



(12)发明专利申请

(10)申请公布号 CN 111367491 A

(43)申请公布日 2020.07.03

(21)申请号 202010136026.4

G10L 15/26(2006.01)

(22)申请日 2020.03.02

(71)申请人 成都极米科技股份有限公司

地址 610000 四川省成都市高新区世纪城路1129号天府软件园A区4栋1单元2层2号

(72)发明人 姜彦兮

(74)专利代理机构 北京超凡宏宇专利代理事务所(特殊普通合伙) 11463

代理人 王立红

(51)Int.Cl.

G06F 3/16(2006.01)

G06F 1/3234(2019.01)

G06F 1/3206(2019.01)

G06F 40/30(2020.01)

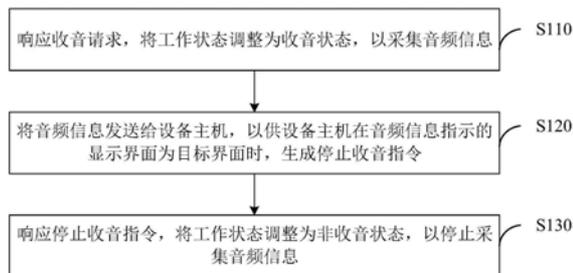
权利要求书2页 说明书12页 附图2页

(54)发明名称

语音交互方法、装置、电子设备及存储介质

(57)摘要

本申请涉及信息处理技术领域,具体而言,涉及一种语音交互方法、装置、电子设备及存储介质。本申请实施例提供的语音交互方法,包括:响应收音请求,将工作状态调整为收音状态,以采集音频信息,并将音频信息发送给设备主机,以供设备主机在音频信息指示的显示界面为目标界面时,生成停止收音指令,此后,响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。本申请实施例提供的语音交互方法、装置、电子设备及存储介质能够通过减少语音控制器处于收音状态的时间,实现降低语音控制器功耗,从而提高续航能力。



1. 一种语音交互方法,其特征在于,包括:  
响应收音请求,将工作状态调整为收音状态,以采集音频信息;  
将所述音频信息发送给设备主机,以供所述设备主机在所述音频信息指示的显示界面为目标界面时,生成停止收音指令;  
响应所述停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。
2. 根据权利要求1所述的语音交互方法,其特征在于,所述语音交互方法,还包括:  
当收音控制按键被触发时,生成所述收音请求。
3. 根据权利要求1所述的语音交互方法,其特征在于,所述将所述音频信息发送给设备主机,以供所述设备主机在所述音频信息指示的显示界面为目标界面时,生成停止收音指令,包括:  
将所述音频信息发送给设备主机,以供所述设备主机从所述音频信息中提取出语音信息,对所述语音信息进行解码,获得文本信息,并对所述文本信息进行语义分析,获得分析结果,以根据所述分析结果判断所述音频信息指示的显示界面是否为目标界面。
4. 一种语音交互方法,其特征在于,包括:  
当接收到语音控制器发送的音频信息时,判断所述音频信息指示的显示界面是否为目标界面;  
当所述音频信息指示的显示界面为目标界面时,对所述音频信息指示的显示界面进行显示,并生成停止收音指令,以供所述语音控制器响应所述停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。
5. 根据权利要求4所述的语音交互方法,其特征在于,所述当接收到语音控制器发送的音频信息时,判断所述音频信息指示的显示界面是否为目标界面,包括:  
当接收到语音控制器发送的音频信息时,从所述音频信息中提取出语音信息;  
对所述语音信息进行解码,获得文本信息;  
对所述文本信息进行语义分析,获得分析结果,以根据所述分析结果判断所述音频信息指示的显示界面是否为目标界面。
6. 根据权利要求5所述的语音交互方法,其特征在于,所述对所述文本信息进行语义分析,获得分析结果,以根据所述分析结果判断所述音频信息指示的显示界面是否为目标界面,包括:  
对所述文本信息进行语义分析,获得分析结果;  
根据所述分析结果,确定所述音频信息指示的显示界面,并获取所述显示页面对应的界面标签;  
根据所述界面标签,判断所述音频信息指示的显示界面是否为目标界面。
7. 根据权利要求4所述的语音交互方法,其特征在于,所述语音交互方法,还包括:  
当所述音频信息指示的显示界面非目标界面时,对所述音频信息指示的显示界面进行显示,并生成再次判断指令,所述再次判断指令用于控制设备主机再次执行所述当接收到语音控制器发送的音频信息时,判断所述音频信息指示的显示界面是否为目标界面的步骤。
8. 一种语音交互装置,其特征在于,包括:  
第一调整模块,用于响应收音请求,将工作状态调整为收音状态,以采集音频信息;

发送模块,用于将所述音频信息发送给设备主机,以供所述设备主机在所述音频信息指示的显示界面为目标界面时,生成停止收音指令;

第二调整模块,用于响应所述停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。

9. 一种语音交互装置,其特征在于,包括:

判断模块,用于当接收到语音控制器发送的音频信息时,判断所述音频信息指示的显示界面是否为目标界面;

第一指令生成模块,用于当所述音频信息指示的显示界面为目标界面时,对所述音频信息指示的显示界面进行显示,并生成停止收音指令,以供所述语音控制器响应所述停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。

10. 一种电子设备,其特征在于,包括处理器和存储器,所述存储器上存储有计算机程序,所述处理器用于执行所述计算机程序,以实现权利要求1~7中任意一项所述的语音交互方法。

11. 一种存储介质,所述存储介质上存储有计算机程序,其特征在于,所述计算机程序被执行时,实现权利要求1~7中任意一项所述的语音交互方法。

## 语音交互方法、装置、电子设备及存储介质

### 技术领域

[0001] 本申请涉及信息处理技术领域,具体而言,涉及一种语音交互方法、装置、电子设备及存储介质。

### 背景技术

[0002] 语音控制器是人机交互系统中,以语言驱动的控制单元。现有技术中,语音控制器往往只能通过唤醒词唤醒,以便进行后续语音交互。基于此,现有语音控制器在开机后,需要始终处于收音状态,以便进行语音检测唤醒识别,因此,功耗较高,导致续航能力较弱。

