



(12) 发明专利申请

(10) 申请公布号 CN 112750452 A

(43) 申请公布日 2021.05.04

(21) 申请号 202011598381.X

G10L 25/30 (2013.01)

(22) 申请日 2020.12.29

G10L 15/26 (2006.01)

G10L 19/12 (2013.01)

(71) 申请人 北京字节跳动网络技术有限公司

地址 100041 北京市石景山区实兴大街30
号院3号楼2层B-0035房间

(72) 发明人 杨智焯

(74) 专利代理机构 上海光栅知识产权代理有限
公司 31340

代理人 关浩 马雯雯

(51) Int. Cl.

G10L 21/0208 (2013.01)

G10L 21/0216 (2013.01)

G10L 21/0232 (2013.01)

G10L 21/0272 (2013.01)

G10L 25/24 (2013.01)

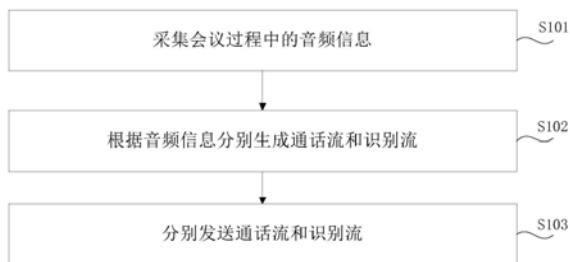
权利要求书2页 说明书13页 附图7页

(54) 发明名称

语音处理方法、装置、系统、智能终端以及电子设备

(57) 摘要

本公开实施例提供一种语音处理方法、装置、系统、智能终端、电子设备以及存储介质,包括:采集会议过程中的音频信息,根据音频信息分别生成通话流和识别流,其中,通话流用于语音通话,识别流用于语音识别,发送通话流和识别流,通过基于音频信息分别生成通话流和识别流的技术方案,使得确定出的音频信息对应的会议内容的展现维度较多、较丰富,进而提高了会议的准确性,且提高了会议的智能化和质量,还提高了用户的会议体验。



1. 一种语音处理方法,所述方法包括:
采集会议过程中的音频信息;
根据所述音频信息分别生成通话流和识别流,其中,所述通话流用于语音通话,所述识别流用于语音识别;
发送所述通话流和所述识别流。
2. 根据权利要求1所述的方法,其中,根据所述音频信息分别生成通话流和识别流,包括:
根据不同的处理方式对所述音频信息进行处理,获得所述通话流和所述识别流。
3. 根据权利要求2所述的方法,其中,根据不同的处理方式对所述音频信息进行处理,得到所述通话流和所述识别流,包括:
对所述音频信息进行增强清晰度处理,获得所述通话流;以及,
对所述音频信息进行保真处理,获得所述识别流。
4. 根据权利要求3所述的方法,其中,对所述音频信息进行增强清晰度处理,获得所述通话流,包括:
对所述音频信息进行降噪处理和自动增益控制,获得所述通话流。
5. 根据权利要求3所述的方法,其中,对所述音频信息进行保真处理,获得所述识别流,包括:
对所述音频信息进行波束选择处理,获得所述识别流。
6. 根据权利要求3所述的方法,在对所述音频信息进行增强清晰度处理,获得所述通话流;以及,对所述音频信息进行保真处理,获得所述识别流之前,所述方法还包括:
对所述音频信息进行回声消除处理。
7. 根据权利要求1至6中任一项所述的方法,其中,所述方法应用于智能终端;发送所述通话流和所述识别流,包括:
所述智能终端向云端服务器发送所述识别流,所述识别流用于,被所述云端服务器进行语音识别,并通过所述云端服务器,向参与所述会议的第一终端设备发送所述识别流和/或对所述识别进行语音识别的识别结果;
所述智能终端向云端服务器发送所述通话流,并通过所述云端服务器向所述第一终端设备分发所述通话流。
8. 一种智能终端,所述智能终端包括:麦克风阵列、处理器和通信模块;
所述麦克风阵列,用于采集会议过程中的音频信息;
所述处理器,用于根据所述音频信息分别生成通话流和识别流,其中,所述通话流用于语音通话,所述识别流用于语音识别;
所述通信模块,用于发送所述通话流和所述识别流。
9. 根据权利要求8所述的智能终端,其中,所述处理器用于,根据不同的处理方式对所述音频信息进行处理,获得所述通话流和所述识别流。
10. 根据权利要求9所述的智能终端,其中,所述处理器用于,对所述音频信息进行增强清晰度处理,获得所述通话流;以及,对所述音频信息进行保真处理,获得所述识别流。
11. 根据权利要求10所述的智能终端,其中,所述处理器用于,对所述音频信息进行降噪处理和自动增益控制,获得所述通话流。

12. 根据权利要求10所述的智能终端,其中,所述处理器用于,对所述音频信息进行波束选择处理,获得所述识别流。

13. 根据权利要求10所述的智能终端,其中,所述处理器用于,对所述音频信息进行回声消除处理。

14. 根据权利要求8至13中任一项所述的智能终端,所述智能终端还包括:
扬声器,用于对参与所述会议的第一终端设备发送的通话流进行语音播报。

15. 一种语音处理装置,所述装置包括:

采集模块,用于采集会议过程中的音频信息;

生成模块,用于根据所述音频信息分别生成通话流和识别流,其中,所述通话流用于语音通话,所述识别流用于语音识别;

发送模块,用于发送所述通话流和所述识别流。

16. 一种电子设备,其特征在于,包括:至少一个处理器和存储器;

所述存储器存储计算机执行指令;

所述至少一个处理器执行所述存储器存储的计算机执行指令,使得所述至少一个处理器执行如权利要求1至7中任一项所述的语音处理方法。

17. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有计算机执行指令,当处理器执行所述计算机执行指令时,实现如权利要求1至7中任一项所述的语音处理方法。

语音处理方法、装置、系统、智能终端以及电子设备

技术领域

[0001] 本公开实施例涉及计算机、语音处理以及网络通信技术领域,尤其涉及一种语音处理方法、装置、系统、智能终端、电子设备以及存储介质。

背景技术

[0002] 会议是指利用现代化的通讯手段,实现召开会议的目的,会议可以包括远程会议,且远程会议主要可以包含电话会议,网络会议,视频会议。

[0003] 目前,应用于会议场景中的语音处理方法包括:本端会议设备采集本端用户对应的音频信息,并将本端用户对应的音频信息发送给对端会议设备,相应地,对端会议设备采集对端用户的音频信息,并将对端用户的音频信息发送给本端会议设备,其中,音频信息用于语音通话。

[0004] 然而,传统的语音处理方法至少存在以下技术问题:通过用于语音通话的音频信息实现会议,可能导致会议的会议内容展现维度较少、丰富程度相对较低,从而导致会议的质量相对偏低。

发明内容

[0005] 本公开实施例提供一种语音处理方法、装置、系统、智能终端、电子设备以及存储介质,以克服相关技术中会议的质量偏低的问题。

[0006] 第一方面,本公开实施例提供一种语音处理方法,包括:

[0007] 采集会议过程中的音频信息;

[0008] 根据所述音频信息分别生成通话流和识别流,其中,所述通话流用于语音通话,所述识别流用于语音识别;

[0009] 发送所述通话流和所述识别流。

[0010] 第二方面,本公开实施例提供一种智能终端,所述智能终端包括:麦克风阵列、处理器和通信模块;

[0011] 所述麦克风阵列,用于采集会议过程中的音频信息;

[0012] 所述处理器,用于根据所述音频信息分别生成通话流和识别流,其中,所述通话流用于语音通话,所述识别流用于语音识别;

[0013] 所述通信模块,用于发送所述通话流和所述识别流。

[0014] 第三方面,本公开实施例提供一种语音处理装置,所述装置包括:

[0015] 采集模块,用于采集会议过程中的音频信息;

[0016] 生成模块,用于根据所述音频信息分别生成通话流和识别流,其中,所述通话流用于语音通话,所述识别流用于语音识别;

[0017] 发送模块,用于发送所述通话流和所述识别流。

[0018] 第四方面,本公开实施例提供一种语音处理系统,所述系统包括:第一终端设备和如上第二方面所述的智能终端;或者,第一终端设备和如上第三方面所述的装置,其中,所

