



(12) 发明专利

(10) 授权公告号 CN 112201246 B

(45) 授权公告日 2023. 11. 28

(21) 申请号 202011308093.6

(22) 申请日 2020.11.19

(65) 同一申请的已公布的文献号  
申请公布号 CN 112201246 A

(43) 申请公布日 2021.01.08

(73) 专利权人 深圳市欧瑞博科技股份有限公司  
地址 518055 广东省深圳市南山区学苑大道1001号南山智园A7栋7楼

(72) 发明人 何海亮

(74) 专利代理机构 深圳市智圈知识产权代理事务所(普通合伙) 44351  
专利代理师 徐川

(51) Int. Cl.

G10L 15/22 (2006.01)

G10L 17/08 (2013.01)

(56) 对比文件

CN 105872688 A, 2016.08.17

CN 110579977 A, 2019.12.17

CN 108520743 A, 2018.09.11

CN 109545206 A, 2019.03.29

CN 110047481 A, 2019.07.23

CN 110232916 A, 2019.09.13

CN 111223490 A, 2020.06.02

CN 111816192 A, 2020.10.23

CN 111819626 A, 2020.10.23

US 2017116994 A1, 2017.04.27

审查员 李胜男

权利要求书2页 说明书16页 附图4页

(54) 发明名称

基于语音的智能控制方法、装置、电子设备及存储介质

(57) 摘要

本申请实施例公开了一种基于语音的智能控制方法、装置、电子设备及存储介质,涉及语音识别领域,该方法包括:获取当前场景的模式信息;检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词;当检测为有时,根据关键词,确定对应的第一控制指令;执行所述第一控制指令。本申请实施例通过获取当前场景的模式信息,当检测到第一语音信息中包括有模式信息对应的关键词时,可以根据关键词确定第一控制指令,从而用户可以通过模式信息对应的关键词输入控制指令,避免了用户每一次发出语音指令时都需要先说唤醒词,从而有效地提高了用户的使用体验。



1. 一种基于语音的智能控制方法,其特征在于,所述方法包括:
  - 获取当前场景的模式信息;
  - 检测用户输入的第一语音信息中是否包括有与所述模式信息对应的关键词;
  - 当检测为有时,根据所述关键词,确定对应的第一控制指令;
  - 执行所述第一控制指令;
  - 当检测所述第一语音信息中没有与所述模式信息对应的关键词时,检测所述第一语音信息是否包含预设唤醒词;
  - 若包含,则对所述第一语音信息进行语义解析以获取第二控制指令;
  - 执行所述第二控制指令;
  - 在所述获取当前场景的模式信息之前,所述方法还包括:
    - 获取用户输入的第二语音信息;
    - 若所述第二语音信息包括预设唤醒词,执行唤醒操作,其中,所述预设唤醒词与所述关键词不同;
    - 获取所述第二语音信息对应的场景,将所述场景作为所述当前场景;
    - 每一个场景包括至少一个模式信息,每一个模式信息包括至少一个关键词和所述关键词对应的第一控制指令;
  - 在所述对所述第一语音信息进行语义解析以获取第二控制指令之后,所述方法还包括:
    - 获取所述第二控制指令对应的场景类别;
    - 若所述场景类别与所述当前场景不同,则将所述当前场景切换为所述场景类别对应的场景。
2. 根据权利要求1所述的方法,其特征在于,所述检测用户输入的第一语音信息中是否包括有与所述模式信息对应的关键词,包括:
  - 对所述第一语音信息进行声纹识别以获取目标声纹特征;
  - 获取所述目标声纹特征与指定声纹特征之间的相似度,其中,所述指定声纹特征为所述第二语音信息的声纹特征;
  - 若所述相似度大于第一预设阈值,则检测用户输入的所述第一语音信息中是否包括有与所述模式信息对应的关键词。
3. 根据权利要求1至2任一项所述的方法,其特征在于,所述检测用户输入的所述第一语音信息中是否包括有与所述模式信息对应的关键词,包括:
  - 基于预设的声学模型对用户输入的所述第一语音信息进行分析,以获取所述第一语音信息与所述模式信息对应的关键词的声学特征相似度;
  - 若所述声学特征相似度大于第二预设阈值,则判定所述第一语音信息包括有与所述模式信息对应的关键词。
4. 一种基于语音的智能控制装置,其特征在于,所述装置包括:
  - 获取模块,用于获取当前场景的模式信息;
  - 检测模块,用于检测用户输入的第一语音信息中是否包括有与所述模式信息对应的关键词;
  - 指令确定模块,用于当检测为有时,根据所述关键词,确定对应的第一控制指令;

执行模块,用于执行所述第一控制指令;

唤醒词检测模块,用于当检测所述第一语音信息中没有与所述模式信息对应的关键词时,检测所述第一语音信息是否包含预设唤醒词;

语义解析模块,用于若包含,则对所述第一语音信息进行语义解析以获取第二控制指令;

第二指令执行模块,用于执行所述第二控制指令;

在所述获取当前场景的模式信息之前,所述装置还包括:

第二语音获取模块,用于获取用户输入的第二语音信息;

唤醒执行模块,用于若所述第二语音信息包括预设唤醒词,执行唤醒操作,其中,所述预设唤醒词与所述关键词不同;

场景获取模块,用于获取所述第二语音信息对应的场景,将所述场景作为所述当前场景;

每一个场景包括至少一个模式信息,每一个模式信息包括至少一个关键词和所述关键词对应的第一控制指令;

在所述对所述第一语音信息进行语义解析以获取第二控制指令之后,所述装置还包括:

场景获取模块,用于获取所述第二控制指令对应的场景类别;

场景切换模块,用于若所述场景类别与所述当前场景不同,则将所述当前场景切换为所述场景类别对应的场景。

5.一种电子设备,其特征在于,包括:

一个或多个处理器;

存储器;

一个或多个应用程序,其中所述一个或多个应用程序被存储在所述存储器中并被配置为由所述一个或多个处理器执行,所述一个或多个应用程序配置用于执行如权利要求1至3任一项所述的方法。

6.一种计算机可读取存储介质,其特征在于,所述计算机可读取存储介质中存储有程序代码,所述程序代码可被处理器调用执行如权利要求1至3任一项所述的方法。

## 基于语音的智能控制方法、装置、电子设备及存储介质

### 技术领域

[0001] 本申请涉及语音识别技术领域,更具体地,涉及一种基于语音的智能控制方法、装置、电子设备及存储介质。

### 背景技术

[0002] 随着语音识别技术的逐渐发展,越来越多的终端设备具备了语音交互的功能,终端设备可以通过对用户语音的信息理解,实现对终端设备的控制。但当前的终端设备通常需要用户先输入一个预先设置的唤醒词来唤醒设备,再进行用户与终端设备之间的交互。用户每次需要输入控制指令时,都需要先输入唤醒词,增加了交互成本,降低了用户的交互体验。如何对语音交互过程进行优化,以提升用户的语音交互体验是亟待解决的问题。

### 发明内容

[0003] 鉴于上述问题,本申请提供了一种基于语音的智能控制方法、装置、电子设备及存储介质,以改善上述问题。

[0004] 第一方面,本申请实施例提供了一种基于语音的智能控制方法,该方法包括:获取当前场景的模式信息;检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词;当检测为有时,根据关键词,确定对应的第一控制指令;执行第一控制指令。

[0005] 进一步地,在获取当前场景的模式信息之前,包括:获取用户输入的第二语音信息;若第二语音信息包括预设唤醒词,执行唤醒操作,其中,预设唤醒词与关键词不同;获取第二语音信息对应的场景,将场景作为当前场景。

[0006] 进一步地,每一个场景包括至少一个模式信息,每一个模式信息包括至少一个关键词和关键词对应的第一控制指令

[0007] 进一步地,检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词,包括:对第一语音信息进行声纹识别以获取目标声纹特征;获取目标声纹特征与指定声纹特征之间的相似度,其中,指定声纹特征为第二语音信息的声纹特征;若相似度大于第一预设阈值,则检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词。

[0008] 进一步地,检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词,包括:基于预设的声学模型对用户输入的第一语音信息进行分析,以获取第一语音信息与模式信息对应的关键词的声学特征相似度;若声学特征相似度大于第二预设阈值,则判定第一语音信息包括有与模式信息对应的关键词。

[0009] 进一步地,当检测第一语音信息中没有与模式信息对应的关键词时,检测第一语音信息是否包含预设唤醒词;若包含,则对第一语音信息进行语义解析以获取第二控制指令;执行第二控制指令。

