



(12) 发明专利

(10) 授权公告号 CN 110610699 B

(45) 授权公告日 2023. 03. 24

(21) 申请号 201910829645.9

G10L 15/26 (2006.01)

(22) 申请日 2019.09.03

G10L 15/30 (2013.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 110610699 A

(56) 对比文件

CN 107591151 A, 2018.01.16

CN 110176235 A, 2019.08.27

(43) 申请公布日 2019.12.24

CN 108307268 A, 2018.07.20

(73) 专利权人 北京达佳互联信息技术有限公司

CN 108010526 A, 2018.05.08

地址 100085 北京市海淀区上地西路6号1

CN 109767774 A, 2019.05.17

幢1层101D1-7

CN 108074561 A, 2018.05.25

CN 109215653 A, 2019.01.15

(72) 发明人 张晓波 张奇 周舟

审查员 李春雨

(74) 专利代理机构 北京柏杉松知识产权代理事

务所(普通合伙) 11413

专利代理师 项京 高莺然

(51) Int. Cl.

G10L 15/08 (2006.01)

G10L 15/22 (2006.01)

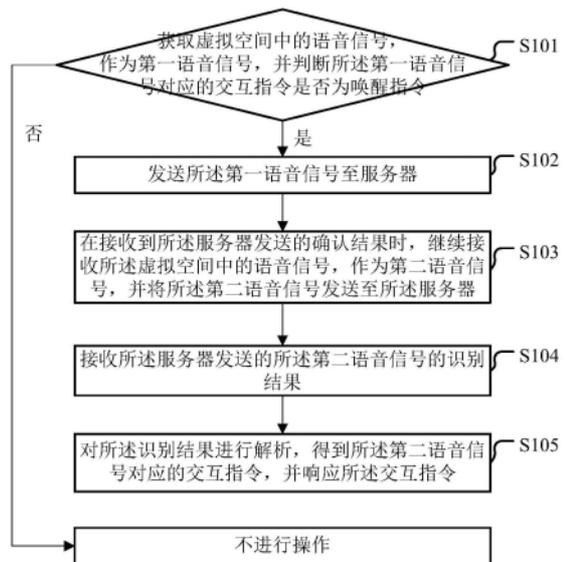
权利要求书5页 说明书19页 附图8页

(54) 发明名称

语音信号处理方法、装置、终端、服务器及存储介质

(57) 摘要

本公开实施例提供了一种语音信号处理方法、装置、终端、服务器及存储介质,所述方法应用于终端,包括:获取虚拟空间中的语音信号,作为第一语音信号,并判断第一语音信号对应的交互指令是否为唤醒指令;若为是,发送第一语音信号至服务器;在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,作为第二语音信号,并将第二语音信号发送至服务器;接收服务器发送的第二语音信号的识别结果;对识别结果进行解析,得到第二语音信号对应的交互指令,并响应交互指令。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。



1. 一种语音信号处理方法,其特征在于,应用于安装有直播应用程序的终端,所述方法包括:

获取虚拟空间中的语音信号,作为第一语音信号,并判断所述第一语音信号对应的交互指令是否为唤醒指令,其中,所述虚拟空间用于进行实时视频行为;所述虚拟空间为直播间;所述唤醒指令为唤醒语音操作终端功能的语音指令;

若为是,发送所述第一语音信号至服务器;

在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号,并将所述第二语音信号发送至所述服务器,其中,所述确认结果用于指示所述第一语音信号对应的交互指令为唤醒指令;

接收所述服务器发送的所述第二语音信号的识别结果;

对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令,以实现需要执行的直播间中的各功能;

所述对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令的步骤,包括:

对所述识别结果进行解析,得到操作命令及操作对象;

对所述操作对象执行所述操作命令对应的动作;

所述操作命令为需要执行的动作对应的命令;所述操作对象为需要执行的动作所针对的目标。

2. 如权利要求1所述的方法,其特征在于,发送目标语音信号至服务器的方式,包括:

按照预设时长将所述目标语音信号封装成多个数据包,其中,所述目标语音信号为所述第一语音信号或所述第二语音信号;

为每个数据包添加顺序标识;

通过预定通道发送添加顺序标识后的多个数据包至所述服务器。

3. 如权利要求1所述的方法,其特征在于,所述在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号的步骤,包括:

在接收到所述服务器发送的确认结果时,输出唤醒成功提示信息;

获取用户在所述虚拟空间中基于所述唤醒成功提示信息输入的语音信号,作为第二语音信号。

4. 如权利要求1所述的方法,其特征在于,所述识别结果包括操作命令标识;

所述对所述识别结果进行解析,得到操作命令及操作对象的步骤,包括:

对所述识别结果进行解析,得到操作命令标识及操作对象;

根据预设的操作命令标识与操作命令的对应关系,确定解析得到的操作命令标识对应的操作命令,其中,所述对应关系包括的操作命令为所述虚拟空间中各功能对应操作命令。

5. 一种语音信号处理方法,其特征在于,应用于服务器,所述方法包括:

接收终端发送的第一语音信号;所述第一语音信号为终端获取虚拟空间中的语音信号;所述虚拟空间为直播间;所述终端安装有直播应用程序;

对所述第一语音信号进行语音识别,确定所述第一语音信号是否包括预设唤醒词;

如果包括,确定所述第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端;所述唤醒指令为唤醒语音操作终端功能的语音指令;

接收终端发送的第二语音信号；

对所述第二语音信号进行语音识别，得到识别结果，并发送所述识别结果至所述终端，以使得所述终端对所述识别结果进行解析，得到所述第二语音信号对应的交互指令，并响应所述交互指令，以实现需要执行的直播间中的各功能；

所述对所述识别结果进行解析，得到所述第二语音信号对应的交互指令，并响应所述交互指令的步骤，包括：

对所述识别结果进行解析，得到操作命令及操作对象；

对所述操作对象执行所述操作命令对应的动作；

所述操作命令为需要执行的动作对应的命令；所述操作对象为需要执行的动作所针对的目标。

6. 如权利要求5所述的方法，其特征在于，接收所述终端发送的目标语音信号的方式，包括：

通过预定通道接收所述终端发送的多个数据包，其中，每个数据包携带顺序标识；

根据每个数据包携带的顺序标识所标识的顺序，将所述多个数据包进行拼接，得到所述目标语音信号，其中，所述目标语音信号为所述第一语音信号或所述第二语音信号。

7. 如权利要求5或6所述的方法，其特征在于，所述对所述第二语音信号进行语音识别，得到识别结果的步骤，包括：

对所述第二语音信号进行语音识别，确定所述第二语音信号的语义信息；

根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象，作为识别结果。

8. 如权利要求7所述的方法，其特征在于，所述根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象，作为识别结果的步骤，包括：

根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象；

根据预设的操作命令标识与操作命令的对应关系，确定所确定的操作命令对应的操作命令标识，并将所述操作命令标识及所述操作对象作为识别结果，其中，所述对应关系包括的操作命令为所述终端的虚拟空间中各功能对应操作命令，所述虚拟空间用于进行实时视频行为。

9. 一种语音信号处理装置，其特征在于，应用于安装有直播应用程序的终端，所述方法包括：

第一语音信号获取模块，被配置为执行获取虚拟空间中的语音信号，作为第一语音信号，并判断所述第一语音信号对应的交互指令是否为唤醒指令，其中，所述虚拟空间用于进行实时视频行为；所述虚拟空间为直播间；所述唤醒指令为唤醒语音操作终端功能的语音指令；

第一语音信号发送模块，被配置为执行若所述第一语音信号对应的交互指令为唤醒指令，并通过发送模块发送所述第一语音信号至服务器；

第二语音信号获取模块，被配置为执行在接收到所述服务器发送的确认结果时，继续接收所述虚拟空间中的语音信号，作为第二语音信号，并通过发送模块将所述第二语音信号发送至所述服务器，其中，所述确认结果用于指示所述第一语音信号对应的交互指令为唤醒指令；

识别结果接收模块,被配置为执行接收所述服务器发送的所述第二语音信号的识别结果;

识别结果解析模块,被配置为执行对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令,以实现需要执行的直播间中的各功能;

所述识别结果解析模块包括:

识别结果解析单元,被配置为执行对所述识别结果进行解析,得到操作命令及操作对象;

动作执行单元,被配置为执行对所述操作对象执行所述操作命令对应的动作;

所述操作命令为需要执行的动作对应的命令;所述操作对象为需要执行的动作所针对的目标。

10.如权利要求9所述的装置,其特征在于,所述发送模块包括:

数据包封装单元,被配置为执行按照预设时长将所述目标语音信号封装成多个数据包,其中,所述目标语音信号为所述第一语音信号或所述第二语音信号;

顺序标识添加单元,被配置为执行为每个数据包添加顺序标识;

数据包发送单元,被配置为执行通过预定通道发送添加顺序标识后的多个数据包至所述服务器。

11.如权利要求9所述的装置,其特征在于,所述第二语音信号获取模块包括:

提示信息输出单元,被配置为执行在接收到所述服务器发送的确认结果时,输出唤醒成功提示信息;

第二语音信号获取单元,被配置为执行获取用户在所述虚拟空间中基于所述唤醒成功提示信息输入的语音信号,作为第二语音信号。

12.如权利要求11所述的装置,其特征在于,所述识别结果包括操作命令标识;

所述识别结果解析单元包括:

识别结果解析子单元,被配置为执行对所述识别结果进行解析,得到操作命令标识及操作对象;

操作命令确定子单元,被配置为执行根据预设的操作命令标识与操作命令的对应关系,确定解析得到的操作命令标识对应的操作命令,其中,所述对应关系包括的操作命令为所述虚拟空间中各功能对应操作命令。

13.一种语音信号处理装置,其特征在于,应用于服务器,所述装置包括:

第一语音信号接收模块,被配置为执行接收终端发送的第一语音信号;所述第一语音信号为终端获取虚拟空间中的语音信号;所述虚拟空间为直播间;所述终端安装有直播应用程序;

第一语音信号识别模块,被配置为执行对所述第一语音信号进行语音识别,确定所述第一语音信号是否包括预设唤醒词;

确认结果发送模块,被配置为执行如果所述第一语音信号包括预设唤醒词,确定所述第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端;所述唤醒指令为唤醒语音操作终端功能的语音指令;

第二语音信号接收模块,被配置为执行接收终端发送的第二语音信号;

第二语音信号识别模块,被配置为执行对所述第二语音信号进行语音识别,得到识别

结果,并发送所述识别结果至所述终端,以使得所述终端对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令,以实现需要执行的直播间中的各功能;

所述对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令的步骤,包括:

对所述识别结果进行解析,得到操作命令及操作对象;

对所述操作对象执行所述操作命令对应的动作;

所述操作命令为需要执行的动作对应的命令;所述操作对象为需要执行的动作所针对的目标。

14.如权利要求13所述的装置,其特征在于,所述第一语音信号接收模块和第二语音信号接收模块包括:

