



(12)发明专利申请

(10)申请公布号 CN 111369990 A

(43)申请公布日 2020.07.03

(21)申请号 202010091573.5

(22)申请日 2020.02.13

(71)申请人 北京达佳互联信息技术有限公司  
地址 100085 北京市海淀区上地西路6号1  
幢1层101D1-7

(72)发明人 张奇

(74)专利代理机构 北京润泽恒知识产权代理有  
限公司 11319

代理人 李娜

(51) Int. Cl.

G10L 15/22(2006.01)

G10L 15/30(2013.01)

G11B 20/10(2006.01)

G11C 7/16(2006.01)

G06F 3/16(2006.01)

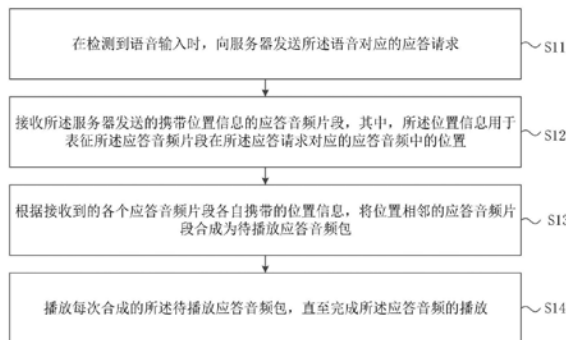
权利要求书2页 说明书19页 附图4页

(54)发明名称

音频播放方法、装置、终端、服务器及存储介  
质

(57)摘要

本申请提供了一种音频播放方法、装置、终  
端、服务器及存储介质。应用于终端的音频播放  
方法包括：在检测到语音输入时，向服务器发送  
所述语音对应的应答请求；接收所述服务器发送  
的携带位置信息的应答音频片段，其中，所述位  
置信息用于表征所述应答音频片段在所述应答  
请求对应的应答音频中的位置；根据接收到的各  
个应答音频片段各自携带的位置信息，将位置相  
邻的应答音频片段合成为待播放应答音频包；播  
放每次合成的所述待播放应答音频包，直至完成  
所述应答音频的播放。本申请采用音频片段边  
合成边播放的策略，显著提升了响应用户的语音  
请求的速度。



1. 一种音频播放方法,其特征在于,应用于终端,所述方法包括:  
检测到语音输入时,向服务器发送所述语音对应的应答请求;  
接收所述服务器发送的携带位置信息的应答音频片段,其中,所述位置信息用于表征所述应答音频片段在所述应答请求对应的应答音频中的位置;  
根据接收到的各个应答音频片段各自携带的位置信息,将位置相邻的应答音频片段合成为待播放应答音频包;  
播放每次合成的所述待播放应答音频包,直至完成所述应答音频的播放。
2. 根据权利要求1所述的方法,其特征在于,所述播放每次合成的所述待播放应答音频包,包括:  
当本次合成的所述待播放应答音频包包括的应答音频片段的数量大于预设数值时,播放本次合成的所述待播放应答音频包;  
当本次合成的所述待播放应答音频包包括的应答音频片段的数量小于所述预设数值,且当前时刻与上一次播放结束的结束时刻之间的时长达到第一预设时长时,播放本次合成的所述待播放应答音频包。
3. 根据权利要求1所述的方法,其特征在于,所述方法还包括:  
在接收到的各个应答音频片段位置不相邻,且当前时刻与上一次播放结束的结束时刻之间的时长达到第二预设时长时,根据位置信息,播放接收到的各个应答音频片段;  
在播放接收到的各个应答音频片段的过程中,若接收到位于已经播放或者正在播放的应答音频片段位置之后的应答音频片段,丢弃该接收到的应答音频片段。
4. 一种音频播放方法,其特征在于,应用于服务器,所述方法包括:  
接收终端发送的应答请求;  
根据所述应答请求获得多个应答音频数据;  
将所述多个应答音频数据中的至少一个应答音频数据合成为应答音频片段,向所述终端发送所述应答音频片段,直至完成所述多个应答音频数据的发送。
5. 一种音频播放装置,其特征在于,应用于终端,所述装置包括:  
第一发送模块,用于在检测到语音输入时,向服务器发送所述语音对应的应答请求;  
接收模块,用于接收所述服务器发送的携带位置信息的应答音频片段,其中,所述位置信息用于表征所述应答音频片段在所述应答请求对应的应答音频中的位置;  
合成模块,用于根据接收到的各个应答音频片段各自携带的位置信息,将位置相邻的应答音频片段合成为待播放应答音频包;  
第一播放模块,用于播放每次合成的所述待播放应答音频包,直至完成所述应答音频的播放。
6. 一种终端,其特征在于,包括:  
处理器;  
用于存储所述处理器可执行指令的存储器;  
其中,所述处理器被配置为执行以实现如权利要求1至3中任一项所述的音频播放方法所执行的操作。
7. 一种非临时性计算机可读存储介质,当所述存储介质中的指令由终端的处理器执行时,使得终端能够执行一种以实现如权利要求1至3中任一项所述的音频播放方法所执行的

操作。

8. 一种音频播放装置,其特征在于,应用于服务器,所述装置包括:

第一接收模块,用于接收终端发送的应答请求;

数据获取模块,用于根据所述应答请求获得多个应答音频数据;

第一发送模块,用于每将所述多个应答音频数据中的至少一个应答音频数据合成为应答音频片段,向所述终端发送所述应答音频片段,直至完成所述多个应答音频数据的发送。

9. 一种服务器,其特征在于,包括:

处理器;

用于存储所述处理器可执行指令的存储器;

其中,所述处理器被配置为执行以实现如权利要求4所述的音频播放方法所执行的操作。

10. 一种非临时性计算机可读存储介质,当所述存储介质中的指令由服务器的处理器执行时,使得服务器能够执行一种以实现如权利要求4所述的音频播放方法所执行的操作。

## 音频播放方法、装置、终端、服务器及存储介质

### 技术领域

[0001] 本公开涉及音频处理技术领域,尤其涉及一种音频播放方法、装置、终端、服务器及存储介质。

### 背景技术

[0002] 现如今,语音识别技术已被广泛地应用于各个方面,为人们的日常生活带来了诸多便利。例如其中一种应用是可安装于各个终端设备(例如:手机)上的语音助手,用户在外出前,通过该语音助手可以及时地了解户外天气状况,又例如用户在行驶过程中,通过该语音助手可以及时地了解到前方道路的交通拥堵情况。

[0003] 在相关技术中,用户通过终端向服务器发起语音请求后,需从服务器下载完整的音频文件并播放,而服务器通常需要较长的时间整理得到音频文件,因此,在相关技术中,需要用户等待较长的时间才能获得完整的音频文件,导致用户无法尽快地做出决策,极大地影响了用户的使用体验。

### 发明内容

[0004] 本公开提供一种音频播放方法、装置、终端、服务器及存储介质,旨在缩短响应用户的语音请求的延时,优化用户的使用体验。

[0005] 根据本公开实施例的第一方面,提供了一种音频播放方法,应用于终端,所述方法包括:

[0006] 在检测到语音输入时,向服务器发送所述语音对应的应答请求;

[0007] 接收所述服务器发送的携带位置信息的应答音频片段,其中,所述位置信息用于表征所述应答音频片段在所述应答请求对应的应答音频中的位置;

[0008] 根据接收到的各个应答音频片段各自携带的位置信息,将位置相邻的应答音频片段合成为待播放应答音频包;

[0009] 播放每次合成的所述待播放应答音频包,直至完成所述应答音频的播放。

[0010] 可选地,所述播放每次合成的所述待播放应答音频包,包括:

[0011] 当本次合成的所述待播放应答音频包包括的应答音频片段的数量大于预设数值时,播放本次合成的所述待播放应答音频包;

[0012] 当本次合成的所述待播放应答音频包包括的应答音频片段的数量小于所述预设数值,且当前时刻与上一次播放结束的结束时刻之间的时长达到第一预设时长时,播放本次合成的所述待播放应答音频包。

[0013] 可选地,所述方法还包括:

[0014] 在接收到的各个应答音频片段位置不相邻,且当前时刻与上一次播放结束的结束时刻之间的时长达到第二预设时长时,根据位置信息,播放接收到的各个应答音频片段;

[0015] 在播放接收到的各个应答音频片段的过程中,若接收到位于已经播放或者正在播放的应答音频片段位置之后的应答音频片段,丢弃该接收到的应答音频片段。

[0016] 可选地,所述方法还包括:

[0017] 在再次接收到应答音频片段的接收时刻与上一次播放结束的结束时刻之间的时长大于所述第二预设时长时,延长所述第一预设时长和/或增加所述预设数值。

[0018] 可选地,所述方法还包括:

[0019] 在接收到的各个应答音频片段位置不相邻,且未接收到的应答音频片段是目标音频片段时,向所述服务器请求所述目标音频片段,所述目标音频片段是所述应答音频中携带应答关键词的应答音频片段;

