

(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
2022年8月4日 (04.08.2022)



(10) 国际公布号
WO 2022/161077 A1

- (51) 国际专利分类号:
G10L 15/22 (2006.01)
- (21) 国际申请号: PCT/CN2021/142083
- (22) 国际申请日: 2021年12月28日 (28.12.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110130831.0 2021年1月29日 (29.01.2021) 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: 王晓博 (WANG, Xiaobo); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 许嘉璐 (XU, Jialu); 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: 北京润泽恒知识产权代理有限公司 (BEIJING RUN ZEHENG INTELLECTUAL PROPERTY LAW FIRM); 中国北京市海淀区中关村南大街甲18号北京国际C座6层606, Beijing 100081 (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, IT,

(54) Title: SPEECH CONTROL METHOD, AND ELECTRONIC DEVICE

(54) 发明名称: 语音控制方法和电子设备

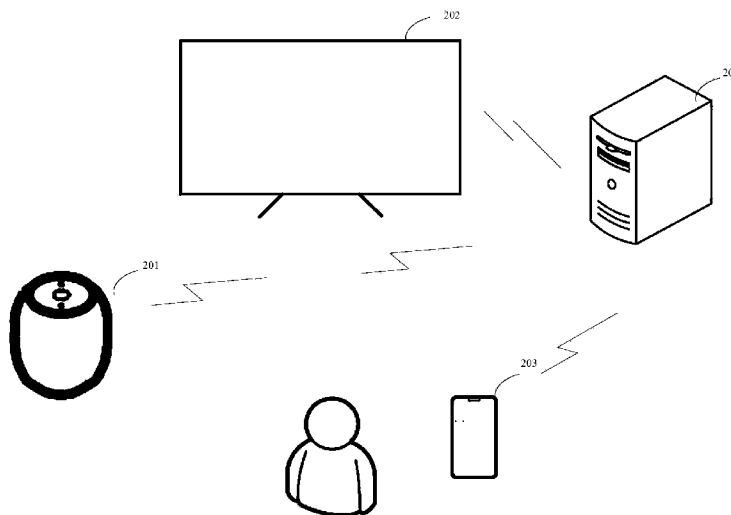


图1

(57) Abstract: A speech control method, and an electronic device The speech control method is applied to a speech control system, wherein the speech control system at least comprises a first electronic device and a second electronic device, which have a speech control function. The speech control method comprises: a first electronic device and a second electronic device respectively receiving a first speech instruction input by a user, and the first electronic device responding to the first speech instruction; the second electronic device performing recording and storing recording data, wherein the recording is used for recording a second speech instruction input



WO 2022/161077 A1

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))。

by the user; the second electronic device sending the recording data of the second electronic device to the first electronic device; the first electronic device responding to the second speech instruction according to recording data of the first electronic device and/or the recording data of the second electronic device, wherein the recording data of the first electronic device comprises recording data of when the first electronic device records a second speech instruction input by the user. By means of the method, the problem of false recognition of speech control in a multi-device scenario can be solved, thereby improving the accuracy of speech control.

(57) 摘要: 一种语音控制方法和电子设备。语音控制方法应用于语音控制系统, 语音控制系统至少包括具备语音控制功能的第一电子设备和第二电子设备, 语音控制方法包括: 第一电子设备和第二电子设备分别接收用户输入的第一语音指令, 第一电子设备应答第一语音指令; 第二电子设备录音, 并保存录音数据, 录音用于录制用户输入的第二语音指令; 第二电子设备向第一电子设备发送第二电子设备的录音数据; 第一电子设备根据第一电子设备的录音数据和/或第二电子设备的录音数据, 应答第二语音指令; 其中, 第一电子设备的录音数据包括第一电子设备录制用户输入的第二语音指令的录音数据。该方法可以解决多设备场景中语音控制的误识别问题, 提升语音控制的准确率。

语音控制方法和电子设备

5 本申请要求于 2021 年 01 月 29 日提交中国专利局、申请号为 202110130831.0、
申请名称为“语音控制方法和电子设备”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

10 本申请涉及计算机技术，尤其涉及一种语音控制方法和电子设备。

背景技术

15 语音助手作为一种基于语音语义算法的新型终端应用程序（application，APP），通过
接收和识别用户发送的语音信号，提供交互对话、信息查询、设备控制等服务功能。随着
深度学习理论的不断发展和智能语音硬件的成熟，语音助手应用程序已经成为智能手机、
平板电脑、智能电视、智能音箱等终端设备必备的软件功能。

20 随着搭载语音助手的终端设备大量普及，众多用户已持有多台相同或相异类型的终端
设备。在用户并发使用多台终端设备的场景下，或用户语音交互发生在多台终端设备有效
工作范围内的场景下，通过终端设备之间的信号检测和交互协商，选择拾音最清晰（即距
离用户最近）的终端设备作为拾音入口，供语音助手应用程序调用，可以提升语音助手应
25 用程序的识别准确率。例如，用户家客厅有音箱，电视机和手机三台设备，这三台设备均
安装有语音助手应用程序，且唤醒词均为“小 E 小 E”。那么，当用户说出唤醒词“小 E 小 E”
后，音箱，电视机以及手机的语音助手应用程序通过检测唤醒词的音频能量信息，在三台
设备中选择一台设备作为应答设备。由于音箱距离用户最近，所以三台设备基于唤醒词的
音频能量信息，协商选择出音箱作为应答设备。音箱唤醒自身的语音助手应用程序，其他
30 设备则对唤醒词不响应，即不唤醒各自的语音助手应用程序。这样，在用户继续说出语音
信号后，就会只有音箱对用户的语音信号进行识别并响应。例如，用户说出语音信号“播
放歌曲 112222”后，音箱对该语音信号进行识别并响应。例如，音箱响应输出语音信号“将
为你播放歌曲 112222”。

35 上述多设备语音控制过程中，由应答设备对用户的语音信号进行识别并响应，然而，
由于使用场景的多样性和复杂性，这种处理方式会存在应答设备误识别的问题，即存在应
答设备不能准确识别用户在唤醒词之后输入的语音信号的问题。

发明内容

本申请提供一种语音控制方法和电子设备，以解决多设备场景中语音控制的误识别问
40 题，提升语音控制的准确率。

第一方面，本申请实施例提供一种语音控制方法，该语音控制方法可以应用于语音控
制系统，该语音控制系统至少可以包括具备语音控制功能的第一电子设备和第二电子设备，
该语音控制方法可以包括：第一电子设备和第二电子设备分别接收用户输入的第一语音指

令，该第一电子设备应答该第一语音指令。第二电子设备录音，并保存录音数据，该录音用于录制用户输入的第二语音指令。第二电子设备向第一电子设备发送第二电子设备的录音数据。第一电子设备根据第一电子设备的录音数据和/或第二电子设备的录音数据，应答第二语音指令。其中，第一电子设备的录音数据包括第一电子设备录制用户输入的第二语音指令的录音数据。

5 第二电子设备录音可以开始于第一电子设备应答该第一语音指令之前，将应答设备的选择过程与电子设备的录音过程解耦合，无论多个电子设备之间是否决策出第一电子设备作为应答设备，第二电子设备都可以对用户输入第二语音指令进行录音，并保存，在决策出第一电子设备作为应答设备之后，将第二电子设备的录音数据发送给第一电子设备，由

10 第一电子设备应答第二语音指令。

本实现方式，第一电子设备作为应答设备应答第一语音指令，第一电子设备和第二电子设备均对第二语音指令进行录音，并保存录音数据，第二电子设备将自身的录音数据发送给第一电子设备，第一电子设备根据第一电子设备的录音数据和/或第二电子设备的录音数据，应答第二语音指令。本实现方式通过非应答设备对用户输入的语音指令进行录音，

15 应答设备基于应答设备的录音数据和/或非应答设备的录音数据，进行 SE、ASR 等处理，有效消除选取应答设备过程中设备之间的通信时延，从而解决多设备场景中因时延导致的语音控制的丢帧问题。应答设备通过多设备协同收音的录音数据，应答第二语音指令，可解决电子设备所拾取的语音指令的音频质量对 ASR 识别准确率的影响问题，提升语音控制的准确率。

20 一种可能的设计中，该方法还可以包括：第一电子设备向第二电子设备调用拾音指令，该拾音指令用于第二电子设备返回第二电子设备的录音数据。

一种可能的设计中，该第二电子设备录音，可以包括：在第二电子设备接收到用户输入的第一语音指令时或之后，第二电子设备录音。

本实现方式，通过在第二电子设备接收到用户输入的第一语音指令时或之后，第二电

25 子设备录音，即在确定应答设备之前第二电子设备开始录音，第二电子设备可以录制到用户输入的第二语音指令。这样可以有效消除选取应答设备过程中设备之间的通信时延，从而解决多设备场景中因时延导致的语音控制的丢帧问题。

一种可能的设计中，该方法还可以包括：在第一电子设备接收到用户输入的第一语音指令时或之后，第一电子设备录音，该录音用于录制用户输入的第二语音指令。

30 一种可能的设计中，该第一语音指令用于唤醒第一电子设备和/或第二电子设备的语音控制功能。

为了便于理解，这里的第一语音指令可以是下述图 3 所示实施例的步骤 401 的语音指令。

一种可能的设计中，该方法还可以包括：第一电子设备和第二电子设备分别根据各自接收到的第一语音指令的音频质量信息，确定第一电子设备为语音控制系统的应答设备。

一种可能的设计中，在第一电子设备应答第一语音指令之后，在录制到用户输入的第二语音指令之前，该方法还可以包括：在第一电子设备和第二电子设备录音过程中，该第一电子设备在预设时间段内未检测到用户输入的第二语音指令，第一电子设备删除已保存的录音数据，并继续录音。第一电子设备向第二电子设备调用多轮对话暂停指令，该多轮

对话暂停指令用于指示多轮对话暂时停止。该第二电子设备删除已保存的录音数据，并继续录音。

为了便于理解，这里的第一语音指令可以是下述图 6 所示实施例的步骤 701 之前的语音指令。这里的第二语音指令可以是下述图 6 所示实施例的步骤 703 的语音指令。

5 一种可能的设计中，该方法还可以包括：第一电子设备接收第二电子设备发送的第二电子设备的录音数据的音频质量信息。

本实现方式，可以加快最优收音设备的决策，从而提升语音控制响应速度。

一种可能的设计中，第一电子设备根据第一电子设备的录音数据和/或第二电子设备的录音数据，应答第二语音指令，可以包括：第一电子设备根据第一电子设备的录音数据的音频质量信息和第二电子设备的录音数据的音频质量信息，从语音控制系统中确定最优收音设备。当最优收音设备为第一电子设备时，第一电子设备根据第一电子设备的录音数据，或者，根据第一电子设备的录音数据和第二电子设备的录音数据，应答第二语音指令。当最优收音设备为第二电子设备时，第一电子设备根据第二电子设备的录音数据，或者，根据第二电子设备的录音数据和第一电子设备的录音数据，应答第二语音指令。其中，该音频质量信息用于表示录音数据的音频质量。

15 本实现方式，通过使用最优收音设备的录音数据，应答第二语音指令，可以降低噪声对语音控制准确率的影响。

一种可能的设计中，第一电子设备根据第一电子设备的录音数据和/或第二电子设备的录音数据，应答第二语音指令，可以包括：第一电子设备根据第一电子设备的录音数据的音频内容信息和/或第二电子设备的录音数据的音频内容信息，应答第二语音指令。其中，音频内容信息用于表示录音数据的音频内容。

例如，当第一电子设备的录音数据的音频内容信息多于第二电子设备的录音数据的音频内容信息时，根据第一电子设备的录音数据的音频内容信息，应答第二语音指令。当第一电子设备的录音数据的音频内容信息少于第二电子设备的录音数据的音频内容信息时，根据第二电子设备的录音数据的音频内容信息，应答第二语音指令。再例如，当第一电子设备的录音数据的音频内容信息与第二电子设备的录音数据的音频内容信息存在部分相同内容时，第一电子设备可以对第一电子设备的录音数据的音频内容信息和第二电子设备的录音数据的音频内容信息进行拼接，根据拼接后的音频内容信息，应答第二语音指令。

25 本实现方式，通过使用多设备协同收音的录音数据，应答第二语音指令，可以避免丢帧，提升语音控制的准确率。

第二方面，本申请实施例提供一种语音控制方法，该方法可以应用于语音控制系统的第一电子设备，该语音控制系统还可以至少包括第二电子设备，该语音控制方法可以包括：第一电子设备接收用户输入的第一语音指令，第一电子设备应答第一语音指令。第一电子设备接收第二电子设备发送的第二电子设备的录音数据，第二电子设备的录音数据包括第二电子设备录制用户输入的第二语音指令的录音数据。第一电子设备根据第一电子设备的录音数据和/或第二电子设备的录音数据，应答第二语音指令，第一电子设备的录音数据包括第一电子设备录制用户输入的第二语音指令的录音数据。

一种可能的设计中，该方法还可以包括：第一电子设备向第二电子设备调用拾音指令，拾音指令用于第二电子设备返回第二电子设备的录音数据。

一种可能的设计中，该方法还可以包括：在第一电子设备接收用户输入的第一语音指令时或之后，第一电子设备录音，录音用于录制用户输入的第二语音指令。

一种可能的设计中，第一语音指令用于唤醒第一电子设备和/或第二电子设备的语音控制功能。

5 一种可能的设计中，该方法还可以包括：第一电子设备根据第一电子设备接收到的第一语音指令的音频质量信息和第二电子设备接收到的第一语音指令的音频质量信息，确定第一电子设备为语音控制系统的应答设备。

一种可能的设计中，在第一电子设备应答第一语音指令之后，在录制到用户输入的第二语音指令之前，该方法还可以包括：在第一电子设备录音过程中，第一电子设备在预设
10 时间段内未检测到用户输入的第二语音指令，第一电子设备删除已保存的录音数据，并继续录音；第一电子设备向第二电子设备调用多轮对话暂停指令，多轮对话暂停指令用于指示多轮对话暂时停止；第二电子设备删除已保存的录音数据，并继续录音。

一种可能的设计中，该方法还可以包括：第一电子设备接收第二电子设备发送的第二电子设备的录音数据的音频质量信息。

15 一种可能的设计中，第一电子设备根据第一电子设备的录音数据和/或第二电子设备的录音数据，应答第二语音指令，可以包括：第一电子设备根据第一电子设备的录音数据的音频质量信息和第二电子设备的录音数据的音频质量信息，从语音控制系统中确定最优收音设备。当最优收音设备为第一电子设备时，第一电子设备根据第一电子设备的录音数据，
20 应答第二语音指令。当最优收音设备为第二电子设备时，第一电子设备根据第二电子设备的录音数据，或者，根据第二电子设备的录音数据和第一电子设备的录音数据，应答第二语音指令。其中，音频质量信息用于表示录音数据的音频质量。