数据包接收单元,被配置为执行通过预定通道接收所述终端发送的多个数据包,其中,每个数据包携带顺序标识;

数据包拼接单元,被配置为执行根据每个数据包携带的顺序标识所标识的顺序,将所述多个数据包进行拼接,得到所述目标语音信号,其中,所述目标语音信号为所述第一语音信号或所述第二语音信号。

15.如权利要求13或14所述的装置,其特征在于,所述第二语音信号识别模块包括:

语义信息确定单元,被配置为执行对所述第二语音信号进行语音识别,确定所述第二语音信号的语义信息;

识别结果确定单元,被配置为执行根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象,作为识别结果。

16.如权利要求15所述的装置,其特征在于,所述识别结果确定单元包括:

第一识别子单元,被配置为执行根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象;

第二识别子单元,被配置为根据预设的操作命令标识与操作命令的对应关系,确定所确定的操作命令对应的操作命令标识,并将所述操作命令标识及所述操作对象作为识别结果,其中,所述对应关系包括的操作命令为所述终端的虚拟空间中各功能对应操作命令,所述虚拟空间用于进行实时视频行为。

17.一种终端,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所述指令,以实现如权利要求1至4中任一项所述的语音信号处理方法。

18.一种服务器,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所述指令,以实现如权利要求5至8中任一项所述的语音信号处理方法。

19.一种存储介质,其特征在于,当所述存储介质中的指令由终端的处理器执行时,使

得终端能够执行权利要求1至4中任一项所述的语音信号处理方法。

20.一种存储介质,其特征在于,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权利要求5至8中任一项所述的语音信号处理方法。

## 语音信号处理方法、装置、终端、服务器及存储介质

### 技术领域

[0001] 本公开涉及语音信号处理技术领域,特别是涉及一种语音信号处理方法、装置、终端、服务器及存储介质。

### 背景技术

[0002] 目前网络直播已成为人们生活中非常常见的网络行为,几乎每个人都接触过网络直播,利用直播应用程序,通过网络进行直播或者观看直播的人也越来越多。

[0003] 用户在进行网络直播时,由于一般需要频繁使用直播应用程序的各个功能,例如,连麦、播放音乐、送礼物等,并且用户经常会进行跳舞、演唱等直播行为,所以在网络直播过程中,用户一般需要频繁手动操作直播应用程序的各个功能,这样的手动操作终端会给用户带来非常多的不便。

### 发明内容

[0004] 为克服相关技术中存在的问题,本公开实施例提供一种语音信号处理方法、装置、终端、服务器及存储介质。具体技术方案如下:

[0005] 根据本公开实施例的第一方面,提供一种语音信号处理方法,应用于终端,所述方法包括:

[0006] 获取虚拟空间中的语音信号,作为第一语音信号,并判断所述第一语音信号对应的交互指令是否为唤醒指令,其中,所述虚拟空间用于进行实时视频行为;

[0007] 若为是,发送所述第一语音信号至服务器;

[0008] 在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号,并将所述第二语音信号发送至所述服务器,其中,所述确认结果用于指示所述第一语音信号对应的交互指令为唤醒指令;

[0009] 接收所述服务器发送的所述第二语音信号的识别结果;

[0010] 对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令。

[0011] 作为一种实施方式,发送目标语音信号至服务器的方式,包括:

[0012] 按照预设时长将所述目标语音信号封装成多个数据包,其中,所述目标语音信号为所述第一语音信号或所述第二语音信号;

[0013] 为每个数据包添加顺序标识;

[0014] 通过预定通道发送添加顺序标识后的多个数据包至所述服务器。

[0015] 作为一种实施方式,所述在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号的步骤,包括:

[0016] 在接收到所述服务器发送的确认结果时,输出唤醒成功提示信息;

[0017] 获取用户在所述虚拟空间中基于所述唤醒成功提示信息输入的语音信号,作为第二语音信号。

[0018] 作为一种实施方式,所述对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令的步骤,包括:

[0019] 对所述识别结果进行解析,得到操作命令及操作对象;

[0020] 对所述操作对象执行所述操作命令对应的动作。

[0021] 作为一种实施方式,所述识别结果包括操作命令标识;

[0022] 所述对所述识别结果进行解析,得到操作命令及操作对象的步骤,包括:

[0023] 对所述识别结果进行解析,得到操作命令标识及操作对象;

[0024] 根据预设的操作命令标识与操作命令的对应关系,确定解析得到的操作命令标识对应的操作命令,其中,所述对应关系包括的操作命令为所述虚拟空间中各功能对应操作命令。

[0025] 根据本公开实施例的第二方面,提供一种语音信号处理方法,应用于服务器,所述方法包括:

[0026] 接收终端发送的第一语音信号;

[0027] 对所述第一语音信号进行语音识别,确定所述第一语音信号是否包括预设唤醒词;

[0028] 如果包括,确定所述第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端;

[0029] 接收终端发送的第二语音信号;

[0030] 对所述第二语音信号进行语音识别,得到识别结果,并发送所述识别结果至所述终端。

[0031] 作为一种实施方式,接收所述终端发送的目标语音信号的方式,包括:

[0032] 通过预定通道接收所述终端发送的多个数据包,其中,每个数据包携带顺序标识;

[0033] 根据每个数据包携带的顺序标识所标识的顺序,将所述多个数据包进行拼接,得到所述目标语音信号,其中,所述目标语音信号为所述第一语音信号或所述第二语音信号。

[0034] 作为一种实施方式,所述对所述第二语音信号进行语音识别,得到识别结果的步骤,包括:

[0035] 对所述第二语音信号进行语音识别,确定所述第二语音信号的语义信息;

[0036] 根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象,作为识别结果。

[0037] 作为一种实施方式,所述根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象,作为识别结果的步骤,包括:

[0038] 根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象;

[0039] 根据预设的操作命令标识与操作命令的对应关系,确定所确定的操作命令对应的操作命令标识,并将所述操作命令标识及所述操作对象作为识别结果,其中,所述对应关系包括的操作命令为所述终端的虚拟空间中各功能对应操作命令,所述虚拟空间用于进行实时视频行为。

[0040] 根据本公开实施例的第三方面,提供一种语音信号处理装置,应用于终端,所述方法包括:

[0041] 第一语音信号获取模块,被配置为执行获取虚拟空间中的语音信号,作为第一语

音信号,并判断所述第一语音信号对应的交互指令是否为唤醒指令,其中,所述虚拟空间用于进行实时视频行为;

[0042] 第一语音信号发送模块,被配置为执行若所述第一语音信号对应的交互指令为唤醒指令,并通过发送模块发送所述第一语音信号至服务器;

[0043] 第二语音信号获取模块,被配置为执行在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号,并通过发送模块将所述第二语音信号发送至所述服务器,其中,所述确认结果用于指示所述第一语音信号对应的交互指令为唤醒指令;

[0044] 识别结果接收模块,被配置为执行接收所述服务器发送的所述第二语音信号的识别结果;

[0045] 识别结果解析模块,被配置为执行对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令。

[0046] 作为一种实施方式,所述发送模块包括:

[0047] 数据包封装单元,被配置为执行按照预设时长将所述目标语音信号封装成多个数据包,其中,所述目标语音信号为所述第一语音信号或所述第二语音信号;

[0048] 顺序标识添加单元,被配置为执行为每个数据包添加顺序标识;

[0049] 数据包发送单元,被配置为执行通过预定通道发送添加顺序标识后的多个数据包至所述服务器。

[0050] 作为一种实施方式,所述第二语音信号获取模块包括:

[0051] 提示信息输出单元,被配置为执行在接收到所述服务器发送的确认结果时,输出唤醒成功提示信息;

[0052] 第二语音信号获取单元,被配置为执行获取用户在所述虚拟空间中基于所述唤醒成功提示信息输入的语音信号,作为第二语音信号。

[0053] 作为一种实施方式,所述识别结果解析模块包括:

[0054] 识别结果解析单元,被配置为执行对所述识别结果进行解析,得到操作命令及操作对象;

[0055] 动作执行单元,被配置为执行对所述操作对象执行所述操作命令对应的动作。

[0056] 作为一种实施方式,所述识别结果包括操作命令标识;

[0057] 所述识别结果解析单元包括:

[0058] 识别结果解析子单元,被配置为执行对所述识别结果进行解析,得到操作命令标识及操作对象;

[0059] 操作命令确定子单元,被配置为执行根据预设的操作命令标识与操作命令的对应关系,确定解析得到的操作命令标识对应的操作命令,其中,所述对应关系包括的操作命令为所述虚拟空间中各功能对应操作命令。

[0060] 根据本公开实施例的第四方面,一种语音信号处理装置,应用于服务器,所述装置包括:

[0061] 第一语音信号接收模块,被配置为执行接收终端发送的第一语音信号;

[0062] 第一语音信号识别模块,被配置为执行对所述第一语音信号进行语音识别,确定所述第一语音信号是否包括预设唤醒词;

[0063] 确认结果发送模块,被配置为执行如果所述第一语音信号包括预设唤醒词,确定所述第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端;

[0064] 第二语音信号接收模块,被配置为执行接收终端发送的第二语音信号;

[0065] 第二语音信号识别模块,被配置为执行对所述第二语音信号进行语音识别,得到识别结果,并发送所述识别结果至所述终端。

[0066] 作为一种实施方式,所述第一语音信号接收模块和第二语音信号接收模块包括:

[0067] 数据包接收单元,被配置为执行通过预定通道接收所述终端发送的多个数据包,其中,每个数据包携带顺序标识;

[0068] 数据包拼接单元,被配置为执行根据每个数据包携带的顺序标识所标识的顺序,将所述多个数据包进行拼接,得到所述目标语音信号,其中,所述目标语音信号为所述第一语音信号或所述第二语音信号。

[0069] 作为一种实施方式,所述第二语音信号识别模块包括:

[0070] 语义信息确定单元,被配置为执行对所述第二语音信号进行语音识别,确定所述第二语音信号的语义信息;

[0071] 识别结果确定单元,被配置为执行根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象,作为识别结果。

[0072] 作为一种实施方式,所述识别结果确定单元包括:

[0073] 第一识别子单元,被配置为执行根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象;

[0074] 第二识别子单元,被配置为根据预设的操作命令标识与操作命令的对应关系,确定所确定的操作命令对应的操作命令标识,并将所述操作命令标识及所述操作对象作为识别结果,其中,所述对应关系包括的操作命令为所述终端的虚拟空间中各功能对应操作命令,所述虚拟空间用于进行实时视频行为。

[0075] 根据本公开实施例的第五方面,提供一种终端,包括:

[0076] 处理器;

[0077] 用于存储所述处理器可执行指令的存储器;

[0078] 其中,所述处理器被配置为执行所述指令,以实现上述第一方面任一所述的语音信号处理方法。

[0079] 根据本公开实施例的第六方面,提供一种服务器,包括:

[0080] 处理器;

[0081] 用于存储所述处理器可执行指令的存储器;

