



(12)发明专利申请

(10)申请公布号 CN 109215659 A

(43)申请公布日 2019.01.15

(21)申请号 201710531700.7

(22)申请日 2017.06.30

(71)申请人 北京国双科技有限公司

地址 100086 北京市海淀区双榆树小区知春路76号翠宫饭店8层A间

(72)发明人 石鹏 梁文波

(74)专利代理机构 北京康信知识产权代理有限公司 11240

代理人 赵囡囡 褚敏

(51) Int. Cl.

G10L 15/26(2006.01)

H04L 29/08(2006.01)

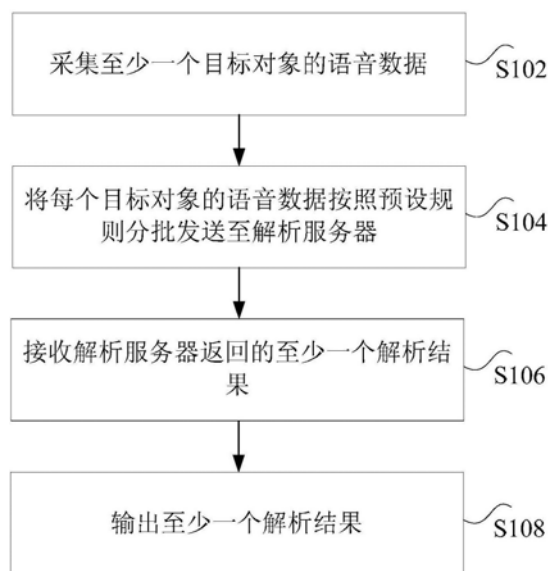
权利要求书2页 说明书11页 附图7页

(54)发明名称

语音数据的处理方法、装置和系统

(57)摘要

本发明公开了一种语音数据的处理方法、装置和系统。其中,该方法包括:采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;将每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,解析服务器用于将接收到的语音数据解析为对应的文字数据;接收解析服务器返回的至少一个解析结果;输出至少一个解析结果。本发明解决了现有的庭审过程中采用人工记录的方式将人员发言转为文字导致时效性和准确度不高的技术问题。



1. 一种语音数据的处理方法,其特征在于,包括:
 - 采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;
 - 将所述每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,所述解析服务器用于将接收到的语音数据解析为对应的文字数据;
 - 接收所述解析服务器返回的至少一个解析结果;
 - 输出所述至少一个解析结果。
2. 根据权利要求1所述的方法,其特征在于,将每个目标对象的语音数据按照预设规则分批发送至解析服务器,包括:
 - 获取传输每个目标对象的语音数据的声道的波形文件;
 - 将所述波形文件按照预设音频时长进行分段处理,得到多个子波形文件;
 - 将所述多个子波形文件依次存储至与所述声道对应的发送队列中;
 - 将所述发送队列中的子波形文件依次发送至所述解析服务器。
3. 根据权利要求2所述的方法,其特征在于,将所述波形文件按照预设音频时长进行分段处理,得到多个子波形文件,包括:
 - 基于所述预设音频时长,从所述波形文件上依次截取所述预设音频时长的多个子波形文件,其中,后一次截取的子波形文件的开始时间为前一次截取的子波形文件的结束时间。
4. 根据权利要求2所述的方法,其特征在于,将所述波形文件按照预设音频时长进行分段处理,得到多个子波形文件,包括:
 - 基于所述预设音频时长,从所述波形文件上按照所述预设音频时长的递增倍数依次截取多个子波形文件,其中,后一次截取的子波形文件包含前一次截取的子波形文件,且所述后一次截取的子波形文件的开始时间与所述前一次截取的子波形文件的开始时间相同。
5. 根据权利要求4所述的方法,其特征在于,在接收所述解析服务器返回的至少一个解析结果之后,所述方法还包括:
 - 通过所述后一次截取的子波形文件对应的解析结果对前一次截取的子波形文件对应的解析结果进行校正。
6. 根据权利要求2所述的方法,其特征在于,在获取传输每个目标对象的语音数据的声道的波形文件之前,所述方法还包括:
 - 设置所述每个目标对象与声道的对应关系。
7. 根据权利要求1所述的方法,其特征在于,所述解析结果包括:所述目标对象的标识信息和所述语音数据对应的文字数据,其中,在接收所述解析服务器返回的至少一个解析结果之后,所述方法还包括:
 - 按照所述目标对象的标识信息,将解析得到的文字数据存储至对应声道的文字队列中。
8. 根据权利要求7所述的方法,其特征在于,输出所述至少一个解析结果,包括:
 - 按照每个目标对象对应的声道的文字队列中存储的文字数据的顺序,输出所述每个目标对象的语音数据对应的文字数据。
9. 一种语音数据的处理装置,其特征在于,包括:
 - 采集单元,用于采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;

发送单元,用于将所述每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,所述解析服务器用于将接收到的语音数据解析为对应的文字数据;

接收单元,用于接收所述解析服务器返回的至少一个解析结果;

输出单元,用于输出所述至少一个解析结果。

10. 一种语音数据的处理系统,其特征在于,包括:

多声道声卡,用于采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;

解析服务器,用于将语音数据解析为对应的文字数据;

处理器,与所述解析服务器通信,用于将所述每个目标对象的语音数据按照预设规则分批发送至解析服务器,并接收所述解析服务器返回的至少一个解析结果;

显示器,与所述处理器连接,用于输出所述至少一个解析结果。

11. 根据权利要求10所述的系统,其特征在于,所述系统还包括:

至少一个语音输入设备,与所述多声道声卡连接,用于接收所述至少一个目标对象的输入的语音数据。

12. 一种存储介质,其特征在于,所述存储介质包括存储的程序,其中,所述程序执行权利要求1至8中任意一项所述的语音数据的处理方法。

13. 一种处理器,其特征在于,所述处理器用于运行程序,其中,所述程序运行时执行权利要求1至8中任意一项所述的语音数据的处理方法。

语音数据的处理方法、装置和系统

技术领域

[0001] 本发明涉及语音识别领域,具体而言,涉及一种语音数据的处理方法、装置和系统。

背景技术

[0002] 随着语音识别技术日趋成熟,语音识别技术已经被广泛应用于各个行业领域,例如,工业、家电、通信、汽车电子、医疗、家庭服务、消费电子产品等各个领域。伴随各种语音识别软件的出现,用户可以通过语音识别软件,将语音转换为相应的文字,使得用户无需再手动输入文字,给用户的生活和工作带来很大的便利。

[0003] 在法院庭审中,需要将法官及当事人等各角色的发言信息记录下来,传统的庭审过程中,主要靠一名速录员对各角色的发言进行记录,这对速录员的打字水平有一定的要求,同时还需要速录员对庭审的案件内容有一定的了解,对法律法规有一定程度的专业深度。但是,在记录过程中,由于速录员打字速度或对案件理解程度不同,难免会出现漏记、错记等现象。

