



(12) 发明专利

(10) 授权公告号 CN 107578777 B

(45) 授权公告日 2021.08.03

(21) 申请号 201610523339.9

G06F 16/34 (2019.01)

(22) 申请日 2016.07.05

G06F 3/14 (2006.01)

(65) 同一申请的已公布的文献号
申请公布号 CN 107578777 A

(56) 对比文件

CN 101309390 A, 2008.11.19

CN 101827256 A, 2010.09.08

(43) 申请公布日 2018.01.12

US 7908145 B2, 2011.03.15

(73) 专利权人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

CN 101309390 A, 2008.11.19

审查员 涂亮梅

(72) 发明人 高杰 周贇

(74) 专利代理机构 北京安信方达知识产权代理
有限公司 11262

代理人 李红爽 栗若木

(51) Int. Cl.

G10L 15/26 (2006.01)

G06F 16/35 (2019.01)

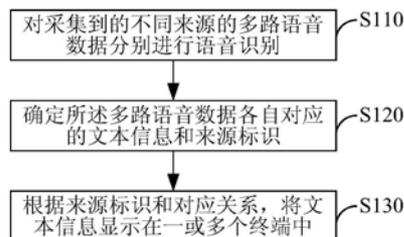
权利要求书4页 说明书21页 附图3页

(54) 发明名称

文字信息显示方法、装置及系统、语音识别
方法及装置

(57) 摘要

一种文字信息显示方法、装置及系统、语音
识别方法及装置;所述文本信息显示方法,预先
设置有来源标识和终端的对应关系,所述方法包
括:对采集到的不同来源的多路语音数据分别进
行语音识别;确定所述多路语音数据各自对应
的文本信息和来源标识;根据所述对应关系,将所
述多路语音数据各自对应的文本信息,分别显示
在所述多路语音数据各自对应的来源标识所对
应的一或多个终端中。本申请能够快速、准确地
实现在终端中只显示某个或某些来源的语音数
据对应的文本信息。



1. 一种文本信息显示方法,预先设置有来源标识和终端的对应关系,所述方法包括:
对采集到的不同来源的多路语音数据分别进行语音识别;
确定所述多路语音数据各自对应的文本信息和来源标识,包括:
通过对所述多路语音数据进行语音识别,分别确定所述多路语音数据各自对应的文本信息和来源标识;

根据所述对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中;

其中,所述对应关系能够根据所述终端所播放的视频的来源或内容自适应更新。

2. 如权利要求1所述的方法,其特征在于,所述将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中包括:

将所述多路语音数据各自对应的文本信息分别转换成字幕,显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中。

3. 如权利要求2所述的方法,其特征在于,所述将所述多路语音数据各自对应的文本信息分别转换成字幕包括:

在所述多路语音数据各自对应的文本信息中,将对应于预定的来源标识的语音数据所对应的文本信息转换成字幕。

4. 如权利要求1所述的方法,其特征在于,还包括:

根据终端所播放的视频,确定所述来源标识和终端的对应关系。

5. 如权利要求1所述的方法,其特征在于,所述确定多路语音数据各自对应的来源标识,还包括:

分别根据所述多路语音数据的采集端确定多路语音数据各自对应的来源标识。

6. 如权利要求1~5中任一项所述的方法,其特征在于,所述对采集到的不同来源的多路语音数据分别进行语音识别包括:

分别将采集到的不同来源的多路语音数据发送到云端进行语音识别,从云端接收对所述多路语音数据进行语音识别得到的文本信息。

7. 如权利要求1~5中任一项所述的方法,其特征在于,所述确定所述多路语音数据各自对应的文本信息和来源标识后还包括:

将所述多路语音数据各自对应的文本信息和来源标识对应保存在数据库中;

所述根据所述对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个所述终端中包括:

根据所述对应关系,确定不同终端对应的来源标识;

从所述数据库中分别获取不同终端对应的来源标识所对应的文本信息,并显示在相应终端中。

8. 一种语音识别方法,包括:

对采集到的一或多路语音数据进行语音识别;

通过对所述一或多路语音数据进行语音识别,分别确定所述一或多路语音数据各自对应的文本信息和来源标识;

其中,所述文本信息用于显示在所对应的语音数据的来源标识对应的一或多个终端中,所述来源标识和终端的对应关系预先设置且能够根据所述终端所播放的视频的来源或

内容自适应更新。

9. 如权利要求8所述的方法,其特征在于,所述对采集到的一或多路语音数据进行语音识别包括:

将采集到的一或多路语音数据发送到云端进行语音识别,从云端接收对所述一或多路语音数据进行语音识别得到的文本信息。

10. 如权利要求8所述的方法,其特征在于,所述确定所述一或多路语音数据各自对应的文本信息和来源标识后还包括:

将所述一或多路语音数据各自对应的文本信息和来源标识对应保存在数据库中。

11. 一种文本信息显示方法,包括:

获取一个或多个来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;

根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中;

其中,所述对应关系能够根据所述终端所播放的视频的来源或内容自适应更新。

12. 如权利要求11所述的方法,其特征在于,所述将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中包括:

将所述一个或多个来源标识对应的文本信息分别转换成字幕,显示在所述一个或多个来源标识所对应的终端中。

13. 如权利要求11所述的方法,其特征在于,所述获取一个或多个来源标识对应的文本信息包括:

从对应保存来源标识和文本信息的数据库中,获取一个或多个来源标识各自对应的文本信息。

14. 如权利要求11所述的方法,其特征在于,所述获取一个或多个来源标识对应的文本信息包括:

根据预置的来源标识和终端的对应关系,获取对应于预定的一或多个来源标识的文本信息。

15. 一种文本信息传输方法,预先保存有相互对应的文本信息和来源标识,其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;所述方法包括:

根据预置的来源标识和终端的对应关系,分别确定不同终端所对应的来源标识;

分别向不同终端发送与该终端对应的来源标识所对应的文本信息;

其中,所述对应关系能够根据所述终端所播放的视频的来源或内容自适应更新。

16. 一种文本信息显示系统,其特征在于,包括:

一个或多个语音识别装置,一个或多个文本信息显示装置;

至少一个所述语音识别装置用于对一或多路采集到的语音数据进行语音识别,确定所述一或多路采集到的语音数据各自对应的文本信息和来源标识,包括:通过对所述一或多路语音数据进行语音识别,分别确定所述一或多路语音数据各自对应的文本信息和来源标识;

至少一个所述文本信息显示装置用于根据预置的来源标识和终端的对应关系,将一或

多路语音数据对应的文本信息,显示在所述一或多路语音数据对应的来源标识所对应的终端中;

其中,所述对应关系能够根据所述终端所播放的视频的来源或内容自适应更新。

17.如权利要求16所述的系统,其特征在于,所述将一或多路语音数据对应的文本信息,显示在所述一或多路语音数据对应的来源标识所对应的终端中包括:

将一或多路语音数据对应的文本信息转换成字幕,显示在所述一或多路语音数据对应的来源标识所对应的终端中。

18.如权利要求16所述的系统,其特征在于,所述语音识别装置对一或多路采集到的语音数据进行语音识别包括:

将一或多路采集到的语音数据发送到云端进行语音识别,从云端接收对所述一或多路语音数据进行语音识别得到的文本信息。

19.如权利要求16所述的系统,其特征在于:

至少一个所述语音识别装置还用于将所述一或多路语音数据各自对应的文本信息和来源标识对应保存在数据库中;

至少一个所述文本信息显示装置还用于从所述数据库中获取一或多路语音数据对应的文本信息。

20.如权利要求19所述的系统,其特征在于,所述文本信息显示装置从所述数据库中获取一或多路语音数据对应的文本信息包括:

根据预置的来源标识和终端的对应关系,从所述数据库中获取对应于预定的一或多个来源标识的文本信息。

21.一种语音识别装置,包括:存储器和处理器;

其特征在于:

所述存储器用于存储用于语音识别的程序;所述用于语音识别的程序在被所述处理器读取执行时,执行如下操作:

对采集到的一或多路语音数据进行语音识别;

通过对所述一或多路语音数据进行语音识别,分别确定所述一或多路语音数据各自对应的文本信息和来源标识;

其中,所述文本信息用于显示在所对应的语音数据的来源标识对应的一或多个终端中,所述来源标识和终端的对应关系预先设置且能够根据所述终端所播放的视频的来源或内容自适应更新。

22.一种语音采集装置,包括:话筒、存储器、处理器;

所述话筒用于采集语音数据;

其特征在于:

所述存储器用于存储用于语音识别的程序;所述用于语音识别的程序在被所述处理器读取执行时,执行如下操作:

对所述话筒所采集到的语音数据进行语音识别;

通过对所述话筒所采集到的语音数据进行语音识别,确定所述语音数据对应的文本信息和来源标识;

其中,所述文本信息用于显示在所对应的语音数据的来源标识对应的一或多个终端

中,所述来源标识和终端的对应关系预先设置且能够根据所述终端所播放的视频的来源或内容自适应更新。

23.一种文本信息处理装置,包括:存储器、处理器;

其特征在于:

所述存储器用于存储用于文本信息显示的程序;所述用于文本信息显示的程序在被所述处理器读取执行时,执行如下操作:

获取一个或多个来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;

根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中;

其中,所述对应关系能够根据所述终端所播放的视频的来源或内容自适应更新。

24.一种文本信息显示装置,包括:显示屏、存储器、处理器;

其特征在于:

所述存储器用于存储用于文本信息显示的程序;所述用于文本信息显示的程序在被所述处理器读取执行时,执行如下操作:

获取预定的来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;

根据所述来源标识和终端的对应关系,将所获取的文本信息显示在所述显示屏中;所述对应关系能够根据所述终端所播放的视频的来源或内容自适应更新。

文字信息显示方法、装置及系统、语音识别方法及装置

技术领域

[0001] 本发明涉及多媒体领域,尤其涉及一种文字信息显示方法、装置及系统、语音识别方法及装置。

背景技术

[0002] 语音识别是一种使用机器学习的方法将语音数据转换为相应文本信息的技术。通过语音识别,可以在传递语音消息、交谈、会议等情况下,将语音以文本形式展现出来。

