

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2023 年 3 月 16 日 (16.03.2023)



(10) 国际公布号
WO 2023/035397 A1

- (51) 国际专利分类号:
G10L 17/02 (2013.01) *G10L 25/30* (2013.01)
G10L 17/04 (2013.01) *G10L 25/51* (2013.01)
G10L 25/24 (2013.01)
- (21) 国际申请号: PCT/CN2021/129733
- (22) 国际申请日: 2021 年 11 月 10 日 (10.11.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202111042821.8 2021 年 9 月 7 日 (07.09.2021) CN
- (71) 申请人: 科大讯飞股份有限公司 (IFLYTEK CO., LTD.) [CN/CN]; 中国安徽省合肥市高新区望江西路 666 号, Anhui 230088 (CN)。
- (72) 发明人: 方昕 (FANG, Xin); 中国安徽省合肥市高新区望江西路 666 号, Anhui 230088 (CN)。 刘俊华 (LIU, Junhua); 中国安徽省合肥市高新区望江西路 666 号, Anhui 230088 (CN)。
- (74) 代理人: 北京集佳知识产权代理有限公司 (UNITALEN ATTORNEYS AT LAW); 中国北京市朝阳区建国门外大街 22 号赛特广场 7 层, Beijing 100004 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,

(54) Title: SPEECH RECOGNITION METHOD, APPARATUS AND DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种语音识别方法、装置、设备及存储介质

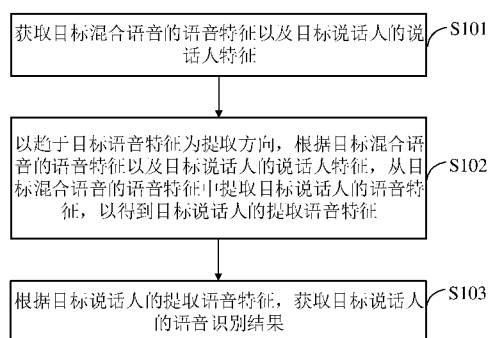


图 1

- S101 Acquire a speech feature of target mixed speech and a speaker feature of a target speaker
- S102 Take the direction of tending to a target speech feature as an extraction direction, and according to the speech feature of the target mixed speech and the speaker feature of the target speaker, extract a speech feature of the target speaker from the speech feature of the target mixed speech, so as to obtain an extracted speech feature of the target speaker
- S103 Acquire a speech recognition result of the target speaker according to the extracted speech feature of the target speaker

(57) Abstract: Provided in the present application are a speech recognition method, apparatus and device, and a storage medium. The method comprises: acquiring a speech feature of target mixed speech and a speaker feature of a specified speaker; taking the direction of tending to a target speech feature as an extraction direction, and according to the speech feature of the target mixed speech and a speaker feature of a target speaker, extracting a speech feature of the target speaker from the speech feature of the target mixed speech, so as to obtain an extracted speech feature of the target speaker, wherein the target speech feature is a speech feature that is used to obtain a speech recognition result consistent with real speech content of the target speaker; and acquiring a speech recognition result of the specified speaker according to an extracted speech feature of the specified speaker. By means of the speech recognition method provided in the present application, speech content of a specified speaker can be recognized more accurately from mixed speech including speech of the specified speaker, thereby achieving a better user experience.

IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：本申请提供了一种语音识别方法、装置、设备及存储介质，其中，方法包括：获取目标混合语音的语音特征以及指定说话人的说话人特征；以趋于目标语音特征为提取方向，根据目标混合语音的语音特征以及目标说话人的说话人特征，从目标混合语音的语音特征中提取目标说话人的语音特征，以得到目标说话人的提取语音特征，其中，目标语音特征为用于获得与目标说话人的真实语音内容一致的语音识别结果的语音特征；根据指定说话人的提取语音特征，获取指定说话人的语音识别结果。经由本申请提供的语音识别方法可从包含指定说话人语音的混合语音中较为准确的识别出指定说话人的语音内容，用户体验较好。

一种语音识别方法、装置、设备及存储介质

本申请要求于 2021 年 09 月 07 日提交中国专利局、申请号为 CN202111042821.8、发明名称为“一种语音识别方法、装置、设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及语音识别技术领域，尤其涉及一种语音识别方法、装置、设备及存储介质。

背景技术

随着人工智能技术的飞速发展，智能设备在人们的生活中扮演着越来越重要的角色，语音交互作为最方便自然的人机交互方式深受用户喜爱。

在用户使用智能设备时，其可能处在一个存在其他人声的复杂环境中，在这种情况下，智能设备采集的语音为混合语音。在进行语音交互时，为了能够获得较好的用户体验，就需要从混合语音中识别出目标说话人的语音内容，而如何从混合语音中识别出目标说话人的语音内容是目前亟需解决的问题。

发明内容

有鉴于此，本申请提供了一种语音识别方法、装置、设备及存储介质，用以从混合语音中较为准确地识别出目标说话人的语音内容，其技术方案如下：

一种语音识别方法，包括：

获取目标混合语音的语音特征以及目标说话人的说话人特征；

以趋于目标语音特征为提取方向，根据所述目标混合语音的语音特征以及所述目标说话人的说话人特征，从所述目标混合语音的语音特征中提取所述目标说话人的语音特征，以得到所述目标说话人的提取语音特征，其中，所述目标语音特征为用于获得与所述目标说话人的真实语音内容一致的语音识别结果的语音特征；