述第一终端设备为参与所述会议的终端设备。

[0019] 第五方面,本公开实施例提供一种电子设备,包括:至少一个处理器和存储器;

[0020] 所述存储器存储计算机执行指令;

[0021] 所述至少一个处理器执行所述存储器存储的计算机执行指令,使得所述至少一个处理器执行如上第一方面以及第一方面各种可能的设计所述的信息显示方法。

[0022] 第六方面,本公开实施例提供一种计算机可读存储介质,所述计算机可读存储介质中存储有计算机执行指令,当处理器执行所述计算机执行指令时,实现如上第一方面以及第一方面各种可能的设计所述的信息显示方法。

[0023] 本实施例提供的语音处理方法、装置、系统、智能终端、电子设备以及存储介质,包括:采集会议过程中的音频信息,根据音频信息分别生成通话流和识别流,其中,通话流用于语音通话,识别流用于语音识别,分别发送通话流和识别流,通过基于音频信息分别生成通话流和识别流的技术特征的技术方案,避免了相关技术中,会议的会议内容的展现维度单一、丰富程度相对较低的问题,实现了确定出的音频信息对应的会议内容的展现维度较多、较丰富,提高了会议的用户对会议内容的理解,进而提高了会议的准确性,且提高了会议的智能化和质量,还提高了用户的会议体验。

附图说明

[0024] 为了更清楚地说明本公开实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图是本公开的一些实施例,对于本领域普通技术人员来讲,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。

[0025] 图1为公开一个实施例的语音处理方法的应用场景示意图;

[0026] 图2为本公开一个实施例的语音处理方法的流程示意图;

[0027] 图3为本公开另一实施例的语音处理方法的流程示意图;

[0028] 图4为本公开另一实施例的语音处理方法的应用场景示意图;

[0029] 图5为本公开实施例的语音处理方法的原理示意图;

[0030] 图6为图5所示的处理器原理图;

[0031] 图7为本公开一个实施例的语音处理装置的示意图;

[0032] 图8为本公开另一实施例的语音处理装置的示意图;

[0033] 图9为本公开实施例提供的电子设备的硬件结构示意图。

具体实施方式

[0034] 为使本公开实施例的目的、技术方案和优点更加清楚,下面将结合本公开实施例中的附图,对本公开实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本公开一部分实施例,而不是全部的实施例。基于本公开中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本公开保护的范围。

[0035] 本公开实施例提供的语音处理方法可以应用于会议的应用场景,且具体可以应用于远程会议的应用场景,其中,远程会议是指利用现代化的通讯手段,实现跨区域召开会议的目的,且远程会议系统可以包含电话会议、网络会议以及视频会议等。

[0036] 图1为公开一个实施例的语音处理方法的应用场景示意图。

[0037] 如图1所示,应用场景可以包括:服务器、至少两个终端设备、以及与各终端设备各自对应用户。其中,图1中示范性地展示了n台终端设备,即与会方的数量为n。

[0038] 示例性地,服务器可以与每一终端设备建立通信链路,并基于通信链路实现与各终端设备的信息交互,从而实现各终端设备各自对应的用户基于远程会议实现交流。

[0039] 远程会议包括多方用户,一方用户可以对应一个终端设备,且每一方用户的数量可以为一个,也可以为多个,本实施例不做限定。例如,远程会议包括多方用户,且多方用户分别为不同企业的多个工作人员;又如,远程会议包括两方用户,且两方用户分别为同一企业的不同部门的多个工作人员;再如,远程会议包括两方用户,且一方用户为企业的多个工作人员,另一方面用户为个体用户,等等。

[0040] 终端设备可以是移动终端,如移动电话(或称为“蜂窝”电话)和具有移动终端的计算机,例如,可以是便携式、袖珍式、手持式、计算机内置的或者车载的移动装置,它们与无线接入网交换语言和/或数据;终端设备还可以是智能音箱、个人通信业务(Personal Communication Service,PCS)电话、无绳电话、会话发起协议(Session Initiation Protocol,SIP)话机、无线本地环路(Wireless Local Loop,WLL)站、个人数字助理(Personal Digital Assistant,PDA),平板型电脑、无线调制解调器(modem)、手持设备(handset)、膝上型电脑(laptop computer)、机器类型通信(Machine Type Communication,MTC)终端等设备;终端设备也可以称为系统、订户单元(Subscriber Unit)、订户站(Subscriber Station),移动站(Mobile Station)、移动台(Mobile)、远程站(Remote Station)、远程终端(Remote Terminal)、接入终端(Access Terminal)、用户终端(User Terminal)、用户代理(User Agent)、用户设备(User Device or User Equipment),等等,在此不作限定。

[0041] 值得说明地是,上述示例只是用于示范性地说明本公开实施例的语音处理方法可能适用的应用场景,而不能理解为对应用场景的限定。例如,可以在上述示例的基础上对应用场景中的元素进行适应性的增加,如增加终端设备的数量等;又如,可以在上述示例的基础上对应用场景中的元素进行适应性的删减,如减少终端设备的数量,和/或,删减服务器,等等。

[0042] 在相关技术中,每一终端设备可以采集与所述每一终端设备对应的用户的音频信息,并根据音频信息生成通话流(用于语音通话),并可以基于所述每一终端设备与服务器之间的通信链路,向服务器发送通话流,服务器可以基于其他通信链路分别向其他通信链路各自对应的终端设备发送通话流,其他终端设备均可以输出通话流,使得其他终端设备各自对应的用户均可以听到所述每一终端设备对应的用户的声音及内容。

[0043] 然而,所述每一终端设备传输的仅包括通话流,导致远程会议中会议内容的展现维度比较单一,智能化偏低。

[0044] 值得说明地是,上述示例只是用于示范性地说明,本实施例的语音处理方法可以适用的应用场景,而不能理解为对本公开实施例的语音处理方法的应用场景地限定。如本实施例的语音处理方法还可以应用于其他会议场景(如本地会议场景);或者,应用于其他需要对音频信息进行语音处理的场景。

[0045] 本公开的发明人经过创造性地劳动,得到了本公开的发明构思:基于音频信息分

别生成通话流和识别流,通话流用于语音通话,识别流用于语音识别,从而实现用于会议的会议内容的多样性,提高用户的会议体验。

[0046] 下面以具体地实施例对本公开的技术方案以及本公开的技术方案如何解决上述技术问题进行详细说明。下面这几个具体的实施例可以相互结合,对于相同或相似的概念或过程可能在某些实施例中不再赘述。下面将结合附图,对本公开的实施例进行描述。

[0047] 请参阅图2,图2为本公开一个实施例的语音处理方法的流程示意图。

[0048] 如图2所示,该方法包括:

[0049] S101:采集会议过程中的音频信息。

[0050] 示例性地,本实施例的执行主体可以语音处理装置,且语音处理装置可以为终端设备、服务器、处理器以及芯片等,本实施例不做限定。

[0051] 例如,当本实施例的语音处理方法应用于如图1所示的应用场景时,语音处理装置可以为如图1中所示的终端设备,如图1中终端设备1至终端设备n中的至少一个。

[0052] 相应地,以终端设备n为例,当用户n发表言论时,终端设备n可以采集对应的音频信息。

[0053] S102:根据音频信息分别生成通话流和识别流,其中,通话流用于语音通话,识别流用于语音识别。

[0054] 在本实施例中,语音处理装置基于采集到的音频信息分别生成:用于语音通话的通话流和用于语音识别的识别流。结合上述示例,该步骤可以理解为:终端设备n对音频信息进行处理,分别生成通话流和识别流。

[0055] 值得说明地是,本实施例提供的包括基于音频信息分别生成通话流和识别流的技术特征的技术方案,避免了相关技术中,用于会议的会议内容较为单一,导致对端用户接收到的会议内容可能会不准确,即用户获取会议内容错误的问题,提高了会议的用户对会议内容的理解,进而提高了会议的准确性,且提高了会议的智能化和质量,还提高了用户的会议体验。

[0056] S103:发送通话流和识别流。

[0057] 结合如图1所示的应用场景,若语音处理装置为终端设备n,则在一种可能实现的技术方案中,终端设备n可以向服务器分别发送通话流和识别流,服务器可以向终端设备1发送通话流和识别流。相应地,终端设备1输出通话流,用户1可以听到与通话流对应的远程会议的语音内容,即用户1可以听到用户n的发言内容;且终端设备1输出与识别流对应的文本内容,即用户1可以看到与用户n的发言内容。