[0003] 相关技术中的方案至少存在如下的不足之处:

[0004] 在多方参与会话的情况下,如果需要在不同终端中显示某个或某些用户的发言所转换成的文本信息,通常需要依靠人工整理的方式完成;但是依靠人工整理的方式需要耗费较长的时间,无法满足实时性需求,而且受制于人类对声音的辨识能力,在区分不同用户的发言时存在出错的可能性。

发明内容

[0005] 本申请提供一种文字信息显示方法、装置及系统、语音识别方法及装置,能够快速、准确地实现在终端中只显示某个或某些来源的语音数据对应的文本信息。

[0006] 本申请采用如下技术方案。

[0007] 一种文本信息显示方法,预先设置有来源标识和终端的对应关系,所述方法包括:

[0008] 对采集到的不同来源的多路语音数据分别进行语音识别;

[0009] 确定所述多路语音数据各自对应的文本信息和来源标识;

[0010] 根据所述对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中。

[0011] 其中,所述将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中可以包括:

[0012] 将所述多路语音数据各自对应的文本信息分别转换成字幕,显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中。

[0013] 其中,所述将所述多路语音数据各自对应的文本信息分别转换成字幕可以包括:

[0014] 在所述多路语音数据各自对应的文本信息中,将对应于预定的来源标识的语音数据所对应的文本信息转换成字幕。

[0015] 其中,所述的方法还可以包括:

[0016] 根据终端所播放的视频,确定所述来源标识和终端的对应关系。

[0017] 其中,所述确定多路语音数据各自对应的来源标识可以包括:

[0018] 分别根据所述多路语音数据的采集端确定多路语音数据各自对应的来源标识;或者,通过对所述多路语音数据进行语音识别,分别确定所述多路语音数据各自对应的来源标识。

[0019] 其中,所述对采集到的不同来源的多路语音数据分别进行语音识别可以包括:

[0020] 分别将采集到的不同来源的多路语音数据发送到云端进行语音识别,从云端接收对所述多路语音数据进行语音识别得到的文本信息。

[0021] 其中,所述确定所述多路语音数据各自对应的文本信息和来源标识后还可以包括:

[0022] 将所述多路语音数据各自对应的文本信息和来源标识对应保存在数据库中;

[0023] 所述根据所述对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个所述终端中可以包括:

[0024] 根据所述对应关系,确定不同终端对应的来源标识;

[0025] 从所述数据库中分别获取不同终端对应的来源标识所对应的文本信息,并显示在相应终端中。

[0026] 一种语音识别方法,包括:

[0027] 对采集到的一或多路语音数据进行语音识别;

[0028] 确定所述一或多路语音数据各自对应的文本信息和来源标识。

[0029] 其中,所述对采集到的一或多路语音数据进行语音识别可以包括:

[0030] 将采集到的一或多路语音数据发送到云端进行语音识别,从云端接收对所述一或多路语音数据进行语音识别得到的文本信息。

[0031] 其中,所述确定所述一或多路语音数据各自对应的文本信息和来源标识后还可以包括:

[0032] 将所述一或多路语音数据各自对应的文本信息和来源标识对应保存在数据库中。

[0033] 一种文本信息显示方法,包括:

[0034] 获取一个或多个来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;

[0035] 根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中。

[0036] 其中,所述将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中可以包括:

[0037] 将所述一个或多个来源标识对应的文本信息分别转换成字幕,显示在所述一个或多个来源标识所对应的终端中。

[0038] 其中,所述获取一个或多个来源标识对应的文本信息可以包括:

[0039] 从对应保存来源标识和文本信息的数据库中,获取一个或多个来源标识各自对应的文本信息。

[0040] 其中,所述获取一个或多个来源标识对应的文本信息可以包括:

[0041] 根据预置的来源标识和终端的对应关系,获取对应于预定的一或多个来源标识的文本信息。

[0042] 一种文本信息传输方法,预先保存有相互对应的文本信息和来源标识,其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;所述方法包括:

[0043] 根据预置的来源标识和终端的对应关系,分别确定不同终端所对应的来源标识;

[0044] 分别向不同终端发送与该终端对应的来源标识所对应的文本信息。

- [0045] 一种文本信息显示系统,包括:
- [0046] 一个或多个语音识别装置,一个或多个文本信息显示装置;
- [0047] 至少一个所述语音识别装置用于对一或多路采集到的语音数据进行语音识别,确定所述一或多路采集到的语音数据各自对应的文本信息和来源标识;
- [0048] 至少一个所述文本信息显示装置用于根据预置的来源标识和终端的对应关系,将一或多路语音数据对应的文本信息,显示在所述一或多路语音数据对应的来源标识所对应的终端中。
- [0049] 其中,所述将一或多路语音数据对应的文本信息,显示在所述一或多路语音数据对应的来源标识所对应的终端中可以包括:
- [0050] 将一或多路语音数据对应的文本信息转换成字幕,显示在所述一或多路语音数据对应的来源标识所对应的终端中。
- [0051] 其中,所述语音识别装置对一或多路采集到的语音数据进行语音识别可以包括:
- [0052] 将一或多路采集到的语音数据发送到云端进行语音识别,从云端接收对所述一或多路语音数据进行语音识别得到的文本信息。
- [0053] 其中,至少一个所述语音识别装置还可以用于将所述一或多路语音数据各自对应的文本信息和来源标识对应保存在数据库中;
- [0054] 至少一个所述文本信息显示装置还可以用于从所述数据库中获取一或多路语音数据对应的文本信息。
- [0055] 其中,所述文本信息显示装置从所述数据库中获取一或多路语音数据对应的文本信息可以包括:
- [0056] 根据预置的来源标识和终端的对应关系,从所述数据库中获取对应于预定的一或多个来源标识的文本信息。
- [0057] 一种语音识别装置,包括:存储器和处理器;
- [0058] 所述存储器用于存储用于语音识别的程序;所述用于语音识别的程序在被所述处理器读取执行时,执行如下操作:
- [0059] 对采集到的一或多路语音数据进行语音识别;
- [0060] 确定所述一或多路语音数据各自对应的文本信息和来源标识。
- [0061] 一种语音采集装置,包括:话筒、存储器、处理器;
- [0062] 所述话筒用于采集语音数据;
- [0063] 所述存储器用于存储用于语音识别的程序;所述用于语音识别的程序在被所述处理器读取执行时,执行如下操作:
- [0064] 对所述话筒所采集到的语音数据进行语音识别;
- [0065] 确定所述语音数据对应的文本信息和来源标识。
- [0066] 一种文本信息处理装置,包括:存储器、处理器;
- [0067] 所述存储器用于存储用于文本信息显示的程序;所述用于文本信息显示的程序在被所述处理器读取执行时,执行如下操作:
- [0068] 获取一个或多个来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;
- [0069] 根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本

信息,显示在所述一个或多个来源标识所对应的终端中。

[0070] 一种文本信息显示装置,包括:显示屏、存储器、处理器;

[0071] 所述存储器用于存储用于文本信息显示的程序;所述用于文本信息显示的程序在被所述处理器读取执行时,执行如下操作:

[0072] 获取预定的来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;

[0073] 将所获取的文本信息显示在所述显示屏中。

[0074] 一种文本信息显示系统,预先设置有来源标识和终端的对应关系,所述系统包括:

[0075] 语音识别模块,用于对采集到的不同来源的多路语音数据分别进行语音识别;

[0076] 确定模块,用于确定所述多路语音数据各自对应的文本信息和来源标识;

[0077] 显示模块,用于根据所述对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个所述终端中。

[0078] 本申请包括以下优点:

[0079] 本申请至少一个实施例可以根据语音数据的来源标识,将不同来源的语音数据进行语音识别得到的文本信息分别显示在对应的终端中,由于能够从采集语音数据的源头开始进行区分,因此能够快速、准确地实现在终端中只显示某个或某些来源的语音数据对应的文本信息。

[0080] 当然,实施本申请的任一产品必不一定需要同时达到以上所述的所有优点。

附图说明

[0081] 图1是实施例1的文本信息显示方法的流程图;

[0082] 图2是应用实施例1的系统架构示意图;

[0083] 图3是实施例2的语音识别方法的流程图;

[0084] 图4是实施例3的文本信息显示方法的流程图;

[0085] 图5是实施例4的文本信息传输方法的流程图;

[0086] 图6是实施例5的文本信息显示系统的示意图;

[0087] 图7是实施例10的文本信息显示系统的示意图;

[0088] 图8是实施例15的字幕加配系统的示意图;

[0089] 图9是实施例15的实现方式之一的示意图;

[0090] 图10是实施例15的实现方式之二的示意图。

具体实施方式

[0091] 下面将结合附图及实施例对本申请的技术方案进行更详细的说明。

[0092] 需要说明的是,如果不冲突,本申请实施例以及实施例中的各个特征可以相互结合,均在本申请的保护范围之内。另外,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。

[0093] 本文中,所述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)等。

[0094] 计算机可读介质包括永久性和非永久性、可移动和非可移动的存储介质。存储介

质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。

[0095] 实施例1

[0096] 一种文本信息显示方法,预先设置有来源标识和终端的对应关系,如图1所示,所述方法包括步骤S110~S130。

[0097] S110、对采集到的不同来源的多路语音数据分别进行语音识别;

[0098] S120、确定所述多路语音数据各自对应的文本信息和来源标识;

[0099] S130、根据所述对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中。

[0100] 应用本实施例方法的一个系统架构如图2所示,系统中可包括一个或多个识别端设备21、以及一个或多个显示端设备22;

[0101] 识别端设备21可对语音采集设备采集到的不同来源的多路语音数据分别进行语音识别,确定多路语音数据各自对应的文本信息和来源标识。

[0102] 其中,对采集到的不同来源的多路语音数据分别进行语音识别可包括:

[0103] 分别将采集到的不同来源的多路语音数据发送到云端进行语音识别,从云端接收对多路语音数据进行语音识别得到的文本信息。

[0104] 其中,所述的云端可以是指一个具有语音识别能力的独立部署的服务器或者服务器集群,具体可以是运行在云计算中的服务器或服务器集群。

[0105] 其中,上述确定多路语音数据各自对应的文本信息和来源标识后还可包括:

[0106] 将多路语音数据各自对应的文本信息和来源标识对应保存在数据库中。通过将识别出来的文本信息以及来源标识保存在数据库中,可以减少识别设备的存储能力,同时也可以比较适用于图2所示的系统架构中,在该系统中由于识别端设备有多个,因此,设置数据库,可以便于将所有的语音识别数据进行保存,便于显示终端设备对文本信息的显示处理。

[0107] 显示端设备22可根据上述的对应关系,将多路语音数据各自对应的文本信息,分别显示在多路语音数据各自对应的来源标识所对应的一或多个终端中。其中,显示端设备对文本信息在终端进行显示,具体可以是显示端设备从数据库主动获取所有的文本信息及对应的来源标识,并基于对应关系来进行终端显示,也可以是根据上述的对应关系,主动从数据库中获取所需要显示的文本信息在终端进行显示,或者,也可以是由数据库主动推送所有的文本信息给显示端设备来进行上述操作,或者,也可以是由数据库根据对应关系向各显示端设备推送特定的文本信息,根据不同的应用场景可具有不同的处理方式。

[0108] 一种实现方式中,根据上述的对应关系,将多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个所述终端中具体可包括:

[0109] 根据上述的对应关系,确定不同终端对应的来源标识;

[0110] 从数据库中分别获取不同终端对应的来源标识所对应的文本信息,并显示在相应终端中。

[0111] 在从数据库中获取文本信息时,可以根据文本信息对应的语音数据的采集时间的先后顺序进行获取,也可以根据文本信息对应的来源标识的优先级从高到低的顺序进行获取。

[0112] 该图2所示的系统架构中,不同来源的多路语音数据可以是指由不同的语音采集设备采集到的语音数据,比如将每个话筒采集的语音数据作为一路语音数据,或者将每个话筒集合采集的语音数据作为一路语音数据(一个话筒集合包含一或多个话筒)。所述不同来源的多路语音数据也可以是指由同样的语音采集设备采集到的、属于不同用户的语音数据;比如多个用户使用同一个话筒说话时,可以通过声纹识别将一个话筒采集到的语音数据区分成不同来源的多路语音数据。在这里,所述的来源标识具体可以是指说话者。当然在其它场景下,该来源标识可以是指一类用户或一组用户,例如在一个多方沟通会议中,每个会议参与方可以有几个人,那么该来源标识具体可以是指一个会议参与方,而不是指参与方中的某个人,等。

[0113] 图2所示的系统架构中,语音采集设备和识别端设备21可以是不同的设备,比如识别端设备21外接的话筒作为语音采集设备。语音采集设备和识别端设备21也可以是同一个设备,比如智能手机或平板电脑,本身自带话筒,并能实现识别端设备21的功能。

[0114] 图2所示的系统架构中,终端和显示端设备22可以是不同设备,比如显示端设备22外接的显示屏作为终端。终端和显示端设备22也可以是同一个设备,比如是智能手机或平板电脑,本身自带显示屏,并能实现显示端设备22的功能。

[0115] 图2所示的系统架构中,每个识别端设备21可以只对一路语音数据进行语音识别。每个识别端设备21也可以通过运行多个识别线程,对多路语音数据分别进行语音识别。

[0116] 图2所示的系统架构中,每个显示端设备22对应一个终端,根据所对应的终端对应的来源标识,从数据库中获取相应的文本信息显示在所对应的终端中。每个显示端设备22也可对应一个或多个终端,可以通过运行多个文本信息显示线程,分别根据不同终端对应的来源标识,从数据库中获取相应的文本信息显示在相应终端中。

[0117] 在一种实施方式中,上述将所述多路语音数据各自对应的文本信息,分别显示在多路语音数据各自对应的来源标识所对应的一或多个终端中具体可包括:

[0118] 将多路语音数据各自对应的文本信息分别转换成字幕,显示在多路语音数据各自对应的来源标识所对应的一或多个终端中。其中,转换成字幕,具体是将语音识别出的文本信息转换成字幕格式的文件信息,以便与终端上显示的视频进行合成显示,在生成的字幕中,可以加入来源标识的相关信息,例如来源标识是某一个人时,可以加入该人的信息,例如李三说的语音,可以在字幕中加入李三,以提示其他用户这是李三说的话等,例如来源标识是一个设备或一组用户时,同样可以加入相关的信息在字幕中。

[0119] 一个直播应用场景下,可在直播过程中进行视频字幕加配,比如在会议、庭审、发布会等场合中,采集用户的语音数据后,通过实时的语音识别得到文本信息并转换成字幕;将字幕和对用户采集的视频数据一起显示在终端中。由于语音数据从产生到最终显示字幕之间所间隔的时间长度可以小到观众无法感知的程度,因此从观众的观感上来说,看到的是包含字幕的直播视频。

[0120] 将字幕和视频数据一起显示的方式可以但不限于包括：生成纯绿色背景的字幕图像，和视频数据中的图像进行合成；或者，生成外挂的字幕文件，在播放视频数据时加载。在一些应用场景中，也可以不显示视频，只显示文本信息。还可以用类似弹幕的方式在播放视频数据的过程中显示文本信息。

[0121] 在一些实施方式中，可以由终端将字幕和视频进行合成，也可以由另外的设备完成合成后在终端进行显示。

[0122] 在一些实施方式中，字幕转换和合成可以由同一个设备实现，也可以由不同设备实现。

[0123] 在一些实施方式中，完成合成的设备可以只针对一个终端要播放的视频和字幕进行合成，也可以通过运行多个合成线程的方式，针对不同终端分别合成视频和字幕。

[0124] 在一些实施方式中，各终端播放的字幕可以相同但视频不同，或者播放的字幕不同但视频相同；或者播放的视频和字幕均不同或均相同。在不同的应用场景下，可具有不同的显示处理方式。

[0125] 在一些实施方式中，可以在转换成的字幕中增加来源标识；比如一路语音数据对应于来源标识“用户A”，该路语音数据对应的文本信息转换成字幕后，可以在字幕中加上前缀或后缀“用户A”。如上述所述的来源标识为某个人时，可以在字幕中加入人的名字，或者某一组用户的名称等。

[0126] 在一些实施方式中，可以根据来源标识设置字幕的显示参数，来源标识不同时，字幕的显示参数也不同。比如两路语音数据的来源标识分别是“用户A”和“用户B”，将两路语音数据对应的文本信息分别转换成两份字幕后，可以一份字幕为蓝色，另一份字幕为红色。其中，显示参数可以包括以下一个或多个：颜色、字体、字号、格式（比如但不限于加粗、倾斜、加下划线、突出显示等）。

[0127] 在显示时加入来源标识，可以通过字幕一目了然地看出每个来源标识所代表的用户各自说了什么。

[0128] 在一些实施方式中，将多路语音数据各自对应的文本信息分别转换成字幕可以包括：

[0129] 在多路语音数据各自对应的文本信息中，将对应于预定的来源标识的语音数据所对应的文本信息转换成字幕。

[0130] 比如当需要只针对某个或某些用户的语音数据显示字幕时，可以将这个或这些用户的语音数据对应的来源标识，设置为所述预定的来源标识。

[0131] 在一些实施方式中，可以将多路语音数据各自对应的文本信息全部转换成字幕，当需要显示部分来源的语音数据生成的字幕时，根据来源标识在转换成的字幕中选择部分字幕显示。

[0132] 一种实施方式中，还可包括：根据终端所播放的视频，确定来源标识和终端的对应关系。

[0133] 对应关系的确定可以在步骤S130之前进行，以便作为预置信息。该对应关系中，来源标识和终端之间的对应方式可以包括以下任一种：多对多、多对一、一对多、一对一。

[0134] 在一些实施方式中，可通过设置对应关系，能够设置不同终端要显示哪些语音数据对应的文本信息，也能够设置不同来源的语音数据对应的文本信息将会显示在哪些终端

中。

[0135] 在一些实施方式中,该对应关系可以是静态配置的,比如每个终端固定显示一或多路语音数据对应的文本信息。该对应关系也可以是能够动态调整的;比如可以人为修改所述对应关系,也可以根据终端所播放的视频的来源或内容自适应更新该对应关系。

[0136] 其中,静态配置的情况举例如下:当某个终端播放针对用户A采集的视频数据时,或是该终端的视频源为对用户A进行采集的摄像头时,可以将针对用户A采集的语音数据对应的来源标识与该终端对应。

[0137] 其中,根据终端所播放的视频的内容自适应更新所述对应关系的情况举例如下:比如通过人脸识别技术,发现某个终端播放的视频中的人物从用户A变更为用户B,可以在该终端对应的来源标识中,增加针对用户B采集的语音数据对应的来源标识,删除针对用户A采集的语音数据对应的来源标识。再比如通过图像识别发现某个终端播放的视频中具有某个区域的标志物时,可以在该终端对应的来源标识中,增加该区域采集的语音数据对应的来源标识与该终端之间的对应。

[0138] 其中,根据终端所播放的视频的来源自适应更新所述对应关系的情况举例如下:当某个终端将视频源从摄像头X切换到摄像头Y时,可以在该终端对应的来源标识中,增加来自于摄像头Y的采集区域或对象的语音数据对应的来源标识,删除来自于摄像头Y的采集区域或对象的语音数据对应的来源标识。

[0139] 本可选实施方式中,所述对应关系可以集中保存,比如由服务器保存所述对应关系,即:服务器可以知道不同终端对应的来源标识(或者不同来源标识对应的终端)。所述对应关系也可以分散保存,比如每个设备分别只保存一个终端对应的一或多个来源标识,或只保存一个来源标识对应的一或多个终端。

[0140] 在其它可选实施方式中,所述终端对应的来源标识也可以和该终端播放的视频无关;比如无论播出何种视频,或无论是否播出视频,都不影响来源标识和终端的对应关系。

[0141] 本实施例的一种可选实施方式中,所述确定多路语音数据各自对应的来源标识包括:

[0142] 分别根据所述多路语音数据的采集端确定多路语音数据各自对应的来源标识;或者,通过对所述多路语音数据进行语音识别,分别确定所述多路语音数据各自对应的来源标识。