[0010] 进一步地,在对第一语音信息进行语义解析以获取第二控制指令之后,包括:获取第二控制指令对应的场景类别;若场景类别与当前场景不同,则将当前场景切换为场景类别对应的场景。

[0011] 第二方面,本申请实施例提供了一种基于语音的智能控制装置,该装置包括:获取模块,用于获取当前场景的模式信息;检测模块,用于检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词;指令确定模块,用于当检测为有时,根据关键词,确定对应的第一控制指令;执行模块,用于执行第一控制指令。

[0012] 第三方面,本申请提供了一种电子设备,包括:一个或多个处理器;存储器;一个或多个应用程序,其中一个或多个应用程序被存储在存储器中被配置为由一个或多个处理器执行,一个或多个应用程序配置用于执行上述第一方面所述的方法。

[0013] 第四方面,本申请提供了一种计算机可读存储介质,所述计算机可读取存储介质中存储有程序代码,所述程序代码可被处理器调用执行上述第一方面所述的方法。

[0014] 本申请实施例公开了一种基于语音的智能控制方法、装置、电子设备及存储介质,可以获取终端设备当前场景,用户可以仅通过关键词发出当前场景中的控制指令,避免了用户每一次发出语音指令时都需要先说唤醒词。具体地,获取当前场景的模式信息,其中,模式信息为当前场景下用于语音交互的配置信息;检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词,其中,每个关键词可以对应有当前场景的控制指令,终端设备可以通过检测该关键词获取用户的控制指令;当检测为有时,根据关键词,确定对应的第一控制指令;执行第一控制指令。因此,终端设备可以通过检测关键词获取用户输入的控制指令,而不是仅仅是检测到唤醒词后才能对语音中的控制指令进行识别,用户可以通过关键词发出相应的控制指令,避免了每一次发出语音指令时用户都需要先说唤醒词,从而有效地提高了用户的使用体验。

## 附图说明

[0015] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0016] 图1示出了一种适用于本申请实施例的应用环境示意图。

[0017] 图2示出了本申请一实施例提供的一种基于语音的智能控制方法的方法流程图。

[0018] 图3示出了本申请另一实施例提供的一种基于语音的智能控制方法的方法流程图。

[0019] 图4示出了本申请另一实施例提供的一种基于语音的智能控制方法中步骤S350的流程图。

[0020] 图5示出了本申请又一实施例提供的一种基于语音的智能控制方法的方法流程图。

[0021] 图6示出了本申请实施例提供的一种基于语音的智能控制装置的结构框图;

[0022] 图7示出了本申请实施例提供的电子设备的结构框图。

[0023] 图8示出了本申请实施例的用于保存或者携带实现根据本申请实施例的基于语音的智能控制方法的程序代码的存储单元。

## 具体实施方式

[0024] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0025] 随着语音交互技术的不断发展,越来越多的终端设备可以通过识别用户语音数据来进行交互。在语音交互中,为了降低误唤醒,通常采用一次语音唤醒对应一次语音识别,即用户每次输入语音控制命令时都需要先输入预设的唤醒词,当用户输入的语音与预设的唤醒词匹配时,终端设备再对用户输入的语音进行语音识别,以获取该语音对应的控制命令。例如,某设备的唤醒词为“AA”,当用户想与该设备进行交互时,需要先说出“AA”来唤醒休眠状态下的终端设备,再说出控制命令“请播放音乐”,或者说出带有唤醒词的控制命令“AA,请播放音乐”,而如果直接说“请播放音乐”则设备无法识别用户的语音。因此,用户每次需要输入控制指令时,都需要先输入唤醒词,降低了用户的交互体验。

[0026] 为了解决上述问题,发明人经过长期的研究,提出了本申请实施例中基于语音的智能控制方法、装置、电子设备及存储介质。本申请实施例获取当前场景的模式信息,检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词,当检测为有时,根据关键词,确定对应的第一控制指令,执行第一控制指令。本申请实施例可以通过获取当前场景的模式信息,当检测到第一语音信息中包括有模式信息对应的关键词时,根据关键词确定第一控制指令,因此用户可以仅通过关键词发出相应的控制指令,避免了用户每一次发出语音指令时都需要先说唤醒词,从而有效地提高了用户的使用体验。终端设备可以在免唤醒的状态直接执行第一控制指令,从而一定程度上提高了终端设备的控制效率。

[0027] 为了更好地理解本申请实施例提供的一种基于语音的智能控制方法、装置、电子设备及存储介质,下面先对适用于本申请实施例的应用环境进行描述。

[0028] 请参阅图1,图1示出了一种适用于本申请实施例的应用环境示意图。本申请实施例提供的虚拟试用方法可以应用于如图1所示的多态交互系统100。多态交互系统100包括终端设备101以及服务器102,服务器102与终端设备101通信连接。其中,服务器102可以是单独的服务器,也可以是服务器集群,可以是本地服务器,也可以是云端服务器,在此不作具体限定。

[0029] 其中,终端设备101可以是具有语音交互装置的各种电子设备,包括但不限于智能家居设备、智能音响、智能网关、机器人、车载设备、智能手机、平板电脑、膝上型便携计算机、台式计算机和可穿戴式电子设备等。具体地,终端设备101可以包括麦克风等语音输入模块、扬声器等语音输出模块以及处理器。语音交互装置可以内置在终端设备101中,也可以作为一个独立的模块,通过API或者其它方式与终端设备101进行通信。作为一种方式,终端设备101上还可以设置有用于输入字符的字符输入模块、用于输入图像的图像输入模块、用于输入视频的视频输入模块等,以实现多态交互。

[0030] 其中,终端设备101上可以安装有客户端应用程序,用户可以基于客户端应用程序与服务器102进行通信,具体地,服务器102上安装有对应的服务端应用程序,用户可以基于客户端应用程序在服务器102注册一个用户帐号,并基于该用户帐号与服务器102进行通信,例如用户在客户端应用程序登录用户帐号,并基于该用户帐号通过客户端应用程序进

行输入,可以输入文字信息、语音信息、图像信息或视频信息等,客户端应用程序接收到用户输入的信息后,可以将该信息发送至服务器102,使得服务器102可以接收该信息并进行处理及存储,服务器102还可以接收该信息并根据该信息返回一个对应的输出信息至终端设备101。

[0031] 在一些实施方式中,终端设备101在获取与用户输入的信息对应的回复信息后,可以在终端设备101的显示屏或与其连接的其他图像输出设备上显示该回复信息。作为一种方式,在显示该回复信息的同时,也可以通过终端设备101的扬声器或与其连接的其他音频输出设备播放对应的音频,还可以在终端设备101的显示屏上显示与该回复信息对应的文字或图形,实现在图像、语音、文字等多个方面上与用户的多态交互。

[0032] 在一些实施方式中,服务器102可以对终端设备101上接收的语音数据提供识别服务,以得到用户输入语音数据的文本表示,并基于文本表示得到用户意图表示,从而生成相应的控制指令,将该控制指令返回至终端设备101上。终端设备101根据该控制指令相应的操作,为用户提供服务。例如,播放歌曲、拨打电话、设置闹钟等。

[0033] 在一些实施方式中,终端设备101可以是智能音响、智能网关、智能家居面板等终端设备,该终端设备可以与至少一个受控设备连接,其中,受控设备可以但不仅限于包括空调、地暖、新风、窗帘、灯、电视、冰箱、电扇等智能家居设备,终端设备和智能家居设备可以通过蓝牙、WiFi或者ZigBee等方式进行连接。终端设备101获取用户输入的信息后,对用户输入的信息进行识别,当确定该信息为受控设备的控制指令时,控制该受控设备根据控制指令进行响应。例如,智能网关与窗帘相连接,用户输入语音指令“AA,打开窗帘”,智能网关在检测到唤醒词“AA”后识别该语音指令,控制窗帘打开。

[0034] 在一些实施方式中,对用户输入的信息进行处理的装置也可以设置于终端设备101上,使得终端设备101无需依赖与服务器102建立通信即可实现与用户的交互,此时多态交互系统100可以只包括终端设备101。

[0035] 上述应用环境仅为方便理解所作的示例,可以理解的是,本申请实施例不仅局限于上述应用环境。

[0036] 下面将结合附图具体描述本申请的各实施例。

[0037] 请参阅图2,图2为本申请一实施例提供的一种基于语音的智能控制方法的方法流程图,该方法包括步骤S210至步骤S240。

[0038] 步骤S210:获取当前场景的模式信息。

[0039] 其中,模式信息为终端设备当前工作场景对应的用于语音交互的配置信息。具体地,每个场景对应的模式信息可以包括可用作控制命令的关键词、用于检测语音信息的声学检测模型、进行响应的语音信号强度阈值、等待用户输入语音信息的时长等信息中的一项或者几项。

