一第二对象关联的第一音频数据之前，音频数据播放方法还可以包括：

将图片发送至服务器，服务器用于识别图片，得到图片中 N 个第二对象中每一第二对象对应的图像区域的坐标；

响应于第一输入，播放每一第二对象关联的第一音频数据，包括：

响应于第一输入，将第一输入的输入参数发送至服务器，服务器用于根据输入参数以及每一第二对象对应的图像区域的坐标，生成音频播放规则，音频播放规则包括每一第二对象的关联的第一音频数据及其播放参数；

接收服务器发送的音频播放规则；

根据音频播放规则播放每一第二对象关联的第一音频数据。

本实施例中，对图片的识别以及各个音频播放规则的确定，可以是在服务器中进行的，如此，可以降低对电子设备的硬件配置的要求，降低电

子设备计算资源的消耗。

结合一个应用场景，电子设备在显示图片时，可以将图片发送至服务器，服务器则可以对图片进行识别，得到图片中的各个第二对象，以及各个第二对象对应的图像区域在图片中的坐标。

各个第二对象可以通过文字或者其他形式的标识进行代表。相应地，服务器可以将各第二对象的标识以及在图片中的坐标进行关联存储。为简化说明，可以认为服务器将各第二对象的标识以及在图片中的坐标存储在第一映射表中。

电子设备在接收到第一输入时，可以将第一输入的输入参数发送至服务器。举例来说，第一输入的输入参数可以包括用户点击的图像区域相对于图片的位置。

服务器根据第一输入的输入参数，以及上述的第一映射表，可以确定用户点击的图像区域对应的第二对象，即确定上述的第一对象。

结合一个举例，服务器在建立音频播放规则时，可以主要进行如下处理过程：

一是根据上述第一映射表中各第二对象对应的图像区域的坐标，确定各第二对象与第一对象之间的距离，以进一步确定各个第二对象对应的音频播放音量。如上文所示的，这里的距离与音频播放音量可以是负相关的，即距离越大，音频播放音量越低，反之亦然。

二是根据各个第二对象的标识，以及对象音频数据对应关系（该对应关系可以认为是存储在第二映射表中的），从预设的音频数据库中查询各第二对象关联的第一音频数据。

如此，服务器可以将第二对象—第一音频数据—音频播放音量的对应关系，作为音频播放规则发送至电子设备。电子设备则可以根据该音频播放规则，播放每一第二对象关联的第一音频数据。

当然，在一些实施方式中，服务器还可以进一步根据各个第二对象对应的图像区域的颜色，确定各个第二对象关联的第一音频数据的播放音色，并将播放音色加入到上述的音频播放规则中。

或者，服务器还可以将根据第二对象与第一对象之间的距离，确定各

个第二对象对应的音频播放速度，并将播放速度加入到上述的音频播放规则中。

如图 3 所示，以下结合一个具体应用例，对本申请实施例提供的音频数据播放方法进行说明。

该具体应用例中，音频数据播放方法可以应用于电子设备中，该电子设备可以与服务器进行数据交互。音频数据播放方法包括：

步骤 301，服务器解析图片中内容，提取图片中的第二对象；

容易理解的是，服务器解析的图片，可以是由电子设备发送至服务器的。而服务器可以是采用深度学习模型对图片进行解析。

步骤 302，电子设备接收用户的第一输入，将第一输入的输入参数发送至服务器；

比如，用户可以点击图片中的某一图像区域，电子设备可以将点击的图像区域相对图片整体的位置信息，作为输入参数发送至服务器。

步骤 303，服务器获取输入参数，确定第一对象，并按照相对第一对象从远到近的顺序，将各第二对象存储至预设数组中；

步骤 304，服务器根据预设数组中各第二对象的顺序，按从小到大规则为各第二对象分配音频播放音量；

当然，在一些实施方式中，服务器也可以为第二对象分配音频播放速度或者其他类型的播放参数。

步骤 305，计算各第二对象的图像区域的颜色的冷暖程度 CW（以下可以简称第二对象的 CW）；

CW 的一种可行的计算方式如下：

$$CW=r*0.299+g*0.578+b*0.114$$

其中，CW 是用于衡量颜色冷暖程度的值，r 是 RGB 值中红色通道的数值，g 是 RGB 值中绿色通道的数值，b 是 RGB 值中蓝色通道的数值。r、g 以及 b 的取值范围通常均为 0~255。

步骤 306，判断 CW 是否大于或等于 192，若是，执行步骤 307，若否，执行步骤 308；

步骤 307，根据 255 与第二对象的 CW 的差值，确定第二对象对应的

音频播放音色的沉闷程度，执行步骤 309；

步骤 308，根据第二对象的 CW 与 0 的差值，确定第二对象对应的音频播放音色的清脆程度，执行步骤 309；

步骤 309，按照对各第二对象确定音频播放音量与音频播放音色，播放各第二对象关联的音频数据。

结合以上具体应用例可见，本申请实施例提供的音频数据播放方法，通过确定图片中各个第二对象音频播放音量与音频播放音色，可以比较准确地向用户传达图片的整体内容，满足残障用户对图片内容的理解需求。