[0004] 针对上述现有的庭审过程中采用人工记录的方式将人员发言转为文字导致时效性和准确度不高的问题,目前尚未提出有效的解决方案。

发明内容

[0005] 本发明实施例提供了一种语音数据的处理方法、装置和系统,以至少解决现有的庭审过程中采用人工记录的方式将人员发言转为文字导致时效性和准确度不高的技术问题。

[0006] 根据本发明实施例的一个方面,提供了一种语音数据的处理方法,包括:采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;将每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,解析服务器用于将接收到的语音数据解析为对应的文字数据;接收解析服务器返回的至少一个解析结果;输出至少一个解析结果。

[0007] 根据本发明实施例的另一方面,还提供了一种语音数据的处理装置,包括:采集单元,用于采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;发送单元,用于将每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,解析服务器用于将接收到的语音数据解析为对应的文字数据;接收单元,用于接收解析服务器返回的至少一个解析结果;输出单元,用于输出至少一个解析结果。

[0008] 根据本发明实施例的另一方面,还提供了一种语音数据的处理系统,包括:多声道声卡,用于采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;解析服务器,用于将语音数据解析为对应的文字数据;处理器,与解析服务器通信,用于将每个目标对象的语音数据按照预设规则分批发送至解析服务器,并接收解析服务器返回的至少一个解析结果;显示器,与处理器连接,用于输出至少一个解析结果。

[0009] 根据本发明实施例的另一方面,还提供了一种存储介质,存储介质包括存储的程

序,其中,程序执行上述的语音数据的处理方法。

[0010] 根据本发明实施例的另一方面,还提供了一种处理器,处理器用于运行程序,其中,程序运行时执行上述的语音数据的处理方法。

[0011] 在本发明实施例中,通过采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;将每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,解析服务器用于将接收到的语音数据解析为对应的文字数据;接收解析服务器返回的至少一个解析结果;输出至少一个解析结果,达到了将发言人的语音数据即时转换为文字并进行显示的目的,从而实现了在庭审过程中持续记录各角色的发言信息的技术效果,进而解决了现有的庭审过程中采用人工记录的方式将人员发言转为文字导致时效性和准确度不高的技术问题。

附图说明

[0012] 此处所说明的附图用来提供对本发明的进一步理解,构成本申请的一部分,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定。在附图中:

[0013] 图1是根据本发明实施例的一种语音数据的处理方法流程图;

[0014] 图2(a)是根据本发明实施例的一种可选的庭审现场示意图;

[0015] 图2(b)是根据本发明实施例的一种可选的庭审现场各角色话筒与声道的对应关系示意图;

[0016] 图3是根据本发明实施例的一种可选的语音数据的处理方法流程图;

[0017] 图4(a)是根据本发明实施例的一种可选的某一声道上传输的语音信号示意图;

[0018] 图4(b)是根据本发明实施例的一种可选的语音信号片段示意图;

[0019] 图5(a)是根据本发明实施例的一种可选的截取波形文件的示意图;

[0020] 图5(b)是根据本发明实施例的又一种可选的截取波形文件的示意图;

[0021] 图6是根据本发明实施例的一种可选的基于语音识别的即时展现各角色的发言信息的流程示意图;

[0022] 图7是根据本发明实施例的一种语音数据的处理系统示意图;

[0023] 图8是根据本发明实施例的一种优选的用于庭审过程中记录发言人发言记录的语音系统示意图;以及

[0024] 图9是根据本发明实施例的一种语音数据的处理装置示意图。

具体实施方式

[0025] 为了使本技术领域的人员更好地理解本发明方案,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分的实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都应当属于本发明保护的范围。

[0026] 需要说明的是,本发明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,以便这里描述的本发明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。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0027] 实施例1

[0028] 根据本发明实施例,提供了一种语音数据的处理方法实施例,需要说明的是,在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0029] 图1是根据本发明实施例的一种语音数据的处理方法流程图,如图1所示,该方法包括如下步骤:

[0030] 步骤S102,采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道。

[0031] 具体地,在上述步骤中,上述目标对象可以为能够发出声音的人或物体,在目标对象为人的情况下,用户可以通过话筒、麦克风等语音输入设备,或者带有麦克风的智能设备(例如,手机、平板电脑、笔记本等)输入语音数据,可以通过建立每个目标对象与语音设备的对应的关系,使得每个目标对象对应一个声道;语音数据是指通过语音来记录或传输的数据。

[0032] 此处需要说明的是,在传统的庭审过程中,都会安排一个速录员,对法官以及当事人等各角色的发言进行记录,由于庭审中的各角色的发言记录是要即时展现在大屏幕上,因而,要求速录员的记录速度和记录准确度都有很高的要求。在庭审过程中,通过各种语音输入设备采集法官以及当事人等各角色在发言过程中的语音数据,并将语音数据转换为相应的文字,显示在大屏幕上,可以大大节省人力成本,并且可以更快速、更准确地记录各角色的发言。

[0033] 一种可选的实施例中,图2(a)是根据本发明实施例的一种可选的庭审现场示意图,如图2(a)所示,在庭审现场的人员主要有:审判长、人民陪审员、书记员、法官助理、原告、被告,以及第三人。法庭现场各角色的大致位置如图2(a)所示。每个人员通过其位置旁的话筒进行发言,因而,可以设置一个多声道声卡,该声卡连接至速录员电脑上,声卡上每个声道对应一个人员(即上述目标对象)的话筒,将声卡的各个声道与庭审角色关系对应,并按照此对应方式将各角色的话筒连接至声卡。作为一种可选的实施方式,各个声道与各角色的话筒的对应关系如图2(b)所示,声道1连接审判长的话筒,声道2和声道3分别连接两个人民陪审员的话筒,声道4连接法官助理的话筒,声道5和声道6分别连接两个原告的话筒,声道7连接被告的话筒,声道8连接第三人的话筒。

[0034] 步骤S104,将每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,解析服务器用于将接收到的语音数据解析为对应的文字数据。

[0035] 具体地,在上述步骤中,上述解析服务器可以是用于将语音数据转换为文字数据的语音解析服务器;在采集到一个或多个目标对象的语音数据后,将采集到的语音数据按照预设规则分批发送至语音解析服务器,例如,在接收到一句话的语音数据后,可以将这句话的语音数据按照输入的时间顺序,分多次发送至解析服务器进行解析,解析一部分语音

数据,则将这部分的语音数据对应的文字解析结果显示出来,从而达到即时展现的目的。

[0036] 此处需要说明的是,现有的语音识别软件,采集完发言人的一句话或一段话的语音数据后,通常是将整个语音数据对应的波形文件发送到服务器进行解析,由于文件大,解析时间长,发言人一句话已经说完,需要等一段时间才能看到文字信息,无法满足庭审过程中即时展现发言记录的要求,不能达到即时目的。而上述步骤S104将采集到的语音数据分批发送至解析服务器进行解析,由于解析的数据量小,解析速度快,因而,可以实现边发言边显示转换文字的效果。

