



# (12) 发明专利申请

(10) 申请公布号 CN 115457945 A

(43) 申请公布日 2022. 12. 09

(21) 申请号 202211402578.0

G06F 40/30 (2020.01)

(22) 申请日 2022.11.10

G06F 16/332 (2019.01)

G06F 16/33 (2019.01)

(71) 申请人 广州小鹏汽车科技有限公司

地址 510000 广东省广州市天河区岑村松岗大街8号

(72) 发明人 韩传宇 宋旭阳 赵耀 易晖  
翁志伟

(74) 专利代理机构 北京清亦华知识产权代理事  
务所(普通合伙) 11201

专利代理师 雷玉龙

(51) Int. Cl.

G10L 15/18 (2013.01)

G10L 15/06 (2013.01)

G10L 15/22 (2006.01)

G10L 15/26 (2006.01)

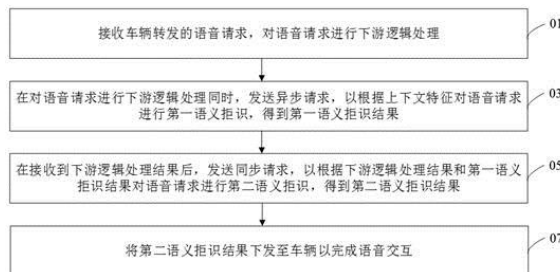
权利要求书2页 说明书13页 附图4页

## (54) 发明名称

语音交互方法、服务器和存储介质

## (57) 摘要

本发明公开了一种语音交互方法、服务器和存储介质。该语音交互方法包括：接收车辆转发的语音请求，对语音请求进行下游逻辑处理；在对语音请求进行下游逻辑处理同时，发送异步请求，以根据上下文特征对语音请求进行第一语义拒识，得到第一语义拒识结果；在接收到下游逻辑处理结果后，发送同步请求，以根据下游逻辑处理结果和第一语义拒识结果对语音请求进行第二语义拒识，得到第二语义拒识结果；将第二语义拒识结果下发至车辆以完成语音交互。本发明根据上下文特征对语音请求进行第一语义拒识并根据第一语义拒识结果和下游逻辑处理结果对语音请求进行第二语义拒识得到第二语义拒识结果，进而完成语音交互，能够降低语音交互过程的端到端时延。



1. 一种语音交互方法,其特征在于,包括:

接收车辆转发的语音请求,对所述语音请求进行下游逻辑处理;

在对所述语音请求进行所述下游逻辑处理同时,发送异步请求,以根据上下文特征对所述语音请求进行第一语义拒识,得到第一语义拒识结果;

在接收到下游逻辑处理结果后,发送同步请求,以根据所述下游逻辑处理结果和所述第一语义拒识结果对所述语音请求进行第二语义拒识,得到第二语义拒识结果;

将所述第二语义拒识结果下发至所述车辆以完成语音交互。

2. 根据权利要求1所述的语音交互方法,其特征在于,所述接收车辆转发的语音请求,对所述语音请求进行下游逻辑处理,包括:

将所述语音请求发送至中控服务以使所述中控服务将所述语音请求发送至对话系统下游服务进行入参处理,由所述对话系统下游服务实现所述下游逻辑处理。

3. 根据权利要求2所述的语音交互方法,其特征在于,所述将所述语音请求发送至中控服务以使所述中控服务将所述语音请求发送至对话系统下游服务进行入参处理,由所述对话系统下游服务实现所述下游逻辑处理,包括:

将所述语音请求发送至所述中控服务以使所述中控服务将所述语音请求发送至对话系统下游服务进行自然语言理解、对话管理和/或业务机器人入参,从而由所述对话系统下游服务实现自然语言理解、对话管理和业务机器人对应的业务逻辑处理。

4. 根据权利要求1所述的语音交互方法,其特征在于,所述接收车辆转发的语音请求,对所述语音请求进行下游逻辑处理的步骤之后,所述语音交互方法包括:

通过中控服务将所述语音请求的语音识别文本特征发送至上下文服务,以将所述语音识别文本特征作为所述上下文特征存入数据存储服务。

5. 根据权利要求4所述的语音交互方法,其特征在于,所述在对所述语音请求进行所述下游逻辑处理同时,发送异步请求,以根据上下文特征对所述语音请求进行第一语义拒识,得到第一语义拒识结果的步骤之前,所述语音交互方法包括:

在对所述语音请求进行下游逻辑处理同时,将所述语音请求发送至声学拒识服务进行处理,以得到声学特征;

根据所述声学特征对所述语音请求进行拒识得到声学拒识结果;

将所述声学特征和所述声学拒识结果发送至所述上下文服务,以将所述声学特征和所述声学拒识结果作为上下文特征存入所述数据存储服务。

6. 根据权利要求5所述的语音交互方法,其特征在于,所述在对所述语音请求进行所述下游逻辑处理同时,发送异步请求,以根据上下文特征对所述语音请求进行第一语义拒识,得到第一语义拒识结果,包括:

通过所述中控服务发送所述异步请求至语义拒识服务进行入参;

通过所述语义拒识服务获取所述语音识别文本特征、所述声学特征和所述声学拒识结果对所述语音请求进行第一语义拒识,得到第一语义拒识结果。

7. 根据权利要求6所述的语音交互方法,其特征在于,所述在对所述语音请求进行所述下游逻辑处理同时,发送异步请求,以根据上下文特征对所述语音请求进行第一语义拒识,得到第一语义拒识结果的步骤之后,所述语音交互方法包括:

将所述第一语义拒识结果存入所述数据存储服务。

8. 根据权利要求7所述的语音交互方法,其特征在于,所述在接收到下游逻辑处理结果后,发送同步请求,以根据所述下游逻辑处理结果和所述第一语义拒识结果对所述语音请求进行第二语义拒识,得到第二语义拒识结果,包括:

在接收到所述下游逻辑处理结果后,通过所述中控服务基于所述下游逻辑处理结果发送所述同步请求至语义拒识服务进行入参;

通过所述语义拒识服务获取所述数据存储服务的所述第一语义拒识结果;

根据所述第一语义拒识结果和所述下游逻辑处理结果进行第二语义拒识,融合得到第二语义拒识结果。

9. 一种服务器,其特征在于,所述服务器包括处理器和存储器,所述存储器上存储有计算机程序,当所述计算机程序被所述处理器执行时,实现权利要求1-8任一项所述的语音交互方法。

10. 一种包含有计算机程序的非易失性计算机可读存储介质,其特征在于,当所述计算机程序被一个或多个处理器执行时,实现权利要求1-8任一项所述的语音交互方法。

## 语音交互方法、服务器和存储介质

### 技术领域

[0001] 本发明涉及语音交互技术领域,特别涉及一种语音交互方法、服务器和存储介质。

### 背景技术

[0002] 多模态特征语义拒识在学术界发展迅速,但工业落地鲜有方案。目前,多种特征提取与模型推理的相关语音交互方案在端到端上耗时较大,不能满足语音交互实时性的要求,用户体验不佳。