需要说明的是，本申请实施例提供的音频数据播放方法，执行主体可以为音频数据播放装置，或者该音频数据播放装置中的用于执行音频数据播放方法的控制模块。本申请实施例中以音频数据播放装置执行音频数据播放方法为例，说明本申请实施例提供的音频数据播放装置。

如图 4 所示，本申请实施例提供的音频数据播放装置 400，包括：

第一接收模块 401，用于在显示图片的情况下，接收对图片中的第一对象的第一输入，其中，图片包括 N 个第二对象，第一对象为 N 个第二对象中的任一第二对象，N 为大于 1 的整数；

第一播放模块 402，用于响应于第一输入，播放每一第二对象关联的第一音频数据，其中，每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关。

可选地，每一第二对象关联的第一音频数据的播放音色，与每一第二对象对应的图像区域的颜色相匹配。

可选地，音频数据播放装置 400 还可以包括：

第二接收模块，用于接收对第一对象的第二输入；

第二播放模块，用于在第一对象的对象类型为预设对象类型的情况下，响应于第二输入，播放第一对象关联的第二音频数据；

其中，第二音频数据包括以下至少一项：

用于提示至少一个第二对象的状态的音频数据；

用于提示至少一个第三对象至第一对象的距离的音频数据，第三对象为 N 个第二对象中除第一对象之外的第二对象。

可选地，音频数据播放装置 400 还可以包括：

更新模块，用于在接收到第三输入的情况下，响应于第三输入，将第一对象更新为距离第三输入的终止位置最近，且对象类型为预设对象类型的第二对象；

第三播放模块，用于播放更新后的第一对象所关联的第二音频数据。

可选地，音频数据播放装置 400 还可以包括以下至少一项：

第一停止播放模块，用于在接收到对第一对象的第四输入的情况下，响应于第四输入，停止播放对象类型与第一对象的对象类型不同的第二对象所关联的第一音频数据；

第二停止播放模块，用于在接收到对图片的第五输入的情况下，响应于第五输入，从图片中确定与第五输入匹配的目标图像区域，停止播放第四对象关联的第一音频数据，第四对象为对应的图像区域位于目标图像区域之外的第二对象；

第四播放模块，用于在接收到对第一对象的第六输入的情况下，响应于第六输入，播放于第一对象关联的第三音频数据；

调整模块，用于在接收到第七输入的情况下，响应于第七输入，根据第七输入的输入参数，调整第一音频数据的播放音量。

可选地，音频数据播放装置 400 还可以包括：

第一确定模块，用于确定图片中的背景图像区域以及每一第二对象对应的图像区域；

第二确定模块，用于从背景图像区域中确定出直线线条；

第四确定模块，用于根据每一第二对象对应的图像区域与直线线条的位置关系，确定任两个第二对象之间的距离。

可选地，音频数据播放装置 400 还可以包括：

发送模块，用于将图片发送至服务器，服务器用于识别图片，得到图片中 N 个第二对象中每一第二对象对应的图像区域的坐标；

相应地，第一播放模块 401，可以包括：

发送单元，用于响应于第一输入，将第一输入的输入参数发送至服务器，服务器用于根据输入参数以及每一第二对象对应的图像区域的坐标，

生成音频播放规则，音频播放规则包括每一第二对象的关联的第一音频数据及其播放参数；

接收单元，用于接收服务器发送的音频播放规则；

播放单元，用于根据音频播放规则播放每一第二对象关联的第一音频数据。

本申请实施例提供的音频数据播放装置，在显示图片的情况下，接收对图片中的第一对象的第一输入，响应于第一输入，播放每一第二对象关联的第一音频数据，每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关，如此，通过对各第二对象关联的第一音频数据的播放参数的处理，可以比较准确地向用户传达图片的内容。各第二对象关联的第一音频数据的播放音色，与各第二对象对应的图像区域的颜色相匹配，进一步方便了用户对图片内容的理解。此外，音频数据播放装置还可以响应用户相关的输入，来调整音频播放的焦点，从而满足用户对比较关注的图片内容的获取需求，提升用户使用体验。

本申请实施例中的音频数据播放装置可以是装置，也可以是终端中的部件、集成电路、或芯片。该装置可以是移动电子设备，也可以为非移动电子设备。示例性的，移动电子设备可以为手机、平板电脑、笔记本电脑、掌上电脑、车载电子设备、可穿戴设备、超级移动个人计算机（ultra-mobile personal computer, UMPC）、上网本或者个人数字助理（personal digital assistant, PDA）等，非移动电子设备可以为服务器、网络附属存储器（Network Attached Storage, NAS）、个人计算机（personal computer, PC）、电视机（television, TV）、柜员机或者自助机等，本申请实施例不作具体限定。

本申请实施例中的音频数据播放装置可以为具有操作系统的装置。该操作系统可以为安卓（Android）操作系统，可以为 iOS 操作系统，还可以为其他可能的操作系统，本申请实施例不作具体限定。

本申请实施例提供的音频数据播放装置能够实现图 1 至图 3 的方法实施例实现的各个过程，为避免重复，这里不再赘述。

可选地，如图 5 所示，本申请实施例还提供一种电子设备 500，包括

处理器 501，存储器 502，存储在存储器 502 上并可在处理器 501 上运行的程序或指令，该程序或指令被处理器 501 执行时实现上述音频数据播放方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