一种可能的设计中，第一电子设备根据第一电子设备的录音数据和/或第二电子设备的录音数据，应答第二语音指令，可以包括：第一电子设备根据第一电子设备的录音数据的音频内容信息和/或第二电子设备的录音数据的音频内容信息，应答第二语音指令。其中，
25 音频内容信息用于表示录音数据的音频内容。

第三方面，本申请实施例提供一种语音控制方法，该语音控制方法可以应用于语音控制系统的第二电子设备，该语音控制系统还可以至少包括第一电子设备，该语音控制方法
30 可以包括：第二电子设备录音，并保存录音数据，录音用于录制用户输入的第二语音指令。第二电子设备向第一电子设备发送第二电子设备的录音数据，第二电子设备的录音数据包括第二电子设备录制用户输入的第二语音指令的录音数据，录音数据用于第一电子设备在
应答第一语音指令之后，应答第二语音指令。

一种可能的设计中，该方法还可以包括：第二电子设备接收第一电子设备调用拾音指令，拾音指令用于第二电子设备返回第二电子设备的录音数据。

一种可能的设计中，第二电子设备录音，可以包括：在第二电子设备接收到用户输入
35 的第一语音指令时或之后，第二电子设备录音。

一种可能的设计中，该方法还可以包括：第二电子设备根据第二电子设备接收到的第一语音指令的音频质量信息和第一电子设备接收到的第一语音指令的音频质量信息，确定第一电子设备为语音控制系统的应答设备。

一种可能的设计中，在第一电子设备应答第一语音指令之后，该方法还可以包括：在

第二电子设备录音过程中，第二电子设备接收第二电子设备调用多轮对话暂停指令，多轮对话暂停指令用于指示多轮对话暂时停止；第二电子设备删除已保存的录音数据，并继续录音。

5 一种可能的设计中，该方法还可以包括：第二电子设备向第一电子设备发送第二电子设备的录音数据的音频质量信息。

第四方面，本申请实施例提供一种语音控制装置，该装置具有实现上述第二方面或第二方面任一种可能的设计的功能。所述功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。所述硬件或软件包括一个或多个与上述功能相对应的模块，例如，收发单元或模块，处理单元或模块。

10 第五方面，本申请实施例提供一种语音控制装置，该装置具有实现上述第三方面或第三方面任一种可能的设计的功能。所述功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。所述硬件或软件包括一个或多个与上述功能相对应的模块，例如，收发单元或模块，处理单元或模块。

15 第六方面，本申请实施例提供一种电子设备，该电子设备可以包括：一个或多个处理器；一个或多个存储器；其中，所述一个或多个存储器用于存储一个或多个程序；所述一个或多个处理器用于运行所述一个或多个程序，以实现如第二方面或第二方面任一种可能的设计所述的方法。

20 第七方面，本申请实施例提供一种电子设备，该电子设备可以包括：一个或多个处理器；一个或多个存储器；其中，所述一个或多个存储器用于存储一个或多个程序；所述一个或多个处理器用于运行所述一个或多个程序，以实现如第三方面或第三方面任一种可能的设计所述的方法。

第八方面，本申请实施例提供一种计算机可读存储介质，其特征在于，包括计算机程序，所述计算机程序在计算机上被执行时，使得所述计算机执行如第二方面或第二方面任一种可能的设计所述的方法。

25 第九方面，本申请实施例提供一种计算机可读存储介质，其特征在于，包括计算机程序，所述计算机程序在计算机上被执行时，使得所述计算机执行如第三方面或第三方面任一种可能的设计所述的方法。

30 第十方面，本申请实施例提供一种芯片，其特征在于，包括处理器和存储器，所述存储器用于存储计算机程序，所述处理器用于调用并运行所述存储器中存储的计算机程序，以执行如第二方面或第二方面任一种可能的设计所述的方法。

第十一方面，本申请实施例提供一种芯片，其特征在于，包括处理器和存储器，所述存储器用于存储计算机程序，所述处理器用于调用并运行所述存储器中存储的计算机程序，以执行如第三方面或第三方面任一种可能的设计所述的方法。

35 第十二方面，本申请实施例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二方面或第二方面任一种可能的设计所述的方法。

第十三方面，本申请实施例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三方面或第三方面任一种可能的设计所述的方法。

第十四方面，本申请实施例提供一种语音控制系统，所述语音控制系统至少包括具备语音控制功能的第一电子设备和第二电子设备。第一电子设备用于执行如第二方面或第二

方面任一种可能的设计所述的方法。第二电子设备用于执行如第三方面或第三方面任一种可能的设计所述的方法。

5 本申请实施例的语音控制方法和电子设备，在上述多设备场景下，通过多个设备之间不进行跨设备通信直接录音的方式，解决多设备场景中语音控制的丢帧问题，提升语音控制的准确率。之后，通过多设备协同收音的录音数据，应答用户输入的语音指令，可有效解决电子设备所拾取的语音指令的音频质量对 ASR 识别准确率的影响问题，提升语音控制的准确率。

附图说明

- 10 图 1 为本申请实施例提供一种语音控制系统的示意图；
图 2 为本申请实施例提供的一种电子设备的硬件结构示意图；
图 3 为本申请实施例提供的一种语音控制方法的流程示意图；
图 4 为本申请实施例提供的一种多设备语音控制的场景示意图；
图 5 为本申请实施例提供的另一种多设备语音控制的场景示意图；
15 图 6 为本申请实施例提供的另一种语音控制方法的流程示意图；
图 7 为本申请实施例提供的另一种多设备语音控制的场景示意图；
图 8 为本申请实施例提供的一种语音控制装置的结构示意图；
图 9 为本申请实施例提供的一种语音控制装置的结构示意图；
图 10 为本申请实施例提供的一种电子设备的结构示意图。

20

具体实施方式

本申请实施例涉及的术语“第一”、“第二”等仅用于区分描述的目的，而不能理解为指示或暗示相对重要性，也不能理解为指示或暗示顺序。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元。方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

应当理解，在本申请中，“至少一个（项）”是指一个或者多个，“多个”是指两个或两个以上。“和/或”，用于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：只存在 A，只存在 B 以及同时存在 A 和 B 三种情况，其中 A，B 可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项（个）”或其类似表达，是指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 和 b”，“a 和 c”，“b 和 c”，或“a 和 b 和 c”，其中 a，b，c 可以是单个，也可以是多个。

30 语音助手：一种基于人工智能构建的应用程序，借助语音语义识别算法，通过与用户进行即时问答式的语音交互，帮助用户完成信息查询、设备控制、文本输入等操作。语音助手通常采用分阶段级联处理，依次通过语音唤醒、语音前端处理、自动语音识别（automatic speech recognition, ASR）、自然语言理解（natural language understanding, NLU）、对话管理（dialog management, DM）、自然语言生成（natural language generation, NLG）、文本转语音（text to speech, TTS）等基本工作流程提供服务功能。其中，语音前端处理可以

包括但不限于语音增强（speech enhancement, SE）。ASR 可以以 SE 降噪处理后的语音信号为输入，输出用户语音信号的文本化描述结果。ASR 是语音助手应用程序准确完成后续识别处理任务的基础。输入 ASR 的用户语音信号的音频质量，直接决定了 ASR 识别结果的准确率。本申请实施例的语音控制方法可以保障输入 ASR 的用户语音信号的准确性和可靠性，从而提升 ASR 识别结果的准确率，进而准确完成后续识别处理任务。

语音唤醒：电子设备在锁屏或语音助手休眠状态下，接收并检测特定的用户语音信号（即唤醒词），激活或启动语音助手，使语音助手进入等待语音信号输入状态。

回声消除（acoustic echo cancellation, AEC）：一种语音前端处理技术，通过音波干扰方式消除麦克风与扬声器因空气产生回受路径而产生的杂音，可有效缓解由于扬声器播放音频或声波空间反射所引发的噪声干扰问题。

多设备语音控制过程中，多个电子设备通过相互通信协商选择出应答设备，由应答设备对用户的语音信号进行识别并响应。这种处理方式存在误识别的原因有两个方面：音频质量和时延。对于音频质量，由于使用场景的多样性和复杂性，电子设备所拾取并处理的用户语音指令不可避免地会受到各类外部噪声和内部噪声的干扰。噪声的干扰会影响电子设备所拾取用户语音指令的音频质量。例如，外部噪声可以是设备周边的空调风机、无关人声等噪声，内部噪声可以是电子设备自身所播放的音/视频。另外，电子设备与用户之间的距离、方位，以及电子设备自身摆放姿态和麦克风模组性能等，也会影响电子设备所拾取用户语音指令的音频质量。当电子设备所拾取用户语音指令的音频质量较差时，会引发误识别。对于时延，多个电子设备在协商选择应答设备的过程中，多个电子设备之间跨设备通信产生的通信时延和应答设备选择产生的时延，都会引发丢帧问题，进而引发误识别。例如，上述时延，会导致用户说出语音信号“播放歌曲 112222”，而应答设备仅识别到语音信号“2222”，即未接收并识别到语音信号“播放歌曲 11”，进而使得应答设备无法对用户语音指令进行准确识别和响应。

本申请实施例的语音控制方法可以从提升音频质量和/或降低时延，解决多设备语音控制过程中的语音指令误识别的问题。通过多个电子设备之间不进行跨设备通信直接开始录音的方式，消除通过通信实现多设备唤醒和数据传输，而产生的时延，进而消除时延对 ASR 识别准确性的影响，解决多设备场景中语音控制的丢帧问题，提升语音控制的准确率。通过在多个电子设备中选择一个或多个电子设备作为最优收音设备，最优收音设备的录音数据的音频质量好于其他电子设备。基于最优收音设备的录音数据，对用户输入的语音指令进行响应。通过多设备协同收音，可解决电子设备所拾取的语音指令的音频质量对 ASR 识别准确率的影响问题，提升语音控制的准确率。

本申请实施例的语音控制方法，可以应用于多设备场景。多设备场景可以包括用户并发使用多个电子设备的场景，或用户语音交互发生在多个电子设备的有效工作范围内的场景。其中，多个电子设备各自具备语音控制功能。该语音控制功能可以由语音助手提供。在该多设备场景下，用户在说出唤醒词和语音指令后，采用本实施例的方法，可以保障输入 ASR 的语音指令的准确性和可靠性，从而提升 ASR 识别结果的准确率，进而准确完成后续识别处理任务，完成对语音指令的响应。使得电子设备更加智能，实现了电子设备与用户之间的高效准确互动。同时，提高了用户的使用体验。

本申请实施例的语音指令，指用户以声音形式向电子设备输入的指令。该语音指令用于使得电子设备向用户提供交互对话、信息查询、设备控制等服务功能。例如，该语音指令可以是用户通过电子设备的麦克风输入的一段语音信号。

5 在一些实施例中，可以通过在电子设备中安装语音助手，以使该电子设备实现语音控制功能。语音助手一般情况下是处于休眠状态的。用户在使用电子设备的语音控制功能之前，可以对语音助手进行语音唤醒。其中，唤醒语音助手的语音信号可以称为唤醒词(或唤醒语音)。该唤醒词可以预先注册在电子设备中。举例而言，该唤醒词可以是“小 E 小 E”，当然可以理解的，唤醒词也可以是其他任意词语或语句，其可以根据需求进行灵活设置，本申请实施例不一一举例说明。

10 另外，上述语音助手可以是电子设备中的嵌入式应用(即电子设备的系统应用)，也可以是可下载应用。嵌入式应用是作为电子设备(如手机)实现的一部分提供的应用程序。可下载应用是一个可以提供自己的因特网协议多媒体子系统(internet protocol multimedia subsystem, IMS)连接的应用程序。可下载应用可以预先安装在电子设备中，也可由用户下载并安装在电子设备中的第三方应用。

15 下面将结合附图对本申请实施例的实施方式进行详细描述。

图 1 为本申请实施例提供的一种语音控制系统的示意图。该语音控制系统可以包括多个电子设备，且多个电子设备满足以下条件中的一个或多个：连接了同一个无线接入点(如 WiFi 接入点)，或登录了同一个账号，或被用户设置在同一个组中，或用户语音交互发生在该多个电子设备的有效工作范围内。

20 其中，作为一种示例，该语音控制系统可以包括三个电子设备，例如，第一电子设备 201、第二电子设备 202 和第三电子设备 203。第一电子设备 201、第二电子设备 202 和第三电子设备 203 均具备语音控制功能，如均安装有语音助手。

在一些实施例中，第一电子设备 201、第二电子设备 202 和第三电子设备 203 唤醒语音助手的唤醒词可以相同，如均为“小 E 小 E”。

25 示例性的，本申请实施例所述的电子设备，如上述第一电子设备 201，第二电子设备 202 以及第三电子设备 203 可以为手机、平板电脑、桌面型、膝上型、手持计算机、笔记本电脑、台式电脑、超级移动个人计算机(ultra-mobile personal computer, UMPC)、上网本，以及蜂窝电话、个人数字助理(personal digital assistant, PDA)、增强现实(augmented reality, AR)虚拟现实(virtual reality, VR)设备、媒体播放器、电视机、智能音箱、智能手表、智能耳机等设备。本申请实施例对电子设备的具体形态不作特殊限制。电子设备的具体结构可以参考图 2 对应实施例的描述。

30 另外，在一些实施例中，上述第一电子设备 201，第二电子设备 202 及第三电子设备 203 可以为相同类型的电子设备，如第一电子设备 201，第二电子设备 202 及第三电子设备 203 均为手机。在其他一些实施例中，上述第一电子设备 201，第二电子设备 202 及第三电子设备 203 可以为不同类型的电子设备，如第一电子设备 201 为手机，第二电子设备 202 为智能音箱，第三电子设备 203 为电视机(如图 1 所示)。

在本申请实施例中，通过第一电子设备 201，第二电子设备 202 及第三电子设备 203 之间不进行跨设备通信直接开始录音的方式，解决多设备场景中语音控制的丢帧问题，提升语音控制的准确率。

第一电子设备 201, 第二电子设备 202 及第三电子设备 203 各自无需其他设备(例如, 中心设备)调用, 便可以录音, 实现了去中心化的录音方式。这种去中心化的录音方式, 无需执行选择一台设备作为调用设备的过程, 可以有效消除设备之间通信所产生的时延, 提升后续语音控制的准确率。

