(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114596859 A
(43)申请公布日 2022.06.07

(21)申请号 202210249744.1

(22)申请日 2022.03.14

(71)申请人 云知声智能科技股份有限公司
地址 100096 北京市海淀区西三旗建材城内1幢一层101号

(72)发明人 李伟敬 孙立新 王明明 李鹏 梁家恩

(51)Int.Cl.
G10L 15/26(2006.01)
G10L 15/04(2013.01)
G10L 21/0208(2013.01)

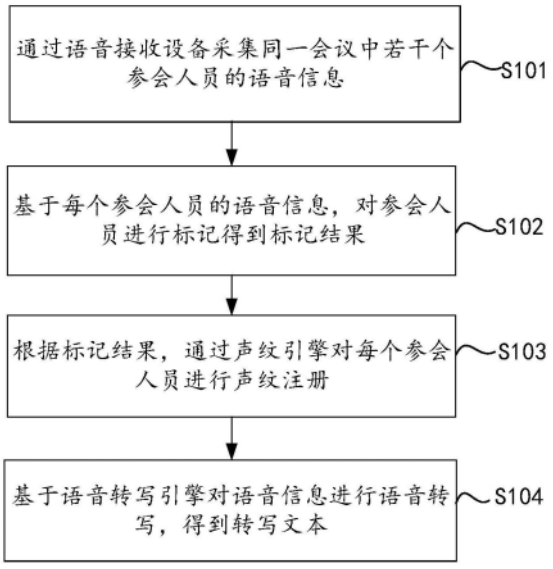
权利要求书2页 说明书6页 附图2页

(54)发明名称

会议语音转写方法、系统、设备及存储介质

(57)摘要

本申请公开了一种会议语音转写方法、系统、设备及存储介质，该方法通过语音接收设备采集同一会议中若干个参会人员的语音信息；基于每个所述参会人员的语音信息，对所述参会人员进行标记得到标记结果；根据所述标记结果，通过声纹引擎对每个所述参会人员进行声纹注册；基于语音转写引擎对所述语音信息进行语音转写，得到转写文本。本申请无需用户提前注册，采用用户无感的注册方案，来更好的提升角色标记性能，提高标记准确率。



CN 114596859 A

1.一种会议语音转写方法,其特征在于,所述方法包括以下步骤:

通过语音接收设备采集同一会议中若干个参会人员的语音信息;

基于每个所述参会人员的语音信息,对所述参会人员进行标记得到标记结果;

根据所述标记结果,通过声纹引擎对每个所述参会人员进行声纹注册;

基于语音转写引擎对所述语音信息进行语音转写,得到转写文本。

2.根据权利要求1所述的方法,其特征在于,在通过语音接收设备采集同一会议中若干个参会人员的语音信息之后,还包括:

将所述语音信息进行降噪处理。

3.根据权利要求1所述的方法,其特征在于,所述方法还包括:

确定所述转写文本中的长句文本;

对所述长句文本进行断句,得到若干个文本片段;

对所述若干个语音片段进行目标标记,得到目标会议文本并输出。

4.根据权利要求3所述的方法,其特征在于,对所述长句文本进行断句,得到若干个文本片段,包括:

采用智能断句,按照文本语义对所述长句文本进行断句,得到若干个语音片段;

根据所述若干个语音片段的片段起始位置和片段终止位置,对所述若干个语音片段添加标点符号。

5.根据权利要求1所述的方法,其特征在于,对所述参会人员进行标记得到标记结果,包括:

对所述参会人员的姓名进行标记得到姓名标记结果;或者,

对所述参会人员的角色进行标记得到角色标记结果。

6.一种会议语音转写系统,其特征在于,所述系统包括:

信息采集单元,用于通过语音接收设备采集同一会议中若干个参会人员的语音信息;

标记单元,用于基于每个所述参会人员的语音信息,对所述参会人员进行标记得到标记结果;

声纹注册单元,用于根据所述标记结果,通过声纹引擎对每个所述参会人员进行声纹注册;以及

语音转写单元,用于基于语音转写引擎对所述语音信息进行语音转写,得到转写文本。

7.根据权利要求6所述的系统,其特征在于,所述系统还包括:

降噪单元,用于将所述语音信息进行降噪处理。

8.根据权利要求6所述的系统,其特征在于,所述系统还包括:

长句确定单元,用于确定所述转写文本中的长句文本;