[0058] 基于上述分析可知,本实施例提供了一种语音处理方法,该方法包括:采集会议过程中的音频信息,根据音频信息分别生成通话流和识别流,其中,通话流用于语音通话,识别流用于语音识别,发送通话流和识别流,通过既生成用于语音通话的通话流,又生成用于语音识别的识别流,可以避免相关技术中,对音频信息进行处理,得到用于表征会议的会议内容的方式较为单一,实现了确定出的音频信息对应的会议内容较多、较丰富,进而提高了会议的准确性,且提高了会议的智能化和质量,还提高了用户的会议体验。

[0059] 请参阅图3,图3为本公开另一实施例的语音处理方法的流程示意图。

[0060] 如图3所示,该方法包括:

[0061] S201:采集会议过程中的音频信息。

[0062] 为使读者更加深刻地理解本实施例的技术方案,以及本实施例的技术方案与相关技术方案的区别等,现结合图4和图5,对图3所示的语音处理方法进行更为详细地阐述。其中,图4为本公开另一实施例的语音处理方法的应用场景示意图,图5为本公开实施例的语音处理方法的原理示意图。

[0063] 如图4所示,应用场景包括:智能终端、第一终端设备、第二终端设备以及云端服务器,第一终端设备为第一与会用户与第二与会用户进行远程会议的设备,智能终端和第二终端设备为,第二与会用户与第一与会用户进行远程会议的设备。

[0064] 值得说明地是,在如图4所示的应用场景中,智能终端和第二终端设备为两个独立的设备,而在另一些实施例中,智能终端可以集成于第二终端设备中,且对智能终端的外在展现形式,本实施例不做限定。

[0065] 结合如图4所示的应用场景,该步骤可以理解为:当第二与会用户发表言论时,智能终端设备可以采集相应的音频信息。

[0066] 结合图5可知,在一种可能实现的方案中,智能终端中可以设置一个麦克风或者麦克风阵列,并通过一个麦克风或者麦克风阵列采集音频信息。

[0067] 值得说明地是,麦克风阵列中的麦克风的数量可以基于需求、历史记录和试验等进行设置,例如,麦克风的数量为6个。

[0068] S202:对音频信息的信号类型进行转换,其中,信号类型包括模拟信号和数字信号,且转换前的音频信息的信号类型为模拟信号,转换后的音频信息的信号类型为数字信号。

[0069] 结合图5可知,在一种可能实现的方案中,智能终端中可以设置模数转换器,麦克风阵列将采集到的模拟信号的音频信息发送给模数转换器,模数转换器将模拟信号的音频信息转换为,数字信号的音频信息,以便提高后续对音频信息的处理效率和准确性。

[0070] S203:对转换后的音频信息进行回声消除处理,得到残差信号。

[0071] 结合图5和图6可知,在一种可能实现的方案中,智能终端中可以设置处理器,处理器与模数转换器连接,用于接收由模数转换器发送的转换后的音频信息,且处理器可以对转换后的音频信息进行回声消除处理,其中,图6为图5所示的处理器原理图。

[0072] 一个示例中,回声消除处理的方法可以包括:确定音频信息对应的回声信号,根据获取到的参考信号对回声信号进行消除处理,得到残差信号。

[0073] 例如,可以根据麦克风阵列和智能终端的扬声器,估计音频信息对应的回声路径,根据回声路径和获取到的参考信号(如从扬声器中的功率放大器处获取到的参考信号),估计麦克风阵列接收到的回声信号,计算参考信号与回声信号之间的差值,该差值即为残差信号,且该残差信号即消除过回声的信号。

[0074] 另一个示例中,回声消除处理的方法还可以包括:处理器中可以设置自适应滤波器,自适应滤波器可以估计一个近似的回声路径来逼近真实回声路径,从而得到估计的回声信号,并在纯净语音和回声的混合信号中除去回声信号来实现回声的消除,且自适应滤波器具体可以为FIR滤波器(Finite Impulse Response)。

[0075] S204:对残差信号进行回声残留抑制处理,得到抑制残留回声的信号。

[0076] 在一些实施例中,回声残留抑制处理的方法可以包括:对残差信号进行傅立叶变换,得到频域信号,确定频域信号对应的频域调整参数,根据频域调整参数对频域信号进行

调整,并对调整后的频域信号进行逆傅立叶变换,得到抑制残留回声的信号。

[0077] 例如,处理器中可以预先设置深度学习神经网络,处理器对残差信号作傅立叶变换,得到频域信号,将频域信号送入给深度学习神经网络,深度学习神经网络输出频域上的掩码(掩码表征,频域的背景噪声的概率),在频域信号的基础上乘以掩码,得到处理后的频域信号,对处理后的频域信号进行逆傅立叶变换,得到抑制残留回声的信号。

[0078] S205:对抑制残留回声的信号进行去混响处理,得到去混响后的信号。

[0079] 一个示例中,去混响处理的方法可以包括:构建多通道线性预测模型(Multichannel Linear Prediction,MCLP),多通道线性预测模型表征,抑制残留回声的信号由当前信号(即抑制残留回声的信号)与过去若干帧的信号线性组合而成,基于多通道线性预测模型对过去若干帧的信号进行卷积,可以得到当前信号中的混响部分的信号,通过当前信号减去混响部分的信号即可得到去混响后的信号。

[0080] 另一个示例中,去混响处理的方法还可以包括:确定麦克风阵列中每一麦克风对应的频率倒谱系数(Mel Frequency Cepstrum Coefficient,MFCC),并确定相邻麦克风之间的频率倒谱系数差,且基于频率到普系数差构建去混响后的信号。

[0081] S206:根据不同的处理方式对去混响后的信号进行处理,获得通话流和识别流。

[0082] 在一些实施例中,S206可以包括:对去混响后的信号进行增强清晰度处理,获得通话流;以及,对去混响后的信号进行保真处理,获得识别流。

[0083] 值得说明地是,在如图5所示的原理图中,处理器可以包括预处理器、增强清晰度处理器以及保真处理器,其中,预处理器用于执行回声消除处理、回声残留抑制处理以及去混响处理的预处理器,增强清晰度处理器用于对由预处理器进行处理后的信号进行增强清晰度处理,保真处理器用于对由预处理器进行处理后的信号进行保真处理。

[0084] 示例性地,关于增强清晰度处理的描述如下:

[0085] 一个示例中,增强清晰度处理可以包括:基本谱减法。其中,基本谱减法可以理解为:预先设置基本频域,并对基本频域外的去混响后的信号进行剔除。

[0086] 另一个示例中,降噪处理可以包括:维纳滤波降噪。其中,维纳滤波降噪可以理解为:基于预设的均方误差训练滤波器,并基于滤波器对去混响后的信号进行过滤,使得过滤后的去混响后的信号与纯净的去混响后的信号之间的误差小于预设的误差阈值。

[0087] 结合图6可知,再一个示例中,增强清晰度处理可以包括:依次对去混响后的信号进行波束处理、合成降噪处理、最小波束处理、抑制降噪处理、人声均衡处理以及自动增益控制,获得通话流。

[0088] 示例性地,波束处理的方法可以包括:确定与去混响后的信号对应的多组波束信号。

[0089] 例如,建立广义旁瓣相消器(General sidelobe canceller,GSC)模型,将去混响后的信号输入至广义旁瓣相消器模型,输出水平空间的多组波束信号。

[0090] 示例性地,合成降噪处理的方法可以包括:根据多组波束信号确定去混响后的信号(即音频信息的纯净信号)的期望估计,对去混响后的信号的期望估计、以及多组波束进行相位合成,得到降噪后的波束信号。

[0091] 例如,对去混响后的信号的幅值谱建模,建设得到的语音和噪声的幅度谱服从高斯分布,获取会议的稳态噪声,估计去混响后的信号的后验信噪比,根据贝叶斯原理和后验

信噪比,得到去混响后的信号的期望估计,对去混响后的信号的期望估计与多组波束的相位进行合成处理,得到降噪后的波束信号。

[0092] 示例性地,最小波束处理的方法可以包括:确定降噪后的波束信号中最大能量的波束信号与最小能量的波束信号之间的能量比值,根据能量比值确定归一化的波束信号。