[0082] 其中,所述处理器被配置为执行所述指令,以实现上述第二方面任一所述的语音信号处理方法。

[0083] 根据本公开实施例的第七方面,提供一种存储介质,当所述存储介质中的指令由终端的处理器执行时,使得终端能够执行上述第一方面任一所述的语音信号处理方法。

[0084] 根据本公开实施例的第八方面,提供一种存储介质,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上述第二方面任一所述的语音信号处理方法。

[0085] 本公开实施例所提供的方案中,终端可以获取虚拟空间中的语音信号,作为第一

语音信号,并判断第一语音信号对应的交互指令是否为唤醒指令,若为是,发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,作为第二语音信号,并将第二语音信号发送至服务器,其中,确认结果用于指示第一语音信号对应的交互指令为唤醒指令,接收服务器发送的第二语音信号的识别结果,对识别结果进行解析,得到第二语音信号对应的交互指令,并响应交互指令。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

### 附图说明

[0086] 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,并不构成对本公开的不当限定。

[0087] 图1是根据一示例性实施例示出的第一种语音信号处理方法的流程图;

[0088] 图2是根据一示例性实施例示出的直播间界面的一种示意图;

[0089] 图3是根据一示例性实施例示出的目标语音信号发送方式的一种流程图;

[0090] 图4是根据一示例性实施例示出的直播间界面的另一种示意图;

[0091] 图5是根据一示例性实施例示出的第二种语音信号处理方法的流程图;

[0092] 图6是根据一示例性实施例示出的目标语音信号接收方式的一种流程图;

[0093] 图7是根据一示例性实施例示出的语音信号处理方法的一种信令交互图;

[0094] 图8是根据一示例性实施例示出的第一种语音信号处理装置的结构框图;

[0095] 图9是根据一示例性实施例示出的第二种语音信号处理装置的结构框图;

[0096] 图10是根据一示例性实施例示出的一种终端的结构框图;

[0097] 图11是根据一示例性实施例示出的一种终端的具体结构框图;

[0098] 图12是根据一示例性实施例示出的一种服务器的结构框图;

[0099] 图13是根据一示例性实施例示出的一种服务器的具体结构框图。

### 具体实施方式

[0100] 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。

[0101] 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。

[0102] 为了提高在网络直播过程中用户操作终端的便利性,本公开实施例提供了一种语音信号处理方法、装置、终端、服务器及计算机可读存储介质。

[0103] 本公开实施例所提供的第一种语音信号处理方法可以应用于安装有直播应用程序的终端。下面首先对本公开实施例所提供的第一种语音信号处理方法进行介绍。

[0104] 如图1所示,一种语音信号处理方法,应用于终端,所述方法包括:

[0105] 在步骤S101中,获取虚拟空间中的语音信号,作为第一语音信号,并判断所述第一语音信号对应的交互指令是否为唤醒指令;若为是,执行步骤S102;若为否,不进行操作;

[0106] 其中,所述虚拟空间用于进行实时视频行为。

[0107] 在步骤S102中,发送所述第一语音信号至服务器;

[0108] 在步骤S103中,在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号,并将所述第二语音信号发送至所述服务器;

[0109] 其中,所述确认结果用于指示所述第一语音信号对应的交互指令为唤醒指令。

[0110] 在步骤S104中,接收所述服务器发送的所述第二语音信号的识别结果;

[0111] 在步骤S105中,对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令。

[0112] 可见,本公开实施例所提供的方案中,终端可以获取虚拟空间中的语音信号,作为第一语音信号,并判断第一语音信号对应的交互指令是否为唤醒指令,若为是,发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,作为第二语音信号,并将第二语音信号发送至服务器,其中,确认结果用于指示第一语音信号对应的交互指令为唤醒指令,接收服务器发送的第二语音信号的识别结果,对识别结果进行解析,得到第二语音信号对应的交互指令,并响应交互指令。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0113] 用户在利用终端安装的直播应用程序进行直播或者观看直播时,可以发出各种语音信号,终端便可以获取虚拟空间中的语音信号,作为第一语音信号。其中,虚拟空间用于进行实时视频行为,也就是直播行为,虚拟空间可以为直播间。终端可以通过自身具有的麦克风等器件采集上述第一语音信号。

[0114] 由于用户发出的语音信号时并不一定是需要对终端进行操作,因此为了确定用户是否需要对终端进行操作,在上述步骤S101中,终端获取虚拟空间中的第一语音信号后,可以判断第一语音信号对应的交互指令是否为唤醒指令。其中,唤醒指令为唤醒语音操作终端功能的语音指令。

[0115] 终端在接收到唤醒指令之前,语音操作终端功能可以处于休眠状态,此时终端可以持续获取用户发出的语音信号,但是并不需要进行响应,当终端接收到唤醒指令时,说明用户想要启动语音操作终端功能,语音操作终端功能便进入工作状态。

[0116] 为了告知用户语音操作终端功能的状态,方便用户对其进行操作,终端显示的直播间界面中可以显示预设图标,用于标识语音操作终端功能的状态。例如,直播间界面可以如图2所示,预设图标210可以标识语音操作终端功能的状态,例如,当预设图标中的卡通形象为睡眠状态时,可以表示语音操作终端功能处于休眠状态;当预设图标中的卡通形象为舒醒状态时,可以表示语音操作终端功能处于工作状态。

[0117] 在一种实施方式中,终端可以对第一语音信号进行语音识别,确定其语义信息,若该语义信息包括预设的唤醒词,那么便可以确定第一语音信号对应的交互指令为唤醒指令。其中,唤醒词可以为预设的用于唤醒语音操作终端功能的词语,例如,可以为“小快,小快”、“嗨,小快”、“小快醒醒”等。

[0118] 由于对于直播场景来说,环境比较嘈杂,终端对语音信号的识别可能不够准确,因此为了准确确定第一语音信号对应的交互指令是否为唤醒指令,可以采用DWS (Dual Wakeup System,双重唤醒系统)确定第一语音信号对应的交互指令是否为唤醒指令。DWS是一个采用本地识别模型及远端识别模型的机制。具体来说,本地识别模型用于识别用户发出的语音信号中的唤醒词,进而将该语音信号发送到远端服务器,服务器可以设置一个更精确的唤醒词识别模型,用于唤醒词的二次识别确认,并将确认结果返回给终端。

[0119] 那么,终端确定第一语音信号对应的交互指令为唤醒指令后,便可以将第一语音信号发送至服务器。服务器接收到第一语音信号后,可以对第一语音信号进行语音识别,确定其语义信息,若该语义信息包括预设的唤醒词,那么便可以确定第一语音信号对应的交互指令为唤醒指令,此时服务器可以发送确认结果至终端。其中,该确认结果即用于指示第一语音信号对应的交互指令为唤醒指令。

[0120] 进而,在上述步骤S103中,终端在接收到服务器发送的确认结果时,说明用户想要通过语音操作终端,那么此时终端便可以控制语音操作终端功能进入工作状态,以继续接收虚拟空间中的语音信号,并将该语音信号作为第二语音信号发送至服务器。

[0121] 服务器接收到终端发送的第二语音信号后,便可以对第二语音信号进行语音识别,确定其语义信息,进而根据其语义信息确定识别结果,并将该识别结果发送至终端。那么在上述步骤S104中,终端便可以接收到服务器发送的识别结果。

[0122] 进而,终端便可以对该识别结果进行解析,得到第二语音信号对应的交互指令,并响应该交互指令,即执行步骤S105。终端接收到上述识别结果后,为了确定用户需要进行的操作,可以对识别结果进行解析,得到第二语音信号对应的交互指令。交互指令即为指示终端做出相应动作的指令,例如,可以播放音乐、打开美颜、打开连麦等。进而,终端便可以响应该交互指令,以实现用户需要执行的直播间中的各功能。

[0123] 作为本公开实施例的一种实施方式,如图3所示,终端发送目标语音信号至服务器的方式,可以包括:

[0124] S301,按照预设时长将所述目标语音信号封装成多个数据包;

[0125] 其中,目标语音信号为上述第一语音信号或上述第二语音信号,也就是说,终端向服务器发送语音信号均可以采用该发送方式。

[0126] 终端可以对目标语音信号进行预处理,例如,对语音信号进行编码等,在此不做具体限定。接下来,终端可以将目标语音信号按照预设时长封装成多个数据包,其中,预设时长可以根据网络状况、一般语音信号的长度等因素确定,例如,可以为100毫秒、120毫秒、150毫秒等,在此不做具体限定。

[0127] S302,为每个数据包添加顺序标识;

[0128] 为了方便服务器在接收到数据包后,可以顺利将多个数据包拼接还原为目标语音信号,终端可以为每个数据包添加顺序标识。其中,顺序标识即为能够唯一标识数据包的顺序的标识。

[0129] 在一种实施方式中,终端将目标语音信号封装为多个数据包后,可以按照数据包对应于目标语音信号中的前后顺序,为各个数据包分别添加顺序标识 $1, 2, \dots, N$ ,其中, $N$ 即为数据包的数量。

[0130] S303,通过预定通道发送添加顺序标识后的多个数据包至所述服务器。

[0131] 为每个数据包添加顺序标识后,终端便可以将添加顺序标识后的多个数据包通过预定通道发送至服务器,其中,该预定通道为终端与服务器之间发送语音信号及语音信号识别结果的专用通道。

[0132] 在一种实施方式中,上述预定通道可以为SAS(Socket Audio Service,长链接语音服务)提供的专用Socket通道,这样,可以使终端与服务器之间的数据传输更加快速和稳定。

[0133] 可见,在本实施例中,终端可以按照预设时长将目标语音信号封装成多个数据包,然后为每个数据包添加顺序标识,进而通过预定通道发送添加顺序标识后的多个数据包至服务器,这样,由于终端将目标语音信号封装成多个数据包,可以陆续将数据包发送至服务器,而不是等待目标语音信号结束后再全部发送至服务器,这样,服务器可以在接收到数据包后便开始对目标语音信号进行处理,可以提高响应速度,缩短用户等待实长,可以保证终端与服务器之间的数据传输更加快速和稳定,提高响应速度及用户体验。

[0134] 作为本公开实施例的一种实施方式,上述在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号的步骤,可以包括:

[0135] 在接收到所述服务器发送的确认结果时,输出唤醒成功提示信息;获取用户在所述虚拟空间中基于所述唤醒成功提示信息输入的语音信号,作为第二语音信号。

[0136] 终端在接收到服务器发送的确认结果时,说明此时用户想要通过语音对终端进行操作,为了告知用户此时可以通过语音控制直播间实现直播间的各项功能,终端可以输出唤醒成功提示信息。

[0137] 例如,如果直播间界面如图2所示,那么终端可以改变开通形象的状态,举例来说,卡通形象可以眨眼、微笑、跳动等,以表示语音控制直播间功能已经被唤醒。

[0138] 又例如,终端可以输出文字提示信息,例如,“语音助手已开启”、“请您发出语音指令”、“嗨,您可以通过语音控制直播间的各功能了”等。

[0139] 用户在获知语音控制直播间功能开启后,便可以继续发出语音信号,该语音信号便为第二语音信号,终端获取该第二语音信号后,便可以将其发送至服务器,以对第二语音信号进行识别,确定用户需要执行的操作。