[0143] 本可选实施方式可以快速确定来源标识,基本不会增加额外的工作量。

[0144] 本可选实施方式中,确定所述语音数据的来源标识的各种可能的时间节点包括但不限于以下的一个或多个:采集语音数据时、采集语音数据后进行语音识别前、进行语音识别后。

[0145] 本可选实施方式中,根据采集端确定来源标识时,可以根据采集端本身的硬件标识得到所述来源标识。比如采用手机、平板电脑等智能终端采集语音数据时,来源标识可以是该智能终端的标识,比如但不限于IMSI(International Mobile Subscriber Identification Number,国际移动用户识别码)。再比如采用话筒采集语音数据后,再通过蓝牙模块发送,来源标识可以是蓝牙模块的标识。或者也可以根据采集端上运行的软件的标识确定所述来源标识,比如根据采集客户端的标识确定来源标识。

[0146] 类似地,本备选方案中,通过语音识别确定来源标识时,可以根据识别端设备的硬

件标识,或运行的识别客户端的标识确定所述来源标识,还可以根据对语音数据进行声纹识别所得到的用户身份确定所述来源标识。

[0147] 其中,所述采集客户端、识别客户端的标识可以在初始化时分配,也可以在对采集客户端或识别客户端进行认证的过程中分配。

[0148] 其中,一个识别客户端可以只对应一路语音数据,也可以对应多路语音数据。

[0149] 其中,当一个识别客户端对应多路语音数据,且来源标识是识别客户端的标识时,这多路语音数据对应的来源标识将相同。比如庭审现场中,将针对原告三个话筒采集的三路语音数据都发送到同一个标识为“原告方”的识别客户端,则这三路语音数据对应的来源标识都是“原告方”。

[0150] 其中,当一个识别客户端对应一路语音数据,且来源标识是识别客户端的标识时,多路语音数据对应的来源标识也有可能部分相同。比如庭审现场中,将针对原告三个话筒采集的三路语音数据分别发送到三个识别客户端,如果这三个识别客户端的标识均为“原告方”,则这三路语音数据对应来源标识都是“原告方”。

[0151] 其它可选实施方式中,可以通过用户自行选择来源标识,或者将用户在采集客户端或识别客户端中登录的账号作为来源标识。

[0152] 实施例2

[0153] 一种语音识别方法,如图3所示,包括步骤S210~S220。

[0154] S210、对采集到的一或多路语音数据进行语音识别;

[0155] S220、确定所述一或多路语音数据各自对应的文本信息和来源标识。

[0156] 一些实施方式中,确定所述一或多路语音数据各自对应的文本信息和来源标识后,可以根据预置的来源标识和终端的对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中;实现细节可参见实施例1。

[0157] 一些实施方式中,所述步骤S210~S220可以由如图2所示的识别端设备执行。

[0158] 一些实施方式中,所述步骤S210~S220可以由运行多个识别客户端的设备或服务器执行。

[0159] 一些可选实施方式中,所述对采集到的一或多路语音数据进行语音识别可以包括:

[0160] 将采集到的一或多路语音数据发送到云端进行语音识别,从云端接收对所述一或多路语音数据进行语音识别得到的文本信息。

[0161] 其中,所述的云端可以是指一个具有语音识别能力的独立部署的服务器或者服务器集群,具体可以是运行在云计算中的服务器或服务器集群。

[0162] 一些可选实施方式中,所述确定所述一或多路语音数据各自对应的文本信息和来源标识后还可以包括:

[0163] 将所述一或多路语音数据各自对应的文本信息和来源标识对应保存在数据库中。

[0164] 如果需要根据预置的来源标识和终端的对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中,则具体可以根据预置的来源标识和终端的对应关系,确定不同终端对应的来源标识;从所述数据库中分别获取不同终端对应的来源标识所对应的文本信息,并显示在相应终端

中。本实施例的步骤S210~S220的其它实现细节可参见实施例1。

[0165] 实施例3

[0166] 一种文本信息显示方法,如图4所示,包括步骤S310~S320:

[0167] S310、获取一个或多个来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;

[0168] S320、根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中。

[0169] 其中,可以先对语音数据进行语音识别,确定语音数据对应的文本信息和来源标识后,将语音数据对应的文本信息和来源标识对应保存。确定语音数据对应的来源标识和文本信息的实现细节可参见实施例1。

[0170] 一些实施方式中,本实施例的方法可以由图2所示的显示端设备执行。

[0171] 一些实施方式中,所述步骤S310~S320可以用于一个终端的文本信息显示,即:步骤S310中获取的一或多个来源标识是对应于一个终端的。

[0172] 一些实施方式中,所述步骤S310~S320可以用于多个终端的文本信息显示;即:步骤S310中获取的是多个终端对应的来源标识。

[0173] 一些实施方式中,所述将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中可以包括:

[0174] 将所述一个或多个来源标识对应的文本信息分别转换成字幕,显示在所述一个或多个来源标识所对应的终端中。

[0175] 其中,转换成字幕,具体可以是将语音识别出的文本信息转换成字幕格式的文件信息,以便与终端上显示的视频进行合成显示。

[0176] 一些实施方式中,所述获取一个或多个来源标识对应的文本信息可以包括:

[0177] 从对应保存来源标识和文本信息的数据库中,获取一个或多个来源标识各自对应的文本信息。

[0178] 其中,可以从数据库主动获取所有的文本信息,并基于对应关系来进行终端显示,也可以是根据上述的对应关系,主动从数据库中获取所需要显示的来源标识对应的文本信息,在终端进行显示,或者,也可以是由数据库主动推送所有的文本信息来进行上述操作,或者,也可以是由数据库根据对应关系针对不同终端分别推送特定的文本信息,根据不同的应用场景可具有不同的处理方式。

[0179] 其中,可以在对语音数据进行语音识别,确定语音数据对应的文本信息和来源标识后,将语音数据对应的文本信息和来源标识对应保存在数据库中,以便在显示时从数据库中获取相应的文本信息。

[0180] 一些实施方式中,所述获取一个或多个来源标识对应的文本信息包括:

[0181] 根据预置的来源标识和终端的对应关系,获取对应于预定的一或多个来源标识的文本信息。

[0182] 其中,所述预定的一个或多个来源标识可以是所述对应关系中对应于预定的一个或多个终端的来源标识。

[0183] 本实施例的方法可以和实施例2配合使用。

[0184] 本实施例的步骤S310~S320的其它实现细节可参见实施例1。

[0185] 实施例4

[0186] 一种文本信息传输方法,预先保存有相互对应的文本信息和来源标识,其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;如图5所示,所述方法包括S410~S420:

[0187] S410、根据预置的来源标识和终端的对应关系,分别确定不同终端所对应的来源标识;

[0188] S420、分别向不同终端发送与该终端对应的来源标识所对应的文本信息。

[0189] 其中,可以先对语音数据进行语音识别,确定语音数据对应的文本信息和来源标识后,将语音数据对应的文本信息和来源标识对应保存。确定语音数据对应的来源标识和文本信息的实现细节可参见实施例1。

[0190] 在分别向不同终端发送与该终端对应的来源标识所对应的文本信息后,就可以使得不同来源的多路语音数据各自对应的文本信息,根据预置的来源标识和终端的对应关系,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中。显示时的实现细节可参见实施例1。一些实施方式中,可以由保存有相互对应的文本信息和来源标识的服务器执行步骤410~420,从而针对不同终端对应的来源标识,向不同终端推送对应于相应来源标识的文本信息。

[0191] 其中,该服务器可以本地保存相互对应的文本信息和来源标识,也可以从其它设备获取相互对应的文本信息和来源标识。

[0192] 其中,该服务器可以本地保存来源标识和终端的对应关系,也可以从其它设备获取来源标识和终端的对应关系。

[0193] 实施例5

[0194] 一种文本信息显示系统,如图6所示,包括:

[0195] 一个或多个语音识别装置51,一个或多个文本信息显示装置52;

[0196] 至少一个所述语音识别装置51用于对一或多路采集到的语音数据进行语音识别,确定所述一或多路采集到的语音数据各自对应的文本信息和来源标识;

[0197] 至少一个所述文本信息显示装置52用于根据预置的来源标识和终端的对应关系,将一或多路语音数据对应的文本信息,显示在所述一或多路语音数据对应的来源标识所对应的终端中。

[0198] 图6中所述语音识别装置51和所述文本信息显示装置52之间的连线只代表两者之间存在文本信息的传递关系,但并不代表两者必须是直接传递文本信息也不代表两者必须一一对应。

[0199] 本实施例中,所述语音识别装置51可以用于执行实施例1中的步骤S110~S120;所述语音识别装置51可以作为图2所示的识别端设备。所述文本信息显示装置52可以用于执行实施例1中的步骤S130;所述文本信息显示装置52可以作为图2所示的显示端设备。

[0200] 一些实施方式中,一个所述语音识别装置51可以只运行一个识别客户端。一些实施方式,一个所述语音识别装置51可以运行多个识别客户端。

[0201] 一些实施方式中,一个所述文本信息显示装置52可以只为一个终端服务,根据该终端对应的来源标识,将相应的一或多路语音数据对应的文本信息显示在该终端中。另一种应用场景中,一个所述文本信息显示装置52可以为多个终端服务,分别根据不同终端对

应的来源标识,将相应的一或多路语音数据对应的文本信息显示在相应终端中。

[0202] 一些实施方式中,至少部分所述语音识别装置51和至少部分所述文本信息显示装置52可以复用硬件资源。

[0203] 一些实施方式中,所述文本信息显示装置52将一或多路语音数据对应的文本信息,显示在所述一或多路语音数据对应的来源标识所对应的终端中可以包括:

[0204] 将一或多路语音数据对应的文本信息转换成字幕,显示在所述一或多路语音数据对应的来源标识所对应的终端中。

[0205] 其中,转换成字幕,具体可以是将语音识别出的文本信息转换成字幕格式的文件信息,以便与终端上显示的视频进行合成显示。

[0206] 一些实施方式中,所述语音识别装置51对一或多路采集到的语音数据进行语音识别可以包括:

[0207] 将一或多路采集到的语音数据发送到云端进行语音识别,从云端接收对所述一或多路语音数据进行语音识别得到的文本信息。

[0208] 其中,所述的云端可以是指一个具有语音识别能力的独立部署的服务器或者服务器集群,具体可以是运行在云计算中的服务器或服务器集群。

[0209] 一些实施方式中,至少一个所述语音识别装置51还可以用于将所述一或多路语音数据各自对应的文本信息和来源标识对应保存在数据库中;

[0210] 至少一个所述文本信息显示装置52还可以用于从所述数据库中获取一或多路语音数据对应的文本信息。

[0211] 可以是文本信息显示装置52从数据库主动获取所有的文本信息及对应的来源标识,并基于对应关系来进行终端显示,也可以是根据上述的对应关系,主动从数据库中获取所需要显示的文本信息在终端进行显示,或者,也可以是由数据库主动推送所有的文本信息给文本信息显示装置52来进行上述操作,或者,也可以是由数据库根据对应关系向各文本信息显示装置52推送特定的文本信息,根据不同的应用场景可具有不同的处理方式。

[0212] 一些实施方式中,所述文本信息显示装置52从所述数据库中获取一或多路语音数据对应的文本信息可以包括:

[0213] 根据预置的来源标识和终端的对应关系,从所述数据库中获取对应于预定的一或多个来源标识的文本信息。

[0214] 其中,当一个文本信息显示装置52为一个终端服务时,所述预定的一个或多个来源标识可以是所述对应关系中对应于预定的一个终端的来源标识;当一个文本信息显示装置52为多个终端服务时,所述预定的一个或多个来源标识可以是所述对应关系中对应于多个终端的来源标识。

[0215] 本实施例中语音识别装置51和文本信息显示装置52的其它实现细节可参见实施例1中对图2所示的系统架构的描述。

[0216] 实施例6

[0217] 一种语音识别装置,包括:存储器和处理器;

[0218] 所述存储器用于存储用于语音识别的程序;所述用于语音识别的程序在被所述处理器读取执行时,执行如下操作:

[0219] 对采集到的一或多路语音数据进行语音识别;

[0220] 确定所述一或多路语音数据各自对应的文本信息和来源标识。

[0221] 一些实施方式中,确定所述一或多路语音数据各自对应的文本信息和来源标识后,可以根据预置的来源标识和终端的对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中;实现细节可参见实施例1。

[0222] 一些实施方式中,所述语音识别装置可以作为图2所示的识别端设备。

[0223] 一些实施方式中,所述对采集到的一或多路语音数据进行语音识别可以包括:

[0224] 将采集到的一或多路语音数据发送到云端进行语音识别,从云端接收对所述一或多路语音数据进行语音识别得到的文本信息。

[0225] 其中,所述的云端可以是指一个具有语音识别能力的独立部署的服务器或者服务器集群,具体可以是运行在云计算中的服务器或服务器集群。

[0226] 一些实施方式中,所述确定所述一或多路语音数据各自对应的文本信息和来源标识后还可以包括:

[0227] 将所述一或多路语音数据各自对应的文本信息和来源标识对应保存在数据库中。

[0228] 如果需要根据预置的来源标识和终端的对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中,则具体可以根据预置的来源标识和终端的对应关系,确定不同终端对应的来源标识;从所述数据库中分别获取不同终端对应的来源标识所对应的文本信息,并显示在相应终端中。

[0229] 本实施例的装置实现语音识别、以及确定所述一或多路语音数据各自对应的文本信息和来源标识的细节可参见实施例1。

[0230] 实施例7

[0231] 一种语音采集装置,包括:话筒、存储器、处理器;

[0232] 所述话筒用于采集语音数据;

[0233] 所述存储器用于存储用于语音识别的程序;所述用于语音识别的程序在被所述处理器读取执行时,执行如下操作:

[0234] 对所述话筒所采集到的语音数据进行语音识别;

[0235] 确定所述语音数据对应的文本信息和来源标识。

[0236] 本实施例提供了一种集采集、识别为一体的装置,比如可以是一个带有话筒的智能手机或平板电脑,也可以看成是一个具备处理能力的“智能话筒”。

[0237] 一些实施方式中,确定所述语音数据对应的文本信息和来源标识后,可以根据预置的来源标识和终端的对应关系,将所述语音数据对应的文本信息,显示在所述语音数据对应的来源标识所对应的一或多个终端中;实现细节可参见实施例1。

[0238] 一些实施方式中,所述语音采集装置可以作为对图2所示的语音采集设备和识别端设备进行集成的设备。

[0239] 本实施例中,进行语音识别、确定所述语音数据对应的文本信息和来源标识的实现细节可以参见实施例1。

[0240] 实施例8

[0241] 一种文本信息处理装置,包括:存储器、处理器;

[0242] 所述存储器用于存储用于文本信息显示的程序;所述用于文本信息显示的程序在被所述处理器读取执行时,执行如下操作:

[0243] 获取一个或多个来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;

[0244] 根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中。

[0245] 其中,可以由其它装置(比如但不限于实施例6或7的装置)对语音数据进行语音识别,确定语音数据对应的文本信息和来源标识,并将语音数据对应的文本信息和来源标识对应保存。确定语音数据对应的来源标识和文本信息的实现细节可参见实施例1。

[0246] 一些实施方式中,所述文本信息处理装置可以作为图2所示的显示端设备。

[0247] 一些实施方式中,所述将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中可以包括:

[0248] 将所述一个或多个来源标识对应的文本信息分别转换成字幕,显示在所述一个或多个来源标识所对应的终端中。

[0249] 其中,转换成字幕,具体可以是将语音识别出的文本信息转换成字幕格式的文件信息,以便与终端上显示的视频进行合成显示。

[0250] 一些实施方式中,所述获取一个或多个来源标识对应的文本信息可以包括:

[0251] 从对应保存来源标识和文本信息的数据库中,获取一个或多个来源标识各自对应的文本信息。

[0252] 其中,可以在对语音数据进行语音识别,确定语音数据对应的文本信息和来源标识后,将语音数据对应的文本信息和来源标识对应保存在数据库中,以便在显示时从数据库中获取相应的文本信息。

[0253] 其中,可以是数据库主动获取所有的文本信息,并基于对应关系来进行终端显示,也可以是根据上述的对应关系,主动从数据库中获取所需要显示的来源标识对应的文本信息,在终端进行显示,或者,也可以是由数据库主动推送所有的文本信息来进行上述操作,或者,也可以是由数据库根据对应关系针对不同终端分别推送特定的文本信息,根据不同的应用场景可具有不同的处理方式。

[0254] 一些实施方式中,所述获取一个或多个来源标识对应的文本信息可以包括:

[0255] 根据预置的来源标识和终端的对应关系,获取对应于预定的一或多个来源标识的文本信息。

[0256] 其中,所述预定的一个或多个来源标识可以是所述对应关系中对应于预定的一个或多个终端的来源标识。本实施例中,根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中的实现细节可以参见实施例1。

[0257] 实施例9

[0258] 一种文本信息显示装置,包括:显示屏、存储器、处理器;

[0259] 所述存储器用于存储用于文本信息显示的程序;所述用于文本信息显示的程序在被所述处理器读取执行时,执行如下操作:

[0260] 获取预定的来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对

该来源标识对应的语音数据进行语音识别得到；

[0261] 将所获取的文本信息显示在所述显示屏中。

[0262] 本实施例提供了一种可以根据来源标识在视频中显示相应语音数据对应的文本信息的显示装置,预定的来源标识即与所述显示装置对应的来源标识。该显示装置比如可以是一个智能手机或平板电脑,也可以看成是一个具备处理能力的“智能显示屏”。

[0263] 其中,可以由其它装置(比如但不限于实施例6或7的装置)对语音数据进行语音识别,确定语音数据对应的文本信息和来源标识,并将语音数据对应的文本信息和来源标识对应保存。确定语音数据对应的来源标识和文本信息的实现细节可参见实施例1。

[0264] 一些实施方式中,所述文本信息显示装置可以作为对图2所示的显示端设备和终端进行集成的设备。

[0265] 一些实施方式中,所述文本信息显示装置中的存储器还可以用于存储用于进行字幕生成、合成的程序;所述将所获取的文本信息显示在所述显示屏中可以包括:

[0266] 将获取的文本信息转换成字幕,与所述装置接收的视频数据一起显示。

[0267] 其中,转换成字幕,具体可以是将语音识别出的文本信息转换成字幕格式的文件信息,以便与终端上显示的视频进行合成显示。

[0268] 本实施例中,根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中的实现细节可以参见实施例1。

[0269] 实施例10、一种处理装置,包括:

[0270] 第一装置,是实施例6所述的语音识别装置和实施例7所述的语音采集装置中的任一个;

[0271] 第二装置,是实施例8所述的文本信息显示装置和实施例9所述的显示装置中的任一个。

[0272] 其中,第一装置和第二装置可以复用存储器和/或处理器。

[0273] 所述处理装置可以是一个智能手机或平板电脑,自带话筒、显示屏,还包含网卡、蓝牙模块等通信模块。在一种应用场景中,所述处理装置可以通过自带的话筒采集用户的语音数据,识别为文本信息后发送到数据库;另外从数据库获取预定的来源标识的文本信息,直接或转换字幕后通过自带的显示屏进行显示。在另一种应用场景中,所述处理装置也可以外接话筒和显示屏;其它操作与上述应用场景类似。

[0274] 本实施例中第一装置、第二装置的实现细节可参见实施例6~9。

[0275] 实施例11

[0276] 一种文本信息显示系统,如图7所示,预先设置有来源标识和终端的对应关系,所述系统包括:

[0277] 语音识别模块71,用于对采集到的不同来源的多路语音数据分别进行语音识别;

[0278] 确定模块72,用于确定所述多路语音数据各自对应的文本信息和来源标识;

[0279] 显示模块73,用于根据所述对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个所述终端中。

