

(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
2024年9月6日 (06.09.2024)



(10) 国际公布号
WO 2024/179425 A1

- (51) 国际专利分类号:
G10L 15/22 (2006.01)
- (21) 国际申请号: PCT/CN2024/078662
- (22) 国际申请日: 2024年2月27日 (27.02.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202310224268.2 2023年2月28日 (28.02.2023) CN
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 龙水平 (LONG, Shuiping); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。许强 (XU, Qiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。吴文昊 (WU, Wenhao); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。曾

宇浩 (ZENG, Yuhao); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。李琛贺 (LI, Chenhe); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 广州三环专利商标代理有限公司 (SCIHEAD IP LAW FIRM); 中国广东省广州市越秀区先烈中路80号汇华商贸大厦1508室, Guangdong 510070 (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, MG, MK, MN, MU, 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。

(54) Title: VOICE INTERACTION METHOD AND RELATED DEVICE

(54) 发明名称: 语音交互方法及相关设备

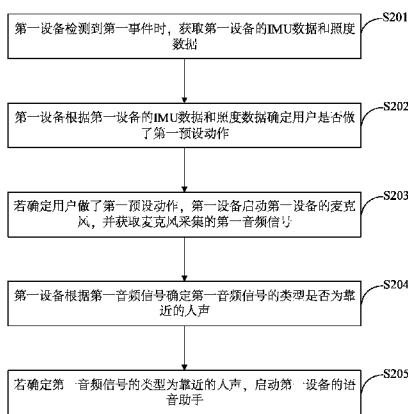


图 2

- S201 When a first device detects a first event, the first device acquires IMU data and illuminance data of the first device
- S202 The first device, according to the IMU data and illuminance data of the first device, determines whether a user performed a first preset action
- S203 If it is determined that the user performed the first preset action, the first device turns on a microphone of the first device and acquires a first audio signal acquired by the microphone
- S204 The first device, according to the first audio signal, determines whether the type of the first audio signal is an approaching human voice
- S205 If it is determined that the type of the first audio signal is an approaching human voice, turn on a voice assistant of the first device

(57) Abstract: A voice interaction method and a related device, which relate to the field of artificial intelligence. When a first device detects a first event, the first device acquires IMU data and illuminance data of the first device (S201); according to the IMU data and illuminance data of the first device, it is determined whether a user performed a first preset action (S202); if it is determined that the user performed the first preset action, the first device turns on a microphone of the first device and acquires a first audio signal acquired by the microphone (S203); according to the first audio signal, it is determined whether the type of the first audio signal is an approaching human voice (S204); and if it is determined that the type of the first audio signal is an approaching human voice, the first device turns on a voice assistant of the first device (S205). An event in which the user approaches the first device may be more accurately detected by means of the IMU data and the illuminance data, instead of only using acceleration data, which requires the user to significantly raise a wrist/hand to be able to detect the event in which the user approaches the device. Therefore, the solution is more sensitive, and the detection may be achieved when the user naturally raises his or her wrist or hand.

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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: 一种语音交互方法及相关设备, 涉及人工智能领域。第一设备在检测到第一事件时, 获取第一设备的IMU数据和照度数据 (S201); 根据第一设备的IMU数据和照度数据确定用户是否做了第一预设动作 (S202); 若确定用户做了第一预设动作, 第一设备启动第一设备的麦克风, 获取麦克风采集的第一音频信号 (S203); 根据第一音频信号确定第一音频信号的类型是否为靠近的人声 (S204); 若确定第一音频信号的类型为靠近的人声, 第一设备启动第一设备的语音助手 (S205)。通过IMU数据和照度数据可以更加准确的检测用户靠近第一设备的事件, 而不是仅使用加速度数据时需要用户显著抬腕/抬手才能检测到用户靠近设备的事件, 该方案更加的灵敏, 并且在自然抬腕或者抬手时即可实现检测。

语音交互方法及相关设备

本申请要求于 2023 年 02 月 28 日提交中国国家知识产权局、申请号为 202310224268.2、发明名称为“语音交互方法及相关设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及人工智能领域，尤其涉及一种语音交互方法及相关设备。

背景技术

当前语音助手交互的方式主要是两种。一种方式下，用户讲出特定的唤醒词（比如“小艺小艺”），智能终端识别到该语音信号后发起语音会话，这种方式在公共场所所有隐私的问题，而且交互较冗长。另一种方式，用户执行特定的动作，比如显著的抬腕动作，或长按实体键（比如电源键或运动快捷键），按键操作需要较大力度和延迟，一不小心还可能调出重启/关机界面，交互过程不够简洁。

发明内容

本申请提供一种语音交互方法及相关设备，可以提高检测交互操作的灵敏度。

第一方面，本申请提供一种语音交互方法，应用于第一设备，该方法包括：

第一设备在检测到第一事件时，获取第一设备的惯性测量单元（inertial measurement unit, IMU）数据和照度数据；根据第一设备的 IMU 数据和照度数据确定用户是否做了第一预设动作；若确定用户做了第一预设动作，第一设备启动第一设备的麦克风，获取麦克风采集的第一音频信号；根据第一音频信号确定第一音频信号的类型是否为靠近的人声；若确定第一音频信号的类型为靠近的人声，第一设备启动所述第一设备的语音助手。

其中，第一预设动作为抬腕动作或者抬手动作。

通过 IMU 数据和照度数据可以更加准确的检测用户靠近第一设备的事件，而不是仅使用加速度数据时需要用户显著抬腕/抬手才能检测到用户靠近设备的事件，本申请的方案更加的灵敏，并且在自然抬腕或者抬手时即可实现检测。

在一个可能的实现方式中，第一事件为抬腕硬件中断事件、第一设备的抬手硬件中断事件、第一设备的按键亮屏事件、第一设备的抬手亮屏事件或者第一设备的抬腕亮屏事件。

在一个可能的实现方式中，本申请的方法还包括：

在启动第一设备的麦克风之后，持续获取第一设备采集的 IMU 数据和照度数据，根据第一设备采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作；

若确定第一音频信号的类型为靠近的人声，启动第一设备的语音助手，包括：

若确定用户执行了所述第二预设动作，且确定第一音频信号的类型为靠近的人声，启动第一设备的语音助手。

其中，第二预设动作为抬腕保持动作或者抬手保持动作。

通过在启动第一设备的麦克风之后，持续获取第一设备采集的 IMU 数据和照度数据，并基于采集的 IMU 数据和照度数据确定用户是否保持了抬腕动作或者是否保持了抬手动作，在确定用户保持了抬腕动作或者保持了抬手动作，且第一音频信号的类型为靠近的人声时，才会启动第一设备的语音助手，有利于降低误唤醒率和干扰。

在一个可能的实现方式中，本申请的方法还包括：

若确定用户未执行第二预设动作，则第一设备关闭第一设备的麦克风。

在确定用户未执行第二预设动作，也就是用户未保持抬手动作或者未保持抬手动作，表示用户不是准确唤醒语音助手，在这种情况下第一设备关闭麦克风，可以降低第一设备功率，同时也降低了语音助手的误唤醒率。

在一个可能的实现方式中，确定用户是否执行了第一预设动作所使用的 IMU 数据和照度数据的采集时长为第一预设时长，确定用户是否执行了第二预设动作所使用的 IMU 数据和照度数据的采集时长为第二预设时长，其中，第一预设时长大于第二预设时长。

通过设置第二预设时长小于第一预设时长，使得由于在确定用户是否执行了第二预设动作之前不用采集较长的 IMU 数据和照度数据，进而可以快速的检测用户是否执行了第二预设动作。

在确定用户是否执行了第二预设动作所使用的 IMU 数据和照度数据的采集时长小于确定用户是否执行了第一预设动作所使用的 IMU 数据和照度数据的采集时长时，可以通过分别对确定用户是否执行了第二预设动作所使用的 IMU 数据和照度数据进行插值处理，使得确定用户是否执行了第二预设动作所使用的 IMU 数据和照度数据的等效时长与确定用户是否执行了第一预设动作所使用的 IMU 数据和照度数据的采集时长一致，进而可以使用同一个预测模型预测确定用户是否执行了第一预设动作和第二预设动作。

在一个可能的实现方式中，第一事件为抬腕硬件中断事件、第一设备的抬手硬件中断事件、第一设备的抬手亮屏事件或者第一设备的抬腕亮屏事件，获取所述第一设备的 IMU 数据和照度数据，包括：

启动第一设备的环境光传感器，采集所述第一设备的环境的照度数据和继续采集 IMU 数据，并获取与所述第一事件相关的缓存的 IMU 数据，并对与所述第一事件相关的照度数据置零补齐。

通过对照度数据进行置零补齐，可以使得照度数据的等效时长与 IMU 数据的采集时长一致，进而可以在判断用户是否执行了第一预设动作时可以使用完整的抬腕阶段或者抬手阶段的数据，使得判断结果更加准确，并且在用户体验上提高了唤醒的成功率，降低了误唤醒率。

在一个可能的实现方式中，所述第一音频信号的时长小于或者等于 0.5s，且根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声是所述第一设备的数字信号处理器 DSP 实现的。

通过将第一音频信号的时长设定为小于或者等于 0.5s，是为了后续启动语音助手识别第一语音信号时，可以实时回显第一音频信号对应的文本。并且第一音频信号是未经过自动增益控制 (automatic gain control, AGC)、降噪、去混响或者压缩处理的音频信号，不存在靠近人声信息的丢失，使用未经过 AGC、降噪、去混响或者压缩处理的原始音频信号来确定第一音频信号的声音类型，可以提高确定的第一音频信号的声音类型的精度。

在一个可能的实现方式中，所述第一音频信号是通过所述第一设备的多个麦克风采集得到的。

针对风噪场景，不同麦克风采集的音频信号中存在风噪信号，可以利用一个麦克风采集的音频信号中的风噪对另一麦克风采集的音频信号中的风噪进行降风噪处理，再利用将风噪处理会后的音频信号进行声音类型的预测，可以得到更加准确的预测结果。

在一个可能的实现方式中，本申请的方法还包括：

第一设备根据第一设备的麦克风采集的音频信号确定待执行任务；若待执行任务为敏感任务，第一设备获取用户的身份信息；在基于用户的身份信息确定用户为目标用户，第一设备执行待处理任务。

其中，上述用户为第一设备的使用者，目标用户为第一设备的拥有者。

通过上述方式可以避免非第一设备拥有者通过第一设备执行安全敏感的语音任务，保证了第一设备拥有者的设备的信息安全。

在一个可能的实现方式中，本申请的方法还包括：

在启动语音助手后，第一设备实时显示采集的音频信号对应的文本。

第二方面，本申请提供一种电子设备，包括：

获取单元，用于在检测到第一事件时，获取电子设备的 IMU 数据和照度数据；

确定单元，用于根据电子设备的 IMU 数据和照度数据确定用户是否做了第一预设动作；

启动单元，用于若确定用户做了所述第一预设动作，启动电子设备的麦克风，获取单元，还用于获取麦克风采集的第一音频信号；

确定单元，还用于根据第一音频信号确定第一音频信号的类型是否为靠近的人声；

启动单元，还用于若确定第一音频信号的类型为靠近的人声，启动电子设备的语音助手。

在一个可能的实现方式中，第一事件为抬腕硬件中断事件、所述电子设备的抬手硬件中断事件、电子设备的按键亮屏事件、所述电子设备的抬手亮屏事件或者所述电子设备的抬腕亮屏事件。

在一个可能的实现方式中，获取单元，还用于在启动单元启动电子设备的麦克风之后，持续获取电子设备采集的 IMU 数据和照度数据，

确定单元，还用于根据电子设备采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作；

在若确定第一音频信号的类型为靠近的人声，启动电子设备的语音助手的方面，启动单元具体用于：若确定用户执行了第二预设动作，且确定第一音频信号的类型为靠近的人声，启动电子设备的语音助手。

在一个可能的实现方式中，启动单元还用于：

若确定用户未执行第二预设动作，则关闭电子设备的麦克风。

在一个可能的实现方式中，确定用户是否执行了第一预设动作所使用的 IMU 数据和照度数据的采集

时长为第一预设时长，确定用户是否执行了第二预设动作所使用的 IMU 数据和照度数据的采集时长为第二预设时长，其中，第一预设时长大于第二预设时长。

在一个可能的实现方式中，第一事件为抬腕硬件中断事件、电子设备的抬手硬件中断事件、电子设备的抬手亮屏事件或者电子设备的抬腕亮屏事件，在获取电子设备的 IMU 数据和照度数据的方面，获取单元具体用于：

启动电子设备的环境光传感器，采集电子设备的环境的照度数据和继续采集 IMU 数据，并获取与第一事件相关的缓存的 IMU 数据，并对与第一事件相关的照度数据置零补齐。

在一个可能的实现方式中，第一音频信号的时长小于或者等于 0.5s，且根据第一音频信号确定第一音频信号的类型是否为靠近的人声是电子设备的数字信号处理器(digital signal processor, DSP)实现的。

在一个可能的实现方式中，第一音频信号是通过电子设备的多个麦克风采集得到的。

在一个可能的实现方式中，确定单元，还用于根据电子设备的麦克风采集的音频信号确定待执行任务；获取单元，还用于若待执行任务为敏感任务，获取用户的身份信息；

电子设备还包括：

执行单元，用于在基于用户的身份信息确定用户为目标用户，执行待处理任务。

在一个可能的实现方式中，电子设备还包括：

显示单元，用于在启动语音助手后，实时显示采集的音频信号对应的文本。

第三方面，本申请提供另一种语音交互方法，应用于第一设备，该方法包括：

第一设备在检测到第一事件时，获取第一设备的 IMU 数据；根据第一设备的 IMU 数据确定用户是否做了第一预设动作；若确定用户做了第一预设动作，第一设备启动第一设备的麦克风，获取麦克风采集的第一音频信号；根据第一音频信号确定第一音频信号的类型是否为靠近的人声；若确定第一音频信号的类型为靠近的人声，第一设备启动所述第一设备的语音助手。

其中，第一事件为第一设备的抬腕事件、第一设备的抬手事件或者第一设备的转腕事件；且第一设备的抬腕事件、第一设备的抬手事件和第一设备的转腕事件均是通过同一事件预测模型得到的。

上述事件预测模型不仅可以应用于抬腕事件的检测、转腕事件的检测和抬手事件的检测，还可以应用于其他应用里的事件的检测，比如抬腕即说应用的事件、抬腕亮屏应用的事件，在此不做限定。因此事件预测模型可以看成第一设备的公共能力。通过使用第一设备的公共能力来检测第一事件，对于第一设备来说，不需要再单独训练一个专门用于检测第一事件的神经网络模型，减少了第一设备的工作量，降低了第一设备的算力消耗。并且在检测到第一事件后，进一步判断用户是否做了第一预设动作，若确定用户做了第一预设动作，才会启动第一设备的麦克风，获取麦克风采集的第一音频信号；并根据第一音频信号确定第一音频信号的类型是否为靠近的人声；若确定第一音频信号的类型为靠近的人声，第一设备启动第一设备的语音助手。采用该方式，有利于降低第一设备的语音助手被启动的概率。