断句单元,用于对所述长句文本进行断句,得到若干个文本片段;

语音片段标记单元,用于对所述若干个语音片段进行目标标记,得到目标会议文本并输出。

9.一种电子设备,其特征在于,所述电子设备包括:至少一个处理器、存储器、至少一个网络接口和用户接口;

所述至少一个处理器、存储器、至少一个网络接口和用户接口通过总线系统耦合在一起;

所述处理器通过调用所述存储器存储的程序或指令，用于执行如权利要求1至5中任意一项所述的会议语音转写方法的步骤。

10.一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上存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求1至5中任意一项所述的会议语音转写方法的步骤。

会议语音转写方法、系统、设备及存储介质

技术领域

[0001] 本申请涉及人工智能技术领域，尤其涉及一种会议语音转写方法、系统、设备及存储介质。

背景技术

[0002] 随着AI技术发展，语音转写功能在智能产品中的应用也越来越普遍，当前的会议转写系统，主要处理流程如下：语音拾音，即通过麦克风拾取发音人数据；语音识别，将麦克风采集到的数据，发送到语音识别模型，进行语音转文字处理；角色标记，根据识别结果及语音进行角色信息标记。

[0003] 在实现本公开实施例的过程中，发现相关技术中至少存在如下问题：相关技术中的角色标记准确率较低，经常出现角色标记错误，尤其是相同性别角色的说话场景，同性别角色不容易分辨，标记信息不准确，文本不易理解，导致标记结果参考意义不大。有些相关技术方案通过声纹注册的方式，提前采集讲话者的声纹信息，以提高角色标记准确率，但该方案需要用户提前注册，而且即使同一个人在不同时间不同地点的声纹特征也有所区别，因此该方案也依然存在角色标记准确率低的问题。

[0004] 针对上述的问题，目前尚未提出有效的解决方案。

发明内容

[0005] 为了解决上述技术问题，本申请提供了一种会议语音转写方法、系统、设备及存储介质。

[0006] 本申请提供了一种会议语音转写方法，所述方法包括以下步骤：

[0007] 通过语音接收设备采集同一会议中若干个参会人员的语音信息；

[0008] 基于每个所述参会人员的语音信息，对所述参会人员进行标记得到标记结果；

[0009] 根据所述标记结果，通过声纹引擎对每个所述参会人员进行声纹注册；

[0010] 基于语音转写引擎对所述语音信息进行语音转写，得到转写文本。

[0011] 在一些实施例中，在通过语音接收设备采集同一会议中若干个参会人员的语音信息之后，还包括：

[0012] 将所述语音信息进行降噪处理。

[0013] 在一些实施例中，所述方法还包括：

[0014] 确定所述转写文本中的长句文本；

[0015] 对所述长句文本进行断句，得到若干个文本片段；

[0016] 对所述若干个语音片段进行目标标记，得到目标会议文本并输出。

[0017] 在一些实施例中，对所述长句文本进行断句，得到若干个文本片段，包括：

[0018] 采用智能断句，按照文本语义对所述长句文本进行断句，得到若干个语音片段；

[0019] 根据所述若干个语音片段的片段起始位置和片段终止位置，对所述若干个语音片段添加标点符号。

[0020] 在一些实施例中，对所述参会人员进行标记得到标记结果，包括：

[0021] 对所述参会人员的姓名进行标记得到姓名标记结果；或者，

[0022] 对所述参会人员的角色进行标记得到角色标记结果。

[0023] 本申请还提供了一种会议语音转写系统，所述系统包括：

[0024] 信息采集单元，用于通过语音接收设备采集同一会议中若干个参会人员的语音信息；

[0025] 标记单元，用于基于每个所述参会人员的语音信息，对所述参会人员进行标记得到标记结果；

[0026] 声纹注册单元，用于根据所述标记结果，通过声纹引擎对每个所述参会人员进行声纹注册；以及

[0027] 语音转写单元，用于基于语音转写引擎对所述语音信息进行语音转写，得到转写文本。

[0028] 在一些实施例中，所述系统还包括：

[0029] 降噪单元，用于将所述语音信息进行降噪处理。

[0030] 在一些实施例中，所述系统还包括：

[0031] 长句确定单元，用于确定所述转写文本中的长句文本；

[0032] 断句单元，用于对所述长句文本进行断句，得到若干个文本片段；

[0033] 语音片段标记单元，用于对所述若干个语音片段进行目标标记，得到目标会议文本并输出。

[0034] 本申请还提供了一种电子设备，该电子设备包括：至少一个处理器、存储器、至少一个网络接口和用户接口；

[0035] 所述至少一个处理器、存储器、至少一个网络接口和用户接口通过总线系统耦合在一起；

[0036] 所述处理器通过调用所述存储器存储的程序或指令，用于执行如上述的会议语音转写方法的步骤。

[0037] 本申请还提供了一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现上述的会议语音转写方法的步骤。