### 发明内容

[0003] 本发明实施方式提供一种语音交互方法、服务器和存储介质。

[0004] 本发明实施方式提供一种语音交互方法。所述语音交互方法包括:接收车辆转发的语音请求,对所述语音请求进行下游逻辑处理;在对所述语音请求进行所述下游逻辑处理同时,发送异步请求,以根据上下文特征对所述语音请求进行第一语义拒识,得到第一语义拒识结果;在接收到下游逻辑处理结果后,发送同步请求,以根据所述下游逻辑处理结果和所述第一语义拒识结果对所述语音请求进行第二语义拒识,得到第二语义拒识结果;将所述第二语义拒识结果下发至所述车辆以完成语音交互。

[0005] 如此,本发明的语音交互方法根据上下文特征对语音请求进行第一语义拒识并根据第一语义拒识结果和下游逻辑处理结果对语音请求进行第二语义拒识得到第二语义拒识结果,以完成语音交互,能够使语音交互过程的端到端时延大大降低,满足语音交互的实时性需求。

[0006] 所述接收车辆转发的语音请求,对所述语音请求进行下游逻辑处理,包括:将所述语音请求发送至中控服务以使所述中控服务将所述语音请求发送至对话系统下游服务进行入参处理,由所述对话系统下游服务实现所述下游逻辑处理。

[0007] 如此,本发明通过将语音请求发送至对话系统下游服务实现对语音请求的下游逻辑处理。

[0008] 所述将所述语音请求发送至中控服务以使所述中控服务将所述语音请求发送至对话系统下游服务进行入参处理,由所述对话系统下游服务实现所述下游逻辑处理,包括:将所述语音请求发送至所述中控服务以使所述中控服务将所述语音请求发送至对话系统下游服务进行自然语言理解、对话管理和/或业务机器人入参,从而由所述对话系统下游服务实现自然语言理解、对话管理和业务机器人对应的业务逻辑处理。

[0009] 如此,本发明可以通过自然语言理解、对话管理或业务机器人入参方式,实现对语音请求的自然语言理解、对话管理或业务机器人对应的业务逻辑处理。

[0010] 所述接收车辆转发的语音请求,对所述语音请求进行下游逻辑处理的步骤之后,所述语音交互方法包括:通过中控服务将所述语音请求的语音识别文本特征发送至上下文服务,以将所述语音识别文本特征作为所述上下文特征存入数据存储服务。

[0011] 如此,本发明的语音交互方法可以在上下文特征中写入ASR特征,并将存储有ASR

特征的上下文特征存储在数据存储服务中,以便于后续的异步请求时从上下文服务中直接获取具有ASR特征的上下文特征。

[0012] 所述在对所述语音请求进行所述下游逻辑处理同时,发送异步请求,以根据上下文特征对所述语音请求进行第一语义拒识,得到第一语义拒识结果的步骤之前,所述语音交互方法包括:在对所述语音请求进行下游逻辑处理同时,将所述语音请求发送至声学拒识服务进行处理,以得到声学特征;根据所述声学特征对所述语音请求进行拒识得到声学拒识结果;将所述声学特征和所述声学拒识结果发送至所述上下文服务,以将所述声学特征和所述声学拒识结果作为上下文特征存入所述数据存储服务。

[0013] 如此,本发明的语音交互方法可以在上下文特征中写入声学特征和声学拒识结果,并将存储有声学特征和声学拒识结果的上下文特征存储在数据存储服务中,以便于后续的异步请求时从上下文服务中直接获取具有声学特征和声学拒识结果的上下文特征。

[0014] 所述在对所述语音请求进行所述下游逻辑处理同时,发送异步请求,以根据上下文特征对所述语音请求进行第一语义拒识,得到第一语义拒识结果,包括:通过所述中控服务发送所述异步请求至语义拒识服务进行入参;通过所述语义拒识服务获取所述语音识别文本特征、所述声学特征和所述声学拒识结果对所述语音请求进行第一语义拒识,得到第一语义拒识结果。

[0015] 如此,本发明通过中控服务发送异步请求至语义拒识服务进行入参,并通过语义拒识服务对语音请求进行第一语义拒识,从而得到第一语义拒识结果,将时延隐藏在主干链路中,能够降低本方案进行语音交互的端到端时延。

[0016] 所述在对所述语音请求进行所述下游逻辑处理同时,发送异步请求,以根据上下文特征对所述语音请求进行第一语义拒识,得到第一语义拒识结果的步骤之后,所述语音交互方法包括:将所述第一语义拒识结果存入所述数据存储服务。

[0017] 如此,便于后续第二语义拒识时及时从数据存储服务中读取第一语音拒识结果。

[0018] 所述在接收到下游逻辑处理结果后,发送同步请求,以根据所述下游逻辑处理结果和所述第一语义拒识结果对所述语音请求进行第二语义拒识,得到第二语义拒识结果,包括:在接收到所述下游逻辑处理结果后,通过所述中控服务基于所述下游逻辑处理结果发送所述同步请求至语义拒识服务进行入参;通过所述语义拒识服务获取所述数据存储服务的所述第一语义拒识结果;根据所述第一语义拒识结果和所述下游逻辑处理结果进行第二语义拒识,融合得到第二语义拒识结果。

[0019] 如此,本发明的语音交互方法在进行第二次同次请求时,通过直接读取拒识模型结果并做规则预测,可以进一步降低语音交互过程的端到端时延,同时保证语音交互的准确性。

[0020] 本发明提供一种服务器。所述服务器包括处理器和存储器,所述存储器上存储有计算机程序,当所述计算机程序被所述处理器执行时,实现上述实施方式任一项所述的语音交互方法。

[0021] 如此,本发明的服务器应用上述语音交互方法根据上下文特征对语音请求进行第一语义拒识并根据第一语义拒识结果和下游逻辑处理结果对语音请求进行第二语义拒识得到第二语义拒识结果,以完成语音交互,能够使语音交互过程的端到端时延大大降低。

[0022] 本发明还提供一种包含有计算机程序的非易失性计算机可读存储介质。当所述计

计算机程序被一个或多个处理器执行时,实现上述实施方式任一项所述的语音交互方法。

[0023] 如此,本发明的存储介质应用上述语音交互方法根据上下文特征对语音请求进行第一语义拒识并根据第一语义拒识结果和下游逻辑处理结果对语音请求进行第二语义拒识得到第二语义拒识结果,以完成语音交互,能够使语音交互过程的端到端时延大大降低。

[0024] 本发明实施方式的附加方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本发明的实践了解到。