在一个可能的实现方式中，第一设备根据第一设备的 IMU 数据确定用户是否做了第一预设动作，包括：

第一设备根据第一设备的 IMU 数据计算得到第一设备的姿态信息；从第一设备的 IMU 数据中获取第一设备的加速度信息；若第一设备的姿态信息处于预设姿态范围、第一设备的加速度信息处于预设加速度范围内，且第一事件时长处于预设时长范围内，则第一设备确定用户做了第一预设动作。

采用该方式，有利于提高判断用户是否做了第一预设动作的精度。

在一个可能的实现方式中，本申请的方法还包括：

在启动第一设备的麦克风之后，持续获取第一设备采集的 IMU 数据和照度数据，根据第一设备采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作；

若确定第一音频信号的类型为靠近的人声，启动第一设备的语音助手，包括：

若确定用户执行了所述第二预设动作，且确定第一音频信号的类型为靠近的人声，启动第一设备的语音助手。

其中，第二预设动作为抬腕保持动作或者抬手保持动作。

通过在启动第一设备的麦克风之后，持续获取第一设备采集的 IMU 数据和照度数据，并基于采集的 IMU 数据和照度数据确定用户是否保持了抬腕动作或者是否保持了抬手动作，在确定用户保持了抬腕动作或者保持了抬手动作，且第一音频信号的类型为靠近的人声时，才会启动第一设备的语音助手，有利于降低误唤醒率和干扰。

在一个可能的实现方式中，本申请的方法还包括：

若确定用户未执行第二预设动作，则第一设备关闭第一设备的麦克风。

在确定用户未执行第二预设动作，也就是用户未保持抬手动作或者未保持抬手动作，表示用户不是准确唤醒语音助手，在这种情况下第一设备关闭麦克风，可以降低第一设备功率，同时也降低了语音助手的误唤醒率。

在一个可能的实现方式中，所述第一音频信号的时长小于或者等于 0.5s，且根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声是所述第一设备的数字信号处理器 DSP 实现的。

通过将第一音频信号的时长设定为小于或者等于 0.5s，是为了后续启动语音助手识别第一语音信号时，可以实时回显第一音频信号对应的文本。并且第一音频信号是未经过 AGC、降噪、去混响或者压缩处理的音频信号，不存在靠近人声信息的丢失，使用未经过 AGC、降噪、去混响或者压缩处理的原始音频信号来确定第一音频信号的声音类型，可以提高确定的第一音频信号的声音类型的精度。

在一个可能的实现方式中，所述第一音频信号是通过所述第一设备的多个麦克风采集得到的。

针对风噪场景，不同麦克风采集的音频信号中存在风噪信号，可以利用一个麦克风采集的音频信号中的风噪对另一麦克风采集的音频信号中的风噪进行降风噪处理，再利用将风噪处理会后的音频信号进行声音类型的预测，可以得到更加准确的预测结果。

在一个可能的实现方式中，本申请的方法还包括：

第一设备根据第一设备的麦克风采集的音频信号确定待执行任务；若待执行任务为敏感任务，第一设备获取用户的身份信息；在基于用户的身份信息确定用户为目标用户，第一设备执行待处理任务。

其中，上述用户为第一设备的使用者，目标用户为第一设备的拥有者。

通过上述方式可以避免非第一设备拥有者通过第一设备执行安全敏感的语音任务，保证了第一设备拥有者的设备的信息安全。

在一个可能的实现方式中，本申请的方法还包括：

在启动语音助手后，第一设备实时显示采集的音频信号对应的文本。

第四方面，本申请提供另一种电子设备，包括：

获取单元，用于在检测到第一事件时，获取第一设备的 IMU 数据；

确定单元，用于根据第一设备的 IMU 数据确定用户是否做了第一预设动作；

启动单元，用于若确定用户做了第一预设动作，启动第一设备的麦克风，获取单元，还用于获取麦克风采集的第一音频信号；

确定单元，还用于根据第一音频信号确定第一音频信号的类型是否为靠近的人声；

启动单元，还用于若确定第一音频信号的类型为靠近的人声，启动所述第一设备的语音助手。

其中，第一事件为第一设备的抬腕事件、第一设备的抬手事件或者第一设备的转腕事件；且第一设备的抬腕事件、第一设备的抬手事件和第一设备的转腕事件均是通过同一事件预测模型得到的。

在一个可能的实现方式中，在根据第一设备的 IMU 数据确定用户是否做了第一预设动作的方面，确定单元，还用于包括：

第一设备根据第一设备的 IMU 数据计算得到第一设备的姿态信息；从第一设备的 IMU 数据中获取第一设备的加速度信息；若第一设备的姿态信息处于预设姿态范围、第一设备的加速度信息处于预设加速度范围内，且第一事件时长处于预设时长范围内，则第一设备确定用户做了第一预设动作。

在一个可能的实现方式中，获取单元，还用于在启动单元启动电子设备的麦克风之后，持续获取电子设备采集的 IMU 数据和照度数据，

确定单元，还用于根据电子设备采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作；

在若确定第一音频信号的类型为靠近的人声，启动电子设备的语音助手的方面，启动单元具体用于：若确定用户执行了第二预设动作，且确定第一音频信号的类型为靠近的人声，启动电子设备的语音助手。

在一个可能的实现方式中，第一音频信号的时长小于或者等于 0.5s，且根据第一音频信号确定第一音频信号的类型是否为靠近的人声是电子设备的 DSP 实现的。

在一个可能的实现方式中，第一音频信号是通过电子设备的多个麦克风采集得到的。

在一个可能的实现方式中，确定单元，还用于根据电子设备的麦克风采集的音频信号确定待执行任务；

获取单元，还用于若待执行任务为敏感任务，获取用户的身份信息；

电子设备还包括：

执行单元，用于在基于用户的身份信息确定用户为目标用户，执行待处理任务。

在一个可能的实现方式中，电子设备还包括：

显示单元，用于在启动语音助手后，实时显示采集的音频信号对应的文本。

第五方面，本申请提供另一种电子设备，包括处理器和存储器。存储器用于存储程序代码。处理器用于调用存储于存储器的程序代码，以执行第一方面、第三方面、第一方面的任一种可能的实施方式提供的方法，或者第三方面的任一种可能的实施方式提供的方法。

第六方面，本申请提供了一种计算机存储介质，包括计算机指令，当所述计算机指令在电子设备上运行时，使得所述电子设备执行如第一方面任一种可能的实施方式提供的方法。

第七方面，本申请提供一种计算机程序产品，当计算机程序产品在计算机上运行时，使得计算机执行如第一方面、第三方面、第一方面的任一种可能的实施方式提供的方法，或者第三方面的任一种可能的实施方式提供的方法。

可以理解地，上述提供的第二方面所述的电子设备、第三方面所述的电子设备、第四方面所述的计算机存储介质或者第五方面所述的计算机程序产品均用于执行第一方面、第三方面、第一方面的任一种可能的实施方式提供的方法，或者第三方面的任一种可能的实施方式提供的方法。因此，其所能达到的有益效果可参考对应方法中的有益效果，此处不再赘述。

附图说明

图 1a 为本申请实施例提供的一种语音交互方法的场景示意图；

图 1b 为本申请实施例提供的另一种语音交互方法的场景示意图；

图 1c 为本申请实施例提供的另一种语音交互方法的场景示意图；

图 1d 为本申请实施例提供的另一种语音交互方法的场景示意图；

图 2 为本申请实施例提供的一种语音交互方法的流程示意图；

图 3 为抬腕过程中照度数据和 IMU 数据的变化示意图；

图 4 为本申请实施例提供的腕戴设备在偏转角范围内所采集的音频信号强度示意图；

图 5 为本申请实施例提供的主麦和副麦所采集的音频信号强度示意图；

图 6 为本申请实施例提供的预设区域示意图；

图 7 为本申请实施例提供的唇高和唇宽示意图；

图 8 为本申请实施例提供的一种电子设备的结构示意图；

图 8a 为本申请实施例提供的另一种电子设备的结构示意图；

图 9 为本申请实施例提供的另一种电子设备的结构示意图。

具体实施方式

本申请的说明书和权利要求书及所述附图中的术语“第一”、“第二”、“第三”和“第四”等是用于区别不同对象，而不是用于描述特定顺序。此外，术语“包括”和“具有”以及它们任何变形，意图在于覆盖不排除他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。

在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

“多个”是指两个或两个以上。“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 这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

首先对本申请涉及的术语进行解释。

智能语音助理 (intelligent personal assistant, IPA)，是由人工智能驱动的个人虚拟助手，也可以称为智能助理、语音助手等。允许用户通过语音交互轻松地操作智能终端，比如智能手机、智能手表、平板电脑、笔记本电脑、AI 音箱、智能大屏和智能座舱等)的各种功能或进行互联网搜索等，比如在智能手机上设闹钟、使用文本转语音技术朗读电子邮件、播放和搜索音乐以及发送短信等。具体的智能个人助理包括华为的小艺 (对应的海外版为 Celia)，苹果的 Siri，亚马逊的 Alexa，微软的 Cortana 和谷歌的 Google Assistant。

腕戴设备 (wrist wearable device)：是指佩戴在用户手腕上的智能设备，一般根据其有无屏幕或屏幕大

小，应用处理器的计算能力以及智能化程度（比如是否可以自行下载和安装 App，是否支持语音助手）高低，可分为智能手表、运动手表、智能手环。

惯性测量单元（IMU，inertial measurement unit）：是测量刚性物体在三维空间的角速度以及加速度的装置。一般用来识别和跟踪设备的姿态和运动，比如设备相对水平面的倾斜角度，设备的方位（即与地磁北极的偏转角度），设备的累积旋转角度、相对运动速度和位移等。

IMU 抬腕硬件中断：IMU 器件内部实现的低功耗初级抬腕检测，一旦检测到抬腕事件就产生硬件中断。

环境光传感器（ambient light sensor，Als）：环境光传感器可以感知周围光线情况，可采集环境的照度数据，并告知处理芯片自动调节屏幕亮度，降低智能终端的功耗。

下面结合附图对本申请的实施例进行描述。

参见图 1a，图 1a 为本申请实施例提供的一种语音交互方法的应用场景示意图。如图 1a 所示：

用户在腕戴设备灭屏的状态下，直接自然移腕靠近嘴唇。用户讲出语音任务，腕戴设备的麦克风采集用户的音频信号，腕戴设备的语音助手基于音频信号识别出语音任务，并针对该语音任务作出响应。

参见图 1b，图 1b 为本申请实施例提供的另一种语音交互方法的应用场景示意图。如图 1b 所示：

用户在腕戴设备亮屏的状态下，比如用户通过腕戴设备的屏幕阅读信息或者用户正在操控腕戴设备，用户抬腕使得用户的嘴唇朝向腕戴设备的屏幕并靠近腕戴设备，然后用户讲出语音任务，比如“今天天气”，腕戴设备采集用户的音频信号，腕戴设备的语音助手基于音频信号识别出语音任务，并针对该语音任务作出响应，比如腕戴设备输出“今天天气晴朗”。

参见图 1c，图 1c 为本申请实施例提供的另一种语音交互方法的应用场景示意图。如图 1c 所示：

用户在手机灭屏的情况下，通过按压手机的电源键点亮手机屏幕，然后从桌面拿起手机，将手机底端靠近嘴边，然后用户讲出语音任务，手机采集用户的音频信号，手机的语音助手基于音频信号识别出语音任务，并针对该语音任务作出响应。

参见图 1d，图 1d 为本申请实施例提供的一种语音交互方法的应用场景示意图。如图 1d 所示：

用户脸朝向大屏设备，可以是正面朝向大屏，也可以是向左偏转预设角度或者向右偏转预设角度朝向大屏，然后用户讲出语音任务，比如“今天天气”，腕戴设备采集用户的音频信号，并基于音频信号识别出语音任务，并针对该语音任务作出响应，比如腕戴设备输出“今天天气晴朗”。

应理解，本申请的语音交互方法不仅可应用于上述四种场景，还可以应用于其他场景，在此不做限定。

参见图 2，图 2 为本申请实施例提供的一种语音交互方法的流程示意图。如图 2 所示，该方法包括：

S201、第一设备检测到第一事件时，获取第一设备的 IMU 数据和照度数据。

应理解，第一设备可以为用户的腕戴设备或者用户终端设备，用户终端设备可以为智能手机、平板电脑等。

其中，第一事件为第一设备的抬腕硬件中断事件、第一设备的抬手硬件中断事件、第一设备的按键亮屏事件、第一设备的抬手亮屏事件或者第一设备的抬腕亮屏事件。

对于第一设备的抬腕硬件中断事件和抬手硬件中断事件，是指 IMU 器内部实现的低功耗初级抬腕检测，一旦检测到较小程度的抬腕或转腕行为（比如，基于第一设备的 IMU 采集的加速度数据，若加速度超过重力加速度（ 9.8m/s^2 ）的 30% 以上，则判定为抬腕或转腕行为或者抬手行为），则第一设备的 IMU 产生抬手硬件中断或者抬腕硬件中断，此中断事件可被其它软件模块订阅，以触发其它软件启动运行，比如，启动抬腕亮屏软件再做一次抬腕或转腕行为检测，以实现准确可靠的抬腕亮屏功能（比如，实现尽可能高的成功率以及尽可能低的误触率）。

对于第一设备的按键亮屏事件，是指第一设备检测到用户针对第一设备的按键（比如电源按键、音量+按键、音量-按键）的操作时，产生按键亮屏事件，此事件可被其它软件模块订阅，以触发其它软件启动运行，比如渲染软件开始界面渲染计算，之后将产生的画面信息提交给屏幕做显示。应理解，产生第一设备的按键亮屏事件是第一设备确定达到了点亮第一设备的屏幕的条件，此时第一设备还并未控制点亮第一设备的屏幕。

对于第一设备的抬腕亮屏事件或者抬手事件，是指第一设备检测到用户试图点亮屏幕的抬腕事件或者抬手事件，抬腕事件或者抬手事件可以是第一设备基于 IMU 采集的 IMU 数据确定的。应理解，产生第一设备的抬腕亮屏事件或者抬手亮屏事件是第一设备确定达到了点亮第一设备的屏幕的条件，此时第一设备

还并未控制点亮第一设备的屏幕。应理解，第一设备检测到用户试图点亮屏幕的抬腕事件或者抬手事件是第一设备的抬腕或者抬手亮屏软件来检测用户试图点亮屏幕的抬腕事件或者抬手事件。

在此需要指出的是，对于上述第一事件，可能是用户误触发或用户出于其它目的而导致第一设备检测到相应的第一事件，比如在用户摆动手臂时，腕戴设备的 IMU 数据也会发生变化，可能会触发抬腕硬件中断事件或者抬腕亮屏事件，但用户不是想抬起手腕与腕戴设备进行语音交互。再比如，在用户拿着手机跑步过程中，手机的 IMU 数据会发生变化，可能会触发抬手中断事件或者抬手亮屏事件。再比如，用户想使用智能手机上的应用程序，用户可能会通过智能手机的电源键点亮智能手机的屏幕，这样会触发手机的抬手亮屏事件，但用户不是想抬手与手机进行语音交互。