[0038] 本申请实施例提供的上述技术方案与现有技术相比具有如下优点：

[0039] 本申请实施例提供的一种会议语音转写方法、系统、设备及存储介质，该方法通过语音接收设备采集同一会议中若干个参会人员的语音信息；基于每个参会人员的语音信息，对参会人员进行标记得到标记结果；根据标记结果，通过声纹引擎对每个参会人员进行声纹注册；基于语音转写引擎对语音信息进行语音转写，得到转写文本。本申请无需用户提前注册，采用用户无感的注册方案，来更好的提升角色标记性能，提高标记准确率。

附图说明

[0040] 此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本发明的实施例，并与说明书一起用于解释本发明的原理。

[0041] 为了更清楚地说明本发明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简单地介绍，显而易见地，对于本领域普通技术人员而

言，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。

[0042] 图1为本申请实施例提供的一种会议语音转写方法的流程示意图；

[0043] 图2为本申请实施例提供的一种会议语音转写系统的结构示意图；

[0044] 图3为本申请实施例提供的一种电子设备的结构示意图。

具体实施方式

[0045] 为使本申请实施例的目的、技术方案和优点更加清楚，下面将结合本申请实施例中的附图，对本申请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申请的一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0046] 图1为本申请实施例提供的一种会议语音转写方法的流程示意图，该方法具体包括以下步骤：

[0047] S101、通过语音接收设备采集同一会议中若干个参会人员的语音信息。

[0048] 本申请通过采集同一会议中的参会人员的语音信息，无外界环境的影响，避免相同发音人在不同时刻声纹信息不同而导致提前注册的声纹信息与角色不对应而导致的标记准确率降低的问题。避免了现有的不同时间点注册的声纹，受参会人员的身体健康状况、情绪、地点、噪声环境、背景噪声等对声纹的干扰。本申请采用同一语音接收设备进行语音信息的采集，避免了不同设备采集信息对声纹的影响。

[0049] 可选地，语音接收设备可以是例如麦克风、录音笔等设备。

[0050] S102、基于每个参会人员的语音信息，对参会人员进行标记得到标记结果。

[0051] 例如，相应的操作人员可以选取本次会议中每个发言的参会人员的3～5句话，进行用户标记，可以标记参会人员的姓名或角色。

[0052] S103、根据标记结果，通过声纹引擎对每个参会人员进行声纹注册。

[0053] 将标记结果发送到声纹引擎，通过声纹引擎进行声纹注册，在此过程中，对于用户(参会人员)来说是在不知不觉中完成的，无需用户过多的配合，用户体验感较高，相对来说声纹注册成功率也较高。

[0054] S104、基于语音转写引擎对语音信息进行语音转写，得到转写文本。

[0055] 本申请无需用户提前注册，采用用户无感的注册方案，来更好的提升角色标记性能，提高标记准确率。

[0056] 在一些实施例中，在通过语音接收设备采集同一会议中若干个参会人员的语音信息之后，还包括：

[0057] 将语音信息进行降噪处理。

[0058] 在一些实施例中，该方法还包括：

[0059] 确定转写文本中的长句文本；

[0060] 对长句文本进行断句，得到若干个文本片段；

[0061] 对若干个语音片段进行目标标记，得到目标会议文本并输出。

[0062] 本申请通过将长句文本进行断句，以及进行用户标记，可以进一步提高标记准确率。

[0063] 在一些实施例中，对长句文本进行断句，得到若干个文本片段，包括：

[0064] 采用智能断句，按照文本语义对长句文本进行断句，得到若干个语音片段；

[0065] 根据若干个语音片段的片段起始位置和片段终止位置，对若干个语音片段添加标点符号。

[0066] 可选地，本申请还可以对转写文本进行校对，例如校对转写的文本信息是否正确，或者校对标记的角色信息是否正确，并编辑修改错误的内容，最终将正确的会议文本输出。