[0140] 可见,在本实施例中,终端在接收到服务器发送的确认结果时,可以输出唤醒成功提示信息,进而,获取用户在虚拟空间中基于唤醒成功提示信息输入的语音信号,作为第二语音信号。这样,可以提示用户通过语音控制直播间,以实现直播间的各功能,用户体验更佳。

[0141] 作为本公开实施例的一种实施方式,上述对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令的步骤,可以包括:

[0142] 对所述识别结果进行解析,得到操作命令及操作对象;对所述操作对象执行所述操作命令对应的动作。

[0143] 终端接收到服务器发送的识别结果后,为了确定用户需要进行的操作,可以对识

别结果进行解析,得到操作命令及操作对象。其中,操作命令即为需要执行的动作对应的命令,操作对象即为需要执行的动作所针对的目标。这样,终端便可以确定用户需要进行的操作。

[0144] 接下来,终端便可以对该操作对象执行操作命令对应的动作。例如,操作命令为:播放歌曲,操作对象为:歌曲A。那么,终端便可以执行播放歌曲A的操作。

[0145] 在一种实施方式中,终端在确定上述操作命令及操作对象后,可以在直播间界面显示操作命令及操作对象对应的提示信息,以使用户查看,确定终端即将执行的操作是否为自己需要的,如果不是,用户可以重新发出语音信号以实现所需的操作。

[0146] 例如,如图4所示的直播间界面的示意图,当终端确定操作命令为播放歌曲,操作对象为歌曲A时,可以在界面中显示提示信息410“播放歌曲”,以供用户查看。

[0147] 可见,在本实施例中,终端可以对服务器发送的识别结果进行解析,得到操作命令及操作对象,进而对操作对象执行操作命令对应的动作。这样,可以准确解析识别结果,准确响应用户的交互指令。

[0148] 作为本公开实施例的一种实施方式,上述识别结果可以包括操作命令标识。在这种情况下,上述对所述识别结果进行解析,得到操作命令及操作对象的步骤,可以包括:

[0149] 对所述识别结果进行解析,得到操作命令标识及操作对象;根据预设的操作命令标识与操作命令的对应关系,确定解析得到的操作命令标识对应的操作命令。

[0150] 为了方便确定操作命令,可以预先设定上述操作命令标识及操作命令的对应关系,其中,该对应关系包括的操作命令可以为虚拟空间中各功能对应操作命令。也就是说,由于直播间内的各功能一般是固定的,可以包括拍摄、连麦、送礼物、播放歌曲等,所以为了方便确定操作命令,可以预先设定这些功能对应的操作命令标识。

[0151] 这样,终端对上述识别结果进行解析得到操作命令标识及操作对象后,便可以根据该对应关系,确定解析得到的操作命令标识所对应的操作命令。例如,预设的操作命令标识与操作命令的对应关系如下表所示:

[0152]	序号	操作命令标识	操作命令
	1	A#	播放歌曲
[0153]	2	B#	打开
	3	C#	停止播放

[0154] 那么,如果终端对上述识别结果进行解析后,得到的操作命令标识为C#,便可以确定操作命令为:停止播放。

[0155] 可见,在本实施例中,上述识别结果可以包括操作命令标识,在这种情况下,终端可以对识别结果进行解析,得到操作命令标识及操作对象,进而根据预设的操作命令标识与操作命令的对应关系,确定解析得到的操作命令标识对应的操作命令。可以更加快速地确定操作命令,进一步提高响应速度,提升用户体验。

[0156] 相应于上述第一种语音信号处理方法,本公开实施例还提供了另一种语音信号处理方法,本公开实施例所提供的第二种语音信号处理方法可以上述直播应用程序的服务

器。下面对本公开实施例所提供的另一种语音信号处理方法进行介绍。

[0157] 如图5所示,一种语音信号处理方法,应用于服务器,所述方法包括:

[0158] 在步骤S501中,接收终端发送的第一语音信号;

[0159] 在步骤S502中,对所述第一语音信号进行语音识别,确定所述第一语音信号是否包括预设唤醒词;如果包括,执行步骤S503;如果不包括,不进行操作;

[0160] 在步骤S503中,确定所述第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端;

[0161] 在步骤S504中,接收终端发送的第二语音信号;

[0162] 在步骤S505中,对所述第二语音信号进行语音识别,得到识别结果,并发送所述识别结果至所述终端。

[0163] 可见,本公开实施例所提供的方案中,服务器可以接收终端发送的第一语音信号,对第一语音信号进行语音识别,确定第一语音信号是否包括预设唤醒词;如果包括,确定第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端,接收终端发送的第二语音信号,进而对第二语音信号进行语音识别,得到识别结果,并发送识别结果至所述终端。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0164] 用户在利用终端安装的直播应用程序进行直播或者观看直播时,可以发出各种语音信号,终端便可以获取虚拟空间中的语音信号,作为第一语音信号。终端获取虚拟空间中的第一语音信号后,可以判断第一语音信号对应的交互指令是否为唤醒指令。

[0165] 终端可以对第一语音信号进行语音识别,确定其语义信息,若该语义信息包括预设的唤醒词,那么便可以确定第一语音信号对应的交互指令为唤醒指令。由于对于直播场景来说,环境比较嘈杂,终端对语音信号的识别可能不够准确,因此为了准确确定第一语音信号对应的交互指令是否为唤醒指令,终端可以将第一语音信号发送至服务器。

[0166] 在步骤S501中,服务器便可以接收到终端发送的第一语音信号,服务器接收到第一语音信号后,可以确定第一语音信号是否包括预设唤醒词,若包括预设唤醒词,那么便可以确定第一语音信号对应的交互指令为唤醒指令,此时服务器可以发送确认结果至终端。

[0167] 终端在接收到服务器发送的确认结果时,可以继续接收虚拟空间中的语音信号,并将该语音信号作为第二语音信号发送至服务器。进而,在上述步骤S504中,服务器便可以接收终端发送的第二语音信号,并在步骤S505中,对第二语音信号进行语音识别,得到识别结果,并发送该识别结果至所述终端。在一种实施方式中,服务器可以采用protobuf编码方式对识别结果进行编码后,发送至终端。

[0168] 终端便在接收到服务器发送的识别结果后,可以对该识别结果进行解析,得到第二语音信号对应的交互指令,进而响应该交互指令,以完成用户想要执行的操作。

[0169] 作为本公开实施例的一种实施方式,如图6所示,服务器接收终端发送的目标语音信号的方式,可以包括:

[0170] S601,通过预定通道接收所述终端发送的多个数据包;

[0171] 上述目标语音信号为第一语音信号或第二语音信号,也就是说,服务器接收终端

发送的语音信号均可以采用该接收方式。

[0172] 服务器可以通过预定通道接收终端发送的多个数据包,终端在向服务器发送目标语音信号时,可以将目标语音信号按照预设时长封装成多个数据包,并为每个数据包添加顺序标识。进而通过预定通道将添加顺序标识后的多个数据包发送至服务器。这样,服务器接收到的多个数据包便携带顺序标识。

[0173] 在一种实施方式中,上述预定通道可以为SAS(Socket Audio Service,长链接语音服务)提供的专用Socket通道,这样,可以使终端与服务器之间的数据传输更加快速和稳定。

[0174] S602,根据每个数据包携带的顺序标识所标识的顺序,将所述多个数据包进行拼接,得到所述目标语音信号。

[0175] 进而,服务器可以根据每个数据包携带的顺序标识所标识的顺序,将多个数据包进行拼接,也就可以得到目标语音信号。

[0176] 例如,服务器接收到10个数据包,该多个数据包携带的顺序标识分别为1-10,那么服务器便可以按照顺序标识从1至10的顺序,将对应的数据包进行拼接,也就可以还原得到目标语音信号。

[0177] 可见,在本实施例中,服务器可以通过预定通道接收终端发送的多个数据包,并根据每个数据包携带的顺序标识所标识的顺序,将多个数据包进行拼接,得到目标语音信号。这样可以准确拼接得到目标语音信号,同时由于终端将目标语音信号封装成多个数据包,可以陆续将数据包发送至服务器,而不是等待目标语音信号结束后再全部发送至服务器,这样,服务器可以在接收到数据包后便开始对目标语音信号进行处理,可以提高响应速度,缩短用户等待实长,可以保证终端与服务器之间的数据传输更加快速和稳定,提高响应速度及用户体验。

[0178] 作为本公开实施例的一种实施方式,上述对所述第二语音信号进行语音识别,得到识别结果的步骤,可以包括:

[0179] 对所述第二语音信号进行语音识别,确定所述第二语音信号的语义信息;根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象,作为识别结果。

[0180] 由于语音信息可以标识用户想要进行的操作的内容,所以服务器可以对第二语音信号进行语音识别,确定第二语音信号的语义信息。进而根据该语义信息便可以确定第二语音信号对应的操作命令及操作对象,并将其作为识别结果。其中,操作命令即为需要执行的动作对应的命令,操作对象即为需要执行的动作所针对的目标。

[0181] 例如,服务器对第二语音信号进行语音识别,确定第二语音信号的语义信息为:发送笑脸表情,那么服务器可以确定第二语音信号对应的操作命令为:发送,操作对象为:笑脸表情。

[0182] 可见,在本实施例中,服务器可以对第二语音信号进行语音识别,确定第二语音信号的语义信息,进而根据语义信息确定第二语音信号对应的操作命令及操作对象,作为识别结果。可以准确确定识别结果,以便终端可以准确响应用户的交互指令。

[0183] 作为本公开实施例的一种实施方式,上述根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象,作为识别结果的步骤,可以包括:

[0184] 根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象;根据预设

的操作命令标识与操作命令的对应关系,确定所确定的操作命令对应的操作命令标识,并将所述操作命令标识及所述操作对象作为识别结果。

[0185] 为了方便确定操作命令,可以预先设定上述操作命令标识及操作命令的对应关系,其中,该对应关系包括的操作命令可以为虚拟空间中各功能对应操作命令。也就是说,由于直播间内的各功能一般是固定的,可以包括拍摄、连麦、送礼物、播放歌曲等,所以为了方便确定操作命令,可以预先设定这些功能对应的操作命令标识。

[0186] 这样,服务器根据第二语音信号的语义信息确定第二语音信号对应的操作命令及操作对象后,可以根据预设的操作命令标识与操作命令的对应关系,确定所确定的操作命令对应的操作命令标识,进而,将操作命令标识及所述操作对象作为识别结果发送至终端。

[0187] 可见,在本实施例中,服务器可以根据语义信息确定第二语音信号对应的操作命令及操作对象,进而根据预设的操作命令标识与操作命令的对应关系,确定所确定的操作命令对应的操作命令标识,并将操作命令标识及操作对象作为识别结果。这样可以使终端更加快速地确定操作命令,进一步提高响应速度,提升用户体验。