5 之后, 基于第一电子设备 201, 第二电子设备 202 及第三电子设备 203 各自的设备信息、各自的录音数据等一个或多个维度, 在第一电子设备 201, 第二电子设备 202 及第三电子设备 203 中选择一个或多个电子设备作为最优收音设备。基于最优收音设备的录音数据, 对用户输入的语音指令进行响应。本申请实施例通过多设备协同收音, 可解决电子设备所拾取的语音指令的音频质量对 ASR 识别准确率的影响问题。

10 在一些实施例中, 该语音控制系统还可以包括服务器 204。该服务器 204 能够可以提供智能语音服务。

请参考图 2, 为本申请实施例提供的一种电子设备的结构示意图。

如图 2 所示, 电子设备可以包括处理器 110, 外部存储器接口 120, 内部存储器 121, 通用串行总线(universal serial bus, USB)接口 130, 充电管理模块 140, 电源管理模块 141, 15 电池 142, 天线 1, 天线 2, 移动通信模块 150, 无线通信模块 160, 音频模块 170, 扬声器 170A, 受话器 170B, 麦克风 170C, 耳机接口 170D, 传感器模块 180, 按键 190, 马达 191, 指示器 192, 摄像头 193, 显示屏 194, 以及用户标识模块(subscriber identification module, SIM)卡接口 195 等。其中, 传感器模块 180 可以包括压力传感器 180A, 陀螺仪传感器 180B, 气压传感器 180C, 磁传感器 180D, 加速度传感器 180E, 距离传感器 180F, 接近光传感器 20 180G, 指纹传感器 180H, 温度传感器 180J, 触摸传感器 180K, 环境光传感器 180L, 骨传导传感器 180M 等。

可以理解的是, 本实施例示意的结构并不构成对电子设备的具体限定。在另一些实施例中, 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, 或者拆分某些部件, 或者不同的部件布置。图示的部件可以以硬件, 软件或软件和硬件的组合实现。

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

控制器可以是电子设备的神经中枢和指挥中心。控制器可以根据指令操作码和时序信号, 产生操作控制信号, 完成取指令和执行指令的控制。

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

在一些实施例中, 处理器 110 可以包括一个或多个接口。接口可以包括集成电路(integrated circuit, I2C)接口, 集成电路内置音频(inter-integrated circuitsound, I2S)接口, 脉冲编码调制(pulse code modulation, PCM)接口, 通用异步收发传输器(universal asynchronous

receiver/transmitter, UART)接口, 移动产业处理器接口(mobile industry processor interface, MIPI), 通用输入输出(general-purpose input/output, GPIO)接口, 用户标识模块(subscriber identity module, SIM)接口, 和/或通用串行总线(universal serial bus, USB)接口等。

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

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

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

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

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

30 无线通信模块 160 可以提供应用在电子设备上的包括无线局域网(wireless local area networks, WLAN)(如无线保真(wireless fidelity, Wi-Fi)网络), 蓝牙(blueetooth, BT), 全球导航卫星系统(global navigation satellite system, GNSS), 调频(frequency modulation, FM), 近距离无线通信技术(near field communication, NFC), 红外技术(infrared, IR)等无线通信的解决方案。无线通信模块 160 可以是集成至少一个通信处理模块的一个或多个器件。无线通信模块 160 经由天线 2 接收电磁波, 将电磁波信号调频以及滤波处理, 将处理后的信号发送到处理器 110。无线通信模块 160 还可以从处理器 110 接收待发送的信号, 对其进行调频, 放大, 经天线 2 转为电磁波辐射出去。例如, 在本申请一些实施例中, 无线通信模块 160 可以与其他电子设备进行交互, 如在检测到与唤醒词匹配的语音信号后, 向其他电子设备发送检测到的语音信号的能量信息。例如, 本申请实施例的电子设备可以通过移动通信模块 150 和/或无线通信模块 160 与其他电子设备通信。举例而言, 第一电子设备 201 通过通信模块 150 和/或无线通信模块 160 向第二电子设备 202 发送调用拾音指令等。

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

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

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

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

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

摄像头 193 用于捕获静态图像或视频。物体通过镜头生成光学图像投射到感光元件。感光元件可以是电荷耦合器件(charge coupled device, CCD)或互补金属氧化物半导体
30 (complementary metal-oxide-semiconductor, CMOS)光电晶体管。感光元件把光信号转换成电信号，之后将电信号传递给 ISP 转换成数字图像信号。ISP 将数字图像信号输出到 DSP 加工处理。DSP 将数字图像信号转换成标准的 RGB，YUV 等格式的图像信号。在一些实施例中，电子设备可以包括 1 个或 N 个摄像头 193，N 为大于 1 的正整数。

数字信号处理器用于处理数字信号，除了可以处理数字图像信号，还可以处理其他数字信号。例如，当电子设备在频点选择时，数字信号处理器用于对频点能量进行傅里叶变换等。
35

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

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

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

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

15 盘存储器件,闪存器件,通用闪存存储器(universal flash storage, UFS)等。电子设备可以通过音频模块 170,扬声器 170A,受话器 170B,麦克风 170C,耳机接口 170D,以及应用处理器等实现音频功能。例如音乐播放,录音等。

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

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

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

30 麦克风 170C,也称“话筒”,“传声器”,用于将声音信号转换为电信号。当拨打电话或发送语音信息或需要通过语音助手触发电子设备执行某些事件时,用户可以通过人嘴靠近麦克风 170C 发声,将声音信号输入到麦克风 170C。电子设备可以设置至少一个麦克风 170C。在另一些实施例中,电子设备可以设置两个麦克风 170C,除了采集声音信号,还可以实现降噪功能。在另一些实施例中,电子设备还可以设置三个,四个或更多麦克风 170C,实现采集声音信号,降噪,还可以识别声音来源,实现定向录音功能等。例如,本申请实施例的电子设备可以通过麦克风 170C 接收用户输入的语音指令。

35 耳机接口 170D 用于连接有线耳机。耳机接口 170D 可以是 USB 接口 130,也可以是 3.5mm 的开放移动电子设备平台(open mobile terminal platform, OMTP)标准接口,美国蜂窝电信工业协会(cellular telecommunications industry association of the USA, CTIA)标准接口。

压力传感器 180A 用于感受压力信号,可以将压力信号转换成电信号。在一些实施例中,压力传感器 180A 可以设置于显示屏 194。压力传感器 180A 的种类很多,如电阻式压力传感器,电感式压力传感器,电容式压力传感器等。电容式压力传感器可以是包括至少两个具有导电材料的平行板。当有力作用于压力传感器 180A,电极之间的电容改变。电子设备根据电容的变化确定压力的强度。当有触摸操作作用于显示屏 194,电子设备根据压

力传感器 180A 检测所述触摸操作强度。电子设备也可以根据压力传感器 180A 的检测信号计算触摸的位置。在一些实施例中，作用于相同触摸位置，但不同触摸操作强度的触摸操作，可以对应不同的操作指令。例如：当有触摸操作强度小于第一压力阈值的触摸操作作用于短消息应用图标时，执行查看短消息的指令。当有触摸操作强度大于或等于第一压力阈值的触摸操作作用于短消息应用图标时，执行新建短消息的指令。

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

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

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

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

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

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

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

指纹传感器 180H 用于采集指纹。电子设备可以利用采集的指纹特性实现指纹解锁，访问应用锁，指纹拍照，指纹接听来电等。

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

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

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

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

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

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

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

以下实施例中的方法均可以在具有上述硬件结构的电子设备中实现。

在本申请实施例中，在上述多设备场景下，通过多个设备之间不进行跨设备通信直接开始录音的方式，解决多设备场景中语音控制的丢帧问题，提升语音控制的准确率。

之后，基于多个电子设备的设备信息、录音数据等一个或多个维度，在多个电子设备中选择一个或多个电子设备作为最优收音设备。基于最优收音设备的录音数据，对用户输入的语音指令进行响应。通过最优收音设备的选择，选择满足拾音最清晰(距离用户最近)、受噪声干扰情况最低(距离噪声源最远)、或 SE 处理效果最优(麦克风降噪性能最优或支持 AEC)中至少一项的电子设备作为拾音入口，供语音助手调用，可有效解决电子设备所拾取的语音指令的音频质量对 ASR 识别准确率的影响问题。其中，该设备信息可以包括但不限于电子设备的静态属性信息或动态属性信息。该静态属性信息可以包括但不限于设备型号、系统版本、麦克风的能力信息等。该动态属性信息可以包括但不限于电子设备的

电量信息、耳机状态信息、麦克风状态信息、扬声器状态信息、录音数据的音频质量信息等。其中，扬声器状态信息可以用于指示电子设备的扬声器是否被占用。音频质量信息用于表示录音数据的音频质量好坏。音频质量信息的具体形式可以包括音强信息、噪声声强信息、信噪比信息等一项或多项。

5

图3为本申请实施例提供的一种语音控制方法的流程示意图。本实施例以如图1所示的三个电子设备，音箱201，电视机202和手机203为例进行举例说明。如图3所示，本实施例的方法可以包括：

步骤401、音箱201，电视机202和手机203分别接收用户输入的第一语音指令。

10 该第一语音指令用于唤醒电子设备的语音助手。例如，该第一语音指令可以是上述唤醒词“小E小E”。本实施例中，该第一语音指令用于唤醒音箱201，电视机202和手机203各自的语音助手。

15 对于安装有语音助手的电子设备，在该电子设备没有其他软硬件使用麦克风采集语音信号的情况下，电子设备可以通过麦克风实时监测用户是否有语音信号输入。一般情况下，在用户想要使用电子设备的语音控制功能时，可以在电子设备的拾音范围内发声，以将发出的声音输入到麦克风。此时，若电子设备没有其他软硬件正在使用麦克风采集语音信号，则电子设备可以通过麦克风监测到对应的语音信号，如第一语音指令。

20 例如，结合图4所示，用户在想要使用语音控制功能时，可以说出唤醒词“小E小E”。如用户的发声位置位于音箱201，电视机202和手机203各自的拾音范围内，且均没有其他软硬件正在使用麦克风采集语音信号，则音箱201，电视机202和手机203便可通过各自的麦克风检测到唤醒词“小E小E”对应的第一语音指令。

步骤402、响应于第一语音指令，音箱201，电视机202和手机203分别唤醒各自的语音助手，并开始录音。

25 当电子设备检测到该第一语音指令时，响应于该第一语音指令，电子设备唤醒语音助手。一种示例，在电子设备接收到上述第一语音指令后，可以对该第一语音指令进行校验，即判断接收到的该第一语音指令是否是注册在电子设备中的唤醒词。如果校验通过，则表明接收到的第一语音指令是唤醒词，唤醒语音助手。如果校验未通过，则表明接收到的第一语音指令不是唤醒词，此时电子设备可以不唤醒语音助手，即保持语音助手的休眠状态。

30 本实施例中，当音箱201，电视机202和手机203分别检测到第一语音指令时，音箱201，电视机202和手机203分别唤醒各自的语音助手，并开始录音。音箱201，电视机202和手机203分别开始录音后，可以通过各自的麦克风检测用户是否输入其他语音指令，当检测到用户输入的其他语音指令时，生成录音数据，并保存在自身设备中。

35 例如，结合图3和图4所示，音箱201，电视机202和手机203在开始录音之后，分别接收到用户输入的第二语音指令。例如，以用户说出的第二语音指令为“播放歌曲112222”为例。音箱201，电视机202和手机203分别对第二语音指令进行录音，生成各自的录音数据，该录音数据的内容即为“播放歌曲112222”。

需要说明的是，一种可实现方式，录音数据可以是每录制0.5s，生成录音数据。其中，0.5还可以是其他数值，例如，0.6，1等，本申请实施例不一一举例说明。保存录音数据时，可以是使用新的录音数据覆盖之前的录音数据，也可以是不使用新的录音数据覆盖之前的

录音数据，保存之前的录音数据和新的录音数据。本申请实施例以保存之前的录音数据和新的录音数据为例进行举例说明。

5 在一些实施例中，电子设备还可以根据录音数据，确定录音数据对应的音频质量信息。换言之，电子设备还对自身的录音数据进行质量评价。如上所述音频质量信息可以包括音强信息、噪声声强信息、信噪比信息等一项或多项。

以本实施例的三个电子设备为例，音箱 201，电视机 202 和手机 203 可以分别对各自的录音数据进行质量评价，确定各自的录音数据对应的音频质量信息。

步骤 403、音箱 201，电视机 202 和手机 203 分别执行应答设备选择，确定应答设备，应答设备播放第一语音指令对应的应答语音。

10 其中，步骤 402 和步骤 403 的执行顺序不以序号大小作为限制，其还可以是其他执行顺序。例如，在开始录音的同时，执行应答设备选择。

15 本实施例的应答设备是指用于播放与用户输入的语音指令对应的应答语音。例如，应答设备播放第一语音指令对应的应答语音，即唤醒应答语音，如“我在”。而其他不作为应答设备的电子设备虽然唤醒了语音助手，但不播放与用户输入的语音指令对应的应答语音。

20 电子设备可以基于第一语音指令对应的音频质量信息，执行应答设备选择，确定应答设备。一种可实现方式，电子设备可以对接收到的第一语音指令进行质量评价，确定自身接收到的第一语音指令对应的音频质量信息，并广播自身接收到的第一语音指令对应的音频质量信息和自身的设备信息。电子设备接收到其他电子设备广播的自身接收到的第一语音指令对应的音频质量信息和自身的设备信息。电子设备根据所有电子设备的音频质量信息和设备信息，从中选择一个电子设备作为应答设备。例如，选择音频质量最好的电子设备作为应答设备。

25 结合步骤 402 中的示例，在音箱 201 检测到第一语音指令时，音箱 201 还可以对第一语音指令进行质量评价，确定音箱 201 接收到的第一语音指令对应的音频质量信息，并广播音箱 201 接收到的第一语音指令对应的音频质量信息和音箱 201 的设备信息。类似的处理方式，在电视机 202 检测到第一语音指令时，电视机 202 还可以对第一语音指令进行质量评价，确定电视机 202 接收到的第一语音指令对应的音频质量信息，并广播电视机 202 接收到的第一语音指令对应的音频质量信息和电视机 202 的设备信息。在手机 203 检测到第一语音指令时，手机 203 还可以对第一语音指令进行质量评价，确定手机 203 接收到的第一语音指令对应的音频质量信息，并广播手机 203 接收到的第一语音指令对应的音频质量信息和手机 203 的设备信息。这样，音箱 201 可以接收到电视机 202 和手机 203 的第一语音指令对应的音频质量信息和设备信息，音箱 201 根据音箱 201、电视机 202 和手机 203 的第一语音指令对应的音频质量信息和设备信息，在音箱 201、电视机 202 和手机 203 中选择一个电子设备作为应答设备。类似的，电视机 202 可以接收到音箱 201 和手机 203 的第一语音指令对应的音频质量信息和设备信息，电视机 202 根据音箱 201、电视机 202 和手机 203 的第一语音指令对应的音频质量信息和设备信息，在音箱 201、电视机 202 和手机 203 中选择一个电子设备作为应答设备。手机 203 可以接收到音箱 201 和电视机 202 的第一语音指令对应的音频质量信息和设备信息，手机 203 根据音箱 201、电视机 202 和手机 203 的第一语音指令对应的音频质量信息和设备信息，在音箱 201、电视机 202 和手机 203 中选择一个电子设备作为应答设备。