[0067] 在一些实施例中，对参会人员进行标记得到标记结果，包括：

[0068] 对参会人员的姓名进行标记得到姓名标记结果；或者，

[0069] 对参会人员的角色进行标记得到角色标记结果。

[0070] 本申请通过人工介入的方式，通过操作员选取会议过程中参会人员的一些语音信息，并做声纹标记，进行声纹注册，在此操作过程中对使用者无感，避免了使用者不配合声纹注册的事情；此外，声纹注册的语音信息选取自同一会议中的数据，无外界环境和人的影响，避免相同发音人不同时刻声纹信息不同而导致提前注册角色分离准确率低的问题。通过以上两种方式，可以很好的提高角色标记的准确率，同时标记参会人员角色或者姓名，使会议文本更方便理解。

[0071] 参见图2，本申请还提供了一种会议语音转写系统，该系统包括：

[0072] 信息采集单元21，用于通过语音接收设备采集同一会议中若干个参会人员的语音信息；

[0073] 标记单元22，用于基于每个参会人员的语音信息，对参会人员进行标记得到标记结果；

[0074] 声纹注册单元23，用于根据标记结果，通过声纹引擎对每个参会人员进行声纹注册；以及

[0075] 语音转写单元24，用于基于语音转写引擎对语音信息进行语音转写，得到转写文本。

[0076] 在一些实施例中，该系统还包括：

[0077] 降噪单元，用于将语音信息进行降噪处理。

[0078] 在一些实施例中，该系统还包括：

[0079] 长句确定单元，用于确定转写文本中的长句文本；

[0080] 断句单元，用于对长句文本进行断句，得到若干个文本片段；

[0081] 语音片段标记单元，用于对若干个语音片段进行目标标记，得到目标会议文本并输出。

[0082] 本申请实施例还提供了一种计算机可读存储介质，计算机可读存储介质上存储有计算机程序，计算机程序被处理器执行时实现如各方法实施例的步骤，例如包括：

[0083] 通过语音接收设备采集同一会议中若干个参会人员的语音信息；

[0084] 基于每个参会人员的语音信息，对参会人员进行标记得到标记结果；

[0085] 根据标记结果，通过声纹引擎对每个参会人员进行声纹注册；

[0086] 基于语音转写引擎对语音信息进行语音转写，得到转写文本。

[0087] 图3是本发明另一个实施例提供的电子设备的结构示意图。图3所示的电子设备300包括：至少一个处理器301、存储器302、至少一个网络接口304和其他用户接口303。电子设备300中的各个组件通过总线系统305耦合在一起。可理解，总线系统305用于实现这些组

件之间的连接通信。总线系统305除包括数据总线之外，还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见，在图3中将各种总线都标为总线系统305。

[0088] 其中，用户接口303可以包括显示器、键盘或者点击设备(例如，鼠标，轨迹球(trackball)、触感板或者触摸屏等。

[0089] 可以理解，本发明实施例中的存储器302可以是易失性存储器或非易失性存储器，或可包括易失性和非易失性存储器两者。其中，非易失性存储器可以是只读存储器(Read-OnlyMemory，ROM)、可编程只读存储器(ProgrammableROM，PROM)、可擦除可编程只读存储器(ErasablePROM，EPROM)、电可擦除可编程只读存储器(ElectricallyEPROM，EEPROM)或闪存。易失性存储器可以是随机存取存储器(RandomAccessMemory，RAM)，其用作外部高速缓存。通过示例性但不是限制性说明，许多形式的RAM可用，例如静态随机存取存储器(StaticRAM，SRAM)、动态随机存取存储器(DynamicRAM，DRAM)、同步动态随机存取存储器(SynchronousDRAM，SDRAM)、双倍数据速率同步动态随机存取存储器(DoubleDataRate SDRAM，DDRSDRAM)、增强型同步动态随机存取存储器(Enhanced SDRAM，ESDRAM)、同步连接动态随机存取存储器(SynchlinkDRAM，SLDRAM)和直接内存总线随机存取存储器(DirectRambusRAM，DRRAM)。本文描述的存储器302旨在包括但不限于这些和任意其它适合类型的存储器。