[0020] 在接收到所述目标音频片段后,将所述目标音频片段以及与所述目标音频片段位置相邻的应答音频片段合成为待播放应答音频包,并播放该待播放应答音频包。

[0021] 可选地,所述根据接收到的各个应答音频片段各自携带的位置信息,将位置相邻的应答音频片段合成为待播放应答音频包,包括:

[0022] 根据接收到的各个应答音频片段各自携带的位置信息,确定位置相邻的多个应答音频片段;

[0023] 对所述位置相邻的多个应答音频片段进行语义分析,在所述位置相邻的多个应答音频片段组成短句时,将所述位置相邻的多个应答音频片段合成为待播放应答音频包。

[0024] 可选地,在所述接收所述服务器发送的携带位置信息的应答音频片段之前,所述方法还包括:

[0025] 向所述服务器发送长连接建立请求,以提示所述服务器所述终端准备接收应答音频片段;

[0026] 接收所述服务器发送的长连接建立响应,与所述服务器建立长连接;

[0027] 接收所述服务器发送的携带位置信息的应答音频片段,包括:

[0028] 通过长连接接收所述服务器发送的携带位置信息的应答音频片段。

[0029] 根据本公开实施例的第二方面,提供了一种音频播放方法,应用于服务器,所述方法包括:

[0030] 接收终端发送的应答请求;

[0031] 根据所述应答请求获得多个应答音频数据;

[0032] 每将所述多个应答音频数据中的至少一个应答音频数据合成为应答音频片段,向所述终端发送所述应答音频片段,直至完成所述多个应答音频数据的发送。

[0033] 可选地,所述方法还包括:

[0034] 接收所述终端发送的目标音频片段请求,所述目标音频片段是所述应答音频中携带应答关键词的应答音频片段;

[0035] 向所述终端发送所述目标音频片段,以使所述终端将所述目标音频片段以及与所述目标音频片段位置相邻的应答音频片段合成为待播放应答音频包,并播放该待播放应答音频包。

[0036] 可选地,在所述向所述终端发送所述应答音频片段之前,所述方法还包括:

[0037] 接收所述终端发送长连接建立请求,所述长连接建立请求表征所述终端准备接收应答音频片段;

[0038] 向所述终端发送的长连接建立响应,与所述终端建立长连接;

[0039] 向所述终端发送所述应答音频片段,包括:

- [0040] 通过长连接向所述终端发送所述应答音频片段。
- [0041] 根据本公开实施例的第三方面,提供了一种音频播放装置,应用于终端,所述装置包括:
- [0042] 第一发送模块,用于在检测到语音输入时,向服务器发送所述语音对应的应答请求;
- [0043] 接收模块,用于接收所述服务器发送的携带位置信息的应答音频片段,其中,所述位置信息用于表征所述应答音频片段在所述应答请求对应的应答音频中的位置;
- [0044] 合成模块,根据接收到的各个应答音频片段各自携带的位置信息,将位置相邻的应答音频片段合成为待播放应答音频包;
- [0045] 第一播放模块,用于播放每次合成的所述待播放应答音频包,直至完成所述应答音频的播放。
- [0046] 可选地,所述播放模块,包括:
- [0047] 第一播放子模块,用于当本次合成的所述待播放应答音频包包括的应答音频片段的数量大于预设数值时,播放本次合成的所述待播放应答音频包;
- [0048] 第二播放子模块,用于当本次合成的所述待播放应答音频包包括的应答音频片段的数量小于所述预设数值,且当前时刻与上一次播放结束的结束时刻之间的时长达到第一预设时长时,播放本次合成的所述待播放应答音频包。
- [0049] 可选地,所述装置还包括:
- [0050] 第二播放模块,用于在接收到的各个应答音频片段位置不相邻,且当前时刻与上一次播放结束的结束时刻之间的时长达到第二预设时长时,根据位置信息,播放接收到的各个应答音频片段;
- [0051] 丢弃模块,用于在播放接收到的各个应答音频片段的过程中,若接收到位于已经播放或者正在播放的应答音频片段位置之后的应答音频片段,丢弃该接收到的应答音频片段。
- [0052] 可选地,所述装置还包括:
- [0053] 延迟模块,用于在再次接收到应答音频片段的接收时刻与上一次播放结束的结束时刻之间的时长大于所述第二预设时长时,延长所述第一预设时长和/或增加所述预设数值。
- [0054] 可选地,所述装置还包括:
- [0055] 请求模块,用于在接收到的各个应答音频片段位置不相邻,且未接收到的应答音频片段是目标音频片段时,向所述服务器请求所述目标音频片段,所述目标音频片段是所述应答音频中携带应答关键词的应答音频片段;
- [0056] 第三播放模块,用于在接收到所述目标音频片段后,将所述目标音频片段以及与所述目标音频片段位置相邻的应答音频片段合成为待播放应答音频包,并播放该待播放应答音频包。
- [0057] 可选地,所述合成模块,包括:
- [0058] 确定子模块,用于根据接收到的各个应答音频片段各自携带的位置信息,确定位置相邻的多个应答音频片段;
- [0059] 合成子模块,用于对所述位置相邻的多个应答音频片段进行语义分析,在所述位

置相邻的多个应答音频片段组成短句时,将所述位置相邻的多个应答音频片段合成为待播放应答音频包。

[0060] 可选地,所述装置还包括:

[0061] 第二发送模块,用于向所述服务器发送长连接建立请求,以提示所述服务器所述终端准备接收应答音频片段;

[0062] 建立长连接模块,用于接收所述服务器发送的长连接建立响应,与所述服务器建立长连接;

[0063] 所述接收模块,包括:

[0064] 接收子模块,用于通过长连接接收所述服务器发送的携带位置信息的应答音频片段。

[0065] 根据本公开实施例的第四方面,提供了一种音频播放装置,应用于服务器,所述装置包括:

[0066] 第一接收模块,用于接收终端发送的应答请求;

[0067] 数据获取模块,用于根据所述应答请求获得多个应答音频数据;

[0068] 第一发送模块,用于每将所述多个应答音频数据中的至少一个应答音频数据合成为应答音频片段,向所述终端发送所述应答音频片段,直至完成所述多个应答音频数据的发送。

[0069] 可选地,所述装置还包括:

[0070] 第二接收模块,用于接收所述终端发送的目标音频片段请求,所述目标音频片段是所述应答音频中携带应答关键词的应答音频片段;

[0071] 第二发送模块,用于向所述终端发送所述目标音频片段,以使所述终端将所述目标音频片段以及与所述目标音频片段位置相邻的应答音频片段合成为待播放应答音频包,并播放该待播放应答音频包。

[0072] 可选地,所述装置还包括:

[0073] 第三接收模块,用于接收所述终端发送长连接建立请求,所述长连接建立请求表征所述终端准备接收应答音频片段;

[0074] 建立长连接模块,用于向所述终端发送的长连接建立响应,与所述终端建立长连接;

[0075] 所述第一发送模块,包括:

[0076] 发送子模块,用于通过长连接向所述终端发送所述应答音频片段。

[0077] 根据本公开实施例的第五方面,提供一种终端,包括:

[0078] 处理器;

[0079] 用于存储该处理器可执行指令的存储器;

[0080] 其中,所述处理器被配置为执行以实现如第一方面所述的音频播放方法所执行的操作。

[0081] 根据本公开实施例的第六方面,提供一种非临时性计算机可读存储介质,当该存储介质中的指令由终端的处理器执行时,使得终端能够执行本公开实施例第一方面所述的音频播放方法所执行的操作。

[0082] 根据本公开实施例的第七方面,提供一种服务器,所述服务器包括:

[0083] 处理器；

[0084] 用于存储所述处理器可执行指令的存储器；

[0085] 其中,所述处理器被配置为执行本公开实施例第二方面所述的音频播放方法所执行的操作。

[0086] 根据本公开实施例的第八方面,提供一种非临时性计算机可读存储介质,当该存储介质中的指令由服务器的处理器执行时,使得服务器能够执行本公开实施例第二方面所述的音频播放方法所执行的操作。

[0087] 在本申请实施例中,终端在检测到语音输入时,向服务器发送所述语音对应的应答请求;接收所述服务器发送的携带位置信息的应答音频片段,其中,所述位置信息用于表征所述应答音频片段在所述应答请求对应的应答音频中的位置;根据接收到的各个应答音频片段各自携带的位置信息,将位置相邻的应答音频片段合成为待播放应答音频包;播放每次合成的所述待播放应答音频包,直至完成所述应答音频的播放。在本申请中,终端在向服务器发送音频数据后,无需等待服务器整理得到完整的音频文件后再从服务器中读取该完整的音频文件,而是直接根据服务器发送的各个应答音频片段各自携带的位置信息,合成应答音频包,从而实现音频片段的边合成边播放,提升了用户的语音请求的响应速度,能帮助用户更快速地做出决策,优化了用户的使用体验。

## 附图说明

[0088] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。

[0089] 图1是本申请一实施例示出的一种音频播放方法的流程图;

[0090] 图2是本申请一实施例示出的处理携带应答关键词的应答音频片段的方法的流程图;