[0040] 其中,终端设备可以存储有至少一个场景和每个场景对应的模式信息,场景和模式信息的对应关系可以由用户预先设置后存储在终端设备或服务器中的,也可以是由终端设备出厂时预先默认设置并存储的,还可以是由服务器预先设定完成后发送至终端设备的,在此不做限定。

[0041] 在一些实施方式中,可以根据用户实际使用的需求为终端设备设置多个场景。作为一种方式,可以根据不同的功能设置多个场景。例如,将场景设置为音乐播放、新闻播放、

购物、订票、智能家居设备控制等应用场景。作为另一种方式,可以根据不同的交互对象设置多个场景。例如,可以根据交互对象的数量将场景设置为多人交互场景和单人交互场景,也可以根据交互对象的性质将场景设置为老人、儿童、成年人等交互场景。作为又一种方式,还可以根据外部环境信息设置多个工作场景。例如,可以根据时间将场景设置为清晨场景、傍晚场景、夜晚场景等使用场景。在此不对场景的设置进行限定。

[0042] 在一些实施方式中,可以对每个场景设置一个模式信息,当终端设备工作在该场景时,基于该场景对应的模式信息进行语音交互。例如,不同场景的模式信息中可作为控制命令的关键词可以是不同的,以实现不同场景下通过语音输入该关键词执行该场景下对应的控制命令。又例如,不同场景的模式信息中可进行响应的语音信号强度阈值可以是不同的,具体地,可以在多人交互场景的模式信息中设置较高的语音信号强度阈值以防止误唤醒,在单人交互场景的模式信息中设置较低的语音信号强度阈值以提高响应的灵敏度。

[0043] 在一些实施方式中,每一个场景包括至少一个模式信息,每一个模式信息包括至少一个关键词,该关键词对应有一个用于进行控制的控制指令。作为一种方式,对于非连续对话的场景,当唤醒设备后语音等待超过预设时长后,仍然可以通过检测关键词来获取用户输入的控制指令,而不需要将用户输入的语音流发送至服务器,对用户说的整段语音进行语音识别和语义解析,来获取用户意图输入的控制指令。从而对于非连续对话场景,可以在降低功耗的同时,实现免唤醒词的命令控制。

[0044] 作为一种方式,可以对每个场景设置多个模式信息,通过对当前工作场景进一步细分,可以实现同一个工作场景下可以具有多种不同的与用户进行语音交互的模式。例如,对根据功能设置得到的多个场景,可以根据不同的交互对象对每个场景设置多种模式信息,每个模式信息对应不同的关键词。例如,智能家居设备控制的场景,可以包括儿童、成人等不同的模式信息,不同的模式信息对应的具有操控权限的设备不同,不同的模式信息对应的关键词也不同,具体地,用户可以设定不允许儿童使用电视机、投影仪等娱乐设施,则儿童模式信息不包括有可用作控制命令的关键词“开电视”,而成人模式则具有娱乐设施的操控权限,成人模式对应的关键词包括“开电视”。

[0045] 本申请实施例通过获取场景的模式信息,可以在不同的场景下,根据不同的模式信息进行该场景下的语音交互,从而提升用户的语音交互体验。进一步地,终端设备还可以通过多种方式获取当前场景的模式信息。

[0046] 在一些实施方式中,终端设备可以通过获取当前的环境信息从而获取当前场景的模式信息,其中,当前的环境信息可以通过终端设备的传感器获取的光照信息,也可以是通过终端设备的时钟获取的当前的时间信息,还可以是通过声音采集装置获取的当前场景的噪音信息。

[0047] 在另一些实施方式中,终端设备可以通过获取用户输入的语音信息获取当前场景的模式信息,具体地,请参见后续实施例。

[0048] 在又一些实施方式中,终端设备上可以配置有选择控件,作为一种方式,用户可以通过工作场景的选择控件直接选择当前工作场景,从而获取该场景对应的模式信息。作为另一种方式,用户也可以通过模式选择控件选择以何种方式获取当前场景的模式信息。例如,当用户选择语音模式时,终端设备通过获取用户输入的语音信息获取当前场景的模式信息;当用户选择语音模式和环境模式时,可以结合获取的语音信息和当前的环境信息来



获取当前场景的模式信息。例如,当用户选择语音模式和环境模式时,当用户输入“AA,播一首周杰伦的歌曲吧”,识别到用户意图是播放音乐,并通过时钟获取当前时间为晚上,则可以将获取夜间播放音乐对应的模式信息。

[0049] 步骤S220:检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词。

[0050] 在获取当前场景的模式信息之后,终端设备通过语音采集模块可以持续采集用户输入的语音信号。作为一种方式,可以在终端设备中设置语音检测模块,该语音检测模块采用语音活动检测技术(Voice activity detection,VAD)检测麦克风等语音采集模块获取的用户输入的语音信号。可选地,还可以在检测到语音信号后通过时间延迟补偿来确定获取的语音信号的完整性,避免漏掉部分语音信号。

[0051] 在一些实施方式中,当获取到用户输入的第一语音信息后,可以对第一语音信息进行一定的预处理操作,然后再进行检测以判断第一语音信息中是否包括有与模式信息对应的关键词。其中,预处理操作可以包括噪声抑制处理、回声消除处理、信号增强处理等,通过预处理操作可以提高检测的准确率。

[0052] 具体地,通过获取当前场景的模式信息,终端设备可以仅检测当前场景的模式信息对应的关键词。可以理解的是,关键词之间的区别性越大,关键词检测的准确率也越高。其中,可以表现为关键词的长度、各个关键词包括的音节之间的差异等。通过这种方式,在每个场景下需要检测的关键词的数量较少,从而所需的功耗也较少,由于减少了待检测的关键词数量,一定程度上也降低了由于关键词之间的区别性较小导致的识别错误率。

[0053] 在一些实施方式中,每个场景可以包括多个模式信息,可以根据获取的用户输入的第一语音信息,确定当前场景对应的模式信息。例如,智能家居设备控制的场景下,可以根据不同的交互对象设置多种模式信息,每个模式信息对应不同的关键词,可以通过识别语音信息确定当前交互对象,然后根据交互对象确定当前的模式信息对应的关键词,从而检测第一语音信息中是否包含该关键词。例如,智能家居设备控制的场景,可以包括儿童、成人等模式信息,儿童模式信息不具有娱乐设施的操控权限,即不具备娱乐设施相关的关键词,当通过检测获取的语音确定交互对象为儿童时,不对如“开电视”等娱乐设施相关的关键词进行检测。

[0054] 其中,检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词可以采用多种方式,例如,使用声学模型、基于模板匹配、利用神经网络等方式。本申请实施例对此不做限定。

[0055] 在一些实施方式中,可以基于预设的声学模型对用户输入的第一语音信息进行分析,以获取第一语音信息与模式信息对应的关键词的声学特征相似度,根据声学特征相似度来判断第一语音信息中是否包括有与模式信息对应的关键词,若声学特征相似度大于第二预设阈值,则判定第一语音信息包括有与模式信息对应的关键词。

[0056] 其中,第二预设阈值为预先设置的数值,第二预设阈值越高,检测关键词的准确率越高,终端设备的反应灵敏度越低;第二预设阈值越低,检测关键词的准确率越低,相应地,终端设备的反应灵敏度会越高。例如,当关键词为“切歌”时,当第二预设阈值较高时只有用户清晰准确地说出“切歌”两个字才会判定语音信息中包含该关键词,由于第二预设阈值较高,在一些情况下,即使用户说出该关键词,也可能因为口音或者噪音导致声学特征的相似

度较低,被判定为不包括该关键词;而当第二预设阈值较低时,一些相近的语音信息也会被判定为包含该关键词,例如“七哥”也可能被误判为包含“切歌”,带有口音的“切歌”语音也可以被判定为包含“切歌”。

[0057] 可以理解的是,由于不同场景的模式信息中关键词可以是不同的,所以不同的场景的模式可以对应不同的声学模型,其中,每一个声学模型中包括该场景的模式信息中每个关键词对应的声学特征。