为了避免上述误触发，在检测到第一事件时，第一设备获取第一设备的 IMU 数据和照度数据，并通过第一设备的 IMU 数据和照度数据确定用户是否做了第一预设动作。

在一个示例中，当第一事件为第一设备的按键亮屏事件时，第一设备获取第一设备的 IMU 数据和（环境）照度数据，包括第一设备启动环境光传感器采集照度数据和启动 IMU 采集 IMU 数据，并获取采集的照度数据和 IMU 数据，其中，照度数据的采集时长与 IMU 数据的采集时长相同。

应理解，在采集 IMU 数据之前，若 IMU 未启动，则启动 IMU；若 IMU 里面的加速度计已启动而角度计未启动，在需要采集角度数据之前，启动 IMU 数据的角度计。在采集照度数据之前，若环境光传感器未启动，则启动环境光传感器。

在另一个示例中，当第一事件为第一设备的抬腕硬件中断事件、第一设备的抬手硬件中断事件、第一设备的抬腕亮屏事件或第一设备的抬手亮屏事件，第一设备获取第一设备的 IMU 数据和照度数据，包括：

启动第一设备的环境光传感器，采集第一设备的（环境）照度数据和继续采集 IMU 数据，并获取与第一事件相关的缓存的 IMU 数据，并对与第一事件相关的照度数据置零补齐（因第一事件过程中尚未启动环境光传感器），置零是考虑到下部空间一般照度较低。

应理解，检测第一设备的抬腕硬件中断事件、第一设备的抬手硬件中断事件、第一设备的抬腕亮屏事件或第一设备的抬手亮屏事件是利用第一设备的 IMU 数据，也就是说在检测到第一设备的抬腕硬件中断事件、第一设备的抬手硬件中断事件、第一设备的抬腕亮屏事件或第一设备的抬手亮屏事件之前，第一设备的 IMU 已启动并采集了 IMU 数据。因此，确定用户是否做了第一预设动作使用的 IMU 数据可以看成两部分，一部分是在检测到第一事件之前采集的，这部分数据可称为与第一事件相关的 IMU 数据，另一部分是检测到第一事件时开始采集的，两部分的 IMU 数据是一个完整的抬腕或者抬手的 IMU 数据，使用一个完整的抬腕或者抬手的 IMU 数据可以更加准确的判断用户是否执行了第一预设动作。

由于确定用户是否做了第一预设动作使用的照度数据的采集时长需要与 IMU 数据的采集时长相等，因此在使用照度数据和 IMU 数据确定用户是否做了第一预设动作之前，对照度数据进行置零补齐，进行置零补齐得到的数据为与第一事件相关的照度数据。置零补齐后的照度数据的等效时长与 IMU 数据的采集时长相等。其中，IMU 数据的采集时长为第一预设时长。

应理解，传感器在工作一段时间后可以采集一定量的数据，数据量的多少可以通过时长来表征，这里的等效时长就是表征数据量的大小。

S202、第一设备根据第一设备的 IMU 数据和照度数据确定用户是否做了第一预设动作。

其中，第一预设动作可以为抬腕动作或者抬手动作。

应理解，参照图 1a 和图 1b，在用户抬腕将腕戴设备靠近嘴唇的过程中，由于上部空间光线比下部空间的光线强，以及用户嘴唇靠近朝向腕戴设备的屏幕时，屏幕下方的环境光传感器的进光被遮挡，因此环境光传感器采集的照度数据的变化如图 3 左图的曲线所示，环境光传感器采集的照度数值先上升后显著下降，并稳定在一个较低值。参照图 1a 和图 1b，在用户抬腕将腕戴设备靠近嘴唇过程中，腕戴设备随着用户的手腕运动，产生特定的姿态运动（比如表盘与水平面的倾角发生改变以朝向用户嘴唇）和位置运动（比如从下部空间上升到上部空间，与用户身体垂直面的距离缩小），相应的加速度的变化如图 3 右图的曲线所示。同样的，对于手持设备（比如手机），在用户拿起手持设备，抬手以使手持设备靠近用户的嘴唇过程中，手持设备的环境光传感器采集的照度数值也会有上升的趋势，但是在手持设备靠近用户的嘴唇后，手持设备下方的环境光传感器的进光不会被遮挡，因此手持设备下方的环境光传感器采集的照度数值不会出现明显的下降。手持设备随着用户的手部运动，也会产生特定的姿态运动（比如手机屏幕与水平面的倾角变为接近零度，以让手机底端的麦克风朝向用户的嘴唇）和位置运动。

其中，图 3 右图示意出了用户在抬腕将腕戴设备靠近嘴唇的过程中，腕戴设备在本机坐标系下 x 轴方向，y 轴方向和 z 轴方向上的加速度示意图。本机坐标系的原点为腕戴设备的重心，x 轴指向腕戴设备表

盘的3点钟方向，y轴指向表盘的12点钟方向，z轴指向表盘的正上方。

基于上述可知，第一设备在靠近用户的嘴唇过程中会产生特定的姿态运动和位置运动，因此可以基于腕戴设备的照度数据和IMU数据确定用户是否执行了抬腕动作或者抬手动作。

在一个可能的实现方式中，预先训练一个预测模型，用于预测用户是否执行了抬腕动作或者抬手动作。该预测模型可以是基于决策树或者卷积神经网络实现的。可选的，对于腕戴设备和手持设备，可以使用同一个预测模型进行预测，也可以使用不同的预测模型进行预测。

训练数据包括采集时长为固定时长（比如0.5s）的加速度数据和照度数据及其衍生数据，比如加速度数据的最大值、均值、偏度、峰度、直方图、按照一定窗口大小生成的滑动平均值序列、加速度向量的幅度值序列。可选的，训练数据是针对不同群体的用户采集的，包括老年群体和青少年群体、男性群体和女性群体、和/或左手佩戴/持握群体和右手佩戴/持握群体。训练数据中的照度数据也可以包括在不同的环境下采集的数据，包括户外晴天环境、室内正常照明环境、暗光环境。

在一个示例中，第一设备将IMU数据和照度数据输入到第一预测模型中进行预测，可以得到第一预测结果，比如第一预测结果为抬腕动作。第一预测模型可以为四分类网络，第一设备将IMU数据和照度数据输入到第一预测模型中进行预测，第一预测模型输出四个概率，分别为抬腕动作的概率、落腕动作的概率、抬腕保持动作的概率及自由手位的概率；第一预测结果为最大概率对应的动作，比如抬腕动作。其中，这里的自由手位是指除了抬腕动作、落腕动作和抬腕保持动作之外的动作。第一预测模型也可分为两分类网络，第一设备将IMU数据和照度数据输入到第一预测模型中进行预测，第一预测模型输出两个概率，分别为抬腕动作的概率和其他动作的概率，第一预测结果为最大概率对应的动作；这里的其他动作是指处理抬腕动作之外的动作。

应理解，抬腕保持动作是指用户抬起手腕后，手腕在一段时间内保持一定的姿态（比如表盘朝向自己，与水平面的pitch角为正负15度之内，roll角为30~60度）且照度保持一定的稳定（比如照度最大值/最小值的倍数小于2），且没有落下。对于第一设备来说，用户抬起手腕后采集的IMU数据和照度数据在预设范围内，则判定用户执行了抬腕保持动作。

在另一个示例中，第一设备将IMU数据和照度数据输入到第二预测模型中进行预测，可得到第二预测结果，比如第二预测结果为抬手动作。第二预测模型可以是四分类网络，第一设备将IMU数据和照度数据输入到第二预测模型中进行预测，第二预测模型输出四个概率，分别为抬手动作的概率、落臂动作的概率、抬手保持动作的概率和自由手位的概率。第二预测结果为最大概率对应的动作，比如抬手动作。这里的自由手位是除抬手动作、落臂动作和抬手保持动作之外的动作。第二预测模型也可分为两分类网络，第一设备将IMU数据和照度数据输入到第二预测模型中进行预测，第二预测模型输出两个概率，分别为抬手动作的概率和其他动作的概率，第二预测结果为最大概率对应的动作；这里的其他动作是指除抬手动作之外的动作。

其中，上述第一预测模型和第二预测模型的训练数据可以参见前述相关描述，在此不再叙述。

S203、若确定用户做了第一预设动作，第一设备启动第一设备的麦克风，并获取麦克风采集的第一音频信号。

其中，第一音频信号是通过第一设备的多个麦克风采集的，也就是说，若确定了用户做了第一预设动作，第一设备启动第一设备的多个麦克风，比如包括主麦克风和副麦克风。单个麦克风采集的音频信号存在风噪，并且基于单个麦克风采集的音频信号来预测音频信号的声音类型，准确率低，通过第一设备的多个麦克风采集第一音频信号，再利用多个麦克风采集的音频信号来预测音频信号的声音类型，可以得到准确率高的预测结果。

对于多麦克风采集的音频信号，主麦采集的音频信号和副麦采集的音频信号可能都存在风噪，但两者采集的音频信号中风噪强度不一样，第一设备可以利用其中一个音频信号的风噪对另一音频信号中的风噪进行将风噪处理，比如利用副麦采集的音频信号中风噪对主麦采集的音频信号中风噪进行降风噪处理，再利用将风噪处理会后的音频信号进行声音类型的预测，可以得到更加准确的预测结果。

第一设备的麦克风采集到第一音频信号后，第一音频信号会被传输至第一设备的DSP。DSP在硬件实现基础的语音激活检测（voice activity detection, VAD）；当第一音频信号的能量超过预设能量阈值时，第一设备的DSP产生硬件VAD中断，触发DSP中的软件算法运行，也就是确定第一音频信号的类型是否为靠近的人声。

S204、第一设备根据第一音频信号确定第一音频信号的类型是否为靠近的人声。

在一个示例中，第一设备将第一音频信号输入到第三预测模型中进行预测，可以得到第三预测结果，

比如第三预测结果为靠近的人声。第三预测模型可以为四分类网络，第一设备将第一音频信号输入到第一预测模型中进行预测，第三预测模型输出四个概率，分别为靠近的人声的概率、附近的人声的概率、远处人声的概率及非人声的概率；第三预测结果为最大概率对应的声音类型，比如靠近的人声。第三预测模型也可分为两分类网络，第一设备将第一音频信号输入到第三预测模型中进行预测，第三预测模型输出两个概率，分别为靠近的人声的概率和非靠近人声的概率，第三预测结果为最大概率对应的声音类型。

具体的，第一设备通过第三预测模型对第一音频信号进行特征提取，得到第一音频信号的音频特征。可选的，音频特征包括能量特征、时域特征、频域特征、乐理特征和感知特征中的至少一项。

其中，提取音频特征的策略是基于设备场景和用户场景设定的，具体到用户抬手将腕戴设备靠近嘴边说出语音任务的场景：考虑到人声离开口腔后的高低频分布特点（即在嘴唇正对方向的偏转方位上的音频信号能量衰减，如图4所示，用户脑后方音频信号的能量衰减最大，其中，频率越高衰减越大；而嘴唇靠近腕戴设备时，比如距离腕戴设备的屏幕3-5cm时，腕戴设备的麦克风的偏转角度为15-55度。）嘴唇靠近屏幕产生特定的反射混响特征，以及腕戴设备的麦克风位于腕戴设备的边缘位置（比如，11点方向）。由于一般靠近讲话时会压低声音，具体的预设策略可以为高低频比例（high low band ratio, HLBR）特征，比如以（线性频率或者梅尔刻度频率）1kHz为界的高低频能量比，128点快速傅里叶变换（fast Fourier transform, FFT）三阶多项式拟合的4个系数和线性回归拟合的2个系数、混响特征（比如，自相关最大峰与正负10ms之内其他峰的平均值的比值、自相关的最大峰与次的9个峰的平均值的比值、自相关标准差和曲线下面积、绝对一阶导自相关标准差和曲线下面积、语音混响调制能量比（speech-to-reverberation modulation energy ratio, SRMR）、低语特征（比如音高和音色等）。在DSP运行内容和计算资源允许的条件下，具体的预设策略可以为可以包括受人耳听觉认知启发改变量化特征尺度得到的特征，比如84个梅尔频率倒谱系数（Mel-Frequency Cepstral Coefficients, MFCCs）。

对于用户抬手将手持设备靠近嘴边说出语音任务的场景，手持设备靠近用户嘴边3-5cm时存在明显的喷麦现象，如图5所示，对于喷麦现象，手持设备的主麦克风可以采集到有较强的低频气息风噪；而且对于运动风噪或者环境风噪，主麦克风和副麦克风可以采集到强的低频风噪。对于这类情况，具体的预设策略可以为主副麦的HLBR特征，比如以（线性频率或者梅尔刻度频率）200Hz为界的高低频能量比，128点FFT的三级多项式拟合的4个系数、线性回归拟合的2个系数，或者双通道MFCCs。

第一设备根据预设策略提取采集的音频信号的音频特征集数值，具体为：按照预设策略对采集的音频数据进行计算，生成各个音频特征的具体数据。比如MFCCs数值的计算方法：

对采集的音频信号进行分帧，一般25ms为一帧，帧交叠为10ms；采用周期图法对每帧音频信号进行功率谱估计；对估计得到的功率谱采用85个梅尔滤波器进行滤波，计算每个滤波器里的能量；对每个滤波器里的能量取log对数，得到85个对数结果；对85个对数结果进行离散余弦变换（discrete cosine transform, DCT），最后保留DCT的第2-85个系数，去掉第1个系数，最后得到84个系数，该84个系数即为84个梅尔频率倒谱系数。

在一个可能的实施例中，第三预测模型可以按照如下方法得到：

获取腕戴设备或者手持设备采集的一定数量的嘴唇靠近腕戴设备或者手持设备时用户发出的音频信号，及腕戴设备或者手持设备附近或者远处的人声对应的音频信号，然后按照上述预设策略提取采集的音频信号的音频特征集数值，再将音频特征集数值作为训练样本，对机器学习模型或者卷积神经网络进行训练，以得到第三预测模型。

其中，机器学习模型可以为决策树、随机森林算法、Xgboost或者AdaBoost。

其中，第一音频信号的时长小于或者等于0.5s。比如第一音频信号的时长可以为0.25s、0.3s、0.4s和0.5s。

通过将第一音频信号的时长设定为小于或者等于0.5s，是为了后续启动语音助手识别第一语音信号时，可以实时回显第一音频信号对应的文本。

其中，根据第一音频信号确定第一音频信号的类型是否为靠近的人声是在第一设备的DSP实现的。并且第一音频信号是未经过AGC、降噪、去混响或者压缩处理的音频信号，不存在靠近人声信息的丢失，使用未经过AGC、降噪、去混响或者压缩处理的原始音频信号来确定第一音频信号的声音类型，可以提高确定的第一音频信号的声音类型的精度。

应理解，一个完整语音任务对应的音频信号不仅仅包括第一音频信号，还包括第一音频信号之后的音频信号。麦克风采集音频信号是按照小于或者等于0.5s的采集时长周期性采集的。

S205、若确定第一音频信号的类型为靠近的人声，启动第一设备的语音助手。

应理解，在确定第一音频信号的类型后，第一设备的麦克风持续采集音频信号，并提取采集的音频信号的音频特征集数值，按照预设时间间隔基于采集的音频信号的音频特征集数值确定采集的音频信号的类型。其中，预设时间间隔可以为 0.1s、0.2s 或者其他时间。