203 中选择一个电子设备作为应答设备。这里，以音箱 201、电视机 202 和手机 203 均确定音箱 201 为应答设备作为示例性说明。

例如，如图 4 所示，音箱 201 作为应答设备，播放唤醒应答语音，如“我在”。而电视机 202 和手机 203 不播放唤醒应答语音，但是如上步骤 402 所述电视机 202 和手机 203 各自的语音助手处于唤醒状态，并且可以录音。

需要说明的是，在执行应答设备选择过程中，还可以结合其他信息选择应答设备，例如各个电子设备的优先级等。另外，执行应答设备选择的具体实现方式也可以采用其他方式，本申请实施例不以上述方式作为限制。例如，可以采用用户上一次使用过程中的应答设备或者用户设置的应答设备作为本实施例的应答设备。

10 步骤 404、音箱 201 分别向电视机 202 和手机 203 调用拾音指令，该拾音指令用于指示返回录音数据。

在上述步骤 403 之后，音箱 201 开始执行分布式收音任务。应答设备可以分别向其他非应答设备调用拾音指令，该拾音指令用于指示非应答设备向应答设备返回录音数据。

15 结合上述步骤的示例，音箱 201 的语音助手可以调用电视机 202 的语音助手与音箱 201 的语音助手之间的接口，以向电视机 202 传递拾音指令。音箱 201 的语音助手可以调用手机 203 的语音助手与音箱 201 的语音助手之间的接口，以向音箱 201 传递拾音指令。该拾音指令可以携带应答设备的标识信息。应答设备的标识信息可以是应答设备的媒体访问控制(media access control, MAC)地址。例如，该拾音指令可以携带音箱 201 的标识信息，以指示电视机 202 向音箱 201 返回录音数据。

20 步骤 405、电视机 202 和手机 203 分别向音箱 201 发送录音数据。

应答设备接收其他非应答设备发送的录音数据。其他非应答设备在发送各自录音数据后，可以继续录音，并向应答设备发送新的录音数据。

25 结合上述步骤的示例，电视机 202 向音箱 201 发送电视机 202 的录音数据。手机 203 向音箱 201 发送手机 203 的录音数据。该录音数据可以包括上述第二语音指令。例如，该录音数据的内容为“播放歌曲 112222”。

一种可实现方式，音箱 201 对接收到的电视机 202 的录音数据进行质量评价，确定电视机 202 的录音数据对应的音频质量信息。音箱 201 对接收到的手机 203 的录音数据进行质量评价，确定手机 203 的录音数据对应的音频质量信息。

30 另一种可实现方式，音箱 201 还可以接收电视机 202 发送的电视机 202 的录音数据对应的音频质量信息。音箱 201 还可以接收手机 203 发送的手机 203 的录音数据对应的音频质量信息。

步骤 406、音箱 201 根据音频质量信息，在音箱 201、电视机 202 和手机 203 中确定最优收音设备，并根据最优收音设备的录音数据，播放第二语音指令对应的应答语音。

35 应答设备根据多个电子设备（包括自身和其他非应答设备）的录音数据对应的音频质量信息，从多个电子设备中选择一个最优收音设备，使用该最优收音设备的录音数据，进行 SE、ASR 等处理，以正确识别用户输入的语音指令，进而对用户输入的语音指令进行准确响应。其中，对用户输入的语音指令进行准确响应包括播放用户输入的语音指令对应的应答语音。在一些实施例中，对用户输入的语音指令进行准确响应还可以包括触发应答设备或其他非应答设备执行语音指令对应的事件。该事件可以是播放歌曲、播放视频、拨

打电话等。

需要说明的是，在一些实施例中，音箱 201 也可以将最优收音设备的录音数据发送给如图 1 所示的服务器 204，由服务器 204 使用该最优设备的录音数据，进行 SE、ASR 等处理，以正确识别用户输入的语音指令，进而对用户输入的语音指令进行准确响应。

5 例如，结合图 4 和图 5 所示，用户虽然距离手机 203 最近，但是由于用户在使用吹风机 205，该吹风机 205 会产生噪音，影响手机 203 的收音质量。本实施例的音箱 201 根据音箱 201、电视机 202 和手机 203 的录音数据的音频质量信息，在音箱 201、电视机 202 和手机 203 中确定最优收音设备为音箱 201。例如，如图 5 所示，音箱 201 可以播放应答语音“将在这里为您播放歌曲 112222”。歌曲 112222 的多媒体资源可以由服务器 204 或手机 203 提供。

10 可选的，另一种可实现方式，音箱 201 还可以根据自身的录音数据和最优收音设备的录音数据，播放第二语音指令对应的应答语音。例如，音箱 201 可以将自身的录音数据和最优收音设备的录音数据进行拼接，基于拼接后的录音数据，播放第二语音指令对应的应答语音。

15 可选的，在步骤 406 之后，还可以再次执行步骤 404 至步骤 406，以采用类似的方式对新的录音数据进行处理，以正确识别用户输入的新的语音指令，进而对用户输入的新的语音指令进行准确响应。

可选的，在一些实施例中，本申请实施例的语音控制方法还可以通过如下步骤对新的录音数据进行处理。

20 步骤 407、音箱 201 向电视机 202 和手机 203 分别发送停止录音指令。

应答设备向其他非应答设备发送停止录音指令，该停止录音指令用于指示停止录音，并丢弃录音数据。

步骤 408、电视机 202 和手机 203 分别停止录音，并丢弃录音数据。

其他非应答设备基于停止录音指令停止录音，以减少功率消耗。

25 例如，音箱 201 向电视机 202 和手机 203 分别发送停止录音指令。电视机 202 和手机 203 分别停止录音，并丢弃录音数据。例如，丢弃第二语音指令对应的录音数据。之后，由音箱 201 接收用户输入的新的语音指令。例如，以用户说出的第三语音指令为“换一首歌”为例。音箱 201 对第三语音指令进行录音，生成录音数据，该录音数据的内容即为“换一首歌”。音箱 201 使用该录音数据，进行 SE、ASR 等处理，以正确识别用户输入的语音指令，进而对用户输入的语音指令进行准确响应。例如，音箱 201 可以播放应答语音“好的，为您切换歌曲”，播放切换后的歌曲。

30 需要说明的是，本实施例以应答设备和最优收音设备均为音箱 201 为例进行举例说明，应答设备和最优收音设备可以是同一设备，也可以是不同设备，例如，应答设备为音箱 201，最优收音设备为电视机 202，本申请实施例不以上述举例作为限定。当应答设备和最优收音设备是不同设备时，应答设备可以调用最优收音设备的录音数据。

35 在一些实施例中，当应答设备接收到的语音指令用于关闭语音助手时，应答设备可以停止调用其他非应答设备的录音数据，然后停止自身分布式收音任务，丢弃录音数据。

本申请实施例，在多个电子设备分别接收到用户输入的第一语音指令时，多个电子设备分别唤醒各自的语音助手，并开始录音，该第一语音指令用于唤醒电子设备的语音助手。

在多个电子设备协商确定应答设备之后，应答设备可以根据各个电子设备的录音数据，确定最优收音设备，根据最优收音设备的录音数据，播放第二语音指令对应的应答语音。与通过中心设备调用后开始录音的方式不同，本实施例通过从电子设备唤醒后直接开始录音，不再依赖中心设备调用，实现了去中心化的协同收音方式。在没有确定应答设备之前已经
5 开始录音，使用录音数据，进行 SE、ASR 等处理，有效消除设备之间的通信时延，从而解决多设备场景中因时延导致的语音控制的丢帧问题。

通过使用最优收音设备的录音数据，进行 SE、ASR 等处理，可以正确识别用户输入的语音指令，进而对用户输入的语音指令进行准确响应，提升语音控制的准确率。

通过将唤醒和收音两个过程结合起来，使得音频录制提前开始，电子设备可以对自身的录音数据进行质量评价，进而可以加快电子设备的音频评价的速度，缩短后续决策最优收音设备所需时间，加快语音控制方法处理流程，提升语音控制响应速度。
10

需要说明的是，上述图 3 实施例以通过唤醒词唤醒语音助手，并开始录音为例进行举例说明，本申请实施例不以此作为限制，本申请实施例也可以没有上述唤醒过程，通过其
15 他方式触发电子设备录音，并基于多设备协同收音，提升语音控制的准确率。举例而言，该其他方式可以是电子设备检测到人声，或者电子设备检测到特定用户的声音等，本申请实施例不一一举例说明。对于没有上述唤醒过程触发电子设备录音，实现语音控制方法的具体实现方式，与图 3 所示实施例类似，例如，开始录音之后，应答设备调用拾音指令，非应答设备返回录音数据，应答设备根据各个电子设备的录音数据，确定最优收音设备，
20 根据最优收音设备的录音数据，播放第二语音指令对应的应答语音。其实现原理和技术效果可以参见上述实施例的解释说明。

图 6 为本申请实施例提供的另一种语音控制方法的流程示意图。本实施例以如图 1 所示的三个电子设备，音箱 201，电视机 202 和手机 203，且应答设备为音箱 201 为例进行举
25 例说明。本实施例为电子设备唤醒之后的非第一次调用，例如，语音助手的多轮对话的第二次调用、第三次调用、第四次调用等。如图 6 所示，本实施例的方法可以包括：

步骤 701、音箱 201 分别向电视机 202 和手机 203 调用多轮对话暂停指令，该多轮对话暂停指令用于指示多轮对话暂时停止。

应答设备在预设时间段内未检测到用户输入的新的语音指令，即用户输入的语音指令
30 之间存在时间间隔。应答设备检测到该时间间隔，触发多轮对话暂停操作。应答设备可以分别向其他非应答设备调用多轮对话暂停指令，该多轮对话暂停指令用于指示多轮对话暂时停止。

例如，音箱 201 的语音助手可以调用电视机 202 的语音助手与音箱 201 的语音助手之间的接口，以向电视机 202 传递多轮对话暂时停止指令。音箱 201 的语音助手可以调用手机
35 203 的语音助手与音箱 201 的语音助手之间的接口，以向音箱 201 传递多轮对话暂时停止指令。音箱 201 删除之前保存的录音数据，并继续保持录音。

步骤 702、电视机 202 和手机 203 分别删除各自保存的录音数据，并分别保持录音。

电视机 202 和手机 203 分别删除调用多轮对话暂停指令之前的录音数据，并继续保持录音。

步骤 703、音箱 201，电视机 202 和手机 203 分别接收用户输入的第四语音指令，并分别对该第四语音指令进行录音，生成各自的录音数据。

在一些实施例中，音箱 201，电视机 202 和手机 203 还可以分别对各自接收到的录音数据进行质量评价，确定各自接收到的录音数据对应的音频质量信息。

5 例如，如图 7 所示，以用户说出的第四语音指令可以是“播放电影 333333”为例。音箱 201，电视机 202 和手机 203 分别对第四语音指令进行录音，生成各自的录音数据，该录音数据的内容即为“播放电影 333333”。

步骤 704、音箱 201 分别向电视机 202 和手机 203 调用拾音指令，该拾音指令用于指示返回录音数据。

10 在上述步骤 703 之后，音箱 201 开始重新执行分布式收音任务。应答设备可以分别向其他非应答设备调用拾音指令，该拾音指令用于指示非应答设备向应答设备返回录音数据。

步骤 705、电视机 202 和手机 203 分别向音箱 201 发送录音数据。

结合上述步骤的示例，电视机 202 向音箱 201 发送电视机 202 的录音数据。手机 203 向音箱 201 发送手机 203 的录音数据。例如，该录音数据的内容为“播放电影 333333”。

15 步骤 706、音箱 201 根据音频质量信息，在音箱 201、电视机 202 和手机 203 中确定最优收音设备，并根据最优收音设备的录音数据，对第四语音指令进行响应。

20 应答设备根据多个电子设备（包括自身和其他非应答设备）的录音数据对应的音频质量信息，从多个电子设备中选择一个最优收音设备，使用该最优收音设备的录音数据，进行 SE、ASR 等处理，以正确识别用户输入的语音指令，进而对用户输入的语音指令进行准确响应。其中，对用户输入的语音指令进行准确响应包括播放用户输入的语音指令对应的应答语音。在一些实施例中，对用户输入的语音指令进行准确响应还可以包括触发应答设备或其他非应答设备执行语音指令对应的事件。该事件可以是播放歌曲、播放视频、拨打电话等。

25 例如，结合图 7 所示，本实施例的音箱 201 根据音箱 201、电视机 202 和手机 203 的录音数据的音频质量信息，在音箱 201、电视机 202 和手机 203 中确定最优收音设备为音箱 201。例如，如图 7 所示，音箱 201 可以播放应答语音“将在电视机上播放播放电影 333333”，并由电视机 202 开始播放电影 333333。

30 之后如果用户再次触发多轮对话暂停，则可以重新执行上述步骤 701 至步骤 706。在这个过程中，最优收音设备可以发生变化。例如，结合图 7 所示示例，电视机开始播放电影之后，以用户说出的第五语音指令可以是“声音小点”为例。音箱 201，电视机 202 和手机 203 分别对第五语音指令进行录音，生成各自的录音数据，该录音数据的内容即为“声音小点”。之后，通过上述步骤所涉及的流程，在音箱 201、电视机 202 和手机 203 中确定最优收音设备为电视机 202。音箱 201 可以基于电视机 202 的录音数据，对第五语音指令进行响应。本实施例可以在用户的环境发生改变时，根据录音效果选择不同的设备进行收音。例如，电视机 202 开始播放电影后，用户家中出现强烈自噪（如电影播放过程中产生的声音），此时音箱 201 的语音助手的收音也会混入电视机播放的语句，如果使用音箱 201 的录音数据会造成 ASR 识别错误，本实施例的语音控制方法通过动态调用电视机进行收音并完成回声消除，可以提升 ASR 识别准确率，进而对用户输入的语音指令进行准确响应，提升语音控制的准确率。