## 附图说明

[0025] 本发明的上述和/或附加的方面和优点可以从结合下面附图对实施方式的描述中将变得明显和容易理解,其中:

图1是相关技术中语音交互方案的工作流程的示意图;

图2是本发明的语音交互方法的流程示意图之一;

图3是本发明的语音交互方法的工作流程的示意图;

图4是本发明的语音交互方法的流程示意图之二;

图5是本发明的语音交互方法的流程示意图之三;

图6是本发明的语音交互方法的流程示意图之四。

## 具体实施方式

[0026] 下面详细描述本发明的实施方式,所述实施方式的示例在附图中示出,其中,相同或类似的标号自始至终表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施方式是示例性的,仅用于解释本发明的实施方式,而不能理解为对本发明的实施方式的限制。

[0027] 对于多种特征提取与模型推理的相关方案例如可以为如图1所示的语音交互方案。图1中的语音交互方案仅通过一次同步请求对语音请求进行语义拒识,图1中标粗部分为该语义拒识方案的主干链路时延。

[0028] 具体地,图1中,首先通过云端的自动语音识别(Automatic Speech Recognition, ASR)服务获取请求音频和ASR特征,然后自动语音识别服务广播ASR特征结果请求中控服务,由中控服务将ASR特征写入对话管理上下文服务。对话管理上下文服务可以存储上下文多模态特征,并将上下文多模态特征处理结果返回中控服务,将ASR特征结果存入数据存储库中。当下游服务进行下游逻辑处理完成后,可以将下游逻辑处理结果返回至中控服务,此时中控服务向语义拒识服务进行同步请求字段及音频,对话管理上下文服务再向语义拒识服务读取上下文ASR特征并返回上下文ASR特征结果至语义拒识服务。语义拒识服务再将音频请求传递给声学拒识服务。声学拒识服务可以根据音频请求返回声学特征和声学拒识结果至语义拒识服务。最后,将声学特征和声学拒识结果存入对话管理上下文服务,同时将声学特征和声学拒识结果存入数据存储服务(redis)中返回多模态特征拒识至语义拒识服务中。

[0029] 其中,图1中从ASR识别服务到对话系统中其他下游服务之间的语音处理过程的时延大于200ms,耗时较大,不能满足语音交互实时性的要求,用户体验不佳。

[0030] 有鉴于此,请参阅图2,本发明提供一种语音交互方法。该语音交互方法包括:

01:接收车辆转发的语音请求,对语音请求进行下游逻辑处理;

03:在对语音请求进行下游逻辑处理同时,发送异步请求,以根据上下文特征对语音请求进行第一语义拒识,得到第一语义拒识结果;

05:在接收到下游逻辑处理结果后,发送同步请求,以根据下游逻辑处理结果和第一语义拒识结果对语音请求进行第二语义拒识,得到第二语义拒识结果;

07:将第二语义拒识结果下发至车辆以完成语音交互。

[0031] 本发明还提供一种服务器。服务器包括处理器和存储器,存储器上存储有计算机程序,处理器用于接收车辆转发的语音请求,对语音请求进行下游逻辑处理;在对语音请求进行下游逻辑处理同时,发送异步请求,以根据上下文特征对语音请求进行第一语义拒识,得到第一语义拒识结果;在接收到下游逻辑处理结果后,发送同步请求,以根据下游逻辑处理结果和第一语义拒识结果对语音请求进行第二语义拒识,得到第二语义拒识结果;将第二语义拒识结果下发至车辆以完成语音交互。

[0032] 也即是,请参阅图3,本发明对语音请求进行语义拒识方案流程图如图3所示。

[0033] 本发明的语音交互方法先接收车辆转发的语音请求,对语音请求进行下游逻辑处理。请结合图3,如图3所示的步骤1-7,本发明先通过ASR识别服务获取车辆转发的语音请求的请求音频和ASR特征,然后对该语音请求进行一系列的下游逻辑处理。下游逻辑处理指的是ASR识别服务请求中控服务,并通过中控服务将语音请求发送至对话系统中其他下游服务进行下游逻辑处理。图3中标粗部分为本发明的语义拒识方案的主干链路时延。

[0034] 在对语音请求进行下游逻辑处理同时,发送异步请求,以根据上下文特征对语音请求进行第一语义拒识,得到第一语义拒识结果。具体地,如图3中所示的步骤8-11,在对语音请求进行下游逻辑处理同时,中控服务向语义拒识服务发送异步请求,并向对话管理上下文服务读取上下文声学特征、ASR特征和声学拒识结果,从而根据上下文特征对语音请求进行第一语义拒识,得到第一语义拒识结果。其中,拒识指的是判断语音请求是否为集外词或者无效输入。本发明的语音交互方法考虑了上下文信息,可以使得语音请求的特征提取的结果更准确。

[0035] 在接收到下游逻辑处理结果后,发送同步请求,以根据下游逻辑处理结果和第一语义拒识结果对语音请求进行第二语义拒识,得到第二语义拒识结果。具体地,如图3所示的步骤14-15,中控服务在接收到由对话系统中其他下游服务的下游逻辑处理结果后,可以向语义拒识服务发送同步请求,语义拒识服务根据下游逻辑处理结果和第一语义拒识结果对语音请求进行第二语义拒识,得到第二语义拒识结果。

[0036] 最后,语义拒识服务从数据存储服务中读取多模态拒识模型结果,得到三代拒识模型结果,语义拒识服务将三代拒识模型结果,三代拒识模型结果即为第二语义拒识结果,即直接读取模型结果并做规则预测得到第二语义拒识结果返回中控服务中,从而将第二语义拒识结果下发至车辆以完成语音交互。

[0037] 也即是,请结合图3,本发明可以通过第一次异步请求(步骤8),将时延隐藏在主干链路中,二次同步请求(步骤15)时,直接读取模型结果并做规则预测,本发明的方案中图3的步骤1-18对比相关方案中图1的步骤7-14,端到端时延能从200ms降低到6-10ms。

[0038] 如此,本发明的语音交互方法根据上下文特征对语音请求进行第一语义拒识并根据第一语义拒识结果和下游逻辑处理结果对语音请求进行第二语义拒识得到第二语义拒

识结果,以完成语音交互,能够使语音交互过程的端到端时延大大降低,满足语音交互的实时性需求。

[0039] 步骤01包括:

011:将语音请求发送至中控服务以使中控服务将语音请求发送至对话系统下游服务进行入参处理,由对话系统下游服务实现下游逻辑处理。

[0040] 处理器用于将语音请求发送至中控服务以使中控服务将语音请求发送至对话系统下游服务进行入参处理,由对话系统下游服务实现下游逻辑处理。

[0041] 具体地,如图3中的步骤1.2和步骤2.1所示,ASR识别服务将语音请求发送至中控服务,以使中控服务将语音请求发送至对话系统下游服务进行入参处理,由对话系统下游服务实现下游逻辑处理。