[0037] 步骤S106,接收解析服务器返回的至少一个解析结果。

[0038] 具体地,在上述步骤中,在将采集到的语音数据分批发送给解析服务器后,解析服务器会将每次接收到的语音数据转换为对应的文字,并返回每次解析得到的文字数据。

[0039] 此处需要说明的是,在采集到的至少一个目标对象的语音数据为同时发言的语音数据的情况下,上述解析服务器可以并行的方式同时处理多个目标对象的语音数据。

[0040] 步骤S108,输出至少一个解析结果。

[0041] 具体地,在上述步骤中,在接收到服务器返回的对语音数据进行分批解析的至少一个解析结果后,输出这些解析结果,则得到完整语音数据的解析结果。

[0042] 作为一种可选的实施例,可以按照解析顺序输出对语音数据进行分批解析后的至少一个解析结果。

[0043] 以图2(a)和2(b)所示的庭审现场为例,假设书记员对着话筒讲了一句“请审判长入庭…全体起立”,即向其位置旁的话筒输入了一段对应的语音数据,则通过上述步骤S104,可以按照预设规则将“请审判长入庭…全体起立”这段语音数据对应的音频文件划分为“请”、“审判长”、“入庭”、“全体”和“起立”多个子音频文件依次发送至解析服务器进行解析,则随着书记员的讲话,庭审现场的屏幕上逐次显示“请”、“请审判长”、“请审判长入庭”、“请审判长入庭,全体”和“请审判长入庭,全体起立”的界面,而不会像传统的语音识别技术,将一句话的语音数据全部发送至解析服务器进行解析,由于一句话的语音数据文件较大,解析需要一定的时间,会出现,书记员讲完话后,等待一段时间后出现“请审判长入庭,全体起立”的一个界面。

[0044] 由上可知,在本申请上述实施例中,当采集到一个或多个目标对象的语音数据后,将采集到的每个目标对象的语音数据按照预设规则划分为多个文件较小的子语音数据,分批发送至解析服务器,通过解析服务器对接收到的子语音数据分别进行解析并转换为对应的文字解析结果,最后将分批发送的多个子语音数据对应的文字解析结果合并后输出,达到了将发言人的语音数据即时转换为文字并进行显示的目的,从而实现了在庭审过程中持续记录各角色的发言信息的技术效果,进而解决了现有的庭审过程中采用人工记录的方式将人员发言转为文字导致时效性和准确度不高的技术问题。

[0045] 在一种可选的实施例中,如图3所示,将每个目标对象的语音数据按照预设规则分批发送至解析服务器,可以包括如下步骤:

[0046] 步骤S302,获取传输每个目标对象的语音数据的声道的波形文件;

[0047] 步骤S304,将波形文件按照预设音频时长进行分段处理,得到多个子波形文件;

[0048] 步骤S306,将多个子波形文件依次存储至与声道对应的发送队列中;

[0049] 步骤S308,将发送队列中的子波形文件依次发送至解析服务器。

[0050] 具体地,在上述步骤中,通过多声道声卡进行声音采集,每个声道都对应一个波形文件,该波形文件用于传输与声道连接的目标对象的语音数据;为了实现将每个目标对象的语音数据分批发送至解析服务器,可以对与该目标对象对应的声道的波形文件进行分段处理,由于波形文件为一个语音数据对应的时序波,因而,按照预设音频时长将波形文件可以划分为多个子波形文件,并将划分后的多个子波形文件按照顺序排入声道对应的发送队列中。

[0051] 一种可选的实施例中,图4(a)是根据本发明实施例的一种可选的某一声道上传输的语音信号示意图,如图4(a)所示,当该声道上的发言人发言的过程中,波形会出现上下波动,当该声道上的发言人停止发言后,则波形呈平缓趋势。

[0052] 此处需要说明的是,作为一种可选的实施方式,可以根据波形文件的波形来确定发言人是否在讲话,以及每次讲话的时间。通过波形文件上波形的振动情况下,可以确定一句话或一段话的起始时间和结束时间。除了上述根据波形文件的波形来确定一句话或一段话的起始时间和结束时间外,还可以通过语义信息来确定一句话或一段话的起始与结束。

[0053] 在一种可选的实施例中,在获取传输每个目标对象的语音数据的声道的波形文件之前,上述方法还可以包括:设置每个目标对象与声道的对应关系。

[0054] 具体地,在上述实施例中,可以将用于采集每个目标对象的语音数据的语音输入设备分别与多声道声卡的每个声道连接,从而将每个目标对象的角色与声道进行一一对应。

[0055] 基于上述图4(a)所示的语音信号,一种可选的实施例中,可以设置每25ms为一个片段,将波形文件按时间进行划分,并将划分后的子波形文件排入声道对应的发送队列中。如图4(b)所示为根据本发明实施例的一种可选的语音信号片段示意图,该片段的时长为25ms。

[0056] 通过上述实施例,可以实现即时展示发言人的发言信息的目的。

[0057] 基于上述实施例,在第一种可选的实施方式中,将波形文件按照预设音频时长进行分段处理,得到多个子波形文件,可以包括:基于预设音频时长,从波形文件上依次截取预设音频时长的多个子波形文件,其中,后一次截取的子波形文件的开始时间为前一次截取的子波形文件的结束时间。

[0058] 具体地,在上述实施方式中,按照预设音频时长可以将目标对象的语音数据对应的波形文件进行等分分段,按照时间顺序依次截取多个子波形文件,其中,多个子波形文件的音频时长相同,都等于预设音频时长。图5(a)是根据本发明实施例的一种可选的截取波形文件的示意图,如图5(a)所示,对于一个0.1s时长的波形文件,按照预设音频时长25ms依次截取,可以截取4个子波形文件。

[0059] 通过上述第一种实施方式,可以实现即时展现数据的目的。

[0060] 基于上述实施例,在第二种可选的实施方式中,将波形文件按照预设音频时长进行分段处理,得到多个子波形文件,包括:基于预设音频时长,从波形文件上按照预设音频时长的递增倍数依次截取多个子波形文件,其中,后一次截取的子波形文件包含前一次截取的子波形文件,且后一次截取的子波形文件的开始时间与前一次截取的子波形文件的开始时间相同。

[0061] 具体地,在上述实施方式中,按照预设音频时长可以将目标对象的语音数据对应

的波形文件按照递增音频时长进行截取,每次截取的开始时间都是该波形文件的起始时间,即,第一次截取的音频长度可以是预设音频时长,第二次截取的音频长度是预设音频时长的两倍,以此类推,将该波形文件截取成多个波形文件,可以看出,后一次截取的子波形文件包含了前一次截取的子波形文件。图5(b)是根据本发明实施例的又一种可选的截取波形文件的示意图,如图5(b)所示,对于一个0.1s时长的波形文件,按照预设音频时长25ms截取,则截取的4个子波形文件的长度分别为25ms、50ms、75ms和100ms。

[0062] 此处需要说明的是,由于语音数据越完整,解析出的文字数据越准确。在上述第二种划分方式中,后一次解析的波形文件对应的语音数据包含了前一次解析的子波形文件对应的语音数据,则后一次的解析结果可以用于对前一次的解析结果进行校对。

