(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113053380 B
(45)授权公告日 2023.12.01

(21)申请号 202110335864.9

(22)申请日 2021.03.29

(65)同一申请的已公布的文献号
申请公布号 CN 113053380 A

(43)申请公布日 2021.06.29

(73)专利权人 海信电子科技(武汉)有限公司
地址 430073 湖北省武汉市东湖新技术开发区软件园东路1号软件产业4.1期B2栋13层02号-2

(72)发明人 胡帆 雷将 徐侃

(74)专利代理机构 北京弘权知识产权代理有限公司 11363
专利代理师 逯长明 许伟群

(51)Int.Cl.
G10L 15/22(2006.01)
G10L 15/26(2006.01)
G10L 21/0208(2013.01)
G10L 15/04(2013.01)
H04N 21/472(2011.01)
H04N 21/422(2011.01)

(56)对比文件
CN 101510815 A,2009.08.19
CN 101510886 A,2009.08.19
CN 102868635 A,2013.01.09
CN 104616652 A,2015.05.13
CN 107404446 A,2017.11.28
CN 108683635 A,2018.10.19
CN 110324303 A,2019.10.11
CN 110971352 A,2020.04.07
JP 2018049058 A,2018.03.29
US 2006029102 A1,2006.02.09
US 2018096695 A1,2018.04.05
US 2019253477 A1,2019.08.15
WO 2016129188 A1,2016.08.18

审查员 王欣

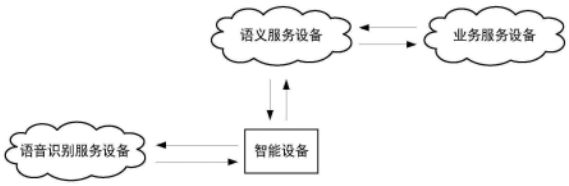
权利要求书2页 说明书13页 附图7页

(54)发明名称

服务器及语音识别方法

(57)摘要

本申请实施例提供了一种服务器及语音识别方法,服务器被配置为:接收来自显示设备的语音分片数据;若所述语音分片数据不是语音会话的最后一片数据,根据服务器未向语音识别服务设备发送所述语音分片数据的上一片数据,暂不发送所述语音分片数据,直到所述服务器已发送所述上一片数据,或所述上一片数据被标记为忽略状态,再将所述语音分片数据发送给所述语音识别服务设备;若所述语音分片数据是所述最后一片数据,在所述最后一片数据之前非忽略状态的语音分片数据均已发送,或等待的时间超过预设时间阈值时,再将已接收还未发送、且未被标记为忽略状态的语音分片数据发送给语音识别服务设备。本申请解决了语音识别准确性低的技术问题。



CN 113053380 B

1.一种服务器,其特征在于,所述服务器被配置为:

接收来自显示设备的语音分片数据;

若所述语音分片数据不是语音会话的最后一片数据,根据服务器未向语音识别服务设备发送所述语音分片数据的上一片数据,暂不发送所述语音分片数据,直到所述服务器已发送所述上一片数据,或所述上一片数据被标记为忽略状态,再将所述语音分片数据发送给所述语音识别服务设备,其中,对于未接收到的语音分片数据,若接收到排序在后的语音分片数据的数量达到预设数量阈值,则将未接收到的语音分片数据标记为忽略状态,所述忽略状态表示不再将所述语音分片数据发送给所述语音识别服务设备;

若所述语音分片数据是所述最后一片数据,在所述最后一片数据之前的非忽略状态的语音分片数据均已发送,或等待的时间超过预设时间阈值时,再将已接收还未发送、且未被标记为忽略状态的语音分片数据发送给语音识别服务设备。

2.根据权利要求1所述的服务器,其特征在于,所述服务器还被配置为:

对于未接收到的语音分片数据,若接收到排序在后的语音分片数据的数量达到预设数量阈值,将所述未接收到的语音分片数据标记为忽略状态。

3.根据权利要求1所述的服务器,其特征在于,所述服务器还被配置为:

对于未接收到且未被标记为忽略状态的语音分片数据,若距离所述最后一片数据的顺序越近,则所述语音分片数据对应的预设等待时间越大,所述预设时间阈值为全部预设等待时间中的最大值。

4.根据权利要求1所述的服务器,其特征在于,所述语音分片数据包括语音数据和分片参数,所述分片参数包括分片序号,所述分片序号用于确定所述语音分片数据在语音会话中的顺序。

5.根据权利要求4所述的服务器,其特征在于,所述分片序号包括数组,所述数组包括两个数值,所述数组的第一个数值表示所述语音分片数据在语音会话中的顺序,所述数组的第二个数值表示所述语音分片数据的下一分片数据在所述语音会话中的顺序。

6.根据权利要求4所述的服务器,其特征在于,所述分片序号表示所述语音分片数据在语音会话中的顺序,最后一片数据的分片参数还包括所述语音会话的结束标识。

7.根据权利要求1所述的服务器,其特征在于,所述服务器还被配置为:

从接收到所述语音会话的第一片数据开始,检测最后一次接收到语音分片数据的时间距与当前时间的差值是否大于预设超时阈值,若大于,则确定所述语音会话已结束。

8.根据权利要求7所述的服务器,其特征在于,所述检测最后一次接收到语音分片数据的时间距与当前时间的差值是否大于预设超时阈值,包括:每隔预设周期检测最后一次接收到语音分片数据的时间距与当前时间的差值是否大于预设超时阈值。

9.根据权利要求1所述的服务器,其特征在于,所述服务器还被配置为:

在接收到所述语音分片数据后,将所述语音分片数据存储到所述语音会话对应的缓存文件中。

10.一种语音识别方法,其特征在于,包括:

显示设备将语音分片数据发送给服务器;

服务器接收来自显示设备的语音分片数据;

若所述语音分片数据不是语音会话的最后一片数据,所述服务器根据未向语音识别服

务设备发送所述语音分片数据的上一片数据,暂不发送所述语音分片数据,直到所述服务器已发送所述上一片数据,或所述上一片数据被标记为忽略状态,再将所述语音分片数据发送给所述语音识别服务设备,其中,对于未接收到的语音分片数据,若接收到排序在后的语音分片数据的数量达到预设数量阈值,则将未接收到的语音分片数据标记为忽略状态,所述忽略状态表示不再将所述语音分片数据发送给所述语音识别服务设备;

若所述语音分片数据是所述最后一片数据,所述服务器在所述最后一片数据之前的非忽略状态的语音分片数据均已发送,或等待的时间超过预设时间阈值时,再将已接收还未发送、且未被标记为忽略状态的语音分片数据发送给所述语音识别服务设备;

所述语音识别服务设备根据接收到的语音分片数据进行实时语音识别。

服务器及语音识别方法

技术领域

[0001] 本申请涉及语音交互技术领域,尤其涉及一种服务器及语音识别方法。

背景技术

[0002] 随着人工智能在显示设备领域的飞速发展,越来越多的显示设备如智能电视开始支持语音控制功能,用户可向电视输入一个语音会话,电视可通过语义引擎对语音会话进行语义识别,从而得到用户意图,进而针对用户意图进行响应。在一些场景下,用户输入的语音会话为一段较长的语音,如果电视在语音接收完毕后,再通过语义引擎对语音会话进行语义识别,最终可能需要花费较长的时间才能进行响应。相关技术中,为了提高语音识别速度,电视可在用户输入语音会话的过程中,就实时以分片的形式将用户语音发送给语义引擎进行语音识别,从而可提高语音交互的响应速度。然而,由于网络波动等原因,语义引擎接收到的语音分片数据的顺序可能并不是电视发送的顺序,这将导致语音识别的准确性降低,降低了用户体验。

发明内容

[0003] 为解决语音交互的准确性低的技术问题,本申请提供了一种服务器及语音识别方法。

[0004] 第一方面,本申请提供了一种服务器,该服务器被配置为:

[0005] 接收来自显示设备的语音分片数据;