[0091] 图3是本申请一实施例示出的另一种音频播放方法的流程图;

[0092] 图4是本申请一实施例示出的一种应用于终端的音频播放装置的结构示意图;

[0093] 图5是本申请一实施例示出的一种应用于服务器的音频播放装置的结构示意图;

[0094] 图6是本申请一实施例示出的一种终端的结构框图;

[0095] 图7是本申请一实施例示出的一种服务器的结构框图。

## 具体实施方式

[0096] 为了使本领域普通人员更好地理解本公开的技术方案,下面将结合附图,对本公开实施例中的技术方案进行清楚、完整地描述。

[0097] 需要说明的是,本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0098] 为克服相关技术中响应客用户的语音请求延时较长的问题,本申请一实施例提供



了一种音频播放方法,应用于终端,参考图1。图1是本申请一实施例示出的一种音频播放方法的流程图。如图1所示,本申请的音频播放方法包括以下步骤:

[0099] 步骤S11:在检测到语音输入时,向服务器发送所述语音对应的应答请求。

[0100] 在本实施例中,终端可以是:计算机、智能手机、平板等。终端上可以安装有客户端,通过客户端自动识别用户发出的语音,将语音转换为对应的音频数据,并将该音频数据作为应答请求发送至服务器,同时等待服务器针对该音频数据的应答。

[0101] 步骤S12:接收所述服务器发送的携带位置信息的应答音频片段,其中,所述位置信息用于表征所述应答音频片段在所述应答请求对应的应答音频中的位置。

[0102] 在本实施例中,服务器在接收到终端发送的音频数据后,首先整理得到对应该音频数据的目标文件,然后将目标文件划分为多个片段,针对每一个片段,生成一个应答音频片段,最终将生成的多个应答音频片段发送至终端,对终端发送的应答请求进行应答。每个应答音频片段均携带有位置信息,所述位置信息用于表征所述应答音频片段在所述应答请求对应的应答音频中的位置。

[0103] 示例地,用户发出的语音为“今天天气好吗?”,终端将该语音转换为音频数据并发送至服务器,服务器整理得到的目标文件可以是:“今天天气很好,气温25度”;然后将该目标文件划分为多个片段,例如可以是:“今天”、“天气”、“很好”、“气温”、“25”、“度”共6个片段,接着再为每一个片段添加位置信息,得到6个应答音频片段(关于服务器添加位置信息得到应答音频片段的详细过程请见下文),该6个应答音频片段组成了应答用户的此次语音的完整的应答音频包。

[0104] 步骤S13:根据接收到的各个应答音频片段各自携带的位置信息,将位置相邻的应答音频片段合成为待播放应答音频包。

[0105] 步骤S14:播放每次合成的所述待播放应答音频包,直至完成所述应答音频的播放。

[0106] 在本实施例中,终端提供接收服务和播放服务,接收服务用于接收服务器发送的应答音频片段并将其缓存到接收队列中,播放服务用于从接收队列中读取位置相邻多个的应答音频片段,并将位置相邻多个的应答音频片段合成为待播放应答音频包,然后播放该待播放音频包。

[0107] 在本实施例中,各个应答音频片段各自携带的位置信息可以用编号标识。终端每接收到一个应答音频片段,将该应答音频片段缓存到接收队列中。该接收队列按照各个应答音频片段的编号的大小顺序缓存有多个应答音频片段。示例地,假设当前时刻终端已经接收到了4个应答音频片段,分别是编号为3、5、6、7的音频片段,终端按照对接收到的多个应答音频片段编号的大小顺序,即:3、5、6、7,将位置信息相邻的多个应答音频片段合成为待播放应答音频包,即将上述4个应答音频片段合成待播放应答音频包。

[0108] 在一种可能的实施方式中,步骤S13包括:

[0109] 根据接收到的各个应答音频片段各自携带的位置信息,确定位置相邻的多个应答音频片段;

[0110] 对所述位置相邻的多个应答音频片段进行语义分析,在所述位置相邻的多个应答音频片段组成短句时,将所述位置相邻的多个应答音频片段合成为待播放应答音频包。

[0111] 在另一种可能的实施方式中,步骤S13包括:

[0112] 当本次合成的所述待播放应答音频包包括的应答音频片段的数量大于预设数值时,播放本次合成的所述待播放应答音频包。

[0113] 为使得合成的待播放应答音频包的语句意思流畅,本实施例提出将位置相邻的多个应答音频片段组成短句,或者编号连续的预设数值个音频片段,是终端执行“合成待播放应答音频包”的触发条件。预设数值可根据终端当前所处的网络的质量随机调整。

[0114] 在位置相邻的多个应答音频片段组成短句,是终端执行“合成待播放应答音频包”的触发条件的情况下,终端对接收到的应答音频片段进行语义检测,一旦检测到某些相邻的应答音频片段可以组成一个完整的短句,直接读取该位置相邻的多个应答音频片段,并合成为待播放应答音频包,然后播放所合成的应答音频包。

[0115] 示例地,假设接收队列中已经有三个位置相邻的应答音频片段,分别为“今天”、“天气”、“很好”。终端检测到这三个片段可以组成一个完整的短句“今天天气很好”,则将这三个应答音频片段合成待播放应答音频包。

[0116] 在位置相邻的预设数值个音频片段,是终端执行“合成待播放应答音频包”的触发条件的情况下,终端周期性检测接收队列中是否缓存有编号连续的预设数值个应答音频片段,一旦检测到接收队列中缓存有编号连续的预设数值个音频片段,直接读取该编号连续的预设数值个音频片段并合成为待播放应答音频包,然后播放所合成的应答音频包;在播放后,继续检测接收队列中是否缓存有编号连续的预设数值个音频片段,重复执行上述过程,直至播放完毕服务器发送的所有的应答音频片段。

[0117] 示例地,假设接收队列中已经有5个编号连续的应答音频片段,分别为:位置编号为3-7的应答音频片段,若预设数值被设置为5,终端检测到该接收队列中已缓存有编号连续的预设数值个应答音频片段,直接读取该位置编号为3-7的应答音频片段并合成为待播放应答音频包,然后播放所合成的应答音频包。在本申请实施例中,终端根据用户发出的语音向服务器发送音频数据,并接收服务器返回的应答音频片段,在接收到编号连续的预设数值个应答音频片段后,按照编号从小到大的顺序,预设数值将编号连续的预设数值个应答音频片段合成为待播放应答音频包并播放,直至完成应答音频的播放。

[0118] 确定位置相邻的应答音频片段,可以通过以下步骤实现:

[0119] 步骤A1:从已接收到的多个所述应答音频片段中确定音频片段组;

[0120] 步骤A2:确定所述音频片段组中最大的编号信息;

[0121] 步骤A3:将小于所述最大的编号信息且当前未接收的应答音频片段确定为与已接收到的多个所述应答音频片段编号信息连续的音频片段。

[0122] 其中,确定音频片段组时可以采取如下方式:

[0123] 针对已接收的多个音频片段,按照编号从小到大的顺序排列,并按照编号从小到大的顺序判断位置上相邻的两个音频片段的编号差值是否大于预设差值,当该编号差值大于预设编号差值时,在该相邻的两个音频片段之间添加间隔符,将该间隔符之前的 $N(1 \leq N < \text{预设数值})$ 个音频片段确定为音频片段组。

[0124] 在实际实施场景中,因网络的波动,终端一般不是严格按照编号的大小顺序接收到音频片段的,因此接收队列中已缓存的音频片段存在多种情况,下面将以预设数值为5为例,列举出几种常见的实施场景中确定与已接收到的多个音频片段编号连续的音频片段的方式:

[0125] 在第一种实施场景中,接收队列中已有的音频片段为:编号分别为8、9、11、12的音频片段,按照上述步骤A1-A3,以预设编号差值为3且预设数值为5为例,已接收的编号分别为8、9、11、12的音频片段中不存在间隔符,因此,将编号分别为8、9、11、12的多个音频片段确定为音频片段组;然后确定该音频片段组中最大的编号为12;最后将小于编号12的未接收的所有的音频片段确定为与已接收到的音频片段编号连续音频片段,例如:未接收的所有的音频片段可以是编号为10的音频片段,也可以是编号6、7、10的音频片段或者其它音频片段。

[0126] 在第二种实施场景中,接收队列中已有的音频片段为:编号分别为9、11、12、16、17的音频片段的音频片段,按照上述步骤A1-A3,以预设编号差值为3且预设数值为5为例,已接收的编号分别为9、11、12、16、17的音频片段中,12与16之间的编号差值为4,大于预设编号差值为3,存在间隔符,因此,将编号分别为9、11、12的多个音频片段确定为音频片段组;然后确定该音频片段组中最大的编号为12;最后将小于编号12的未接收的所有的音频片段确定为与已接收到的音频片段编号连续音频片段,例如:未接收的所有的音频片段可以是编号为10的音频片段,也可以是编号为8、10的音频片段,也可以是编号为7、8、10的音频片段。