[0063] 由此,基于上述第二种实施方式,在接收解析服务器返回的至少一个解析结果之后,上述方法还可以包括:通过后一次截取的子波形文件对应的解析结果对前一次截取的子波形文件对应的解析结果进行校正。具体地,由于每个目标对应一个声道,则每个目标对象的语音数据都对应一个波形文件,在将每个目标对象的波形文件按照上述第二种实施方式截取后,可以使用后一个解析结果对前一个解析结果进行校正。

[0064] 通过上述第二种实施方式,既可以满足即时展现数据的目的,还可以通过校对的方式,提高语音转文字的准确度。

[0065] 此处还需要说明的是,由于上述第二种实施方式中,将语音数据对应的波形文件按照递增音频时长进行截取后的多个子波形文件是依次存储在发送队列中进行发送的,虽然发送每个数据包的时延很小,但是为了进一步提高语音转文字的速度,作为一种优选的实施方式,在语音数据对应的波形文件按照递增音频时长进行截取后的多个子波形文件后可以以并发的方式发送到解析服务器,这样避免了分段发送数据包的时延,进一步提高了语音转文字的速度。在实际实施场景中,即将发送队列中的数据包通过多个发送模块同时发送,这样,前面截取的音频时长较短的数据包由于时长较短,还是先达到解析服务器,优先进行解析,并将解析后的结果优先呈现出来。

[0066] 在一种可选的实施例中,上述解析结果至少包括如下任意之一:目标对象的标识信息和语音数据对应的文字数据,其中,在接收解析服务器返回的至少一个解析结果之后,上述方法还可以包括:按照目标对象的标识信息,将解析得到的文字数据存储至对应声道的文字队列中。

[0067] 基于上述实施例,输出至少一个解析结果,可以包括:按照每个目标对象对应的声道的文字队列中存储的文字数据的顺序,输出每个目标对象的语音数据对应的文字数据。

[0068] 图6是根据本发明实施例的一种可选的基于语音识别的即时展现各角色的发言信息的流程示意图,如图6所示,通过多声道声卡采集到至少一个目标对象(图中示出了3个)的语音数据后,对采集到的每个目标对象的语音数据对应的波形文件进行分段处理,分段后按顺序排入每个声道对应的发送队列中;然后将发送队列中的片段按顺序依次发送至解析服务器,并接收已经解析完成的文字,存储在对应的声道的接收队列中,最后从每个目标对象对应的声道的文字队列中取出文字,展示在大屏幕上。由于进行了分段处理,发言人持续说话的过程中,整个流程都在持续不断的进行解析展示,由此完成了各角色发言信息的即时展示。

[0069] 通过上述实施例公开的方案,提供了一种新的、即时的庭审过程中各角色发言信

息的展示方式,通过对语音识别技术的运用,将庭审过程中各角色的发言进行了创新性的即时展示,发言人在发言的过程中可以持续看到自己的发言信息,并且新的展示方式规避了人为错记漏记的可能,极大的提高了信息的完整性与准确性。

[0070] 实施例2

[0071] 根据本发明实施例,还提供了一种用于实现上述语音数据的处理方法的系统实施例,图7是根据本发明实施例的一种语音数据的处理系统示意图,如图7所示,该系统包括:多声道声卡701、解析服务器703、处理器705和显示器707。

[0072] 其中,多声道声卡701,用于采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;

[0073] 解析服务器703,用于将语音数据解析为对应的文字数据;

[0074] 处理器705,与解析服务器通信,用于将每个目标对象的语音数据按照预设规则分批发送至解析服务器,并接收解析服务器返回的至少一个解析结果;

[0075] 显示器707,与处理器连接,用于输出至少一个解析结果。

[0076] 具体地,上述多声道声卡可以将至少一个声道的语音信号转换为数字信号,以用于解析服务器进行处理;上述解析服务器是用于将语音转文字的服务器;上述处理器可以是计算机、笔记本电脑、平板电脑、手机等任意一种智能设备的处理器,与多声道声卡连接,用于接收至少一个目标对象的语音数据,并将接收到的每个目标对象的语音数据分批发送至解析服务器进行解析,并接收解析服务器解析后的文字数据;上述显示器可以为用于解析结果的显示屏;可选地,上述处理器和上述显示器可以同一设备(例如,计算机、笔记本电脑、平板电脑、手机等)的处理器和显示器,也可以是外置的显示屏或投影屏幕。

[0077] 由上可知,在本申请上述实施例中,当多声道声卡701采集到一个或多个目标对象的语音数据后,处理器705将采集到的每个目标对象的语音数据按照预设规则划分为多个文件较小的子语音数据,分批发送至解析服务器703,解析服务器703对接收到的子语音数据分别进行解析并转换为对应的文字解析结果,最后显示器707将分批发送的多个子语音数据对应的文字解析结果合并后输出,达到了将发言人的语音数据即时转换为文字并进行显示的目的,从而实现了在庭审过程中持续记录各角色的发言信息的技术效果,进而解决了现有的庭审过程中采用人工记录的方式将人员发言转为文字导致时效性和准确度不高的技术问题。

[0078] 在一种可选的实施例中,上述系统还包括:至少一个语音输入设备,与多声道声卡连接,用于接收至少一个目标对象的输入的语音数据。

[0079] 具体地,在上述实施例中,上述语音输入设备为用于目标对象输入语音数据的设备,可以话筒或麦克风;每个目标对象的语音输入设备分别与多声道声卡的每个声道连接。

[0080] 作为一种优选的实施例,图8是根据本发明实施例的一种优选的用于庭审过程中记录发言人发言记录的语音系统示意图,如图8所示,庭审软件安装在速录员的电脑上,审判长、原告和被告等各角色的话筒分别连接至多声道声卡的各个声道中,通过多声道声卡采集每个角色发言的语音信号,并语音转化为数字信号,发送至语音解析服务器;语音解析服务器将接收到语音数据转换为文字信息后,返回至连接庭审软件的大屏显示器上,进行即时展示。

[0081] 通过上述实施例,在庭审过程中,引入基于语音识别的即时展现各角色的发言信

息装置以后,可以在法庭即时展现发言人信息,并且此项工作完全不需要人为参与,不仅提高了效率节约了人力成本,更在准确性、完整性方面有了大大的提高。

[0082] 实施例3

[0083] 根据本发明实施例,还提供了一种用于实现上述语音数据的处理方法的装置实施例,图9是根据本发明实施例的一种语音数据的处理装置示意图,如图9所示,该装置包括:采集单元901、发送单元903、接收单元905和输出单元907。

[0084] 其中,采集单元901,用于采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;

[0085] 发送单元903,用于将每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,解析服务器用于将接收到的语音数据解析为对应的文字数据;

[0086] 接收单元905,用于接收解析服务器返回的至少一个解析结果;