[0280] 本实施例中,所述语音识别模块71是所述系统中负责进行语音识别的部分,可以是软件、硬件或两者的结合。

[0281] 本实施例中,所述确定模块72是所述系统中负责确定语音数据对应的文本信息和来源标识的部分,可以是软件、硬件或两者的结合。

[0282] 本实施例中,所述显示模块73是所述系统中负责将语音数据对应的文本信息显示在终端中的部分,可以是软件、硬件或两者的结合。

[0283] 一些实施方式中,所述显示模块73将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中可以包括:

[0284] 将所述多路语音数据各自对应的文本信息分别转换成字幕,显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中。

[0285] 一些实施方式中,所述将所述多路语音数据各自对应的文本信息分别转换成字幕可以包括:

[0286] 在所述多路语音数据各自对应的文本信息中,将对应于预定的来源标识的语音数据所对应的文本信息转换成字幕。

[0287] 其中,转换成字幕,具体可以是将语音识别出的文本信息转换成字幕格式的文件信息,以便与终端上显示的视频进行合成显示。

[0288] 一些实施方式中,所述系统还可以包括:

[0289] 设置模块,用于根据终端所播放的视频,确定所述来源标识和终端的对应关系。

[0290] 一些实施方式中,所述确定模块72确定多路语音数据各自对应的来源标识可以包括:

[0291] 分别根据所述多路语音数据的采集端确定多路语音数据各自对应的来源标识;或者,通过对所述多路语音数据进行语音识别,分别确定所述多路语音数据各自对应的来源标识。

[0292] 一些实施方式中,所述语音识别模块71对采集到的不同来源的多路语音数据分别进行语音识别可以包括:

[0293] 分别将采集到的不同来源的多路语音数据发送到云端进行语音识别,从云端接收对所述多路语音数据进行语音识别得到的文本信息。

[0294] 其中,所述的云端可以是指一个具有语音识别能力的独立部署的服务器或者服务器集群,具体可以是运行在云计算中的服务器或服务器集群。

[0295] 一些实施方式中,所述确定模块72确定所述多路语音数据各自对应的文本信息和来源标识后还可以包括:

[0296] 将所述多路语音数据各自对应的文本信息和来源标识对应保存在数据库中;

[0297] 所述根据所述对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个所述终端中可以包括:

[0298] 根据所述对应关系,确定不同终端对应的来源标识;

[0299] 从所述数据库中分别获取不同终端对应的来源标识所对应的文本信息,并显示在相应终端中。

[0300] 本实施例中,各模块分别对应实现实施例1中的步骤S110~S130,各模块的实现细节可参见实施例1中对应的步骤的实现细节。

[0301] 实施例12

[0302] 一种语音识别装置,包括:

- [0303] 识别模块,用于对采集到的一或多路语音数据进行语音识别;
- [0304] 文本信息及来源标识确定模块,用于确定所述一或多路语音数据各自对应的文本信息和来源标识。
- [0305] 一些实施方式中,确定所述一或多路语音数据各自对应的文本信息和来源标识后,可以根据预置的来源标识和终端的对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中;实现细节可参见实施例1。
- [0306] 一些实施方式中,本实施例的所述语音识别装置可以作为图2所示的识别端设备。
- [0307] 本实施例中,所述识别模块是所述装置中负责进行语音识别的部分,可以是软件、硬件或两者的结合。
- [0308] 本实施例中,所述文本信息及来源标识确定模块是所述装置中负责确定语音数据对应的文本信息和来源标识的部分,可以是软件、硬件或两者的结合。
- [0309] 一些实施方式中,所述识别模块对采集到的一或多路语音数据进行语音识别可以包括:
- [0310] 将采集到的一或多路语音数据发送到云端进行语音识别,从云端接收对所述一或多路语音数据进行语音识别得到的文本信息。
- [0311] 其中,所述的云端可以是指一个具有语音识别能力的独立部署的服务器或者服务器集群,具体可以是运行在云计算中的服务器或服务器集群。
- [0312] 一些实施方式中,所述文本信息及来源标识确定模块确定所述一或多路语音数据各自对应的文本信息和来源标识后还可以包括:
- [0313] 将所述一或多路语音数据各自对应的文本信息和来源标识对应保存在数据库中。
- [0314] 如果需要根据预置的来源标识和终端的对应关系,将所述多路语音数据各自对应的文本信息,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中,则具体可以根据预置的来源标识和终端的对应关系,确定不同终端对应的来源标识;从所述数据库中分别获取不同终端对应的来源标识所对应的文本信息,并显示在相应终端中。
- [0315] 本实施例中,各模块分别对应实现实施例2中的步骤S210~S220,各模块的实现细节可参见实施例2。
- [0316] 实施例13
- [0317] 一种文本信息显示装置,包括:
- [0318] 获取模块,用于获取一个或多个来源标识对应的文本信息;其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;
- [0319] 文本信息显示模块,用于根据预置的来源标识和终端的对应关系,将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中。
- [0320] 其中,可以由其它装置(比如但不限于为实施例12的装置)对语音数据进行语音识别,确定语音数据对应的文本信息和来源标识,并将语音数据对应的文本信息和来源标识对应保存。确定语音数据对应的来源标识和文本信息的实现细节可参见实施例1。
- [0321] 一些实施方式中,本实施例的方法可以由图2所示的显示端设备执行。
- [0322] 本实施例中,所述获取模块是所述系统中负责获取来源标识对应的文本信息的部

分,可以是软件、硬件或两者的结合。

[0323] 本实施例中,所述文本信息显示模块是所述装置中负责将语音数据对应的文本信息显示在终端中的部分,可以是软件、硬件或两者的结合。

[0324] 一些实施方式中,所述显示单元将所述一个或多个来源标识对应的文本信息,显示在所述一个或多个来源标识所对应的终端中包括:

[0325] 将所述一个或多个来源标识对应的文本信息分别转换成字幕,显示在所述一个或多个来源标识所对应的终端中。

[0326] 其中,转换成字幕,具体可以是将语音识别出的文本信息转换成字幕格式的文件信息,以便与终端上显示的视频进行合成显示。

[0327] 一些实施方式中,所述获取单元获取一个或多个来源标识对应的文本信息可以包括:

[0328] 从对应保存来源标识和文本信息的数据库中,获取一个或多个来源标识各自对应的文本信息。

[0329] 一些实施方式中,所述获取单元所述获取一个或多个来源标识对应的文本信息可以包括:

[0330] 根据预置的来源标识和终端的对应关系,获取对应于预定的一或多个来源标识的文本信息。

[0331] 其中,所述预定的一个或多个来源标识可以是所述对应关系中对应于预定的一个或多个终端的来源标识。

[0332] 本实施例中,各模块分别对应实现实施例3中的步骤S310~S320,各模块的实现细节可参见实施例3。

[0333] 实施例14

[0334] 一种文本信息传输装置,预先保存有相互对应的文本信息和来源标识,其中,来源标识所对应的文本信息通过对该来源标识对应的语音数据进行语音识别得到;所述装置包括:

[0335] 来源标识确定模块,用于根据预置的来源标识和终端的对应关系,分别确定不同终端所对应的来源标识;

[0336] 发送模块,用于分别向不同终端发送与该终端对应的来源标识所对应的文本信息。

[0337] 其中,可以先对语音数据进行语音识别,确定语音数据对应的文本信息和来源标识后,将语音数据对应的文本信息和来源标识对应保存。确定语音数据对应的来源标识和文本信息的实现细节可参见实施例1。

[0338] 在分别向不同终端发送与该终端对应的来源标识所对应的文本信息后,就可以使得不同来源的多路语音数据各自对应的文本信息,根据预置的来源标识和终端的对应关系,分别显示在所述多路语音数据各自对应的来源标识所对应的一或多个终端中。显示时的实现细节可参见实施例1。

[0339] 本实施例中,所述来源标识确定模块是所述装置中负责确定来源标识的部分,可以是软件、硬件或两者的结合。

[0340] 本实施例中,所述发送模块是所述装置中负责发送来源标识对应的文本信息的部

分,可以是软件、硬件或两者的结合。

[0341] 本实施例中,各模块分别对应实现实施例4中的步骤S410~S420,各模块的实现细节可参见实施例4。

[0342] 实施例15

[0343] 一种视频字幕加配系统,包括两大部分功能:第一是采集现场音频输入,并通过云端的语音识别服务完成语音数据到文本信息的实时语音识别功能;在语音识别中将确定语音数据的来源标识,来源标识和文本信息将对应保存在识别结果数据库中,相当于实现图2所示的语音采集设备和识别端设备的功能;第二是从识别结果数据库中实时获取文本信息,产生字幕,并跟直播现场采集的视频数据进行合成后播出,完成最终的直播图像,相当于实现图2所示的显示端设备和终端的功能。

[0344] 本实施例的系统可以应用于演讲、会议等场景的视频直播的实时字幕加配场景。

[0345] 所述系统如图8所示,包括:

[0346] 音频输入硬件(即采集端),完成音频捕捉的功能,本实施例中,音频输入硬件可以包括话筒,音频传输线路和声卡;它完成的功能是采集现场的语音输入,并转换成为数字化的语音数据;

[0347] 识别客户端,负责确定语音数据对应的文本信息和来源标识,并将同意语音数据对应的文本信息和来源标识对应保存到识别结果数据库中;

[0348] 云端语音识别服务,负责接收识别客户端发送过来的语音数据,并将其转换为文本信息发送回给识别客户端;

[0349] 识别结果数据库,完成文本信息和来源标识对应保存的功能;识别结果数据库作为各个识别客户端和字幕产生客户端的信息交换中介;

[0350] 字幕产生客户端,负责根据来源标识和终端的对应关系,确定所连接的终端对应的来源标识,获取识别结果数据库中与所确定的来源标识对应的文本信息;将获取的文本信息转换为纯字幕的图像,比如通常为纯绿色屏幕中的文字字幕等;

[0351] 现场图像采集模块,作为视频源,负责采集现场的视频数据;

[0352] 合成图像模块,完成的功能是将现场图像采集模块返回的视频数据中的图像跟字幕产生客户端产生的字幕图像进行叠加合成,生成图像叠加字幕的最终直播视频,用于视频直播或者投屏等终端进行播放。