[0042] 如此,本发明通过将语音请求发送至对话系统下游服务实现对语音请求的下游逻辑处理。

[0043] 更具体地,步骤011包括:

0111:将语音请求发送至中控服务以使中控服务将语音请求发送至对话系统下游服务进行自然语言理解、对话管理和/或业务机器人入参,从而由对话系统下游服务实现自然语言理解、对话管理和业务机器人对应的业务逻辑处理。

[0044] 处理器用于将语音请求发送至中控服务以使中控服务将语音请求发送至对话系统下游服务进行自然语言理解、对话管理和/或业务机器人入参,从而由对话系统下游服务实现自然语言理解、对话管理和业务机器人对应的业务逻辑处理。

[0045] 具体地,对话系统下游服务对语音请求进行入参处理的方式包括自然语音理解入参、对话管理入参和业务机器人入参的三种入参处理方式中的一种或多种。

[0046] 当入参处理的方式包括自然语音理解入参的处理方式时,可以实现对语音请求的自然语言理解对应的业务逻辑处理;当入参处理的方式包括对话管理入参的处理方式时,可以实现对语音请求的对话管理对应的业务逻辑处理;当入参处理的方式包括业务机器人入参的处理方式时,可以实现对语音请求的业务机器人对应的业务逻辑处理。

[0047] 如此,本发明可以通过自然语言理解、对话管理或业务机器人入参方式,实现对语音请求的自然语言理解、对话管理或业务机器人对应的业务逻辑处理。

[0048] 在步骤01之后,语音交互方法包括:

02:通过中控服务将语音请求的语音识别文本特征发送至上下文服务,以将语音识别文本特征作为上下文特征存入数据存储服务。

[0049] 处理器用于通过中控服务将语音请求的语音识别文本特征发送至上下文服务,以将语音识别文本特征作为上下文特征存入数据存储服务。

[0050] 具体地,如图3中的步骤1.2、2.1、2.2、3.1和3.2所示,中控服务接收ASR识别服务的语音识别文本(Automatic Speech Recognition,ASR)特征后,可以将ASR特征发送至上下文服务,并将ASR特征作为上下文特征存入数据存储服务。

[0051] 在特征处理阶段(步骤1~7),以用户的上一轮语音请求为“打开空调”,当前轮语音请求为“风量三挡”为例进行说明,对于当前轮语音请求“风量三挡”而言,将该语音请求的ASR特征存储至数据存储服务的步骤,可以表述为如下表1所示的表格形式。

[0052] 表1



关联步骤	输入	输出	耗时
1.2 2.1 2.2 3.1 3.2	用户语音请求 【风量三挡】	【三挡风】的ASR特征存入数据存储服务： {“asrAlignment”:[{“conf”:1.0,“end”:820,“pinyin”: “feng”,“start”:640,“word”：“风”},{“conf”:1.0, “end”:940,“pinyin”：“liang”,“start”:820,“word”: “量”},{“conf”:1.0,“end”:1120,“pinyin”：“san”, “start”:940,“word”：“三”},{“conf”:1.0,“end”:1340, “pinyin”：“dang”,“start”:1120,“word”：“档”}], “eof”:true}	异步 50ms

如此,本发明的语音交互方法可以在上下文特征中写入ASR特征,并将存储有ASR特征的上下文特征存储在数据存储服务中,以便于后续的异步请求时从上下文服务中直接获取具有ASR特征的上下文特征。

[0053] 请参阅图4,在步骤03之前,语音交互方法包括:

021:在对语音请求进行下游逻辑处理同时,将语音请求发送至声学拒识服务进行处理,以得到声学特征;

022:根据声学特征对语音请求进行拒识得到声学拒识结果;

023:将声学特征和声学拒识结果发送至上下文服务,以将声学特征和声学拒识结果作为上下文特征存入数据存储服务。

[0054] 处理器用于在对语音请求进行下游逻辑处理同时,将语音请求发送至声学拒识服务进行处理,以得到声学特征;根据声学特征对语音请求进行拒识得到声学拒识结果;将声学特征和声学拒识结果发送至上下文服务,以将声学特征和声学拒识结果作为上下文特征存入数据存储服务。

[0055] 具体地,如图3所示中的步骤1.1、4、5、6和7所示,在对语音请求进行下游逻辑处理同时,ASR识别服务还向声学拒识服务发送语音请求,声学拒识服务先进行处理得到声学特征,然后根据该声学特征判定语音请求是否拒识,即根据声学特征对语音请求进行拒识得到声学拒识结果。其中,声学特征指的是纯音频特征。

[0056] 声学拒识结果包括语音请求是无效输入、语音请求不是无效输入和不确定该语音请求是否为无效输入三种结果声学拒识结果可以用0、1和2表示,0表示放行,该语音请求不是无效输入,1表示拒识,该语音请求是无效输入,2表示不确定,不确定该语音请求是否为无效输入。

[0057] 在特征处理阶段(步骤1~7),以用户的上一轮语音请求为“打开空调”,当前轮语音请求为“风量三挡”为例进行说明,对于当前轮语音请求“风量三挡”而言,将该语音请求的声学特征和声学拒识结果存储至数据存储服务的步骤,可以表述为如下表2所示的表格形式。

[0058] 表2

关联步骤	输入	输出	耗时

1.145 67	用户 音频 【风量 三挡】	声学拒识结果:acoustic rejection=0存入上下文服务(0:放行;1:拒识;2:不确定) 声学MFCC特征:N-M特征矩阵存入上下文服务 与ASR特征组成上下文特征: {“data”; [{"MFCCFeat": [], "acousticRejection": "0", "asrAlignment": [{"conf": 1.0, "end": 1120, "pinyin": "da kai", "start": 660, "word": "打开"}, {"conf": 0.795, "end": 1820, "pinyin": "kongtiao", "start": 1120, "word": "空调"}], "eof": true}, {"MFCCFeat": [], "acousticRejection": "0", "asrAlignment": [{"conf": 1.0, "end": 820, "pinyin": "feng", "start": 640, "word": "风"}, {"conf": 1.0, "end": 940, "pinyin": "liang", "start": 820, "word": "量"}, {"conf": 1.0, "end": 1120, "pinyin": "san", "start": 940, "word": "三"}, {"conf": 1.0, "end": 1340, "pinyin": "dang", "start": 1120, "word": "档"}], "eof": true}]}	异步 100ms
-------------	------------------------	---	-------------

之后,声学拒识服务可以将声学特征和声学拒识结果作为上下文特征存入数据存储服务以便于后续的异步请求时从上下文服务中直接获取具有声学特征和声学拒识结果的上下文特征。

[0059] 如此,本发明的语音交互方法可以在上下文特征中写入声学特征和声学拒识结果,并将存储有声学特征和声学拒识结果的上下文特征存储在数据存储服务中,以便于后续的异步请求时从上下文服务中直接获取具有声学特征和声学拒识结果的上下文特征。