需要说明的是，上述图 3 和图 6 所示实施例以应答设备根据音频质量信息，选择最优收音设备，根据最优收音设备的录音数据，应答第二语音指令为例进行举例说明，其还可以是其他处理方式，例如，应答设备直接根据接收的录音数据，或者根据接收的录音数据和自身的录音数据，应答第二语音指令。其中，根据接收的录音数据和自身的录音数据，
5 应答第二语音指令的具体实现方式可以是，应答设备对接收到的录音数据的音频内容信息和自身的录音数据的音频内容信息进行拼接，基于拼接后的音频内容信息，应答第二语音指令。例如，用户说出语音信号“播放歌曲 112222”，应答设备仅识别到语音信号“2222”，应答设备的录音数据的音频内容信息用于表示语音信号“2222”，应答设备接收到其他设备的录音数据的音频内容信息用于表示语音信号“播放歌曲 112”，应答设备可以对二者
10 进行拼接，得到拼接后的音频内容信息，拼接后的音频内容信息用于表示语音信号“播放歌曲 112222”。

图 8 为本申请实施例的一种语音控制装置的结构示意图。如图 8 所示，该装置可以应用于语音控制系统的电子设备（如上述第一电子设备 201），该语音控制系统还可以至少
15 包括第二电子设备（如第二电子设备 202 或第三电子设备 203），该装置可以包括：收发模块 81 和处理模块 82。举例而言，收发模块 81 具体可以是如图 2 所示实施例的移动通信模块 150 和/或无线通信模块 160。处理模块 82 可以是如图 2 所示实施例的处理器 110。

收发模块 81 用于接收用户输入的第一语音指令，处理模块 82 用于应答第一语音指令。收发模块 81 还用于接收第二电子设备发送的第二电子设备的录音数据，第二电子设备的
20 录音数据包括第二电子设备录制用户输入的第二语音指令的录音数据。处理模块 82 还用于根据第一电子设备的录音数据和/或第二电子设备的录音数据，应答第二语音指令，第一电子设备的录音数据包括第一电子设备录制用户输入的第二语音指令的录音数据。

在一些实施例中，收发模块 81 还用于向第二电子设备调用拾音指令，拾音指令用于第二
25 电子设备返回第二电子设备的录音数据。

在一些实施例中，处理模块 82 还用于在第一电子设备接收用户输入的第一语音指令
时或之后，录音，录音用于录制用户输入的第二语音指令。

在一些实施例中，第一语音指令用于唤醒第一电子设备和/或第二电子设备的语音控制
30 功能。

在一些实施例中，处理模块 82 还用于根据第一电子设备接收到的第一语音指令的音
35 频质量信息和第二电子设备接收到的第一语音指令的音频质量信息，确定第一电子设备为语音控制系统的应答设备。

在一些实施例中，处理模块 82 还用于在第一电子设备应答第一语音指令之后，在录制
到用户输入的第二语音指令之前，在第一电子设备录音过程中，在预设时间段内未检测到
用户输入的第二语音指令，删除已保存的录音数据，并继续录音。收发模块 81 还用于向第
40 二电子设备调用多轮对话暂停指令，多轮对话暂停指令用于指示多轮对话暂时停止。

在一些实施例中，收发模块 81 还用于接收第二电子设备发送的第二电子设备的录音
数据的音频质量信息。

在一些实施例中，处理模块 82 用于根据第一电子设备的录音数据的音频质量信息和
第二电子设备的录音数据的音频质量信息，从语音控制系统中确定最优收音设备。当最优

收音设备为第一电子设备时，根据第一电子设备的录音数据，应答第二语音指令。当最优收音设备为第二电子设备时，根据第二电子设备的录音数据，或者，根据第二电子设备的录音数据和第一电子设备的录音数据，应答第二语音指令。其中，音频质量信息用于表示录音数据的音频质量。

5 在一些实施例中，处理模块 82 用于根据第一电子设备的录音数据的音频内容信息和/或第二电子设备的录音数据的音频内容信息，应答第二语音指令。其中，音频内容信息用于表示录音数据的音频内容。

本申请实施例的语音控制装置可以用于执行上述方法实施例中应答设备（如音箱 201）的步骤，其技术原理和技术效果可以参见上述方法实施例的解释说明，此处不再赘述。

10

图 9 为本申请实施例的一种语音控制装置的结构示意图。如图 9 所示，该装置可以应用于语音控制系统的电子设备（如第二电子设备 202 或第三电子设备 203），该语音控制系统还可以至少包括第一电子设备（如第一电子设备 201），该装置可以包括：收发模块 91 和处理模块 92。举例而言，收发模块 91 具体可以是如图 2 所示实施例的移动通信模块 150 和/或无线通信模块 160。处理模块 92 可以是如图 2 所示实施例的处理器 110。

15

处理模块 92 用于录音，并保存录音数据，录音用于录制用户输入的第二语音指令。收发模块 91 用于向第一电子设备发送第二电子设备的录音数据，第二电子设备的录音数据包括第二电子设备录制用户输入的第二语音指令的录音数据，录音数据用于第一电子设备在应答第一语音指令之后，应答第二语音指令。

20

一种可能的设计中，收发模块 91 还用于接收第一电子设备调用拾音指令，拾音指令用于第二电子设备返回第二电子设备的录音数据。

一种可能的设计中，处理模块 92 用于在第二电子设备接收到用户输入的第一语音指令时或之后，录音。

25

一种可能的设计中，处理模块 92 还用于根据第二电子设备接收到的第一语音指令的音频质量信息和第一电子设备接收到的第一语音指令的音频质量信息，确定第一电子设备为语音控制系统的应答设备。

30

一种可能的设计中，处理模块 92 还用于在第一电子设备应答第一语音指令之后，在第二电子设备录音过程中，通过收发模块 91 接收第二电子设备调用多轮对话暂停指令，多轮对话暂停指令用于指示多轮对话暂时停止。处理模块 92 还用于删除已保存的录音数据，并继续录音。

一种可能的设计中，收发模块 91 还用于向第一电子设备发送第二电子设备的录音数据的音频质量信息。

35

本申请实施例的语音控制装置可以用于执行上述方法实施例中任意非应答设备（如电视 202 或手机 203）的步骤，其技术原理和技术效果可以参见上述方法实施例的解释说明，此处不再赘述。

本申请实施例另一些实施例还提供了一种电子设备，用于执行以上各方法实施例中电子设备的方法。如图 10 所示，该电子设备可以包括：麦克风 1001、一个或多个处理器 1002；一个或多个存储器 1003；上述各器件可以通过一个或多个通信总线 1005 连接。其中上述

存储器 1003 中存储一个或多个计算机程序 1004，一个或多个处理器 1002 用于执行一个或多个计算机程序 1004，该一个或多个计算机程序 1004 包括指令，上述指令可以用于执行上述方法实施例中任一电子设备执行的各个步骤。该电子设备可以是上述任一形式的电子设备，例如，智能手机、智能手表等。

5 当然，图 10 所示的电子设备还可以包含如显示屏等其他器件，本申请实施例对此不做任何限制。当其包括其他器件时，具体可以为图 2 所示的电子设备。

本申请实施例的电子设备可以用于执行上述任一方法实施例中电子设备的步骤，其技术原理和技术效果可以参见上述方法实施例的解释说明，此处不再赘述。

10 本申请实施例另一些实施例还提供一种计算机存储介质，该计算机存储介质可包括计算机指令，当该计算机指令在电子设备上运行时，使得该电子设备执行上述方法实施例中电子设备执行的各个步骤。

本申请实施例另一些实施例还提供一种计算机程序产品，当该计算机程序产品在计算机上运行时，使得该计算机执行上述方法实施例中电子设备执行的各个步骤。

15 本申请实施例还提供一种语音控制系统，该语音控制系统可以至少包括：第一电子设备和第二电子设备，其中，第一电子设备可以采用图 8 或图 10 所示实施例的结构，第二电子设备可以采用图 9 或图 10 所示实施例的结构，其对应地，可以执行上述任一方法实施例的技术方案，其实现原理和技术效果类似，此处不再赘述。

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

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

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

35 以上各实施例中提及的处理器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，上述方法实施例的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。处理器可以是通用处理器、数字信号处理器 (digital signal processor, DSP)、特定应用集成电路 (application-specific integrated circuit, ASIC)、现场可编程门阵列 (field programmable gate array, FPGA) 或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。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。本申请实施例公开的方法的步骤可以直接体现为硬件编码处理器执行完成，或者用编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，处理器读取存储器中的信息，结合其硬件完成上述方法的步骤。

上述各实施例中提及的存储器可以是易失性存储器或非易失性存储器，或可包括易失性和非易失性存储器两者。

5 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

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过程，在此不再赘述。

10 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，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，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。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以
15 是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是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。

20 另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。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或者该技术方案的部分可以以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中，包括若干指令用以使得一台计算机设备（个人计算机，服务器，或者网络
25 设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（read-only memory，ROM）、随机存取存储器（random access memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

30 以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以所述权利要求的保护范围为准。

权 利 要 求 书

1.一种语音控制方法，其特征在于，应用于语音控制系统，所述语音控制系统至少包括具备语音控制功能的第一电子设备和第二电子设备，所述方法包括：

5 第一电子设备和第二电子设备分别接收用户输入的第一语音指令，所述第一电子设备应答所述第一语音指令；

所述第二电子设备录音，并保存录音数据，所述录音用于录制用户输入的第二语音指令；

所述第二电子设备向所述第一电子设备发送所述第二电子设备的录音数据；

10 所述第一电子设备根据所述第一电子设备的录音数据和/或所述第二电子设备的录音数据，应答所述第二语音指令；

其中，所述第一电子设备的录音数据包括所述第一电子设备录制用户输入的所述第二语音指令的录音数据。

2.根据权利要求1所述的方法，其特征在于，所述方法还包括：

15 所述第一电子设备向所述第二电子设备调用拾音指令，所述拾音指令用于所述第二电子设备返回所述第二电子设备的录音数据。

3.根据权利要求1或2所述的方法，其特征在于，所述第二电子设备录音，包括：

在所述第二电子设备接收到用户输入的第一语音指令时或之后，所述第二电子设备录音。

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：

20 在所述第一电子设备接收到用户输入的第一语音指令时或之后，所述第一电子设备录音，所述录音用于录制用户输入的第二语音指令。

5.根据权利要求1至4任一项所述的方法，其特征在于，所述第一语音指令用于唤醒所述第一电子设备和/或所述第二电子设备的语音控制功能。

6.根据权利要求1至5任一项所述的方法，其特征在于，所述方法还包括：

25 所述第一电子设备和所述第二电子设备分别根据各自接收到的第一语音指令的音频质量信息，确定所述第一电子设备为所述语音控制系统的应答设备。

7.根据权利要求1至6任一项所述的方法，其特征在于，在所述第一电子设备应答所述第一语音指令之后，在录制到用户输入的第二语音指令之前，所述方法还包括：

30 在所述第一电子设备和所述第二电子设备录音过程中，所述第一电子设备在预设时间段内未检测到用户输入的第二语音指令，所述第一电子设备删除已保存的录音数据，并继续录音；所述第一电子设备向所述第二电子设备调用多轮对话暂停指令，所述多轮对话暂停指令用于指示多轮对话暂时停止；所述第二电子设备删除已保存的录音数据，并继续录音。

8.根据权利要求1至7任一项所述的方法，其特征在于，所述方法还包括：

35 所述第一电子设备接收所述第二电子设备发送的所述第二电子设备的录音数据的音频质量信息。

9.根据权利要求1至8任一项所述的方法，其特征在于，所述第一电子设备根据所述第一电子设备的录音数据和/或所述第二电子设备的录音数据，应答所述第二语音指令，包括：

所述第一电子设备根据所述第一电子设备的录音数据的音频质量信息和所述第二电子设备的录音数据的音频质量信息，从所述语音控制系统中确定最优收音设备；

5 当所述最优收音设备为第一电子设备时，所述第一电子设备根据所述第一电子设备的录音数据，或者，根据所述第一电子设备的录音数据和所述第二电子设备的录音数据，应答所述第二语音指令；

当所述最优收音设备为第二电子设备时，所述第一电子设备根据所述第二电子设备的录音数据，或者，根据所述第二电子设备的录音数据和所述第一电子设备的录音数据，应答所述第二语音指令；

其中，所述音频质量信息用于表示所述录音数据的音频质量。

10 10.根据权利要求 1 至 8 任一项所述的方法，其特征在于，所述第一电子设备根据所述第一电子设备的录音数据和/或所述第二电子设备的录音数据，应答所述第二语音指令，包括：

所述第一电子设备根据所述第一电子设备的录音数据的音频内容信息和/或所述第二电子设备的录音数据的音频内容信息，应答所述第二语音指令；

15 其中，所述音频内容信息用于表示所述录音数据的音频内容。

11.一种语音控制方法，其特征在于，应用于语音控制系统的第一电子设备，所述语音控制系统还至少包括第二电子设备，所述方法包括：

20 所述第一电子设备接收用户输入的第一语音指令，所述第一电子设备应答所述第一语音指令；

所述第一电子设备接收所述第二电子设备发送的所述第二电子设备的录音数据，所述第二电子设备的录音数据包括所述第二电子设备录制用户输入的第二语音指令的录音数据；

25 所述第一电子设备根据所述第一电子设备的录音数据和/或所述第二电子设备的录音数据，应答所述第二语音指令，所述第一电子设备的录音数据包括所述第一电子设备录制用户输入的所述第二语音指令的录音数据。

12.根据权利要求 11 所述的方法，其特征在于，所述方法还包括：

所述第一电子设备向所述第二电子设备调用拾音指令，所述拾音指令用于所述第二电子设备返回所述第二电子设备的录音数据。

30 13.根据权利要求 12 所述的方法，其特征在于，所述方法还包括：

在所述第一电子设备接收用户输入的所述第一语音指令时或之后，所述第一电子设备录音，所述录音用于录制用户输入的所述第二语音指令。

14.根据权利要求 11 至 13 任一项所述的方法，其特征在于，所述第一语音指令用于唤醒所述第一电子设备和/或所述第二电子设备的语音控制功能。

35 15.根据权利要求 11 至 14 任一项所述的方法，其特征在于，所述方法还包括：

所述第一电子设备根据所述第一电子设备接收到的第一语音指令的音频质量信息和所述第二电子设备接收到的第一语音指令的音频质量信息，确定所述第一电子设备为所述语音控制系统的应答设备。