[0353] 本实施例中,识别客户端和音频输入硬件是一一对应的关系,所述系统可以包含多个“音频输入硬件+识别客户端”的组合,比如图9中有N组。其中每个组合对应于一个用户或用户集合的语音数据,以该组合中识别客户端被标注的角色信息作为该组合对应的语音数据的来源标识。每个组合所输出的语音数据对应的文本信息,与该语音数据对应的来源标识也相互对应,从而可以在文本信息中区分不同用户/用户集合的说话内容。

[0354] 本实施例中,字幕产生客户端、现场图像采集模块、合成图像模块是一一对应的关系,所述系统可以包含多个“字幕产生客户端+现场图像采集模块+合成图像模块”的组合,比如图9中有M组。其中每个组合可以表示一个显示需求,每个显示需求可以显示不同的“字幕+视频”组合;一个显示需求可以对应于一或多个终端。比如一个例子中,发布会的现场会有大屏幕投影显示当前发言的用户的视频,并加配该用户的语音数据对应的文本信息所转换成的字幕;同时会做视频网络直播,包含发言的用户及直播主持人的画面,分别加配发言

用户和直播主持人的语音数据对应的文本信息所转换成的字幕。

[0355] 本实施例中,进行视频字幕加配的过程如下:

[0356] 在系统启动后,识别客户端将获取音频输入硬件采集到的语音数据,并将获取的语音数据通过网络实时的发送到云端的语音识别服务,完成语音识别的功能,获取云端服务器返回的作为识别结果的文本信息;

[0357] 识别客户端在获取语音识别得到的语音数据对应的文本信息之后,将语音数据对应的文本信息以及来源标识通过网络发送给识别结果数据库对应保存;其中,来源标识是识别客户端的标识,本实施例中是为识别客户端标注的用户角色信息,比如“发言人A”或“原告方”等;

[0358] 字幕产生客户端从识别结果数据库获取预定的一个或多个来源标识对应的文本信息,将获取的文本信息转换成字幕;

[0359] 合成图像模块将字幕产生客户端转换成的字幕,与现场图像采集模块采集的视频数据进行合成后,发送给相应终端进行播放。

[0360] 本实施例的一种应用场景是法庭的现场庭审过程;庭审现场划分为四个区域:被告席、原告席、审判席、证人席;每个区域各配备有话筒作为音频输入硬件,配备有摄像头作为现场图像采集模块。另外还可以配备能对全局进行拍摄的摄像头。现场可以配备多个显示屏作为所述终端,比如一个全体人员可以看到的大屏幕,一个专供审判席查看的小屏幕,一个在法庭之外用于直播的屏幕。

[0361] 每个区域的话筒均可以为一个或多个。当一个区域中分别为不同人员准备多个话筒时,这多个话筒采集的语音数据对应的来源标识可以相同,也可以不同。比如将原告席的多个话筒所采集的语音信息对应的来源标识都设置成“原告方”,也可以将多个话筒各自采集的语音信息对应的来源标识分别设置成“原告”、“原告代理律师”等。

[0362] 每个区域的摄像头也可以有多个,分别对不同人物进行拍摄。摄像头可以直接连接到合成图像模块,也可以通过数据库或其它中介为合成图像模块提供视频数据。

[0363] 每个合成图像模块可以固定将四个区域的摄像头中的一个作为视频源,也可以将多个摄像头作为视频源,播放时从多个摄像头中选择一个拍摄的视频数据进行字幕合成,播放过程中还可以根据指令或需求将视频源切换到另外的摄像头。

[0364] 所述显示屏播放的视频数据即对应的合成图像模块所提供的视频数据。所述显示屏与来源标识之间的对应关系可以根据显示屏锁播放的视频数据确定或修改。比如在一个播放对全局进行拍摄的画面的显示屏中,显示的是针对法庭中不同用户采集的语音数据所对应的文本信息转换成的字幕;在一个播放对某个区域拍摄的画面中的显示屏中,显示的是针对该区域中的用户采集的语音数据所对应的文本信息转换成的字幕。

[0365] 当显示屏中播放不同用户的语音数据所对应的文本信息转换成的字幕时,字幕中包含来源标识;或者,不同用户的语音数据所对应的文本信息转换成的字幕,采用不同的显示参数加以区分。

[0366] 本应用场景包含两种较为典型的实现方式,以被告席、原告席、审判席、证人席各有一个话筒、一个摄像头和一个显示屏的情况为例进行说明。

[0367] 一种实现方式如图9所示,每个区域的话筒、摄像头、显示屏均连接到该区域的一个处理装置(也可以话筒、摄像头、显示屏分别连接到该区域的不同处理装置);每个处理装

置均可以与服务器之间进行数据传输。所述服务器上针对每一个处理装置运行一个识别客户端、字幕产生客户端、以及用于合成字幕和视频数据的合成线程。所述服务器维护所述识别结果数据库(也可以由另外的设备维护所述识别结果数据库,并与该服务器之间进行数据传输),通过云端语音识别服务得到语音数据对应的文本信息并保存。所述摄像头拍摄的视频数据通过所连接的处理装置发送到所述服务器,由所述服务器提供给相应的合成线程使用。

[0368] 另一种可选的实施方式如图10所示,每个区域的话筒、摄像头、显示屏均连接到该区域的一个处理装置(也可以话筒、摄像头、显示屏分别连接到该区域的不同处理装置);话筒所连接的处理装置中运行所述识别客户端,通过云端语音识别服务得到语音数据对应的文本信息并保存在识别结果数据库中。显示屏所连接的处理装置中运行所述字幕产生客户端及合成图像模块。每个处理装置均与服务器上的识别结果数据库进行数据交互。所述摄像头拍摄的视频数据通过所连接的处理装置,直接发送或通过其它设备中转到需要播放该视频数据的显示屏所连接的处理装置。

[0369] 图9所示的实现方式相当于将语音识别、字幕生成、合成都集中在服务器中完成,每个区域的处理装置只负责数据传输。来源标识和终端的对应关系由服务器维护;服务器可以监控全局的数据流向,在需要时根据预定规则更新所述对应关系,或根据指令更新所述对应关系。通过对服务器中的所述对应关系进行更新,可以一次性调整系统中字幕的显示位置。

[0370] 图10所示的实现方式相当于将语音识别、字幕生成、合成都放在每个区域的处理装置中自行完成,服务器只用于维护所述识别结果数据库。来源标识和终端的对应关系需要由每个区域的处理装置分别去维护,每个处理装置中只包含所述对应关系中和本处理装置相关的部分,比如本处理装置连接的显示屏所对应的来源标识;所述处理装置可以自发或根据指令更新所述对应关系,但如果要改变全局的显示策略,则需要对多个处理装置里的所述对应关系分别进行更新。

[0371] 当然,也可以采用图9、图10之外的实现方式,比如将语音识别、字幕生成、合成中的一种或多种操作放到处理装置里完成,另外的操作放到服务器中完成。

[0372] 本实施例中,识别客户端实现图2所示的识别端设备的功能,字幕产生客户端和合成图像模块(或字幕产生客户端、合成图像模块及现场图像采集模块)共同实现图2所示的显示端设备的功能,实现细节可参见实施例1。

[0373] 本领域普通技术人员可以理解上述方法中的全部或部分步骤可通过程序来指令相关硬件完成,所述程序可以存储于计算机可读存储介质中,如只读存储器、磁盘或光盘等。上述实施例的全部或部分步骤也可以使用一个或多个集成电路来实现。相应地,上述实施例中的各模块/单元可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。本申请不限制于任何特定形式的硬件和软件的结合。

[0374] 当然,本申请还可有其他多种实施例,在不背离本申请精神及其实质的情况下,熟悉本领域的技术人员当可根据本申请作出各种相应的改变和变形,但这些相应的改变和变形都应属于本申请的权利要求的保护范围。