### 发明内容

[0003] 本申请实施例的目的在于,提供一种语音交互方法、装置、电子设备及存储介质以解决上述问题。

[0004] 第一方面,本申请实施例提供的语音交互方法,包括:

[0005] 响应收音请求,将工作状态调整为收音状态,以采集音频信息;

[0006] 将音频信息发送给设备主机,以供设备主机在音频信息指示的显示界面为目标界面时,生成停止收音指令;

[0007] 响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。

[0008] 本申请实施例提供的语音交互方法,包括:响应收音请求,将工作状态调整为收音状态,以采集音频信息,并将音频信息发送给设备主机,以供设备主机在音频信息指示的显示界面为目标界面时,生成停止收音指令,此后,响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。如此,便能够通过减少语音控制器处于收音状态的时间,实现降低语音控制器功耗,从而提高续航能力。

[0009] 结合第一方面,本申请实施例还提供了第一方面的第一种可选的实施方式,语音交互方法,还包括:

[0010] 当收音控制按键被触发时,生成收音请求。

[0011] 本申请实施例提供的语音交互方法,还包括:当收音控制按键被触发时,生成收音请求。由于判断收音控制按键是否被触发具有较高的准确性,因此,能够保证生成收音请求的可靠性。

[0012] 结合第一方面,本申请实施例还提供了第一方面的第二种可选的实施方式,将音频信息发送给设备主机,以供设备主机在音频信息指示的显示界面为目标界面时,生成停止收音指令,包括:

[0013] 将音频信息发送给设备主机,以供设备主机从音频信息中提取出语音信息,对语音信息进行解码,获得文本信息,并对文本信息进行语义分析,获得分析结果,以根据分析结果判断音频信息指示的显示界面是否为目标界面。

[0014] 第二方面,本申请实施例提供的语音交互方法,包括:

[0015] 当接收到语音控制器发送的音频信息时,判断音频信息指示的显示界面是否为目

标界面；

[0016] 当音频信息指示的显示界面为目标界面时，对音频信息指示的显示界面进行显示，并生成停止收音指令，以供语音控制器响应停止收音指令，将工作状态调整为非收音状态，以停止采集音频信息。

[0017] 本申请实施例语音交互方法，包括：当接收到语音控制器发送的音频信息时，判断音频信息指示的显示界面是否为目标界面，当音频信息指示的显示界面为目标界面时，对音频信息指示的显示界面进行显示，并生成停止收音指令，以供语音控制器响应停止收音指令，将工作状态调整为非收音状态，以停止采集音频信息。如此，便能够通过减少语音控制器处于收音状态的时间，实现降低语音控制器功耗，从而提高续航能力。

[0018] 结合第二方面，本申请实施例还提供了第二方面的第一种可选的实施方式，当接收到语音控制器发送的音频信息时，判断音频信息指示的显示界面是否为目标界面，包括：

[0019] 当接收到语音控制器发送的音频信息时，从音频信息中提取出语音信息；

[0020] 对语音信息进行解码，获得文本信息；

[0021] 对文本信息进行语义分析，获得分析结果，以根据分析结果判断音频信息指示的显示界面是否为目标界面。

[0022] 结合第二方面的第一种可选的实施方式，本申请实施例还提供了第二方面的第三种可选的实施方式，对文本信息进行语义分析，获得分析结果，以根据分析结果判断音频信息指示的显示界面是否为目标界面，包括：

[0023] 对文本信息进行语义分析，获得分析结果；

[0024] 根据分析结果，确定音频信息指示的显示界面，并获取显示页面对应的界面标签；

[0025] 根据界面标签，判断音频信息指示的显示界面是否为目标界面。

[0026] 结合第二方面，本申请实施例还提供了第二方面的第四种可选的实施方式，语音交互方法，还包括：

[0027] 当音频信息指示的显示界面非目标界面时，对音频信息指示的显示界面进行显示，并生成再次判断指令，再次判断指令用于控制设备主机再次执行当接收到语音控制器发送的音频信息时，判断音频信息指示的显示界面是否为目标界面的步骤。

[0028] 本申请实施例中，当音频信息指示的显示界面非目标界面时，对音频信息指示的显示界面进行显示，并生成再次判断指令，再次判断指令用于控制设备主机再次执行当接收到语音控制器发送的音频信息时，判断音频信息指示的显示界面是否为目标界面的步骤。如此，便能够通过语音交互装置，实现用户与设备主机的多轮对话，增强设备主机的控制便捷性。

[0029] 第三方面，本申请实施例提供的语音交互装置，包括：

[0030] 第一调整模块，用于响应收音请求，将工作状态调整为收音状态，以采集音频信息；

[0031] 发送模块，用于将音频信息发送给设备主机，以供设备主机在音频信息指示的显示界面为目标界面时，生成停止收音指令；

[0032] 第二调整模块，用于响应停止收音指令，将工作状态调整为非收音状态，以停止采集音频信息。

[0033] 本申请实施例提供的语音交互装置具有与上述第一方面，或第一方面的任意一种

可选的实施方式所提供的语音交互方法相同的有益效果,此处不作赘述。

[0034] 第四方面,本申请实施例提供的语音交互装置,包括:

[0035] 判断模块,用于当接收到语音控制器发送的音频信息时,判断音频信息指示的显示界面是否为目标界面;

[0036] 第一指令生成模块,用于当音频信息指示的显示界面为目标界面时,对音频信息指示的显示界面进行显示,并生成停止收音指令,以供语音控制器响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。