[0188] 为了进一步描述本公开实施例提供的语音信号处理方法,下面结合图7对本公开实施例提供的语音信号处理方法进行介绍。如图7所示,本公开实施例提供的语音信号处理方法可以包括:

[0189] 步骤S701,终端获取虚拟空间中的语音信号,作为第一语音信号,并判断所述第一语音信号对应的交互指令是否为唤醒指令;若为是,执行步骤S702;

[0190] 其中,所述虚拟空间用于进行实时视频行为。

[0191] 步骤S702,终端发送所述第一语音信号至服务器;

[0192] 步骤S703,服务器接收终端发送的第一语音信号;

[0193] 步骤S704,服务器对所述第一语音信号进行语音识别,确定所述第一语音信号是否包括预设唤醒词;如果包括,执行步骤S705;

[0194] 步骤S705,服务器确定所述第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端;

[0195] 步骤S706,终端在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号,并将所述第二语音信号发送至所述服务器;

[0196] 其中,所述确认结果用于指示所述第一语音信号对应的交互指令为唤醒指令。

[0197] 步骤S707,服务器接收所述终端发送的第二语音信号;

[0198] 步骤S708,服务器对所述第二语音信号进行语音识别,得到识别结果,并发送所述识别结果至所述终端;

[0199] 步骤S709,终端接收所述服务器发送的所述第二语音信号的识别结果;

[0200] 步骤S710,终端对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令。

[0201] 可见,在本实施例中,终端可以获取虚拟空间中的语音信号,作为第一语音信号,并判断第一语音信号对应的交互指令是否为唤醒指令;若为是,发送第一语音信号至服务器,服务器接收终端发送的第一语音信号,对第一语音信号进行语音识别,确定第一语音信号是否包括预设唤醒词;如果包括,确定第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端,终端在接收到服务器发送的确认结果时,继续接收虚拟空间中的语

音信号,作为第二语音信号,并将第二语音信号发送至服务器,服务器接收终端发送的第二语音信号,对第二语音信号进行语音识别,得到识别结果,并发送识别结果至终端,终端接收服务器发送的第二语音信号的识别结果,进而对识别结果进行解析,得到第二语音信号对应的交互指令,并响应交互指令。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0202] 由于步骤S701-步骤S710的具体实现方式已经在上述各实施例中进行介绍,因此在此不再赘述。

[0203] 图8是根据一示例性实施例示出的第一种语音信号处理装置框图。

[0204] 如图8所示,一种语音信号处理装置,应用于终端,所述装置包括:

[0205] 第一语音信号获取模块810,被配置为执行获取虚拟空间中的语音信号,作为第一语音信号,并判断所述第一语音信号对应的交互指令是否为唤醒指令;

[0206] 其中,所述虚拟空间用于进行实时视频行为。

[0207] 第一语音信号发送模块820,被配置为执行若所述第一语音信号对应的交互指令为唤醒指令,并通过发送模块(图8中未示出)发送所述第一语音信号至服务器;

[0208] 第二语音信号获取模块830,被配置为执行在接收到所述服务器发送的确认结果时,继续接收所述虚拟空间中的语音信号,作为第二语音信号,并通过发送模块将所述第二语音信号发送至所述服务器;

[0209] 其中,所述确认结果用于指示所述第一语音信号对应的交互指令为唤醒指令。

[0210] 识别结果接收模块840,被配置为执行接收所述服务器发送的所述第二语音信号的识别结果;

[0211] 识别结果解析模块850,被配置为执行对所述识别结果进行解析,得到所述第二语音信号对应的交互指令,并响应所述交互指令。

[0212] 可见,本公开实施例所提供的方案中,终端可以获取虚拟空间中的语音信号,作为第一语音信号,并判断第一语音信号对应的交互指令是否为唤醒指令,若为是,发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,作为第二语音信号,并将第二语音信号发送至服务器,其中,确认结果用于指示第一语音信号对应的交互指令为唤醒指令,接收服务器发送的第二语音信号的识别结果,对识别结果进行解析,得到第二语音信号对应的交互指令,并响应交互指令。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0213] 作为本公开实施例的一种实施方式,上述发送模块可以包括:

[0214] 数据包封装单元(图8中未示出),被配置为执行按照预设时长将所述目标语音信号封装成多个数据包;

[0215] 其中,所述目标语音信号为所述第一语音信号或所述第二语音信号。

- [0216] 顺序标识添加单元(图8中未示出),被配置为执行为每个数据包添加顺序标识;
- [0217] 数据包发送单元(图8中未示出),被配置为执行通过预定通道发送添加顺序标识后的多个数据包至所述服务器。
- [0218] 作为本公开实施例的一种实施方式,上述第二语音信号获取模块830可以包括:
- [0219] 提示信息输出单元(图8中未示出),被配置为执行在接收到所述服务器发送的确认结果时,输出唤醒成功提示信息;
- [0220] 第二语音信号获取单元(图8中未示出),被配置为执行获取用户在所述虚拟空间中基于所述唤醒成功提示信息输入的语音信号,作为第二语音信号。
- [0221] 作为本公开实施例的一种实施方式,上述识别结果解析模块850可以包括:
- [0222] 识别结果解析单元(图8中未示出),被配置为执行对所述识别结果进行解析,得到操作命令及操作对象;
- [0223] 动作执行单元(图8中未示出),被配置为执行对所述操作对象执行所述操作命令对应的动作。
- [0224] 作为本公开实施例的一种实施方式,上述识别结果可以包括操作命令标识;
- [0225] 上述识别结果解析单元可以包括:
- [0226] 识别结果解析子单元(图8中未示出),被配置为执行对所述识别结果进行解析,得到操作命令标识及操作对象;
- [0227] 操作命令确定子单元(图8中未示出),被配置为执行根据预设的操作命令标识与操作命令的对应关系,确定解析得到的操作命令标识对应的操作命令。
- [0228] 其中,所述对应关系包括的操作命令为所述虚拟空间中各功能对应操作命令。
- [0229] 图9是根据一示例性实施例示出的第二种语音信号处理装置框图。
- [0230] 如图9所示,一种语音信号处理装置,应用于服务器,所述装置包括:
- [0231] 第一语音信号接收模块910,被配置为执行接收终端发送的第一语音信号;
- [0232] 第一语音信号识别模块920,被配置为执行对所述第一语音信号进行语音识别,确定所述第一语音信号是否包括预设唤醒词;
- [0233] 确认结果发送模块930,被配置为执行如果所述第一语音信号包括预设唤醒词,确定所述第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端;
- [0234] 第二语音信号接收模块940,被配置为执行接收终端发送的第二语音信号;
- [0235] 第二语音信号识别模块950,被配置为执行对所述第二语音信号进行语音识别,得到识别结果,并发送所述识别结果至所述终端。
- [0236] 可见,本公开实施例所提供的方案中,服务器可以接收终端发送的第一语音信号,对第一语音信号进行语音识别,确定第一语音信号是否包括预设唤醒词;如果包括,确定第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端,接收终端发送的第二语音信号,进而对第二语音信号进行语音识别,得到识别结果,并发送识别结果至所述终端。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。
- [0237] 作为本公开实施例的一种实施方式,上述第一语音信号接收模块910和第二语音

信号接收模块940可以包括：

[0238] 数据包接收单元(图9中未示出),被配置为执行通过预定通道接收所述终端发送的多个数据包,其中,每个数据包携带顺序标识;

[0239] 数据包拼接单元(图9中未示出),被配置为执行根据每个数据包携带的顺序标识所标识的顺序,将所述多个数据包进行拼接,得到所述目标语音信号。

[0240] 其中,所述目标语音信号为所述第一语音信号或所述第二语音信号。

[0241] 作为本公开实施例的一种实施方式,上述第二语音信号识别模块950可以包括:

[0242] 语义信息确定单元(图9中未示出),被配置为执行对所述第二语音信号进行语音识别,确定所述第二语音信号的语义信息;

[0243] 识别结果确定单元(图9中未示出),被配置为执行根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象,作为识别结果。

[0244] 作为本公开实施例的一种实施方式,上述识别结果确定单元可以包括:

[0245] 第一识别子单元(图9中未示出),被配置为执行根据所述语义信息确定所述第二语音信号对应的操作命令及操作对象;

[0246] 第二识别子单元(图9中未示出),被配置为根据预设的操作命令标识与操作命令的对应关系,确定所确定的操作命令对应的操作命令标识,并将所述操作命令标识及所述操作对象作为识别结果。

[0247] 其中,所述对应关系包括的操作命令为所述终端的虚拟空间中各功能对应操作命令,所述虚拟空间用于进行实时视频行为。

[0248] 本公开实施例还提供了一种终端,如图10所示,终端可以包括处理器1001、通信接口1002、存储器1003和通信总线1004,其中,处理器1001,通信接口1002,存储器1003通过通信总线1004完成相互间的通信,

[0249] 存储器1003,用于存放计算机程序;

[0250] 处理器1001,用于执行存储器1003上所存放的程序时,实现上述第一种语音信号处理方法。

[0251] 可见,本公开实施例所提供的方案中,终端可以获取虚拟空间中的语音信号,作为第一语音信号,并判断第一语音信号对应的交互指令是否为唤醒指令,若为是,发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,作为第二语音信号,并将第二语音信号发送至服务器,其中,确认结果用于指示第一语音信号对应的交互指令为唤醒指令,接收服务器发送的第二语音信号的识别结果,对识别结果进行解析,得到第二语音信号对应的交互指令,并响应交互指令。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0252] 上述终端提到的通信总线可以是外设部件互连标准(Peripheral Component Interconnect, PCI)总线或扩展工业标准结构(Extended Industry Standard Architecture, EISA)总线等。该通信总线可以分为地址总线、数据总线、控制总线等。为便于表示,图中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0253] 通信接口用于上述终端与其他设备之间的通信。

[0254] 存储器可以包括随机存取存储器(Random Access Memory, RAM),也可以包括非易失性存储器(Non-Volatile Memory, NVM),例如至少一个磁盘存储器。可选的,存储器还可以是至少一个位于远离前述处理器的存储装置。

[0255] 上述的处理器可以是通用处理器,包括中央处理器(Central Processing Unit, CPU)、网络处理器(Network Processor, NP)等;还可以是数字信号处理器(Digital Signal Processing, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现场可编程门阵列(Field-Programmable Gate Array, FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

[0256] 图11是根据一示例性实施例示出的上述终端的一种具体结构框图。例如,终端可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,医疗设备,健身设备,个人数字助理等。

[0257] 参照图11,终端可以包括以下一个或多个组件:处理组件1102,存储器1104,电源组件1106,多媒体组件1108,音频组件1110,输入/输出(I/O)的接口1112,传感器组件1114,以及通信组件1116。

[0258] 处理组件1102通常控制终端的整体操作,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1102可以包括一个或多个处理器1120来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件1102可以包括一个或多个模块,便于处理组件1102和其他组件之间的交互。例如,处理组件1102可以包括多媒体模块,以方便多媒体组件1108和处理组件1102之间的交互。