[0006] 若所述语音分片数据不是语音会话的最后一片数据,根据服务器未向语音识别服务设备发送所述语音分片数据的上一片数据,暂不发送所述语音分片数据,直到所述服务器已发送所述上一片数据,或所述上一片数据被标记为忽略状态,再将所述语音分片数据发送给所述语音识别服务设备;

[0007] 若所述语音分片数据是所述最后一片数据,在所述最后一片数据之前的非忽略状态的语音分片数据均已发送,或等待的时间超过预设时间阈值时,再将已接收还未发送、且未被标记为忽略状态的语音分片数据发送给语音识别服务设备。

[0008] 在一些实施例中,对于未接收到的语音分片数据,若接收到排序在后的语音分片数据的数量达到预设数量阈值,将所述未接收到的语音分片数据标记为忽略状态。

[0009] 在一些实施例中,所述服务器还被配置为:

[0010] 对于未接收到且未被标记为忽略状态的语音分片数据,若距离所述最后一片数据的顺序越近,则所述语音分片数据对应的预设等待时间越大,所述预设时间阈值为全部预设等待时间中的最大值。

[0011] 在一些实施例中,所述语音分片数据包括语音数据和分片参数,所述分片参数包括分片序号,所述分片序号用于确定所述语音分片数据在语音会话中的顺序。

[0012] 在一些实施例中,所述分片序号包括数组,所述数组包括两个数值,所述数组的第一个数值表示所述语音分片数据在语音会话中的顺序,所述数组的第二个数值表示所述语

音分片数据的下一分片数据在所述语音会话中的顺序。

[0013] 第二方面,本申请提供了一种语音识别方法,该方法包括:

[0014] 显示设备将语音分片数据发送给服务器;

[0015] 服务器接收来自显示设备的语音分片数据;

[0016] 若所述语音分片数据不是语音会话的最后一片数据,所述服务器根据未向语音识别服务设备发送所述语音分片数据的上一片数据,暂不发送所述语音分片数据,直到所述服务器已发送所述上一片数据,或所述上一片数据被标记为忽略状态,再将所述语音分片数据发送给所述语音识别服务设备;

[0017] 若所述语音分片数据是所述最后一片数据,所述服务器在所述最后一片数据之前的非忽略状态的语音分片数据均已发送,或等待的时间超过预设时间阈值时,再将已接收还未发送、且未被标记为忽略状态的语音分片数据发送给语音识别服务设备;

[0018] 所述语音识别服务设备根据接收到的语音分片数据进行实时语音识别本申请提供的服务器及语音识别方法的有益效果包括:

[0019] 本申请实施例通过在接收到语音分片数据之后,根据上一片语音分片数据还没发送到语音识别服务设备,暂不发送该语音分片数据,直到服务器已发送所述上一片数据,或所述上一片数据被标记为忽略状态,再将所述语音分片数据发送给所述语音识别服务设备,从而保障了语音识别服务设备接收到的语音分片数据的时序,有利于提高语音识别服务设备的语音识别的准确性,进而有利于提高语音交互的响应准确性,提升语音交互体验。

附图说明

[0020] 为了更清楚地说明本申请的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,对于本领域普通技术人员而言,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0021] 图1中示例性示出了根据一些实施例的显示设备与控制装置之间操作场景的示意图;

[0022] 图2中示例性示出了根据一些实施例的控制装置100的硬件配置框图;

[0023] 图3中示例性示出了根据一些实施例的显示设备200的硬件配置框图;

[0024] 图4中示例性示出了根据一些实施例的显示设备200中软件配置示意图;

[0025] 图5中示例性示出了根据一些实施例的语音识别网络架构示意图;

[0026] 图6中示例性示出了根据一些实施例的缓存文件的示意图;

[0027] 图7中示例性示出了根据一些实施例的缓存文件的示意图;

[0028] 图8中示例性示出了根据一些实施例的缓存文件的示意图;

[0029] 图9中示例性示出了根据一些实施例的缓存文件的示意图;

[0030] 图10中示例性示出了根据一些实施例的缓存文件的示意图;

[0031] 图11中示例性示出了根据一些实施例的缓存文件的示意图;

[0032] 图12中示例性示出了根据一些实施例的缓存文件的示意图;

[0033] 图13中示例性示出了根据一些实施例的缓存文件的示意图。

具体实施方式

[0034] 为使本申请的目的和实施方式更加清楚,下面将结合本申请示例性实施例中的附图,对本申请示例性实施方式进行清楚、完整地描述,显然,描述的示例性实施例仅是本申请一部分实施例,而不是全部的实施例。

[0035] 需要说明的是,本申请中对于术语的简要说明,仅是为了方便理解接下来描述的实施方式,而不是意图限定本申请的实施方式。除非另有说明,这些术语应当按照其普通和通常的含义理解。

[0036] 本申请中说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”等是用于区别类似或同类的对象或实体,而不必然意味着限定特定的顺序或先后次序,除非另外注明。应该理解这样使用的用语在适当情况下可以互换。

[0037] 术语“包括”和“具有”以及他们的任何变形,意图在于覆盖但不排他的包含,例如,包含了一系列组件的产品或设备不必限于清楚地列出的所有组件,而是可包括没有清楚地列出的或对于这些产品或设备固有的其它组件。

[0038] 术语“模块”是指任何已知或后来开发的硬件、软件、固件、人工智能、模糊逻辑或硬件或/和软件代码的组合,能够执行与该元件相关的功能。

[0039] 图1为根据实施例中显示设备与控制装置之间操作场景的示意图。如图1所示,用户可通过智能设备300或控制装置100操作显示设备200。

[0040] 在一些实施例中,控制装置100可以是遥控器,遥控器和显示设备的通信包括红外协议通信或蓝牙协议通信,及其他短距离通信方式,通过无线或有线方式来控制显示设备200。用户可以通过遥控器上按键、语音输入、控制面板输入等输入用户指令,来控制显示设备200。

[0041] 在一些实施例中,也可以使用智能设备300(如移动终端、平板电脑、计算机、笔记本电脑等)以控制显示设备200。例如,使用在智能设备上运行的应用程序控制显示设备200。

[0042] 在一些实施例中,显示设备200还可以采用除了控制装置100和智能设备300之外的方式进行控制,例如,可以通过显示设备200设备内部配置的获取语音指令的模块直接接收用户的语音指令控制,也可以通过显示设备200设备外部设置的语音控制设备来接收用户的语音指令控制。

[0043] 在一些实施例中,显示设备200还与服务器400进行数据通信。可允许显示设备200通过局域网(LAN)、无线局域网(WLAN)和其他网络进行通信连接。服务器400可以向显示设备200提供各种内容和互动。服务器400可以是一个集群,也可以是多个集群,可以包括一类或多类服务器。

[0044] 图2示例性示出了根据示例性实施例中控制装置100的配置框图。如图2所示,控制装置100包括控制器110、通信接口130、用户输入/输出接口140、存储器、供电电源。控制装置100可接收用户的输入操作指令,且将操作指令转换为显示设备200可识别和响应的指令,起用用户与显示设备200之间交互中介作用。

[0045] 图3示出了根据示例性实施例中显示设备200的硬件配置框图。

[0046] 在一些实施例中,显示设备200包括调谐解调器210、通信器220、检测器230、外部装置接口240、控制器250、显示器260、音频输出接口270、存储器、供电电源、用户接口中的

至少一种。

[0047] 在一些实施例中控制器包括处理器,视频处理器,音频处理器,图形处理器,RAM,ROM,用于输入/输出的第一接口至第n接口。

[0048] 在一些实施例中,显示器260包括用于呈现画面的显示屏组件,以及驱动图像显示的驱动组件,用于接收源自控制器输出的图像信号,进行显示视频内容、图像内容以及菜单操控界面的组件以及用户操控UI界面。

[0049] 在一些实施例中,显示器260可为液晶显示器、OLED显示器、以及投影显示器,还可以为一种投影装置和投影屏幕。

[0050] 在一些实施例中,通信器220是用于根据各种通信协议类型与外部设备或服务器进行通信的组件。例如:通信器可以包括Wifi模块,蓝牙模块,有线以太网模块等其他网络通信协议芯片或近场通信协议芯片,以及红外接收器中的至少一种。显示设备200可以通过通信器220与外部控制设备100或服务器400建立控制信号和数据信号的发送和接收。