在一个可行的实现方式中，第一设备在开启第一设备的麦克风后，控制第一设备的 IMU 和环境光传感器继续采集 IMU 数据和照度数据。可选的，采集时长为第二预设时长。第一设备根据继续采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作；若确定用户执行了第二预设动作，且第一音频信号的类型为靠近的人声，则第一设备启动语音助手。若确定用户未执行第二预设动作，则第一设备关闭第一设备的麦克风。

其中，第二预设动作可以为抬腕保持动作或者抬手保持动作。在一个示例中，若用户的抬腕动作或者抬手动作保持了第三预设时长，第一设备启动语音助手。可选的，第三预设时长可以为 0.3s。

在一个示例中，在判断用户是否保持了抬腕动作时，第一设备将继续采集的 IMU 数据和照度数据输入到第四预测模型中进行预测，可以得到第四预测结果，比如，第四预测结果为抬腕保持动作。第四预测模型可以为四分类网络，第一设备将继续采集的 IMU 数据和照度数据输入到第四预测模型中进行预测，第四预测模型输出四个概率，分别为抬腕动作的概率、落腕动作的概率、抬腕保持动作的概率及自由手位的概率；第四预测结果为最大概率对应的动作，比如抬腕保持动作。其中，这里的自由手位是指除了抬腕动作、落腕动作和抬腕保持动作之外的动作。第四预测模型也可为两分类网络，第一设备将继续采集的 IMU 数据和照度数据输入到第四预测模型中进行预测，第四预测模型输出两个概率，分别为抬腕保持动作的概率和其他动作的概率，第四预测结果为最大概率对应的动作；这里的其他动作是指除了抬腕保持动作之外的动作。

在另一个示例中，在判断用户是否保持了抬手动作时，第一设备将继续采集的 IMU 数据和照度数据输入到第五预测模型中进行预测，可以得到第五预测结果，比如，第五预测结果为抬手保持动作。第五预测模型可以为四分类网络，第一设备将 IMU 数据和照度数据输入到第五预测模型中进行预测，第五预测模型输出四个概率，分别为抬手动作的概率、落臂动作的概率、抬手保持动作的概率及自由手位的概率；第五预测结果为最大概率对应的动作，比如抬手保持动作。其中，这里的自由手位是指除了抬手动作、落臂动作和抬手保持动作之外的动作。第五预测模型也可为两分类网络，第一设备将继续采集的 IMU 数据和照度数据输入到第五预测模型中进行预测，第五预测模型输出两个概率，分别为抬手保持动作的概率和其他动作的概率，第五预测结果为最大概率对应的动作；这里的其他动作是指除了抬手保持动作之外的动作。

在此需要指出的是，第四预测模型与第一预测模型可以为同一个预测模型。当第四预测模型与第一预测模型为同一个预测模型时，用于确定用户是否执行了第二预设动作的 IMU 数据和照度数据的采集时长需要与用于确定用户是否执行了第一预设动作的 IMU 数据和照度数据的采集时长相等，也即是需要第二预设时长与第一预设时长相等。当第二预设时长小于第一预设时长时，在使用继续采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作之前，第一设备分别对继续采集的 IMU 数据和照度数据进行插值处理，以使得用于确定用户是否执行了第二预设动作的 IMU 数据和照度数据的等效时长与用于确定用户是否执行了第一预设动作的 IMU 数据和照度数据的采集时长相等，进而可以使用同一个预测模型进行预测。

同理，第五预测模型与第二预测模型可以为同一个预测模型。当第五预测模型与第二预测模型为同一个预测模型时，第二预设时长小于第一预设时长时，在使用继续采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作之前，第一设备分别对继续采集的 IMU 数据和照度数据进行插值处理，以使得用于确定用户是否执行了第二预设动作的 IMU 数据和照度数据的等效时长与用于确定用户是否执行了第一预设动作的 IMU 数据和照度数据的采集时长相等，进而可以使用同一个预测模型进行预测。

应理解，传感器在工作一段时间后可以采集一定量的数据，数据量的多少可以通过时长来表征，这里的等效时长就是表征数据量大小。

需要说明的是，利用预测模型预测用户是否执行了第一预设动作或者第二预设动作的过程只是一个示例，不是对本申请的限定，当然可以通过其他方式预测用户是否执行了第一预设动作或者第二预设动作。

在开启语音助手后，第一设备根据第一设备的麦克风采集的音频信号确定待执行任务，并识别待处理任务的类型；若待处理任务为敏感任务，为了保证安全性，需要获取用户的身份信息，比如用户的指纹信息，声纹信息、脸部图像信息等。在根据获取的用户的身份信息确定用户为目标用户时，执行待处理任务。

其中，敏感任务可以为转账任务、打开隐私文件夹任务等。目标用户为第一设备的所有者。

在一个示例中，第一设备的当前使用者对第一设备说“向 xx 转账 50 元”，第一设备的麦克风采集对应的音频信号，根据该音频信号确定待处理任务：向 xx 转账，金额为 50 元。第一设备确定该待处理任务为敏感任务，获取第一设备的当前使用者的声纹信息，将当前使用者的声纹信息与第一设备中预存储的第一设备的所有者的声纹信息进行匹配，若匹配成功，则确定第一设备的使用者为第一设备的所有者，第一设备执行待处理任务：向 xx 转账，金额为 50 元。若匹配失败，则表示第一设备的使用者不是第一设备的所有者，第一设备不执行待处理任务。

在一个可能的实施例中，在第一设备的语音助手识别出第一音频信号对应的任务后，第一设备确定执行该任务之前是否需要身份认证：如需要进行身份认证，第一设备向用户发出提示信息，比如提示信息为“请先通过人脸或者指纹解锁屏幕”。第一设备获取使用者的身份信息，比如人脸信息或者指纹信息；在身份认证通过（比如人脸解锁已成功或者第一设备已处于屏幕解锁状态）后，第一设备执行第一音频信号对应的任务。

在一个可能的实施例中，第一设备获取第一置信度和第二置信度，其中第一置信度用于表征用户保持抬腕动作或者抬手动作的时长；第二置信度用于表征第一音频信号为靠近的人声的程度；根据预设权重对第一置信度和第二置信度进行加权求和，以得到启动语音助手的置信度。若启动语音助手的置信度大于预设置信度，则启动第一设备的语音助手。

可选的，第一设备可以如下方式得到第一置信度和第二置信度：

第一设备根据获取的 IMU 数据计算用户保持抬腕靠近嘴边姿态的时长，或者根据获取的 IMU 数据计算用户保持抬手靠近嘴边姿态的时长，也即是根据获取的 IMU 数据计算用户保持抬腕动作或者抬手动作的时长；再基于得到的时长计算得到第一置信度；其中，得到的时长越长，第一置信度越大。可选的，第一置信度的取值区间为 $[0.6, 1.0]$ ，也即是说得到的时长的最低值对应的第一置信度为 0.6，得到的时长的最高值对应的第一置信度为 1.0；基于此可以确定得到的时长与对应的第一置信度之间的线性关系，在得到时长后，第一设备可以基于该线性关系计算确定得到时长对应的第一置信度。

对于第二置信度，在前述步骤 S204 中，第一设备将第三预测模型输出的靠近的人声的概率作为第二置信度。

在一个可能的实施例中，在启动语音助手后，第一设备基于自动语音识别 (automatic speech recognition, ASR) 技术识别第一音频信号对应的语音内容，并实时显示第一音频信号对应的语音内容。比如第一音频信号对应的语音内容为“明天天气怎么样”，在识别出“明天”，就显示“明天”；在识别出“天气”，就显示“天气”，而不是在识别到完整的语音内容后再显示。可选的，第一设备将第一音频信号发送至服务器，服务器基于 ASR 技术识别第一音频信号对应的语音内容，并对语音内容进行过滤，过滤一些不是对语音助手讲的内容，比如“今天下班打球啊”，也就是第一设备的语音助手对不是对语音助手讲的内容不进行反馈，语音助手终止语音会话。采用这种方式，可以减少用户的干扰。

对于过滤规则，可以预先设定一些关键词或者句子；第一设备确定第一音频信号对应的语音内容包含设定的关键词或者句子时，则确定第一音频信号对应的语音内容是需要过滤掉的内容，语音助手不需要对第一音频信号对应的语音内容作出反馈。

应理解，第一设备为如图 1a 和图 1b 中所示的腕戴设备，或者为如图 1c 中所示的手持设备。

可以看出，通过 IMU 数据和照度数据可以更加准确的检测用户靠近第一设备的事件，而不是仅使用加速度数据时需要用户显著抬腕/抬手才能检测到用户靠近设备的事件，本申请的方案更加的灵敏，并且在自然抬腕或者抬手时即可实现检测，给用户提供了自然的语音交互体验，并且动作幅度小，在公共场合改善了用户隐私，操作轻松。通过将第一音频信号的时长设定为小于或者等于 0.5s，是为了后续启动语音助手识别第一语音信号时，可以实时回显第一音频信号对应的文本。并且第一音频信号是未经过 AGC、降噪、去混响或者压缩处理的音频信号，不存在靠近人声信息的丢失，可以最大限度利用第一音频信号中的信息，可以提高确定的第一音频信号的声音类型的精度。在音频信号对应的任务是敏感任务之前时，需要进行身份认证，通过该方式可以避免非第一设备拥有者通过第一设备执行安全敏感的语音任务，保证了第一设备拥有者的设备的信息安全。

在另一个具体的实施例中，第一设备实时采集第一设备的 IMU 数据。应理解，这里的 IMU 数据包括加速度、角速度等数据。第一设备根据 IMU 数据确定是否触发第一事件，也即是第一设备是否检测到第一事件，若第一设备根据 IMU 数据确定触发第一事件，则表示第一设备检测到第一事件；若第一设备根据 IMU 数据确定未触发第一事件，则表示第一设备未检测到第一事件。

其中，第一事件为第一设备的抬腕事件、第一设备的抬手事件或第一设备的转腕事件。应理解，第一事件当然还可以为除了第一设备的抬腕事件、第一设备的抬手事件和第一设备的转腕事件之外的事件。

第一设备的抬腕事件是指用户抬手将第一设备置于可观看或者可操作的位置和姿态。应理解，这里的第一设备为手持设备。

第一设备的转腕事件是指用户直接向内转腕或者先向外转腕再向内转腕将第一设备置于可观看或者可操作的位置和姿态。应理解，这里的第一设备为腕戴设备。

第一设备的抬腕事件是指用户抬腕将第一设备置于可观看或者可操作的位置和姿态。应理解，这里的第一设备为腕戴设备。

第一设备的位置和姿态可以通过第一设备的 IMU 数据确定，第一设备可以基于第一设备的 IMU 数据确定是否触发第一事件。在一个可能的实施例中，第一设备可以一个事件预测模型来预测第一设备是否触发第一事件。具体的，第一设备将采集的 IMU 数据输入到事件预测模型中进行处理，事件预测模型输出四个概率值。该四个概率值分别用于表征触发第一设备的抬腕事件的概率、触发第一设备的抬手事件的概率、触发第一设备的转腕事件的概率及未触发上述事件的概率。其中，最大概率对应结果为事件预测模型的最终输出结果。

在此需要指出的是，上述事件预测模型不仅可以应用于抬腕事件的检测、转腕事件的检测和抬手事件的检测，还可以应用于其他应用里的事件的检测，比如抬腕即说应用的事件、抬腕亮屏应用的事件，在此不做限定。因此事件预测模型可以看成第一设备的公共能力。

通过使用第一设备的公共能力来检测第一事件，对于第一设备来说，不需要再单独训练一个专门用于检测第一事件的神经网络模型，减少了第一设备的工作量，降低了第一设备的算力消耗。

由于第一设备是通过第一设备的公共能力检测第一事件的，在通过第一设备的公共能力检测到第一事件不能表示用户做了第一预设动作，因此第一设备需要进一步判断用户是否做了第一预设动作。

具体的，第一设备根据第一设备的 IMU 数据计算得到第一设备的姿态信息；从第一设备的 IMU 数据中获取第一设备的加速度信息；若第一设备的姿态信息处于预设姿态范围、第一设备的加速度信息处于预设加速度范围内，且第一事件时长处于预设时长范围内，则确定用户做了第一预设动作。

其中，第一事件时长是指检测到第一事件后，第一设备保持将第一设备置于可观看或者可操作的位置和姿态的时长。预设姿态范围、预设加速度范围和预设时长范围均是基于历史经验值得到的。

在判断用户是否做了第一预设动作之后第一设备所做的操作可参见 S203-S205 的相关描述，在此不再描述。

在另一个具体的实施例中，第二设备获取用户的第二音频信号，第二设备根据第二音频信号确定用户相对于第二设备的位置信息；若根据用户相对于第二设备的位置信息确定用户是否位于预设范围内；若确定用户位于预设范围内，则第二设备确定检测到用户靠近第二设备，第二设备启动靠近人声检测。第二设备通过靠近人声检测确认第三音频信号的类型为靠近的人声时，第二设备启动语音助手。其中，第三音频信号的采集时间在第二音频信号的采集时间之后。

在一个示例中，第二设备为大屏设备，用户在大屏设备附近发声时，大屏设备总的多个麦克风（比如 3 个以上的麦克风）均能收到用户的音频数据，第二设备基于音频的到达时间差（time difference of arrival, TDOA）技术对多个麦克风采集的音频信号进行处理，以得到用户相对于第二设备的位置信息。第二设备根据用户相对于第二设备的位置信息确定用户是否位于预设范围内。示例性的，预设区域如图 6 所示，预设区域为大屏设备前 5 米的倒梯形区域，梯形下部左右角度均为 120°。在确定用户处于预设区域内时，第二设备确定启动靠近人声检测。

在一个可能的实施例中，第二设备在启动靠近人声检测后，第二设备持续检测用户保持靠近第二设备的行为，具体是第二设备持续获取用户的音频信号，并根据获取的音频信号确定用户是否仍处于预设区域内；若根据获取的音频信号确定用户未处于预设区域内，第二设备则终止靠近人声检测。

可选的，为了提高语音任务的响应速度，第二设备使用采集时长较短的音频信号确定用户是否处于预设区域内。其中，采集时长较短是指采集时长不大于预设采集时长，预设采集时长可以为 0.1s、0.2s、0.5s 或者其他时长。

可选的，靠近人声检测和设备靠近检测可以同步进行。其中，设备靠近检测就是基于获取的音频信号确定用户是否处于预设范围内。

在一个可能的实施例中，靠近人声检测的具体过程包括：

第二设备获取用户对第二设备讲出语音任务时的视频流，并同时获取用户对第二设备讲出语音任务时对应的音频信号，该音频信号可以称为第三音频信号；第二设备根据视频流和第三音频信号确定音频特征信息，该音频特征信息包括语音与唇动的关联系数，和/或语音唇动视频间时延；第二设备根据音频特征信息确定第三音频信号的类型；若确定的第三音频信号的类型为靠近的人声时，第二设备启动语音助手。