[0259] 存储器1104被配置为存储各种类型的数据以支持在终端的操作。这些数据的示例包括用于在终端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。存储器1104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。

[0260] 电源组件1106为终端的各种组件提供电力。电源组件1106可以包括电源管理系统,一个或多个电源,及其他与为终端生成、管理和分配电力相关联的组件。

[0261] 多媒体组件1108包括在终端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1108包括一个前置摄像头和/或后置摄像头。当终端处于操作模式,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。

[0262] 音频组件1110被配置为输出和/或输入音频信号。例如,音频组件1110包括一个麦克风(MIC),当终端处于操作模式,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存储器1104或经由通信组件

1116发送。在一些实施例中，音频组件1110还包括一个扬声器，用于输出音频信号。

[0263] I/O接口1112为处理组件1102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。

[0264] 传感器组件1114包括一个或多个传感器，用于为终端提供各个方面的状态评估。例如，传感器组件1114可以检测到终端的打开/关闭状态，组件的相对定位，例如所述组件为终端的显示器和小键盘，传感器组件1114还可以检测终端或终端一个组件的位置改变，用户与终端接触的存在或不存在，终端方位或加速/减速和终端的温度变化。传感器组件1114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1114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1114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。

[0265] 通信组件1116被配置为便于终端和其他设备之间有线或无线方式的通信。终端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1116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1116还包括近场通信（NFC）模块，以促进短程通信。例如，在NFC模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

[0266] 在示例性实施例中，终端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，用于执行上述方法。

[0267] 在示例性实施例中，还提供了一种包括指令的非临时性计算机可读存储介质，例如包括指令的存储器1104，上述指令可由终端的处理器1120执行以完成上述方法。例如，所述非临时性计算机可读存储介质可以是ROM、随机存取存储器（RAM）、CD-ROM、磁带、软盘和光数据存储设备等。

[0268] 本公开实施例还提供了一种服务器，如图12所示，服务器可以包括处理器1201、通信接口1202、存储器1203和通信总线1204，其中，处理器1201，通信接口1202，存储器1203通过通信总线1204完成相互间的通信，

[0269] 存储器1203，用于存放计算机程序；

[0270] 处理器1201，用于执行存储器1203上所存放的程序时，实现上述第一种语音信号处理方法。

[0271] 可见，本公开实施例所提供的方案中，服务器可以接收终端发送的第一语音信号，对第一语音信号进行语音识别，确定第一语音信号是否包括预设唤醒词；如果包括，确定第一语音信号对应的交互指令为唤醒指令，并发送确认结果至所述终端，接收终端发送的第二语音信号，进而对第二语音信号进行语音识别，得到识别结果，并发送识别结果至所述终端。由于在虚拟空间中，用户可以通过发出语音信号对终端进行操作，无需频繁手动终端，大大提高在网络直播过程中操作终端的便利性。同时，由于终端确定第一语音信号对应的交互指令为唤醒指令后，会发送第一语音信号至服务器，在接收到服务器发送的确认结果时，继续接收虚拟空间中的语音信号，因此可以在嘈杂直播环境下准确识别唤醒指令。

[0272] 上述服务器提到的通信总线可以是外设部件互连标准 (Peripheral Component Interconnect, PCI) 总线或扩展工业标准结构 (Extended Industry Standard Architecture, EISA) 总线等。该通信总线可以分为地址总线、数据总线、控制总线等。为便于表示,图中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0273] 通信接口用于上述服务器与其他设备之间的通信。

[0274] 存储器可以包括随机存取存储器 (Random Access Memory, RAM),也可以包括非易失性存储器 (Non-Volatile Memory, NVM),例如至少一个磁盘存储器。可选的,存储器还可以是至少一个位于远离前述处理器的存储装置。

[0275] 上述的处理器可以是通用处理器,包括中央处理器 (Central Processing Unit, CPU)、网络处理器 (Network Processor, NP) 等;还可以是数字信号处理器 (Digital Signal Processing, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现场可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

[0276] 图13是根据一示例性实施例示出的服务器的一种具体结构框图。服务器包括处理组件1322,其进一步包括一个或多个处理器,以及由存储器1332所代表的存储器资源,用于存储可由处理组件1322的执行的指令,例如应用程序。存储器1332中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外,处理组件1322被配置为执行指令,以执行上述视频内容主题的确定方法。

[0277] 服务器还可以包括一个电源组件1326被配置为执行服务器的电源管理,一个有线或无线网络接口1350被配置为将服务器连接到网络,和一个输入输出 (I/O) 接口1358。服务器可以操作基于存储在存储器1332的操作系统,例如Windows Server™,Mac OS X™, Unix™,Linux™,FreeBSD™或类似。

[0278] 本公开实施例还提供了一种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,使得终端能够执行上述实施例中任一所述的第一种语音信号处理方法。

[0279] 可见,本公开实施例所提供的方案中,终端可以获取虚拟空间中的语音信号,作为第一语音信号,并判断第一语音信号对应的交互指令是否为唤醒指令,若为是,发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,作为第二语音信号,并将第二语音信号发送至服务器,其中,确认结果用于指示第一语音信号对应的交互指令为唤醒指令,接收服务器发送的第二语音信号的识别结果,对识别结果进行解析,得到第二语音信号对应的交互指令,并响应交互指令。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0280] 本公开实施例还提供了一种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上述实施例中任一所述的第二种语音信号处理方法。

[0281] 可见,本公开实施例所提供的方案中,服务器可以接收终端发送的第一语音信号,

对第一语音信号进行语音识别,确定第一语音信号是否包括预设唤醒词;如果包括,确定第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端,接收终端发送的第二语音信号,进而对第二语音信号进行语音识别,得到识别结果,并发送识别结果至所述终端。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0282] 本公开实施例还提供了一种应用程序产品,该应用程序产品用于在运行时执行上述实施例中任一所述的第一种语音信号处理方法。

[0283] 可见,本公开实施例所提供的方案中,该应用程序产品在运行时可以获取虚拟空间中的语音信号,作为第一语音信号,并判断第一语音信号对应的交互指令是否为唤醒指令,若为是,发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,作为第二语音信号,并将第二语音信号发送至服务器,其中,确认结果用于指示第一语音信号对应的交互指令为唤醒指令,接收服务器发送的第二语音信号的识别结果,对识别结果进行解析,得到第二语音信号对应的交互指令,并响应交互指令。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0284] 本公开实施例还提供了一种应用程序产品,该应用程序产品用于在运行时执行上述实施例中任一所述的第二种语音信号处理方法。

[0285] 可见,本公开实施例所提供的方案中,该应用程序产品在运行时可以接收终端发送的第一语音信号,对第一语音信号进行语音识别,确定第一语音信号是否包括预设唤醒词;如果包括,确定第一语音信号对应的交互指令为唤醒指令,并发送确认结果至所述终端,接收终端发送的第二语音信号,进而对第二语音信号进行语音识别,得到识别结果,并发送识别结果至所述终端。由于在虚拟空间中,用户可以通过发出语音信号对终端进行操作,无需频繁手动终端,大大提高在网络直播过程中操作终端的便利性。同时,由于终端确定第一语音信号对应的交互指令为唤醒指令后,会发送第一语音信号至服务器,在接收到服务器发送的确认结果时,继续接收虚拟空间中的语音信号,因此可以在嘈杂直播环境下准确识别唤醒指令。

[0286] 本领域技术人员在考虑说明书及实践这里公开的申请后,将容易想到本公开的其他实施方案。本公开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上面的权利要求指出。