[0087] 输出单元907,用于输出至少一个解析结果。

[0088] 此处需要说明的是,上述采集单元901、发送单元903、接收单元905和输出单元907对应于实施例1中的步骤S102至S108,上述模块与对应的步骤所实现的示例和应用场景相同,但不限于上述实施例1所公开的内容。需要说明的是,上述模块作为装置的一部分可以在诸如一组计算机可执行指令的计算机系统中执行。

[0089] 由上可知,在本申请上述实施例中,通过采集单元901采集一个或多个目标对象的语音数据,发送单元903将采集到的每个目标对象的语音数据按照预设规则划分为多个文件较小的子语音数据,分批发送至解析服务器,通过接收单元905接收解析服务器对接收到的子语音数据分别进行解析后的文字解析结果,最后通过输出单元907将分批发送的多个子语音数据对应的文字解析结果合并后输出,达到了将发言人的语音数据即时转换为文字并进行显示的目的,从而实现了在庭审过程中持续记录各角色的发言信息的技术效果,进而解决了现有的庭审过程中采用人工记录的方式将人员发言转为文字导致时效性和准确度不高的技术问题。

[0090] 在一种可选的实施例中,上述发送单元903包括:获取模块,用于获取传输每个目标对象的语音数据的声道的波形文件;处理模块,用于将波形文件按照预设音频时长进行分段处理,得到多个子波形文件;存储模块,用于将多个子波形文件依次存储至与声道对应的发送队列中;发送模块,用于将发送队列中的子波形文件依次发送至解析服务器。

[0091] 在一种可选的实施例中,上述处理模块还包括:第一处理子模块,用于基于预设音频时长,从波形文件上依次截取预设音频时长的多个子波形文件,其中,后一次截取的子波形文件的开始时间为前一次截取的子波形文件的结束时间。

[0092] 在一种可选的实施例中,上述处理模块还包括:第二处理子模块,用于基于预设音频时长,从波形文件上按照预设音频时长的递增倍数依次截取多个子波形文件,其中,后一次截取的子波形文件包含前一次截取的子波形文件,且后一次截取的子波形文件的开始时间与前一次截取的子波形文件的开始时间相同。

[0093] 在一种可选的实施例中,上述装置还包括:校正单元,用于通过后一次截取的子波形文件对应的解析结果对前一次截取的子波形文件对应的解析结果进行校正。

[0094] 在一种可选的实施例中,上述装置还包括:设置单元,用于设置每个目标对象与声道的对应关系。

[0095] 在一种可选的实施例中,上述解析结果包括:目标对象的标识信息和语音数据对应的文字数据,其中,上述装置还包括:存储单元,用于按照目标对象的标识信息,将解析得到的文字数据存储至对应声道的文字队列中。

[0096] 在一种可选的实施例中,上述输出单元还用于按照每个目标对象对应的声道的文字队列中存储的文字数据的顺序,输出每个目标对象的语音数据对应的文字数据。

[0097] 上述语音数据的处理装置包括处理器和存储器,上述采集单元、发送单元、接收单元、输出单元、获取模块、处理模块、存储模块、发送模块、第一处理子模块、第二处理子模块、校正单元、设置单元和存储单元等均作为程序单元存储在存储器中,由处理器执行存储在存储器中的上述程序单元来实现相应的功能。

[0098] 处理器中包含内核,由内核去存储器中调取相应的程序单元。内核可以设置一个或以上,通过调整内核参数来达到将发言人的语音数据即时转换为文字并进行显示的目的,从而实现了在庭审过程中持续记录各角色的发言信息的技术效果。

[0099] 存储器可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM),存储器包括至少一个存储芯片。

[0100] 本发明实施例提供了一种存储介质,其上存储有程序,该程序被处理器执行时实现上述实施例1中任意一项可选的或优选的语音数据的处理方法。

[0101] 本发明实施例提供了一种处理器,处理器用于运行程序,其中,程序运行时执行上述实施例1中任意一项可选的或优选的语音数据的处理方法。

[0102] 本发明实施例提供了一种设备,设备包括处理器、存储器及存储在存储器上并可在处理器上运行的程序,处理器执行程序时实现以下步骤:采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;将每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,解析服务器用于将接收到的语音数据解析为对应的文字数据;接收解析服务器返回的至少一个解析结果;输出至少一个解析结果。

[0103] 进一步地,将每个目标对象的语音数据按照预设规则分批发送至解析服务器,包括:获取传输每个目标对象的语音数据的声道的波形文件;将波形文件按照预设音频时长进行分段处理,得到多个子波形文件;将多个子波形文件依次存储至与声道对应的发送队列中;将发送队列中的子波形文件依次发送至解析服务器。

[0104] 进一步地,将波形文件按照预设音频时长进行分段处理,得到多个子波形文件,包括:基于预设音频时长,从波形文件上依次截取预设音频时长的多个子波形文件,其中,后一次截取的子波形文件的开始时间为前一次截取的子波形文件的结束时间。

[0105] 进一步地,将波形文件按照预设音频时长进行分段处理,得到多个子波形文件,包括:基于预设音频时长,从波形文件上按照预设音频时长的递增倍数依次截取多个子波形文件,其中,后一次截取的子波形文件包含前一次截取的子波形文件,且后一次截取的子波形文件的开始时间与前一次截取的子波形文件的开始时间相同。

[0106] 进一步地,在接收解析服务器返回的至少一个解析结果之后,方法还包括:通过后一次截取的子波形文件对应的解析结果对前一次截取的子波形文件对应的解析结果进行校正。

[0107] 进一步地,在获取传输每个目标对象的语音数据的声道的波形文件之前,方法还

包括:设置每个目标对象与声道的对应关系。

[0108] 进一步地,解析结果包括:目标对象的标识信息和语音数据对应的文字数据,其中,在接收解析服务器返回的至少一个解析结果之后,方法还包括:按照目标对象的标识信息,将解析得到的文字数据存储至对应声道的文字队列中。

[0109] 进一步地,输出至少一个解析结果,包括:按照每个目标对象对应的声道的文字队列中存储的文字数据的顺序,输出每个目标对象的语音数据对应的文字数据。

[0110] 需要说明的是,本文中的设备可以是服务器、PC、PAD、手机等。

[0111] 本申请还提供了一种计算机程序产品,当在数据处理设备上执行时,适于执行初始化有如下方法步骤的程序:采集至少一个目标对象的语音数据,其中,每个目标对象对应一个声道;将每个目标对象的语音数据按照预设规则分批发送至解析服务器,其中,解析服务器用于将接收到的语音数据解析为对应的文字数据;接收解析服务器返回的至少一个解析结果;输出至少一个解析结果。

[0112] 进一步地,将每个目标对象的语音数据按照预设规则分批发送至解析服务器,包括:获取传输每个目标对象的语音数据的声道的波形文件;将波形文件按照预设音频时长进行分段处理,得到多个子波形文件;将多个子波形文件依次存储至与声道对应的发送队列中;将发送队列中的子波形文件依次发送至解析服务器。

[0113] 进一步地,将波形文件按照预设音频时长进行分段处理,得到多个子波形文件,包括:基于预设音频时长,从波形文件上依次截取预设音频时长的多个子波形文件,其中,后一次截取的子波形文件的开始时间为前一次截取的子波形文件的结束时间。