需要说明的是，本申请实施例中的电子设备包括上述的移动电子设备和非移动电子设备。

图 6 为实现本申请实施例的一种电子设备的硬件结构示意图。

该电子设备 600 包括但不限于：射频单元 601、网络模块 602、音频输出单元 603、输入单元 604、传感器 605、显示单元 606、用户输入单元 607、接口单元 608、存储器 609、以及处理器 610 等部件。

本领域技术人员可以理解，电子设备 600 还可以包括给各个部件供电的电源（比如电池），电源可以通过电源管理系统与处理器 610 逻辑相连，从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。图 6 中示出的电子设备结构并不构成对电子设备的限定，电子设备可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置，在此不再赘述。

其中，用户输入单元 607，用于在显示图片的情况下，接收对图片中的第一对象的第一输入，其中，图片包括 N 个第二对象，第一对象为 N 个第二对象中的任一第二对象，N 为大于 1 的整数；

音频输出单元 603，用于播放每一第二对象关联的第一音频数据，其中，每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关。

本申请实施例提供的电子设备，在显示图片的情况下，接收对图片中的第一对象的第一输入，该图片包括多个第二对象，第一对象为这多个第二对象中的任一第二对象；响应于第一输入，播放每一第二对象关联的第一音频数据，每一第二对象关联的第一音频数据的播放参数与每一第二对象至第一对象的距离相关。本申请实施例有助于向用户传达图片的整体内容，提高用户体验。

可选地，每一第二对象关联的第一音频数据的播放音色，与每一第二

对象对应的图像区域的颜色相匹配。

可选地，用户输入单元 607，还可用于接收对第一对象的第二输入；

音频输出单元 603，还可用于在第一对象的对象类型为预设对象类型的情况下，响应于第二输入，播放第一对象关联的第二音频数据；

其中，第二音频数据包括以下至少一项：

用于提示至少一个第二对象的状态的音频数据；

用于提示至少一个第三对象至第一对象的距离的音频数据，第三对象为 N 个第二对象中除第一对象之外的第二对象。

可选地，处理器 610，可用于在接收到第三输入的情况下，响应于第三输入，将第一对象更新为距离第三输入的终止位置最近，且对象类型为预设对象类型的第二对象；

音频输出单元 603，还可用于播放更新后的第一对象所关联的第二音频数据。

可选地，处理器 610，可用于在接收到对第一对象的第四输入的情况下，响应于第四输入，停止播放对象类型与第一对象的对象类型不同的第二对象所关联的第一音频数据；

可选地，处理器 610，可用于在接收到对图片的第五输入的情况下，响应于第五输入，从图片中确定与第五输入匹配的目标图像区域，停止播放第四对象关联的第一音频数据，第四对象为对应的图像区域位于目标图像区域之外的第二对象；

可选地，音频输出单元 603，还可用于在接收到对第一对象的第六输入的情况下，响应于第六输入，播放于第一对象关联的第三音频数据；

可选地，处理器 610，可用于在接收到第七输入的情况下，响应于第七输入，根据第七输入的输入参数，调整第一音频数据的播放音量。

可选地，处理器 610，可用于确定图片中的背景图像区域以及每一第二对象对应的图像区域；从背景图像区域中确定出直线线条；根据每一第二对象对应的图像区域与直线线条的位置关系，确定任两个第二对象之间的距离。

可选地，射频单元 601，可用于将图片发送至服务器，响应于第一输

入，将第一输入的输入参数发送至服务器，以及接收服务器发送的音频播放规则；

其中，服务器用于识别图片，得到图片中 N 个第二对象中每一第二对象对应的图像区域的坐标，服务器还用于根据输入参数以及每一第二对象对应的图像区域的坐标，生成音频播放规则，音频播放规则包括每一第二对象的关联的第一音频数据及其播放参数；

相应地，音频输出单元 603，还可用于根据音频播放规则播放每一第二对象关联的第一音频数据。

应理解的是，本申请实施例中，输入单元 604 可以包括图形处理器（Graphics Processing Unit, GPU）6041 和麦克风 6042，图形处理器 6041 对在视频捕获模式或图像捕获模式中由图像捕获装置（如摄像头）获得的静态图片或视频的图像数据进行处理。显示单元 606 可包括显示面板 6061，可以采用液晶显示器、有机发光二极管等形式来配置显示面板 6061。用户输入单元 607 包括触控面板 6071 以及其他输入设备 6072。触控面板 6071，也称为触摸屏。触控面板 6071 可包括触摸检测装置和触摸控制器两个部分。其他输入设备 6072 可以包括但不限于物理键盘、功能键（比如音量控制按键、开关按键等）、轨迹球、鼠标、操作杆，在此不再赘述。存储器 609 可用于存储软件程序以及各种数据，包括但不限于应用程序和操作系统。处理器 610 可集成应用处理器和调制解调处理器，其中，应用处理器主要处理操作系统、用户界面和应用程序等，调制解调处理器主要处理无线通信。可以理解的是，上述调制解调处理器也可以不集成到处理器 610 中。