[0051] 在一些实施例中,用户接口,可用于接收控制装置100(如:红外遥控器等)的控制信号。

[0052] 在一些实施例中,检测器230用于采集外部环境或与外部交互的信号。例如,检测器230包括光接收器,用于采集环境光线强度的传感器;或者,检测器230包括图像采集器,如摄像头,可以用于采集外部环境场景、用户的属性或用户交互手势,再或者,检测器230包括声音采集器,如麦克风等,用于接收外部声音。

[0053] 在一些实施例中,外部装置接口240可以包括但不限于如下:高清多媒体接口接口(HDMI)、模拟或数据高清分量输入接口(分量)、复合视频输入接口(CVBS)、USB输入接口(USB)、RGB端口等任一个或多个接口。也可以是上述多个接口形成的复合性的输入/输出接口。

[0054] 在一些实施例中,调谐解调器210通过有线或无线接收方式接收广播电视信号,以及从多个无线或有线广播电视信号中解调出音视频信号,如以及EPG数据信号。

[0055] 在一些实施例中,控制器250和调谐解调器210可以位于不同的分体设备中,即调谐解调器210也可在控制器250所在的主体设备的外置设备中,如外置机顶盒等。

[0056] 在一些实施例中,控制器250,通过存储在存储器上中各种软件控制程序,来控制显示设备的工作和响应用户的操作。控制器250控制显示设备200的整体操作。例如:响应于接收到用于选择在显示器260上显示UI对象的用户命令,控制器250便可以执行与由用户命令选择的对象有关的操作。

[0057] 在一些实施例中,所述对象可以是可选对象中的任何一个,例如超链接、图标或其他可操作的控件。与所选择的对象有关操作有:显示连接到超链接页面、文档、图像等操作,或者执行与所述图标相对应程序的操作。

[0058] 在一些实施例中控制器包括中央处理器(Central Processing Unit,CPU),视频处理器,音频处理器,图形处理器(Graphics Processing Unit,GPU),RAM Random Access Memory,RAM),ROM(Read-Only Memory,ROM),用于输入/输出的第一接口至第n接口,通信总线(Bus)等中的至少一种。

[0059] CPU处理器。用于执行存储在存储器中操作系统和应用程序指令,以及根据接收外部输入的各种交互指令,来执行各种应用程序、数据和内容,以便最终显示和播放各种音视

频内容。CPU处理器,可以包括多个处理器。如,包括一个主处理器以及一个或多个子处理器。

[0060] 在一些实施例中,图形处理器,用于产生各种图形对象,如:图标、操作菜单、以及用户输入指令显示图形等。图形处理器包括运算器,通过接收用户输入各种交互指令进行运算,根据显示属性显示各种对象;还包括渲染器,对基于运算器得到的各种对象,进行渲染,上述渲染后的对象用于显示在显示器上。

[0061] 在一些实施例中,视频处理器,用于将接收外部视频信号,根据输入信号的标准编解码协议,进行解压缩、解码、缩放、降噪、帧率转换、分辨率转换、图像合成等视频处理,可得到直接可显示设备200上显示或播放的信号。

[0062] 在一些实施例中,视频处理器,包括解复用模块、视频解码模块、图像合成模块、帧率转换模块、显示格式化模块等。其中,解复用模块,用于对输入音视频数据流进行解复用处理。视频解码模块,用于对解复用后的视频信号进行处理,包括解码和缩放处理等。图像合成模块,如图像合成器,其用于将图形生成器根据用户输入或自身生成的GUI信号,与缩放处理后视频图像进行叠加混合处理,以生成可供显示的图像信号。帧率转换模块,用于对转换输入视频帧率。显示格式化模块,用于将接收帧率转换后视频输出信号,改变信号以符合显示格式的信号,如输出RGB数据信号。

[0063] 在一些实施例中,音频处理器,用于接收外部的音频信号,根据输入信号的标准编解码协议,进行解压缩和解码,以及降噪、数模转换、和放大处理等处理,得到可以在扬声器中播放的声音信号。

[0064] 在一些实施例中,用户可在显示器260上显示的图形用户界面(GUI)输入用户命令,则用户输入接口通过图形用户界面(GUI)接收用户输入命令。或者,用户可通过输入特定的声音或手势进行输入用户命令,则用户输入接口通过传感器识别出声音或手势,来接收用户输入命令。

[0065] 在一些实施例中,“用户界面”,是应用程序或操作系统与用户之间进行交互和信息交换的介质接口,它实现信息的内部形式与用户可以接受形式之间的转换。用户界面常用的表现形式是图形用户界面(Graphic User Interface,GUI),是指采用图形方式显示的与计算机操作相关的用户界面。它可以是在电子设备的显示屏中显示的一个图标、窗口、控件等界面元素,其中控件可以包括图标、按钮、菜单、选项卡、文本框、对话框、状态栏、导航栏、Widget等可视的界面元素。

[0066] 在一些实施例中,显示设备的系统可以包括内核(Kernel)、命令解析器(shell)、文件系统和应用程序。内核、shell和文件系统一起组成了基本的操作系统结构,它们让用户可以管理文件、运行程序并使用系统。上电后,内核启动,激活内核空间,抽象硬件、初始化硬件参数等,运行并维护虚拟内存、调度器、信号及进程间通信(IPC)。内核启动后,再加载Shell和用户应用程序。应用程序在启动后被编译成机器码,形成一个进程。

[0067] 显示设备的系统可以包括内核(Kernel)、命令解析器(shell)、文件系统和应用程序。内核、shell和文件系统一起组成了基本的操作系统结构,它们让用户可以管理文件、运行程序并使用系统。上电后,内核启动,激活内核空间,抽象硬件、初始化硬件参数等,运行并维护虚拟内存、调度器、信号及进程间通信(IPC)。内核启动后,再加载Shell和用户应用程序。应用程序在启动后被编译成机器码,形成一个进程。

[0068] 如图4所示,将显示设备的系统分为三层,从上至下分别为应用层、中间件层和硬件层。

[0069] 应用层主要包含电视上的常用应用,以及应用框架(Application Framework),其中,常用应用主要是基于浏览器Browser开发的应用,例如:HTML5 APPs;以及原生应用(NativeAPPs);

[0070] 应用框架(Application Framework)是一个完整的程序模型,具备标准应用软件所需的一切基本功能,例如:文件存取、资料交换...,以及这些功能的使用接口(工具栏、状态列、菜单、对话框)。

[0071] 原生应用(Native APPs)可以支持在线或离线,消息推送或本地资源访问。

[0072] 中间件层包括各种电视协议、多媒体协议以及系统组件等中间件。中间件可以使用系统软件所提供的基础服务(功能),衔接网络上应用系统的各个部分或不同的应用,能够达到资源共享、功能共享的目的。

[0073] 硬件层主要包括HAL接口、硬件以及驱动,其中,HAL接口为所有电视芯片对接的统一接口,具体逻辑由各个芯片来实现。驱动主要包含:音频驱动、显示驱动、蓝牙驱动、摄像头驱动、WIFI驱动、USB驱动、HDMI驱动、传感器驱动(如指纹传感器,温度传感器,压力传感器等)、以及电源驱动等。

[0074] 在一些实施例中的硬件或软件架构可以基于上述实施例中的介绍,在一些实施例中可以是基于相近的其他硬件或软件架构,可以实现本申请的技术方案即可。

[0075] 为清楚说明本申请的实施例,下面结合图5对本申请实施例提供的一种语音识别网络架构进行描述。

[0076] 参见图5,图5为本申请实施例提供的一种语音识别网络架构示意图。图5中,智能设备用于接收输入的信息以及输出对该信息的处理结果。语音识别服务设备为部署有语音识别服务的电子设备,语义服务设备为部署有语义服务的电子设备,业务服务设备为部署有业务服务的电子设备。这里的电子设备可包括服务器、计算机等,这里的语音识别服务、语义服务(也可称为语义引擎)和业务服务为可部署在电子设备上的web服务,其中,语音识别服务用于将音频识别为文本,语义服务用于对文本进行语义解析,业务服务用于提供具体的服务如墨迹天气的天气查询服务、QQ音乐的音乐查询服务等。在一个实施例中,图5所示架构中可存在部署有不同业务服务的多个实体服务设备,也可以一个或多个实体服务设备中集合一项或多项功能服务。