[0037] 本申请实施例提供的语音交互装置具有与上述第二方面,或第二方面的任意一种可选的实施方式所提供的语音交互方法相同的有益效果,此处不作赘述。

[0038] 第五方面,本申请实施例提供的电子设备,包括处理器和存储器,存储器上存储有计算机程序,处理器用于执行计算机程序,以实现上述第一方面,第一方面的第一种可选的实施方式所提供的语音交互方法,或实现上述第二方面,第二方面的任意一种可选的实施方式所提供的语音交互方法。

[0039] 本申请实施例提供的电子设备具有与上述第一方面,第一方面的第一种可选的实施方式所提供的语音交互方法相同的有益效果,或具有与上述第二方面,第二方面的任意一种可选的实施方式所提供的语音交互方法相同的有益效果,此处不作赘述。

[0040] 第六方面,本申请实施例还提供了一种存储介质,存储介质上存储有计算机程序,计算机程序被执行时,实现上述第一方面,第一方面的第一种可选的实施方式所提供的语音交互方法,或实现上述第二方面,第二方面的任意一种可选的实施方式所提供的语音交互方法。

[0041] 本申请实施例提供的存储介质具有与上述第一方面,第一方面的第一种可选的实施方式所提供的语音交互方法相同的有益效果,或具有与上述第二方面,第二方面的任意一种可选的实施方式所提供的语音交互方法相同的有益效果,此处不作赘述。

## 附图说明

[0042] 为了更清楚地说明本申请实施例的技术方案,下面将对本申请实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本申请的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,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相关的附图。

[0043] 图1为本申请实施例提供的一种语音交互方法的步骤流程图。

[0044] 图2为本申请实施例提供的语音交互方法的另一种步骤流程图。

[0045] 图3为本申请实施例提供的一种语音交互装置的示意性结构框图。

[0046] 图4为本申请实施例提供的语音交互装置的另一种示意性结构框图。

[0047] 图5为本申请实施例提供的一种电子设备的示意性结构框图。

[0048] 图6为本申请实施例提供的电子设备的另一种示意性结构框图。

[0049] 附图标记:111-第一调整模块;112-发送模块;113-第二调整模块;121-判断模块;122-第一指令生成模块;200-电子设备;210-处理器;220-存储器;230-显示器。

## 具体实施方式

[0050] 为使本申请实施例的目的、技术方案和优点更加清楚,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描述。此外,应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步定义和解释。

[0051] 第一实施例:

[0052] 请参阅图1,为本申请实施例提供的一种语音交互方法的步骤流程图,该语音交互方法应用于语音控制器。可以理解的是,本申请实施例中,语音控制器能够与控制主机通信,例如,可以通过蓝牙通信模块与控制主机通信。此外,所应说明的是,本申请实施例提供的语音交互方法不以图1及以下所示顺序为限制,以下结合图1对本申请实施例提供的语音交互方法的步骤流程进行描述。

[0053] 步骤S110,响应收音请求,将工作状态调整为收音状态,以采集音频信息。

[0054] 本申请实施例中,语音控制器的工作状态包括收音状态和非收音状态,当语音控制器的工作状态为收音状态时,语音控制器能够采集应用环境的音频信息,并将音频信息发送给设备主机,当语音控制器的工作状态为非收音状态时,语音控制器停止采集应用环境的音频信息。此外,可以理解的是,本申请实施例中,当语音控制器开机或开启语音交互功能之后,工作状态默认为非收音状态,且只有当收音请求生成时,才会响应收音请求,将工作状态调整为收音状态,而关于收音请求的生成,本申请实施例中,作为一种可选的实施方式,其可以包括步骤S011。

[0055] 步骤S011,当收音控制按键被触发时,生成收音请求。

[0056] 本申请实施例中,语音控制器上设置有收音控制按键,收音控制按键可以是机械按键,也可以是触摸式按键,例如,电阻式触摸按键、电容式触摸按键等。用户可以通过触发射音控制按键,生成收音请求,因此,可以对收音控制按键进行监控,并在监控结果指示,收音控制按键被触发时,生成收音请求。

[0057] 此外,本申请实施例中,语音控制器可以通过专用的编码译码器(Coder/Decoder, Codec)芯片采集音频信息,并对音频信息进行模数转换,以将以模拟信号形式采集的音频信息,转换为以数字信号形式存在的音频信息。

[0058] 步骤S120,将音频信息发送给设备主机,以供设备主机在音频信息指示的显示界面为目标界面时,生成停止收音指令。

[0059] 为保证语音交互的及时性,本申请实施例中,语音控制器可以在采集音频信息的同时,将采集的音频信息,实时发送给设备主机。当然,为降低音频信息的数据传输压力,同时,保证信号传输的可靠性,本申请实施例中,也可以每间隔预设时长,将当前时间点之前,预设时长内采集的音频信息发送给设备主机,而预设时长可以是,但不限于,10ms、15ms、20ms。

[0060] 此外,对于音频信息的发送方式,本申请实施例中,可以通过蓝牙模块将音频信息发送给设备主机。基于此,在将音频信息发送给设备主机之前,还将对音频信息进行格式转换,例如,将音频信息转换为高级音频编码(Advanced Audio Coding, ACC)格式、子带编码(Sub-Band Coding, SBC)格式、APT-X格式、LDAC格式、高音质无线音频(Hi-Res Wireless Audio, HWA)格式等蓝牙音频传输格式中的任意一种,其中,APT-X格式为一种通过基于子带

自适应差分脉冲编码调制 (Sub-Band Adaptive Differential Pulse Code Modulation, SB-ADPCM) 技术的数字音频压缩算法对音频信息进行格式转换之后,获得的蓝牙音频传输格式,而LDAC格式为一种接近无损的编码格式。

[0061] 当设备主机接收到语音控制器发送的音频信息时,可以判断音频信息指示的显示界面是否为目标界面,当音频信息指示的显示界面为目标界面时,对音频信息指示的显示界面进行显示,并生成停止收音指令,将停止收音指令发送给语音控制器。