16.根据权利要求 11 至 15 任一项所述的方法，其特征在于，在所述第一电子设备应答

所述第一语音指令之后，在录制到用户输入的第二语音指令之前，所述方法还包括：

在所述第一电子设备录音过程中，所述第一电子设备在预设时间段内未检测到用户输入的第二语音指令，所述第一电子设备删除已保存的录音数据，并继续录音；所述第一电子设备向所述第二电子设备调用多轮对话暂停指令，所述多轮对话暂停指令用于指示多轮对话暂时停止。

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

所述第一电子设备接收所述第二电子设备发送的所述第二电子设备的录音数据的音频质量信息。

18.根据权利要求 11 至 17 任一项所述的方法，其特征在于，所述第一电子设备根据所述所述第一电子设备的录音数据和/或所述第二电子设备的录音数据，应答所述第二语音指令，包括：

所述第一电子设备根据所述第一电子设备的录音数据的音频质量信息和所述第二电子设备的录音数据的音频质量信息，从所述语音控制系统中确定最优收音设备；

当所述最优收音设备为第一电子设备时，所述第一电子设备根据所述第一电子设备的录音数据，应答所述第二语音指令；

当所述最优收音设备为第二电子设备时，所述第一电子设备根据所述第二电子设备的录音数据，或者，根据所述第二电子设备的录音数据和所述第一电子设备的录音数据，应答所述第二语音指令；

其中，所述音频质量信息用于表示所述录音数据的音频质量。

19.根据权利要求 11 至 17 任一项所述的方法，其特征在于，所述第一电子设备根据所述所述第一电子设备的录音数据和/或所述第二电子设备的录音数据，应答所述第二语音指令，包括：

所述第一电子设备根据所述第一电子设备的录音数据的音频内容信息和/或所述第二电子设备的录音数据的音频内容信息，应答所述第二语音指令；

其中，所述音频内容信息用于表示所述录音数据的音频内容。

20.一种语音控制方法，其特征在于，应用于语音控制系统的第二电子设备，所述语音控制系统还至少包括第一电子设备，所述方法包括：

所述第二电子设备录音，并保存录音数据，所述录音用于录制用户输入的第二语音指令；

所述第二电子设备向所述第一电子设备发送所述第二电子设备的录音数据，所述第二电子设备的录音数据包括所述第二电子设备录制用户输入的第二语音指令的录音数据，所述录音数据用于所述第一电子设备在应答第一语音指令之后，应答所述第二语音指令。

21.根据权利要求 20 所述的方法，其特征在于，所述方法还包括：

所述第二电子设备接收所述第一电子设备调用拾音指令，所述拾音指令用于所述第二电子设备返回所述第二电子设备的录音数据。

22.根据权利要求 20 或 21 所述的方法，其特征在于，所述第二电子设备录音，包括：在所述第二电子设备接收到用户输入的第一语音指令时或之后，所述第二电子设备录音。

23.根据权利要求 20 至 22 任一项所述的方法，其特征在于，所述方法还包括：

所述第二电子设备根据所述第二电子设备接收到的第一语音指令的音频质量信息和所述第一电子设备接收到的第一语音指令的音频质量信息，确定所述第一电子设备为所述语音控制系统的应答设备。

5 24.根据权利要求 20 至 23 任一项所述的方法，其特征在于，在所述第一电子设备应答所述第一语音指令之后，所述方法还包括：

在所述第二电子设备录音过程中，所述第二电子设备接收所述第二电子设备调用多轮对话暂停指令，所述多轮对话暂停指令用于指示多轮对话暂时停止；所述第二电子设备删除已保存的录音数据，并继续录音。

10 25.根据权利要求 20 至 24 任一项所述的方法，其特征在于，所述方法还包括：

所述第二电子设备向所述第一电子设备发送所述第二电子设备的录音数据的音频质量信息。

26.一种电子设备，其特征在于，包括：一个或多个处理器和存储器；

15 所述存储器与所述一个或多个处理器耦合，所述存储器用于存储计算机程序代码，所述计算机程序代码包括计算机指令，当所述一个或多个处理器执行所述计算机指令时，所述电子设备执行如权利要求 11 至 19 中任一项所述的语音控制方法，或者，所述电子设备执行如权利要求 20 至 25 任一项所述的语音控制方法。

20 27.一种计算机存储介质，其特征在于，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权利要求 11 至 19 中任一项所述的语音控制方法，或者，使得所述电子设备执行如权利要求 20 至 25 任一项所述的语音控制方法。

28.一种计算机程序产品，其特征在于，当所述计算机程序产品在计算机上运行时，使得所述计算机执行如权利要求 11 至 19 中任一项所述的语音控制方法，或者，使得所述计算机执行如权利要求 20 至 25 任一项所述的语音控制方法。

25 29.一种语音控制系统，其特征在于，所述语音控制系统至少包括具备语音控制功能的第一电子设备和第二电子设备，所述语音控制系统用于执行如权利要求 1 至 10 任一项所述的语音控制方法。