[0127] 在第三种实施场景中,接收队列中已有的音频片段为:编号分别为9、11、12、16、17、18的音频片段的音频片段,按照上述步骤A1-A3,以预设编号差值为3且预设数值为5为例,已接收的编号分别为9、11、12、16、17、18的音频片段中,12与16之间的编号差值为4,大于预设编号差值为3,存在间隔符,因此,将编号分别为9、11、12的多个音频片段确定为音频片段组;然后确定该音频片段组中最大的编号为12;最后将小于编号12的未接收的所有的音频片段确定为与已接收到的音频片段编号连续音频片段,例如:未接收的所有的音频片段可以是编号为10的音频片段,也可以是编号为8、10的音频片段,也可以是编号为7、8、10的音频片段。

[0128] 在第四种实施场景中,接收队列中已有的音频片段为:编号分别为4、9、11、12、16、17、18的音频片段的音频片段,按照上述步骤A1-A3,以预设编号差值为3且预设数值为5为例,已接收的编号分别为4、9、11、12、16、17、18的音频片段中,4与9之间的编号差值为5,大于预设编号差值为3,存在间隔符,因此,将编号分别为4的多个音频片段确定为音频片段组;然后确定该音频片段组中最大的编号为4;最后将小于编号4的未接收的所有的音频片段确定为与已接收到的音频片段编号连续音频片段,例如:未接收的所有的音频片段可以是编号为3的音频片段,也可以是编号为2、3的音频片段。

[0129] 在本申请中,终端在向服务器发送音频数据后,无需等待服务器整理得到完整的音频文件(由应答语音的所有的音频片段合成的音频文件)再从服务器中读取该完整的音频文件,而是直接根据服务器发送的各个应答音频片段各自携带的位置信息,合成应答音频包,从而实现音频片段的边合成边播放,提升了用户的语音请求的响应速度,能帮助用户更快速地做出决策,优化了用户的使用体验。

[0130] 在实际实施过程中,音频文件中的应答关键词丢失将会导致终端通过语音无法正确地向用户传达信息(例如:在上述目标文件“今天天气很好,气温25度”中,“很好”和“25”为应答关键词,当上述两个应答关键词对应的应答音频片段丢失时,终端无法向用户传达准确的天气状况),本申请实施例考虑到上述问题,将应答音频片段分类为至少两种类型,

一种为携带普通信息的应答音频片段,另一种为携带应答关键词的应答音频片段。预设数值应答关键词预设数值。

[0131] 下面将对本申请中处理携带应答关键词的应答音频片段的过程进行详细说明。

[0132] 图2是本申请一实施例示出的处理携带应答关键词的应答音频片段的方法的流程图。参照图2,本申请提供的应答音频片段播放方法还包括以下步骤:

[0133] 步骤S21:在接收到的各个应答音频片段位置不相邻,且未接收到的应答音频片段是目标音频片段时,向所述服务器请求所述目标音频片段,所述目标音频片段是所述应答音频中携带应答关键词的应答音频片段。

[0134] 步骤S22:在接收到所述目标音频片段后,将所述目标音频片段以及与所述目标音频片段位置相邻的应答音频片段合成为待播放应答音频包,并播放该待播放应答音频包。

[0135] 在本实施例中,为便于陈述,将携带应答关键词的应答音频片段命名为目标音频片段,表示其携带应答关键词预设数值。

[0136] 本申请可以采用如下方式获得目标音频片段:在根据目标文件划分得到多个片段后,一种可能的实施方式是:为各个片段标记表征应答重要程度的分数,将分数大于预设分数的片段对应的应答音频片段确定为目标音频片段(即:携带应答关键词的应答音频片段),将分数不大于预设分数的片段对应的应答音频片段确定为普通音频片段(即:未携带应答关键词的音频片段)。另一种可能的实施方式是:对各个片段进行语义分析,根据语义分析结果和终端发送的应答请求,从各个片段中确定目标音频片段(即:携带应答关键词的应答音频片段)。

[0137] 在本申请各个实施例中,与已接收到的多个所述应答音频片段位置相邻的应答音频片段具有以下特征:未被终端接收;与已接收的多个应答音频片段中的部分或者全部应答音频片段构成一段位置相邻的应答音频片段。

[0138] 在确定与已接收到的多个音频片段编号连续的音频片段之后,判断与已接收到的多个音频片段编号连续的音频片段是否是目标音频片段,如果是目标音频片段,表示还有未接收的携带应答关键词的音频片段,为保证向用户准确地传达信息,需进一步获取到该目标音频片段,在获取该目标音频片段的过程中,同时还可能接收到与目标音频片段编号连续的其它的普通音频片段。

[0139] 在实际实施场景中,终端可以向服务器发送已准备好接收音频片段的通知消息,然后服务器可以首先从消息队列中读取所有的目标音频片段发送给终端,然后再读取普通音频片段发送给终端,从而增长了目标音频片段的接收时间,最大概率地保证所有的目标音频片段被终端接收到,提高了终端语音播放时为用户传达的信息的准确度。

[0140] 相应地,上述步骤S13可以是:

[0141] 在接收到所述目标音频片段及与所述目标音频片段的编号连续的多个音频片段时,按照编号从小到大的顺序,将所述目标音频片段及与所述目标音频片段的编号连续的多个音频片段合成为待播放应答音频包。

[0142] 示例地,在上述第一种实施场景中,音频片段组中已有编号为8、9、11、12的音频片段,假设未接收的所有的音频片段是编号为6、7、10的音频片段,且只有编号为7的音频片段是目标音频片段,编号为6、10的音频片段为普通音频片段,那么,终端继续等待并接收编号为7的音频片段,在接收编号为7的音频片段的过程中,可能还会接收到编号为6、10的音频

片段。假设最终接收到编号为6-12的音频片段,终端从接收队列中读取编号为6-10的5个音频片段,并将这5个音频片段合成为待播放音频包。

[0143] 本申请实施例为缩短响应用户的语音请求的延时,还设置了以下的强制播放策略。当满足下述任意条件之一时,需要执行强制播放策略:

[0144] 条件1:达到第一预设时长时,仍未接收到编号连续的预设数值个音频片段,且与已接收到的多个音频片段编号连续的音频片段不是目标音频片段。

[0145] 条件2:在第二预设时长内,仍未接收到目标音频片段。

[0146] 条件3:在第二预设时长内,接收到了目标音频片段,但不满足编号连续的第二预设数量个音频片段的播放条件。

[0147] 针对上述条件1,在上述列举的第一种实施场景中,以预设数值为5为例,假设未接收的音频片段是编号为6、7、10的音频片段,且编号为6、7、10的音频片段均为普通音频片段,那么终端的接收服务不再考虑继续接收编号为6、7、10的音频片段,而是向播放服务发送强制播放指令,使播放服务直接按照编号从小到大的顺序,即:8、9、11、12的顺序,依次从接收队列中读取编号最小的音频片段并合成相应的待播放应答音频包,以进行强制播放。如果在强制播放编号为8的音频片段对应的待播放应答音频包时,接收服务接收到了编号为10、13的音频片段,接收服务向播放服务器发送停止强制播放的指令,并告知播放服务播放编号为9-13的音频片段。由于编号为9-13的音频片段满足编号连续的预设数值个音频片段的条件,因此播放服务读取编号为9-13的音频片段,并将编号为9-13的音频片段合成为待播放应答音频包,并进行播放。

[0148] 针对上述条件2,在上述列举的第二种实施场景中,以预设数值为5为例,假设未接收的音频片段是编号为8、10的音频片段,且编号为8、10的音频片段均为目标音频片段,那么终端的接收服务继续等待第二预设时长以接收编号为8、10的音频片段,若在第二预设时长内仍未接收到编号为8、10的音频片段,接收服务放弃继续接收编号为8、10的音频片段,向播放服务发送强制播放指令,使播放服务直接按照编号从小到大的顺序,即:9、11、12、16、17的顺序,依次从接收队列中读取编号最小的音频片段并合成相应的待播放应答音频包,以进行强制播放。

[0149] 针对上述条件3,在上述列举的第三种实施场景中,以预设数值为5为例,假设未接收的音频片段是编号为8、10的音频片段,且编号为8的音频片段均为目标音频片段,编号为10的音频片段均为普通音频片段,接收服务如果在第二预设时长内接收到编号为8的音频片段,但是未接收到编号为10的音频片段,此时仍不满足编号连续的预设数值个音频片段的条件,但是仍向播放服务发送强制播放指令,使播放服务直接按照编号从小到大的顺序,即8、9、11、12、16、17、18的顺序,依次从接收队列中读取编号最小的音频片段并合成相应的待播放应答音频包,以进行强制播放。

[0150] 本申请实施例设置了强制播放策略,缩短了终端响应用户的语音请求的延时,以便于用户及时获得所请求的信息并做出决策,优化了用户的使用体验。

[0151] 在具体实施时,强制播放策略包括但不限于以下几种:

[0152] 强制播放策略一:当本次合成的所述待播放应答音频包包括的应答音频片段的数量小于所述预设数值,且当前时刻与上一次播放结束的结束时刻之间的时长达到第一预设时长时,播放本次合成的所述待播放应答音频包。