本申请实施例还提供一种可读存储介质，可读存储介质上存储有程序或指令，该程序或指令被处理器执行时实现上述音频数据播放方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

其中，处理器为上述实施例中的电子设备中的处理器。可读存储介质，包括计算机可读存储介质，如计算机只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、磁碟或者光盘等。

本申请实施例另提供了一种芯片，芯片包括处理器和通信接口，通信接口和处理器耦合，处理器用于运行程序或指令，实现上述音频数据播放方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

应理解，本申请实施例提到的芯片还可以称为系统级芯片、系统芯片、芯片系统或片上系统芯片等。

需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。此外，需要指出的是，本申请实施方式中的方法和装置的范围不限按示出或讨论的顺序来执行功能，还可包括根据所涉及的功能按基本同时的方式或按相反的顺序来执行功能，例如，可以按不同于所描述的次序来执行所描述的方法，并且还可以添加、省去、或组合各种步骤。另外，参照某些示例所描述的特征可在其他示例中被组合。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以计算机软件产品的形式体现出来，该计算机软件产品存储在一个存储介质（如ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端（可以是手机，计算机，服务器，或者网络设备等）执行本申请各个实施例的方法。

上面结合附图对本申请的实施例进行了描述，但是本申请并不局限于上述的具体实施方式，上述的具体实施方式仅仅是示意性的，而不是限制性的，本领域的普通技术人员在本申请的启示下，在不脱离本申请宗旨和权利要求所保护的范围情况下，还可做出很多形式，均属于本申请的保护

之内。

权 利 要 求 书

1. 一种音频数据播放方法，包括：

在显示图片的情况下，接收对图片中的第一对象的第一输入，所述图片包括 N 个第二对象，所述第一对象为所述 N 个第二对象中的任一第二对象，N 为大于 1 的整数；

响应于所述第一输入，播放每一所述第二对象关联的第一音频数据，每一所述第二对象关联的第一音频数据的播放参数与每一所述第二对象至所述第一对象的距离相关。

2. 根据权利要求 1 所述的方法，其中，每一所述第二对象关联的第一音频数据的播放音色，与每一所述第二对象对应的图像区域的颜色相匹配。

3. 根据权利要求 1 所述的方法，其中，所述响应于所述第一输入，播放每一所述第二对象关联的第一音频数据之后，所述方法还包括：

接收对所述第一对象的第二输入；

在所述第一对象的对象类型为预设对象类型的情况下，响应于所述第二输入，播放所述第一对象关联的第二音频数据，

其中，所述第二音频数据包括以下至少一项：

用于提示至少一个所述第二对象的状态的音频数据；

用于提示至少一个第三对象至所述第一对象的距离的音频数据，所述第三对象为所述 N 个第二对象中除所述第一对象之外的第二对象。

4. 根据权利要求 3 所述的方法，其中，所述响应于所述第二输入，播放所述第一对象关联的第二音频数据之后，所述方法还包括：

在接收到第三输入的情况下，响应于所述第三输入，将所述第一对象更新为距离所述第三输入的终止位置最近，且对象类型为所述预设对象类型的第二对象；

播放更新后的第一对象所关联的第二音频数据。

5. 根据权利要求 1 所述的方法，其中，所述响应于所述第一输入，播放每一所述第二对象关联的第一音频数据之后，所述方法还包括以下至少

一项：

在接收到对第一对象的第四输入的情况下，响应于所述第四输入，停止播放对象类型与所述第一对象的对象类型不同的第二对象所关联的第一音频数据；

在接收到对所述图片的第五输入的情况下，响应于所述第五输入，从所述图片中确定与所述第五输入匹配的目标图像区域，停止播放第四对象关联的第一音频数据，所述第四对象为对应的图像区域位于所述目标图像区域之外的第二对象；

在接收到对所述第一对象的第六输入的情况下，响应于所述第六输入，播放于所述第一对象关联的第三音频数据；

在接收到第七输入的情况下，响应于所述第七输入，根据第七输入的输入参数，调整所述第一音频数据的播放音量。

6. 根据权利要求 1 所述的方法，其中，所述响应于所述第一输入，播放每一所述第二对象关联的第一音频数据之前，所述方法还包括：

确定所述图片中的背景图像区域以及每一所述第二对象对应的图像区域；

从所述背景图像区域中确定出直线线条；

根据每一所述第二对象对应的图像区域与所述直线线条的位置关系，确定任两个所述第二对象之间的距离。

7. 根据权利要求 1 所述的方法，其中，所述播放每一所述第二对象关联的第一音频数据之前，所述方法还包括：

将所述图片发送至服务器，所述服务器用于识别所述图片，得到所述图片中 N 个第二对象中每一所述第二对象对应的图像区域的坐标；

所述响应于所述第一输入，播放每一所述第二对象关联的第一音频数据，包括：

响应于所述第一输入，将所述第一输入的输入参数发送至所述服务器，所述服务器用于根据所述输入参数以及每一所述第二对象对应的图像区域的坐标，生成音频播放规则，所述音频播放规则包括每一所述第二对象的关联的第一音频数据及其播放参数；