[0090] 在一些实施方式中，存储器302存储了如下的元素，可执行单元或者数据结构，或者他们的子集，或者他们的扩展集：操作系统3021和应用程序3022。

[0091] 其中，操作系统3021，包含各种系统程序，例如框架层、核心库层、驱动层等，用于实现各种基础业务以及处理基于硬件的任务。应用程序3022，包含各种应用程序，例如媒体播放器(MediaPlayer)、浏览器(Browser)等，用于实现各种应用业务。实现本发明实施例方法的程序可以包含在应用程序3022中。

[0092] 在本发明实施例中，通过调用存储器302存储的程序或指令，具体的，可以是应用程序3022中存储的程序或指令，处理器301用于执行各方法实施例所提供的方法步骤，例如包括：

[0093] 通过语音接收设备采集同一会议中若干个参会人员的语音信息；

[0094] 基于每个参会人员的语音信息，对参会人员进行标记得到标记结果；

[0095] 根据标记结果，通过声纹引擎对每个参会人员进行声纹注册；

[0096] 基于语音转写引擎对语音信息进行语音转写，得到转写文本。。

[0097] 上述本发明实施例揭示的方法可以应用于处理器301中，或者由处理器301实现。处理器301可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器301中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器301可以是通用处理器、数字信号处理器(DigitalSignalProcessor，DSP)、专用集成电路(ApplicationSpecific IntegratedCircuit，ASIC)、现成可编程门阵列(FieldProgrammableGateArray，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。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件单元组合执行完成。软件单元可以位于随机存储器，闪存、只读存储器，可

编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器302，处理器301读取存储器302中的信息，结合其硬件完成上述方法的步骤。

[0098] 可以理解的是，本文描述的这些实施例可以用硬件、软件、固件、中间件、微码或其组合来实现。对于硬件实现，处理单元可以实现在一个或多个专用集成电路(Application SpecificIntegratedCircuits,ASIC)、数字信号处理器(DigitalSignalProcessing,DSP)、数字信号处理设备(DSPDevice,DSPD)、可编程逻辑设备(ProgrammableLogicDevice,PLD)、现场可编程门阵列(Field-ProgrammableGateArray,FPGA)、通用处理器、控制器、微控制器、微处理器、用于执行本申请功能的其它电子单元或其组合中。

[0099] 对于软件实现，可通过执行本文功能的单元来实现本文所述的技术。软件代码可存储在存储器中并通过处理器执行。存储器可以在处理器中或在处理器外部实现。

[0100] 为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，在实施本发明时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0101] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置或系统实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。以上所描述的装置及系统实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

[0102] 需要说明的是，在本文中，诸如“第一”和“第二”等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，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、物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、物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，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0103] 以上所述仅是本发明的具体实施方式，使本领域技术人员能够理解或实现本发明。对这些实施例的多种修改对本领域的技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下，在其它实施例中实现。因此，本发明将不会被限制于本文所示的这些实施例，而是要符合与本文所申请的原理和新颖特点相一致的最宽的范围。