根据所述目标说话人的提取语音特征，获取所述目标说话人的语音识别结果。

可选的，获取所述目标说话人的说话人特征，包括：

获取所述目标说话人的注册语音；

对所述目标说话人的注册语音提取短时声纹特征和长时声纹特征，以得到

多尺度声纹特征，作为所述目标说话人的说话人特征。

可选的，所述以趋于目标语音特征为提取方向，根据所述目标混合语音的语音特征以及所述目标说话人的说话人特征，从所述目标混合语音的语音特征中提取所述目标说话人的语音特征，包括：

- 5 利用预先建立的特征提取模型，以所述目标混合语音的语音特征以及所述目标说话人的说话人特征为依据，从所述目标混合语音的语音特征中提取所述目标说话人的语音特征；

10 其中，所述特征提取模型采用包含指定说话人的语音的训练混合语音的语音特征和所述指定说话人的说话人特征，以基于所述指定说话人的提取语音特征获取的语音识别结果为优化目标训练得到，所述指定说话人的提取语音特征为从所述训练混合语音的语音特征中提取的所述指定说话人的语音特征。

可选的，所述特征提取模型同时以所述指定说话人的提取语音特征和基于所述指定说话人的提取语音特征获取的语音识别结果为优化目标训练得到。

- 15 可选的，所述利用预先建立的特征提取模型，以所述目标混合语音的语音特征以及所述目标说话人的说话人特征为依据，从所述目标混合语音的语音特征中提取所述目标说话人的语音特征，包括：

将所述目标混合语音的语音特征以及所述目标说话人的说话人特征输入所述特征提取模型，得到所述目标说话人对应的特征掩码；

- 20 根据所述目标混合语音的语音特征和所述目标说话人对应的特征掩码，从所述目标混合语音的语音特征中提取所述目标说话人的语音特征。

可选的，所述根据所述目标说话人的提取语音特征，获取所述目标说话人的语音识别结果，包括：

根据所述目标说话人的提取语音特征以及所述目标说话人的注册语音特征，获取所述目标说话人的语音识别结果；

- 25 其中，所述目标说话人的注册语音特征为所述目标说话人的注册语音的语音特征。

可选的，所述根据所述目标说话人的提取语音特征，获取所述目标说话人的语音识别结果，包括：

- 30 将至少包括所述目标说话人的提取语音特征的语音识别输入特征输入预先建立的语音识别模型，得到所述目标说话人的语音识别结果；

所述语音识别模型与所述特征提取模型联合训练得到,所述语音识别模型采用所述指定说话人的提取语音特征,以基于所述指定说话人的提取语音特征获得的语音识别结果为优化目标训练得到。

可选的,将所述语音识别输入特征输入所述语音识别模型,得到所述目标
5 说话人的语音识别结果,包括:

基于所述语音识别模型的编码器模块,对所述语音识别输入特征进行编码,以得到编码结果;

基于所述语音识别模型的注意力模块,从所述编码结果中提取解码时刻解码所需的音频相关特征向量;

10 基于所述语音识别模型的解码器模块,对从所述编码结果中提取的所述音频相关特征向量进行解码,得到所述解码时刻的识别结果。

可选的,所述语音识别模型与所述特征提取模型联合训练的过程包括:

利用特征提取模型,从所述训练混合语音的语音特征中提取所述指定说话人的语音特征,以得到所述指定说话人的提取语音特征;

15 利用语音识别模型和所述指定说话人的提取语音特征,获取所述指定说话人的语音识别结果;

根据所述指定说话人的提取语音特征和所述指定说话人的语音识别结果对特征提取模型进行参数更新,并根据所述指定说话人的语音识别结果对语音识别模型进行参数更新。

20 可选的,所述训练混合语音对应有所述指定说话人的语音;

所述根据所述指定说话人的提取语音特征和所述指定说话人的语音识别结果对特征提取模型进行参数更新,并根据所述指定说话人的语音识别结果对语音识别模型进行参数更新,包括:

25 获取所述指定说话人的语音的标注文本,并获取所述指定说话人的语音的语音特征作为所述指定说话人的标准语音特征;

根据所述指定说话人的提取语音特征和所述指定说话人的标准语音特征确定第一预测损失,并根据所述指定说话人的语音识别结果和所述指定说话人的语音的标注文本,确定第二预测损失;

30 根据所述第一预测损失和所述第二预测损失对特征提取模型进行参数更新,并根据所述第二预测损失对语音识别模型进行参数更新。

可选的, 所述训练混合语音以及所述训练混合语音对应的所述指定说话人的语音从预先构建的训练数据集中获取;

所述训练数据集的构建过程包括:

获取多个说话人的多条语音, 其中, 每条语音为单一说话人的语音, 每条
5 语音具有标注文本;

将所述多条语音中的部分语音或全部语音中的每条语音作为指定说话人的语音: 将其它语音中其他说话人的一条或多条语音与该指定说话人的语音进行混合, 以得到一条训练混合语音, 将该指定说话人的语音与通过混合得到的训练混合语音作为一条训练数据;

10 由获得的所有训练数据组成所述训练数据集。

一种语音识别装置, 包括: 特征获取模块、特征提取模块和语音识别模块;

所述特征获取模块, 用于获取目标混合语音的语音特征以及目标说话人的说话人特征;

15 所述特征提取模块, 用于以提取趋于目标语音特征为提取方向, 根据所述目标混合语音的语音特征以及所述目标说话人的说话人特征, 从所述目标混合语音的语音特征中提取所述目标说话人的语音特征, 以得到所述目标说话人的提取语音特征, 其中, 所述目标语音特征为用于获得与所述目标说话人的真实语音内容一致的语音识别结果的语音特征;

20 所述语音识别模块, 用于根据所述目标说话人的提取语音特征, 获取所述目标说话人的语音识别结果。

可选的, 所述特征获取模块包括: 说话人特征获取模块;