[0093] 例如,确定降噪后的波束信号中的最大能量的波束信号,并确定降噪后的波束信号中的最小能量的波束信号,计算最大能量与最小能量之间的能量比值,判断能量比值是否大于预设的比值阈值,若是,则将最大能量的波束信号以归一化的方式累加处理,得到归一化的波束信号。

[0094] 示例性地,抑制降噪处理的方法可以包括:确定归一化的波束信号的掩码,根据归一化的波束信号的掩码,对归一化的波束信号的非稳态噪声进行抑制,获得抑制后的波束信号。

[0095] 例如,可以预先设置循环神经网络,将归一化的波束信号输出给循环神经网络,循环神经网络输出归一化的波束信号的掩码,可以将归一化的波束信号的掩码(掩码表征,非稳态噪声为背景噪声的概率)乘以非稳态噪声,得到抑制后的波束信号。

[0096] 示例性地,人声均衡处理的方法可以包括:对抑制后的波束信号在预设频带进行补偿,获得补偿后的波束信号。

[0097] 例如,可以预先设置分段的峰值滤波器(peak filter),对于降噪输出的抑制后的波束信号在预设频带(可以基于需求、历史记录和试验等进行设置,本实施例不做限定)进行补偿,得到补偿后的波束信号,使得补偿后的波束信号对应的听感的音质更高。

[0098] 一个示例中,自动增益控制的方法可以包括:将补偿后的波束信号作傅立叶变换,得到的功率谱,将功率谱输入给预先设置的卷积神经网络,得到当前帧的语音存在概率,若基于当前帧的语音存在概率大于预设概率阈值,则确定当前帧的语音存在,则给补偿后的波束信号施加一个逐渐增大的增益,直到补偿后的波束信号的增益稳定,得到通话流。

[0099] 另一个示例中,自动增益控制的方法可以包括如下步骤:

[0100] 步骤1:根据补偿后的波束信号和预先设置的等响度曲线,确定增益权重。

[0101] 其中,等响度曲线可以用于,表征基于实验等方式确定出的,用户满足度相对较高的补偿后的波束信号对应的曲线。

[0102] 在该步骤中,具体可以将补偿后的波束信号映射至等响度曲线,并基于两者之间的差异确定增益权重。

[0103] 步骤2:根据增益权重对补偿后的波束信号进行增强处理。

[0104] 示例性地,关于保真处理的描述如下:

[0105] 一个示例中,保真处理的方法可以包括:对去混响后的信号进行声纹识别处理。

[0106] 例如,智能终端对去混响后的信号进行特征提取处理,得到与去混响后的信号的音高、音强、音长以及音色等特征,并基于音高、音强、音长以及音色等特征对去混响后的信号进行还原,得到识别流,使得识别流具有较低失真的特性。

[0107] 结合图6可知,另一个示例中,保真处理的方法可以包括:对去混响后的信号进行波达角估计处理和波束选择处理。

[0108] 示例性地,波达角估计处理的方法可以包括:对去混响后的信号进行多重信号分类处理,获得方向谱,根据方向谱确定去混响后的信号对应的声源方向。

[0109] 例如,对去混响后的信号进行多重信号分类处理,得到去混响后的信号的频率和时间的方向谱,可以根据频率和时间构建方向谱对应的直方图,并基于直方图确定去混响后的信号的声源方向。

[0110] 示例性地,波束选择处理的方法可以包括:根据声源方向确定去混响后的信号的起始点、结束点以及可控功率响应,根据起始点、结束点以及可控功率响应,从去混响后的信号中选择识别流。

[0111] 值得说明地是,在采用上述方法对音频信息进行处理,获得通话流和识别流时,可以仅采用部分方法对音频信息进行处理,且对音频信息进行处理的各方法的顺序可以相应地调整。

[0112] 例如,在对音频信息进行增强清晰度处理时,可以仅采用降噪处理和自动增益处理;又如,在对音频信息进行保真处理时,可以仅采用波束选择处理;再如,在采用如图6所示的各方法对音频信息进行增强清晰度处理时,可以随意调整各处理方法的顺序,如先进行抑制降噪处理,而后进行最小波束处理,等等。

[0113] S207:发送通话流和识别流。

[0114] 结合如图4所示的应用场景和如图5所示的原理图可知,一个示例中,智能终端可以通过通信模块向云端服务器发送通话流,相应地,云端服务器可以向第一终端设备分发送通话流,相应地,第一终端设备可以基于通话流进行语音播报;智能终端可以向云端服务器发送识别流,相应地,云端服务器可以向第一终端设备发送识别流,相应地,第一终端设备可以基于识别流进行文本显示。

[0115] 其中,也可以由云端服务器基于识别流进行语音识别,得到识别结果(即转写文本),并将识别流和/或转写文本发送给第一终端设备,第一终端设备可以对转写文本进行文本显示,当然,也可以由云服务器对识别流和/或转写文本进行存储。

[0116] 如图5所示,在一些实施例中,云服务器还可以向第二终端设备发送识别流和/或转写文本,相应地,第二终端设备可以对转写文本进行文本显示。

[0117] 如图5所示,在一些实施例中,云服务器还可以向第三终端设备发送识别流和/或转写文本,相应地,第二终端设备可以对转写文本进行文本显示。以图4所示的应用场景为例,第三终端设备可以为非远程会议中的终端设备。也就是说,第三终端设备为可以具有显示功能的设备,并可以对转写文本进行显示,且本实施例对第三终端设备的数量不做限定。

[0118] 另一个示例中,智能终端可以通过通信模块向第二终端设备发送通话流,且第二终端设备上运行有进行会议的软件,相应地,第二终端设备可以基于会议的软件向第一终端设备发送通话流,相应地,第一终端设备可以基于通话流进行语音播报;关于智能终端发送识别流的原理可以参见上述示例,此处不再赘述。

[0119] 另一个示例中,可以在如图4所示的应用场景中增加一个服务器的配置,且删减图4中第二终端设备的配置,智能终端可以向增加的服务器发送通话流,相应地,增加的服务器可以向第一终端设备发送通话流,相应地,第一终端设备可以基于通话流进行语音播报;如上述示例所述,智能终端可以向云端服务器发送识别流,相应地,云端服务器可以向第一终端设备发送识别流,相应地,第一终端设备可以基于识别流进行文本显示。

[0120] 同理,也可以由云端服务器基于识别流进行语音识别,得到识别结果(即转写文本),并将识别流和/或转写文本发送给第一终端设备,第一终端设备可以对转写文本进行

文本显示,当然,也可以由云服务器对识别流和/或转写文本进行存储。

[0121] 另一个示例中,可以删减图4中云端服务器的配置,如智能终端可以向第二终端设备发送通话流和识别流,相应地,第二终端设备可以向第一终端设备发送通话流和识别流,相应地,第一终端设备可以基于通话流进行语音播报,并基于识别流确定转写文本,并基于转写文本进行文本显示。

[0122] 再一个示例中,可以删减图4中第二终端设备的配置,如智能终端可以向云端服务器发送通话流和识别流,相应地,云端服务器可以向第一终端设备发送通话流和识别流,相应地,第一终端设备可以基于通话流进行语音播报,并基于识别流确定转写文本,并基于转写文本进行文本显示。

[0123] 还有一个示例中,智能终端可以向第二终端设备发送识别流,相应地,第二终端设备可以向第一终端设备发送识别流,相应地,第一终端设备可以基于识别流确定转写文本,并基于转写文本进行文本显示;智能终端可以向云端服务器发送通话流,相应地,云端服务器可以向第一终端设备发送通话流,相应地,第一终端设备可以基于通话流进行语音播报。

[0124] 又一个示例中,智能终端可以向第二终端设备发送通话流和识别流,相应地,第二终端设备可以向云端服务器发送通话流和识别流,相应地,云端服务器可以向第一终端设备发送通话流和识别流,相应地,第一终端设备可以基于通话流进行语音播报,并基于识别流确定转写文本,并基于转写文本进行文本显示。

[0125] 在一些实施例中,通信模块可以包括:串行总线接口(USB)、无线保真WiFi以及蓝牙。

[0126] 示例性地,智能终端可以基于通用串行总线接口、无线保真以及蓝牙中的任一种与第二终端设备连接;智能终端可以基于无线保真与云端服务器连接;第二终端设备可以基于无线保真与云端服务器连接。