[0114] 进一步地,将波形文件按照预设音频时长进行分段处理,得到多个子波形文件,包括:基于预设音频时长,从波形文件上按照预设音频时长的递增倍数依次截取多个子波形文件,其中,后一次截取的子波形文件包含前一次截取的子波形文件,且后一次截取的子波形文件的开始时间与前一次截取的子波形文件的开始时间相同。

[0115] 进一步地,在接收解析服务器返回的至少一个解析结果之后,方法还包括:通过后一次截取的子波形文件对应的解析结果对前一次截取的子波形文件对应的解析结果进行校正。

[0116] 进一步地,在获取传输每个目标对象的语音数据的声道的波形文件之前,方法还包括:设置每个目标对象与声道的对应关系。

[0117] 进一步地,解析结果包括:目标对象的标识信息和语音数据对应的文字数据,其中,在接收解析服务器返回的至少一个解析结果之后,方法还包括:按照目标对象的标识信息,将解析得到的文字数据存储至对应声道的文字队列中。

[0118] 进一步地,输出至少一个解析结果,包括:按照每个目标对象对应的声道的文字队列中存储的文字数据的顺序,输出每个目标对象的语音数据对应的文字数据。

[0119] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0120] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程

图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0121] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0122] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0123] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0124] 存储器可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。存储器是计算机可读介质的示例。

[0125] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0126] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0127] 本领域技术人员应明白,本申请的实施例可提供为方法、系统或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0128] 以上仅为本申请的实施例而已,并不用于限制本申请。对于本领域技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本申请的权利要求范围之内。

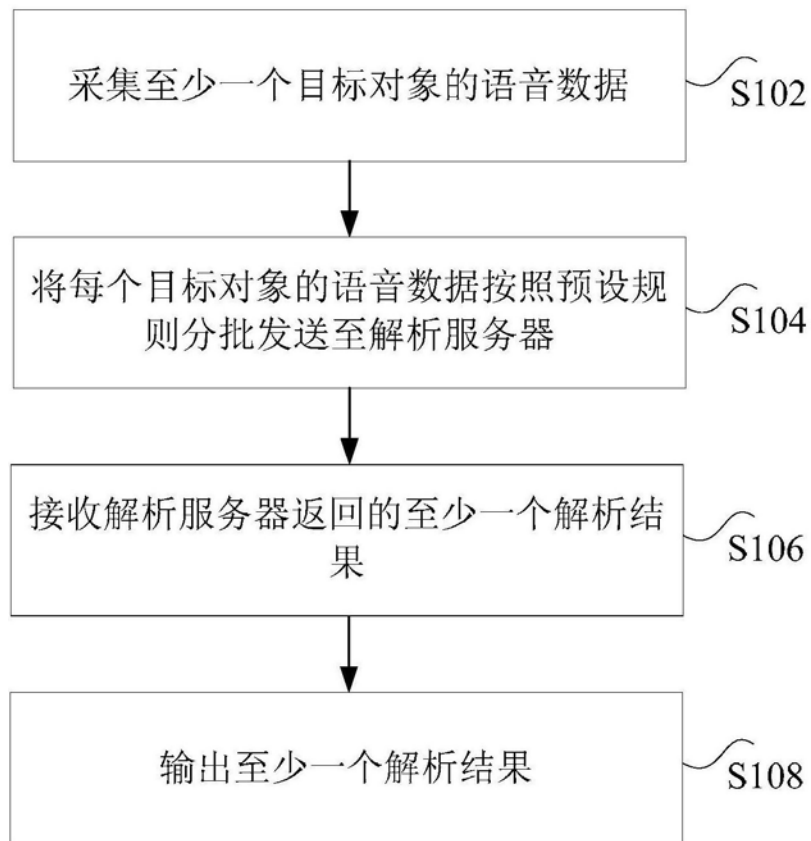


图1

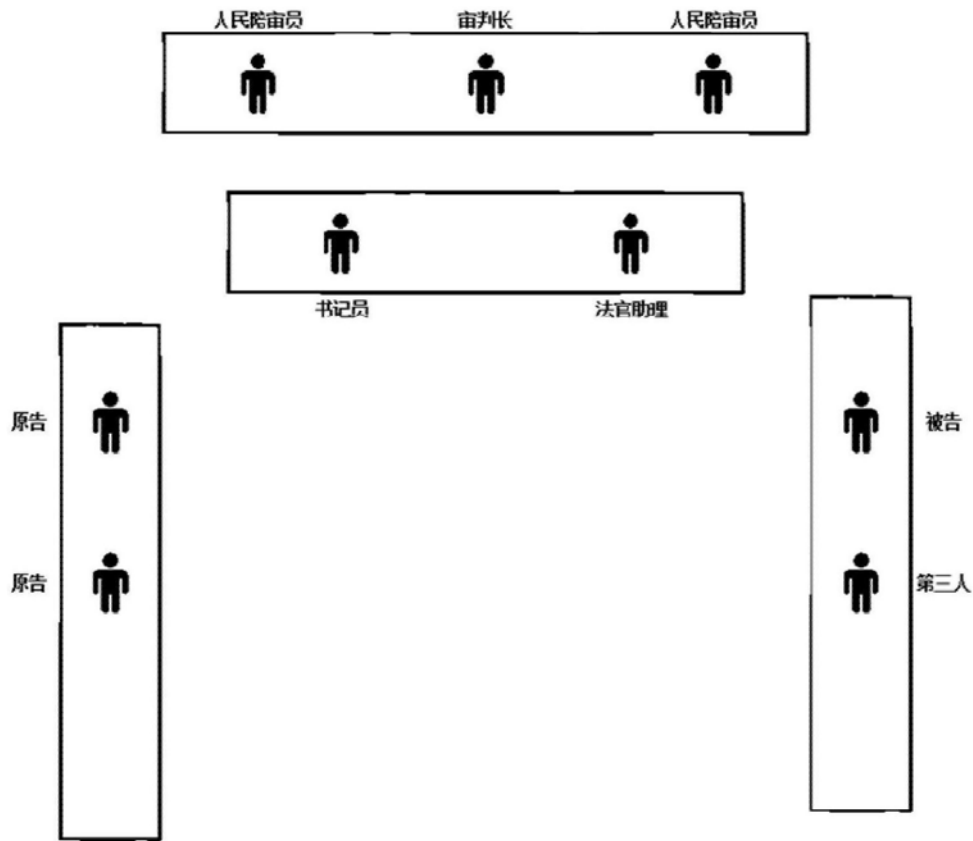


图2 (a)

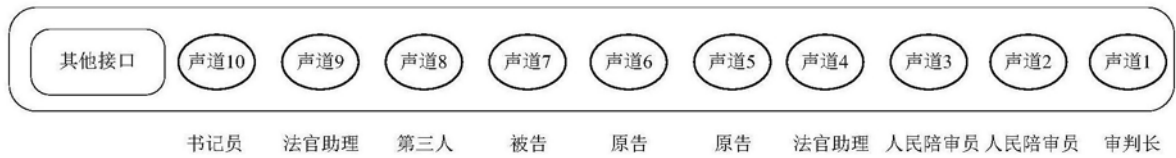


图2 (b)