所述说话人特征获取模块, 用于获取所述目标说话人的注册语音, 对所述目标说话人的注册语音提取短时声纹特征和长时声纹特征, 以得到多尺度声纹特征, 作为所述目标说话人的说话人特征。

25 可选的, 所述特征提取模块具体用于利用预先建立的特征提取模型, 以所述目标混合语音的语音特征以及所述目标说话人的说话人特征为依据, 从所述目标混合语音的语音特征中提取所述目标说话人的语音特征;

其中, 所述特征提取模型采用包含指定说话人的语音的训练混合语音的语音特征和指定说话人的说话人特征, 以基于所述指定说话人的提取语音特征获取的语音识别结果为优化目标训练得到, 所述指定说话人的提取语音特征为从
30

所述训练混合语音的语音特征中提取的所述指定说话人的语音特征。

可选的,所述语音识别模块,具体用于根据所述目标说话人的提取语音特征以及所述目标说话人的注册语音特征,获取所述目标说话人的语音识别结果;

5 其中,所述目标说话人的注册语音特征为所述目标说话人的注册语音的语音特征。

可选的,所述语音识别模块,具体用于将至少包括所述目标说话人的提取语音特征的语音识别输入特征输入预先建立的语音识别模型,得到所述目标说话人的语音识别结果;

10 其中,所述语音识别模型与所述特征提取模型联合训练得到,所述语音识别模型采用所述指定说话人的提取语音特征,以基于所述指定说话人的提取语音特征获得的语音识别结果为优化目标训练得到。

一种语音识别设备,包括:存储器和处理器;

所述存储器,用于存储程序;

15 所述处理器,用于执行所述程序,实现上述任一项所述的语音识别方法的各个步骤。

一种可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时,实现上述任一项所述的语音识别方法的各个步骤。

经由上述方案可知,本申请提供的语音识别方法、装置、设备及存储介质,能够根据目标混合语音的语音特征以及目标说话人的说话人特征,从目标混合语音的语音特征中提取出目标说话人的语音特征,进而能够根据提取出的目标说话人的语音特征,获得目标说话人的语音识别结果,由于本申请在从目标混合语音的语音特征提取目标说话人的语音特征时,以趋于目标语音特征(用于获得与目标目标说话人的真实语音内容一致的语音识别结果的语音特征)为提取方向,因此,提取出的语音特征为目标语音特征或者趋近于目标语音特征的语音特征,可见,经由上述方式提取出的语音特征为有利于语音识别的特征,基于提取出的语音特征进行语音识别,能够获得较好的语音识别效果,即能够获得较为准确的语音识别结果,用户体验较好。

20

25

附图说明

30 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,下面描述中的附图仅仅是本发明的实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据提供的附图获得其他的附图。

图1为本申请实施例提供的语音识别方法的流程示意图;

5 图2为本申请实施例提供的特征提取模型与语音识别模型联合训练的流程示意图;

图3为本申请实施例提供的特征提取模型与语音识别模型联合训练的过程示意图;

图4为本申请实施例提供的语音识别装置的结构示意图;

10 图5为本申请实施例提供的语音识别设备的结构示意图。

具体实施方式

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,都属于本发明保护的范围。

15 在外界环境中,人们通常被许多不同的声源围绕着,比如,多个人同时说话的声音、交通噪声、自然噪声等,经过研究者的不懈努力,以上描述的背景噪声的分离问题即通常意义上的语音增强问题已经得到了较好的解决,而相比之下,多个人同时说话的情况下,如何识别目标说话人的语音内容,即如何从混合语音中识别目标说话人的语音内容为难度更大的问题,其更具备研究意义。

为了能够从混合语音中识别出目标说话人的语音内容,申请人进行了研究,起初的思路是:先训练一个特征提取模型,再训练一个语音识别模型;获取目标说话人的注册语音,并对目标说话人的注册语音提取 d-vector 作为目标说话人的说话人特征;基于预先训练得到的特征提取模型,以目标说话人的说话人特征和目标混合语音的语音特征为依据,从目标混合语音的语音特征中提取出目标说话人的语音特征;通过对提取出的目标说话人的语音特征进行一系列的变换处理来获得目标说话人的语音;将目标说话人的语音输入预先训练得到的语音识别模型进行语音识别,从而获得目标说话人的语音识别结果。

25 30 申请人通过对上述思路进行研究发现,上述思路存在诸多缺陷,主要包括

如下几个方面：其一，对目标说话人的注册语音提取的 d-vector 所包含的声纹信息不足，影响后续特征提取的效果；其二，特征提取模型与语音识别模型是单独训练的，二者是完全割裂的，不能有效联合优化，级联两个独立训练得到的模型进行语音识别会存在级联误差，进而影响语音识别效果；其三，在前端

5 的特征提取部分提取的特征不佳时，后端的语音识别部分没有任何补救措施，可能导致语音识别效果较差。

申请人在上述思路以及上述思路所存在的缺陷的基础上，进一步进行研究，最终提出了一种能够完美克服上述缺陷的语音识别方法，该语音识别方法能够从混合语音中较为准确地识别出目标说话人的语音内容，该语音识别方法

10 可应用于具有数据处理能力的终端，终端可按本申请提供的语音识别方法从目标混合语音中识别出目标说话人的语音内容，该终端可以包括处理组件、存储器、输入/输出接口和电源组件，可选的，该终端还可以包括多媒体组件、音频组件、传感器组件和通信组件等。其中：