[0058] 作为一种方式,不同场景的模式信息可以对应的语音信号强度阈值可以是不同的,当获取的用户输入的语音信号强度大于该场景对应的强度阈值时,通过该场景对应的声学模型对关键词进行检测。例如,可以在多人交互场景的模式信息中设置可以设置较高的语音信号强度阈值以降低嘈杂环境下终端设备被误唤醒的概率,在单人交互场景的模式信息中可以设置较低的语音信号强度阈值以提高响应的灵敏度。

[0059] 作为一种方式,每个声学模型中各个关键词对应的信号强度阈值可以是相同的。作为另一种方式,声学模型中各个关键词分别对应有不同的信号强度阈值,具体地,可以对区别性较差的关键词设置较高的信号强度阈值,对区别性较好的关键词设置较低的信号强度阈值,在每个场景下进一步地兼顾较低的误唤醒概率和较高的反应灵敏度。

[0060] 其中,该声学模型可以是存储在终端设备中的模型,也可以是服务器中的模型。

[0061] 作为一种方式,可以将声学模型预先存储在终端设备中时,终端设备不需要与服务器通信即可对第一语音信息进行语音识别,当终端设备处于网络信号差或者断网的情况下仍然可以实现对关键词的检测。具体地,可以基于声学模型,提取第一语音信息中的声学特征,计算第一语音信息的声学特征和关键词对应的声学特征之间的相似度。例如,可以将第一语音信号中提取得到的梅尔频率倒谱系数作为声学特征,将第一语音信息与模式信息对应的关键词之间的声学特征的最大似然比作为声学特征相似度。具体地,可以获取第一语音信息中的声学特征的每一个特征点,与关键词对应的声学特征的每个特征点进行相似度比较,然后综合所有特征点的相似度得到一个极大似然值作为声学特征相似度。

[0062] 作为另一种方式,可以在终端设备中设置第一声学模型,在服务器中设置第二声学模型,其中,第一声学模型的检测精度小于第二声学模型。当基于第一声学模型,判断第一语音信息与模式信息对应的关键词的声学特征相似度大于指定数值时,将第一语音信息发送至第二声学模型进一步检测。通过这种方式,只有第一声学模型检测到的声学特征相似度大于指定数值时,才会进行精度更高的声学检测,从而可以降低精准检测所需功耗,并提高检测的准确率。

[0063] 步骤S230:当检测为有时,根据关键词,确定对应的第一控制指令。

[0064] 其中,第一控制指令可以是对获取到第一语音信息的终端设备进行控制的指令,也可以是对与该终端设备相连接的其他受控设备进行控制的指令。具体地,控制指令可以包括受控设备和该受控设备的业务技能,根据不同的场景,该业务技能可以是控制技能,也可以是查询技能等。在此不对第一控制指令进行限定。用户可以直接通过语音输入该模式信息对应的关键词来输入第一控制指令,而不需要每次需要输入控制指令时都先输入唤醒词来唤醒终端设备,再识别用户语音信息中的控制指令。

[0065] 在一些实施方式中,第一控制指令可以是仅通过关键词确定的。具体地,终端设备中存储有每个场景的模式信息中的关键词和控制指令之间的对应关系,该对应关系可以由

用户预先设定后存储在终端设备,也可以由终端设备在出厂时默认设置后存储在终端设备本地或者服务器。根据检测到的第一语音信息中包括的与模式信息对应的关键词,和终端设备中存储的关键词和控制指令之间的对应关系,终端设备可以确定用户输入的第一语音信息中的关键词对应的第一控制指令。通过这种方式,终端设备可以在不对第一语音信息进行语义解析的情况下,获取第一语音信息对应的控制指令,从而提高确定第一控制指令的效率。因此,即使终端设备位于较差的网络环境或者断网环境下,也可以根据关键词确定第一控制指令。

[0066] 作为一种方式,不同的关键词可以对应同一个控制指令,即同一个控制指令可以对应语义相近的多个不同的关键词,通过这种方式用户可以采用不同的关键词来实现同一个控制指令。例如,在音乐播放的场景下,切换下一首歌这一控制指令对应的关键词可以为“切歌”和“下一首”。需要说明的是,不同场景下的相同关键词可以对应不同的控制指令。例如,工作场景下关键词“开灯”对应的控制指令为打开书桌的智能台灯,而娱乐场景下关键词“开灯”对应的控制指令为打开房间的彩色氛围灯。

[0067] 在另一些实施方式中,终端设备可以在检测到第一语音信息中包括有与模式信息对应的关键词之后,对第一语音信息进行语义识别,以获取该关键词对应的第一控制指令。具体地,可以通过自动语音识别技术(ASR, Automatic Speech Recognition)将第一语音信息转换为文本后,对该文本执行自然语音理解操作((Natural Language Understanding, NLU), 以实现第一语音信息的解析,根据解析的结果确定第一控制指令。

[0068] 作为一种方式,终端设备可以设置有多轮交互模式,通过开启多轮交互模式,可以结合关键词和终端设备在获取第一语音信息之前的交互内容,来确定第一控制指令。其中,多轮交互模式为可持续获取用户语音并且做出应答的持续对话模式。例如,通过获取的语音信息“我要看综艺”确定当前场景为播放视频的场景时,关键词设置为“播放”,如果这时获取到“播放演员A的小品”的语音,则通过识别确定控制指令为播放演员A的小品,并执行该指令,然后获取到“播放演员B”的语音时,终端设备可以结合之前的交互内容确定是播放演员B的小品,而非播放演员B的综艺。通过这种方式,用户不需要重复输入之前的语音信息,可以提高交互的效率。

[0069] 步骤S240:执行第一控制指令。

[0070] 在一些实施方式中,第一控制指令可以是对获取到第一语音信息的终端设备进行控制的指令,终端设备可以直接执行该第一控制指令。例如,终端设备为智能音响,第一控制指令为播放音乐,则终端设备直接执行第一控制指令播放音乐。

[0071] 在另一些实施方式中,第一控制指令也可以是对与终端设备连接的其他受控设备进行控制的指令,其中,其他受控设备可以是与终端设备通过蓝牙、WiFi或者ZigBee等方式进行本地连接的设备,也可以是与终端设备连接在同一WIFI下的WIFI设备。终端设备可以将第一控制指令发送至第一控制指令对应的受控设备,并指示该受控设备执行第一控制指令。

[0072] 在一些实施方式中,终端设备还可以获取第一控制指令的执行结果,并基于该执行结果输出应答信息。其中,应答信息可以是声音、图像或者是声光结合中的至少一种。例如,用户在智能家居控制场景下,对控制面板输入包含“打开空调”这一关键词的语音信息,控制面板确定该关键词对应的控制指令发送至空调,空调进行对应操作后,可以将执行结

果反馈到控制面板,控制面板可以通过声音提示的方式,告知用户空调打开成功,或者也可以通过震动、灯光闪烁等告知用户空调打开成功。作为一种方式,当第一控制指令执行失败时,也可以输出应答信息向用户反馈执行失败。例如,当空调没有与控制面板连接时,可以告知用户空调未连接,无法执行用户输入的控制指令。

[0073] 需要说明的是,本实施例中未详细描述的部分可以参考前述实施例,在此不再赘述。

[0074] 本实施例提供的一种基于语音的智能控制方法,在获取当前场景的模式信息后,可以检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词,当检测为有时,根据关键词,确定对应的第一控制指令,执行第一控制指令。本申请实施例提供的基于语音的智能控制方法,可以检测语音信息中与模式信息对应的关键词,根据关键词确定控制指令,终端设备可以在免唤醒的状态直接执行第一控制指令,从而一定程度上提高了终端设备的控制效率,用户可以通过关键词发出相应的控制指令,避免了每一次发出语音指令时用户都需要先说唤醒词,从而有效地提高了用户的使用体验。

[0075] 请参阅图3,为本申请一实施例提供的一种基于语音的智能控制方法的方法流程图,该方法从图3可以看出该方法包括步骤S310至步骤S370。

[0076] 步骤S310:获取用户输入的第二语音信息。

[0077] 在休眠状态下,终端设备的语音采集模块可以持续获取外部的声音,当用户发出第二语音信息时,终端设备的获取用户输入的第二语音信息。其中,休眠状态为终端设备在与用户进行语音交互之前的状态,当用户与终端设备进行交互后在预设的等待时间内没有交互行为时,终端设备也可以切换为休眠状态,休眠状态的终端设备中大多数功能模块处于停止工作的状态,消耗的功率较低。