其中，语音与唇动的关联系数为视频流中用户的唇高、唇宽分别与第三音频信号的幅度的互相关系数，比如皮尔逊相关系数，取值范围为[0,1]，1表示强相关，0表示不相关。如图7所示， L_1 表示唇宽， L_2 、 L_3 、 L_4 、 L_5 、 L_6 、 L_7 和 L_8 表示唇高。由于语音传播速度慢于光速，语音信号到达时间相比于视频信号的到达时间存在一个时延，因此在计算唇高、唇宽分别与第三音频信号的音频幅度的互相关系数时，需要在时间轴上对第三音频信号进行右移，以达到第三音频信号与视频的同步，当达到同步时，可以选择唇高、唇宽分别与第三音频信号的音频幅度的互相关系数的最大值作为语音与唇动关联性的表征，也就是取唇高、唇宽分别与第三音频信号的音频幅度的互相关系数的最大值作为语音与唇动的关联系数，而在时间轴上右移的长度可作为语音唇动视频间时延。

在一个可能的实施例，第二设备根据音频特征信息确定第三音频信号的类型，可以是第二设备将音频特征信息输入到第六预测模型中进行处理，可以得到第三音频信号的类型。其中，第六预测模型可以为四分类网络，第二设备将音频特征信息输入到第六预测模型中进行预测，第六预测模型输出四个概率，分别为靠近的人声的概率、附近的人声的概率、远处人声的概率及非人声的概率；第三音频信号的类型为最大概率对应的声音类型，比如靠近的人声。第六预测模型也可为两分类网络，第二设备将音频特征信息输入到第六预测模型中进行预测，第六预测模型输出两个概率，分别为靠近的人声的概率和非靠近人声的概率，第三音频信号的类型为最大概率对应的声音类型。

应理解，在使用第六预测模型之前，第二设备会获取第六预测模型，可以是其他设备训练好第六预测模型，从其他设备中获取的，也可以是第二设备自己训练得到的。

在一个实例中，训练得到第六预测模型的过程，具体包括：

获取第二设备采集的一定数量的用户靠近朝向第二设备的屏幕发出的语音任务时的音频信号，和非朝向第二设备的屏幕的人声音频信号/非人声音频信号，以及对应的视频数据；然后按照上述方式获取对应的音频特征信息；然后将得到的音频特征信息作为训练样本，对机器学习模型或者卷积神经网络进行训练，以得到第六预测模型。

其中，机器学习模型可以为决策树、随机森林算法、Xgboost 或者 AdaBoost。

在一个可能的实施例中，在确定第三音频信号的类型为靠近的人声，且用户靠近第二设备并保持了预设时长，比如 0.3s 时，启动语音助手。通过这种方式，可以避免用户误触发启动语音助手。

在开启语音助手后，第二设备根据第三音频信号确定待执行任务，并识别待处理任务的类型；若待处理任务为敏感任务，为了保证安全性，需要获取用户的身份信息，比如用户的指纹信息，声纹信息、脸部图像信息等。在根据获取的用户的身份信息确定用户为目标用户时，执行待处理任务。

在一个可能的实施例中，在启动语音助手后，第二设备基于 ASR 技术识别第三音频信号对应的语音内容，并实时显示第三音频信号对应的语音内容。比如第三音频信号对应的语音内容为“明天天气怎么样”，在识别出“明天”，就显示“明天”；在识别出“天气”，就显示“天气”，而不是在识别到完整的语音内容后再显示。可选的，第二设备将第三音频信号发送至服务器，服务器基于 ASR 技术识别第三音频信号对应的语音内容，并对语音内容进行过滤，过滤一些不是对语音助手讲的内容，比如“今天下班打球啊”，也就是第二设备的语音助手对不是对语音助手讲的内容不进行反馈，语音助手终止语音会话。采用这种方式，可以减少用户的干扰。

对于过滤规则，可以预先设定一些关键词或者句子；第二设备确定第三音频信号对应的语音内容包含设定的关键词或者句子时，则确定第三音频信号对应的语音内容是需要过滤掉的内容，语音助手不需要对第三音频信号对应的语音内容作出反馈。

可以看出，通过使用大屏设备的多麦克风检测声源的位置实现的用户靠近设备检测，而不是唤醒词来触发麦克风拾取用户的语音任务，给用户提供了自然的语音交互（靠近朝向即说）的体验，操作轻松。通过基于大屏设备的摄像头获取唇动信息，然后计算语音唇动关联性，可以拦截设备附近非朝向大屏的人声和非人声，抑制误唤醒语音助手的用户干扰。在音频信号对应的任务是敏感任务之前时，需要进行身份认证，通过该方式可以避免非第二设备拥有者通过第二设备执行安全敏感的语音任务，保证了第二设备拥有者的设备的信息安全。

参照图 8 所示, 为本申请实施例提供的一种电子设备的结构示意图。如图 8 所示, 该电子设备 800 包括:

获取单元 801, 用于在检测到第一事件时, 获取电子设备的 IMU 数据和照度数据;

确定单元 802, 用于根据电子设备的 IMU 数据和照度数据确定用户是否做了第一预设动作;

启动单元 803, 用于若确定用户做了所述第一预设动作, 启动电子设备的麦克风, 获取单元, 还用于获取麦克风采集的第一音频信号;

确定单元 802, 还用于根据第一音频信号确定第一音频信号的类型是否为靠近的人声;

启动单元 803, 还用于若确定第一音频信号的类型为靠近的人声, 启动电子设备的语音助手。

在一个可能的实现方式中, 第一事件为抬腕硬件中断事件、所述电子设备的抬手硬件中断事件、电子设备的按键亮屏事件、所述电子设备的抬手亮屏事件或者所述电子设备的抬腕亮屏事件。

在一个可能的实现方式中, 获取单元 801, 还用于在启动单元 803 启动电子设备的麦克风之后, 持续获取电子设备采集的 IMU 数据和照度数据,

确定单元 802, 还用于根据电子设备采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作;

在若确定第一音频信号的类型为靠近的人声, 启动电子设备的语音助手的方面, 启动单元 803 具体用于:

若确定用户执行了第二预设动作, 且确定第一音频信号的类型为靠近的人声, 启动电子设备的语音助手。

在一个可能的实现方式中, 启动单元 803 还用于:

若确定用户未执行第二预设动作, 则关闭电子设备的麦克风。

在一个可能的实现方式中, 确定用户是否执行了第一预设动作所使用的 IMU 数据和照度数据的采集时长为第一预设时长, 确定用户是否执行了第二预设动作所使用的 IMU 数据和照度数据的采集时长为第二预设时长, 其中, 第一预设时长大于第二预设时长。

在一个可能的实现方式中, 第一事件为抬腕硬件中断事件、电子设备的抬手硬件中断事件、电子设备的抬手亮屏事件或者电子设备的抬腕亮屏事件, 在获取电子设备的 IMU 数据和照度数据的方面, 获取单元 801 具体用于:

启动电子设备的环境光传感器, 采集电子设备的环境的照度数据和继续采集 IMU 数据, 并获取与第一事件相关的缓存的 IMU 数据, 并对与第一事件相关的照度数据置零补齐。

在一个可能的实现方式中, 第一音频信号的时长小于或者等于 0.5s, 且根据第一音频信号确定第一音频信号的类型是否为靠近的人声是电子设备的数字信号处理器(digital signal processor, DSP)实现的。

在一个可能的实现方式中, 第一音频信号是通过电子设备的多个麦克风采集得到的。

在一个可能的实现方式中, 确定单元 802, 还用于根据电子设备的麦克风采集的音频信号确定待执行任务;

获取单元 801, 还用于若待执行任务为敏感任务, 获取用户的身份信息;

电子设备 800 还包括:

执行单元 804, 用于在基于用户的身份信息确定用户为目标用户, 执行待处理任务。

在一个可能的实现方式中, 电子设备 800 还包括:

显示单元 805, 用于在启动语音助手后, 实时显示采集的音频信号对应的文本。

值得指出的是, 其中, 电子设备的具体功能实现方式可以参见上述语音交互方法的描述, 比如获取单元 801 用于执行 S201 的相关内容, 确定单元 802 用于执行 S202 和 S204 的相关内容、启动单元 803、执行单元 804 和显示单元用于执行 S203 和 S205 的相关内容, 这里不再进行赘述。电子设备中的各个单元或模块可以分别或全部合并为一个或若干个另外的单元或模块来构成, 或者其中的某个(些)单元或模块还可以再拆分为功能上更小的多个单元或模块来构成, 这可以实现同样的操作, 而不影响本发明的实施例的技术效果的实现。上述单元或模块是基于逻辑功能划分的, 在实际应用中, 一个单元(或模块)的功能也可以由多个单元(或模块)来实现, 或者多个单元(或模块)的功能由一个单元(或模块)实现。

参照图 8a 所示, 为本申请实施例提供的一种电子设备的结构示意图。如图 8a 所示, 该自电子设备 800a 包括:

获取单元 801a, 用于在检测到第一事件时, 获取第一设备的 IMU 数据;

确定单元 802a, 用于根据第一设备的 IMU 数据确定用户是否做了第一预设动作;

启动单元 803a, 用于若确定用户做了第一预设动作, 启动第一设备的麦克风, 获取单元, 还用于获取麦克风采集的第一音频信号;

确定单元 802a, 还用于根据第一音频信号确定第一音频信号的类型是否为靠近的人声;

启动单元 803a, 还用于若确定第一音频信号的类型为靠近的人声, 启动所述第一设备的语音助手。

其中, 第一事件为第一设备的抬腕事件、第一设备的抬手事件或者第一设备的转腕事件; 且第一设备的抬腕事件、第一设备的抬手事件和第一设备的转腕事件均是通过同一事件预测模型得到的。

在一个可能的实现方式中, 在根据第一设备的 IMU 数据确定用户是否做了第一预设动作的方面, 确定单元 802a, 还用于包括:

第一设备根据第一设备的 IMU 数据计算得到第一设备的姿态信息; 从第一设备的 IMU 数据中获取第一设备的加速度信息; 若第一设备的姿态信息处于预设姿态范围、第一设备的加速度信息处于预设加速度范围内, 且第一事件时长处于预设时长范围内, 则第一设备确定用户做了第一预设动作。

在一个可能的实现方式中, 获取单元 801a, 还用于在启动单元启动电子设备的麦克风之后, 持续获取电子设备采集的 IMU 数据和照度数据;

确定单元 802a, 还用于根据电子设备采集的 IMU 数据和照度数据确定用户是否执行了第二预设动作;

在若确定第一音频信号的类型为靠近的人声, 启动电子设备的语音助手的方面, 启动单元 803a 具体用于:

若确定用户执行了第二预设动作, 且确定第一音频信号的类型为靠近的人声, 启动电子设备的语音助手。

在一个可能的实现方式中, 第一音频信号的时长小于或者等于 0.5s, 且根据第一音频信号确定第一音频信号的类型是否为靠近的人声是电子设备的 DSP 实现的。

在一个可能的实现方式中, 第一音频信号是通过电子设备的多个麦克风采集得到的。

在一个可能的实现方式中, 确定单元 802a, 还用于根据电子设备的麦克风采集的音频信号确定待执行任务;

获取单元 801a, 还用于若待执行任务为敏感任务, 获取用户的身份信息;

电子设备 800a 还包括:

执行单元 804a, 用于在基于用户的身份信息确定用户为目标用户, 执行待处理任务。

在一个可能的实现方式中, 电子设备 800a 还包括:

显示单元 805a, 用于在启动语音助手后, 实时显示采集的音频信号对应的文本。

值得指出的是, 其中, 电子设备的具体功能实现方式可以参见上述语音交互方法的描述。电子设备中的各个单元或模块可以分别或全部合并为一个或若干个另外的单元或模块来构成, 或者其中的某个(些)单元或模块还可以再拆分为功能上更小的多个单元或模块来构成, 这可以实现同样的操作, 而不影响本发明的实施例的技术效果的实现。上述单元或模块是基于逻辑功能划分的, 在实际应用中, 一个单元(或模块)的功能也可以由多个单元(或模块)来实现, 或者多个单元(或模块)的功能由一个单元(或模块)实现。

基于上述方法实施例以及装置实施例的描述, 请参见图 9, 本发明实施例还提供的一种电子设备 900 的结构示意图。图 9 所示的电子设备 900 (该电子设备 900 具体可以是一种计算机设备) 包括存储器 901、处理器 902、通信接口 903、显示屏 905 以及总线 904。其中, 存储器 901、处理器 902、通信接口 903、显示屏 905 通过总线 904 实现彼此之间的通信连接。

存储器 901 可以是只读存储器 (Read Only Memory, ROM), 静态存储设备, 动态存储设备或者随机存取存储器 (Random Access Memory, RAM)。

存储器 901 可以存储程序, 当存储器 901 中存储的程序被处理器 902 执行时, 处理器 902 和通信接口 903 用于执行本申请实施例的语音交互方法的各个步骤。

处理器 902 可以采用通用的中央处理器 (Central Processing Unit, CPU), 微处理器, 应用专用集成电路 (Application Specific Integrated Circuit, ASIC), 图形处理器 (graphics processing unit, GPU) 或者一个或多个集成电路, 用于执行相关程序, 以实现本申请实施例的电子设备 900 中的单元所需执行的功能, 或者执行本申请方法实施例的语音交互方法。

处理器 902 还可以是一种集成电路芯片, 具有信号的处理能力。在实现过程中, 本申请的语音交互方法的各个步骤可以通过处理器 902 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 902

还可以是通用处理器、数字信号处理器 (Digital Signal Processing, DSP)、专用集成电路 (ASIC)、现成可编程门阵列 (Field Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成, 或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器, 闪存、只读存储器, 可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 901, 处理器 902 读取存储器 901 中的信息, 结合其硬件完成本申请实施例的网络电子设备中包括的单元所需执行的功能, 或者执行本申请方法实施例的语音交互方法。

通信接口 903 使用例如但不限于收发器一类的收发装置, 来实现电子设备 900 与其他设备或通信网络之间的通信。例如, 可以通过通信接口 903 获取数据。

总线 904 可包括在电子设备 900 各个部件 (例如, 存储器 901、处理器 902、通信接口 903) 之间传送信息的通路。

其中, 显示屏 905 用于在启动语音助手后, 显示采集到的音频信号对应的文本。当然还可以显示语音助手针对采集到的音频信号的回复信息的文本。需要指出的是, 显示屏 905 可以为 LCD 屏、LED 屏、OLED 屏, 当然还可以为其他显示屏, 在此不做限定。

应注意, 尽管图 9 所示的电子设备 900 仅仅示出了存储器、处理器、通信接口, 但是在具体实现过程中, 本领域的技术人员应当理解, 电子设备 900 还包括实现正常运行所必须的其他器件, 比如显示器。同时, 根据具体需要, 本领域的技术人员应当理解, 电子设备 900 还可包括实现其他附加功能的硬件器件。此外, 本领域的技术人员应当理解, 电子设备 900 也可仅仅包括实现本申请实施例所必须的器件, 而不必包括图 9 中所示的全部器件。