[0062] 本申请实施例中,设备主机判断音频信息指示的显示界面是否为目标界面的方法,可以为:设备主机从所述音频信息中提取出语音信息,对所述语音信息进行解码,获得文本信息,并对所述文本信息进行语义分析,获得分析结果,以根据所述分析结果判断所述音频信息指示的显示界面是否为目标界面。具体可参考第二实施例提供的语音交互方法中,关于步骤S211、步骤S212和步骤S213的详细描述,此处不作赘述。

[0063] 步骤S130,响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。

[0064] 通过上述设置,能够通过减少语音控制器处于收音状态的时间,也即,在用户有视频播放需求时,通过触发收音控制按键,生成收音请求,以使语音控制器将工作状态调整为收音状态,采集音频信息,同时,将音频信息发送给设备主机,设备主机用于在音频信息指示的显示界面为目标界面时,生成停止收音指令,而语音控制器在接收到设备主机发送的停止收音指令,根据停止收音指令将工作状态调整为非收音状态,实现降低语音控制器功耗,从而提高续航能力。

[0065] 第二实施例:

[0066] 请参阅图2,为本申请实施例提供的一种语音交互方法的步骤流程图,该语音交互方法应用于设备主机。所应说明的是,本申请实施例提供的语音交互方法不以图2及以下所示的顺序为限制,以下结合图2对本申请实施例提供的语音交互方法的步骤流程进行描述。

[0067] 步骤S210,当接收到语音控制器发送的音频信息时,判断音频信息指示的显示界面是否为目标界面。

[0068] 本申请实施例中,设备主机可以是具有视频播放功能的用户终端,而用户终端可以是,但不限于,智能电视、投影控制器、个人电脑(Personal Computer,PC)、个人数字助理(Personal Digital Assistant,PDA)、移动上网设备(Mobile Internet Device,MID)等。当用户有视频播放需求时,可以通过发出语音信息,以使工作状态为收音状态的语音控制器,采集包括前述语音信息的音频信息,并将音频信息发送给设备主机,当设备主机接收到语音控制器发送的音频信息时,便可以判断音频信息指示的显示界面是否为目标界面。

[0069] 需要说明的是,本申请实施例中,目标界面为视频播放的最底层界面,例如,视频选择页面的首页包括“电影”、“电视剧”、“综艺”等选择标签,“电影”标签对应的下层界面包括“华语”、“美国”、“韩国”等地区标签,“喜剧”、“恐怖”、“悬疑”等类型标签,以及“2019”、“2018”、“2017”等年份标签,而“喜剧”标签对应的下层界面包括“料理鼠王”、“小鬼当家”、“楚门的世界”等影片标签,那么,“料理鼠王”、“小鬼当家”、“楚门的世界”等影片标签对应的显示界面即为最底层界面,“电影”、“电视剧”、“综艺”等选择标签,“电影”标签对应的下层界面包括“华语”、“美国”、“韩国”等地区标签,“喜剧”、“恐怖”、“悬疑”等类型标签,以及“2019”、“2018”、“2017”等年份标签对应的显示界面即为非最底层界面,也可以理解为,针

对某个标签而言,当其不存在对应的下层界面时,其即为最底层界面,也即,目标界面,当其存在对应的下层界面时,其即为非最底层界面,也即,非目标界面。

[0070] 此外,对于步骤S210,本申请实施例中,其包括步骤S211、步骤S212和步骤S213。

[0071] 步骤S211,当接收到语音控制器发送的音频信息时,从音频信息中提取出语音信息。

[0072] 由于采集的音频信息中,除用户发出的语音信息外,还包括背景噪音,因此,本申请实施例中,步骤S211可以包括:当接收到语音控制器发送的音频信息时,从音频信息中,检测出人声的起始点和结束点,以根据起始点和结束点,从音频信息中提取出语音信息,而该过程可以基于语音端点检测(Voice Activity Detection,VAD)技术实现。如此,即便在背景噪音严重的复杂应用环境下,也能够从音频信息中提取出语音信息,从而保证语音交互方法的可靠性。

[0073] 对于VAD技术,本申请实施例中,作为第一种可选的实施方式,可以将采集的每段音频信息中,起始处的目标长度段子音频信息作为背景音频,而目标长度段子音频信息对应的时间长度可以位于时间长度区间[200ms,400ms]。由于用户在触发收音控制按键的瞬间,或触发收音控制按键的目标长度段时长内,通常是还没开始发出语音信息的,因此,本申请实施例中,可以将背景音频作为静音音频,基于此,便可以将背景音频中,每帧音频数据的能量均值,作为基准值。此后,将音频信息中,能量值大于基准值的音频数据,作为有效数据,将能量值小于或等于基准值的音频数据,作为无效数据,而语音信息即为所有有效数据的集合。

[0074] 对于VAD技术,本申请实施例中,作为第二种可选的实施方式,也可以对音频信息进行分帧处理,获得多帧音频数据,此后,对每帧音频数据进行音频特征提取,例如,提取对数帧能量、过零率(Zero Crossing Rate,ZCR)提取、在延迟一个位置处标准化自相关系数、Pth阶线性预测的第一个系数、Pth阶线性预测误差的对数等音频特征中的至少一种。此后,通过预设分类模型,对多帧音频数据进行分类,以确定每帧音频数据是否为有效数据,而语音信息即为所有有效数据的集合。

[0075] 可以理解的是,本申请实施例中,上述第二种可选的实施方式中,预设分类模型可以通过构建预设模型,并基于音频样本数据集,对预设模型进行训练获得。其中,预设模型可以是支持向量机(Support Vector Machine,SVM)分类器,而音频样本数据集中,包括多帧样本音频,且每帧样本音频具有对应的音频标签,以及对应的音频特征,其中,音频标签包括语音信息标签和非语音信息标签。