[0078] 步骤S320:若第二语音信息包括预设唤醒词,执行唤醒操作。

[0079] 在获取用户输入的第二语音信息之后,可以基于唤醒词检测模型对第二语音信息进行预设唤醒词检测。其中,预设唤醒词可以是终端设备在出厂时预先默认设置的唤醒词,也可以是用户自行设置的唤醒词,预设唤醒词与关键词是不同的词。例如,预设唤醒词可以是“AA”等与控制命令无关的词语,关键词可以是“开灯”、“开空调”等表征用户控制意图的词语。当第二语音信息包括预设唤醒词时,可以将休眠状态的终端设备切换为唤醒状态,其中,唤醒状态下可以对获取的语音信息进一步识别。

[0080] 在一些实施方式中,可以基于终端设备本地的唤醒词检测模型来检测终端设备,从而节省休眠状态下终端设备进行语音的实时检测所需功耗。具体地,可以基于唤醒词检测模型提取第二语音信息的声学特征后,计算第二语音信息和预设唤醒词的声学特征相似度,进而判断第二语音信息中是否包括预设唤醒词。当第二语音信息和声学特征之间的声学特征相似度大于预设的唤醒阈值时,可以判定第二语音信息包括预设唤醒词,终端设备执行唤醒操作。

[0081] 需要说明的是,进行预设唤醒词检测所使用的唤醒词检测模型与步骤S220中用于获取第一语音信息与模式信息对应的关键词的声学特征相似度的声学模型的实现原理是相同的,但不同的是,终端设备的唤醒词检测模型在休眠状态和不同场景下都是同一个模型,并且是持续对获取的语音信息进行预设唤醒词的检测的;而步骤S220中的声学模型在不同场景的模式信息下是不同的模型,并且只有当终端设备处于该场景下,才基于该场景

对应的声学模型进行关键词的检测。

[0082] 步骤S330:获取第二语音信息对应的场景,将所述场景作为当前场景。

[0083] 在一些实施方式中,可以对获取到的用户输入的第二语音信息进行语音识别和语义解析,以获取第二语音信息对应的场景,将所述场景作为当前场景。

[0084] 作为一种方式,可以预先设置规则模板和场景的对应关系,终端设备采集到用户输入的第二语音信息后,可以将第二语音信息发送至服务器,在服务器对第二语音信息通过ASR将语音转换为文本,对该文本执行NLU以获取第二语音信息对应的场景。具体地,可以通过与规则模板进行匹配、文本分类、信息抽取等方式识别第二语音信息对应的场景。

[0085] 步骤S340:获取当前场景的模式信息。

[0086] 步骤S350:检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词。

[0087] 请参阅图4,在一些实施例中,步骤S350可以包括步骤S351至步骤S353。

[0088] 步骤S351:对第一语音信息进行声纹识别以获取目标声纹特征。

[0089] 其中,不同的语音信息对应不同的声纹特征,每个声纹特征对应一个待响应用户。可选地,可以基于训练好的神经网络模型对获取语音进行声纹特征识别,在此不对进行声纹识别的方式进行限定。

[0090] 作为一种方式,当终端设备获取的第一语音信息仅包含一个声纹特征时,将该声纹特征作为目标声纹特征。作为另一种方式,当终端设备获取的语音信息中包含多个不同的声纹特征,可以获取多个第一语音信息和每一个第一语音信息对应的目标声纹特征。

[0091] 步骤S352:获取目标声纹特征与指定声纹特征之间的相似度。

[0092] 其中,指定声纹特征可以是第二语音信息的声纹特征。终端设备可以对目标声纹特征和指定声纹特征进行匹配,以获取目标声纹特征与指定声纹特征之间的相似度。作为一种方式,当终端设备获取的语音信息中包含多个不同的目标声纹特征时,分别计算每一个目标声纹特征和指定声纹特征之间的相似度。

[0093] 在一些实施方式中,指定声纹特征可以是预先的预设声纹特征,通过预先存储至少一个预设的用户的声纹特征作为预设声纹特征,可以实现只对预设的用户的语音信息进行响应,使其他用户无法随意进行交互,进而节省了识别的功耗。

[0094] 步骤S353:若相似度大于第一预设阈值,则检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词。

[0095] 当相似度大于第一预设阈值时,检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词;而当相似度小于第一预设阈值时,不对第一语音信息进行关键词检测。

[0096] 具体地,当终端设备获取的语音信息中包含多个不同的声纹特征时,只对与指定声纹特征之间的相似度大于第一预设阈值的第一语音信息进行检测。通过仅检测与指定声纹特征匹配的用户语音,即只对发出第二语音信息的用户进行交互,可以提高交互的门槛,使其他用户无法进行交互,从而减少被其他用户打断的概率,解决因对语音信号无差别识别和响应所导致的交互容易被截断的问题。

[0097] 在一些实施方式中,当相似度大于第一预设阈值,还可以检测第一语音信息中是否包含停止响应信息,其中,停止响应信息用于表示第二语音信息对应的用户停止进行交

互。若检测到停止响应信息,则终端设备不再对第一语音信息进行声纹识别,也不进行相似度的判断,即对所有获取的第一语音信息进行无差别的响应。通过这种方式,在该用户停止交互后,终端设备可以与其他用户进行交互,从而提高系统的可用性。

[0098] 步骤S360:当检测为有时,根据关键词,确定对应的第一控制指令。

[0099] 步骤S370:执行第一控制指令。

[0100] 需要说明的是,本实施例中未详细描述的部分可以参考前述实施例,在此不再赘述。

[0101] 本实施例提供的一种基于语音的智能控制方法,获取用户输入的第二语音信息后,若第二语音信息包括预设唤醒词,执行唤醒操作,获取第二语音信息对应的场景,将该场景作为当前场景,获取当前场景的模式信息后,可以检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词,当检测为有时,根据关键词,确定对应的第一控制指令,执行第一控制指令。本实施例提供的一种基于语音的智能控制方法,可以通过识别用户输入的语音信息来获取当前的场景,在用户无感知的情况下获取当前场景,基于该场景的模式信息响应于用户输入的免唤醒词的语音控制命令,在一定程度上可以避免多种场景的免唤醒词的语音控制命令同时存在时出现的串词现象。

[0102] 请参阅图5,为本申请一实施例提供的一种基于语音的智能控制方法的方法流程图,该方法从图5可以看出该方法包括步骤S410至步骤S470。

[0103] 步骤S410:获取当前场景的模式信息。

[0104] 在一些实施方式中,在获取当前场景的模式信息之前,可以获取用户输入的第二语音信息,若所述第二语音信息包括预设唤醒词,则执行唤醒操作,其中,所述预设唤醒词与所述关键词不同,并获取所述第二语音信息对应的场景,将所述场景作为所述当前场景。

[0105] 在一些实施方式中,每一个场景包括至少一个模式信息,每一个模式信息包括至少一个关键词和该关键词对应的第一控制指令。

[0106] 步骤S420:检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词。

[0107] 在一些实施方式中,在获取当前场景的模式信息之后,还可以获取用户输入的第一语音信息,并对第一语音信息进行声纹识别以获取目标声纹特征判断目标声纹特征与指定声纹特征之间的相似度,其中,指定声纹特征为第二语音信息的声纹特征;若相似度大于第一预设阈值,则检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词。

[0108] 当检测用户输入的第一语音信息中包括有与模式信息对应的关键词时,则根据关键词,确定对应的第一控制指令,即进入步骤S430,当检测用户输入的第一语音信息中没有与模式信息对应的关键词时,则检测第一语音信息是否包含预设唤醒词,即进入步骤S450。

[0109] 步骤S430:根据关键词,确定对应的第一控制指令。

[0110] 步骤S440:执行第一控制指令。

[0111] 步骤S450:检测第一语音信息是否包含预设唤醒词。

[0112] 其中,预设唤醒词可以是终端设备在出厂时预先默认设置的唤醒词,也可以是用户自行设置的唤醒词,可以基于唤醒词检测模型对第一语音信息进行预设唤醒词检测。可选地,预设唤醒词与休眠状态下唤醒终端设备所需的唤醒词可以是相同的,也可以是不同的。

[0113] 步骤S460:若包含,则对第一语音信息进行语义解析以获取第二控制指令;

[0114] 若检测到第一语音信息中包含预设唤醒词,则通过ASR将第一语音信息转为文本,并对文本执行NLU以通过语义解析获取第二控制指令。可选地,当对第一语音信息进行语义解析无法获取第二控制指令时,可以输出响应信息与用户交互,以进一步确定用户意图的控制指令。例如,用户输入“AA,倩女幽魂”,终端设备不能确定用户想要下载倩女幽魂的游戏,还是播放倩女幽魂的电影时,可以输出语音信息“请问您想要播放倩女幽魂的电影吗”,并根据用户接下来的输入确定用户的意图的控制指令。