[0153] 本申请各个实施例中,为缩短响应用户的语音请求的延时,根据经验为已等待时长设置一个阈值,即:第一预设时长。

[0154] 在本实施例中,当前时刻与上一次播放结束的结束时刻之间的时长是通过以下方式确定的:终端的播放服务在开始播放本次合成的待播放应答音频包时,会记录起始播放时刻,并将该起始播放时刻发送给接收服务,使接收服务根据该起始播放时刻和当前时刻确定自播放服务上一次播放音频片段的时刻起的已等待时长。

[0155] 示例地,第一预设时长被设置为5秒,假设当前接收到3个应答音频片段,后续没有再收到其他的应答音频片段,且当前时刻与上一次播放结束的结束时刻之间的时长达到5秒,则为了避免用户继续等待,尽管当前接收到的应答音频片段的数量没有达到预设数值,也要将当前接收到的应答音频片段合成为待播放应答音频包,并进行播放。

[0156] 在本实施例中,在进行强制播放时,无需考虑强制播放的音频片段的数量是否符合预设数值,且强制播放一直持续到重新检测到位置相邻的预设数值个音频片段为止。

[0157] 考虑到在实际实施过程中可能导致应答关键词丢失的问题,终端在已等待时长达到第一预设时长时,如果已接收的多个音频片段不满足编号连续的预设数值个音频片段这一条件,判断与已接收到的多个音频片段编号连续的音频片段是否是目标音频片段,如果是目标音频片段,会继续等待目标音频片段,直到接收到该目标音频片段,且接收到与目标音频片段的编号连续的多个音频片段后,将包含该目标音频片段的多个编号连续的音频片段合成为待播放应答音频包,并播放。通过该方式,降低了应答关键词丢失的概率,保证了向用户传达的信息的准确度。

[0158] 为了避免等待目标音频片段的时间过长,还可以采用如下的强制播放策略二。

[0159] 强制播放策略二:在接收到的各个应答音频片段位置不相邻,且当前时刻与上一次播放结束的结束时刻之间的时长达到第二预设时长时,根据位置信息,播放接收到的各个应答音频片段。

[0160] 如前文所述,为保证终端向用户传达的信息的准确度,本申请中的音频片段分为两类:目标音频片段(包含应答关键词的音频片段)和普通音频片段(不包含应答关键词的音频片段)。若在第一预设时长内未接收到目标音频片段时,在第二预设时长内继续接收该目标音频片段,直到接收到包含该目标音频片段的多个编号连续的音频片段时,读取并将包含该目标音频片段的多个编号连续的音频片段合成为待播放应答音频包,并播放。若是在第二预设时长内依然没有接收到该目标音频片段,则不再继续等待,根据位置信息,将已接收到的各个应答音频片段合成待播放应答音频包并播放。

[0161] 示例地,用户发出的语音为“今天天气好吗?”,终端将该语音转换为音频数据并发送至服务器,服务器整理得到的目标文件可以是:“今天天气很好,气温25度”;然后将该目标文件划分为多个片段,例如可以是:“今天”、“天气”、“很好”、“气温”、“25”、“度”共6个片段,接着再为每一个片段添加位置信息,得到6个应答音频片段。因为网络不稳定,出现网络波动,导致服务器发送应答音频片段延迟,到第一预设时长截止,终端仅仅收到了“今天”、“很好”、“气温”、“度”,缺失了携带应答关键词“25”的应答音频片段,因为缺失的应答音频片段是目标音频片段,则在第二预设时长内继续等待接收该目标音频片段。若是到第二预设时长截止,依然没有接收到“25”的应答音频片段,则强制播放收到的四个应答音频片段“今天很好气温度”。避免用户等待时间过长。

[0162] 在本申请实施例中,每一个目标音频片段携带有预先设置的第二预设时长。终端在判断得到与已接收到的多个音频片段编号连续音频片段是目标音频片段时,首先获取该目标音频片段的第二预设时长,该第二预设时长生效的起始时刻为已等待时长超出第一预设时长的时刻,也即:判断与已接收到的多个所述音频片段编号连续音频片段是否是目标音频片段的时刻。

[0163] 在本申请实施例中,还可以将目标音频片段按照应答语音的重要程度再次划分为多个层级,并为不同的层级的目标音频片段设置不同的第二预设时长。是否将目标音频片段划分为多个层级,以及如何设置各个层级对应的重要程度区间可按照实际应用场景灵活设置,本申请对此不作具体限制。

[0164] 本申请通过为每一个目标音频片段设置第二预设时长,在缩短终端响应用户的语音请求的延时的情况下,降低了终端丢失目标音频片段的概率,保证终端向用户传达的信息的准确度,优化用户的使用体验。

[0165] 结合以上各个实施例,由于强制播放策略的使用,终端会不可避免地存在已经播放某音频片段后,接收到位于该已播放的音频片段的位置之后的音频片段的现象,如果不及处理这类音频片段,将导致语音播放混乱。因此,为保证在强制播放策略下语音的正常播放,本申请还提供了配合强制播放策略所使用的过期音频片段清理方法。具体为:

[0166] 在播放接收到的各个应答音频片段的过程中,若接收到位于已经播放或者正在播放的应答音频片段位置之后的应答音频片段,丢弃该接收到的应答音频片段。

[0167] 在本实施例中,终端的接收服务每次接收到新的音频片段时,从播放服务处获得当前正在播放的音频片段的编号,如果新的音频片段的编号小于当前正在播放的音频片段的编号,直接丢弃该新的音频片段,如果新的音频片段的编号不小于当前正在播放的音频片段的编号,将该新的音频片段缓存到接收队列中。

[0168] 示例地,以接收队列中缓存的音频片段为编号为7、8、10、13的音频片段且播放服务正在执行强制播放策略为例,假设播放服务正在强制播放编号为8的音频片段,接收服务接收到一个编号为6的音频片段时,判断该编号为6的音频片段的编号小于正在播放的编号为8的音频片段的编号,因此直接丢弃该编号为6的音频片段;如果接收服务接收到一个编号为9的音频片段,判断该编号为9的音频片段的编号大于正在播放的编号为8的音频片段的编号,因此,将该编号为9的音频片段缓存到接收队列中。

[0169] 本申请实施例针对强制播放策略设置了与其配套的清除过期音频片段的方法,避免了终端语音播放混乱的现象。

[0170] 在实际实施场景中,服务器与终端之间进行音频片段传输时所使用的信息通道的质量会随着网络的波动而波动,因此,本申请为进一步增强用户的使用体验,针对终端还设置了与当前网络质量对应的音频片段的接收策略。

[0171] 本申请实施例还提供一种动态调整音频片段的接收方法。具体为:

[0172] 在再次接收到应答音频片段的接收时刻与上一次播放结束的结束时刻之间的时长大于所述第二预设时长时,延长所述第一预设时长和/或增加所述预设数值。

[0173] 在本申请实施例中,再次接收到应答音频片段的接收时刻与上一次播放结束时刻之间的时长可以反映通信质量。终端的接收服务可以预先设置第二预设时长,将此预设时长对应的网络质量确定为预设质量。

[0174] 在一种实施方式中,当终端的接收服务再次接收到应答音频片段的接收时刻与上一次播放结束时刻之间的时长大于第二预设时长时,确定当前通信质量低于预设质量,通信质量较低,接收服务可以执行以下方式中的任意一种:增大预设数值、调大第一预设时长、增大预设数值且调大第一预设时长。

[0175] 在另一种实施方式中,当终端的接收服务次接收到应答音频片段的接收时刻与上一次播放结束时刻之间的时长小于第二预设时长时,确定当前通信质量不低于预设质量,通信质量较高,接收服务可以执行以下方式中的任意一种:减小预设数值、调小第一预设时长、减小预设数值且调小第一预设时长。

[0176] 在上述第一种实施方式中,当前通信质量较低时,如果不调整音频片段的接收方式,由于接收服务接收音频片段的速率较低,播放服务在播放完待播放应答音频包以后,需要等待较长时间才能再次播放,进而导致最终播放出的语音存在严重的卡顿现象,影响用户的使用体验。而本申请所提出的增大预设数值和/或延长第一预设时长的策略则较好地避免了该现象,在网络质量较差的情况下,接收服务等到接收到较多数量的音频片段时才进行合成和播放,能有效地保证播放出的语音的流畅度。

[0177] 在第二种实施方式中,当前通信质量较高时,由于接收服务接收音频片段的速率较快,播放服务减小预设数值和/或延长第一预设时长,可以充分将接收速度带来的效益最大化,不断地从接收队列中读取音频片段并进行合成和播放,在网络质量高于理想预设质量的情况下,能达到近乎实时响应用户的语音请求的效果。

[0178] 在本申请实施例中,终端实时检测网络质量,根据网络质量对音频片段的接收策略进行适应性调整,实现了音频片段的播放策略的动态调整,即:在网络质量差时,能保证播放的语音的流畅度,在网络质量好时,能有效缩短响应用户的语音请求的延时,加快响应速度,极大地优化了用户的使用体验。