[0076] 步骤S212,对语音信息进行解码,获得文本信息。

[0077] 对于步骤S212,本申请实施例中,可以基于自动语音识别(Automatic Speech Recognition,ASR)技术实现。ASR技术的原理为,对语音信息进行分割,获得多段子语音信息,并对每段子语音信息进行编码,以将每段子语音信息转换为数字向量,此后,通过预设声学模型,将每个数字向量转换为中文文字,此后,对转换获得的每个中文文字进行连词组合,即可获得文本信息。

[0078] 可以理解的是,本申请实施例中,上述预设声学模型可以为基于语言、文本数据集,训练之后的卷积神经网络(Convolutional Neural Network,CNN),本申请实施例,对此不作赘述。

[0079] 步骤S213,对文本信息进行语义分析,获得分析结果,以根据分析结果判断音频信息指示的显示界面是否为目标界面。

[0080] 本申请实施例中,每个显示界面具有对应的界面标签,用于表征显示界面是否为目标界面,而界面标签可以是数字标签,例如,针对某个显示界面而言,其界面标签为“1”,则确定该显示界面为目标界面,若界面标签不为“1”,则确定该显示界面为非目标界面。基于此,本申请实施例中,步骤S213可以包括:对文本信息进行语义分析,获得分析结果,此后,根据分析结果,确定音频信息指示的显示界面,并获取显示页面对应的界面标签,再根据界面标签,判断音频信息指示的显示界面是否为目标界面。

[0081] 对于对文本信息进行语义分析,获得分析结果的步骤,本申请实施例中,可以基于自然语言处理(Natural Language Processing,NLP)技术实现,而NLP技术可以包括句法语义分析技术和信息提取技术。本申请实施例中,可以基于句法语义技术分析对文本信息进行分词、词性标记、命名实体识别和链接、句法分析、语义角色识别,以及多义词消歧等操作,获得第一目标文本。此后,基于信息提取技术从第一目标文本中提取出动作、对象等关键信息,并对关键信息进行连词组合,获得第二目标文本,作为分析结果。例如,当第一目标文本为“我想播放电影料理鼠王”,则基于信息提取技术从第一目标文本中提取出的关键信息包括动作“播放”和对象“料理鼠王”,此后,对关键信息“播放”和“料理鼠王”进行连词组合,获得的第二目标文本为“播放料理鼠王”。

[0082] 在对文本信息进行语义分析,获得分析结果之后,便能够根据分析结果,确定音频信息指示的显示界面,并获取显示页面对应的界面标签,以根据界面标签,判断音频信息指示的显示界面是否为目标界面。

[0083] 步骤S220,当音频信息指示的显示界面为目标界面时,对音频信息指示的显示界面进行显示,并生成停止收音指令,以供语音控制器响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。

[0084] 本申请实施例中,当音频信息指示的显示界面为目标界面时,可以生成界面播放指令,并根据界面播放指令对音频信息指示的显示界面进行显示,同时,生成停止收音指令,并将停止收音指令发送给语音控制器,此后,语音控制器响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。如此,便能够通过减少语音控制器处于收音状态的时间,也即,在用户有视频播放需求时,通过触发收音控制按键,生成收音请求,以使语音控制器将工作状态调整为收音状态,采集音频信息,同时,将音频信息发送给设备主机,设备主机用于在音频信息指示的显示界面为目标界面时,生成停止收音指令,而语音控制器在接收到设备主机发送的停止收音指令,根据停止收音指令将工作状态调整为非收音状态,实现降低语音控制器功耗,从而提高续航能力。

[0085] 为实现用户与设备主机的多轮对话,增强设备主机的控制便捷性,本申请实施例提供的语音交互装置,还可以包括步骤S230,当音频信息指示的显示界面非目标界面时,对音频信息指示的显示界面进行显示,并生成再次判断指令,再次判断指令用于控制设备主机再次执行当接收到语音控制器发送的音频信息时,判断音频信息指示的显示界面是否为目标界面的步骤。

[0086] 此外,基于步骤S230,本申请实施例提供的语音交互装置相对于现有技术中,在“搜索-确认”环节,用户需要再次触发设置于语音控制器上的确认按键进行确认动作的方

案,也即,在搜索确定音频信息指示的显示界面之后,需要触发确认按键,生成确认指令,并将确认指令发送至设备主机,以使设备主机在接收到确认指令时,确认搜索获得的显示界面为用户的意向界面,并对搜索获得的显示界面进行显示的方案而言,在“搜索-确认”环节,用户无需再次按键确认,从而提高了语音交互方法的便捷程度。

[0087] 为方便理解,以下将举例,对结合第一实施例提供的语音交互方法,以及第二实施例提供的语音交互方法所实现的语音交互系统的第一种工作流程进行描述。

[0088] 同样,以视频选择页面的首页包括“电影”、“电视剧”、“综艺”等选择标签,“电影”标签对应的下层界面包括“华语”、“美国”、“韩国”等地区标签,“喜剧”、“恐怖”、“悬疑”等类型标签,以及“2019”、“2018”、“2017”等年份标签,而“喜剧”标签对应的下层界面包括“料理鼠王”、“小鬼当家”、“楚门的世界”等影片标签,且用户还未确定目标影片为例。

[0089] 用户在通过触发收音控制按键,生成收音请求之后,语音控制器将工作状态调整为收音状态,以采集音频信息,此后,用户可以发出第一语音信息“打开视频选择页面”,并将第一音频信息发送给设备主机,由于第一音频信息指示的显示界面非目标界面,因此,设备主机对第一音频信息指示的显示界面进行显示,也即,对视频选择页面的首页进行显示之后,生成再次判断指令。

[0090] 此后,用户可以发出第二语音信息“打开电影选择标签”,语音控制器采集包括第二语音信息的第二音频信息之后,将第二音频信息发送给设备主机,由于第二音频信息指示的显示界面非目标界面,因此,设备主机对第二音频信息指示的显示界面进行显示,也即,对电影选择标签对应的显示界面进行显示之后,生成再次判断指令。