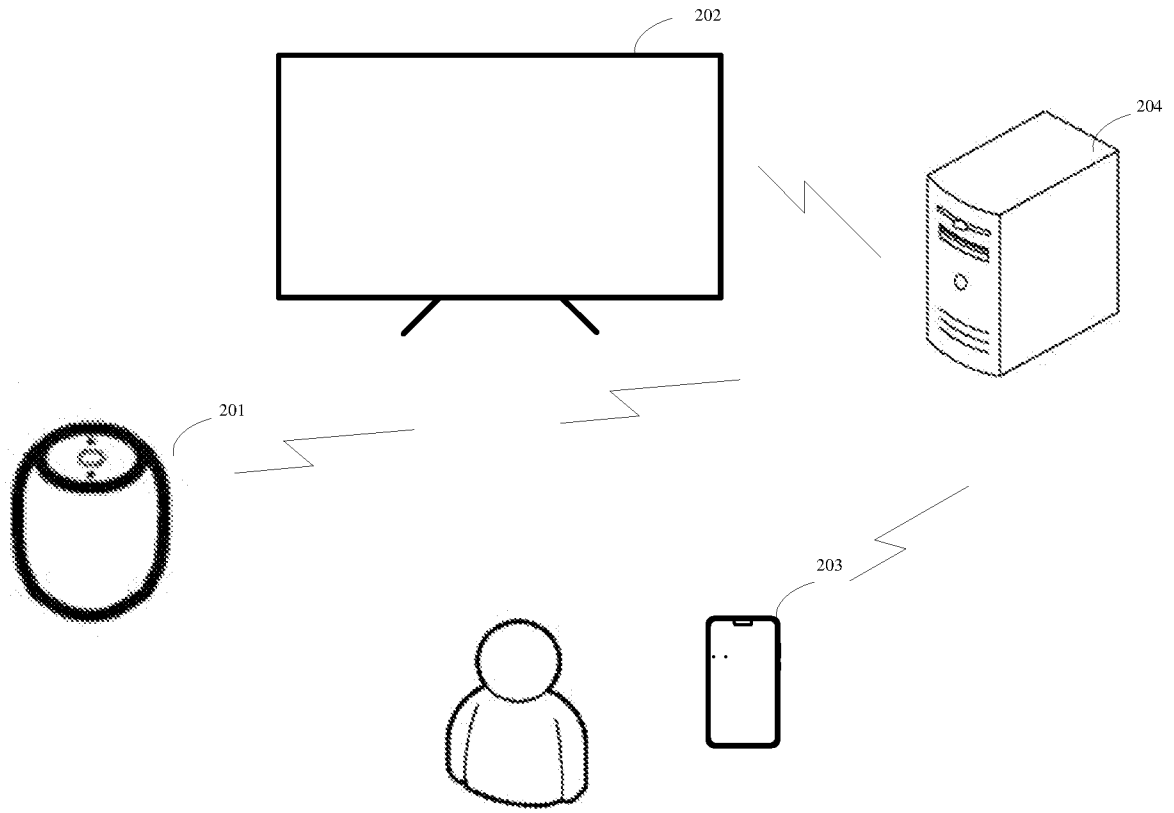


图 1

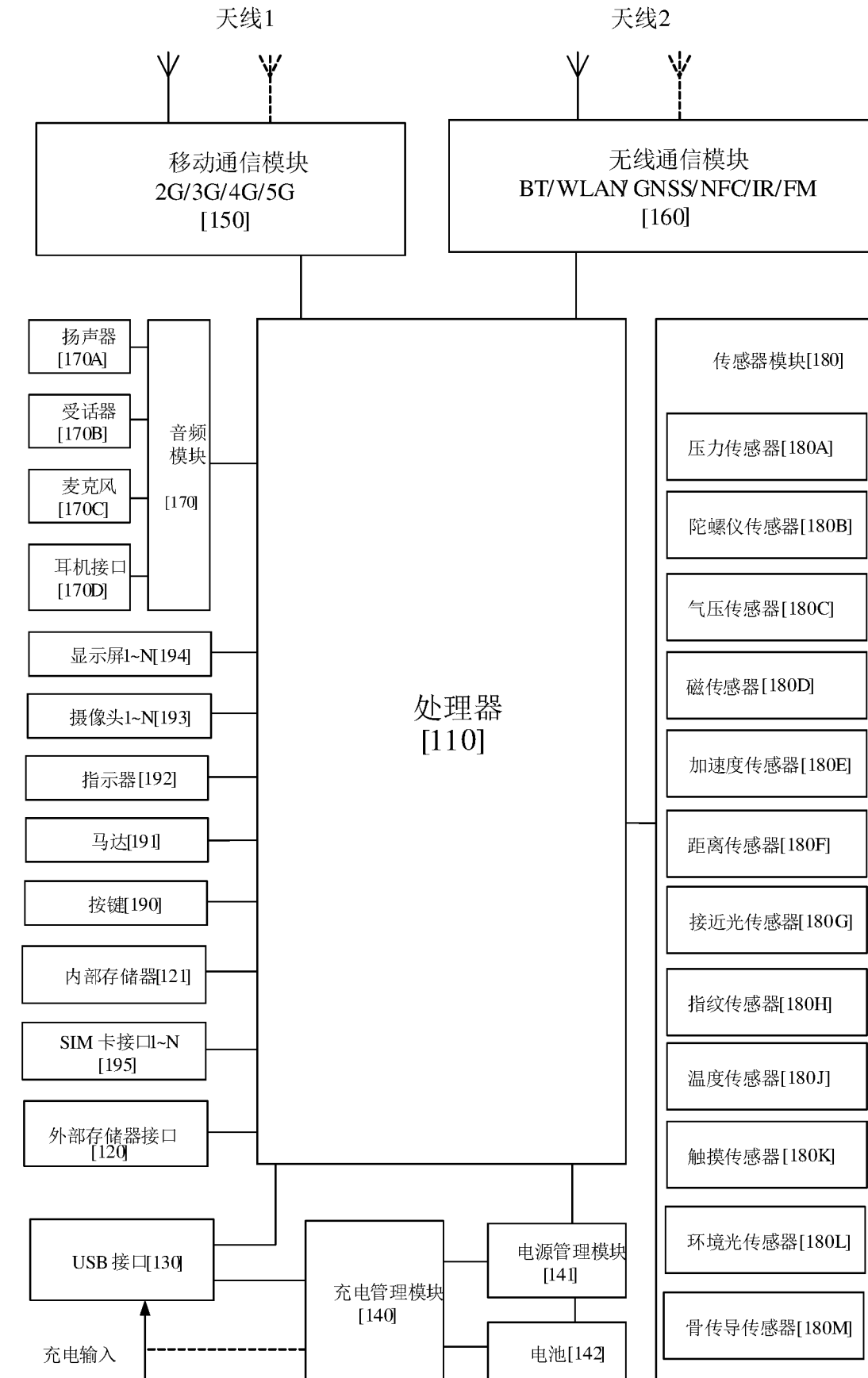


图 2

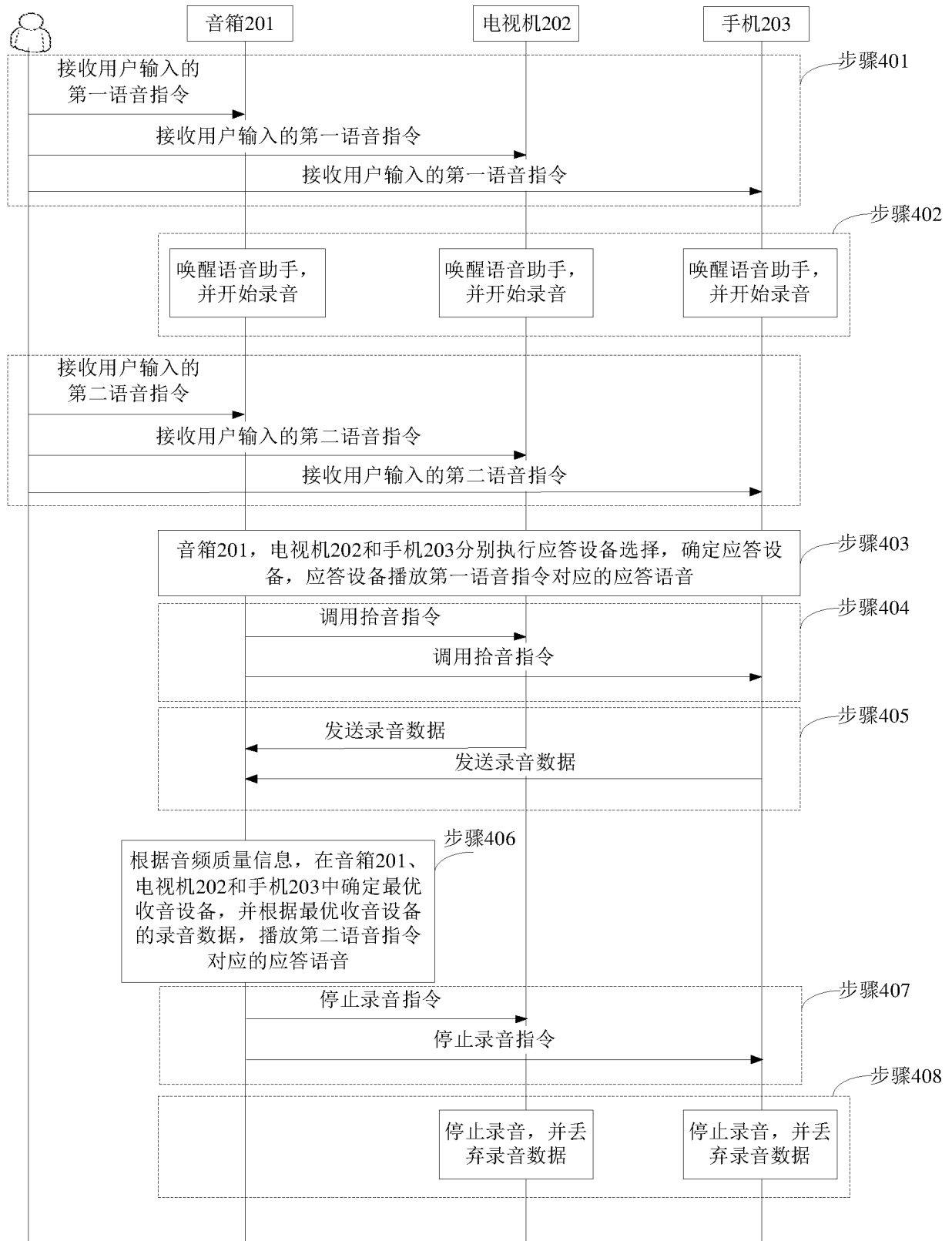


图 3

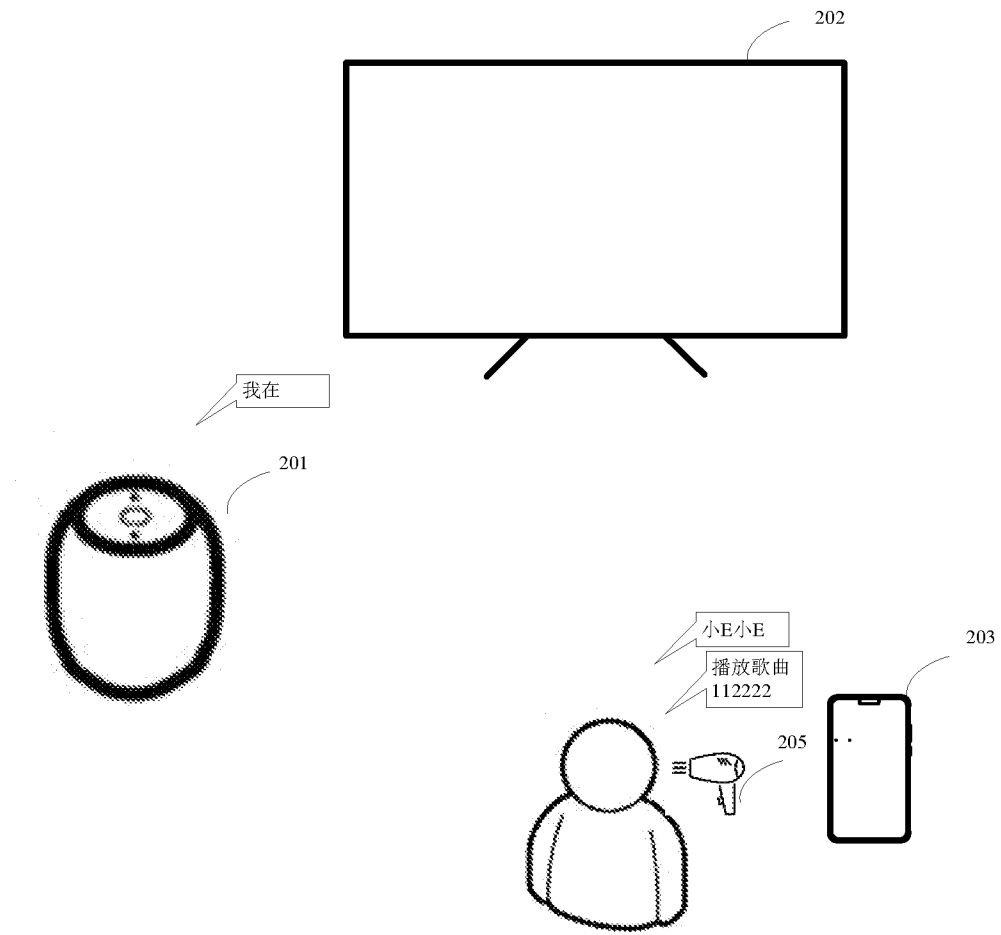


图 4

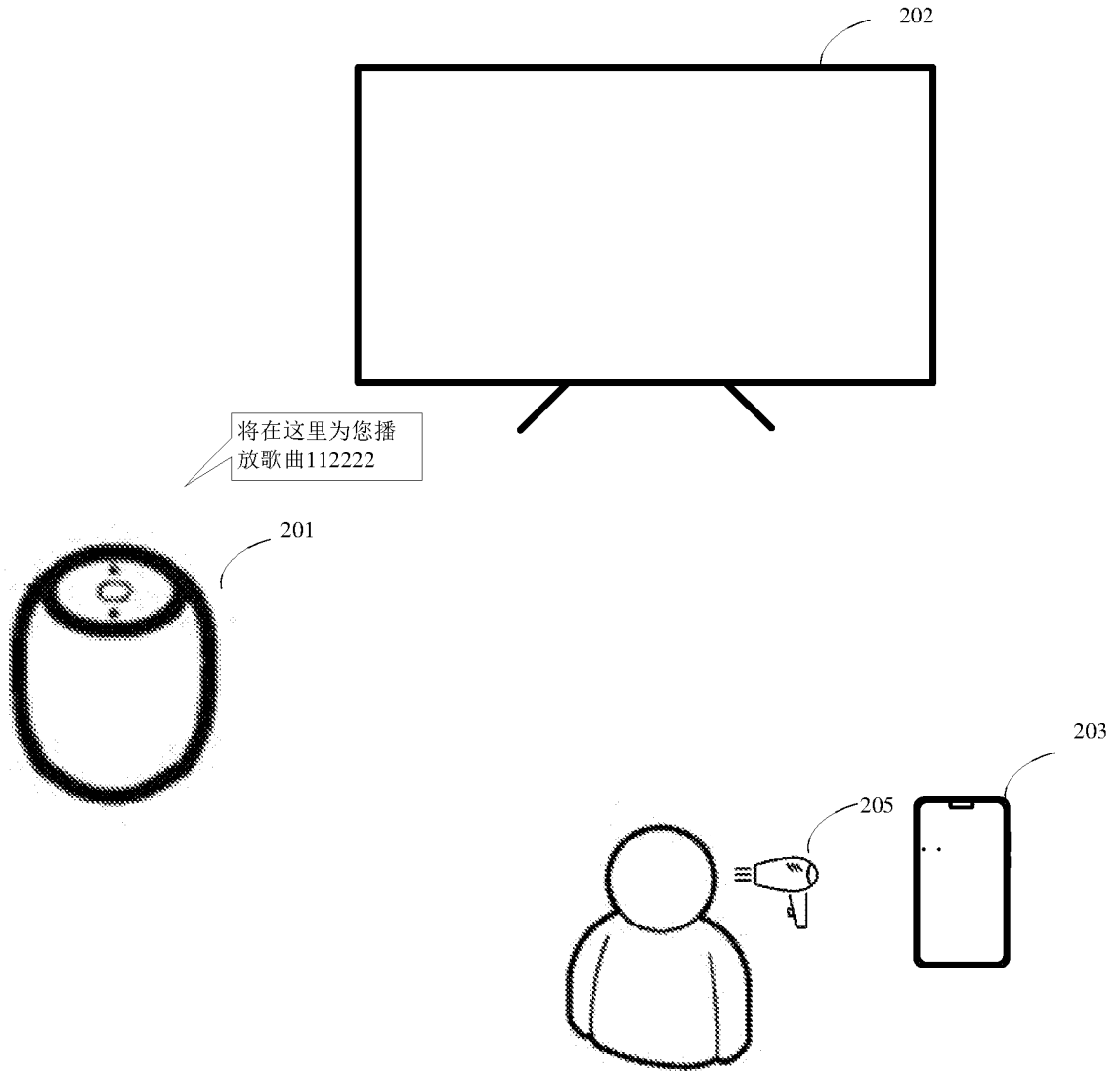


图 5

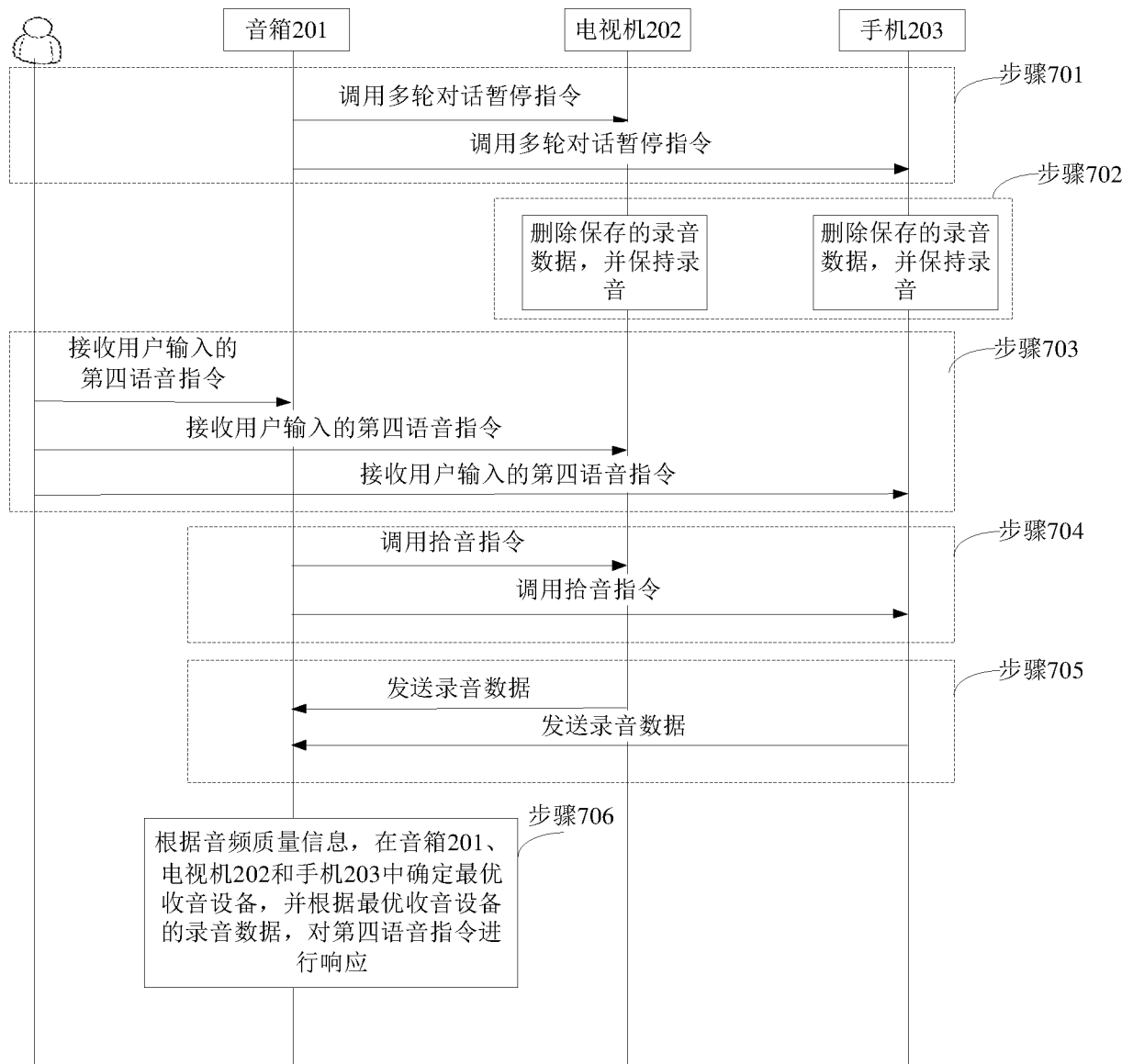


图 6

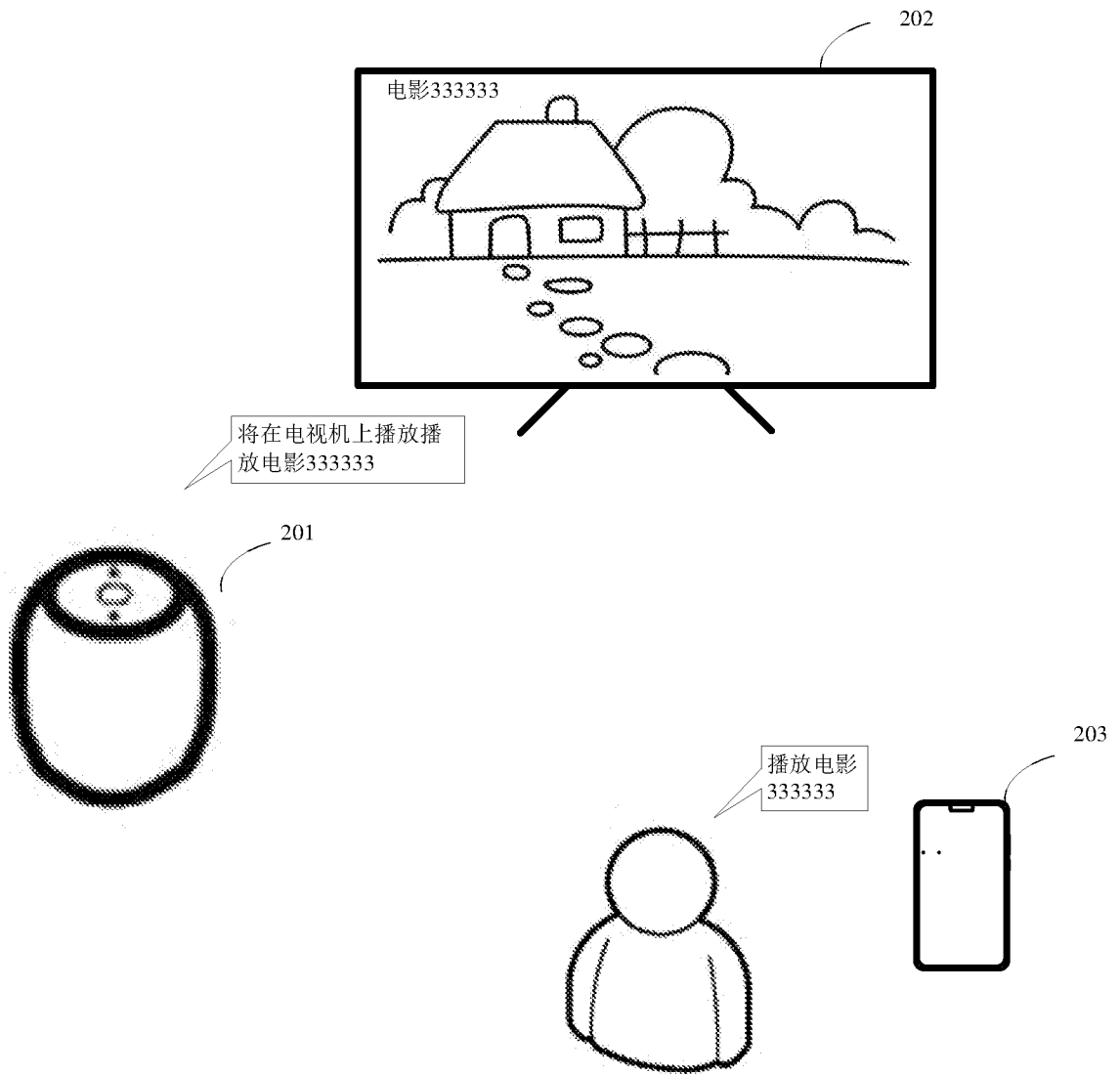


图 7

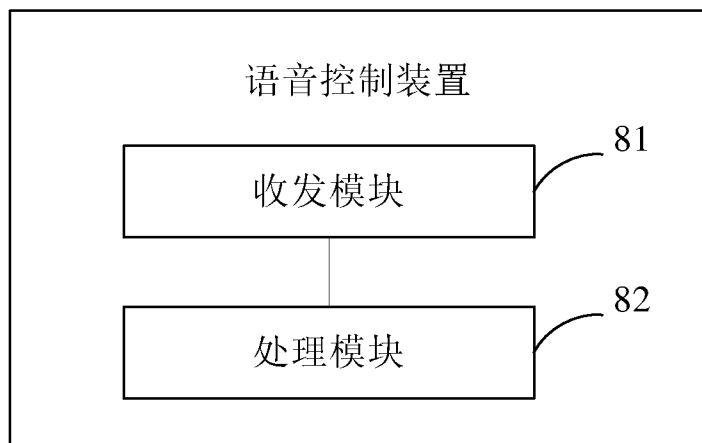


图 8

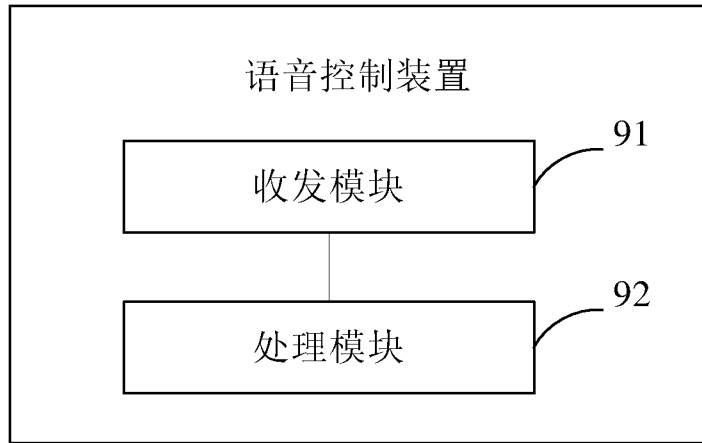


图 9

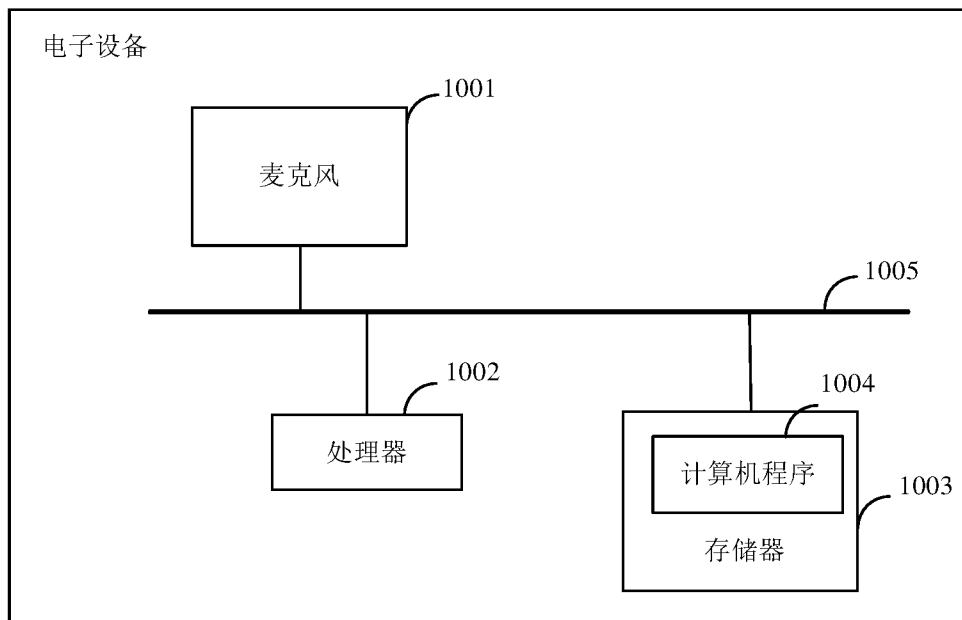


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/142083

A. CLASSIFICATION OF SUBJECT MATTER G10L 15/22(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) G10L 15, G05B 15, G05B 19, H04L 12 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, VEN, WPABSC, ENTXTC, USTXT: 语音, 音频, 录音, 采集, 转录, 多个, 设备, 终端, 装置, 电器, 家用, 家庭, 发送, 选择, 筛选, 挑选, 质量, 最好, 最优, 最高, 识别, audio, speech, voice, recognize+, record+, multip+, Select+, choos+, quality, best.		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 111326151 A (SHANGHAI NOKIA BELL CO., LTD.) 23 June 2020 (2020-06-23) description, paragraphs [0043]-[0058], and figures 2-6	11-14, 16-17, 19-22, 24-28
Y	CN 111326151 A (SHANGHAI NOKIA BELL CO., LTD.) 23 June 2020 (2020-06-23) description, paragraphs [0043]-[0058], and figures 2-6	1-10, 15, 18, 23, 29
Y	CN 107622652 A (QINGDAO HAIER INTELLIGENT TECHNOLOGY RESEARCH AND DEVELOPMENT CO., LTD.) 23 January 2018 (2018-01-23) description, paragraph [0067]	1-10, 15, 18, 23, 29
A	CN 111369994 A (VIVO COMMUNICATION TECHNOLOGY CO., LTD.) 03 July 2020 (2020-07-03) entire document	1-29
A	CN 112002319 A (HAIER YOUJIA INTELLIGENT TECHNOLOGY (BEIJING) CO., LTD.) 27 November 2020 (2020-11-27) entire document	1-29
A	CN 108228699 A (GOOGLE INC.) 29 June 2018 (2018-06-29) entire document	1-29
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>		
Date of the actual completion of the international search 24 February 2022		Date of mailing of the international search report 14 March 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/CN2021/142083

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2016148615 A1 (Samsung Electronics Co., Ltd.) 26 May 2016 (2016-05-26) entire document	1-29
A	US 2018228006 A1 (Lutron Technology Company LLC.) 09 August 2018 (2018-08-09) entire document	1-29

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/142083

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111326151	A	23 June 2020	None			
CN	107622652	A	23 January 2018	CN	107622652	B	02 October 2020
CN	111369994	A	03 July 2020	None			
CN	112002319	A	27 November 2020	None			
CN	108228699	A	29 June 2018	DE	202017105526	U1	24 May 2018
				DE	102017121086	A1	28 June 2018
				US	2018182397	A1	28 June 2018
				WO	2018118136	A1	28 June 2018
				GB	2558342	A	11 July 2018
				SG	10201707702	A1	30 July 2018
				EP	3559945	A1	30 October 2019
US	2016148615	A1	26 May 2016	CN	105632491	A	01 June 2016
				EP	3026667	A1	01 June 2016
				WO	2016085157	A1	02 June 2016
				KR	20160064258	A	08 June 2016
				EP	3026667	B1	07 June 2017
				IN	201717012279	A	07 July 2017
				US	9779732	B2	03 October 2017
US	2018228006	A1	09 August 2018	CA	3052978	A1	16 August 2018
				WO	2018148315	A1	16 August 2018
				CN	110506452	A	26 November 2019
				EP	3580999	A1	18 December 2019
				US	10694608	B2	23 June 2020
				EP	3580999	B1	17 March 2021
				CN	110506452	B	03 December 2021

国际检索报告

国际申请号

PCT/CN2021/142083