[0060] 其中,图3中对语音请求进行下游逻辑处理得到下游逻辑处理结果的步骤,可以表述为如下表3所示的表格形式。

[0061] 表3

关联 步骤	输入	输出	耗时

2.1 8	异步第一次请求体： <pre>{“data”:{“async”:1, “hardwareid”: “xxxxxx”,“msgld”: “xxxxxx”,“query”：“风 量三挡”,“recordid”: “rcidxxxxxx”, “scenelds”:xxxxxx}, “status”：“xxxxxx”}</pre>	得到下游逻辑处理结果后,组成第二次同 步请求体:{"data":{"async":1, "hardwareid":"xxxxxx","msgld": "xxxxxx","query":"风量三挡", "recordid":"rcidxxxxxx","scenelds": xxxxxx},"status":"xxxxxx", "domains":[{"domainConfidence": 0.91380006,"domainName":"ac", "intents":[{"intentConfidence":1.0, "intentName":"ac wind set", "slotConfidence":0.0,"slots": [{"name":"number","pos":[2.2], "rawvalue":"三","value type": "string"}]}]}]}]}	同步 200- 500ms
-------	---	---	---------------------

进一步地,请参阅图5,步骤03包括:

031:通过中控服务发送异步请求至语义拒识服务进行入参;

032:通过语义拒识服务获取语音识别文本特征、声学特征和声学拒识结果对语音请求进行第一语义拒识,得到第一语义拒识结果。

[0062] 处理器用于通过中控服务发送异步请求至语义拒识服务进行入参;通过语义拒识服务获取语音识别文本特征、声学特征和声学拒识结果对语音请求进行第一语义拒识,得到第一语义拒识结果。

[0063] 具体地,如图3的步骤8和步骤9所示,中控服务发送异步请求至语义拒识服务进行入参。然后,如图3中的步骤10和步骤11所示,语义拒识服务从对话管理上下文服务中获取语音识别文本特征、声学特征和声学拒识结果对语音请求进行第一语义拒识,得到第一语义拒识结果。

[0064] 在中控服务对语义拒识服务的第一次异步请求阶段(步骤8~13),以用户的上一轮语音请求为“打开空调”,当前轮语音请求为“风量三挡”为例进行说明,中控服务发送异步请求至语义拒识服务进行入参,语义拒识服务从对话管理上下文服务中获取语音识别文本特征、声学特征和声学拒识结果对语音请求进行第一语义拒识,得到第一语义拒识结果的步骤,可以表述为如下表4所示的表格形式。

[0065] 表4

关联步骤	输入	输出	耗时

8.191 0 11	异步请求体： {“data”： {“async”：1， “hardwareid”： “xxxxxx”， “msgld”： “xxxxxx”， “query”：“风 量三挡”， “recordid”： “rcidxxxxxx ”， “scenelds”： xxxxxx}， “status”： “xxxxxx”}	上下文特征： {“data”；[ {“MFCCFeat”：[ ]， “acousticRejection”：“0”，“asrAlignment”： [ {“conf”：1.0，“end”：1120，“pinyin”：“da kai”，“start”：660，“word”：“打开”}， {“conf”： ：“xxxxxx”，0.795，“end”：1820，“pinyin”：“kongtiao”， “start”：1120，word：“空调”} ]， “eof”：true }， {“MFCCFeat”：[ ]，“acousticRejection”：“0”， “asrAlignment”：[ {“conf”：1.0，“end”：820， “pinyin”：“feng”，“start”：640，“word”： “风”}， {“conf”：1.0，“end”：940，“pinyin”： “liang”，“start”：820，“word”：“量”}， {“conf”：1.0，“end”：1120，“pinyin”：“san”， “start”：940，“word”：“三”}， {“conf”：1.0， “end”：1340，“pinyin”：“dang”，“start”：1120， “word”：“档”} ] ]， “eof”：true } ] }	异步20ms
---------------	--	--	--------

其中，可以通过预先训练的声学拒识模型对语音请求进行第一语义拒识，因此，相对应地，第一语义拒识结果也可以称为多模态拒识模型结果。

[0066] 如此，本发明通过中控服务发送异步请求至语义拒识服务进行入参，并通过语义拒识服务对语音请求进行第一语义拒识，从而得到第一语义拒识结果，将时延隐藏在主干链路中，能够降低本方案进行语音交互的端到端时延。

[0067] 在步骤03之后，语音交互方法包括：

04：将第一语义拒识结果存入数据存储服务。

[0068] 处理器用于将第一语义拒识结果存入数据存储服务。

[0069] 具体地，如图3所示中的步骤12和13所示，本发明在语义拒识服务中得到第一语义拒识结果后，可以将第一语义拒识结果存储至数据存储服务。

[0070] 在将第一语义拒识结果存入数据存储服务的存储阶段（步骤12和13），以用户的上一轮语音请求为“打开空调”，当前轮语音请求为“风量三挡”为例进行说明，将第一语义拒识结果存入数据存储服务的步骤如表5所示。

[0071] 表5

关联步骤	输入	输出	耗时
------	----	----	----

12 13	<p>上下文特征:{"data":[{"MFCCFeat":[], "acousticRejection":"0", "asrAlignment":[{"conf":1.0,"end": 1120,"pinyin":"da kai","start": 660,"word":"打开"}, {"conf":0.795, "end":1820,"pinyin":"kongtiao", "start":1120,word:"空调"}]}, {"eof": true}, {"MFCCFeat":[], "acousticRejection":"0", "asrAlignment":[{"conf":1.0,"end": 820,"pinyin":"feng","start":640, "word":"风"}, {"conf":1.0,"end":940, "pinyin":"liang","start":820, "word":"量"}, {"conf":1.0,"end": 1120,"pinyin":"san","start":940, "word":"三"}, {"conf":1.0,"end": 1340,"pinyin":"dang","start":1120, "word":"档"}]}, {"eof":true}]} 异步 请求体: 异步请求体: {"data": { "async":1, "hardwareid": "xxxxxx", "msgld": "xxxxxx", "query": "风量三 挡", "recordid": "rcidxxxxxx", "scenelds":xxxxxx}, "status": "xxxxxx"}</p>	<p>模型结果: {"query": "风量三挡", "modelConfidence": { "noise":0.0, "clear": 1.0, "taskLabel": "xx", "taskLabel": "xx", "detail": { "taskLabel":0.9, "taskLabel2":0.8}}, "code",10000, "msg": "ok"}</p>	<p>异步 30ms</p>
-------	---	--	--------------------