[0077] 一些实施例中,下面对基于图5所示架构处理输入智能设备的信息的过程进行举例描述,以输入智能设备的信息为通过语音输入的查询语句为例,上述过程可包括如下三个过程:

[0078] [语音识别]

[0079] 智能设备可在接收到通过语音输入的查询语句后,将该查询语句的音频上传至语音识别服务设备,以由语音识别服务设备通过语音识别服务将该音频识别为文本后返回至智能设备。在一个实施例中,将查询语句的音频上传至语音识别服务设备前,智能设备可对查询语句的音频进行去噪处理,这里的去噪处理可包括去除回声和环境噪声等步骤。

[0080] [语义理解]

[0081] 智能设备将语音识别服务识别出的查询语句的文本上传至语义服务设备,以由语

义服务设备通过语义服务对该文本进行语义解析,得到文本的业务领域、意图等。

[0082] [语义响应]

[0083] 语义服务设备根据对查询语句的文本的语义解析结果,向相应的业务服务设备下发查询指令以获取业务服务给出的查询结果。智能设备可从语义服务设备获取该查询结果并输出。作为一个实施例,语义服务设备还可将对查询语句的语义解析结果发送至智能设备,以由智能设备输出该语义解析结果中的反馈语句。

[0084] 需要说明的是,图5所示架构只是一种示例,并非对本申请保护范围的限定。本申请实施例中,也可采用其他架构来实现类似功能,例如:三个过程全部或部分可以由智能终端来完成,在此不做赘述。

[0085] 在一些实施例中,图5所示的智能设备可为显示设备,如智能电视,显示设备可先将采集的用户语音发送到显示设备的服务器,显示设备的服务器再将该用户语音发送到语音识别服务设备进行语音识别。

[0086] 在一些实施例中,图5所示的智能设备也可为其他支持语音交互的设备,如智能音箱、智能手机等等。

[0087] 在一些实施例中,用户单次向显示设备输入的一个查询语句或其他交互语句,可称为一个语音会话。语音交互的场景可细分为问答场景和聊天场景,在问答场景下,用户向显示设备输入一个语音会话,显示设备在做出响应后退出语音交互界面,而在聊天场景下,用户向显示设备输入一个语音会话,显示设备在做出响应后不退出语音交互界面,而是持续采集声音,用户可向显示设备输入一个新的语音会话,显示设备可对新的语音会话进行响应,直到到达终止条件再退出语音交互界面,示例性的,终止条件可为:用户在规定的时间内没有继续输入新的语音会话。

[0088] 在一些实施例中,用户输入的语音会话可能为一段较长时间的语音,显示设备如果在用户说完之后再向服务器发送该语音,需要花费较长时间才能发送完毕,这将导致语音交互的响应时间较长。为缩短语音交互的响应时间,在用户输入语音会话的过程中,显示设备可每隔一段时间就向服务器上传这段时间内接收到的语音数据,这样,一个语音会话就被分成了多片数据,服务器每接收到一片数据,就将该片数据发送给语音识别服务设备,在用户将语音会话输入完毕时,显示设备已经将该语音会话的大部分数据通过服务器发送给了语音识别服务设备,显示设备只需要较短的时间就可将该语音会话剩余的数据通过服务器发送给了语音识别服务设备,从而可缩短语音交互的响应时间。

[0089] 然而,由于网络波动等原因,显示设备上传给服务器的语音分片数据可能不是连续的数据,显示设备先发送的语音分片数据可能后到达服务器,这将导致服务器发送给语音识别服务设备的语音分片数据也不是连续的数据,进而导致语音识别服务设备对语音会话的识别准确性下降,使得语音交互的体验不佳。

[0090] 为解决上述技术问题,在一些实施例中,服务器在接收到一片语音分片数据后,若该语音分片数据不是语音会话的第一片数据,且没接收到上一片数据,先不发送该片数据,等待一段时间后,若接收到上一片数据,则将上一片数据和该语音分片数据发送给服务器,若一直没接收到上一片数据,也可将该语音分片数据发送给服务器,使服务器尽可能接收到较为连续的语音分片数据,这样,服务器向语义服务设备发送的也是较为连续的语音分片数据,从而可提高语音识别准确性,提高语音交互体验。

[0091] 下面以用户向显示设备发出了一次语音会话为例,详细介绍上述提高语音识别准确性的技术方案。

[0092] 在一些实施例中,上述提高语音识别准确性的方法可分为两个阶段,第一个阶段发生在显示设备端,包括语音会话的接收以及处理,第二个阶段发生在服务器端,包括语音会话的处理以及发送。

[0093] 示例性的,第一阶段的过程如下:

[0094] 在一些实施例中,显示设备的遥控器上可设置有语音控制按键,用户按住遥控器上的语音控制按键后,显示设备的控制器可控制显示设备的显示器显示语音交互界面,并控制声音采集器,如麦克风,采集显示设备周围的声音。此时,用户可向显示设备输入语音会话。在用户输入语音会话的过程中,显示设备可将该语音会话以分片形式发送给服务器。

[0095] 在一些实施例中,显示设备可支持语音唤醒功能,显示设备的声音采集器可处于持续采集声音的状态。用户说出唤醒词后,显示设备对用户输入的语音会话进行语音识别,识别出语音会话为唤醒词后,可控制显示设备的显示器显示语音交互界面,对于该内容为唤醒词的语音会话,显示设备可不将该语音会话发送给服务器,在用户向显示设备输入新的语音会话后,显示设备再将该新的语音会话发送给服务器。在用户输入该新的语音会话的过程中,显示设备可将该新的语音会话以分片形式发送给服务器。

[0096] 在一些实施例中,在用户输入一个语音会话后,在显示设备获取该语音会话的响应数据或显示设备根据响应数据进行响应的过程中,显示设备的声音采集器可保持声音采集的状态,用户可随时按住遥控器上的语音控制按键重新输入语音会话,或者说出唤醒词,此时,显示设备可结束上一次的语音交互进程,根据用户新输入的语音会话,开启新的语音交互进程,从而保障语音交互的实时性。

[0097] 以用户通过遥控器进入语音交互界面为例,在一些实施例中,用户可在按住遥控器上的语音控制按键后进行讲话,讲话完毕后松开语音控制按键。显示设备响应于语音控制按键被按下,控制声音采集器采集声音,将采集到的声音进行存储,间隔一段时间,如间隔300ms生成一片语音分片数据,语音分片数据包括语音数据和分片参数,其中,分片参数可包括sessionid(会话标识)和分片序号。会话标识用于区分不同的语音会话,不同的语音会话设置有不同的会话标识,显示设备根据接收到语音控制按键被按下的信号生成一个新的会话标识。分片序号可为一个数组,所述数组包括两个数值,第一个数值表示所述语音分片数据在语音会话中的顺序,所述数组的第二个数值表示所述语音分片数据的下一片数据在语音会话中的顺序。例如,分片序号为1-2,表明当前语音分片数据为第一片数据,当前语音会话还有下一片数据,分片序号为2-2,表明当前语音分片数据为第二片数据,当前语音会话没有下一片数据,即当前语音分片数据为当前语音会话的最后一片数据。在用户松开语音控制按键时,显示设备可根据语音控制按键复位的信号,确定当前语音会话已结束,将最后一片语音分片数据的分片序号设置为前后两个数值相同。

[0098] 在一些实施例中,分片序号也可只包括一个数值,该数值表示语音分片数据在语音会话中的顺序。例如,分片序号为1,表明当前语音分片数据为第一片数据,分片序号为2,表明当前语音分片数据为第二片数据。在用户松开语音控制按键时,显示设备可根据语音控制按键复位的信号,确定当前语音会话已结束,生成一个结束标识,将该结束标识写入到分片参数中,如果一个语音分片数据的分片参数中有结束标识,则表明该语音分片数据为