<p>A. 主题的分类</p> <p>G10L 15/22 (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>G10L 15, G05B 15, G05B 19, H04L 12</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, VEN, WPABSC, ENTXTC, USTXT: 语音, 音频, 录音, 采集, 转录, 多个, 设备, 终端, 装置, 电器, 家用, 家庭, 发送, 选择, 筛选, 挑选, 质量, 最好, 最优, 最高, 识别, audio, speech, voice, recognize+, record+, multip+, Select+, choos+, quality, best.</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>X</td> <td>CN 111326151 A (上海诺基亚贝尔股份有限公司) 2020年6月23日 (2020 - 06 - 23) 说明书第[0043]-[0058]段, 图2-6</td> <td>11-14, 16-17, 19-22, 24-28</td> </tr> <tr> <td>Y</td> <td>CN 111326151 A (上海诺基亚贝尔股份有限公司) 2020年6月23日 (2020 - 06 - 23) 说明书第[0043]-[0058]段, 图2-6</td> <td>1-10, 15, 18, 23, 29</td> </tr> <tr> <td>Y</td> <td>CN 107622652 A (青岛海尔智能技术研发有限公司) 2018年1月23日 (2018 - 01 - 23) 说明书第[0067]段</td> <td>1-10, 15, 18, 23, 29</td> </tr> <tr> <td>A</td> <td>CN 111369994 A (维沃移动通信有限公司) 2020年7月3日 (2020 - 07 - 03) 全文</td> <td>1-29</td> </tr> <tr> <td>A</td> <td>CN 112002319 A (海尔优家智能科技北京有限公司) 2020年11月27日 (2020 - 11 - 27) 全文</td> <td>1-29</td> </tr> <tr> <td>A</td> <td>CN 108228699 A (谷歌有限责任公司) 2018年6月29日 (2018 - 06 - 29) 全文</td> <td>1-29</td> </tr> <tr> <td>A</td> <td>US 2016148615 A1 (Samsung Electronics Co., Ltd.) 2016年5月26日 (2016 - 05 - 26) 全文</td> <td>1-29</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 111326151 A (上海诺基亚贝尔股份有限公司) 2020年6月23日 (2020 - 06 - 23) 说明书第[0043]-[0058]段, 图2-6	11-14, 16-17, 19-22, 24-28	Y	CN 111326151 A (上海诺基亚贝尔股份有限公司) 2020年6月23日 (2020 - 06 - 23) 说明书第[0043]-[0058]段, 图2-6	1-10, 15, 18, 23, 29	Y	CN 107622652 A (青岛海尔智能技术研发有限公司) 2018年1月23日 (2018 - 01 - 23) 说明书第[0067]段	1-10, 15, 18, 23, 29	A	CN 111369994 A (维沃移动通信有限公司) 2020年7月3日 (2020 - 07 - 03) 全文	1-29	A	CN 112002319 A (海尔优家智能科技北京有限公司) 2020年11月27日 (2020 - 11 - 27) 全文	1-29	A	CN 108228699 A (谷歌有限责任公司) 2018年6月29日 (2018 - 06 - 29) 全文	1-29	A	US 2016148615 A1 (Samsung Electronics Co., Ltd.) 2016年5月26日 (2016 - 05 - 26) 全文	1-29
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
X	CN 111326151 A (上海诺基亚贝尔股份有限公司) 2020年6月23日 (2020 - 06 - 23) 说明书第[0043]-[0058]段, 图2-6	11-14, 16-17, 19-22, 24-28																								
Y	CN 111326151 A (上海诺基亚贝尔股份有限公司) 2020年6月23日 (2020 - 06 - 23) 说明书第[0043]-[0058]段, 图2-6	1-10, 15, 18, 23, 29																								
Y	CN 107622652 A (青岛海尔智能技术研发有限公司) 2018年1月23日 (2018 - 01 - 23) 说明书第[0067]段	1-10, 15, 18, 23, 29																								
A	CN 111369994 A (维沃移动通信有限公司) 2020年7月3日 (2020 - 07 - 03) 全文	1-29																								
A	CN 112002319 A (海尔优家智能科技北京有限公司) 2020年11月27日 (2020 - 11 - 27) 全文	1-29																								
A	CN 108228699 A (谷歌有限责任公司) 2018年6月29日 (2018 - 06 - 29) 全文	1-29																								
A	US 2016148615 A1 (Samsung Electronics Co., Ltd.) 2016年5月26日 (2016 - 05 - 26) 全文	1-29																								
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><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 “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>																										
<p>国际检索实际完成的日期</p> <p>2022年2月24日</p>		<p>国际检索报告邮寄日期</p> <p>2022年3月14日</p>																								
<p>ISA/CN的名称和邮寄地址</p> <p>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</p>		<p>授权官员</p> <p>谭雪艳</p> <p>电话号码 (86-10)-62089709</p>																								

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	US 2018228006 A1 (Lutron Technology Company LLC.) 2018年8月9日 (2018 - 08 - 09) 全文	1-29

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/142083

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	111326151	A	2020年6月23日	无			
CN	107622652	A	2018年1月23日	CN	107622652	B	2020年10月2日
CN	111369994	A	2020年7月3日	无			
CN	112002319	A	2020年11月27日	无			
CN	108228699	A	2018年6月29日	DE	202017105526	U1	2018年5月24日
				DE	102017121086	A1	2018年6月28日
				US	2018182397	A1	2018年6月28日
				WO	2018118136	A1	2018年6月28日
				GB	2558342	A	2018年7月11日
				SG	10201707702	A1	2018年7月30日
				EP	3559945	A1	2019年10月30日
US	2016148615	A1	2016年5月26日	CN	105632491	A	2016年6月1日
				EP	3026667	A1	2016年6月1日
				WO	2016085157	A1	2016年6月2日
				KR	20160064258	A	2016年6月8日
				EP	3026667	B1	2017年6月7日
				IN	201717012279	A	2017年7月7日
				US	9779732	B2	2017年10月3日
US	2018228006	A1	2018年8月9日	CA	3052978	A1	2018年8月16日
				WO	2018148315	A1	2018年8月16日
				CN	110506452	A	2019年11月26日
				EP	3580999	A1	2019年12月18日
				US	10694608	B2	2020年6月23日
				EP	3580999	B1	2021年3月17日
				CN	110506452	B	2021年12月3日