接收所述服务器发送的音频播放规则；

根据所述音频播放规则播放每一所述第二对象关联的第一音频数据。

8. 一种音频数据播放装置，包括：

第一接收模块，用于在显示图片的情况下，接收对图片中的第一对象的第一输入，所述图片包括 N 个第二对象，所述第一对象为所述 N 个第二对象中的任一第二对象，N 为大于 1 的整数；

第一播放模块，用于响应于所述第一输入，播放每一所述第二对象关联的第一音频数据，每一所述第二对象关联的第一音频数据的播放参数与每一所述第二对象至所述第一对象的距离相关。

9. 一种电子设备，包括处理器，存储器及存储在所述存储器上并可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如权利要求 1-7 任一项所述的音频数据播放方法的步骤。

10. 一种可读存储介质，所述可读存储介质上存储程序或指令，所述程序或指令被处理器执行时实现如权利要求 1-7 任一项所述的音频数据播放方法的步骤。

11. 一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现如权利要求 1-7 中任一项所述的显示方法的步骤。

12. 一种计算机程序产品，包括有形地包含在计算机可读介质上的计算机程序，所述计算机程序包含用于执行权利要求 1-7 中任一项所述的方法的程序代码。

13. 一种电子设备，被配置成用于执行如权利要求 1-7 中任一项所述的方法的步骤。

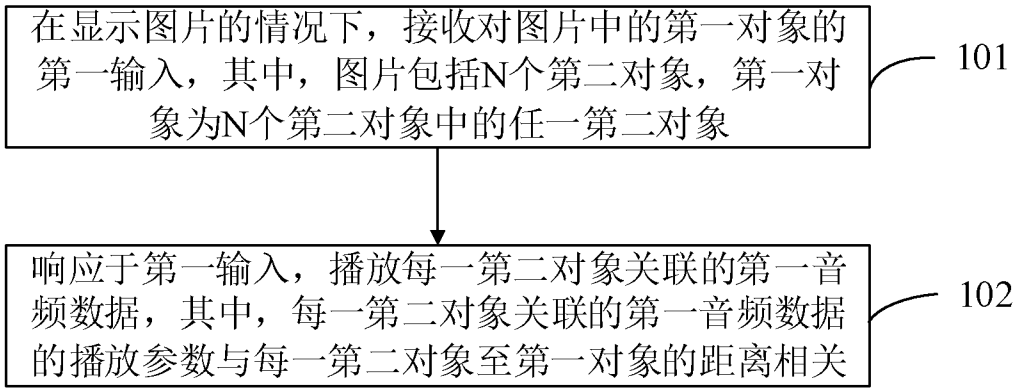


图 1

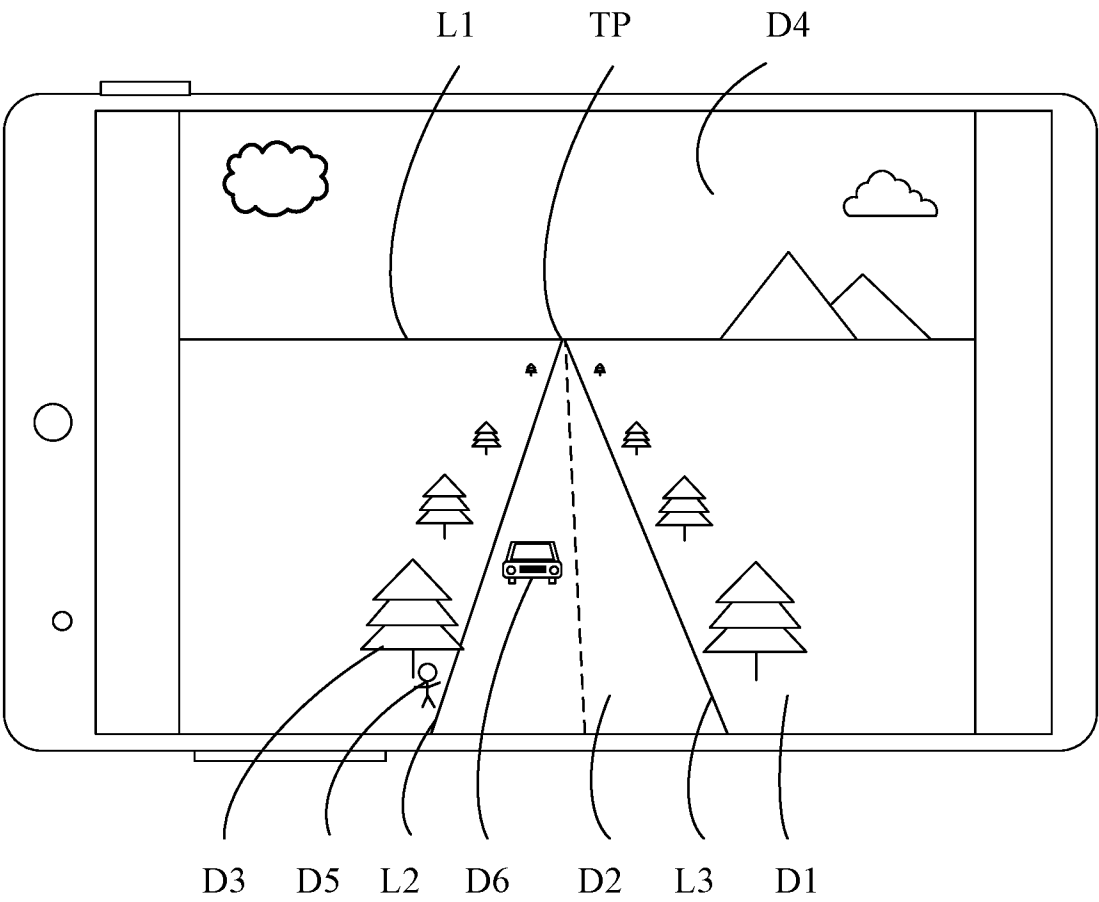


图 2