[0127] S208:对识别流进行编码处理和压缩处理,并对处理后的识别流进行存储。

[0128] 结合如图5所示的原理图,智能终端中可以设置存储器,且存储器可以与处理器连接,一个示例中,存储器可以接收由处理器发送的识别流,并依次对识别流进行编码处理、压缩处理以及存储;另一个示例中,存储器可以接收由处理器进行编码处理和压缩处理后的识别流,并对接收到的处理后的识别流进行存储。

[0129] 值得说明地是,在本实施例中,通过对处理后的识别流进行存储,可以避免由会议记录员进行手动记录造成的成本高,且可靠性偏低的问题,使得会议中的发言内容自动被记录,并方便后续查询和追溯,提高会议的智能化,且提高与会人员的会议体验。

[0130] 应该理解地是,会议是各与会方相互交流的过程,因此,在一些实施例中,智能终端可以接收由第一终端设备发送的通话流,或者通话流和识别流,且当智能终端接收到第一终端设备发送的通话流时,可以基于通话流进行语音播放,当智能终端接收到第一终端设备发送的识别流时,还可以基于识别流进行文本显示。

[0131] 例如,结合如图4所示的应用场景和图5所示的原理图,第一与会用户可以借助于第一终端设备发表言论,则第一终端设备可以采集对应的音频信息,并可以生成通话流,相应地,智能终端中可以设置扬声器,且智能终端可以通过扬声器对通话流进行语音播放。

[0132] 另一些实施例中,第一终端设备也可以基于上述方法生成通话流和识别流,并将通话流和识别流均发送给第二终端设备,第二终端设备可以将通话流发送给智能终端,由

智能终端进行语音播放,并由第二终端设备基于识别流进行文本显示。

[0133] 在一些实施例中,智能终端与第一终端设备可以直接交互,而无需第二终端设备的中间转发,且智能终端中还可以设置显示器,一方面,智能终端可以通过扬声器进行语音播放,另一方面,智能终端可以通过显示器进行文本显示。

[0134] 示例性地,显示器可以用于表征对文本进行显示的设备,如液晶显示器(Liquid Crystal Display,LCD)、发光二极管(Light Emitting Diode,LED)显示器及有机发光(Organic Light Emitting Display,OLED)显示器,等等,本申请实施例不做限定。

[0135] 根据本公开实施例的另一个方面,本公开实施例还提供了一种智能终端。

[0136] 结合图5可知,智能终端可以包括:麦克风阵列、处理器和通信模块(图中未示出);

[0137] 所述麦克风阵列,用于采集会议过程中的音频信息;

[0138] 所述处理器,用于根据所述音频信息分别生成通话流和识别流,其中,所述通话流用于语音通话,所述识别流用于语音识别;

[0139] 所述通信模块,用于发送所述通话流和所述识别流。

[0140] 在一些实施例中,所述处理器用于,根据不同的处理方式对所述音频信息进行处理,获得所述通话流和所述识别流。

[0141] 在一些实施例中,所述处理器用于,对所述音频信息进行增强清晰度处理,获得所述通话流;以及,对所述音频信息进行保真处理,获得所述识别流。

[0142] 在一些实施例中,所述处理器用于,对所述音频信息进行降噪处理和自动增益控制,获得所述通话流。

[0143] 在一些实施例中,所述处理器用于,对所述音频信息进行波束选择处理,获得所述识别流。

[0144] 在一些实施例中,所述处理器用于,对所述音频信息进行回声消除处理。

[0145] 结合图5可知,在一些实施例中,所述智能终端还包括:

[0146] 扬声器,用于对参与所述会议的第一终端设备发送的通话流进行语音播报。

[0147] 结合图5可知,在一些实施例中,所述智能终端还包括:

[0148] 模数转换器,用于对所述音频信息的信号类型进行转换,获得转换后的音频信息,其中,所述转换后的音频信息的信号类型为数字信号。

[0149] 在一些实施例中,所述处理器用于,对所述转换后的音频信息进行回声消除处理。

[0150] 结合图5可知,在一些实施例中,所述音频设备还包括:

[0151] 存储器,用于对所述识别流进行存储。

[0152] 在一些实施例中,所述处理器用于,对所述识别流进行编码处理和压缩处理;

[0153] 所述存储器用于,对处理后的识别流进行存储。

[0154] 在一些实施例中,所述收发器包括:通用串行总线接口、无线保真以及蓝牙中的任一种。

[0155] 根据本公开实施例的另一个方面,本公开实施例还提供了一种语音处理装置。

[0156] 请参阅图7,图7为本公开一个实施例的语音处理装置的示意图。

[0157] 如图7所示,该装置包括:

[0158] 采集模块11,用于采集会议过程中的音频信息;

[0159] 生成模块12,用于根据所述音频信息分别生成通话流和识别流,其中,所述通话流

用于语音通话,所述识别流用于语音识别;

[0160] 发送模块13,用于发送所述通话流和所述识别流。

[0161] 在一些实施例中,所述生成模块12用于,根据不同的处理方式对所述音频信息进行处理,获得所述通话流和所述识别流。

[0162] 在一些实施例中,所述生成模块12用于,对所述音频信息进行增强清晰度处理,获得所述通话流;以及,对所述音频信息进行保真处理,获得所述识别流。

[0163] 在一些实施例中,所述生成模块12用于,对所述音频信息进行降噪处理和自动增益控制,获得所述通话流。

[0164] 在一些实施例中,所述生成模块12用于,对所述音频信息进行波束选择处理,获得所述识别流。

[0165] 在一些实施例中,所述生成模块12用于,对所述音频信息进行回声消除处理。

[0166] 结合图8可知,在一些实施例中,所述音频信息的信号类型为模拟信号;所述装置还包括:转换模块14,用于对所述音频信息的信号类型进行转换,获得转换后的音频信息,其中,所述转换后的音频信息的信号类型为数字信号。

[0167] 结合图8可知,在一些实施例中,所述装置还包括:存储模块15,用于对所述识别流进行存储。

[0168] 在一些实施例中,存储模块15用于,对所述识别流进行编码处理和压缩处理,并对处理后的识别流进行存储。

[0169] 根据本公开实施例的另一个方面,本公开实施例还提供了一种电子设备以及存储介质。

[0170] 参考图9,其示出了适于用来实现本公开实施例的电子设备900的结构示意图,该电子设备900可以为终端设备或服务器。其中,终端设备可以包括但不限于诸如智能音箱、移动电话、笔记本电脑、数字广播接收器、个人数字助理(Personal Digital Assistant,简称PDA)、平板电脑(Portable Android Device,简称PAD)、便携式多媒体播放器(Portable Media Player,简称PMP)、车载终端(例如车载导航终端)等等的移动终端以及诸如数字TV、台式计算机等等的固定终端。图9示出的电子设备仅仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0171] 如图9所示,电子设备900可以包括处理装置(例如中央处理器、图形处理器等)901,其可以根据存储在只读存储器(Read Only Memory,简称ROM)902中的程序或者从存储装置908加载到随机访问存储器(Random Access Memory,简称RAM)903中的程序而执行各种适当的动作和处理。在RAM 903中,还存储有电子设备900操作所需的各种程序和数据。处理装置901、ROM 902以及RAM 903通过总线904彼此相连。输入/输出(I/O)接口905也连接至总线904。

[0172] 通常,以下装置可以连接至I/O接口905:包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置906;包括例如液晶显示器(Liquid Crystal Display,简称LCD)、扬声器、振动器等的输出装置907;包括例如磁带、硬盘等的存储装置908;以及通信装置909。通信装置909可以允许电子设备900与其他设备进行无线或有线通信以交换数据。虽然图9示出了具有各种装置的电子设备900,但是应理解的是,并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

[0173] 特别地,根据本公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信装置909从网络上被下载和安装,或者从存储装置908被安装,或者从ROM902被安装。在该计算机程序被处理装置901执行时,执行本公开实施例的方法中限定的上述功能。

[0174] 需要说明的是,本公开上述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中,计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:电线、光缆、RF(射频)等等,或者上述的任意合适的组合。

[0175] 上述计算机可读介质可以是上述电子设备中所包含的;也可以是单独存在,而未装配入该电子设备中。

[0176] 上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被该电子设备执行时,使得该电子设备执行上述实施例所示的方法。