[0179] 结合以上实施例,在本申请另一实施例中,步骤S12之前,还包括以下步骤:

[0180] 向所述服务器发送长连接建立请求,以提示所述服务器所述终端准备接收应答音频片段;

[0181] 接收所述服务器发送的长连接建立响应,与所述服务器建立长连接;

[0182] 接收所述服务器发送的携带位置信息的应答音频片段,包括:

[0183] 通过长连接接收所述服务器发送的携带位置信息的应答音频片段。

[0184] 具体为,终端在检测到语音输入时,向服务器发送该语音对应的应答请求,终端在接收到服务端的应答指令,确定可以接收并播放应答音频后,向服务器发送长连接建立请求,终端接收服务器发送的长连接建立响应后,与服务器建立长连接。然后终端通过这个长连接服务接收服务器发送的携带位置信息的应答音频片段。

[0185] 在另一种实施方式中,终端还可以执行以下步骤:

[0186] 接收所述服务器发送的所述音频文件对应的文本信息;

[0187] 显示所述文本信息。

[0188] 本申请为了进一步保证终端向用户传达的信息的准确度,终端在向服务器发送语音对应的音频数据之后,还可以接收服务器返回的用于应答该语音的音频文件的文本信息,并通过终端界面展示该文本信息。

[0189] 本申请还提供了一种音频播放方法,应用于服务器,如图3所示。图3是本申请一实



施例示出的另一种音频播放方法的流程图。参照图3,本申请提供的音频播放方法包括以下步骤:

[0190] 步骤S31:接收终端发送的应答请求。

[0191] 步骤S32:根据所述应答请求获得多个应答音频数据。

[0192] 步骤S33:每将所述多个应答音频数据中的至少一个应答音频数据合成为应答音频片段,向所述终端发送所述应答音频片段,直至完成所述多个应答音频数据的发送。

[0193] 在本申请实施例中,按照词语的形式将目标文件划分为多个片段,得到多个词语,为每一个片段编码即:为每一个词语编码,得到多个音频片段。

[0194] 示例地,音频数据为“今天天气好吗”时,获取的目标文件可以是“今天天气很好,气温25度”。划分得到的多个词语可以是:“今天”、“天气”、“很好”、“气温”、“25”、“度”共6个词,因此,为上述6个词语编码即可得到6个音频片段。

[0195] 在本实施例中,服务器接收终端发送的应答请求,还可以包括:

[0196] 接收所述终端发送的目标音频片段请求,所述目标音频片段是所述应答音频中携带应答关键词的应答音频片段;

[0197] 向所述终端发送所述目标音频片段,以使所述终端将所述目标音频片段以及与所述目标音频片段位置相邻的应答音频片段合成为待播放应答音频包,并播放该待播放应答音频包。

[0198] 示例地,服务器接收到终端发送的目标片段请求,目标片段具体为“25”,服务器立即将该片段发送给终端,终端接收该片段后结合相邻位置的应答音频片段“气温”、“度”合成“气温25度”的待播放应答音频包,并播放。

[0199] 在本实施例中,服务器向终端发送应答音频片段之前,还可以包括:

[0200] 接收所述终端发送长连接建立请求,所述长连接建立请求表征所述终端准备接收应答音频片段;

[0201] 向所述终端发送的长连接建立响应,与所述终端建立长连接;

[0202] 向所述终端发送所述应答音频片段,包括:通过长连接向所述终端发送所述应答音频片段。

[0203] 具体为,服务器在接收到终端发送的应答请求之后,开始分段合成应答语音片段,然后向终端发送应答指令,然后接收终端发送的长连接建立请求,向终端发送长连接建立响应,与终端之间建立长连接,最后服务器通过这个长连接服务向终端发送应答音频片段。

[0204] 在本实施例中,服务器在接收终端发送的音频数据,并获取到目标文件后,还可以执行以下步骤:

[0205] 向所述终端发送所述音频文件对应的文本信息,以使所述终端显示所述文本信息。

[0206] 按照生成多个所述音频片段的先后顺序为多个所述音频片段添加位置信息。

[0207] 示例地,针对上述6个音频片段,按照其在目标文件中的顺序,依次得到的位置信息为1-6。

[0208] 将多个携带位置信息的应答音频片段发送给所述终端,以使所述终端按照所述位置信息合成待播放应答音频包,并播放该待播放应答音频包。

[0209] 在一种实施方式中,服务器整理得到目标文件后,可以将该目标文件返回给终端,

使终端可以文字的方式展示给用户;在另一种实施方式中,终端在发起语音请求后,还可以提供语音播放按钮,当检测到用户对语音播放按钮的点击操作时,向服务器发送终端已准备好接收音频片段的通知消息,终端与服务器建立长连接,通过该长连接接收服务器发送的多个音频片段。

[0210] 考虑到用户可能在较长时间后才点击语音播放按钮,因此,在本申请实施例中,服务器在为各个片段编码音频片段时,若未接收到终端发送的已准备好接收音频片段的通知消息,将已编码得到的多个音频片段暂时缓存到消息队列中,并在接收到终端发送的已准备好接收音频片段的通知消息时,从该消息队列中读取各个音频片段并发送给终端。

[0211] 本申请所提供的音频片段具有下述多个有效效果:

[0212] 一、与相关技术中下载完整的音频文件后才能进行语音播放的方式相比,本申请采用边合成边播放的方式,能明显缩短响应用户的语音请求的延时,提升终端响应用户的语音请求的速度,为用户做出决策提供了巨大的便利。

[0213] 二、采用强制播放的策略,避免用户等待较长的时长,使得用户能尽快地获得针对其语音请求的语音响应。

[0214] 三、采用应答关键词策略,将音频分片划分为目标音频分片和普通音频分片两类,针对目标音频分片设置较长的接收时间,在将响应用户的语音请求的延时控制在合理的范围的情况下,降低丢失目标音频分片的概率,保证了向用户传达的信息的准确度。

[0215] 四、根据网络质量调整终端接收音频片段的方式,实现了音频片段的播放策略的动态调整,在网络质量差时,能保证播放的语音的流畅度,在网络质量好时,能有效缩短响应用户的语音请求的延时。

[0216] 基于同一发明构思,本公开一实施例提供一种音频播放装置400。参考图4,图4是本申请一实施例示出的一种应用于终端的音频播放装置的结构示意图。如图4所示,音频播放装置400包括:

[0217] 第一发送模块401,用于在检测到语音输入时,向服务器发送所述语音对应的应答请求;

[0218] 接收模块402,用于接收所述服务器发送的携带位置信息的应答音频片段,其中,所述位置信息用于表征所述应答音频片段在所述应答请求对应的应答音频中的位置;

[0219] 合成模块403,用于根据接收到的各个应答音频片段各自携带的位置信息,将位置相邻的应答音频片段合成为待播放应答音频包;

[0220] 第一播放模块404,用于播放每次合成的所述待播放应答音频包,直至完成所述应答音频的播放。

[0221] 相应地,所述第一播放模块404包括:

[0222] 第一播放子模块,用于当本次合成的所述待播放应答音频包包括的应答音频片段的数量大于预设数值时,播放本次合成的所述待播放应答音频包;

[0223] 第二播放子模块,用于当本次合成的所述待播放应答音频包包括的应答音频片段的数量小于所述预设数值,且当前时刻与上一次播放结束的结束时刻之间的时长达到第一预设时长时,播放本次合成的所述待播放应答音频包。

[0224] 可选地,所述音频播放装置400还包括:

[0225] 第二播放模块,用于在接收到的各个应答音频片段位置不相邻,且当前时刻与上

一次播放结束的结束时刻之间的时长达到第二预设时长时,根据位置信息,播放接收到的各个应答音频片段;

[0226] 清除模块,用于在播放接收到的各个应答音频片段的过程中,若接收到位于已经播放或者正在播放的应答音频片段位置之后的应答音频片段,丢弃该接收到的应答音频片段。

[0227] 可选地,所述音频播放装置400还包括:

[0228] 调节模块,用于在再次接收到应答音频片段的接收时刻与上一次播放结束的结束时刻之间的时长大于所述第二预设时长时,延长所述第一预设时长和/或增加所述预设数值。可选地,所述音频播放装置400还包括:

[0229] 请求模块,用于在接收到的各个应答音频片段位置不相邻,且未接收到的应答音频片段是目标音频片段时,向所述服务器请求所述目标音频片段,所述目标音频片段是所述应答音频中携带应答关键词的应答音频片段;

[0230] 第三播放模块,用于在接收到所述目标音频片段后,将所述目标音频片段以及与所述目标音频片段位置相邻的应答音频片段合成为待播放应答音频包,并播放该待播放应答音频包。可选地,所述合成模块403还包括:

[0231] 确定子模块,用于根据接收到的各个应答音频片段各自携带的位置信息,确定位置相邻的多个应答音频片段;

[0232] 合成子模块,用于对所述位置相邻的多个应答音频片段进行语义分析,在所述位置相邻的多个应答音频片段组成短句时,将所述位置相邻的多个应答音频片段合成为待播放应答音频包。