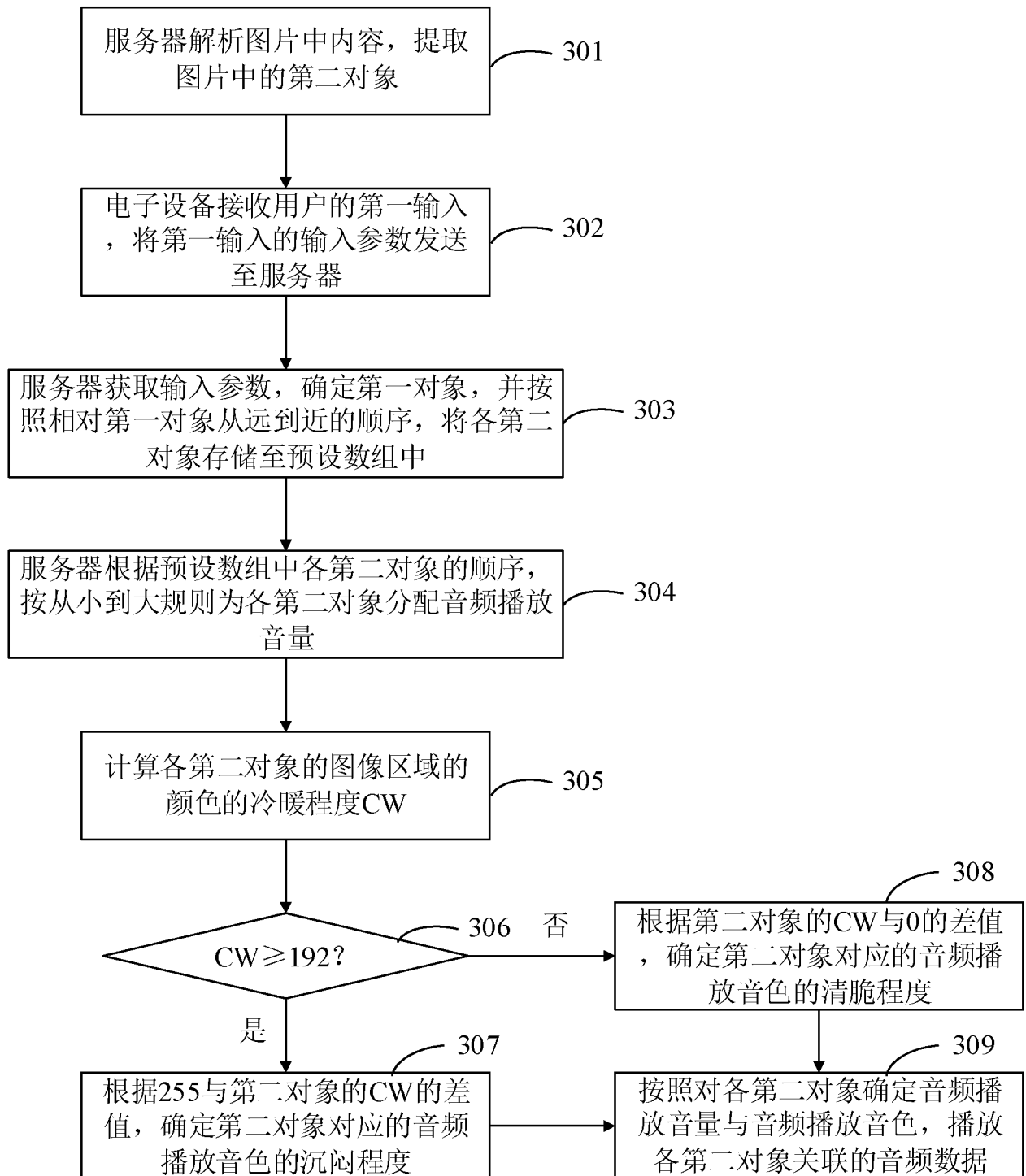


图 3

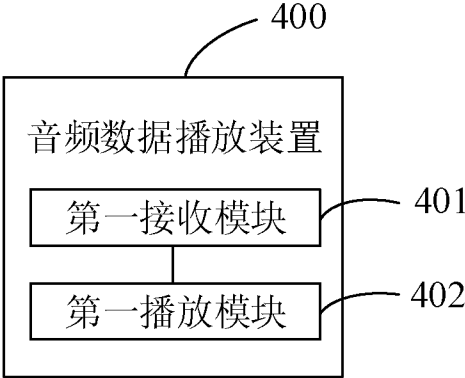


图 4

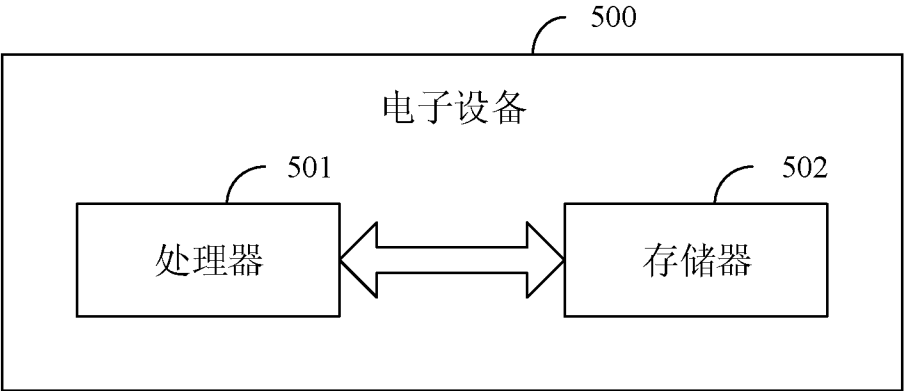


图 5

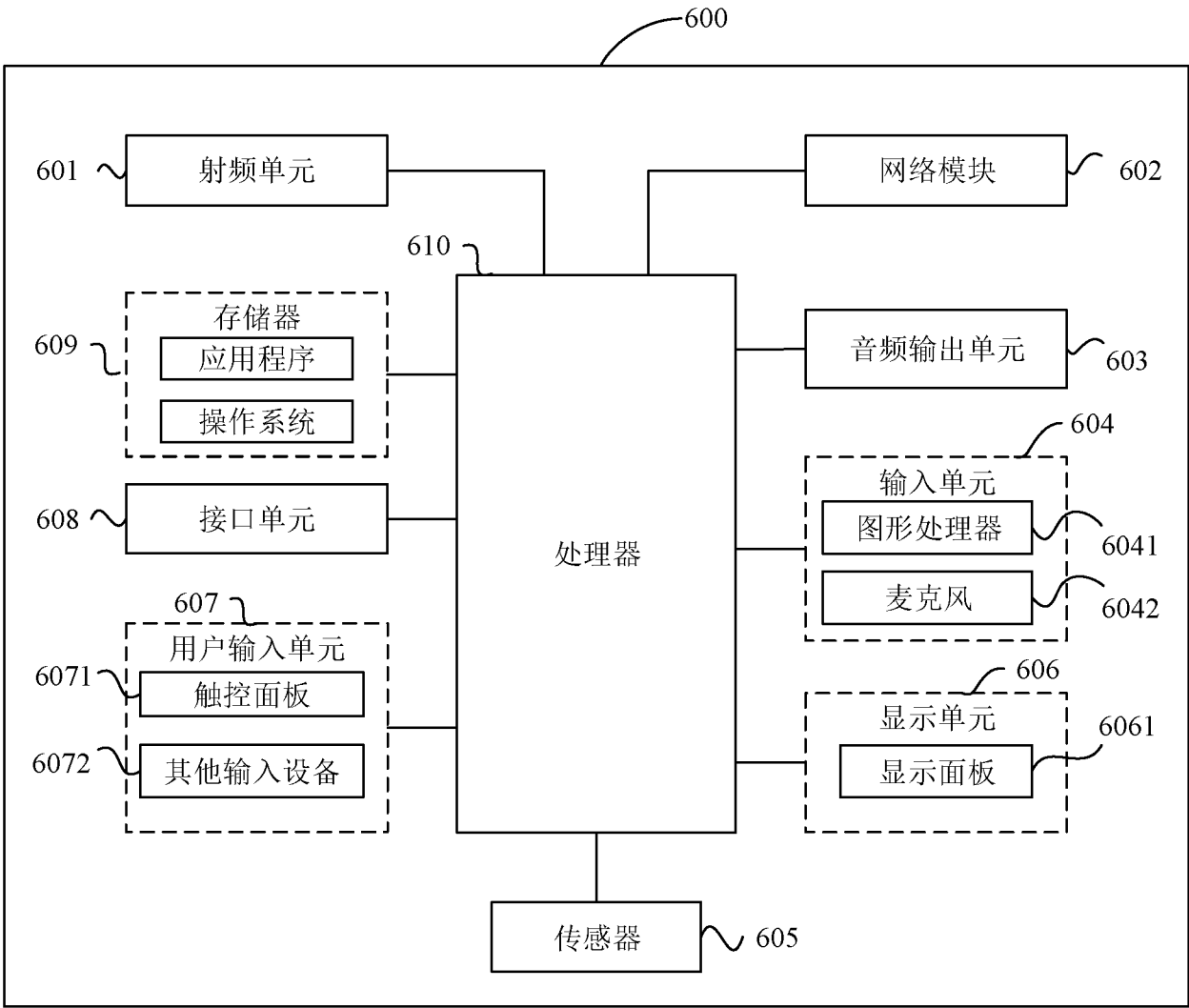


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/113074

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/16(2006.01)i; G06K 9/00(2022.01)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)