[0115] 在一些实施方式中,在对第一语音信息进行语义解析以获取第二控制指令之后,还可以获取第二控制指令对应的场景类别;若场景类别与当前场景不同,则将当前场景切换为场景类别对应的场景。通过对语音信息进行语义解析获取场景类别,可以在用户无感知的情况下进行场景的切换。

[0116] 作为一种方式,终端设备还可以存储有预设唤醒词和控制指令,以及控制指令和场景之间的一一对应的关系,即每一个预设唤醒词对应一个控制指令,每一个控制指令对应一个场景。具体地,当检测到第一语音信息包括预设唤醒词时,将该预设唤醒词对应的控制指令作为第二控制指令,并获取第二控制指令对应的场景的类别。

[0117] 需要说明的是,预设唤醒词在所有的场景下都是有效的,即可以被检测的,而每个场景对应的关键词在其他场景下是无法被识别的。例如,预设唤醒词为“播放音乐”对应音乐场景,而音乐场景对应的关键词可以包括“上一曲”、“下一曲”、“暂停”等,如果在视频场景下检测到用户输入的语音信息包含“播放音乐”,则将音乐场景作为当前场景类别,并将当前场景切换为音乐场景,进而可以采用音乐场景的模式信息对应的“上一曲”等关键词进行控制;而如果终端设备在视频场景下获取包含“上一曲”的语音信息,则终端设备不进行任何响应。

[0118] 作为一种方式,可以根据交互场景是持续交互还是非持续交互,对场景类别进行标识,在获取场景类别的同时获取场景类别的标识,当第二控制指令对应的场景类别与当前场景不同,并且当前场景对应的标识为持续交互,第二控制指令对应的场景类别为非持续交互时,可以仅执行第二控制指令,但不切换当前的场景。只有在第二控制指令对应的场景类别与当前场景不同,并且场景类别和当前场景都是持续性场景时,将当前场景切换为场景类别对应的场景。例如,当前场景为持续交互的音乐播放场景,而第二控制指令“开灯”对应非持续交互的智能设备控制场景,终端设备可以在开灯后仍然保持音乐播放场景,即可以使用该场景对应的关键词进行交互,而不需要再次通过输入语音进入音乐播放场景。通过这种方式,可以避免非持续交互场景下的控制指令打断用户与终端设备的持续交互,从而需要频繁地进行场景切换,进一步降低了交互成本。

[0119] 步骤S470:执行第二控制指令。

[0120] 执行第二控制指令的步骤可参见步骤S240中执行第一控制指令。

[0121] 在一些实施方式中,也可以在获取当前场景的模式信息后,检测第一语音信息是否包含预设唤醒词;若包含,则对第一语音信息进行语义解析以获取第二控制指令,执行第二控制指令;若不包含,则检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词,当检测到包括有与模式信息对应的关键词时,根据关键词确定对应的第一控制指令,并执行第一控制指令。

[0122] 在另一些实施方式中,可以在获取第一语音信息后,同时检测第一语音信息中是否包括有与模式信息对应的关键词、第一语音信息是否包含预设唤醒词,其中预设唤醒词和关键词是不同的词,不会存在关键词和唤醒词同时被检测成功的情况,若检测到关键词,则执行步骤S430,若检测到预设唤醒词,则执行步骤S460。通过同时对预设唤醒词和关键词进行检测,可以提高对用户输入语音的检测效率,更快地进行响应。

[0123] 需要说明的是,本实施例中未详细描述的部分可以参考前述实施例,在此不再赘述。

[0124] 本实施例提供的一种基于语音的智能控制方法,在获取当前场景的模式信息后,可以检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词,当检测为有时,根据关键词,确定对应的第一控制指令,执行第一控制指令;当检测为没有时,根检测第一语音信息中是否包含预设唤醒词,若包含则对第一语音信息进行语义解析以获取第二控制指令,并指令第二控制指令。本申请实施例提供的基于语音的智能控制方法,用户可以直接通过该场景的模式信息对应的关键词发出控制命令,也可以通过带有唤醒词的语音信息进行更加多样性的控制命令,还可以根据带有唤醒词的语音信息进行场景的切换,通过这种方式可以在预设的关键词之外,满足用户更加多样的语音交互需求,进一步地提升用户的使用体验。

[0125] 请参阅图6,为本申请实施例提供的一种基于语音的智能控制装置结构框图,该装置600包括:获取模块610、检测模块620、指令确定模块630以及执行模块640。

[0126] 获取模块610,用于获取当前场景的模式信息。

[0127] 进一步地,在获取当前场景的模式信息之前,装置600还包括:第二语音获取模块,唤醒执行模块,以及场景获取模块。

[0128] 第二语音获取模块,用于获取用户输入的第二语音信息。

[0129] 唤醒执行模块,用于若第二语音信息包括预设唤醒词,执行唤醒操作,其中,预设唤醒词与关键词不同。

[0130] 场景获取模块,用于获取第二语音信息对应的场景,将场景作为当前场景。

[0131] 进一步地,每一个场景包括至少一个模式信息,每一个模式信息包括至少一个关键词和关键词对应的第一控制指令。

[0132] 检测模块620,用于检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词。

[0133] 进一步地,检测模块620还包括:声纹识别子模块,相似度获取模块,第一判断子模块。

[0134] 声纹识别子模块,用于对第一语音信息进行声纹识别以获取目标声纹特征。

[0135] 相似度获取模块,用于判断目标声纹特征与指定声纹特征之间的相似度,其中,指定声纹特征为第二语音信息的声纹特征。

[0136] 第一判断子模块,用于若相似度大于第一预设阈值,则检测用户输入的第一语音信息中是否包括有与模式信息对应的关键词。

[0137] 进一步地,检测模块620还包括:语音分析子模块和第二判断子模块。

[0138] 语音分析子模块,用于基于预设的声学模型对用户输入的第一语音信息进行分析,以获取第一语音信息与模式信息对应的关键词的声学特征相似度;



[0139] 第二判断子模块,用于若声学特征相似度大于第二预设阈值,则判定第一语音信息包括有与模式信息对应的关键词。

[0140] 指令确定模块630,用于当检测为有时,根据关键词,确定对应的第一控制指令。

[0141] 执行模块640,用于执行第一控制指令。

[0142] 进一步地,装置600还包括:唤醒词检测模块,语义解析模块,第二执行模块。

[0143] 唤醒词检测模块,用于当检测第一语音信息中没有与模式信息对应的关键词时,检测第一语音信息是否包含预设唤醒词。

[0144] 语义解析模块,用于若包含,则对第一语音信息进行语义解析以获取第二控制指令。

[0145] 第二指令执行模块,用于执行第二控制指令。

[0146] 进一步地,在对第一语音信息进行语义解析以获取第二控制指令之后,装置600还包括:场景获取模块和场景切换模块。

[0147] 场景获取模块,用于获取第二控制指令对应的场景类别。

[0148] 场景切换模块,用于若场景类别与当前场景不同,则将当前场景切换为场景类别对应的场景。

[0149] 需要说明的是,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,上述描述的装置和模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0150] 在本申请所提供的几个实施例中,模块相互之间的耦合可以是电性,机械或其它形式的耦合。另外,在本申请各个实施例中的各功能模块可以集成在一个处理模块中,也可以是各个模块单独物理存在,也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。

[0151] 综上所述,本申请实施例公开了一种基于语音的智能控制方法、装置、电子设备及存储介质,涉及语音识别领域,该方法包括在获取当前场景的模式信息后,可以检测用户输入的第一语音信息中是否包括有与所述模式信息对应的关键词,当检测为有时,根据所述关键词,确定对应的第一控制指令,行所述第一控制指令。本申请实施例通过获取当前场景的模式信息,当检测到第一语音信息中包括有模式信息对应的关键词时,可以根据关键词确定第一控制指令;终端设备可以在免唤醒的状态直接执行第一控制指令,从而一定程度上提高了终端设备的控制效率;用户可以仅通过关键词发出相应的控制指令,避免了用户每一次发出语音指令时都需要先说唤醒词,从而有效地提高了用户的使用体验。

[0152] 下面将结合图7本申请提供的一种电子设备进行说明。

[0153] 请参阅图7基于上述的基于语音的智能控制方法、装置、电子设备及存储介质,本申请实施例还提供了另一种可以执行前述基于语音的智能控制方法的电子设备700。电子设备700包括相互耦合的一个或多个(图中仅示出一个)处理器710以及存储器720。其中,该存储器720中存储有可以执行前述实施例中内容的程序,而处理器710可以执行该存储器720中存储的程序,存储器720包括前述实施例中所描述的装置。