如此,便于后续第二语义拒识时及时从数据存储服务中读取第一语音拒识结果。

[0072] 请参阅图6,步骤05包括:

051:在接收到下游逻辑处理结果后,通过中控服务基于下游逻辑处理结果发送同步请求至语义拒识服务进行入参;

052:通过语义拒识服务获取数据存储服务的第一语义拒识结果;

053:根据第一语义拒识结果和下游逻辑处理结果进行第二语义拒识,融合得到第二语义拒识结果。

[0073] 处理器用于在接收到下游逻辑处理结果后,通过中控服务基于下游逻辑处理结果发送同步请求至语义拒识服务进行入参;通过语义拒识服务获取数据存储服务的第一语义拒识结果;根据第一语义拒识结果和下游逻辑处理结果进行第二语义拒识,融合得到第二语义拒识结果。

[0074] 请结合图3,如图3中的步骤14-18所示,中控服务在接收到下游逻辑处理结果后,可以基于下游逻辑处理结果发送同步请求至语义拒识服务进行入参,语义拒识服务获取数据存储服务中之前存储的第一语义拒识结果,根据第一语义拒识结果和下游逻辑处理结果进行第二语义拒识,融合得到第二语义拒识结果。

[0075] 也即是,中控服务收到对话系统中其他下游服务(NLU/DM/BOT)返回的下游逻辑处理结果,以业务逻辑向语义拒识服务发出同步请求,语义拒识服务读取数据存储服务(REDIS)中多模态拒识模型结果,结合业务逻辑进行规则推理,然后融合得到第二语义拒识结果返回至中控服务中。

[0076] 在中控服务对语义拒识服务的第二次同步请求阶段(步骤14~18),以用户的上一轮语音请求为“打开空调”,当前轮语音请求为“风量三挡”为例进行说明,中控服务收到对话系统中其他下游服务(NLU/DM/BOT)返回的下游逻辑处理结果,以业务逻辑向语义拒识服务发出第二次同步请求的步骤可以表述为如表6所示的表格形式;语义拒识服务读取数据存储服务(REDIS)中多模态拒识模型结果的步骤可以表述为如表7所示的表格形式;结合业务逻辑进行规则推理,然后融合得到第二语义拒识结果返回至中控服务中的步骤可以表述为如表8所示的表格形式。

[0077] 表6

关联步骤	输入	输出
15	第二次同步请求体:{"data":{"async":1,"hardwareid":"xxxxxx","msgld":"xxxxxx","query":"风量三挡","recordid":"rcidxxxxxx","scenelds":xxxxxx},"status":"xxxxxx","domains":[{"domainConfidence":0.91380006,"domainName":"ac","intents":[{"intentConfidence":1.0,"intentName":"ac wind set","slotConfidence":0.0,"slots":[{"name":"number","pos":[2.2],"rawvalue":"三","value type":"string"}]}]}]}	无

表7

关联步骤	输入	输出	耗时
16 17	Key值: {“recordId”:“rcidxxxxxx”}	数据存储服务中的模型结果:{"query":"风量三挡","modelConfidence":{"noise":0.0,"clear":1.0,"taskLabel":"xx","detail":{"taskLabel":0.9,"taskLabel":0.8}},“code”:10000,“msg”:“ok”}	同步 1-5ms

表8

关联步骤	输入	输出	耗时

18	第二次同步请求体:{"data":{"async":1,"hardwareid":"xxxxxx","msgId":"xxxxxx","query":"风量三挡","recordid":"rcidxxxxxx","scenelds":xxxxxx},"status":"xxxxxx","domains":[{"domainConfidence":0.91380006,"domainName":"ac","intents":[{"intentConfidence":1.0,"intentName":"ac wind set","slotConfidence":0.0,"slots":[{"name":"number","pos":[2.2],"rawvalue":"三","value type":"string"}]}]}]} 数据存储服务中的模型结果:{"query":"风量三挡","modelConfidence":{"noise":0.0,"clear":1.0,"taskLabel":"xx","detail":{"taskLabel":0.9,"taskLabel":0.8}},"code":10000,"msg":"ok"}	最终结果输出: {“query”:“风量三挡”, “queryRewrite”:“风量三挡”,“queryState”, “clear”, “modelConfidence”: {“noise”:0.0,“clear”: 1.0,“taskLabel”: “2.7”:{“taskLabel”: “xx”,“detail”: {“taskLabel”:0.9, “taskLabe2”:0.8}}, “ruleConfidence”:1.0, “filterReason”: “number set”, “code”:10000,“msg”: “ok”}	同步 5ms
----	---	--	-----------

如此,本发明的语音交互方法在进行第二次同次请求时,通过直接读取拒识模型结果并做规则预测,可以进一步降低语音交互过程的端到端时延,同时保证语音交互的准确性。

[0078] 综上,本发明的语音交互方法可以通过第一次异步请求(步骤8),将时延隐藏在主干链路中,在第二次同步请求(步骤15)时,直接读取模型结果并做规则预测,可以使得本发明的语音交互方法的图3中的步骤1-18与相关技术方案的图1中的步骤7-14相比较而言,使用本发明的语音交互方法能够实现端到端时延从200ms降低到6-10ms。其中,图1中的语音交互方案的端到端时延为200ms,本发明的语音交互方案的端到端时延为6-10ms,具体的耗时分析对比表格如表9和表10所示。

[0079] 表9

相关技术的语音交互方案	耗时步骤	耗时
同步方案	图1中的步骤7-14	同步200ms,主干链路增加
	端到端耗时增加	200ms

表10

本发明的语音交互方案	耗时步骤	耗时
异步方案	1.2 2.1 2.2 3.1 3.2	异步50ms,掩藏在主干链路中
	1.1 4 5 6 7	异步100ms,掩藏在主干链路中
	8.1 9 10 11	异步20ms,掩藏在主干链路中
	2.1 8	主干链路200-500ms

	12 13	异步30ms,掩藏在主干链路中
	16 17	同步1-5ms,主干链路增加
	18	同步5ms,主干链路增加
	端到端耗时增加	6-10ms

本发明还提供一种包含有计算机程序的非易失性计算机可读存储介质。当计算机程序被一个或多个处理器执行时,实现上述任意实施例所述的语音交互方法。

[0080] 例如,计算机程序被处理器执行时实现以下语音交互方法的步骤:

01:接收车辆转发的语音请求,对语音请求进行下游逻辑处理;

02:在对语音请求进行下游逻辑处理同时,发送异步请求,以根据上下文特征对语音请求进行第一语义拒识,得到第一语义拒识结果;

03:在接收到下游逻辑处理结果后,发送同步请求,以根据下游逻辑处理结果和第一语义拒识结果对语音请求进行第二语义拒识,得到第二语义拒识结果;