[0287] 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。

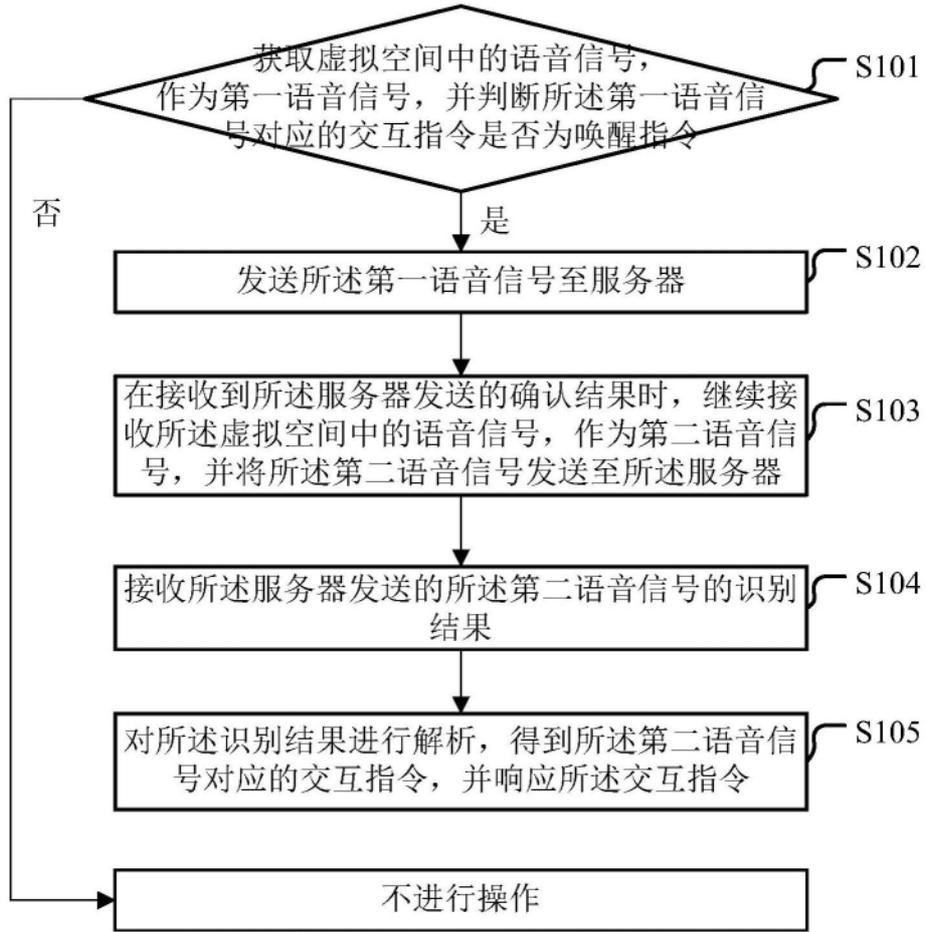


图1

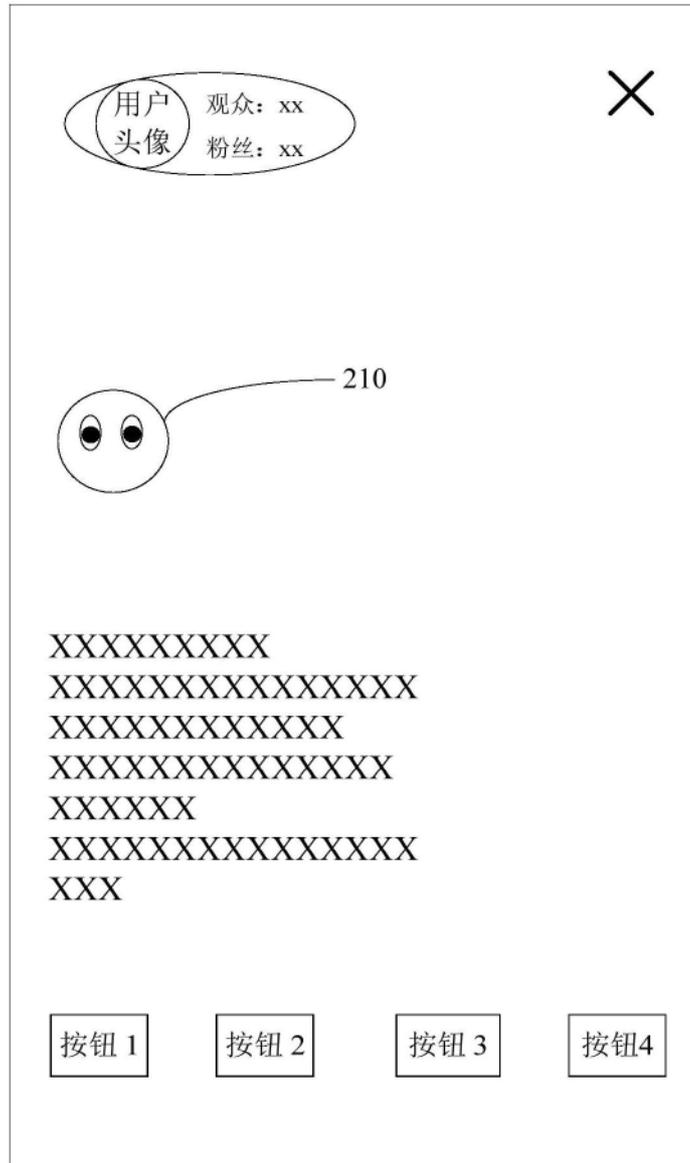


图2

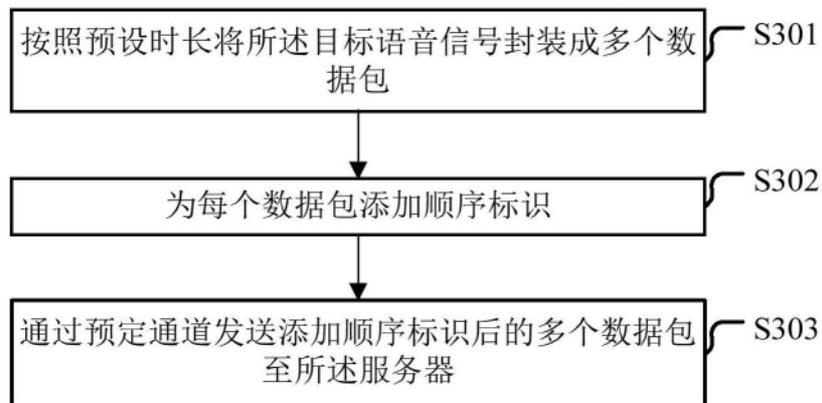


图3

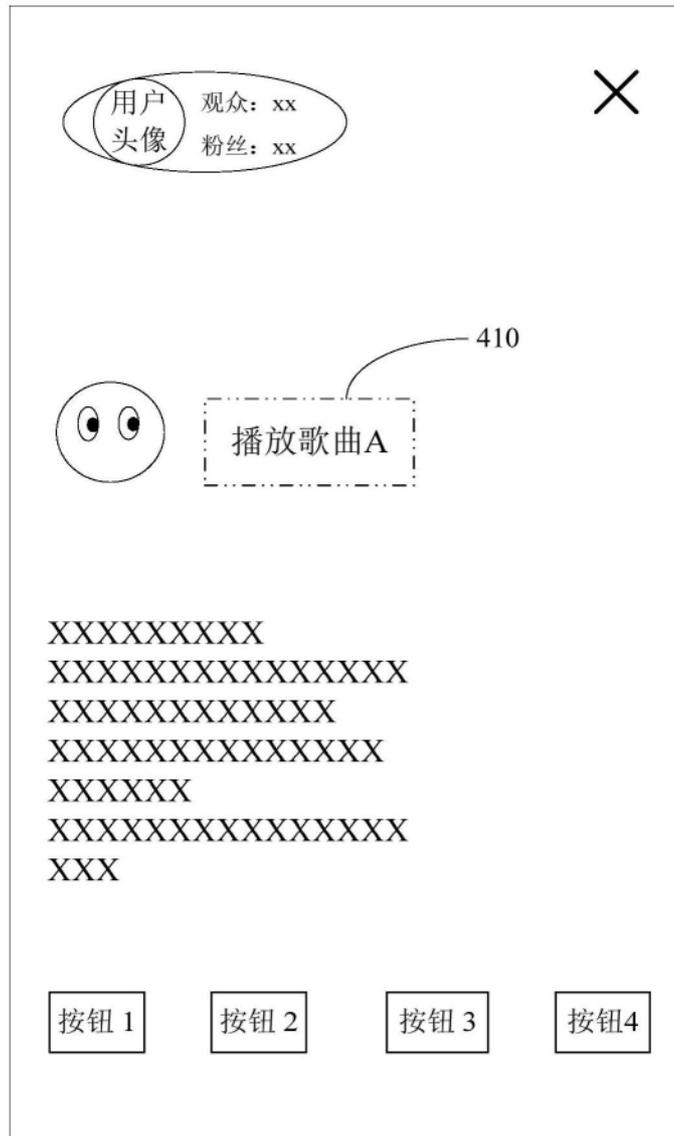


图4

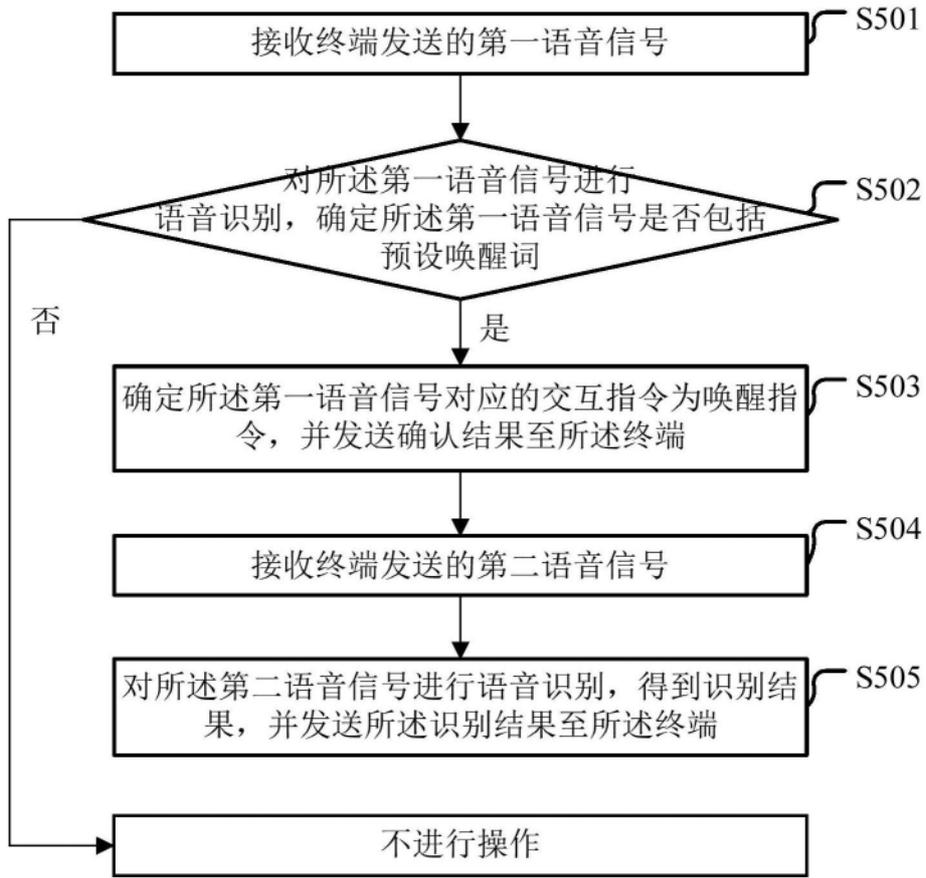


图5