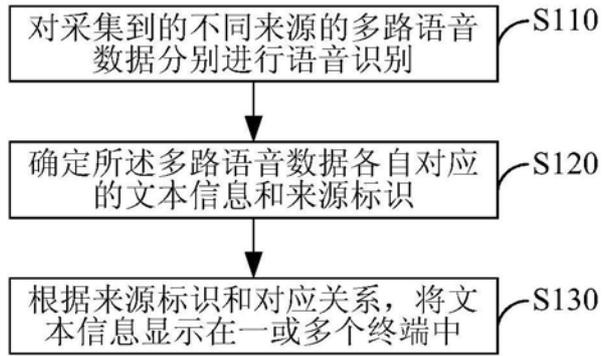


图1

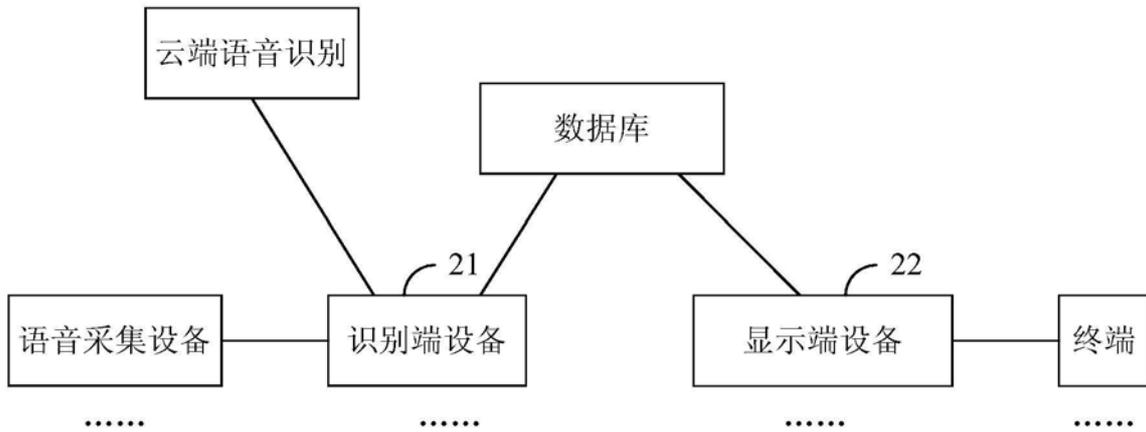


图2

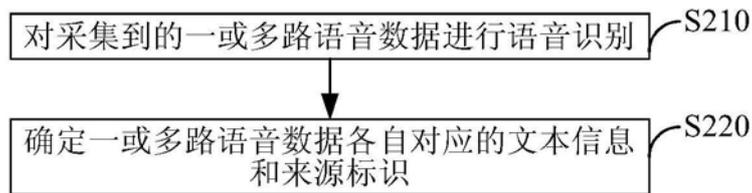


图3

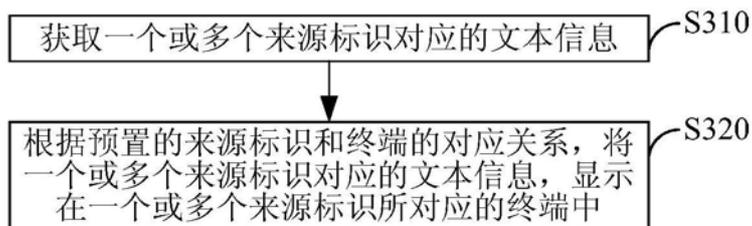


图4

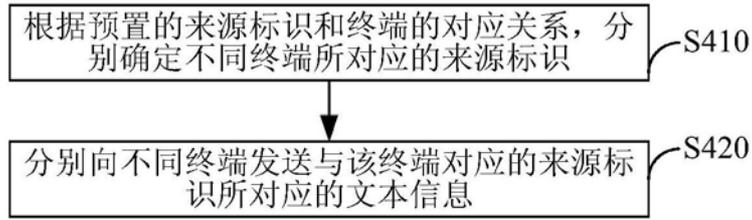


图5

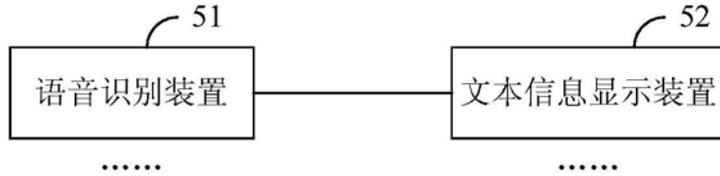


图6



图7

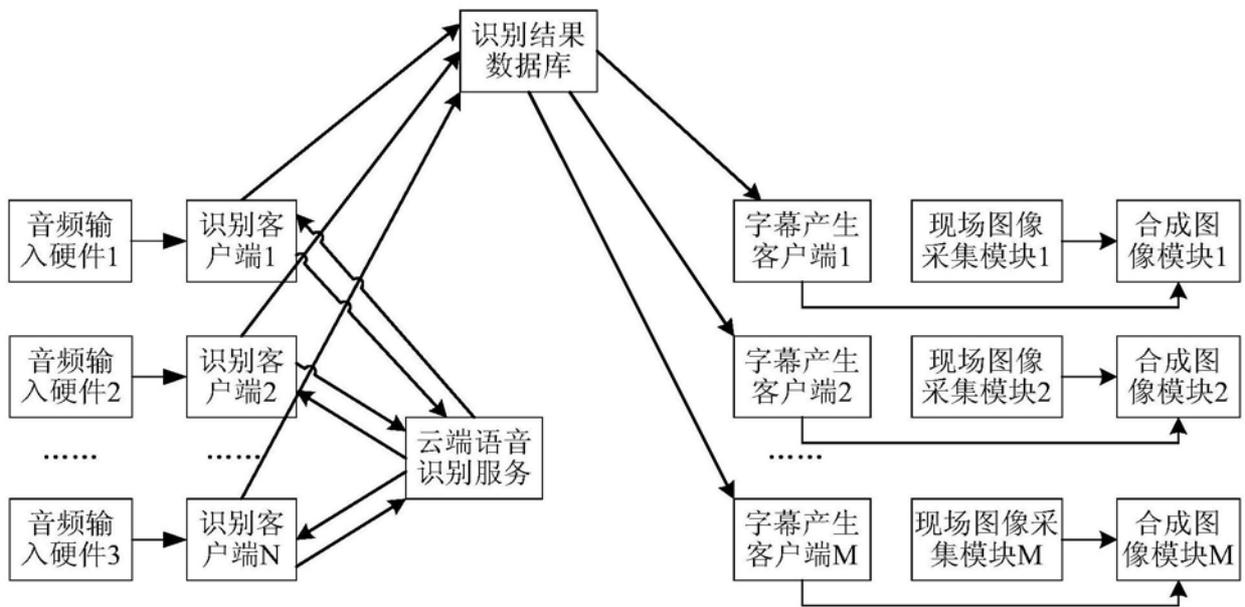


图8

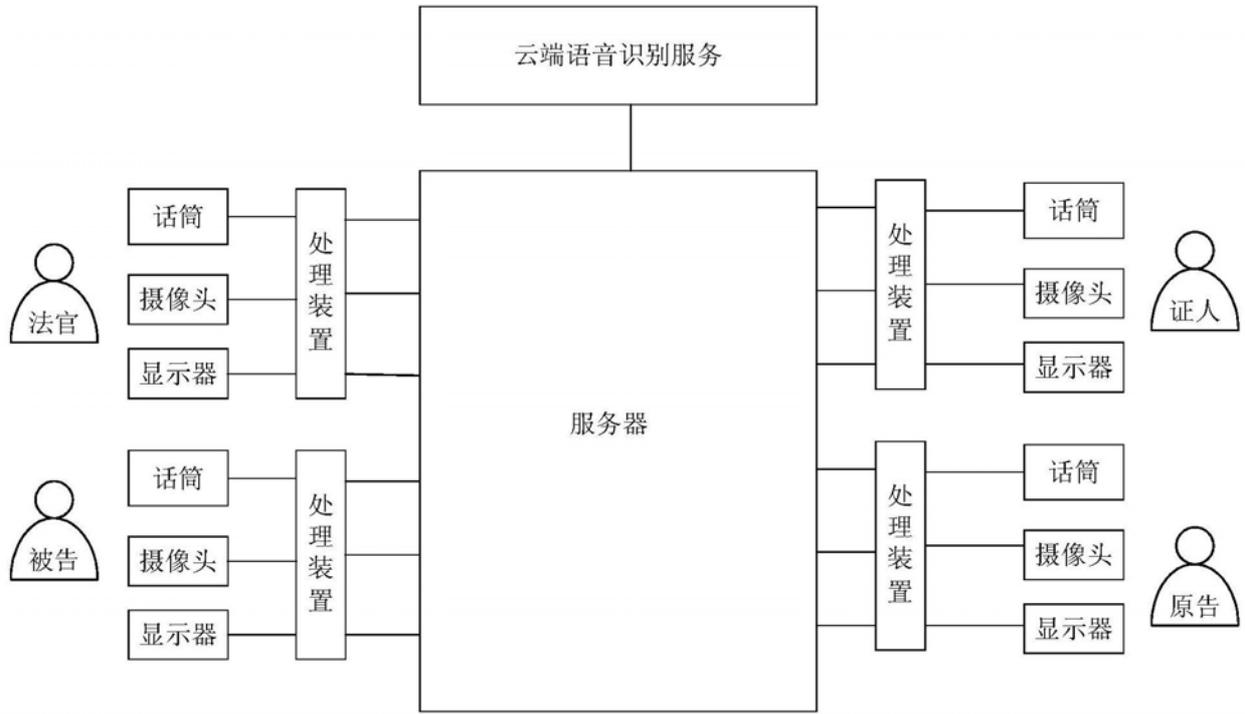


图9

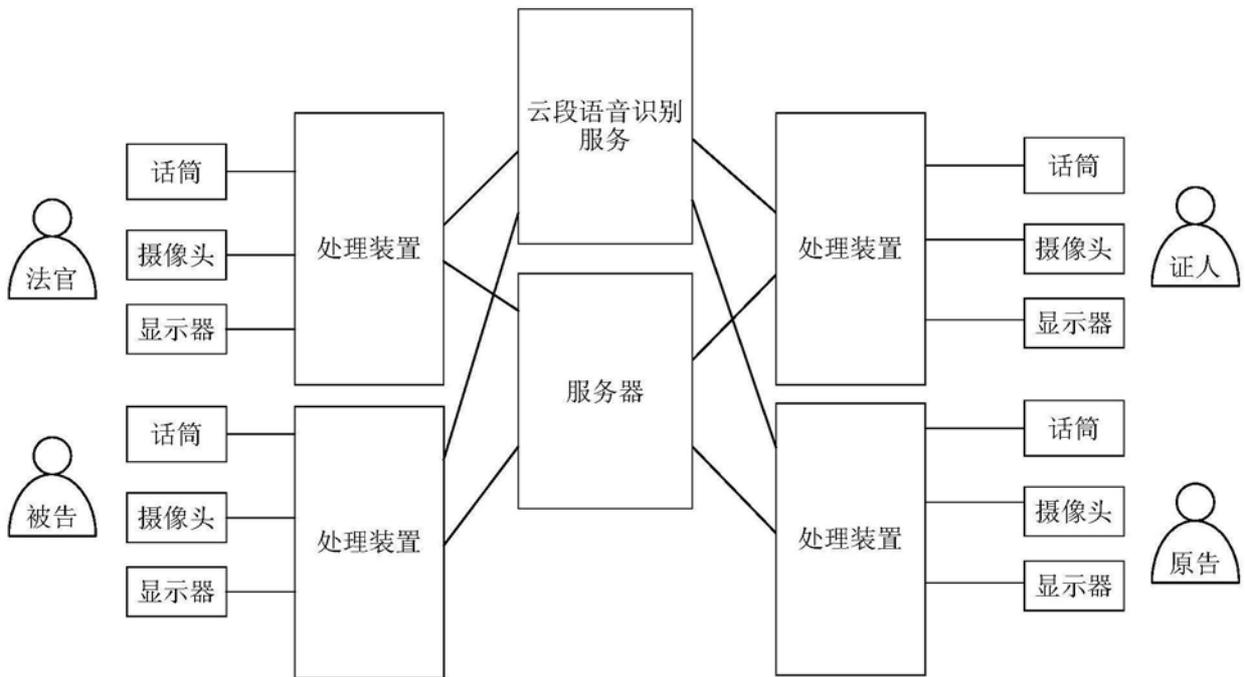


图10