本申请实施例还提供了一种芯片, 所述芯片包括处理器与数据接口, 所述处理器通过所述数据接口读取存储器上存储的指令, 以实现所述的语音交互方法。

可选地, 作为一种实现方式, 所述芯片还可以包括存储器, 所述存储器中存储有指令, 所述处理器用于执行所述存储器上存储的指令, 当所述指令被执行时, 所述处理器用于执行所述的语音交互方法。

本申请实施例还提供了一种计算机可读存储介质, 该计算机可读存储介质中存储有指令, 当其在计算机或处理器上运行时, 使得计算机或处理器执行上述任一个方法中的一个或多个步骤。

本申请实施例还提供了一种包含指令的计算机程序产品。当该计算机程序产品在计算机或处理器上运行时, 使得计算机或处理器执行上述任一个方法中的一个或多个步骤。

本领域技术人员能够领会, 结合本文公开描述的各种说明性逻辑框、模块和算法步骤所描述的功能可以由硬件、软件、固件或其任何组合来实施。如果以软件来实施, 那么各种说明性逻辑框、模块、和步骤描述的功能可作为一或多个指令或代码在计算机可读媒体上存储或传输, 且由基于硬件的处理单元执行。计算机可读媒体可包含计算机可读存储媒体, 其对应于有形媒体, 例如数据存储媒体, 或包括任何促进将计算机程序从一处传送到另一处的媒体 (例如, 基于通信协议) 的通信媒体。以此方式, 计算机可读媒体大体上可对应于 (1) 非暂时性的有形计算机可读存储媒体, 或 (2) 通信媒体, 例如信号或载波。数据存储媒体可为可由一或多个计算机或一或多个处理器存取以检索用于实施本申请中描述的技术的指令、代码和/或数据结构的任何可用媒体。计算机程序产品可包含计算机可读媒体。

作为实例而非限制, 此类计算机可读存储媒体可包括 RAM、ROM、EEPROM、CD-ROM 或其它光盘存储装置、磁盘存储装置或其它磁性存储装置、快闪存储器或用来存储指令或数据结构形式的所要程序代码并且可由计算机存取的任何其它媒体。并且, 任何连接被恰当地称作计算机可读媒体。举例来说, 如果使用同轴缆线、光纤缆线、双绞线、数字订户线 (DSL) 或例如红外线、无线电和微波等无线技术从网站、服务器或其它远程源传输指令, 那么同轴缆线、光纤缆线、双绞线、DSL 或例如红外线、无线电和微波等无线技术包含在媒体的定义中。但是, 应理解, 所述计算机可读存储媒体和数据存储媒体并不包括连接、载波、信号或其它暂时媒体, 而是实际上针对于非暂时性有形存储媒体。如本文中所使用, 磁盘和光盘包含压缩光盘 (CD)、激光光盘、光学光盘、数字多功能光盘 (DVD) 和蓝光光盘, 其中磁盘通常以磁性方式再现数据, 而光盘利用激光以光学方式再现数据。以上各项的组合也应包含在计算机可读媒体的范围内。

可通过例如一或多个数字信号处理器 (DSP)、通用微处理器、专用集成电路 (ASIC)、现场可编程逻辑阵列 (FPGA) 或其它等效集成或离散逻辑电路等一或多个处理器来执行指令。因此, 如本文中所使用的术语“处理器”可指前述结构或适合于实施本文中所描述的技术的任一其它结构中的任一者。另外, 在一些方面中, 本文中所描述的各种说明性逻辑框、模块、和步骤所描述的功能可以提供于经配置以用于编码和解码

的专用硬件和/或软件模块内，或者并入在组合编解码器中。而且，所述技术可完全实施于一或多个电路或逻辑元件中。

本申请的技术可在各种各样的装置或设备中实施，包含无线手持机、集成电路(IC)或一组 IC(例如，芯片组)。本申请中描述各种组件、模块或单元是为了强调用于执行所揭示的技术的装置的功能方面，但未必需要由不同硬件单元实现。实际上，如上文所描述，各种单元可结合合适的软件和/或固件组合在编码硬件单元中，或者通过互操作硬件单元（包含如上文所描述的一或多个处理器）来提供。

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，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应步骤过程的具体描述，在此不再赘述。

应理解，在本申请的描述中，除非另有说明，“/”表示前后关联的对象是一种“或”的关系，例如，A/B 可以表示 A 或 B；其中 A，B 可以是单数或者复数。并且，在本申请的描述中，除非另有说明，“多个”是指两个或两个以上。“以下至少一项(个)”或其类似表达，是指的这些项中的任意组合，包括单项(个)或复数项(个)的任意组合。例如，a，b，或 c 中的至少一项(个)，可以表示：a，b，c，a-b，a-c，b-c，或 a-b-c，其中 a，b，c 可以是单个，也可以是多个。另外，为了便于清楚描述本申请实施例的技术方案，在本申请的实施例中，采用了“第一”、“第二”等字样对功能和作用基本相同的相同项或相似项进行区分。本领域技术人员可以理解“第一”、“第二”等字样并不对数量和执行次序进行限定，并且“第一”、“第二”等字样也并不限定一定不同。同时，在本申请实施例中，“示例性的”或者“例如”等词用于表示作例子、例证或说明。本申请实施例中被描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言，使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念，便于理解。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，该单元的划分，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系统，或一些特征可以忽略，或不执行。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。

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或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。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。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。该计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行该计算机程序指令时，全部或部分地产生按照本申请实施例的流程或功能。该计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。该计算机指令可以存储在计算机可读存储介质中，或者通过该计算机可读存储介质进行传输。该计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线（digital subscriber line, DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。该计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。该可用介质可以是只读存储器（read-only memory, ROM），或随机存取存储器（random access memory, RAM），或磁性介质，例如，软盘、硬盘、磁带、磁碟、或光介质，例如，数字通用光盘（digital versatile disc, DVD）、或者半导体介质，例如，固态硬盘（solid state disk, SSD）等。

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

权利要求书

1. 一种语音交互方法，应用于第一设备，其特征在于，所述方法包括：
在检测到第一事件时，获取所述第一设备的 IMU 数据和照度数据；
根据所述第一设备的 IMU 数据和照度数据确定用户是否做了第一预设动作；
若确定所述用户做了所述第一预设动作，启动所述第一设备的麦克风，获取所述麦克风采集的第一音频信号；
根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声；
若确定所述第一音频信号的类型为靠近的人声，启动所述第一设备的语音助手。
2. 根据权利要求 1 所述的方法，其特征在于，所述第一事件为抬腕硬件中断事件、所述第一设备的抬手硬件中断事件、所述第一设备的按键亮屏事件、所述第一设备的抬手亮屏事件或者所述第一设备的抬腕亮屏事件。
3. 根据权利要求 1 或 2 所述的方法，其特征在于，所述方法还包括：
在启动所述第一设备的麦克风之后，持续获取第一设备采集的 IMU 数据和照度数据，
根据所述第一设备采集的 IMU 数据和照度数据确定所述用户是否执行了第二预设动作；
所述若确定所述第一音频信号的类型为靠近的人声，启动所述第一设备的语音助手，包括：
若确定所述用户执行了所述第二预设动作，且确定所述第一音频信号的类型为靠近的人声，启动所述第一设备的语音助手。
4. 根据权利要求 3 所述的方法，其特征在于，所述方法还包括：
若确定所述用户未执行所述第二预设动作，则关闭所述第一设备的麦克风。
5. 根据权利要求 3 或 4 所述的方法，其特征在于，
确定用户是否执行了所述第一预设动作所使用的 IMU 数据和照度数据的采集时长为第一预设时长，
确定用户是否执行了所述第二预设动作所使用的 IMU 数据和照度数据的采集时长为第二预设时长，
其中，所述第一预设时长大于所述第二预设时长。
6. 根据权利要求 2 所述的方法，其特征在于，所述第一事件为抬腕硬件中断事件、所述第一设备的抬手硬件中断事件、所述第一设备的抬手亮屏事件或者所述第一设备的抬腕亮屏事件，所述获取所述第一设备的 IMU 数据和照度数据，包括：
启动所述第一设备的环境光传感器，采集所述第一设备的环境的照度数据和继续采集 IMU 数据，并获取与所述第一事件相关的缓存的 IMU 数据，并对与所述第一事件相关的照度数据置零补齐。
7. 根据权利要求 1-6 任一项所述的方法，其特征在于，所述第一音频信号的时长小于或者等于 0.5s，且根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声是所述第一设备的数字信号处理器 DSP 实现的。
8. 根据权利要求 1-7 任一项所述的方法，其特征在于，所述第一音频信号是通过所述第一设备的多个麦克风采集得到的。
9. 根据权利要求 1-8 任一项所述的方法，其特征在于，所述方法还包括：
根据所述第一设备的麦克风采集的音频信号确定待执行任务；
若所述待执行任务为敏感任务，获取所述用户的身份信息；
在基于所述用户的身份信息确定所述用户为目标用户，执行所述待处理任务。
10. 根据权利要求 1-9 任一项所述的方法，其特征在于，所述方法还包括：
在开启语音助手后，所述第一设备实时显示采集的音频信号对应的文本。

11. 一种语音交互方法，应用于第一设备，其特征在于，所述方法包括：

在检测到第一事件时，获取所述第一设备的 IMU 数据；其中，所述第一事件为所述第一设备的抬腕事件、所述第一设备的抬手事件或者所述第一设备的转腕事件；且所述第一设备的抬腕事件、所述第一设备的抬手事件和所述第一设备的转腕事件均是通过同一事件预测模型得到的；

根据所述第一设备的 IMU 数据确定用户是否做了第一预设动作；

若确定所述用户做了所述第一预设动作，启动第一设备的麦克风，获取所述麦克风采集的第一音频信号；

根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声；

若确定所述第一音频信号的类型为靠近的人声，启动所述第一设备的语音助手。

12. 根据权利要求 11 所述的方法，其特征在于，所述根据所述第一设备的 IMU 数据确定用户是否做了第一预设动作，包括：

根据所述第一设备的 IMU 数据计算得到所述第一设备的姿态信息；

从所述第一设备的 IMU 数据中获取所述第一设备的加速度信息；

若所述第一设备的姿态信息处于预设姿态范围、所述第一设备的加速度信息处于预设加速度范围内，且所述第一事件时长处于预设时长范围内，则确定所述用户做了所述第一预设动作。

13. 根据权利要求 11 或 12 所述的方法，其特征在于，所述方法还包括：

在启动所述第一设备的麦克风之后，持续获取第一设备采集的 IMU 数据和照度数据，

根据所述第一设备采集的 IMU 数据和照度数据确定所述用户是否执行了第二预设动作；

所述若确定所述第一音频信号的类型为靠近的人声，启动所述第一设备的语音助手，包括：

若确定所述用户执行了所述第二预设动作，且确定所述第一音频信号的类型为靠近的人声，启动所述第一设备的语音助手。

14. 根据权利要求 13 所述的方法，其特征在于，所述方法还包括：

若确定所述用户未执行所述第二预设动作，则关闭所述第一设备的麦克风。

15. 根据权利要求 11-14 任一项所述的方法，其特征在于，所述第一音频信号的时长小于或者等于 0.5s，且根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声是所述第一设备的数字信号处理器 DSP 实现的。

16. 根据权利要求 11-15 任一项所述的方法，其特征在于，所述第一音频信号是通过所述第一设备的多个麦克风采集得到的。

17. 根据权利要求 11-16 任一项所述的方法，其特征在于，所述方法还包括：

根据所述第一设备的麦克风采集的音频信号确定待执行任务；

若所述待执行任务为敏感任务，获取所述用户的身份信息；

在基于所述用户的身份信息确定所述用户为目标用户，执行所述待处理任务。

18. 根据权利要求 11-17 任一项所述的方法，其特征在于，所述方法还包括：

在开启语音助手后，第一设备实时显示采集的音频信号对应的文本。

19. 一种电子设备，其特征在于，包括：

获取单元，用于在检测到第一事件时，获取所述电子设备的 IMU 数据和照度数据；

确定单元，用于根据所述电子设备的 IMU 数据和照度数据确定用户是否做了第一预设动作；

启动单元，用于若确定所述用户做了所述第一预设动作，启动电子设备的麦克风，所述获取单元，还用于获取所述麦克风采集的第一音频信号；

所述确定单元，还用于根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声；

所述启动单元，还用于若确定所述第一音频信号的类型为靠近的人声，启动所述电子设备的语音助手。

20. 根据权利要求 19 所述的电子设备, 其特征在于, 所述第一事件为抬腕硬件中断事件、所述电子设备的抬手硬件中断事件、所述电子设备的按键亮屏事件、所述电子设备的抬手亮屏事件或者所述电子设备的抬腕亮屏事件。

21. 根据权利要求 19 或 20 所述的电子设备, 其特征在于,

所述获取单元, 还用于在所述启动单元启动所述电子设备的麦克风之后, 持续获取电子设备采集的 IMU 数据和照度数据,

所述确定单元, 还用于根据所述电子设备采集的 IMU 数据和照度数据确定所述用户是否执行了第二预设动作;

在所述若确定所述第一音频信号的类型为靠近的人声, 启动所述电子设备的语音助手的方面, 所述启动单元具体用于:

若确定所述用户执行了所述第二预设动作, 且确定所述第一音频信号的类型为靠近的人声, 启动所述电子设备的语音助手。

22. 根据权利要求 21 所述的电子设备, 其特征在于, 所述启动单元还用于:

若确定所述用户未执行所述第二预设动作, 则关闭所述电子设备的麦克风。

23. 根据权利要求 21 或 22 所述的电子设备, 其特征在于,

确定用户是否执行了所述第一预设动作所使用的 IMU 数据和照度数据的采集时长为第一预设时长,

确定用户是否执行了所述第二预设动作所使用的 IMU 数据和照度数据的采集时长为第二预设时长,

其中, 所述第一预设时长大于所述第二预设时长。

24. 根据权利要求 20 所述的电子设备, 其特征在于, 所述第一事件为抬腕硬件中断事件、所述电子设备的抬手硬件中断事件、所述电子设备的抬手亮屏事件或者所述电子设备的抬腕亮屏事件, 在所述获取所述电子设备的 IMU 数据和照度数据的方面, 所述获取单元具体用于:

启动所述电子设备的环境光传感器, 采集所述电子设备的环境的照度数据和继续采集 IMU 数据, 并获取与所述第一事件相关的缓存的 IMU 数据, 并对与所述第一事件相关的照度数据置零补齐。

25. 根据权利要求 19-24 任一项所述的电子设备, 其特征在于, 所述第一音频信号的时长小于或者等于 0.5s, 且根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声是所述电子设备的数字信号处理器 DSP 实现的。

26. 根据权利要求 19-25 任一项所述的电子设备, 其特征在于, 所述第一音频信号是通过所述电子设备的多个麦克风采集得到的。

27. 根据权利要求 19-26 任一项所述的电子设备, 其特征在于,

所述确定单元, 还用于根据所述电子设备的麦克风采集的音频信号确定待执行任务;

所述获取单元, 还用于若所述待执行任务为敏感任务, 获取所述用户的身份信息;

所述电子设备还包括:

执行单元, 用于在基于所述用户的身份信息确定所述用户为目标用户, 执行所述待处理任务。

28. 根据权利要求 19-27 任一项所述的电子设备, 其特征在于, 所述电子设备还包括:

显示单元, 用于在开启语音助手后, 实时显示采集的音频信号对应的文本。

29. 一种电子设备, 其特征在于, 包括:

获取单元, 用于在检测到第一事件时, 获取所述电子设备的 IMU 数据; 其中, 所述第一事件为所述第一设备的抬腕事件、所述第一设备的抬手事件或者所述第一设备的转腕事件; 且所述第一设备的抬腕事件、所述第一设备的抬手事件和所述第一设备的转腕事件均是通过同一事件预测模型得到的;

确定单元，用于根据所述电子设备的 IMU 数据确定用户是否做了第一预设动作；

启动单元，用于若确定所述用户做了所述第一预设动作，启动电子设备的麦克风，获取所述麦克风采集的第一音频信号；

所述确定单元，还用于根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声；

所述启动单元，还用于若确定所述第一音频信号的类型为靠近的人声，启动所述电子设备的语音助手。

30. 根据权利要求 29 所述的电子设备，其特征在于，在所述根据所述电子设备的 IMU 数据确定用户是否做了第一预设动作的方面，所述确定单元具体用于：

根据所述电子设备的 IMU 数据计算得到所述电子设备的姿态信息；

从所述电子设备的 IMU 数据中获取所述电子设备的加速度信息；

若所述电子设备的姿态信息处于预设姿态范围、所述电子设备的加速度信息处于预设加速度范围内，且所述第一事件时长处于预设时长范围内，则确定所述用户做了所述第一预设动作。

31. 根据权利要求 29 或 30 所述的电子设备，其特征在于，

所述获取单元，还用于在所述启动单元启动所述电子设备的麦克风之后，持续获取电子设备采集的 IMU 数据和照度数据，

所述确定单元，还用于根据所述电子设备采集的 IMU 数据和照度数据确定所述用户是否执行了第二预设动作；

在所述若确定所述第一音频信号的类型为靠近的人声，启动所述电子设备的语音助手的方面，所述启动单元具体用于：

若确定所述用户执行了所述第二预设动作，且确定所述第一音频信号的类型为靠近的人声，启动所述电子设备的语音助手。

32. 根据权利要求 31 所述的电子设备，其特征在于，所述启动单元还用于：

若确定所述用户未执行所述第二预设动作，则关闭所述电子设备的麦克风。

33. 根据权利要求 29-32 任一项所述的电子设备，其特征在于，所述第一音频信号的时长小于或者等于 0.5s，且根据所述第一音频信号确定所述第一音频信号的类型是否为靠近的人声是所述电子设备的数字信号处理器 DSP 实现的。

34. 根据权利要求 29-33 任一项所述的电子设备，其特征在于，所述第一音频信号是通过所述电子设备的多个麦克风采集得到的。

35. 根据权利要求 29-34 任一项所述的电子设备，其特征在于，

所述确定单元，还用于根据所述电子设备的麦克风采集的音频信号确定待执行任务；

所述获取单元，还用于若所述待执行任务为敏感任务，获取所述用户的身份信息；

所述电子设备还包括：

执行单元，用于在基于所述用户的身份信息确定所述用户为目标用户，执行所述待处理任务。

36. 根据权利要求 29-35 任一项所述的电子设备，其特征在于，所述电子设备还包括：

显示单元，用于在开启语音助手后，实时显示采集的音频信号对应的文本。

37. 一种电子设备，其特征在于，包括处理器和存储器，其中，所述存储器用于存储程序代码，所述处理器用于执行所述程序代码，以实现权利要求 1 至 18 任一项所述的方法。

38. 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计算机程序，所述计算机程序被处理器执行时，实现权利要求 1 至 18 任一项所述的方法。