CNPAT, IEEE, CNKI, WPI, EPODOC: 图像, 图片, 照片, 读, 对象, 元素, 多, 位置, 距离, 播放, 播报, image, photo, read, object, element, multiple, distance, position, play

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113672193 A (VIVO COMMUNICATION TECHNOLOGY CO., LTD.) 19 November 2021 (2021-11-19) claims 1-10, and description, paragraph 0254	1-13
A	CN 107895006 A (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CO., LTD.) 10 April 2018 (2018-04-10) description, paragraphs 0038-0060, and figures 1-3	1-13
A	CN 104157171 A (SAMSUNG ELECTRONICS (CHINA) R & D CENTER et al.) 19 November 2014 (2014-11-19) entire document	1-13
A	CN 110519636 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 29 November 2019 (2019-11-29) entire document	1-13
A	CN 107885823 A (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CO., LTD.) 06 April 2018 (2018-04-06) entire document	1-13
A	CN 109271628 A (NORTHEASTERN UNIVERSITY) 25 January 2019 (2019-01-25) entire document	1-13



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

02 November 2022

Date of mailing of the international search report

21 November 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
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/CN2022/113074**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2020063614 A1 (NEXTVPU SHANGHAI CO., LTD.) 02 April 2020 (2020-04-02) entire document	1-13
A	US 2020258422 A1 (CAN-U-C LTD.) 13 August 2020 (2020-08-13) entire document	1-13

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/113074

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113672193	A	19 November 2021	None			
CN	107895006	A	10 April 2018	None			
CN	104157171	A	19 November 2014	None			
CN	110519636	A	29 November 2019	None			
CN	107885823	A	06 April 2018	None			
CN	109271628	A	25 January 2019	None			
WO	2020063614	A1	02 April 2020	JP	2020053055	A	02 April 2020
				KR	20200141517	A	18 December 2020
				EP	3640840	A2	22 April 2020
				US	2020097133	A1	26 March 2020
				CN	110955043	A	03 April 2020
				CN	110955325	A	03 April 2020
US	2020258422	A1	13 August 2020	IL	285303	A	30 September 2021
				CN	113678141	A	19 November 2021
				EP	3924873	A1	22 December 2021
				WO	2020165899	A1	20 August 2020
				JP	2022526702	A	26 May 2022
				KR	20210126664	A	20 October 2021
				BR	112021015739	A2	26 October 2021
				IN	202147040067	A	10 September 2021

国际检索报告

国际申请号

PCT/CN2022/113074

A. 主题的分类

G06F 3/16(2006.01)i; G06K 9/00(2022.01)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))

CNPAT, IEEE, CNKI, WPI, EPDOC : 图像, 图片, 照片, 读, 对象, 元素, 多, 位置, 距离, 播放, 播报, image, photo, read, object, element, multiple, distance, position, play

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 113672193 A (维沃移动通信有限公司) 2021年11月19日 (2021 - 11 - 19) 权利要求1-10, 说明书第0254段	1-13
A	CN 107895006 A (广东欧珀移动通信有限公司) 2018年4月10日 (2018 - 04 - 10) 说明书第0038-0060段, 图1-3	1-13
A	CN 104157171 A (三星电子中国研发中心 等) 2014年11月19日 (2014 - 11 - 19) 全文	1-13
A	CN 110519636 A (腾讯科技深圳有限公司) 2019年11月29日 (2019 - 11 - 29) 全文	1-13
A	CN 107885823 A (广东欧珀移动通信有限公司) 2018年4月6日 (2018 - 04 - 06) 全文	1-13
A	CN 109271628 A (东北大学) 2019年1月25日 (2019 - 01 - 25) 全文	1-13
A	WO 2020063614 A1 (NEXTVPU SHANGHAI CO., LTD.) 2020年4月2日 (2020 - 04 - 02) 全文	1-13
A	US 2020258422 A1 (CAN-U-C LTD.) 2020年8月13日 (2020 - 08 - 13) 全文	1-13

☐ 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

2022年11月2日

国际检索报告邮寄日期

2022年11月21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

授权官员

王平

电话号码 86-(10)-53961372

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/113074

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	113672193	A	2021年11月19日	无			
CN	107895006	A	2018年4月10日	无			
CN	104157171	A	2014年11月19日	无			
CN	110519636	A	2019年11月29日	无			
CN	107885823	A	2018年4月6日	无			
CN	109271628	A	2019年1月25日	无			
WO	2020063614	A1	2020年4月2日	JP	2020053055	A	2020年4月2日
				KR	20200141517	A	2020年12月18日
				EP	3640840	A2	2020年4月22日
				US	2020097133	A1	2020年3月26日
				CN	110955043	A	2020年4月3日
				CN	110955325	A	2020年4月3日
US	2020258422	A1	2020年8月13日	IL	285303	A	2021年9月30日
				CN	113678141	A	2021年11月19日
				EP	3924873	A1	2021年12月22日
				WO	2020165899	A1	2020年8月20日
				JP	2022526702	A	2022年5月26日
				KR	20210126664	A	2021年10月20日
				BR	112021015739	A2	2021年10月26日
				IN	202147040067	A	2021年9月10日