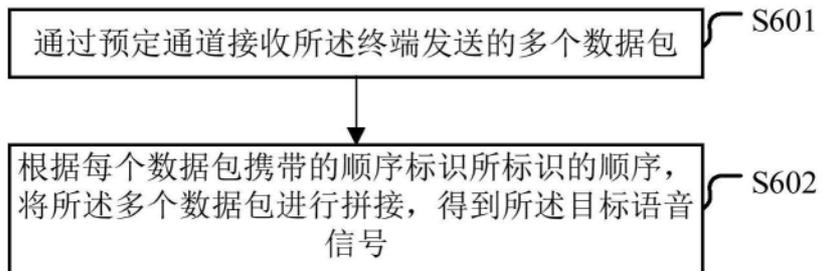


图6

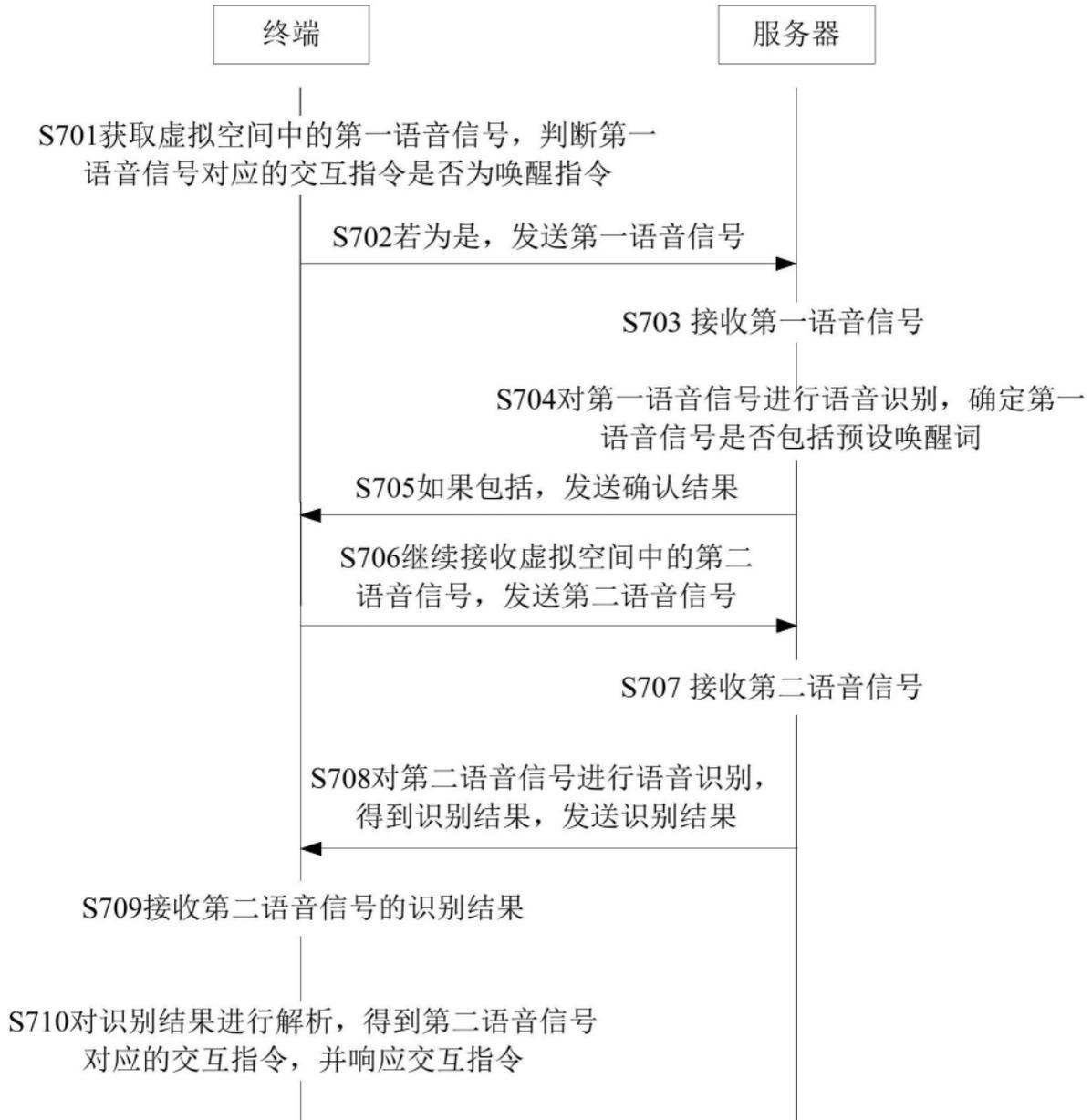


图7

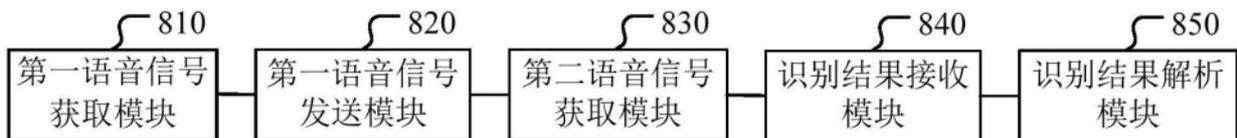


图8



图9

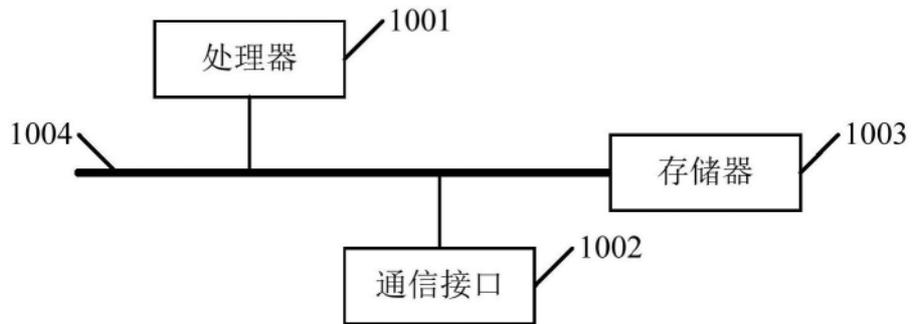


图10

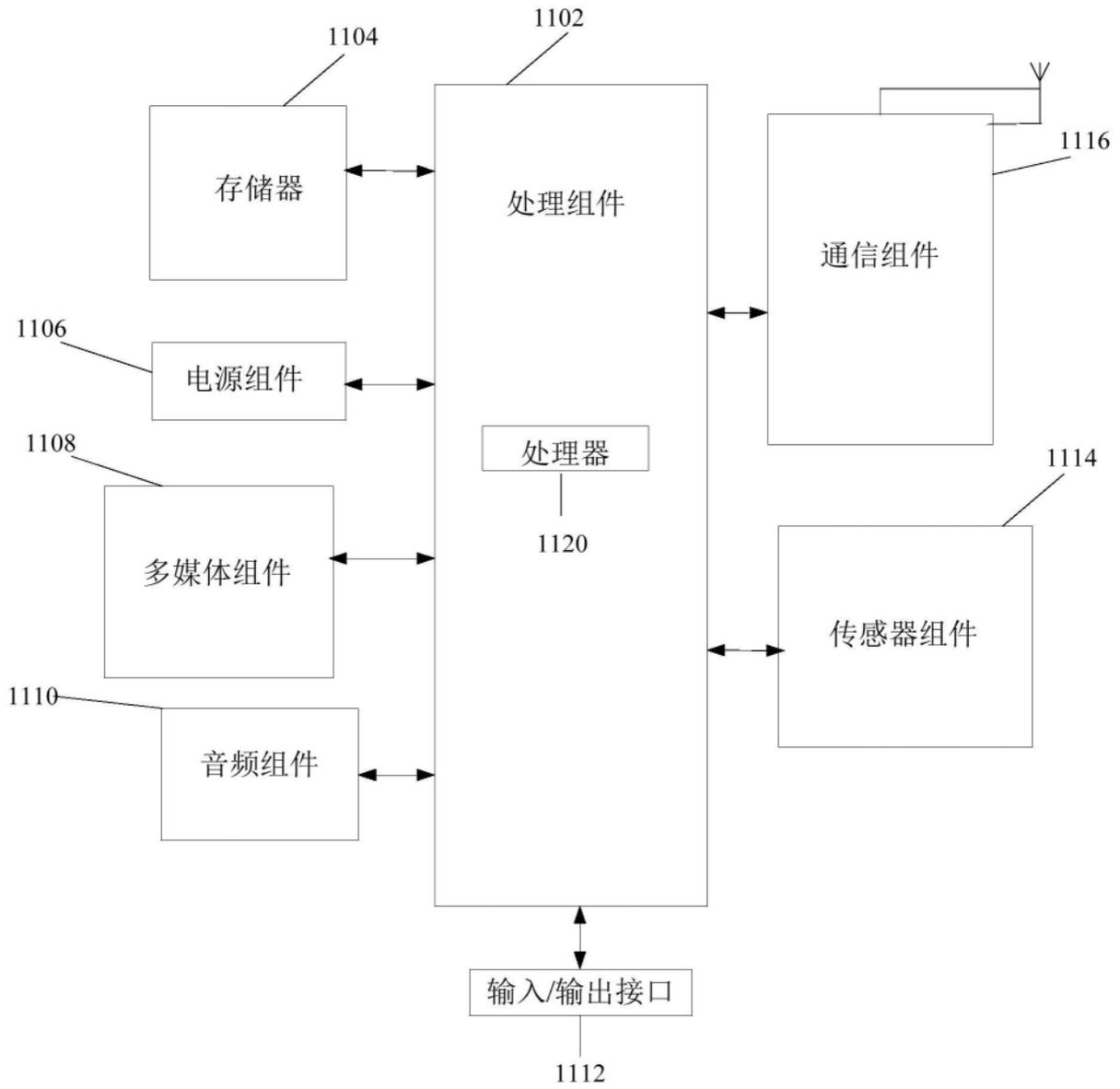


图11

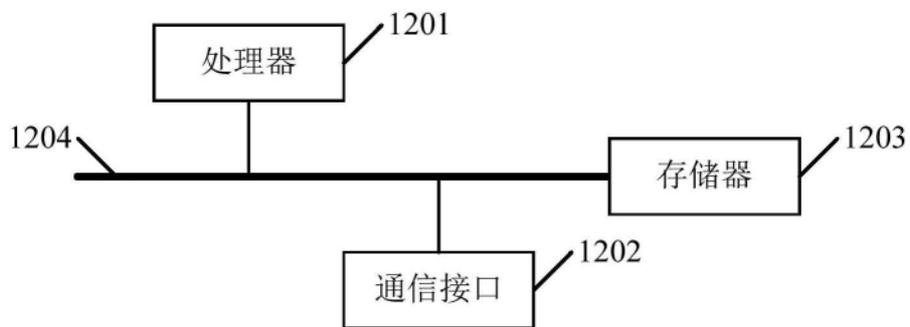


图12

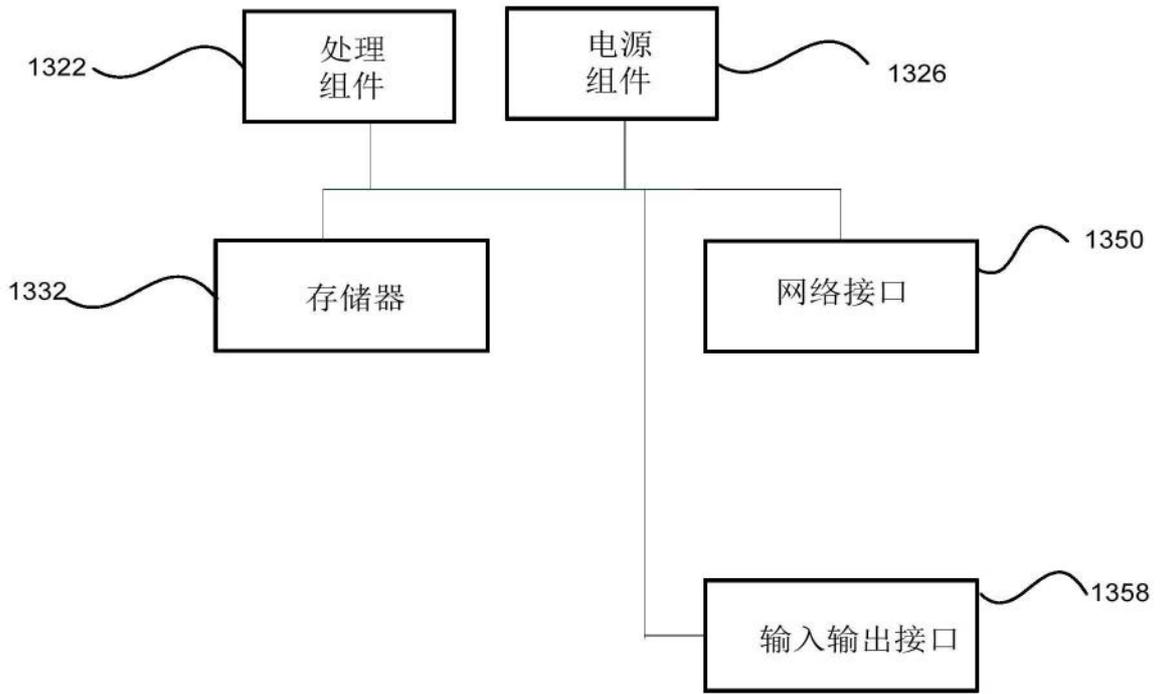


图13