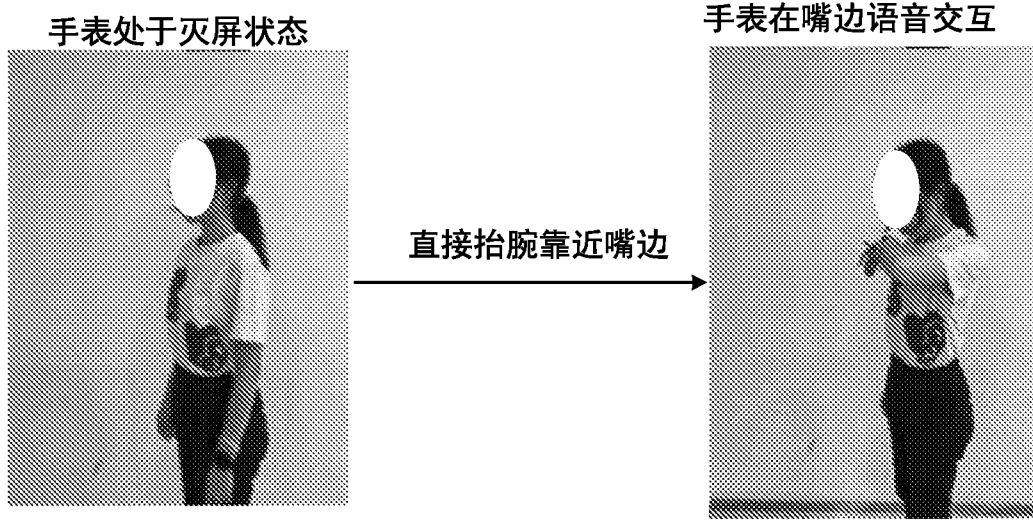


图 1a

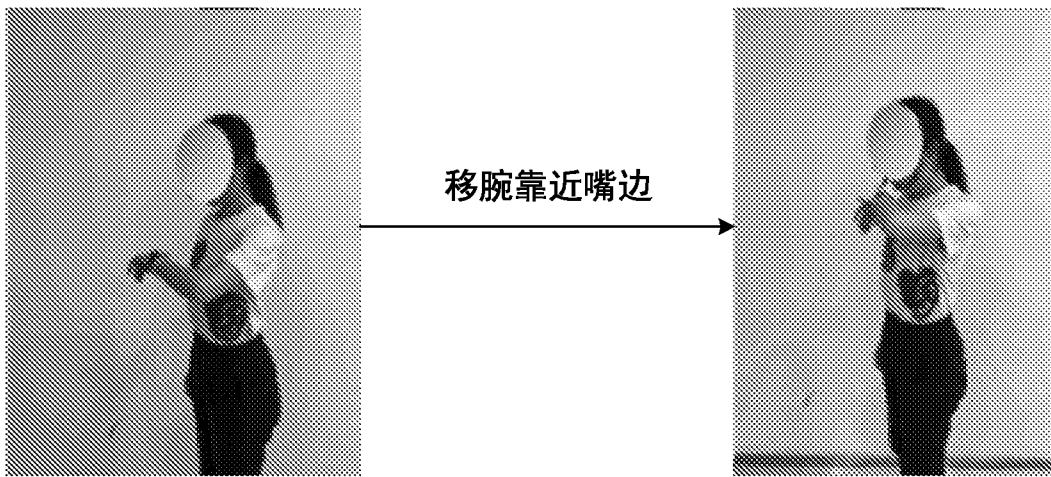


图 1b



图 1c

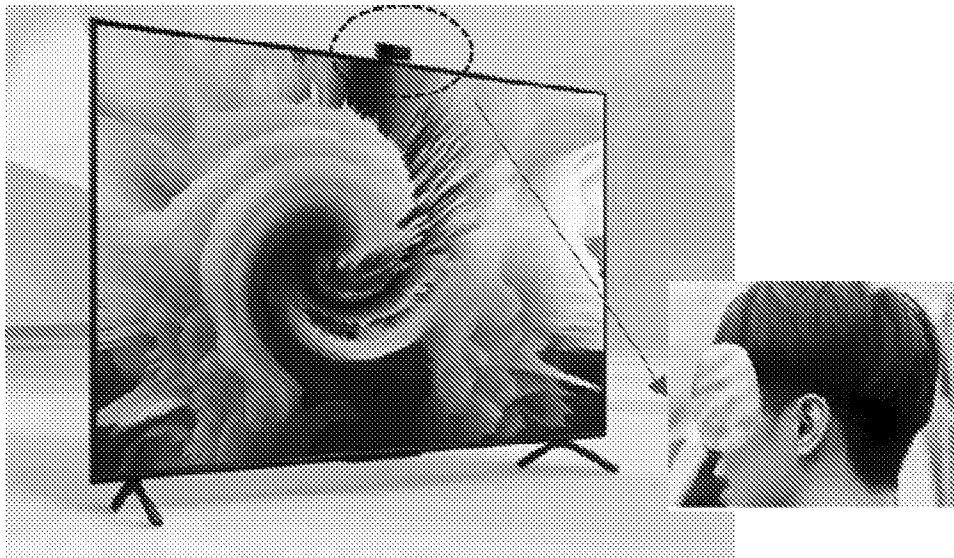


图 1d

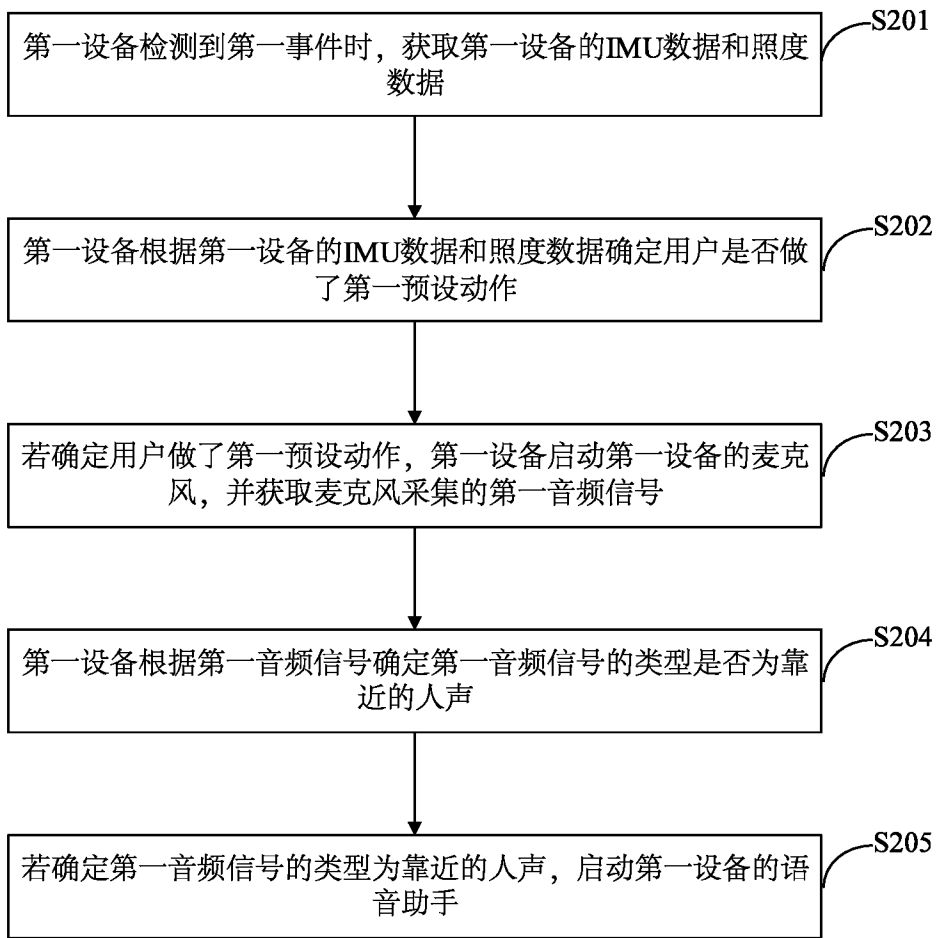


图 2

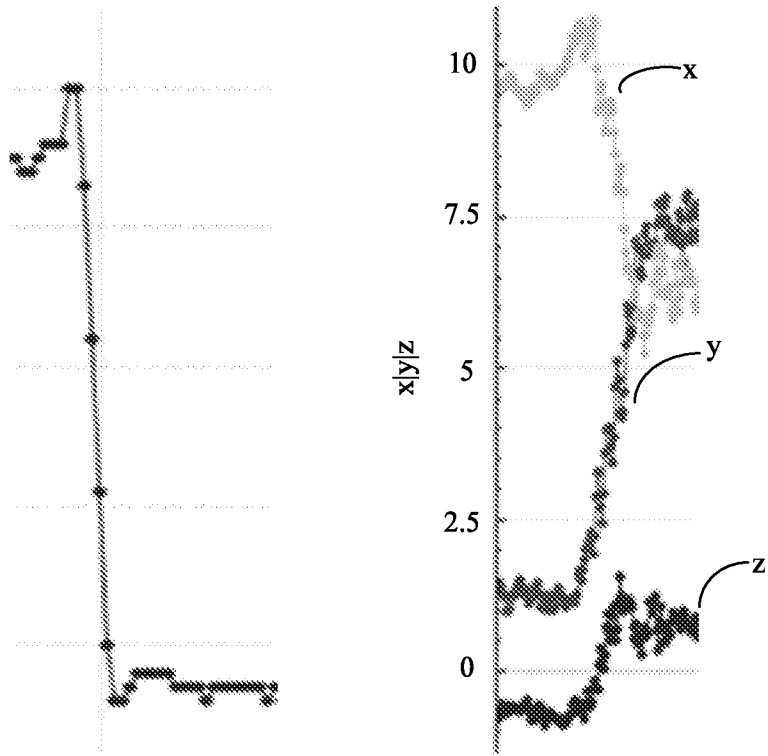


图 3

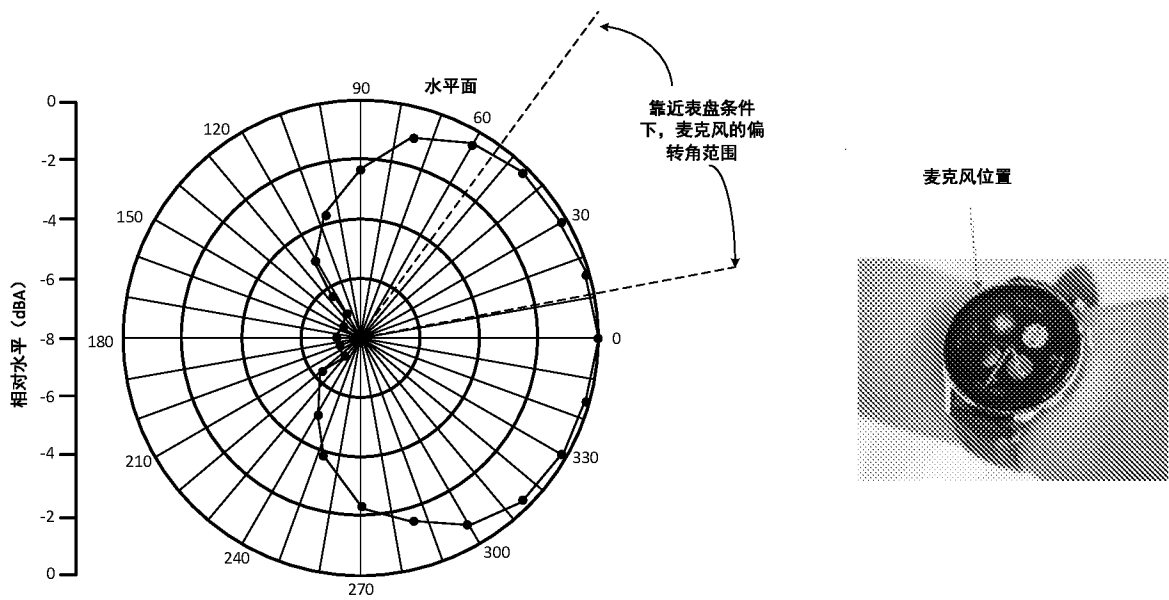


图 4

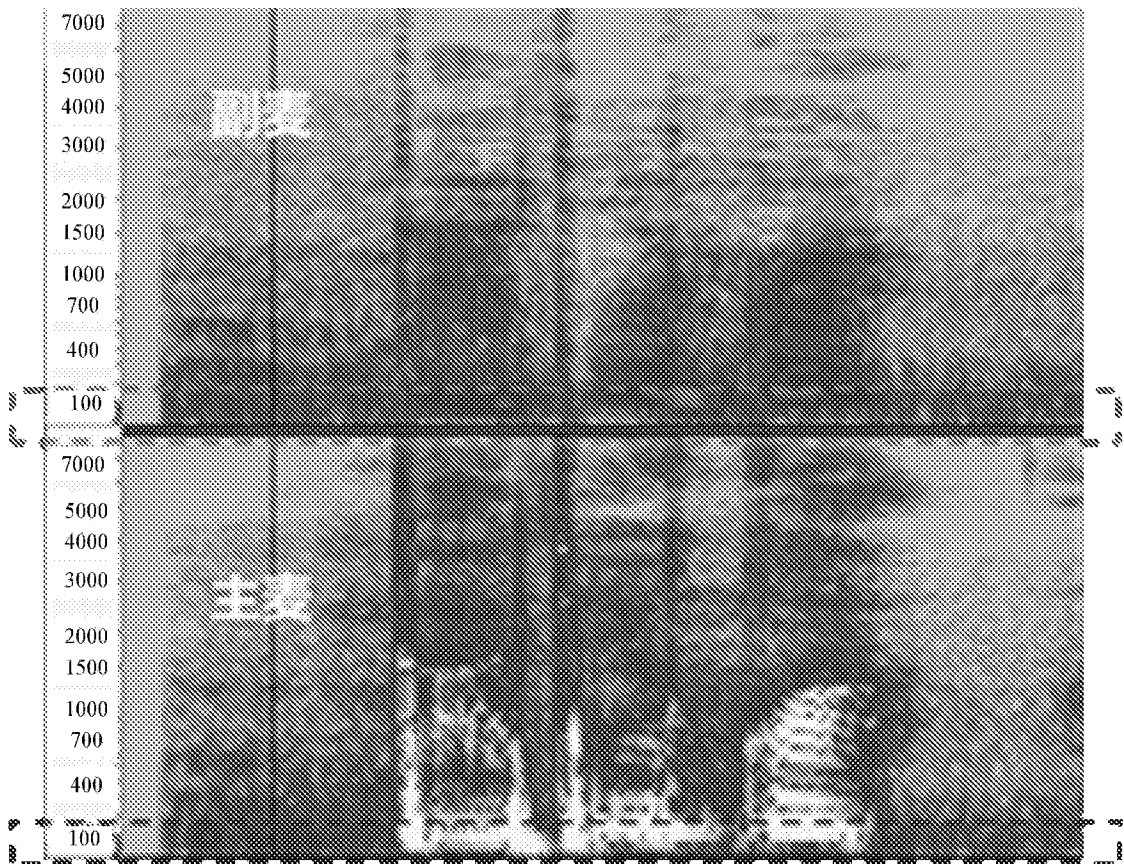


图 5

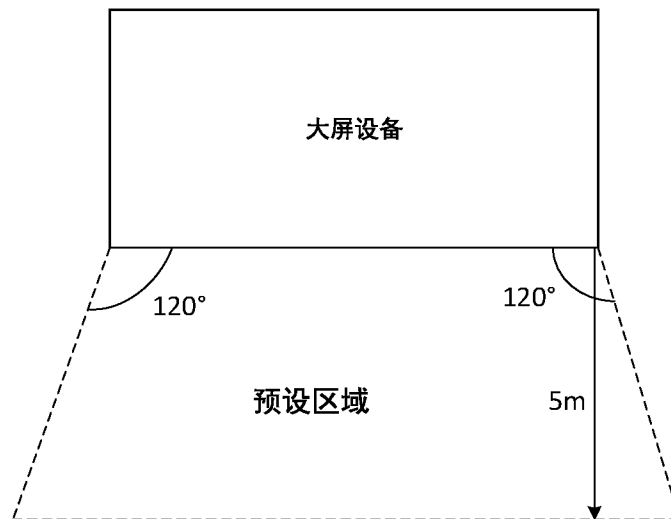


图 6

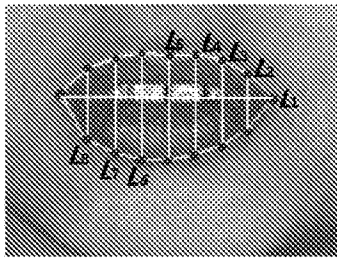


图 7

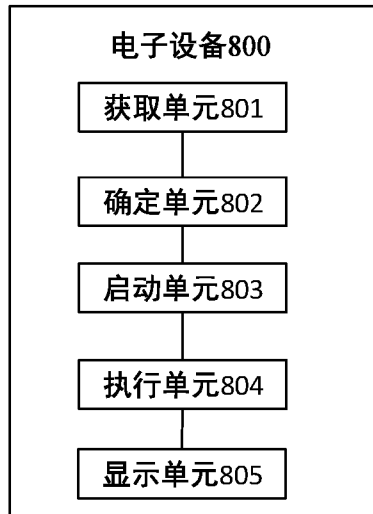


图 8

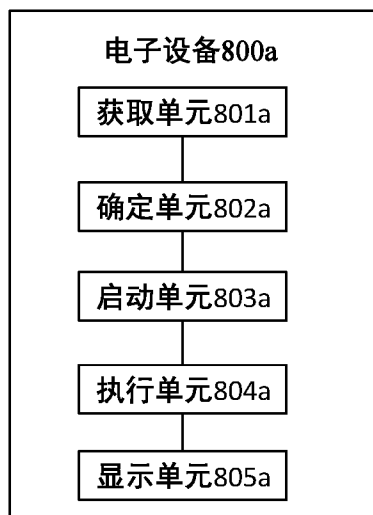


图 8a

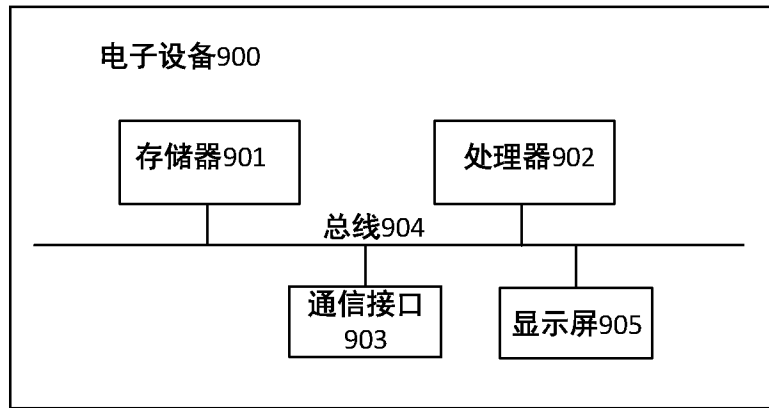


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/078662

A. CLASSIFICATION OF SUBJECT MATTER G10L15/22(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) IPC: G10L 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) CNABS, CNTXT, ENTXTC, WPABSC, ENTXT: 唤醒, 激活, 启动, 苏醒, 叫醒, 语音, 惯性测量, 加速度, 照度, 遮挡, 挡光, 环境光, 免, 无, 不需, 无需, 无须, 唤醒词, 抬手, 抬腕, 抬臂, 转腕, 抬起, awak+, activat+, start+, wak+, voice, speech, assistant, interact+, IMU, inertial, acceleration, illuminance, block+, obscurat+, ambient, environment, no need, without, lift+		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116229953 A (HUAWEI TECHNOLOGIES CO., LTD.) 06 June 2023 (2023-06-06) claims 1-38	1-38
Y	CN 114283798 A (HISENSE VISUAL TECHNOLOGY CO., LTD.) 05 April 2022 (2022-04-05) description, paragraphs 0040-0076, and figures 1-5	11-12, 15-18, 29-30, 33-38
Y	CN 114341779 A (FACEBOOK TECHNOLOGIES, LLC.) 12 April 2022 (2022-04-12) description, paragraph 0086	11-12, 15-18, 29-30, 33-38
A	CN 113377206 A (ANHUI TOYCLOUD TECHNOLOGY CO., LTD.) 10 September 2021 (2021-09-10) entire document	1-38
A	CN 111651041 A (SHANGHAI LONGCHEER TECHNOLOGY CO., LTD.) 11 September 2020 (2020-09-11) entire document	1-38
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 "D" document cited by the applicant in the international application "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 03 June 2024		Date of mailing of the international search report 05 June 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/ CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/078662

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 113724699 A (NREAL (BEIJING) TECHNOLOGY CO., LTD.) 30 November 2021 (2021-11-30) entire document	1-38
A	CN 115588435 A (HONOR TERMINAL CO., LTD.) 10 January 2023 (2023-01-10) entire document	1-38
A	WO 2022228056 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 03 November 2022 (2022-11-03) entire document	1-38

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2024/078662

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116229953	A	06 June 2023	None			
CN	114283798	A	05 April 2022	None			
CN	114341779	A	12 April 2022	JP	2022546179	A	04 November 2022
				EP	4025982	A1	13 July 2022
				WO	2021046267	A1	11 March 2021
				US	2021064132	A1	04 March 2021
				US	11493993	B2	08 November 2022
CN	113377206	A	10 September 2021	None			
CN	111651041	A	11 September 2020	None			
CN	113724699	A	30 November 2021	None			
CN	115588435	A	10 January 2023	None			
WO	2022228056	A1	03 November 2022	None			

<p>A. 主题的分类</p> <p>G10L15/22(2006.01)i</p> <p>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</p>																							
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>IPC: G10L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNABS, CNTXT, ENTXTC, WPABSC, ENTXT: 唤醒, 激活, 启动, 苏醒, 叫醒, 语音, 惯性测量, 加速度, 照度, 遮挡, 挡光, 环境光, 免, 无, 不需, 无需, 无须, 唤醒词, 抬手, 抬腕, 抬臂, 转腕, 抬起, awak+, activat+, start+, wak+, voice, speech, assistant, interact+, IMU, inertial, acceleration, illuminance, block+, obscurat+, ambient, environment, no need, without, lift+</p>																							
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 116229953 A (华为技术有限公司) 2023年6月6日 (2023 - 06 - 06) 权利要求1-38</td> <td>1-38</td> </tr> <tr> <td>Y</td> <td>CN 114283798 A (海信视像科技股份有限公司) 2022年4月5日 (2022 - 04 - 05) 说明书第0040-0076段及图1-5</td> <td>11-12、15-18、 29-30、33-38</td> </tr> <tr> <td>Y</td> <td>CN 114341779 A (脸谱科技有限责任公司) 2022年4月12日 (2022 - 04 - 12) 说明书第0086段</td> <td>11-12、15-18、 29-30、33-38</td> </tr> <tr> <td>A</td> <td>CN 113377206 A (安徽淘云科技股份有限公司) 2021年9月10日 (2021 - 09 - 10) 全文</td> <td>1-38</td> </tr> <tr> <td>A</td> <td>CN 111651041 A (上海龙旗科技股份有限公司) 2020年9月11日 (2020 - 09 - 11) 全文</td> <td>1-38</td> </tr> <tr> <td>A</td> <td>CN 113724699 A (优奈柯恩(北京)科技有限公司) 2021年11月30日 (2021 - 11 - 30) 全文</td> <td>1-38</td> </tr> </tbody> </table> <p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 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 “D” 申请人在国际申请中引证的文件 “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</p>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 116229953 A (华为技术有限公司) 2023年6月6日 (2023 - 06 - 06) 权利要求1-38	1-38	Y	CN 114283798 A (海信视像科技股份有限公司) 2022年4月5日 (2022 - 04 - 05) 说明书第0040-0076段及图1-5	11-12、15-18、 29-30、33-38	Y	CN 114341779 A (脸谱科技有限责任公司) 2022年4月12日 (2022 - 04 - 12) 说明书第0086段	11-12、15-18、 29-30、33-38	A	CN 113377206 A (安徽淘云科技股份有限公司) 2021年9月10日 (2021 - 09 - 10) 全文	1-38	A	CN 111651041 A (上海龙旗科技股份有限公司) 2020年9月11日 (2020 - 09 - 11) 全文	1-38	A	CN 113724699 A (优奈柯恩(北京)科技有限公司) 2021年11月30日 (2021 - 11 - 30) 全文	1-38
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 116229953 A (华为技术有限公司) 2023年6月6日 (2023 - 06 - 06) 权利要求1-38	1-38																					
Y	CN 114283798 A (海信视像科技股份有限公司) 2022年4月5日 (2022 - 04 - 05) 说明书第0040-0076段及图1-5	11-12、15-18、 29-30、33-38																					
Y	CN 114341779 A (脸谱科技有限责任公司) 2022年4月12日 (2022 - 04 - 12) 说明书第0086段	11-12、15-18、 29-30、33-38																					
A	CN 113377206 A (安徽淘云科技股份有限公司) 2021年9月10日 (2021 - 09 - 10) 全文	1-38																					
A	CN 111651041 A (上海龙旗科技股份有限公司) 2020年9月11日 (2020 - 09 - 11) 全文	1-38																					
A	CN 113724699 A (优奈柯恩(北京)科技有限公司) 2021年11月30日 (2021 - 11 - 30) 全文	1-38																					
国际检索实际完成的日期	2024年6月3日	国际检索报告邮寄日期	2024年6月5日																				
ISA/CN的名称和邮寄地址	中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	授权官员	喻新 电话号码 (+86) 010-62089308																				

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 115588435 A (荣耀终端有限公司) 2023年1月10日 (2023 - 01 - 10) 全文	1-38
A	WO 2022228056 A1 (HUAWEI TECH. CO., LTD.) 2022年11月3日 (2022 - 11 - 03) 全文	1-38

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2024/078662

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	116229953	A	2023年6月6日	无			
CN	114283798	A	2022年4月5日	无			
CN	114341779	A	2022年4月12日	JP	2022546179	A	2022年11月4日
				EP	4025982	A1	2022年7月13日
				WO	2021046267	A1	2021年3月11日
				US	2021064132	A1	2021年3月4日
				US	11493993	B2	2022年11月8日
CN	113377206	A	2021年9月10日	无			
CN	111651041	A	2020年9月11日	无			
CN	113724699	A	2021年11月30日	无			
CN	115588435	A	2023年1月10日	无			
WO	2022228056	A1	2022年11月3日	无			