[0233] 可选地,所述音频播放装置400还包括:

[0234] 第二发送模块,用于向所述服务器发送长连接建立请求,以提示所述服务器所述终端准备接收应答音频片段;

[0235] 建立长连接模块,用于接收所述服务器发送的长连接建立响应,与所述服务器建立长连接;

[0236] 所述接收模块402,包括:

[0237] 接收子模块,用于通过长连接接收所述服务器发送的携带位置信息的应答音频片段。

[0238] 关于上述实施例中的音频播放装置,其中各个模块执行操作的具体方式已经在有关应用于终端的音频播放方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0239] 本公开一实施例提供一种终端,包括:

[0240] 处理器;

[0241] 用于存储所述处理器可执行指令的存储器;

[0242] 其中,所述处理器被配置为执行本公开上述应用于终端的音频播放方法所执行的操作。

[0243] 本公开另一实施例提供一种非临时性计算机可读存储介质,当所述存储介质中的指令由终端的处理器执行时,使得终端能够执行一种以实现本公开上述应用于终端的音频播放方法所执行的操作。

[0244] 基于同一发明构思,本公开一实施例提供一种音频播放装置500。参考图5,图5是

本申请一实施例示出的一种应用于服务器的音频播放装置的结构示意图。如图5所示,音频播放装置500包括:

[0245] 第一接收模块501,用于接收终端发送的应答请求;

[0246] 数据获取模块502,用于根据所述应答请求获得多个应答音频数据;

[0247] 第一发送模块503,用于每将所述多个应答音频数据中的至少一个应答音频数据合成为应答音频片段,向所述终端发送所述应答音频片段,直至完成所述多个应答音频数据的发送。

[0248] 可选地,所述音频播放装置500还包括:

[0249] 第二接收模块,用于接收所述终端发送的目标音频片段请求,所述目标音频片段是所述应答音频中携带应答关键词的应答音频片段;

[0250] 第二发送模块,用于向所述终端发送所述目标音频片段,以使所述终端将所述目标音频片段以及与所述目标音频片段位置相邻的应答音频片段合成为待播放应答音频包,并播放该待播放应答音频包。

[0251] 可选地,所述音频播放装置500还包括:

[0252] 第三接收模块,用于接收所述终端发送长连接建立请求,所述长连接建立请求表征所述终端准备接收应答音频片段;

[0253] 建立长连接模块,用于向所述终端发送的长连接建立响应,与所述终端建立长连接;

[0254] 所述第一发送模块503,包括:

[0255] 发送子模块,用于通过长连接向所述终端发送所述应答音频片段。

[0256] 关于上述实施例中的音频播放装置,其中各个模块执行操作的具体方式已经在有关应用于服务器的音频播放方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0257] 本公开一实施例提供一种服务器,包括:

[0258] 处理器;

[0259] 用于存储所述处理器可执行指令的存储器;

[0260] 其中,所述处理器被配置为执行本公开上述应用于服务器的音频播放方法所执行的操作。

[0261] 本公开另一实施例提供一种非临时性计算机可读存储介质,当所述存储介质中的指令由终端的处理器执行时,使得终端能够执行一种以实现本公开上述应用于服务器的音频播放方法所执行的操作。

[0262] 关于上述实施例中的装置,其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0263] 参照图6所示,图6是本申请一实施例示出的一种终端的结构框图,例如,终端600可以包括以下一个或多个组件:处理组件602,存储器604,电力组件606,多媒体组件608,音频组件610,输入/输出(I/O)的接口612,传感器组件614,以及通信组件616。存储器604中存储有非临时性计算机可读存储介质640,可选地,可读存储介质640可以是非临时性计算机可读存储介质,例如,所述非临时性计算机可读存储介质640可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0264] 处理组件602通常控制终端600的整体操作,诸如与显示,电话呼叫,数据通信,相

机操作和记录操作相关联的操作。处理组件602可以包括一个或多个处理器620来执行可读存储介质640的指令,以使终端600能完成上述XX方法的全部或部分步骤。此外,处理组件602可以包括一个或多个模块,便于处理组件602和其他组件之间的交互。例如,处理组件602可以包括多媒体模块,以方便多媒体组件608和处理组件602之间的交互。

[0265] 多媒体组件608被配置为存储各种类型的数据以支持在终端600的操作。这些数据的示例包括用于在终端600上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。多媒体组件608可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。

[0266] 电力组件606为终端600的各种组件提供电力。电力组件606可以包括电源管理系统,一个或多个电源,及其他与为终端600生成、管理和分配电力相关联的组件。

[0267] 多媒体组件608包括在所述终端600和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件608包括一个前置摄像头和/或后置摄像头。当终端600处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0268] 音频组件610被配置为输出和/或输入音频信号。例如,音频组件610包括一个麦克风(MIC),当终端600处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在多媒体组件608或经由通信组件616发送。在一些实施例中,音频组件610还包括一个扬声器,用于输出音频信号。

[0269] I/O接口612为处理组件602和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0270] 传感器组件614包括一个或多个传感器,用于为终端600提供各个方面的状态评估。例如,传感器组件614可以检测到终端600的打开/关闭状态,组件的相对定位,例如所述组件为终端600的显示器和小键盘,传感器组件614还可以检测终端600或终端600一个组件的位置改变,用户与终端600接触的存在或不存在,终端600方位或加速/减速和终端600的温度变化。传感器组件614可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件614还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件614还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0271] 通信组件616被配置为便于终端600和其他设备之间有线或无线方式的通信。终端600可以接入基于通信标准的无线网络,如WiFi,运营商网络(如2G、3G、4G或5G),或它们的组合。在一个示例性实施例中,通信组件616经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信组件616还包括近场通信(NFC)

模块,以促进短程通信。在示例性实施例中,终端600可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述应用于终端的音频播放方法所执行的操作。

[0272] 参照图7所示,图7是根据一示例性实施例示出的一种服务器的结构框图,服务器700可以包括处理器701、电源组件702、网络接口703、存储器704及输入输出接口705,其中,存储器704上配置有非临时性计算机可读存储介质7041,该可读存储介质7041中存储有用于执行上述建立关联关系的方法的指令,存储器704可以与处理器701进行通信,当所述存储介质704中的指令由处理器701执行时,使得服务器700能够执行上述应用于服务器的音频播放方法所执行的操作。其中,网络接口703可以将服务器700连接到网络中,电源组件702可以执行服务器700的电源管理。

[0273] 可选地,可读存储介质740可以是非临时性计算机可读存储介质,例如,所述非临时性计算机可读存储介质740可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0274] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其它实施方案。本申请旨在涵盖本公开的任何变型、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公开的一般性原理并包括本公开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,本公开的真正范围和精神由下面的权利要求指出。