[0154] 其中,处理器710可以包括一个或者多个处理核。处理器710利用各种接口和线路连接整个电子设备700内的各个部分,通过运行或执行存储在存储器720内的指令、程序、代码集或指令集,以及调用存储在存储器720内的数据,执行电子设备700的各种功能和处理

数据。可选地,处理器710可以采用数字信号处理(Digital Signal Processing,DSP)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)、可编程逻辑阵列(Programmable Logic Array,PLA)中的至少一种硬件形式来实现。处理器710可集成中央处理器(Central Processing Unit,CPU)、视频图像处理(Graphics Processing Unit,GPU)和调制解调器等中的一种或几种的组合。其中,CPU主要处理操作系统、用户界面和应用程序等;GPU用于负责显示内容的渲染和绘制;调制解调器用于处理无线通信。可以理解的是,上述调制解调器也可以不集成到处理器710中,单独通过一块通信芯片进行实现。

[0155] 存储器720可以包括随机存储器(Random Access Memory,RAM),也可以包括只读存储器(Read-Only Memory)。存储器720可用于存储指令、程序、代码、代码集或指令集。存储器720可包括存储程序区和存储数据区,其中,存储程序区可存储用于实现操作系统的指令、用于实现至少一个功能的指令(比如触控功能、声音播放功能、视频图像播放功能等)、用于实现上述各个方法实施例的指令等。存储数据区还可以存储电子设备700在使用中所创建的数据(比如电话本、音视频数据、聊天记录数据)等。

[0156] 本领域普通技术人员可以理解,图7示的结构仅为示意,其并不对上述电子设备的结构造成限定。例如,电子设备700还可包括比图7所示更多或者更少的组件,或者具有与图7示不同的配置。

[0157] 请参考图8其示出了本申请实施例提供的一种计算机可读存储介质的结构框图。该计算机可读介质800中存储有程序代码,所述程序代码可被处理器调用执行上述方法实施例中所述的方法。

[0158] 计算机可读存储介质800可以是诸如闪存、EEPROM(电可擦除可编程只读存储器)、EPROM、硬盘或者ROM之类的电子存储器。可选地,计算机可读存储介质800包括非易失性计算机可读介质(non-transitory computer-readable storage medium)。计算机可读存储介质800具有执行上述方法中的任何方法步骤的程序代码810的存储空间。这些程序代码可以从一个或者多个计算机程序产品中读出或者写入到这一个或者多个计算机程序产品中。程序代码810可以例如以适当形式进行压缩。

[0159] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

[0160] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本申请各个实施例所述的方法。

[0161] 上面结合附图对本申请的实施例进行了描述,但是本申请并不局限于上述的具体实施方式,上述的具体实施方式仅仅是示意性的,而不是限制性的,本领域的普通技术人员在本申请的启示下,在不脱离本申请宗旨和权利要求所保护的范围情况下,还可做出很多