04:将第二语义拒识结果下发至车辆以完成语音交互。

[0081] 可以理解,计算机程序包括计算机程序代码。计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。计算机可读存储介质可以包括:能够携带计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、以及软件分发介质等。

[0082] 本发明的存储介质应用上述语音交互方法根据上下文特征对语音请求进行第一语义拒识并根据第一语义拒识结果和下游逻辑处理结果对语音请求进行第二语义拒识得到第二语义拒识结果,以完成语音交互,能够使语音交互过程的端到端时延大大降低。



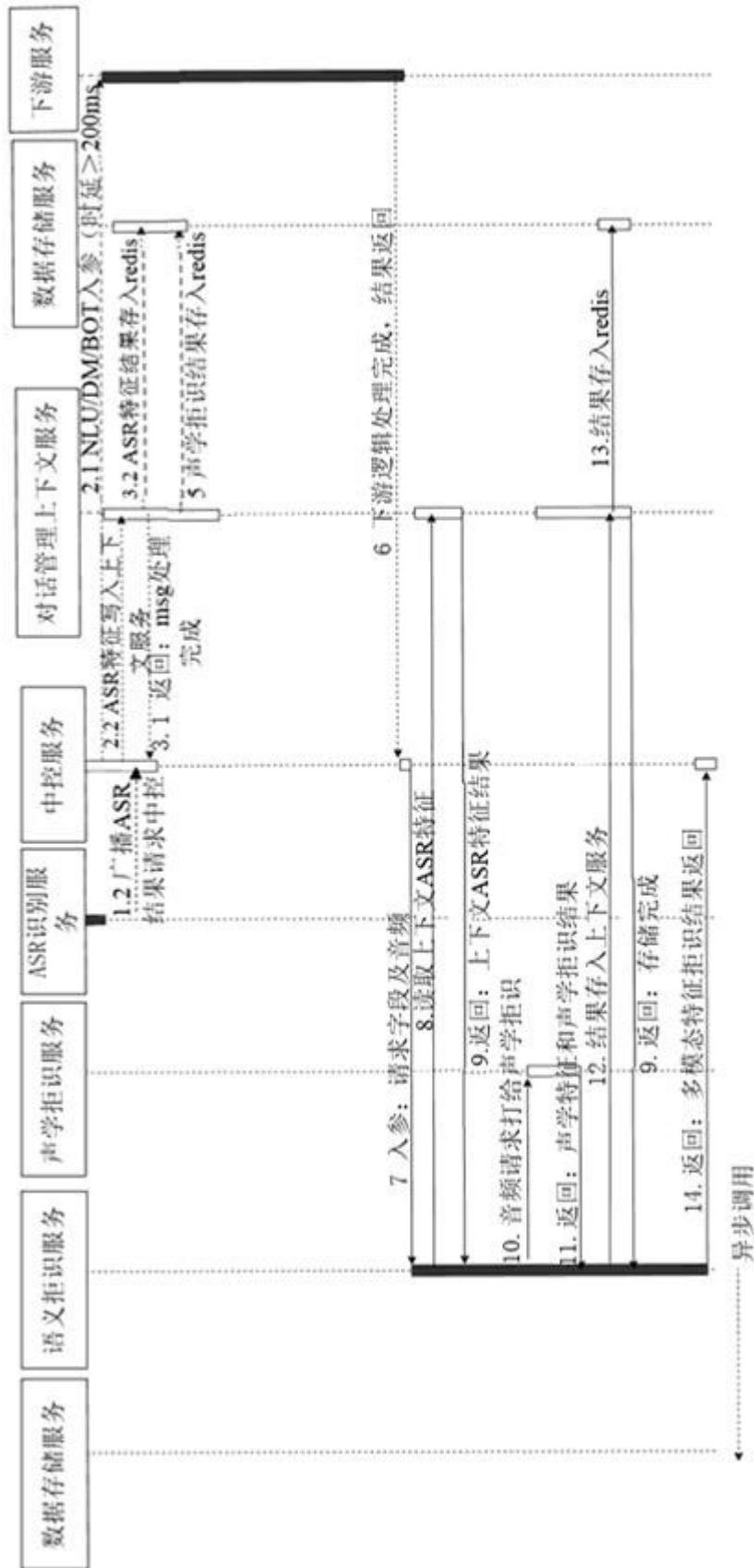


图1

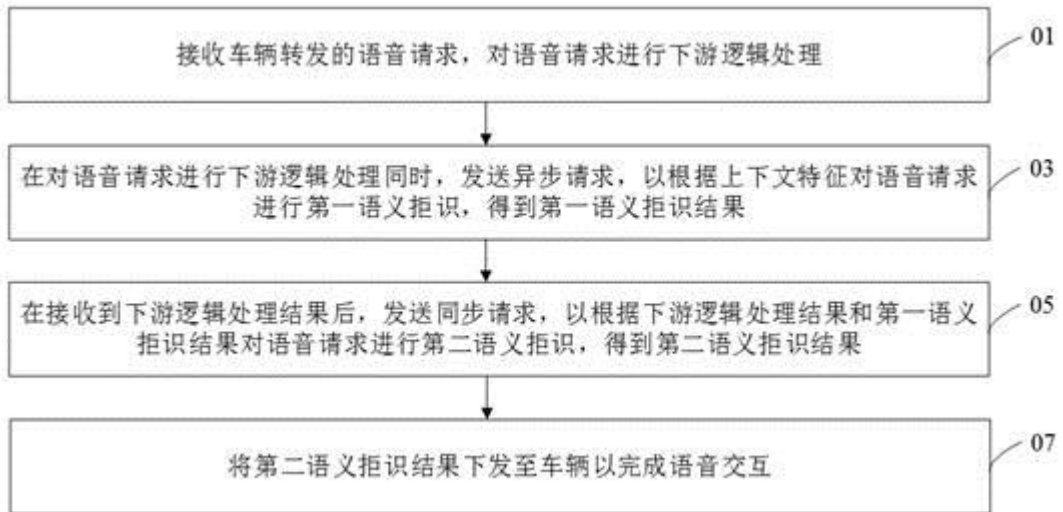


图2

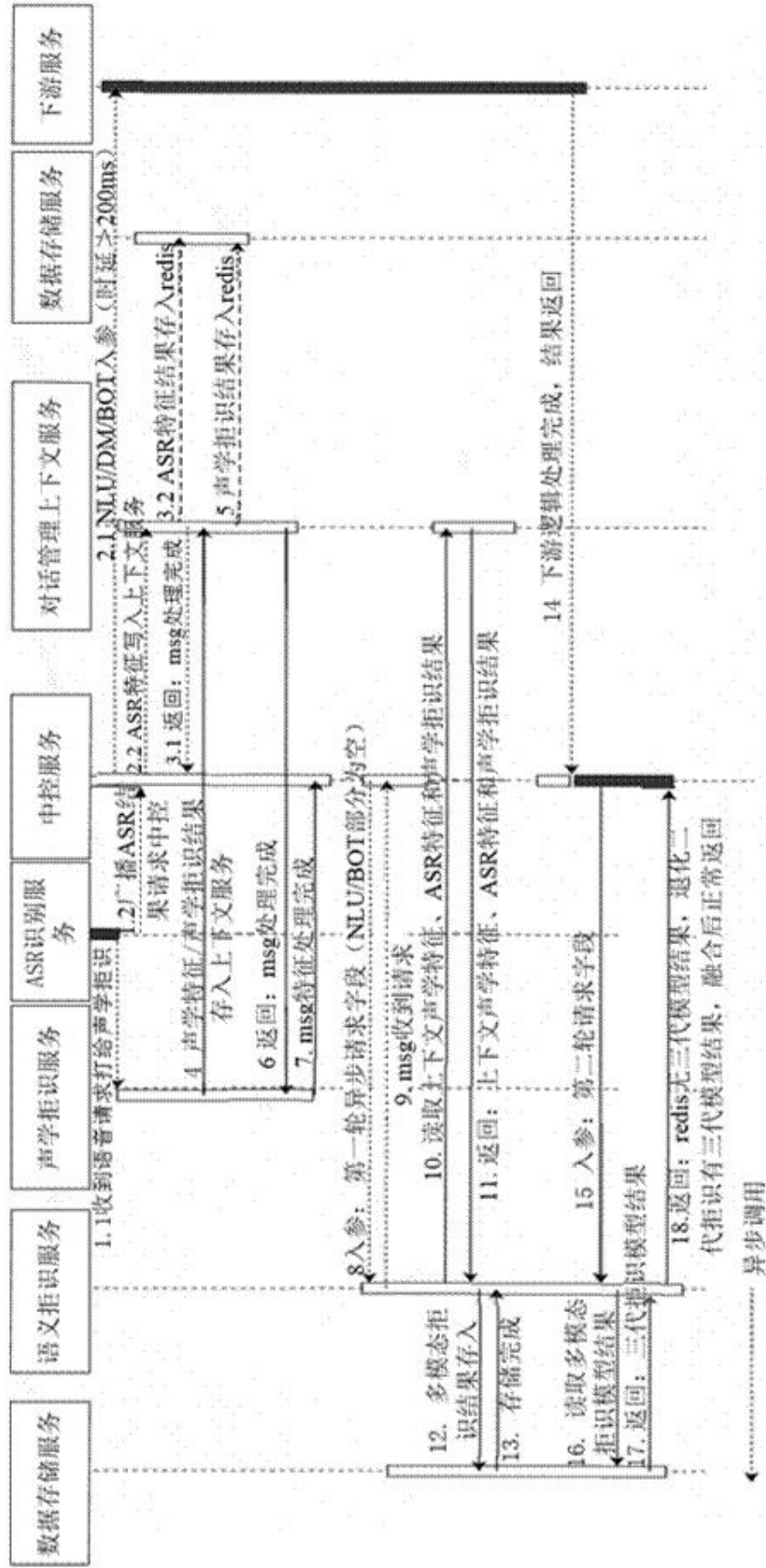


图3

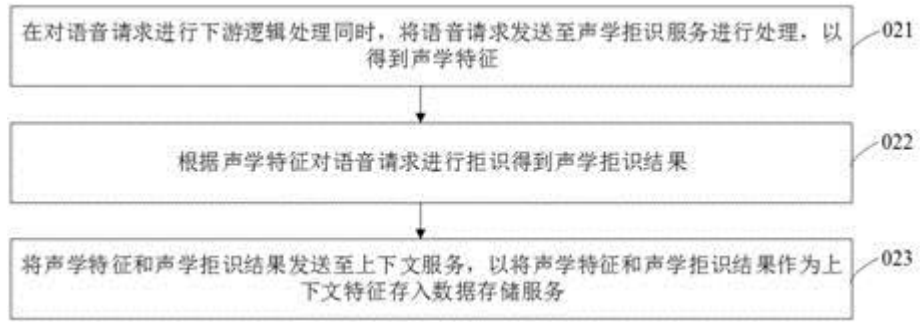


图4

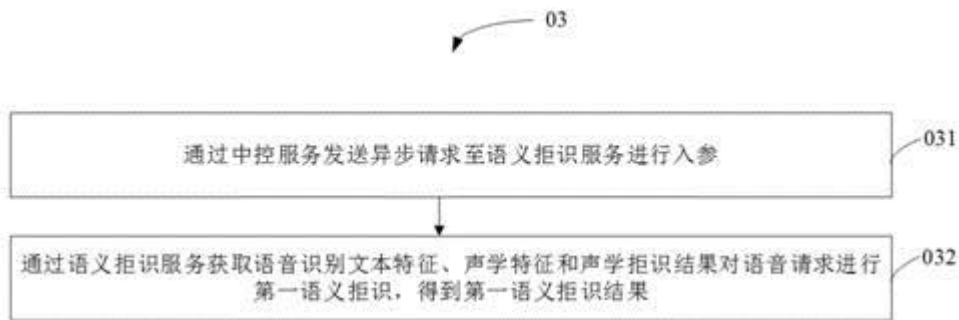


图5

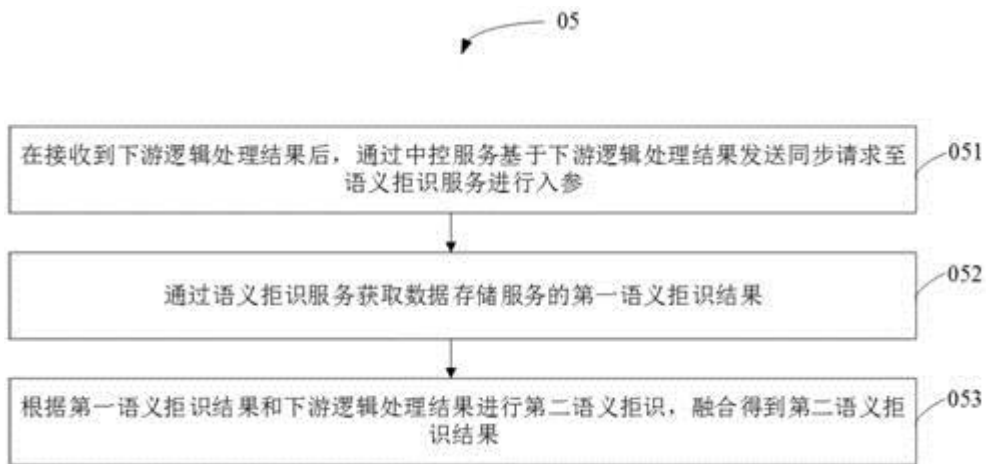


图6