[0275] 应当理解的是,本公开并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限制。

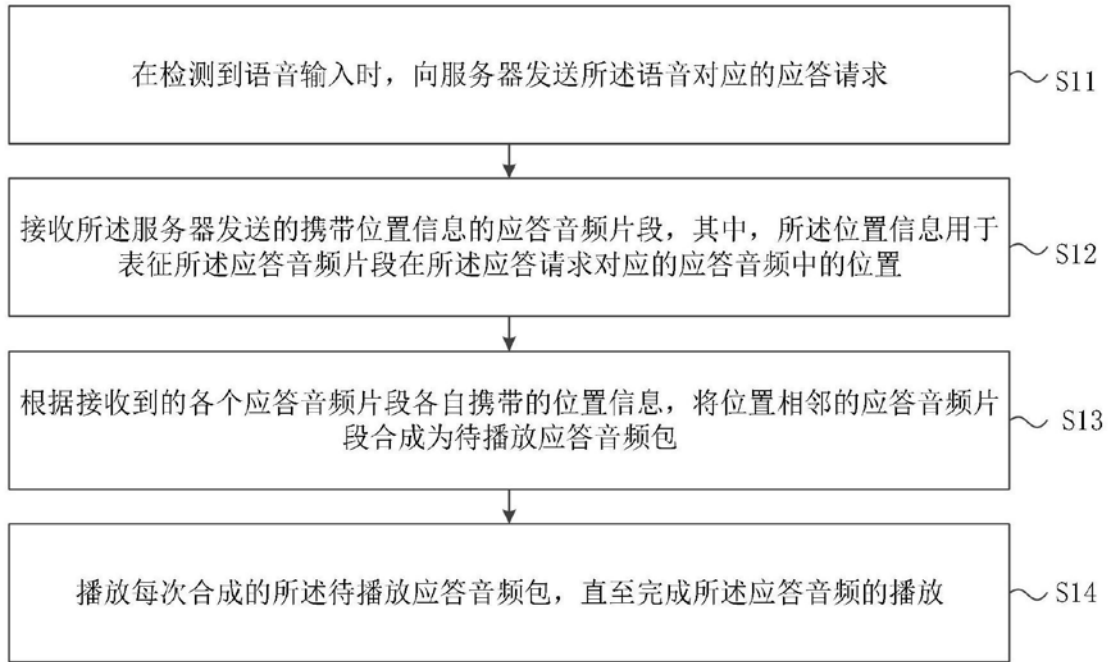


图1

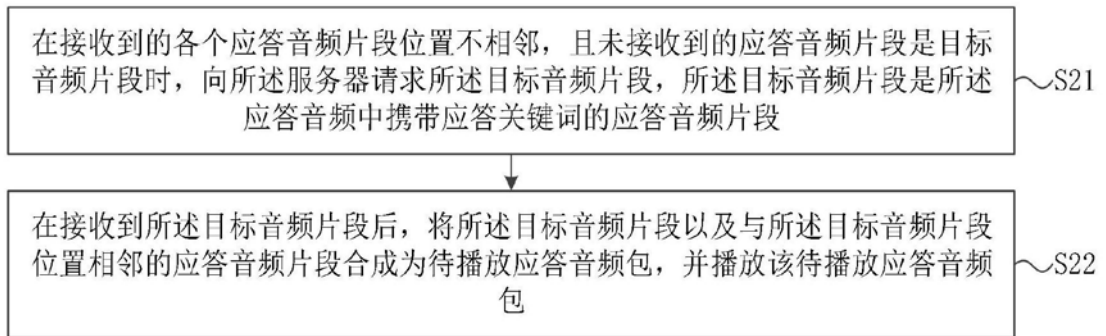


图2

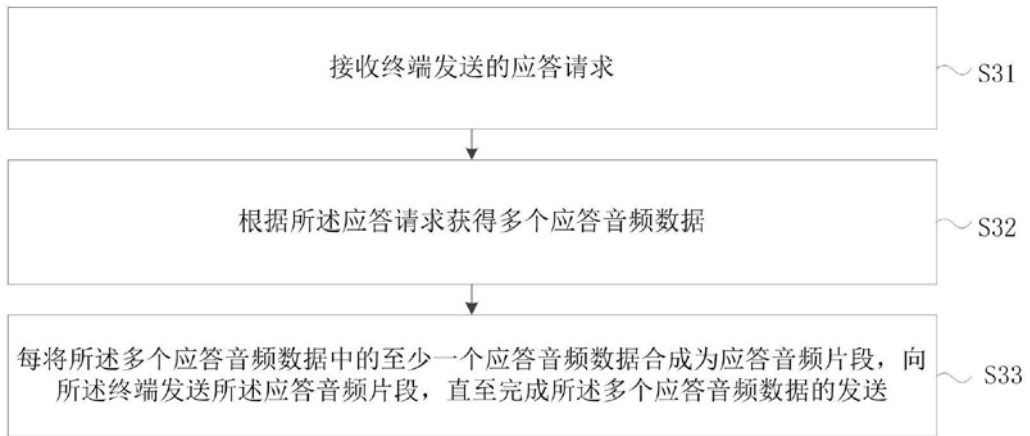


图3



图4

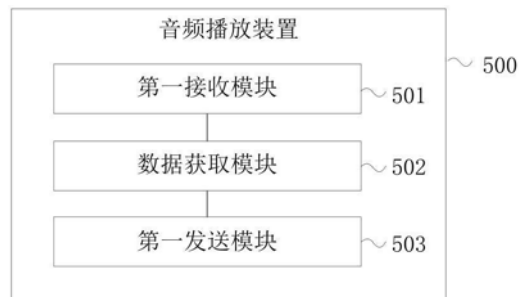


图5



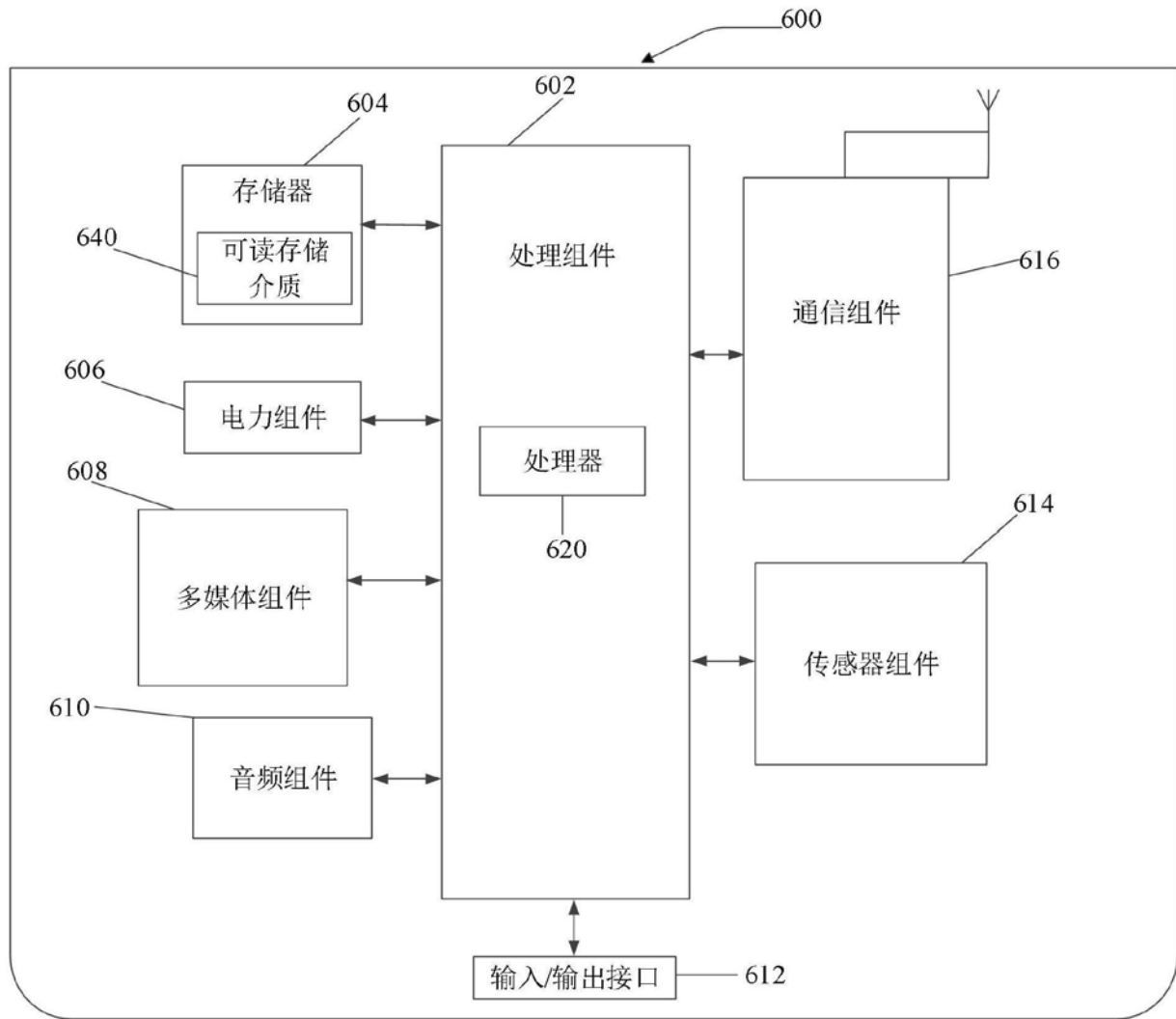


图6

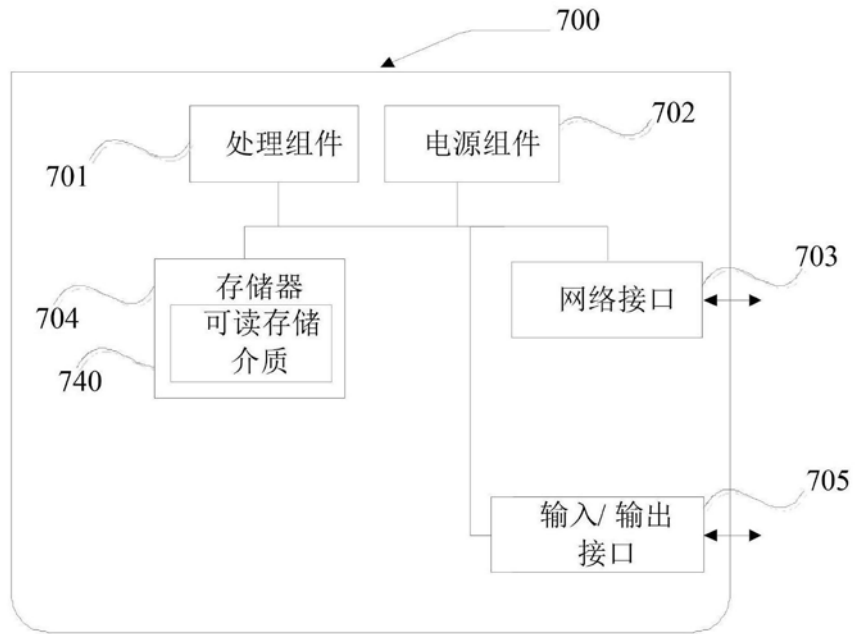


图7