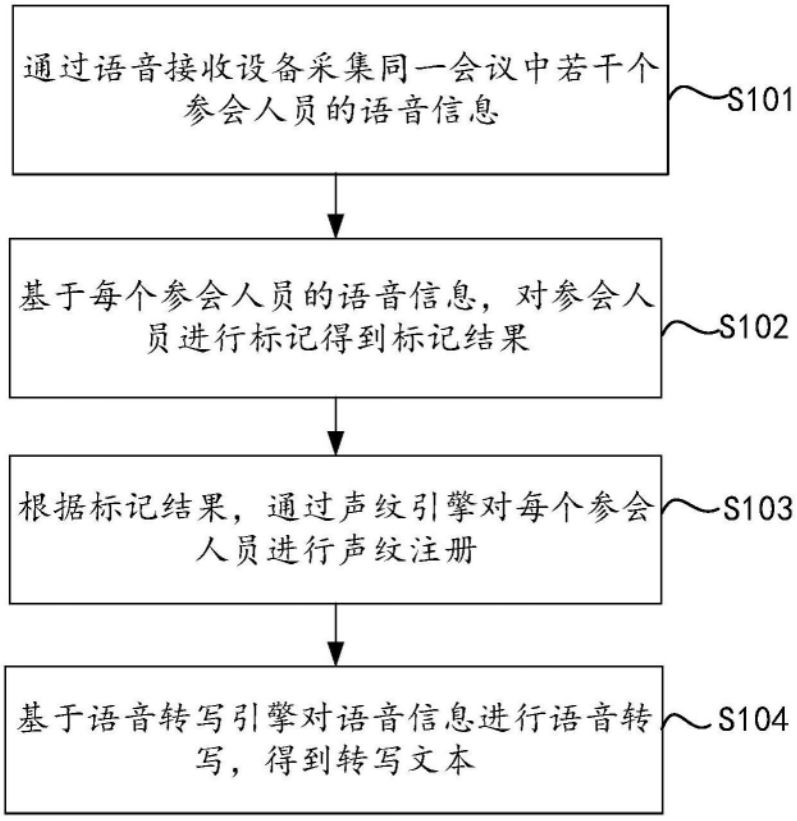

通过语音接收设备采集同一会议中若干个参会人员的语音信息
S101
基于每个参会人员的语音信息，对参会人员进行标记得到标记结果
S102
根据标记结果，通过声纹引擎对每个参会人员进行声纹注册
S103
基于语音转写引擎对语音信息进行语音转写，得到转写文本
S104

图1

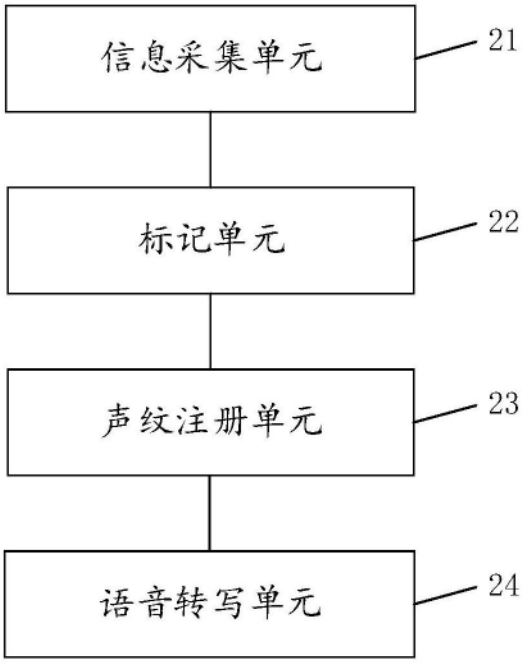

信息采集单元
21
标记单元
22
声纹注册单元
23
语音转写单元
24

图2

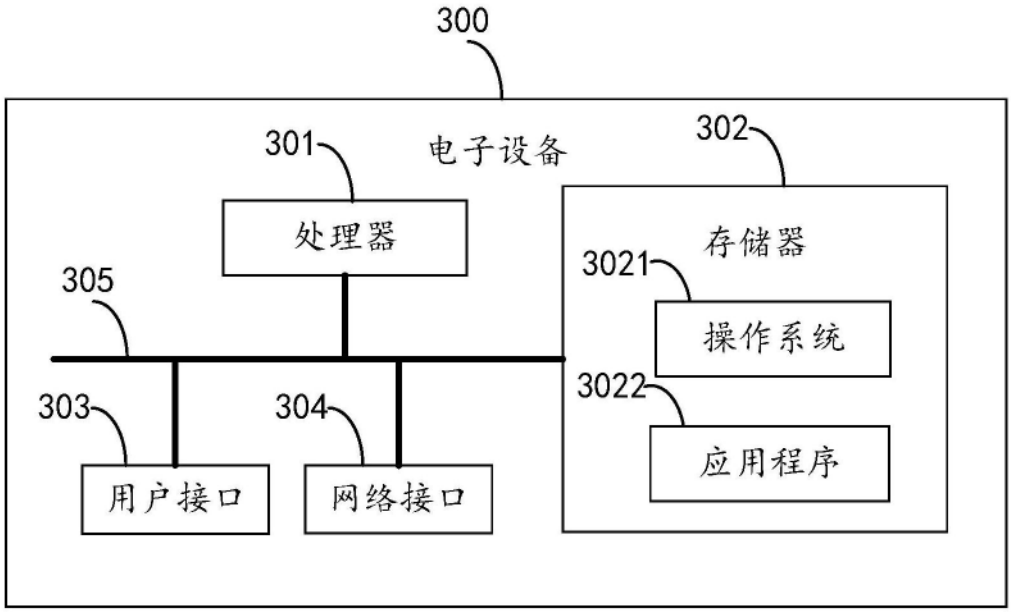

300
电子设备
301
处理器
302
存储器
3021
操作系统
3022
应用程序
305
303
用户接口
304
网络接口

图3