当前语音会话的最后一片数据,如果一个语音分片数据的分片参数中没有结束标识,则表明该语音分片数据不是当前语音会话的最后一片数据。

[0099] 以分片序号为一个数组为例,显示设备每隔一段时间截取这段时间内的语音数据,然后生成分片参数,根据语音数据和分片参数生成语音分片数据,发送给服务器,以使服务器进行第二阶段的处理。

[0100] 示例性的,第二阶段的过程如下:

[0101] 在一些实施例中,服务器在接收到语音分片数据后,可从语音分片数据中提取出会话标识,然后检测是否已经存在该会话标识对应的缓存文件,如果不存在该会话标识对应的缓存文件,则表明服务器第一次接收到该语音会话的语音分片数据。服务器可新建一个该会话标识对应的缓存文件,然后设置该会话标识对应的会话状态参数,会话状态参数包括waitCounter(等待数量)和lastUpdateTime(上次更新时间),其中,waitCounter表示服务器已经接收到但还没有发送给语音识别服务设备、且未被标记为忽略状态的语音分片数据数量,lastUpdateTime表示服务器上次接收到未被标记为忽略状态的语音分片数据的时间,示例性的,lastUpdateTime可为一个时间戳。服务器在每接收一片语音分片数据,可更新一次会话状态参数。服务器将语音分片数据中的语音数据和会话状态参数存储在该缓存文件内,其中,语音数据在存储时,可标注上分片序号和状态,状态可为未收到、待发送、已发送、忽略,其中,“忽略”表示不再将该语音数据发送给语音识别服务设备。

[0102] 根据语音分片数据的分片序号,服务器单次接收到的语音分片数据可能为第一片数据,可能为中间片数据,可能为最后一片数据。从第一次接收到一个语音会话的语音分片数据开始,服务器可根据每次接收到的语音分片数据中的分片序号进行不同的处理。

[0103] 下面先以服务器第一次接收到语音分片数据和第二次接收到语音分片数据为例介绍服务器对语音分片数据的处理。

[0104] 在一些实施例中,在网络畅通的情况下,对于一个会话标识对应的语音会话,服务器第一次接收到的语音分片数据通常为该语音会话的第一片数据,服务器在新建一个缓存文件后,可向语音识别服务设备发送语音识别请求,该语音识别请求包括了与语音识别服务设备开启语音会话的指令以及第一片数据中的语音数据。语音识别服务设备在接收到语音识别请求后,根据开启语音会话的指令开启一个语音识别进程,通过该进程对服务器发送的语音数据进行实时语音识别。

[0105] 在一些实施例中,由于网络波动等原因,对于一个会话标识对应的语音会话,服务器第一次接收到的语音分片数据可能不是该语音会话的第一片数据,而是第二片数据或第三片数据等等,这种情况下,服务器在新建一个缓存文件后,可暂时不向语音识别服务设备发送语音识别请求,而是继续等待接收来自显示设备的语音分片数据。

[0106] 在一些实施例中,对于一个会话标识对应的语音会话,服务器第一次接收到的语音分片数据为该语音会话的第一片数据,第二次接收到的语音分片数据为该语音会话的第二片数据,则服务器已经向语音识别服务设备发送了第一片数据。此时,服务器可将该第二片数据中的语音数据发送给语音识别服务设备,并更新缓存文件;如果该语音分片数据不是该语音会话的第二片数据,而是第三片数据或第四片数据等等,这种情况下,服务器在可暂时不向语音识别服务设备发送该第二次接收到的语音分片数据中的语音数据,然后更新缓存文件。

[0107] 在一些实施例中,对于一个会话标识对应的语音会话,服务器第一次接收到的语音分片数据不是该语音会话的第一片数据,第二次接收到的语音分片数据为该语音会话的第一片数据,则服务器还没有向语音识别服务设备发送语音识别请求。此时,若第一次接收到的语音分片数据恰好是第二片数据,服务器可向语音识别服务设备发送语音识别请求,并更新缓存文件,该语音识别请求包括了与语音识别服务设备开启语音会话的指令以及第一片数据中的语音数据和第二片数据中的语音数据;若第一次接收到的语音分片数据不是第二片数据,服务器可向语音识别服务设备发送语音识别请求并更新缓存文件,该语音识别请求包括了与语音识别服务设备开启语音会话的指令以及第一片数据中的语音数据,第二片数据中的语音数据暂不向语音识别服务设备发送。

[0108] 在一些实施例中,对于一个会话标识对应的语音会话,服务器第一次接收到的语音分片数据不是该语音会话的第一片数据,第二次接收到的语音分片数据也不是该语音会话的第一片数据,则服务器还没有向语音识别服务设备发送语音识别请求。此时,服务器可不向语音识别服务设备发送语音识别请求,只更新缓存文件。

[0109] 下面以服务器第N次接收到语音分片数据,且该语音分片数据为中间片数据为例介绍服务器对语音分片数据的处理,其中,N大于等于2。

[0110] 在一些实施例中,服务器接收到语音分片数据后,根据语音分片数据中的分片序号确定当前语音分片数据是中间片数据后,判断当前语音分片数据的上一片数据是否已发送给语音识别服务设备,若上一片数据已经发送给语音识别服务设备,则将该语音分片数据发送给语音识别服务设备,并更新缓存文件,若服务器未向语音识别服务设备发送所述语音分片数据的上一片数据中的语音数据,暂不向语音识别服务设备发送所述语音分片数据中的语音数据,并更新缓存文件,直到服务器已向语音识别服务设备发送所述上一片数据中的语音数据,或所述上一片数据符合预设的忽略准则,再向所述语音识别服务设备发送所述语音分片数据中的语音数据,并更新缓存文件。例如,服务器已经接收到分片序号为3、4的语音分片数据,该分片序号为3、4的语音分片数据都是中间片数据,但服务器未接收到分片序号为2的语音分片数据,则暂不发送该分片序号为3、4的语音分片数据对应的语音数据,过了一段时间,服务器接收到了分片序号为2的语音分片数据,则可将分片序号为2、3、4的语音分片数据对应的语音数据一并发送给语音识别服务设备。

[0111] 在一些实施例中,服务器可将一些语音分片数据标记为忽略状态:对于未接收到的语音分片数据,若服务器接收到排序在后的语音分片数据的数量达到预设数量阈值,将所述未接收到的语音分片数据标记为忽略状态,示例性的,该预设数量阈值可为3。例如,服务器已经接收到分片序号为6、7、8的语音分片数据,该分片序号为6、7、8的语音分片数据都是中间片数据,但服务器还未接收到分片序号为5的语音分片数据,则将分片序号为5的语音分片数据标记为忽略状态,后续即使接收到了该分片序号为5的语音分片数据,也不将该语音分片数据发送给语音识别服务设备。

[0112] 在显示设备向服务器发送多片语音分片数据的过程中,如果服务器接收到了最后一片数据,无论最后一片数据之前的数据是否接收到,服务器都可判定用户已经完成了语音输入,需要尽快向语音识别服务设备获取语音识别结果,而在一些场景下,显示设备在向服务器发送语音分片数据时,由于网络波动等原因,最后一片数据在发送过程中丢失,服务器不能接收到最后一片数据,不能确定用户是否输入完毕,为解决该问题,在一些实施例

中,服务器可被配置为在第一次接收到该会话标识的语音分片数据后,就启动一个定时任务,每隔5s轮询该会话标识对应的缓存文件,获取lastUpdateTime,将其与当前时间进行比较。如果当前时间与lastUpdateTime的差值大于10秒,则认为请求超时,删除当前缓存文件,向语音识别服务设备发送结束语音会话的指令,若后续再接收到该会话标识的语音分片数据,也不再向语音识别服务设备发送。如果当前时间与lastUpdateTime的差值小于10秒,则认为请求没有超时,不删除当前缓存文件,继续等待接收下一片语音分片数据。当然,服务器也可不启动定时任务,还是实时检测lastUpdateTime。