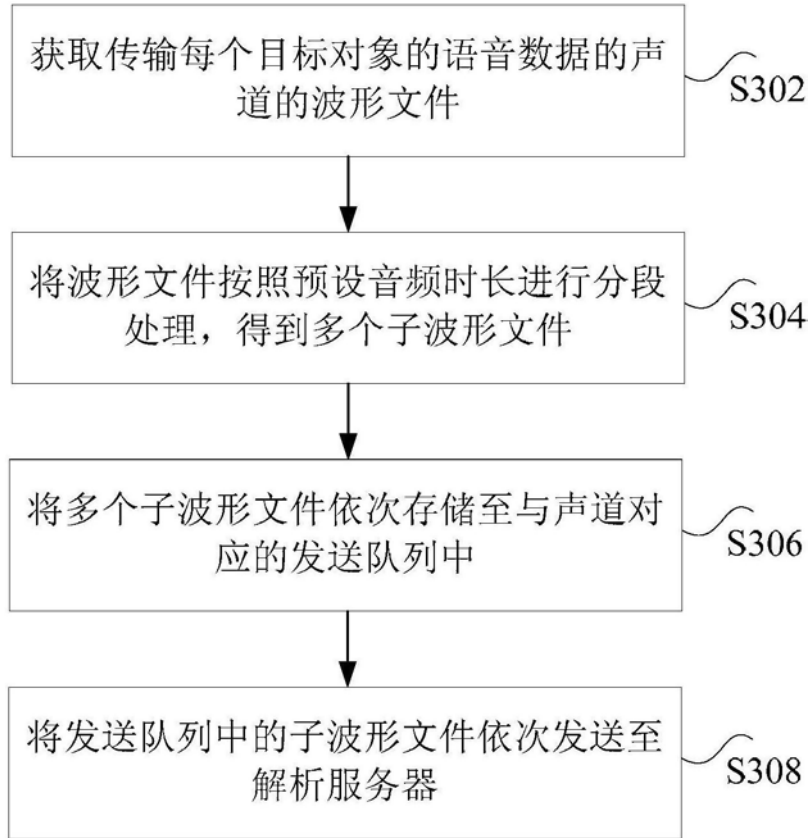


图3



图4 (a)

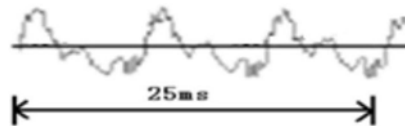


图4 (b)

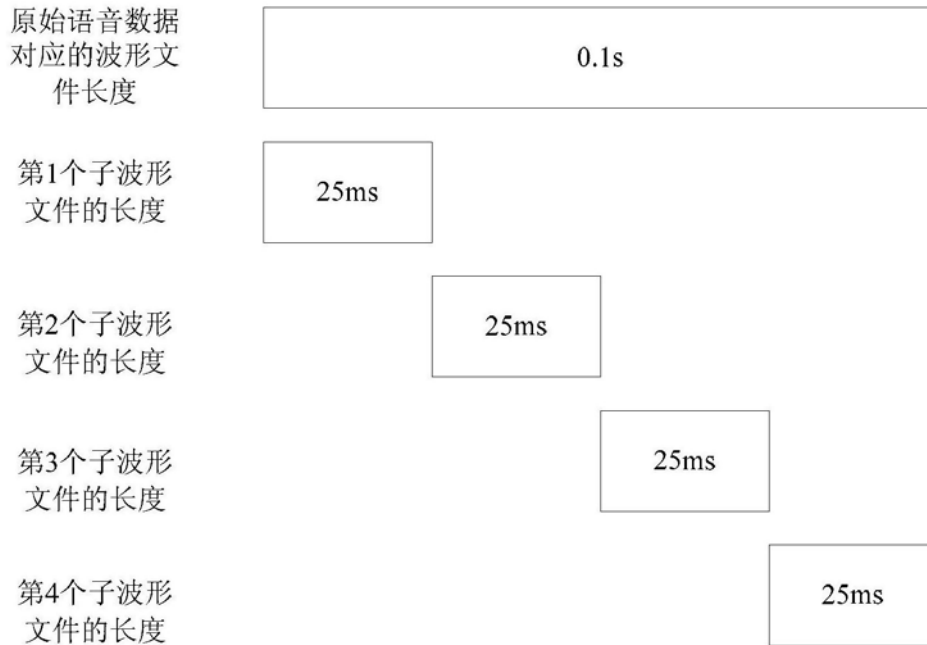


图5 (a)

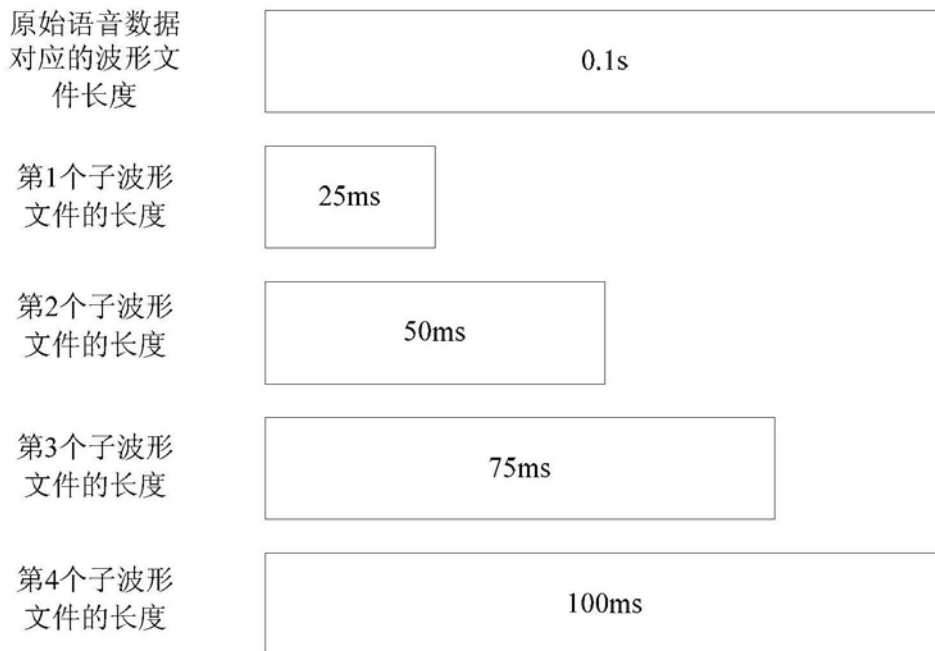


图5 (b)