[0091] 接着,用户可以发出第三语音信息“打开喜剧类型标签”,语音控制器采集包括第三语音信息的第三音频信息之后,将第三音频信息发送给设备主机,由于第三音频信息指示的显示界面非目标界面,因此,设备主机对第三音频信息指示的显示界面进行显示,也即,对喜剧类型标签对应的显示界面进行显示之后,生成再次判断指令。

[0092] 若用户将位于喜剧类型标签对应的显示界面首页的影片标签“料理鼠王”对应的影片作为目标影片,则用户可以发出第四语音信息“播放料理鼠王”,语音控制器采集包括第四语音信息的第四音频信息之后,将第四音频信息发送给设备主机,由于第四音频信息指示的显示界面为目标界面,因此,设备主机对第四音频信息指示的显示界面进行显示,也即,播放影片《料理鼠王》,并生成停止收音指令,将停止收音指令发送给语音控制器,语音控制器则根据停止收音指令将工作状态调整为非收音状态。

[0093] 在对喜剧类型标签对应的显示界面进行显示之后,用户也可以发出类似“翻到下一页”之类的第五语音信息,用于指示翻页,语音控制器采集包括第五语音信息的第五音频信息之后,将第五音频信息发送给设备主机,由于第五音频信息指示的显示界面非目标界面,因此,设备主机对第五音频信息指示的显示界面进行显示,也即,对喜剧类型标签对应的显示界面首页的下一页面进行显示之后,生成再次判断指令,此后,可以重复前述动作,直至用户成功播放目标影片,并生成停止收音指令,将停止收音指令发送给语音控制器,语音控制器则根据停止收音指令将工作状态调整为非收音状态。

[0094] 以下将再举例,对结合第一实施例提供的语音交互方法,以及第二实施例提供的语音交互方法所实现的语音交互系统的第二种工作流程进行描述。

[0095] 以视频选择页面的首页包括“电影”、“电视剧”、“综艺”等选择标签,“电影”标签对

应的下层界面包括“华语”、“美国”、“韩国”等地区标签，“喜剧”、“恐怖”、“悬疑”等类型标签，以及“2019”、“2018”、“2017”等年份标签，而“喜剧”标签对应的下层界面包括“料理鼠王”、“小鬼当家”、“楚门的世界”等影片标签，且用户已经确定将《料理鼠王》作为目标影片为例。

[0096] 用户在通过触发收音控制按键，生成收音请求之后，语音控制器将工作状态调整为收音状态，以采集音频信息，此后，用户可以发出第六语音信息“播放料理鼠王”，并将第六音频信息发送给设备主机，由于第六音频信息指示的显示界面为目标界面，因此，设备主机对第六音频信息指示的显示界面进行显示，也即，播放影片《料理鼠王》，并生成停止收音指令，将停止收音指令发送给语音控制器，语音控制器则根据停止收音指令将工作状态调整为非收音状态。

[0097] 第三实施例：

[0098] 基于与第一实施例提供的语音交互方法同样的发明构思，本申请实施例还提供了一种语音交互装置，请参阅图3，本申请实施例提供的语音交互装置包括第一调整模块111、发送模块112和第二调整模块113。

[0099] 第一调整模块111，用于响应收音请求，将工作状态调整为收音状态，以采集音频信息。

[0100] 关于第一调整模块111的描述具体可参考第一实施例提供的语音交互方法中，关于步骤S110的详细描述，也即，步骤S110可以由第一调整模块111执行。

[0101] 发送模块112，用于将音频信息发送给设备主机，以供设备主机在音频信息指示的显示界面为目标界面时，生成停止收音指令。

[0102] 发送模块112，具体用于将音频信息发送给设备主机，以供设备主机从音频信息中提取出语音信息，对语音信息进行解码，获得文本信息，并对文本信息进行语义分析，获得分析结果，以根据分析结果判断音频信息指示的显示界面是否为目标界面。

[0103] 关于发送模块112的描述具体可参考第一实施例提供的语音交互方法中，关于步骤S120的详细描述，也即，步骤S120可以由发送模块112执行。

[0104] 第二调整模块113，用于响应停止收音指令，将工作状态调整为非收音状态，以停止采集音频信息。

[0105] 关于第二调整模块113的描述具体可参考第一实施例提供的语音交互方法中，关于步骤S130的详细描述，也即，步骤S130可以由第二调整模块113执行。

[0106] 本申请实施例提供的语音交互装置，还可以包括请求生成模块。

[0107] 请求生成模块，用于当收音控制按键被触发时，生成收音请求。

[0108] 关于请求生成模块的描述具体可参考第一实施例提供的语音交互方法中，关于步骤S011的详细描述，也即，步骤S011可以由请求生成模块执行。

[0109] 第四实施例：

[0110] 基于与第二实施例提供的语音交互方法同样的发明构思，本申请实施例还提供了一种语音交互装置，请参阅图4，本申请实施例提供的语音交互装置包括判断模块121和第一指令生成模块122。

[0111] 判断模块121，用于当接收到语音控制器发送的音频信息时，判断音频信息指示的显示界面是否为目标界面。

[0112] 关于判断模块121的描述具体可参考第二实施例提供的语音交互方法中,关于步骤S210的详细描述,也即,步骤S210可以由判断模块121执行。

[0113] 第一指令生成模块122,用于当音频信息指示的显示界面为目标界面时,对音频信息指示的显示界面进行显示,并生成停止收音指令,以供语音控制器响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。

[0114] 关于第一指令生成模块122的描述具体可参考第二实施例提供的语音交互方法中,关于步骤S220的详细描述,也即,步骤S220可以由第一指令生成模块122执行。