[0177] 可以以一种或多种程序设计语言或其组合来编写用于执行本公开的操作的计算机程序代码,上述程序设计语言包括面向对象的程序设计语言——诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言——诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(Local Area Network,简称LAN)或广域网(Wide Area Network,简称WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0178] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注

意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0179] 描述于本公开实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。其中,单元的名称在某种情况下并不构成对该单元本身的限定,例如,第一获取单元还可以被描述为“获取至少两个网际协议地址的单元”。

[0180] 本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如,非限制性地,可以使用的示范类型的硬件逻辑部件包括:现场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、片上系统(SOC)、复杂可编程逻辑设备(CPLD)等等。

[0181] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0182] 以上描述仅为本公开的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本公开中所涉及的公开范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述公开构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

[0183] 此外,虽然采用特定次序描绘了各操作,但是这不应理解为要求这些操作以所示出的特定次序或以顺序次序执行来执行。在一定环境下,多任务和并行处理可能是有利的。同样地,虽然在上面论述中包含了若干具体实现细节,但是这些不应被解释为对本公开的范围的限制。在单独的实施例的上下文中描述的某些特征还可以组合地实现在单个实施例中。相反地,在单个实施例的上下文中描述的各种特征也可以单独地或以任何合适的子组合的方式实现在多个实施例中。

[0184] 尽管已经采用特定于结构特征和/或方法逻辑动作的语言描述了本主题,但是应当理解所附权利要求书中所限定的主题未必局限于上面描述的特定特征或动作。相反,上面所描述的特定特征和动作仅仅是实现权利要求书的示例形式。