形式,均属于本申请的保护之内。

100

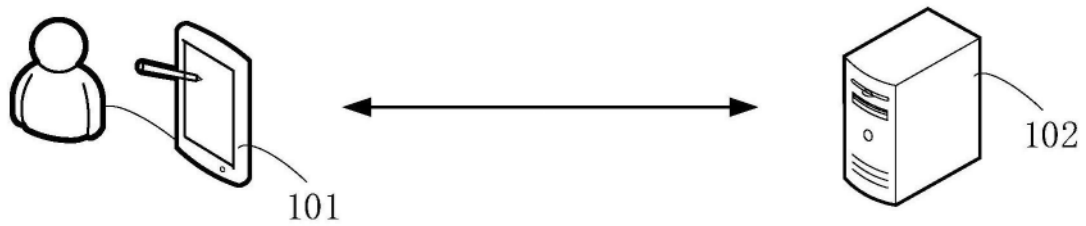


图1

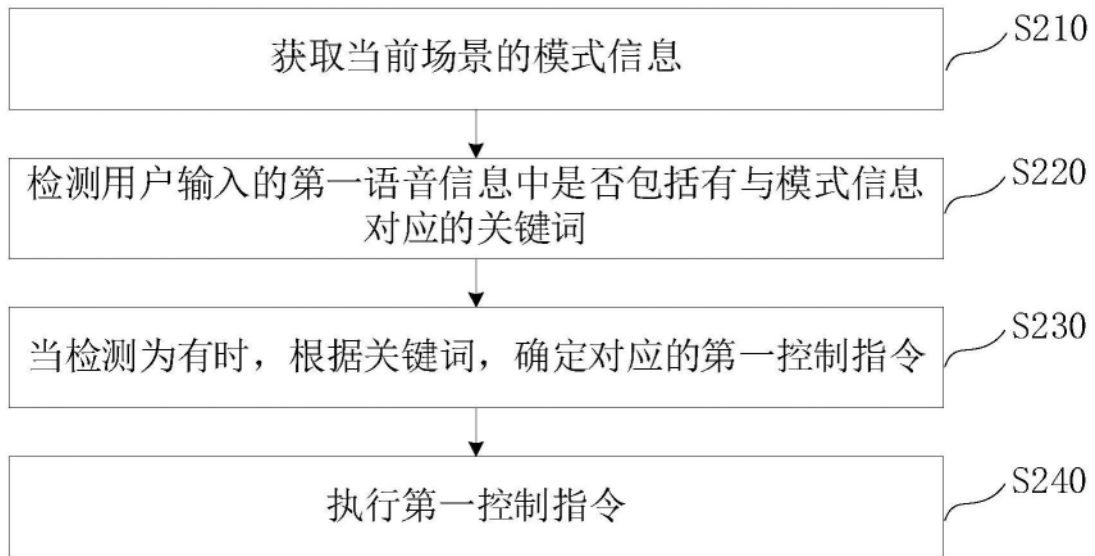


图2

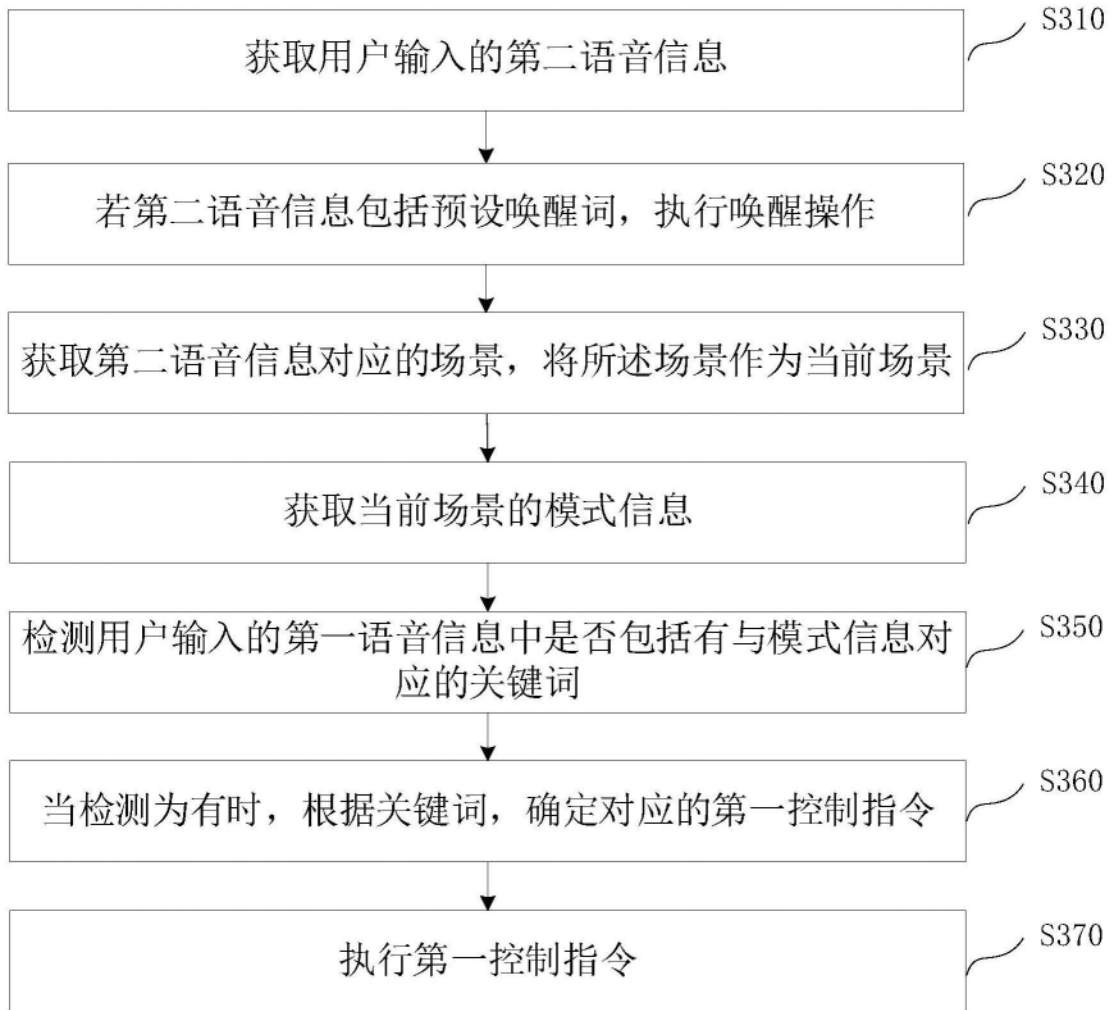


图3

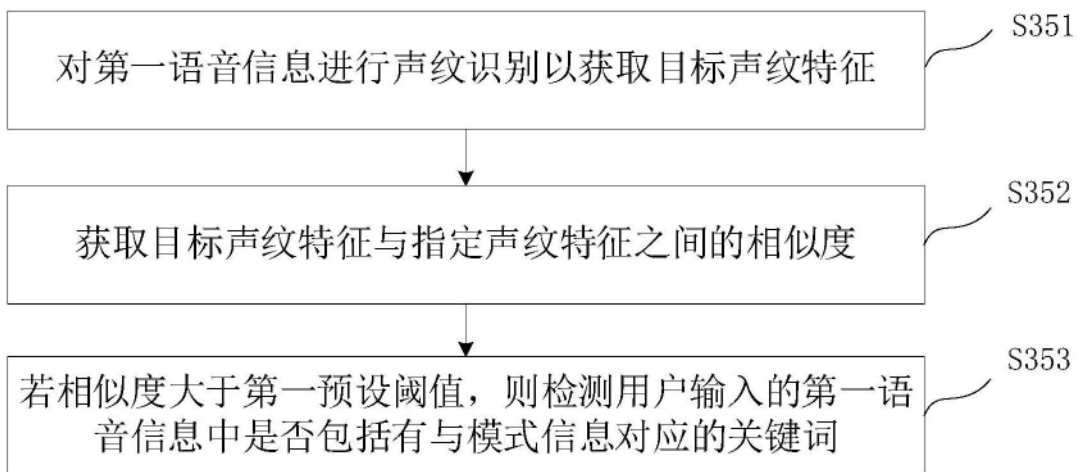


图4

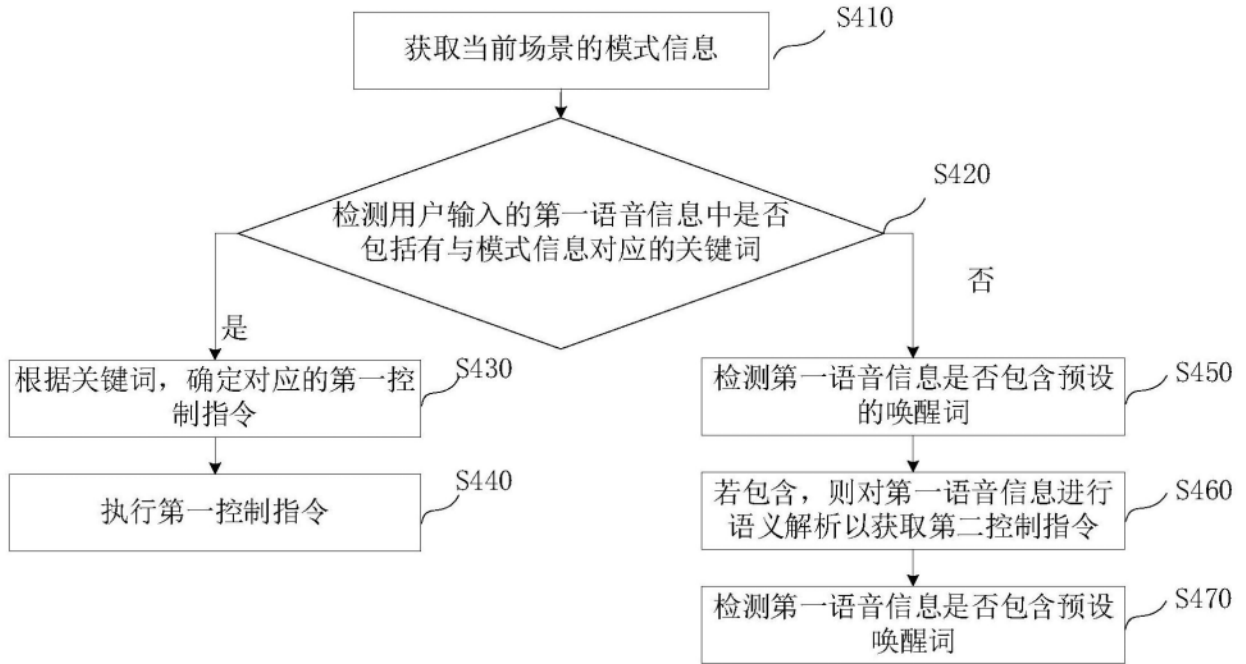


图5



图6

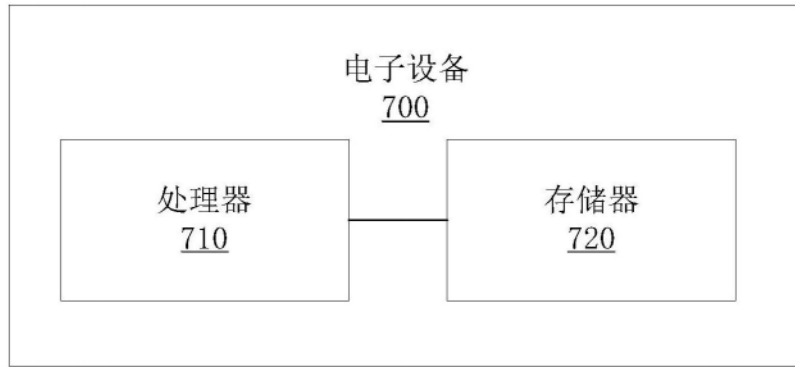


图7

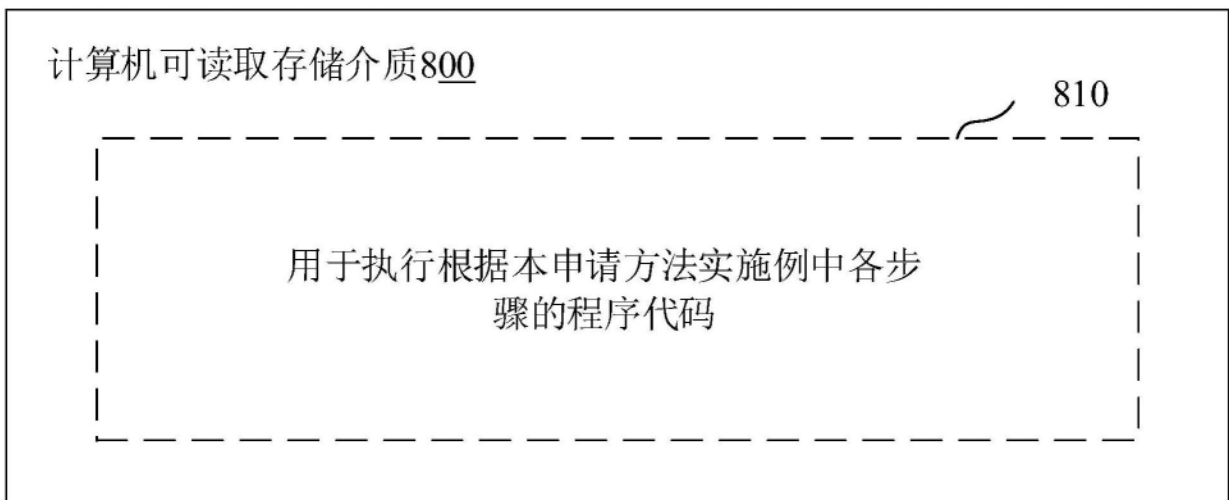


图8