处理组件用于进行数据处理，其可以进行本案的语音合成处理，处理组件

15 可以包括一个或多个处理器，处理组件还可以包括一个或多个模块，便于与其它组件之间的交互。

存储器被配置为存储各种类型的数据，存储器可以有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）、

20 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、光盘等中的一种或多种的组合。

电源组件为终端的各种组件提供电力，电源组件可以包括电源管理系统、一个或多个电源等。

多媒体组件可以包括屏幕，优选的，屏幕可以为触摸显示屏，触摸显示屏

25 可接收来自用户的输入信号。多媒体组件还可以包括前置摄像头和/或后置摄像头。

音频组件被配置为输出和/或输入音频信号，如音频组件可以包括麦克风，麦克风被配置为接收外部音频信号，音频组件还可以包括扬声器，扬声器被配置为输出音频信号，终端合成的语音可通过扬声器输出。

30 输入/输出接口为处理组件与外围接口模块之间的接口，外围接口模块可

以为键盘、按钮等，其中，按钮可包括但不限于主页按钮、音量按钮、启动按钮、锁定按钮等。

传感器组件可以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，传感器组件可以检测终端的打开/关闭状态、用户与终端是否接
5 触、装置的方位、速度、温度等。传感器组件可以包括但不限于图像传感器、加速度传感器、陀螺仪传感器、压力传感器、温度传感器等中的一种或多种的组合。

通信组件被配置为便于终端和其它设备进行有线或无线通信。终端可接入基于通信标准的无线网络，如WiFi、2G、3G、4G、5G中的一种或多种的组合。

10 可选的，终端可被一个或多个应用专用集成电路（ASIC）、数字信号处理器（ASP）、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，用于执行本申请提供的同传翻译方法。

本申请提供的语音识别方法还可应用于服务器，服务器可按本申请提供的
15 语音识别方法从目标混合语音中识别出目标说话人的语音内容，在一种场景中，服务器可通过网络与终端连接，终端获取目标混合语音，将目标混合语音通过与服务器连接的网络传输至服务器，服务器按本申请提供的语音识别方法从目标混合语音中识别出目标说话人的语音内容，再通过网络将目标说话人的语音内容传输至终端。服务器可以包括一个或一个以上的中央处理器和存储
20 器，其中，存储器被配置为存储各种类型的数据，存储器可以有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、光盘等中的一种或多种的组合。服务器还可以包括一个或
25 一个以上电源、一个或一个以上有线网络接口和/或一个或一个以上无线网络接口、一个或一个以上操作系统。

接下来通过下述实施例对本申请提供的语音识别方法进行介绍。

第一实施例

请参阅图1，示出了本申请实施例提供的语音识别方法的流程示意图，该
30 方法可以包括：

步骤 S101: 获取目标混合语音的语音特征以及目标说话人的说话人特征。

其中, 目标混合语音为多个说话人的语音, 其除了包括目标说话人的语音外, 还包括其他说话人的语音, 本申请意在实现, 在存在其他说话人的语音的情况下, 较为准确地识别出目标说话人的语音内容。

5 其中, 获取目标混合语音的语音特征的过程包括: 获取目标混合语音中每个语音帧的特征向量 (比如频谱特征), 以得到特征向量序列, 将获得的特征向量序列作为目标混合语音的语音特征。假设目标混合语音包括 K 个语音帧, 第 k 个语音帧的特征向量表示为 x_k , 则目标混合语音的语音特征可表示为 $[x_1, x_2, \dots, x_k, \dots, x_K]$ 。

10 其中, 获取目标说话人的说话人特征实现方式有多种, 本实施例提供如下两种可选的实现方式: 在一种可能的实现方式中, 可获取目标说话人的注册语音, 对目标说话人的注册语音提取 d-vector, 提取的 d-vector 作为目标说话人的说话人特征; 考虑到 d-vector 包含的声纹信息较为单一, 不够丰富, 为了提升后续特征提取的效果, 本实施例提供另一种较为优选的实现方式, 即, 获取
15 目标说话人的注册语音, 对目标说话人的注册语音提取短时声纹特征和长时声纹特征, 以得到多尺度声纹特征, 将多尺度声纹特征作为目标说话人的说话人特征。

相比于经由上述第一种实现方式获得的说话人特征, 经由上述第二种实现方式获得的说话人特征含有更为丰富的声纹信息, 这使得后续利用经由上述
20 第二种实现方式获得的说话人特征进行特征提取, 能够获得更佳的特征提取效果。

接下来对上述的第二种实现方式中, “对指定说话人的注册语音提取短时声纹特征和长时声纹特征” 的具体实现过程进行介绍。

25 对目标说话人的注册语音提取短时声纹特征和长时声纹特征的过程可以包括: 利用预先建立的说话人表征提取模型, 从目标说话人的注册语音中提取短时声纹特征和长时声纹特征。具体的, 获取目标说话人的注册语音的语音特征序列, 将目标说话人的注册语音的语音特征序列输入预先建立的说话人表征提取模型, 获得目标说话人的短时声纹特征和长时声纹特征。

30 可选的, 说话人表征提取模型可以采用卷积神经网络, 将目标说话人的注册语音的语音特征序列输入卷积神经网络进行特征提取, 以获得浅层特征和深

层特征，其中，浅层特征因感受野较小更能表征短时声纹，因此，将浅层特征作为短时声纹特征，而深层特征因感受野较大更能表征长时声纹，因此，将深层特征作为长时声纹特征。

本实施例中的说话人表征提取模型采用大量带真实说话人标签的训练语音（此处的训练语音优选为单一说话人的语音）训练得到，其中，训练语音的真实说话人标签代表的是训练语音对应的说话人。可选的，可采用交叉熵（Cross Entropy, CE）准则或者度量学习（Metric Learning, ML）准则训练说话人表征提取模型。