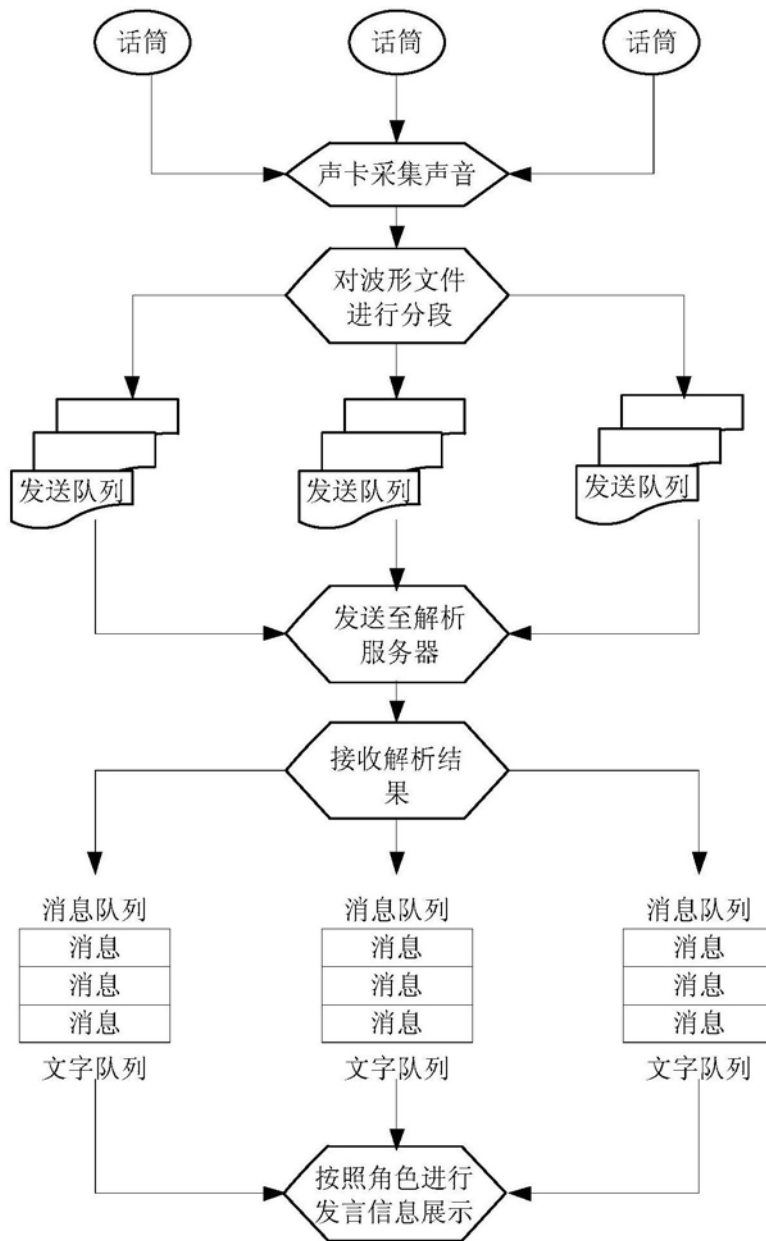


图6

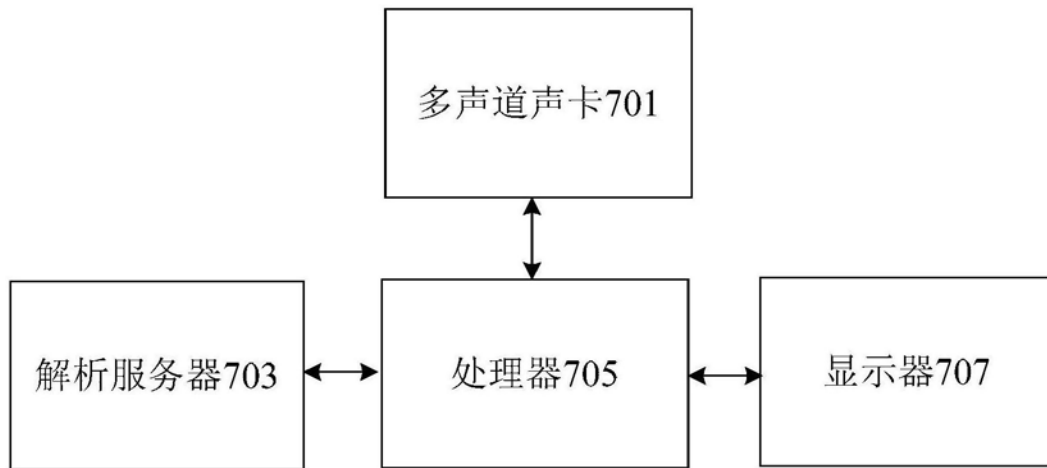


图7

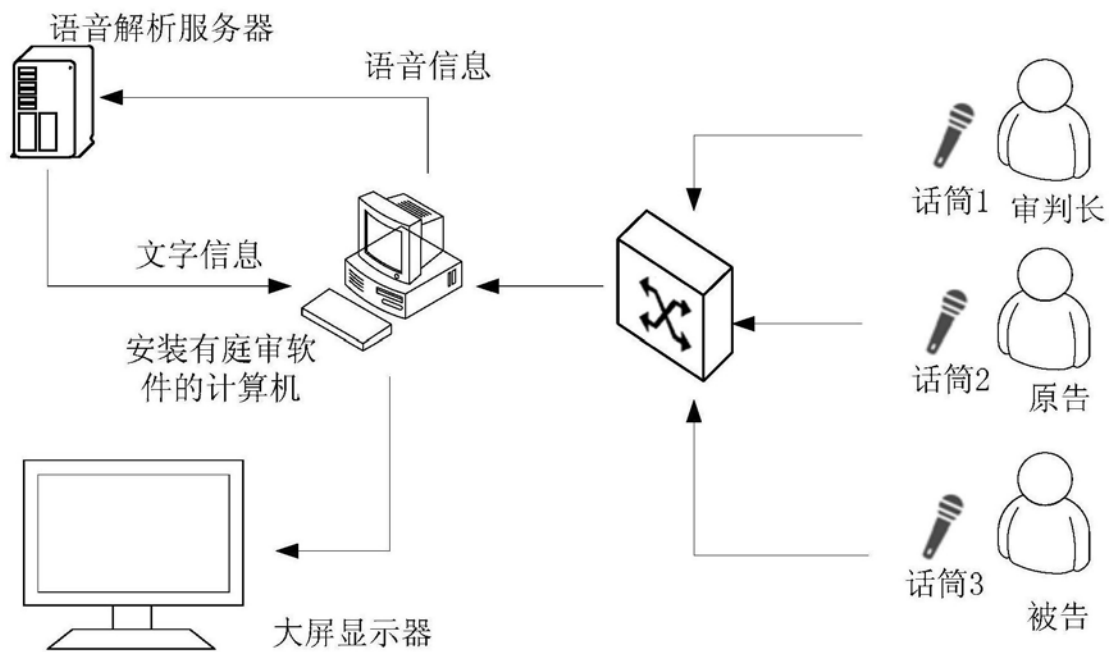


图8



图9