[0113] 下面以服务器第M次接收到语音分片数据,且该语音分片数据为最后一片数据为例介绍服务器对语音分片数据的处理,其中,M大于等于2。

[0114] 在一些实施例中,服务器在接收到一片语音分片数据后,更新缓存文件,根据该语音分片数据的分片序号确定该语音分片数据为最后一片数据,可根据缓存文件中各语音分片数据的状态,判断在该最后一片数据之前的语音分片数据是否均已向语音识别服务设备发送完毕,其中,已忽略的语音分片数据除外。若除了已忽略的语音分片数据,该最后一片数据之前的语音分片数据均已发送完毕,则可将该最后一片数据中的语音数据也发送给语音识别服务设备,并向语音识别服务设备发送结束语音会话的指令。若除了已忽略的语音分片数据,该最后一片数据之前的语音分片数据还有未发送的分片数据,则表明还有没接收到的语音分片数据,此时,服务器可按照预设的等待时间表计算maxWaitTime(最大等待时间),将该最大等待时间作为预设时间阈值,在最大等待时间内继续接收语音分片数据。

[0115] 示例性的,一种语音分片数据的等待时间表可参见表1。

[0116] 表1

[0117]

分片序号差	1	2	3	4	5	6
等待时间(ms)	1200	600	300	300	300	150
分片序号差	7	8	9	10	11	12
等待时间(ms)	150	150	150	0	0	0

[0118] 表1中,分片序号差表示未接收到的语音分片数据与最后一片数据之间的分片序号之间的差值,等到时间表示该分片序号差下,服务器在接收到最后一片数据后,等待接收该语音分片数据的时间阈值。

[0119] 根据表1,分片序号差为1,等待时间为1200ms,表示在接收到最后一片数据后,如果倒数第二片数据没接收到,则最长可等待1200ms,若1200ms内仍然没有接收到该倒数第二片数据,则不再继续等待,将未发送的且未被标记为忽略状态的语音分片数据发送给语音识别服务设备,并向语音识别服务设备发送结束语音会话的指令。如果在1200ms内接收到了该倒数第二片数据,例如,在第100ms内接收到该倒数第二片数据,则不再继续等待,将未发送的且未被标记为忽略状态的语音分片数据发送给语音识别服务设备,并向语音识别服务设备发送结束语音会话的指令。

[0120] 根据表1,如果在接收到最后一片数据后,倒数第二片数据和倒数第三片数据没接收到,则根据表1可得,倒数第二片数据对应的等待时间为1200ms,倒数第三片数据对应的等待时间为600ms,因此,最大等待时间为1200ms,服务器可等待接收语音分片数据,直到等待时间到达1200ms或已经接收到全部的除忽略状态的分片数据。

[0121] 在一些实施例中,由于服务器从接收到最后一片数据到计算出maxWaitTime耗费

了一些时间,而在这段时间,服务器实际也是处于等待状态,因此,为了得到更准确的服务器已等待的时间,服务器可在计算出maxWaitTime时,将CurrentWaitTime(当前等待时间)增加20ms,即设置CurrentWaitTime从20ms开始计时,当CurrentWaitTime达到maxWaitTime时,计时终止,或者当已经接收到全部的除忽略状态的分片数据时,计时终止。其中,上述20ms仅为示例性的,也可设置为其他时间,服务器也可设置CurrentWaitTime从0开始计时。

[0122] 上述实施例示出了服务器在接收到语音分片数据后的处理过程,为了对上述处理过程进行更直观地介绍,以用户输入的语音包括7片语音分片数据为例,图6-图13示例性介绍了一种缓存文件的示意图。

[0123] 参见图6,服务器第一次接收到的一个语音会话的语音分片数据为第一片数据,服务器可将该第一片数据发送给语音识别服务设备,并建立图6所示的缓存文件,在缓存文件中记录lastUpdateTime、waitCounter、分片序号、状态和语音数据,对于状态为已发送的语音分片数据,显示设备可将其删除,从而在缓存文件中不显示该语音分片数据的内容。

[0124] 参见图7,服务器第二次接收到的上述语音会话的语音分片数据为第三片数据,服务器由于还没发送第二片数据,可暂不将该第三片数据发送给语音识别服务设备,并更新图6所示的缓存文件,以得到图7所示的缓存文件。更新过程如下:在缓存文件中记录lastUpdateTime、waitCounter、分片序号、状态和语音数据。

[0125] 参见图8,服务器第三次接收到的上述语音会话的语音分片数据为第四片数据,服务器由于还没发送第三片数据,可暂不将该第四片数据发送给语音识别服务设备,并更新图7所示的缓存文件,以得到图8所示的缓存文件。更新过程如下:在缓存文件中记录lastUpdateTime、waitCounter、分片序号、状态和语音数据。

[0126] 参见图9,服务器第四次接收到的上述语音会话的语音分片数据为第六片数据,服务器由于还没发送第五片数据,可暂不将该第六片数据发送给语音识别服务设备,并更新图8所示的缓存文件,以得到图9所示的缓存文件。更新过程如下:在缓存文件中记录lastUpdateTime、waitCounter、分片序号、状态和语音数据。

[0127] 在得到图9所示的缓存文件后,服务器根据第二片数据虽然没收到,但是收到了第二片数据之后的三片数据,超过了预设数量阈值,将第二片数据标记为忽略状态,再根据第二片数据被标记为忽略状态,将第二片数据之后的第三片数据和第四片数据发送给语音识别服务设备,由于第五片数据没有被标记为忽略状态,因此,暂不发送第六片数据,然后将缓存文件进行更新,从而得到图10所示的缓存文件。

[0128] 参见图11,服务器第五次接收到的上述语音会话的语音分片数据为第五片数据,由于第四片数据已经发送,服务器可将第五片数据和第六片数据发送给语音识别服务设备,并更新图10所示的缓存文件,以得到图11所示的缓存文件。更新过程如下:在缓存文件中记录lastUpdateTime、waitCounter、分片序号、状态和语音数据。

[0129] 参见图12,服务器第六次接收到的上述语音会话的语音分片数据为第二片数据,由于第二片数据已经被标记为忽略状态,因此,服务器可直接将该第二片数据丢弃,不需更新缓存文件。

[0130] 参见图13,服务器第七次接收到的上述语音会话的语音分片数据为第七片数据,也就是最后一片数据,由于第六片数据已经发送,服务器可将七片数据发送给语音识别服务设备,并更新图12所示的缓存文件,以得到图13所示的缓存文件。更新过程如下:在缓存

文件中记录lastUpdateTime、waitCounter、分片序号、状态和语音数据。

[0131] 在一些实施例中,语音识别服务设备可根据接收到的语音分片数据进行实时识别,将识别出的结果实时反馈给服务器,服务器在接收到识别结果后,再对该识别结果进行解析及响应,最终生成响应结果,将响应结果反馈给显示设备,该过程可参见图5中的描述,此处不再赘述。

[0132] 由上述实施例可见,本申请实施例通过在接收到语音分片数据之后,根据上一片语音分片数据还没发送到语音识别服务设备,暂不发送该语音分片数据,直到服务器已发送所述上一片数据,或所述上一片数据被标记为忽略状态,再将所述语音分片数据发送给所述语音识别服务设备,从而保障了语音识别服务设备接收到的语音分片数据的时序,有利于提高语音识别服务设备的语音识别的准确性,进而有利于提高语音交互的响应准确性,提升语音交互体验。

[0133] 由于以上实施方式均是在其他方式之上引用结合进行说明,不同实施例之间均具有相同的部分,本说明书中各个实施例之间相同、相似的部分互相参见即可。在此不再详细阐述。

[0134] 需要说明的是,在本说明书中,诸如“第一”和“第二”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,而不一定要求或暗示这些实体或操作之间存在任何这种实际的关系或顺序。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电路结构、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电路结构、物品或者设备所固有的要素。在没有更多限制的情况下,有语句“包括一个……”限定的要素,并不排除在包括要素的电路结构、物品或者设备中还存在另外的相同要素。