步骤 S102：以趋于目标语音特征为提取方向，根据目标混合语音的语音特征以及目标说话人的说话人特征，从目标混合语音的语音特征中提取目标说话人的语音特征，以得到目标说话人的提取语音特征。

其中，目标语音特征为用于获得与目标说话人的真实语音内容一致的语音识别结果的语音特征。

以趋于目标语音特征为提取方向，能够从目标混合语音的语音特征中提取出目标语音特征或者趋近于目标语音特征的语音特征，即，以趋于目标语音特征为提取方向能够从目标混合语音的语音特征中提取出有利于后续语音识别的语音特征，根据有利于语音识别的语音特征进行语音识别，能够获得较好的语音识别效果。

可选的，以趋于目标语音特征为提取方向，根据目标混合语音的语音特征以及目标说话人的说话人特征，从目标混合语音的语音特征中提取目标说话人的语音特征，以得到目标说话人的提取语音特征的过程可以包括：利用预先建立的特征提取模型，以目标混合语音特征和目标说话人特征为依据，从目标混合语音特征中提取指定说话人的语音特征，以得到目标说话人的提取语音特征。

其中，特征提取模型采用包含指定说话人的语音的训练混合语音的语音特征和指定说话人的说话人特征，以基于指定说话人的提取语音特征获取的语音识别结果为优化目标训练得到。需要说明的是，在训练阶段，特征提取模型的输入为上述的训练混合语音的语音特征和指定说话人的说话人特征，输出为从训练混合语音特征中提取的指定说话人的语音特征。

在一种可能的实现方式中，特征提取模型训练时，以基于指定说话人的提

取语音特征获取的语音识别结果为优化目标。以基于指定说话人的提取语音特征获取的语音识别结果为优化目标训练特征提取模型,使得基于特征提取模型能够从混合语音特征中提取出有利于语音识别的语音特征。

为了提升特征提取效果,在另一种可能的实现方式中,特征提取模型训练时,以指定说话人的提取语音特征以及基于指定说话人的提取语音特征获取的语音识别结果为优化目标。同时以指定说话人的提取语音特征以及基于指定说话人的提取语音特征获取的语音识别结果为优化目标,使得基于特征提取模型能够从混合语音特征中提取出有利于语音识别且趋近于指定说话人的标准语音特征的语音特征。需要说明的是,目标说话人的标准语音特征指的是根据指定说话人的语音(干净语音)获得的语音特征。

步骤 S103: 根据目标说话人的提取语音特征,获取目标说话人的语音识别结果。

根据目标说话人的提取语音特征,获取目标说话人的语音识别结果的实现方式有多种:在一种可能的实现方式中,可只根据目标说话人的提取语音特征,获取目标说话人的语音识别结果;为了提升语音识别效果,在另一种可能的实现方式中,可根据目标说话人的提取语音特征以及目标说话人的注册语音特征(目标说话人的注册语音特征指的是,目标说话人的注册语音的语音特征),获取目标说话人的语音识别结果,其中,目标说话人的注册语音特征作为识别辅助信息,用以提升语音识别效果。

具体的,可利用预先建立的语音识别模型,获取目标说话人的语音识别结果,更为具体的,将目标说话人的提取语音特征作为语音识别输入特征,或者将目标说话人的提取语音特征以及目标说话人的注册语音特征作为语音识别输入特征,将语音识别输入特征输入预先建立的语音识别模型,以得到目标说话人的语音识别结果。

需要说明的是,在将目标说话人的提取语音特征以及目标说话人的注册语音特征作为语音识别输入特征输入语音识别模型时,目标说话人的注册语音特征能够在目标说话人的提取语音特征不准确的情况下,辅助语音识别模型进行语音识别,从而提升语音识别效果。

优选的,语音识别模型可与特征提取模型联合训练得到,语音识别模型采用上述的“指定说话人的提取语音特征”为训练样本,以基于指定说话人的提

取语音特征获得的语音识别结果为优化目标训练得到。将特征提取模型与语音识别模型联合训练，使得特征提取模型能够朝着利于语音识别的方向去优化。

本申请实施例提供的语音识别方法能够从目标混合语音的语音特征中提取出目标说话人的语音特征，进而能够根据提取出的目标说话人的语音特征，
5 获得目标说话人的语音识别结果，由于本申请实施例在从目标混合语音的语音特征提取目标说话人的语音特征时，以趋于目标语音特征（用于获得与目标目标说话人的真实语音内容一致的语音识别结果的语音特征）为提取方向，因此，提取出的语音特征为目标语音特征或者趋近于目标语音特征的语音特征，基于该语音特征进行语音识别，能够获得较好的语音识别效果，即能够获得较为准确
10 的语音识别结果，用户体验较好。

第二实施例

上述实施例提到，用于从目标混合语音的语音特征中提取目标说话人的语音特征的特征提取模型，以及用于根据特征提取模型提取的特征获取目标说话人的语音识别结果的语音识别模型，可通过联合训练方式训练得到。本实施
15 例重点对特征提取模型与语音识别模型的联合训练过程进行介绍。

下面在图 2 的基础上结合图 3 对特征提取模型与语音识别模型的联合训练过程进行介绍，特征提取模型与语音识别模型的联合训练过程可以包括：

步骤 S201：从预先构建的训练数据集 S 中获取训练混合语音 S_m 。

其中，训练数据集 S 中包括多条训练数据，每条训练数据均包括指定说话人的语音（干净语音），以及包含该指定说话人的语音的训练混合语音，其中，
20 指定说话人的语音具有标注文本（标注文本为指定说话人的语音的语音内容）。