[0115] 本申请实施例中,判断模块121可以包括信息提取单元、信息编码单元和语义分析单元。

[0116] 信息提取单元,用于当接收到语音控制器发送的音频信息时,从音频信息中提取出语音信息。

[0117] 本申请实施例中,信息提取单元,具体用于当接收到语音控制器发送的音频信息时,从音频信息中,检测出人声的起始点和结束点,根据起始点和结束点,从音频信息中提取出语音信息。

[0118] 关于信息提取单元的描述具体可参考第二实施例提供的语音交互方法中,关于步骤S211的详细描述,也即,步骤S211可以由信息提取单元执行。

[0119] 信息编码单元,用于对语音信息进行解码,获得文本信息。

[0120] 关于信息编码单元的描述具体可参考第二实施例提供的语音交互方法中,关于步骤S212的详细描述,也即,步骤S212可以由信息编码单元执行。

[0121] 语义分析单元,用于对文本信息进行语义分析,获得分析结果,以根据分析结果判断音频信息指示的显示界面是否为目标界面。

[0122] 本申请实施例中,语义分析单元,具体用于对文本信息进行语义分析,获得分析结果,此后,根据分析结果,确定音频信息指示的显示界面,并获取显示页面对应的界面标签,再根据界面标签,判断音频信息指示的显示界面是否为目标界面。

[0123] 关于语义分析单元的描述具体可参考第二实施例提供的语音交互方法中,关于步骤S213的详细描述,也即,步骤S213可以由语义分析单元执行。

[0124] 本申请实施例提供的语音交互装置,还可以包括第二指令生成模块。

[0125] 第二指令生成模块,用于当音频信息指示的显示界面非目标界面时,对音频信息指示的显示界面进行显示,并生成再次判断指令,再次判断指令用于控制设备主机再次执行当接收到语音控制器发送的音频信息时,判断音频信息指示的显示界面是否为目标界面的步骤。

[0126] 关于第二指令生成模块的描述具体可参考第二实施例提供的语音交互方法中,关于步骤S230的详细描述,也即,步骤S230可以由第二指令生成模块执行。

[0127] 第五实施例:

[0128] 请参阅图5,为本申请实施例提供的一种电子设备200的示意性结构框图。可以理解的是,当电子设备200为应用第一实施例提供的语音交互方法,或应用第三实施例提供的语音交互装置的设备时,其可以为语音控制器,当电子设备200为应用第二实施例提供的语音交互方法,或应用第四实施例提供的语音交互装置的设备时,其为设备主机,设备主机可以是具有视频播放功能的用户终端,而用户终端可以是,但不限于,智能电视、投影控制器、

PC、PDA、MID等。此外,在结构上,电子设备200可以包括处理器210和存储器220。

[0129] 处理器210与存储器220直接或间接地电性连接,以实现数据的传输或交互,例如,这些元件相互之间可通过一条或多条通讯总线或信号线实现电性连接。语音交互装置包括至少一个可以软件或固件(Firmware)的形式存储在存储器220中或固化在电子设备200的操作系统(Operating System,OS)中的软件模块。处理器210用于执行存储器220中存储的可执行模块,例如,语音交互装置所包括的软件功能模块及计算机程序等,以实现语音交互方法。处理器210可以在接收到执行指令后,执行计算机程序。

[0130] 本申请实施例中,处理器210可以是一种集成电路芯片,具有信号处理能力。处理器210也可以是通用处理器,例如,可以是数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、分立门或晶体管逻辑器件、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。此外,通用处理器可以是微处理器或者任何常规处理器等。

[0131] 存储器220可以是,但不限于,随机存取存储器(Random Access Memory, RAM)、只读存储器(Read Only Memory, ROM)、可编程只读存储器(Programmable Read-Only Memory, PROM)、可擦可程序只读存储器(Erasable Programmable Read-Only Memory, EPROM),以及电可擦编程只读存储器(Electric Erasable Programmable Read-Only Memory, EEPROM)。存储器220用于存储程序,处理器210在接收到执行指令后,执行该程序。

[0132] 请参阅图6,本申请实施例中,当电子设备200为应用第二实施例提供的语音交互方法,或应用第四实施例提供的语音交互装置的设备时,其还可以包括显示器230,显示器230与处理器210直接或间接地电性连接,以实现数据的传输或交互,例如,这些元件相互之间可通过一条或多条通讯总线或信号线实现电性连接。

[0133] 本申请实施例中,显示器230可以是,但不限于,阴极射线显像管(Cathode Ray Tube, CRT)显示器、液晶显示器(Liquid Crystal Display, LCD)、等离子平板显示器(Plasma Display Panel, PDP)、有机发光二极管(Organic Light-Emitting Diode, OLED)显示器,用于对音频信息指示的显示界面进行显示。

[0134] 应当理解,图5和图6所示的结构仅为示意,本申请实施例提供的电子设备200还可以具有比图5和图6更少或更多的组件,或是具有与图5和图6所示不同的配置。

[0135] 第六实施例:

[0136] 本申请实施例还提供了一种存储介质,存储介质上存储有计算机程序,计算机程序被执行时,实现第一实施例提供的语音交互方法,或实现第二实施例提供的语音交互方法,具体可参见上述方法实施例,本申请实施例中对此不作赘述。

[0137] 综上所述,本申请实施例中,应用于语音控制器的语音交互方法,包括:响应收音请求,将工作状态调整为收音状态,以采集音频信息,并将音频信息发送给设备主机,以供设备主机在音频信息指示的显示界面为目标界面时,生成停止收音指令,此后,响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。对应的,应用于设备主机的语音交互方法,包括:接收到语音控制器发送的音频信息时,判断音频信息指示的显示界面是否为目标界面,当音频信息指示的显示界面为目标界面时,对音频信息指示的显示界面进行显示,并生成停止收音指令,以供语音控制器响应停止收音指令,将工作状态调整为非收音状态,以停止采集音频信息。基于此,本申请实施例提供的语音交互方法、装置、电子设