[0135] 本领域技术人员在考虑说明书及实践这里发明的公开后,将容易想到本申请的其他实施方案。本申请旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本申请的真正范围和精神由权利要求的内容指出。

[0136] 以上的本申请实施方式并不构成对本申请保护范围的限定。

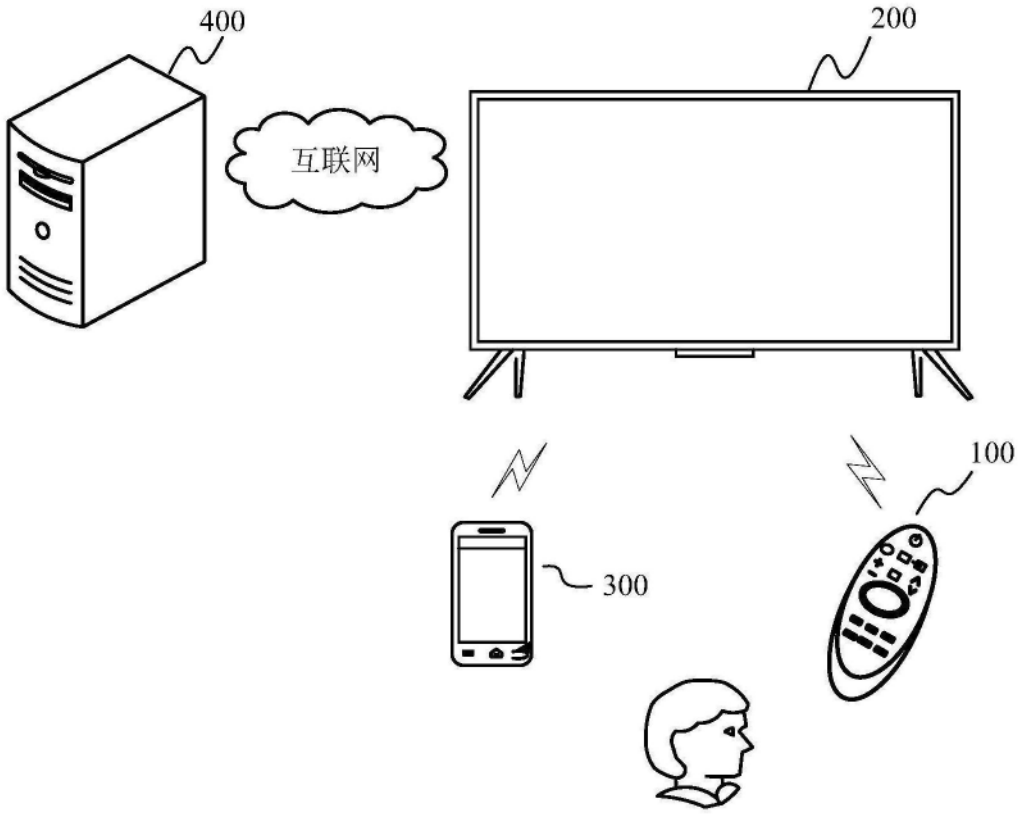

400
200
互联网
100
300

图1

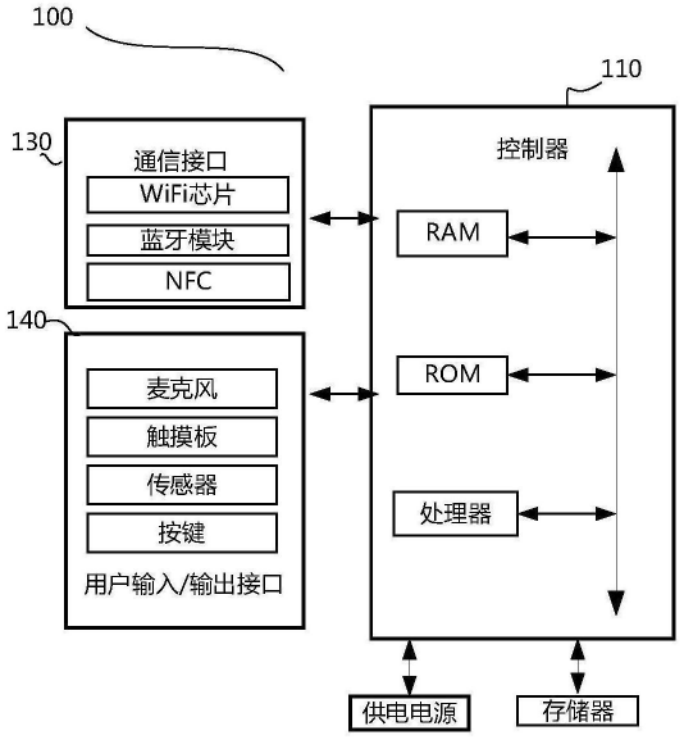

100
110
130
通信接口
WiFi芯片
蓝牙模块
NFC
控制器
RAM
140
麦克风
触摸板
传感器
按键
用户输入/输出接口
ROM
处理器
供电电源
存储器

图2

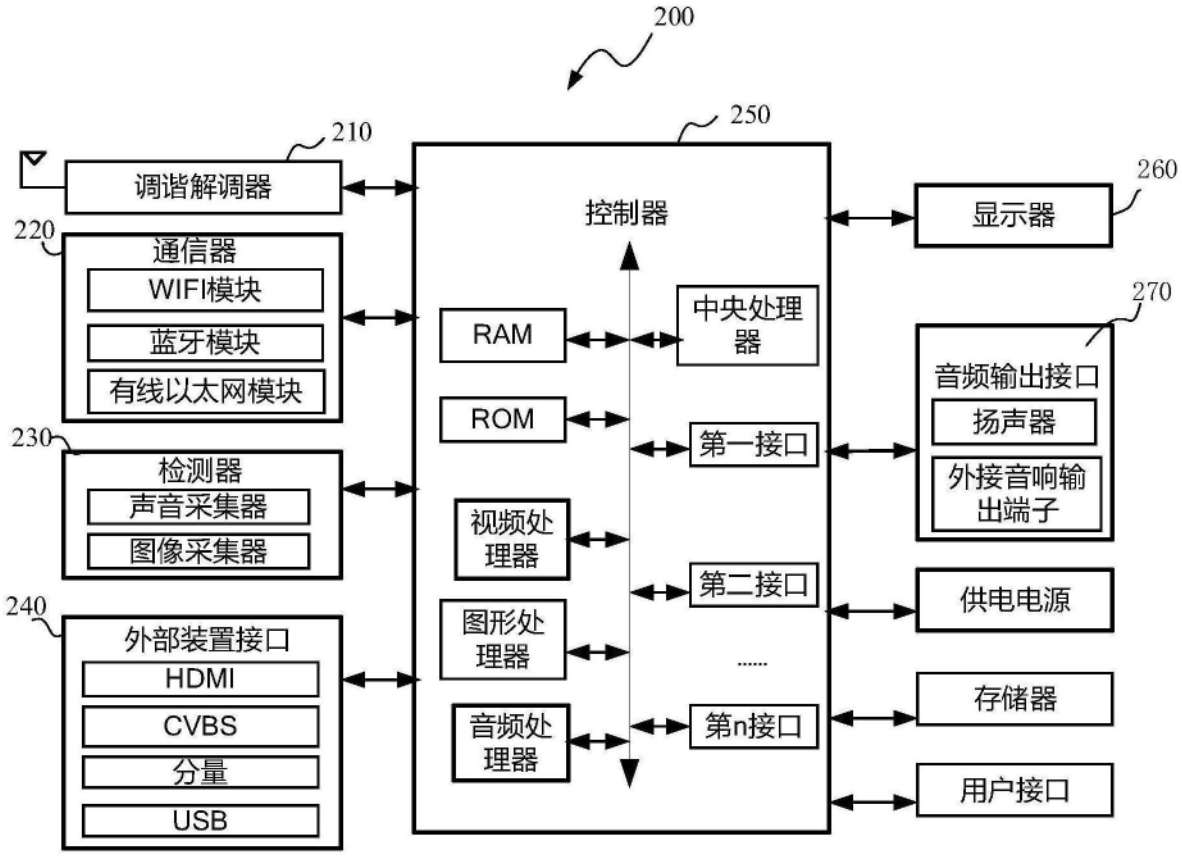

200
210
250
260
220
230
240
270
调谐解调器
控制器
显示器
通信器
WIFI模块
蓝牙模块
有线以太网模块
RAM
中央处理器
ROM
第一接口
音频输出接口
扬声器
外接音响输出端子
检测器
声音采集器
图像采集器
视频处理器
第二接口
供电电源
外部装置接口
HDMI
CVBS
分量
USB
图形处理器
.....
存储器
音频处理器
第n接口
用户接口