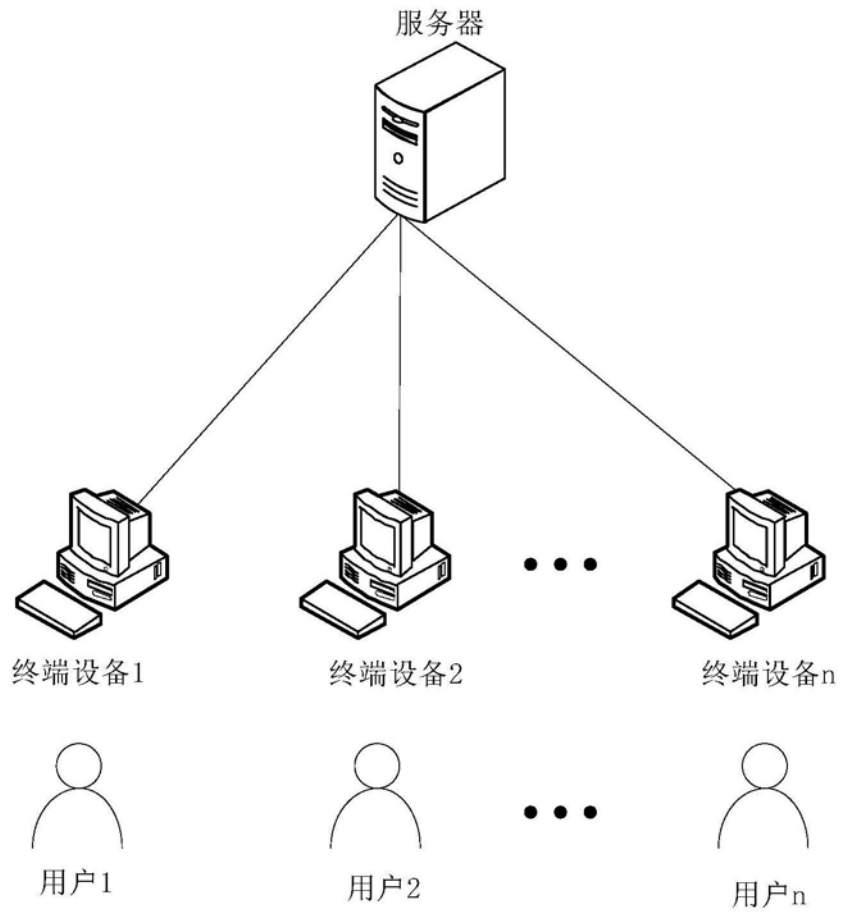


图1

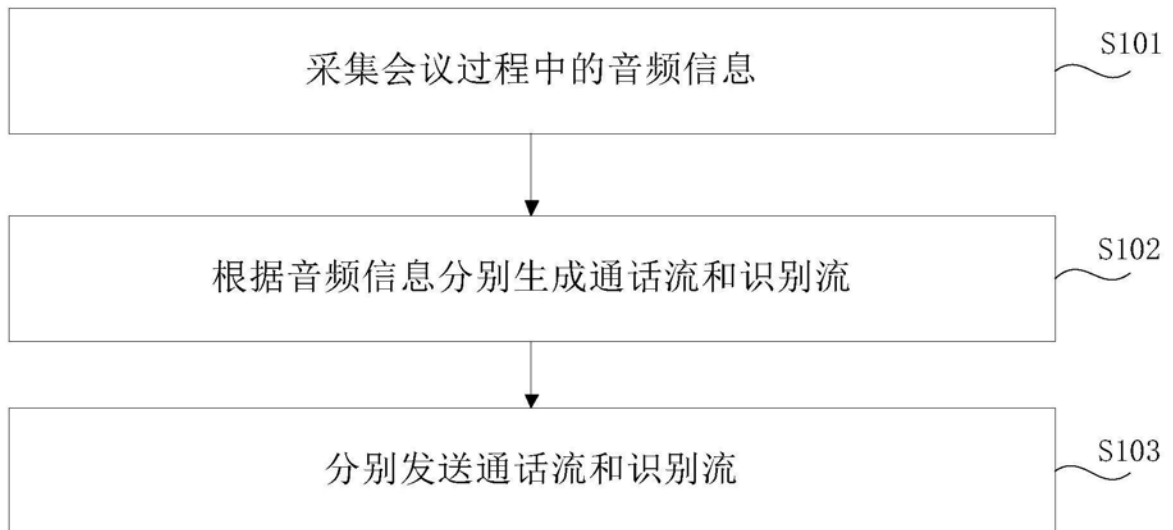


图2

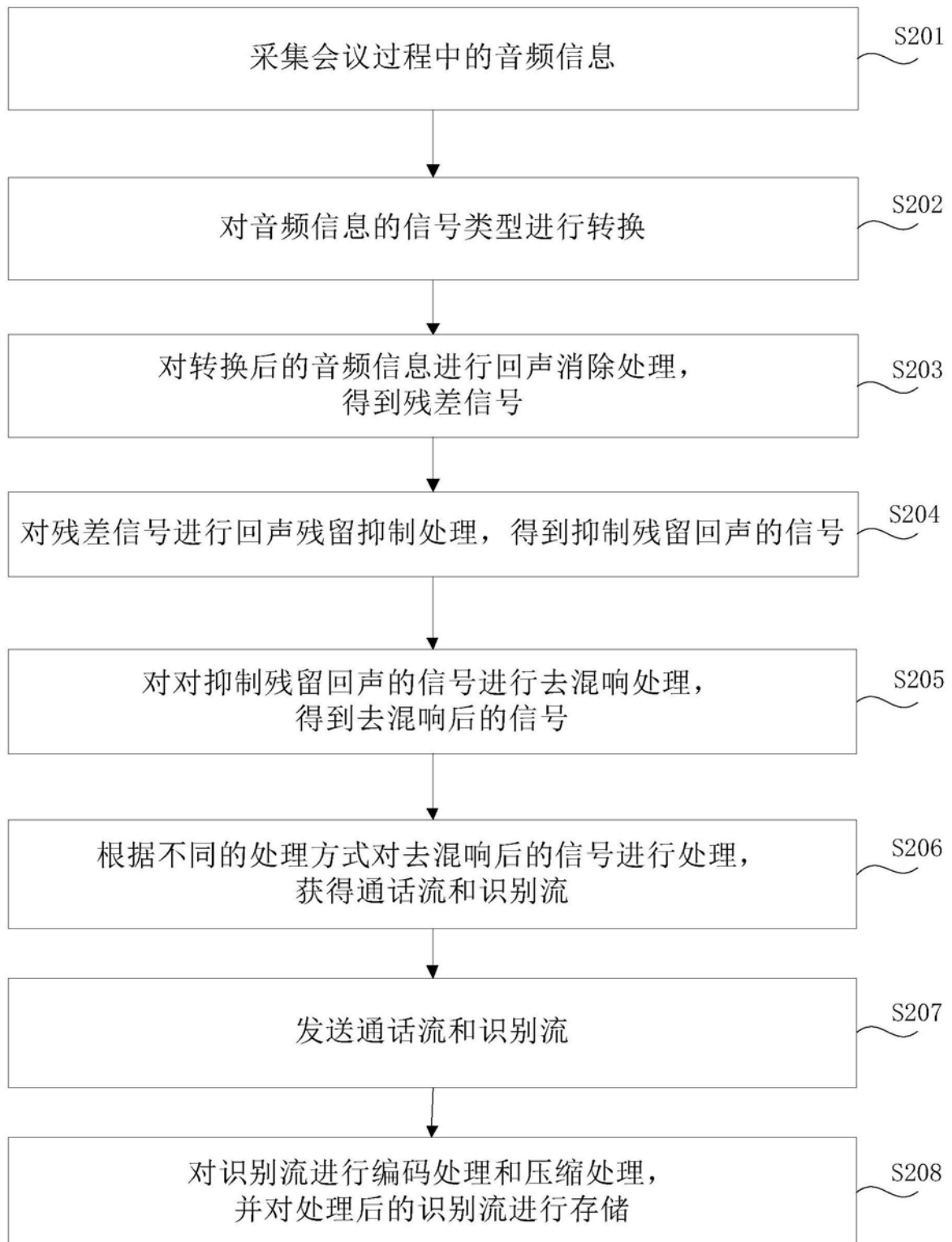


图3

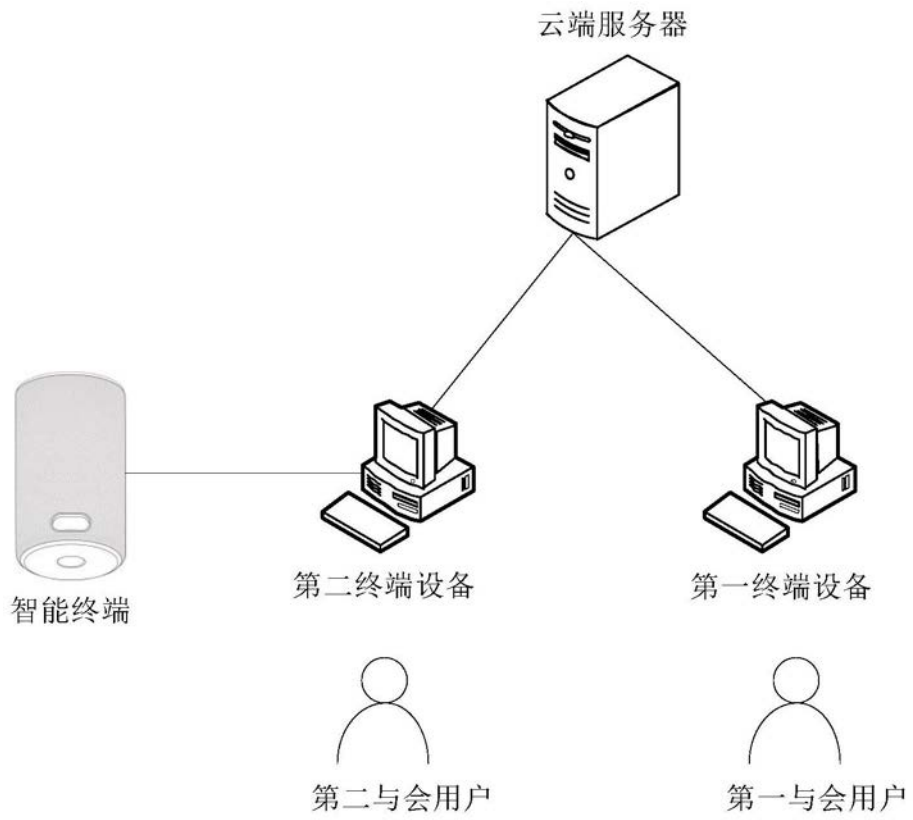


图4

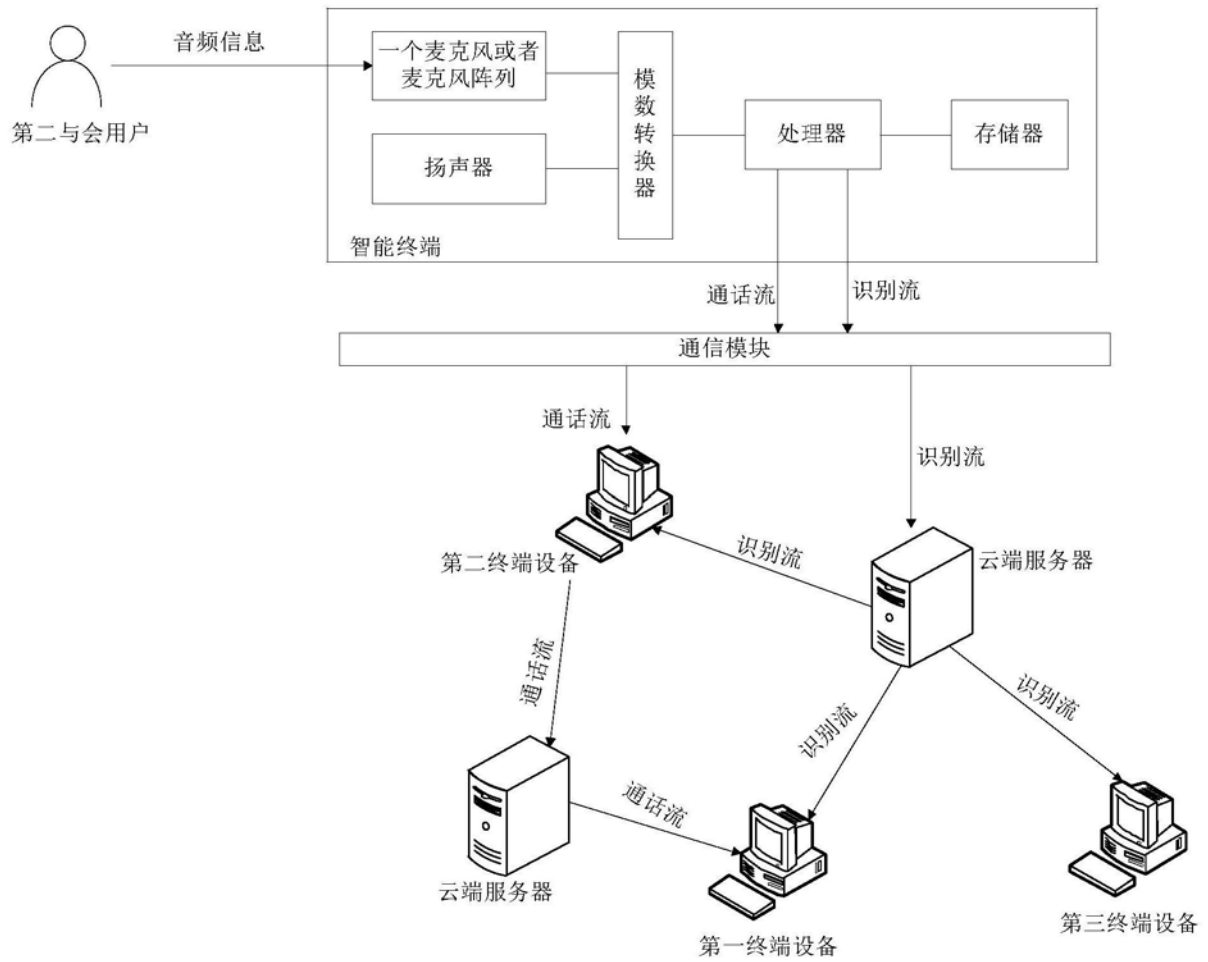