训练数据集 S 的构建过程包括：

步骤 a1、获取多个说话人的多条语音。

本步骤获取的多条语音中的每条语音为单一说话人的语音，每条语音具有
25 标注文本，假设一说话人的语音为内容为“今天天气不错”的语音，则该语音的标注文本为“<s>，今，填，天，气，不，错，</s>”，其中，“<s>”为句子开始符，“</s>”为句子结束符。

需要说明的是，多条语音的数量与多条语音对应的说话人的数量可以相同，也可以不同，假设步骤 a1 获取了 P 个说话人的 Q 多条语音，则 P 与 Q 的关系可以为 $P=Q$ （比如，获取说话人 a 的一条语音、说话人 b 的一条语音、说
30

话人 c 的一条语音), 也可以为 $P < Q$ (比如, 获取说话人 a 的两条语音、说话人 b 的一条语音、说话人 c 的三条语音), 也就是说, 针对每个说话人, 可获取一条语音, 也可以获取多条语音。

步骤 a2、将多条语音中的部分语音或全部语音中的每条语音作为指定说话人的语音: 将其它语音中其他说话人的一条或多条语音与该指定说话人的语音进行混合, 以得到一条训练混合语音, 将该条训练混合语音与该指定说话人的语音作为一条训练数据。

示例性的, 获取的多条语音包括说话人 a 的一条语音、说话人 b 的一条语音、说话人 c 的一条语音和说话人 d 的一条语音, 此处的每条语音都是单一说话人的干净语音, 可将说话人 a 的语音作为指定说话人的语音, 将其他说话人 (一个或多个说话人) 的语音与说话人 a 的语音混合, 以得到一条训练混合语音, 比如, 将说话人 b 的语音与说话人 a 语音混合, 或者, 将说话人 b 的语音、说话人 c 的语音与说话人 a 的语音混合, 将说话人 a 的语音以及通过将说话人 a 的语音与其他说话人的语音混合得到的训练混合语音作为一条训练数据, 同样的, 可将说话人 b 的语音作为指定说话人的语音, 将其他说话人 (一个或多个说话人) 的语音与说话人 b 的语音混合, 以得到一条训练混合语音, 将说话人 b 的语音以及通过将说话人 b 的语音与其他说话人的语音混合得到的训练混合语音作为一条训练数据, 按该方式可获得多条训练数据。

需要说明的是, 在将指定说话人的语音与其他说话人的语音混合时, 若其他说话人的语音的长度与指定说话人的语音的长度不同, 则需要将其他说话人的语音处理成与指定说话人的语音长度相同。假设指定说话人的语音包括 K 个语音帧, 即指定说话人的语音的长度为 K, 若其他说话人的语音的长度大于 K, 则可将其他说话人的语音中的第 K+1 个语音帧以及后面的语音帧删除, 即只保留前 K 个语音帧, 若其他说话人的语音的长度小于 K, 假设为 L, 则从前面复制 K-L 个语音帧进行补充。

步骤 a3、由获得的所有训练数据组成训练数据集。

步骤 S202: 获取训练混合语音 s_m 的语音特征作为训练混合语音特征 X_m , 并获取指定说话人的说话人特征作为训练说话人特征。

如第一实施例所述, 可预先建立说话人表征提取模型, 利用预先建立的说话人表征提取模型对指定说话人的注册语音提取说话人特征, 提取的说话人

特征作为训练说话人特征。如图 3 所示, 利用说话人表征提取模型 300 对指定说话人的注册语音提取短时声纹特征和长时声纹特征, 提取的短时声纹特征和长时声纹特征作为指定说话人的说话人特征。

需要说明的是, 说话人表征提取模型是在对特征提取模型与语音识别模型进行联合训练之前预先训练好的, 在特征提取模型与语音识别模型的联合训练阶段, 其参数固定, 不随特征提取模型与语音识别模型进行参数更新。

步骤 S203: 利用特征提取模型, 以训练混合语音特征 X_m 和训练说话人特征为依据, 从训练混合语音特征 X_m 中提取指定说话人的语音特征, 作为指定说话人的提取语音特征 $\bar{X}_i$ 。

具体的, 首先将训练混合语音特征 X_m 和训练说话人特征输入特征提取模型, 得到指定说话人对应的特征掩码 M , 然后根据指定说话人对应的特征掩码 M , 从训练混合语音特征 X_m 中提取指定说话人的语音特征, 作为指定说话人的提取语音特征 $\bar{X}_i$ 。

如图 3 所示, 将训练混合语音特征 X_m 和训练说话人特征输入特征提取模型 301, 特征提取模型 301 根据输入的训练混合语音特征 X_m 和训练说话人特征确定指定说话人对应的特征掩码 M 并输出。本实施例中的特征提取模型 301 可以但不限于为循环神经网络 (Recurrent Neural Network, RNN)、卷积神经网络 (Convolution Neural Network, CNN)、深度神经网络 (Deep Neural Network, DNN) 等。

需要说明的是, 训练混合语音特征 X_m 为训练混合语音中各语音帧的特征向量组成的特征向量序列 $[X_{m1}, X_{m2}, \dots, X_{mk}, \dots, X_{mK}]$ (K 为训练混合语音的语音帧的总数量), 在将训练混合语音特征 X_m 和训练说话人特征输入特征提取模型 301 时, 可将训练说话人特征与训练混合语音中每一语音帧的特征向量拼接, 拼接后输入特征提取模型 301。示例性的, 训练混合语音中每一语音帧的特征向量为 40 维, 训练说话人特征中的短时声纹特征和长时声纹特征均为 40 维, 则在训练混合语音中每一语音帧的特征向量拼接上短时声纹特征和长时声纹特征后, 可得到 120 维的拼接特征向量。在提取指定说话人的语音特征时, 结合短时声纹特征和长时声纹特征, 增加了输入信息的丰富程度, 这使得特征提取模型更好的提取到指定说话人的语音特征。