图3

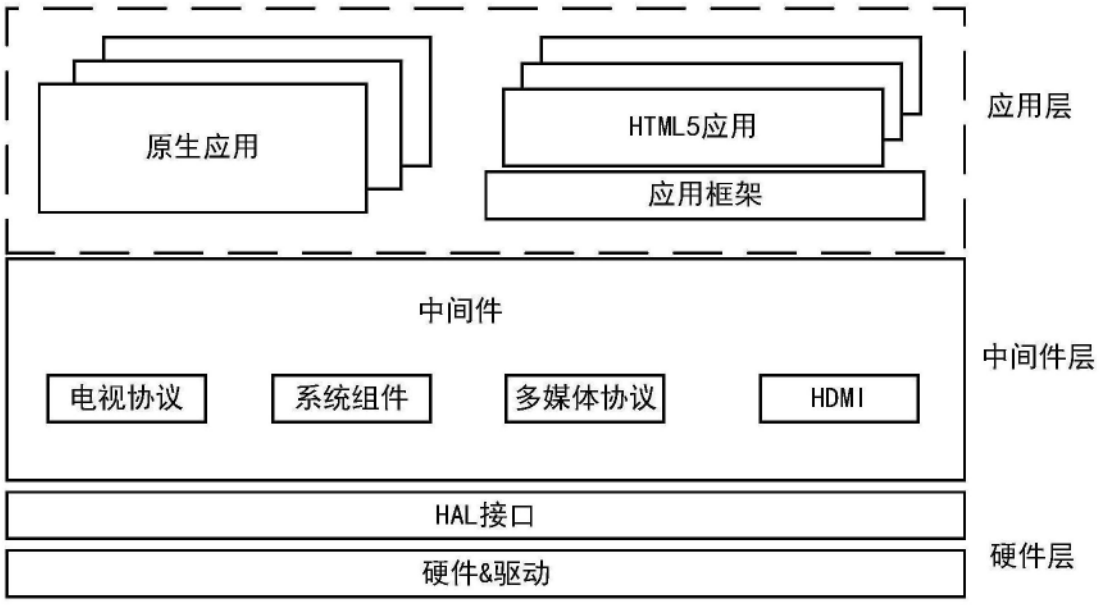

原生应用
HTML5应用
应用框架
应用层
中间件
电视协议
系统组件
多媒体协议
HDMI
中间件层
HAL接口
硬件&驱动
硬件层

图4

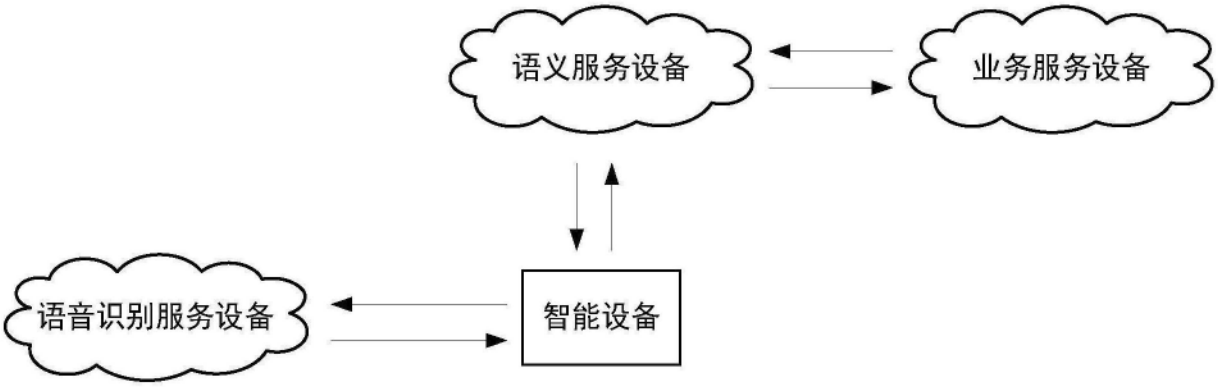

语义服务设备
业务服务设备
语音识别服务设备
智能设备

图5

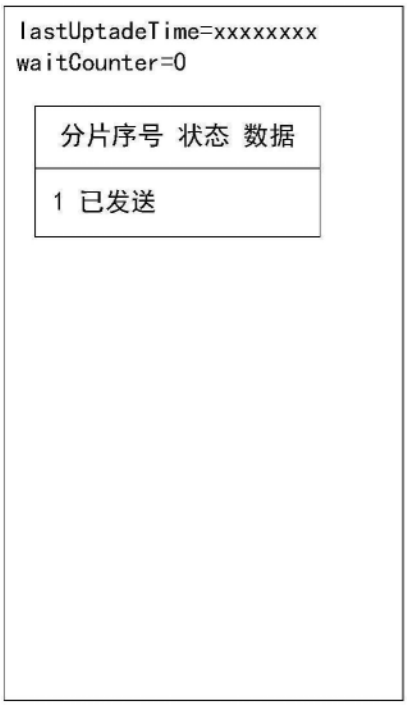

lastUptadeTime=xxxxxxxx
waitCounter=0
分片序号 状态 数据
1 已发送

图6

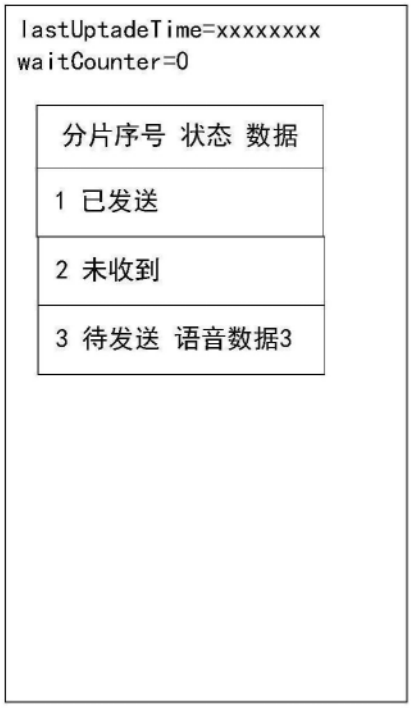

lastUptadeTime=xxxxxxxx
waitCounter=0
分片序号 状态 数据
1 已发送
2 未收到
3 待发送 语音数据3

图7


lastUptadeTime=xxxxxxxx
waitCounter=2
分片序号 状态 数据
1 已发送
2 已发送
3 待发送 语音数据3
4 待发送 语音数据4

图8


lastUptadeTime=xxxxxxxx
waitCounter=3
分片序号 状态 数据
1 已发送
2 未收到
3 待发送 语音数据3
4 待发送 语音数据4
5 未收到
6 待发送 语音数据6

图9


lastUptadeTime=xxxxxxxx
waitCounter=1
分片序号 状态 数据
1 已发送
2 忽略
3 已发送
4 已发送
5 未收到
6 待发送 语音数据6

图10


lastUptadeTime=xxxxxxxx
waitCounter=0
分片序号 状态 数据
1 已发送
2 忽略
3 已发送
4 已发送
5 已发送
6 已发送

图11


lastUptadeTime=xxxxxxxx
waitCounter=0
分片序号 状态 数据
1 已发送
2 忽略
3 已发送
4 已发送
5 已发送
6 已发送

图12


lastUptadeTime=xxxxxxxx
waitCounter=0
分片序号 状态 数据
1 已发送
2 忽略
3 已发送
4 已发送
5 已发送
6 已发送
7 已发送

图13