图5

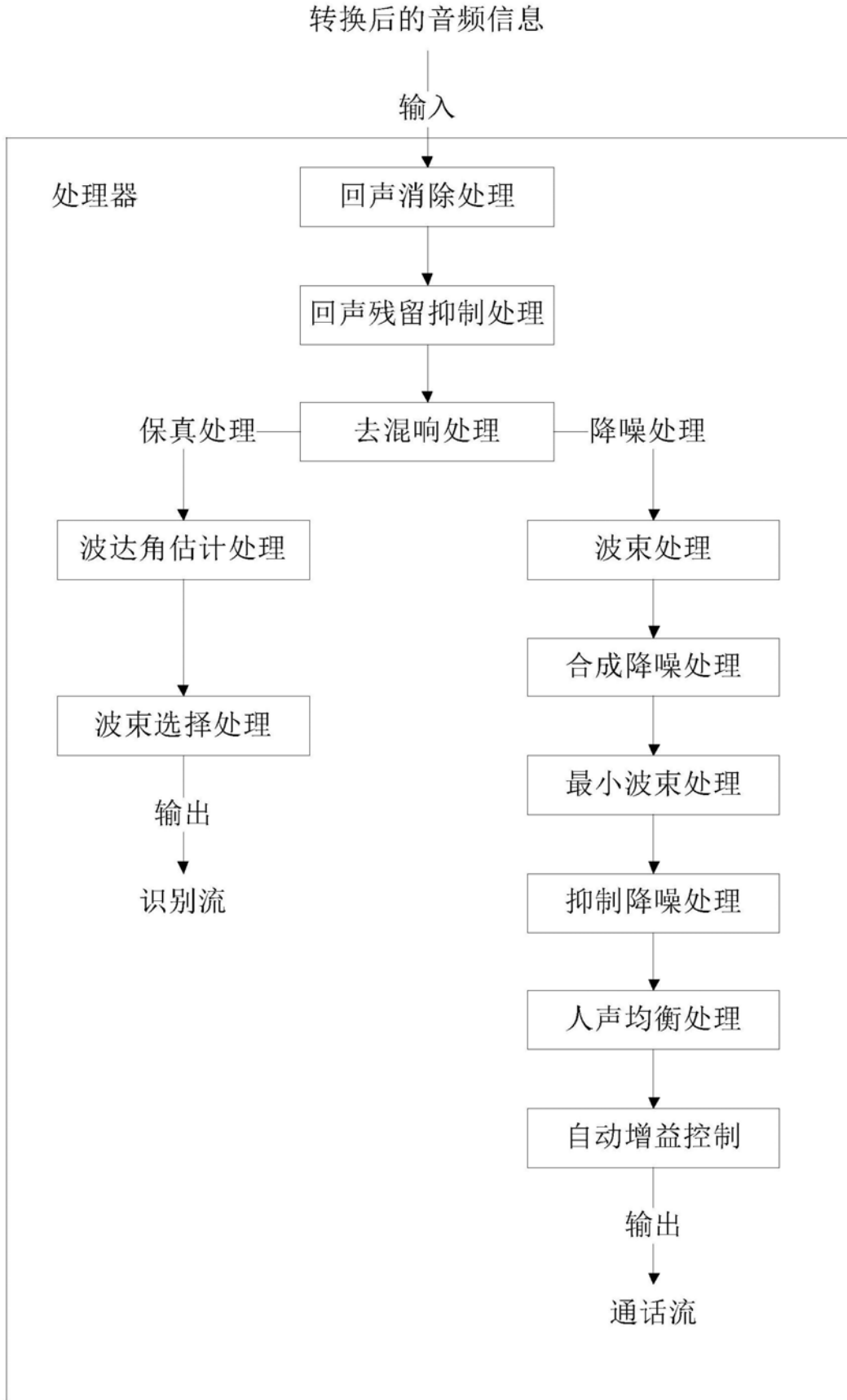


图6

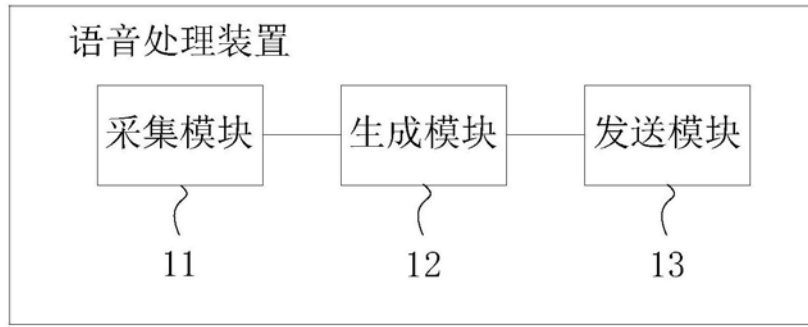


图7

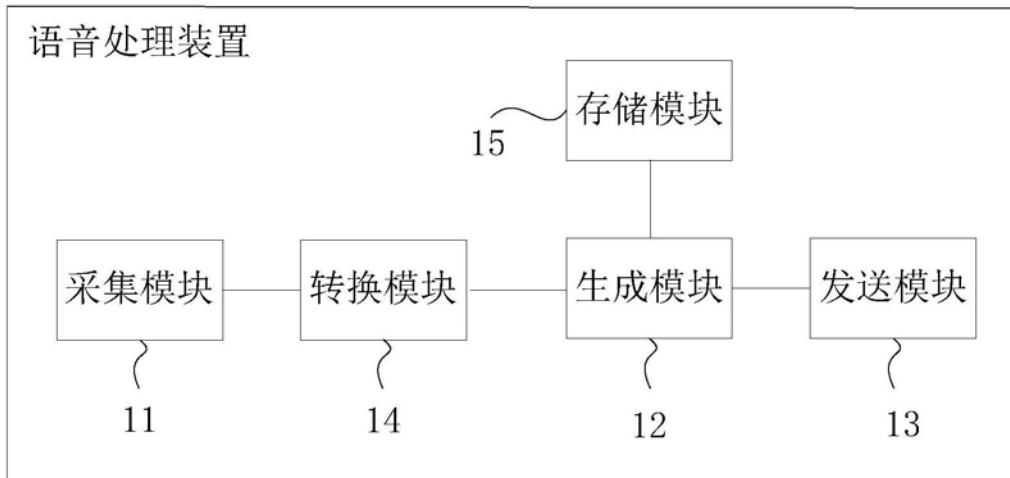


图8

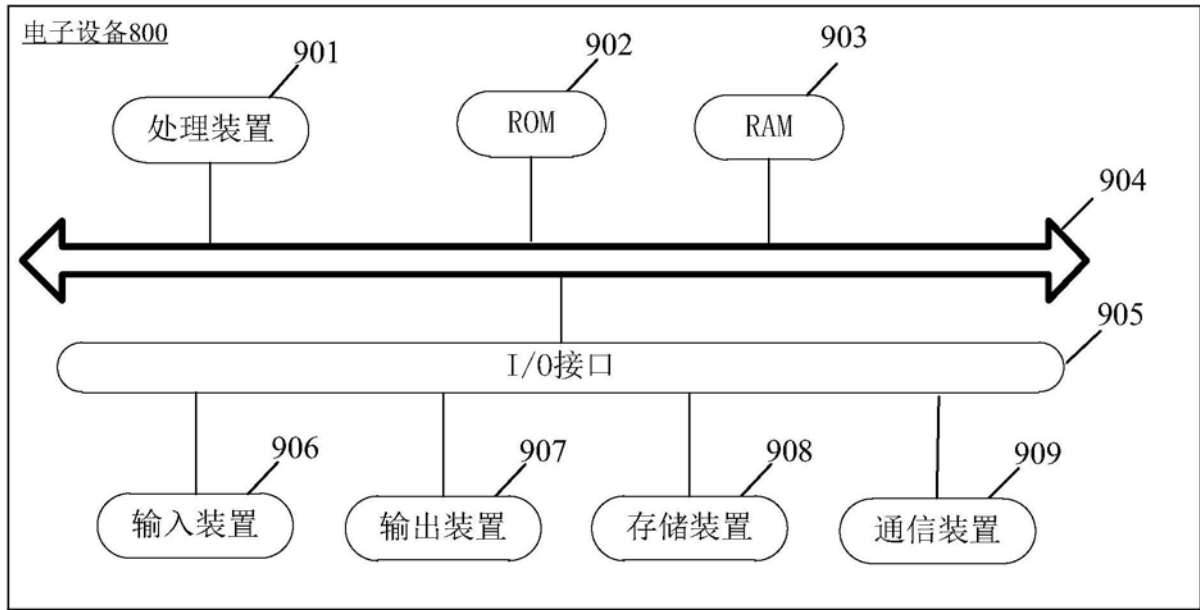


图9