在本实施例中, 指定说话人对应的特征掩码 M 能够表征指定说话人的语

音特征在训练混合语音特征 X_m 中的占比。若将训练混合语音特征 X_m 表示为 $[X_{m1}, X_{m2}, \dots, X_{mk}, \dots, X_{mK}]$ ，将指定说话人对应的特征掩码 M 表示为 $[m_1, m_2, \dots, m_k, \dots, m_K]$ ，则 m_1 表示 x_{m1} 中指定说话人的语音特征的占比， m_2 表示 x_{m2} 中指定说话人的语音特征的占比，以此类推， m_K 表示 x_{mK} 中指定说话人的语音特征的占比， $m_1 \sim m_K$ 均为 $[0,1]$ 之间的值。在获得指定说话人对应的特征掩码 M 后，将训练混合语音特征 X_m 与指定说话人对应的特征掩码 M 逐帧相乘，便可得到从训练混合语音特征 X_m 中提取出的指定说话人的语音特征 $\bar{X}_t = [X_{t1}, X_{t2}, \dots, X_{tk}, \dots, X_{tK}]$ 。

步骤 S204：将指定说话人的提取语音特征 $\bar{X}_t$ 输入语音识别模型，获得指定说话人的语音识别结果 $\bar{T}_t$ 。

优选的，为了提升语音识别模型的识别效果，可获取指定说话人的注册语音特征（指定说话人的注册语音特征指的是，指定说话人的注册语音的语音特征） $X_e = [X_{e1}, X_{e2}, \dots, X_{ek}, \dots, X_{eK}]$ ，除了将指定说话人的提取语音特征 $\bar{X}_t$ 输入语音识别模型外，将指定说话人的注册语音特征 X_e 也输入语音识别模型，用指定说话人的注册语音特征 X_e 辅助语音识别模型进行语音识别。

可选的，本实施例中的语音识别模型可以包括：编码器模块、注意力模块和解码器模块。其中：

编码器模块，用于对指定说话人的提取语音特征 $\bar{X}_t$ 进行编码，以得到 $\bar{X}_t$ 的编码结果 $H^x = [h_1^x, h_2^x, \dots, h_K^x]$ ，以及对指定说话人的注册语音特征 X_e 进行编码，以得到 X_e 的编码结果 $H^e = [h_1^e, h_2^e, \dots, h_K^e]$ 。需要说明的是，若只将指定说话人的提取语音特征 $\bar{X}_t$ 输入语音识别模型，则编码器模块只需要对指定说话人的提取语音特征 $\bar{X}_t$ 进行编码即可。

在编码器模块的输入包括指定说话人的提取语音特征 $\bar{X}_t$ 和指定说话人的注册语音特征 X_e 的情况下：在一种可能的实现方式中，编码器模块中可设置两个编码模块，如图3所示，编码器模块中设置第一编码模块3021和第二编码模块3022，其中，第一编码模块3021用于对指定说话人的提取语音特征 $\bar{X}_t$ 进行编码，第二编码模块用于对指定说话人的注册语音特征 X_e 进行编码；在另一种可能的实现方式中，编码器模块中设置一个编码模块，对指定说话人的提取语音特征 $\bar{X}_t$ 的编码操作和对指定说话人的注册语音特征 X_e 的编码操作均由这一个编码模块执行，即两个编码过程共用一个编码模块。在编码器模块的

输入只包括指定说话人的提取语音特征 $\bar{X}_t$ 的情况下,编码器模块中只需要设置一个编码模块即可。不管编码器模块中设置一个编码模块还是设置两个编码模块,每个编码模块均可以包括一层或多层编码层,编码层可以采用单向或双向长短时记忆神经网络中的长短时记忆层,或者采用卷积神经网络的卷积层。

- 5 注意力模块,用于分别从指定说话人的提取语音特征 $\bar{X}_t$ 的编码结果 H^x 和指定说话人的注册语音特征 X_e 的编码结果中提取解码时刻解码所需的音频相关特征向量。

解码模块,用于对注意力模块提取出的音频相关特征向量进行解码,以得到解码时刻的识别结果。

- 10 如图3所示,注意力模块3023基于注意力机制,在每个解码时刻分别从 $\bar{X}_t$ 的编码结果 $H^x=[h_1^x, h_2^x, \dots, h_K^x]$ 和 X_e 的编码结果 $H^e=[h_1^e, h_2^e, \dots, h_K^e]$ 中提取当前解码时刻所需的音频相关特征向量。对于第 t 个解码时刻,提取的音频相关特征向量表征的是第 t 个解码时刻待解码字符的音频内容。

- 需要说明的是,注意力机制指的是,使用一个向量作为查询项(query),
15 对一组特征向量序列进行注意力机制操作,选出与查询项最匹配的特征向量作为输出,具体为,将查询项与特征向量序列中每个特征向量计算一个匹配系数,然后将这些匹配系数与对应的特征向量相乘并求和,得到一个新的特征向量即为与查询项最匹配的特征向量。