备及存储介质,能够通过减少语音控制器处于收音状态的时间,实现降低语音控制器功耗,从而提高续航能力。

[0138] 在本申请所提供的上述几个实施例中,应该理解到,所揭露的装置和方法,也可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的,例如,附图中的流程图和框图显示了根据本申请的多个实施例的装置、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现方式中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0139] 另外,在本申请每个实施例中的各功能模块可以集成在一起形成一个独立的部分,也可以是每个模块单独存在,也可以两个或两个以上模块集成形成一个独立的部分。

[0140] 所述功能如果以软件功能模块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(可以是个人计算机,服务器,或者网络设备)执行本申请每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、RAM、ROM、磁碟或者光盘等各种可以存储程序代码的介质。

[0141] 以上所述仅为本申请的实施例而已,并不用于限制本申请的保护范围,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

[0142] 此外,需要说明的是,在本文中,诸如“第一”、“第二”、“第三”等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,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、物品或者设备不仅包括那些要素,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、物品或者设备所固有的要素。

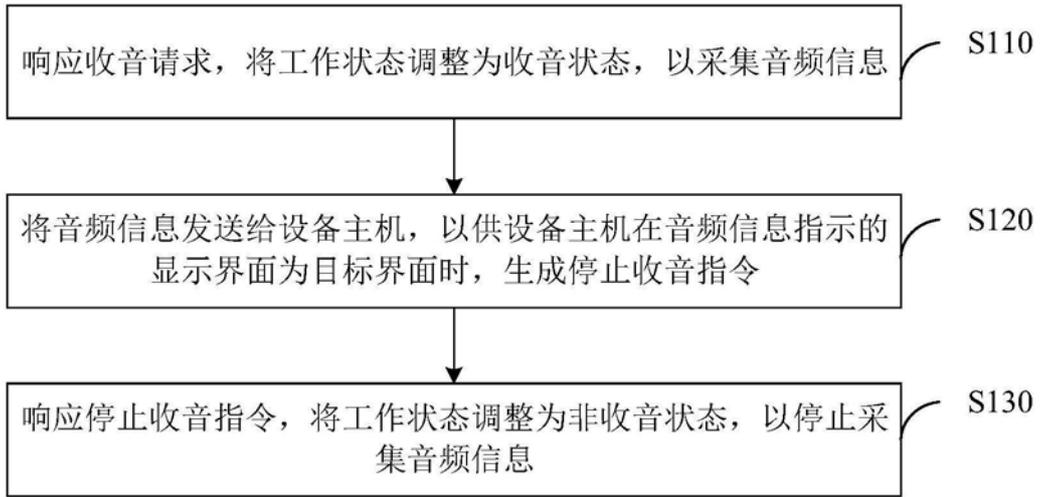


图1

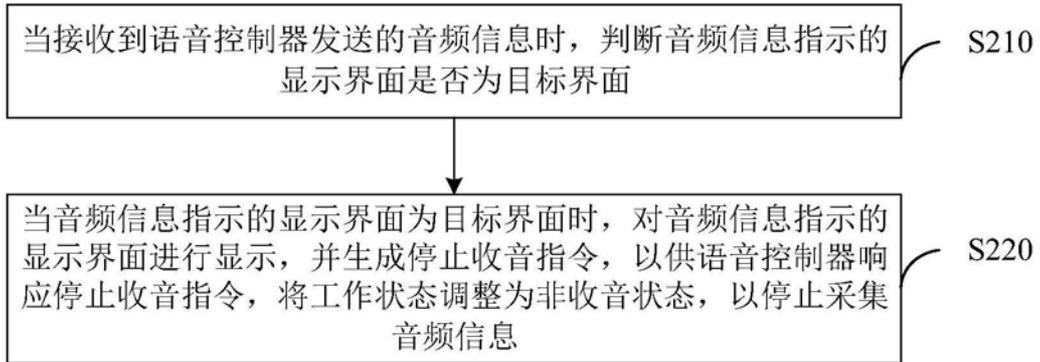


图2

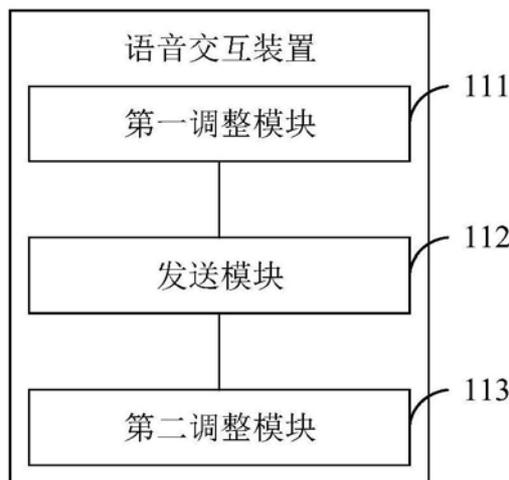


图3

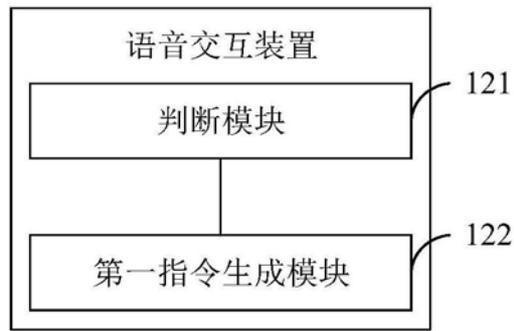


图4

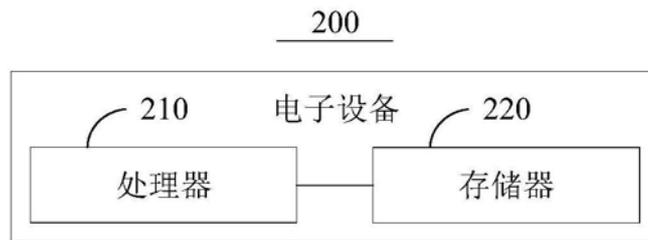


图5



图6