- 对于第 t 个解码时刻:注意力模块3023将解码器模块3024的状态特征向量 d_t 作为查询项,计算 d_t 与 $H^x=[h_1^x, h_2^x, \dots, h_K^x]$ 中每个特征向量的匹配系数 $w_1^x, w_2^x, \dots, w_K^x$,然后将匹配系数 $w_1^x, w_2^x, \dots, w_K^x$ 与 $H^x=[h_1^x, h_2^x, \dots, h_K^x]$ 中对应的特征向量相乘后求和,求和得到的特征向量作为音频相关特征向量 c_t^x ,同样的,注意力模块3023计算 d_t 与 $H^e=[h_1^e, h_2^e, \dots, h_K^e]$ 中每个特征向量的匹配系数 $w_1^e, w_2^e, \dots, w_K^e$,然后将匹配系数 $w_1^e, w_2^e, \dots, w_K^e$ 与
25 $H^e=[h_1^e, h_2^e, \dots, h_K^e]$ 中对应的特征向量相乘后求和,求和得到的特征向量作为音频相关特征向量 c_t^e ,在获得音频相关特征向量 c_t^x 和 c_t^e 后,将音频相关特征向量 c_t^x 和 c_t^e 输入解码器模块3024进行解码,以得到第 t 个解码时刻的识别结果。

- 其中,解码器模块3024的状态特征向量 d_t 根据第 $t-1$ 个解码时刻的识别
30 结果 y_{t-1} 和注意力模块输出的 c_{t-1}^x 和 c_{t-1}^e 确定。可选的,解码器模块3024可以

包括多个神经网络层，比如，两层单向长短时记忆层，在第 t 个解码时刻，解码器模块 3024 的第一层长短时记忆层以第 $t-1$ 个解码时刻的识别结果 y_{t-1} 以及注意力模块 3023 输出的 c_{t-1}^x 和 c_{t-1}^e 作为输入，计算得到解码器的状态特征向量 d_t ， d_t 输入注意力模块 3023，用于计算第 t 个解码时刻的 c_t^x 和 c_t^e ，然后将 c_t^x 和 c_t^e 拼接，拼接后向量作为解码器模块 3024 的第二层长短时记忆层的输入（比如， c_t^x 和 c_t^e 均为 128 维向量，将 c_t^x 与 c_t^e 拼接，可获得 256 维的拼接向量，256 维的拼接向量输入解码器模块 3024 的第二层长短时记忆层），计算得到解码器的输出 h_t^d ，最终根据 h_t^d 计算输出字符的后验概率，从而根据输出字符的后验概率确定第 t 个解码时刻的识别结果。

步骤 S205：根据指定说话人的提取语音特征 $\bar{X}_t$ 和指定说话人的语音识别结果 $\bar{T}_t$ 对特征提取模型进行参数更新，并根据指定说话人的语音识别结果 $\bar{T}_t$ 对语音识别模型进行参数更新。

具体的，步骤 S205 的实现过程可以包括：

步骤 S2051、获取训练混合语音 s_m 对应的指定说话人语音 s_t （指定说话人的语音）的标注文本 T_t ，并获取指定说话人语音 s_t 的语音特征作为指定说话人的标准语音特征 X_t 。

需要说明的是，此处的指定说话人语音 s_t 与上述指定说话人的注册语音为指定说话人的不同语音。

步骤 S2052、根据指定说话人的提取语音特征 $\bar{X}_t$ 和指定说话人的标准语音特征 X_t 确定第一预测损失 Loss1，并根据指定说话人的语音识别结果 $\bar{T}_t$ 和指定说话人语音 s_t 的标注文本 T_t ，确定第二预测损失 Loss2。

可选的，可计算指定说话人的提取语音特征 $\bar{X}_t$ 与指定说话人的标准语音特征 X_t 的最小均方误差，作为第一预测损失 Loss1，根据指定说话人的语音识别结果 $\bar{T}_t$ 和指定说话人语音 s_t 的标注文本 T_t 计算交叉熵损失，作为第二预测损失。

步骤 S2053、根据第一预测损失 Loss1 和第二预测损失 Loss2 对特征提取模型进行参数更新，并根据第二预测损失 Loss2 对语音识别模型进行参数更新。

根据第一预测损失 Loss1 和第二预测损失 Loss2 对特征提取模型进行参数更新使得基于特征提取模型能够从训练混合语音特征中提取出趋近于指定说

话人的标准语音特征且有利于语音识别的语音特征,将该语音特征输入语音识别模型进行语音识别,能够获得较好的语音识别效果。

第三实施例

在上述第三实施例的基础上,本实施例对第一实施例中的“利用预先建立的特征提取模型,以目标混合语音特征和目标说话人特征为依据,从目标混合语音特征中提取指定说话人的语音特征,以得到目标说话人的提取语音特征”的过程进行介绍。

利用预先建立的特征提取模型,以目标混合语音特征和目标说话人特征为依据,从目标混合语音特征中提取指定说话人的语音特征,以得到目标说话人的提取语音特征的过程可以包括:

步骤 b1、将目标混合语音的语音特征和目标说话人的说话人特征输入特征提取模型,得到目标说话人对应的特征掩码。

其中,目标说话人对应的特征掩码能够表征目标混合语音的语音特征中目标说话人的语音特征的占比。

步骤 b2、根据目标说话人对应的特征掩码,从目标混合语音的语音特征中提取目标说话人的语音特征,以得到目标说话人的提取语音特征。

具体的,将目标混合语音的语音特征与目标说话人对应的特征掩码逐帧相乘,从而得到目标说话人的提取语音特征。

在获得目标说话人的提取语音特征后,将目标说话人的提取语音特征和目标说话人的注册语音特征输入语音识别模型,得到目标说话人的语音识别结果,具体的,将目标说话人的提取语音特征和目标说话人的注册语音特征输入语音识别模型,得到目标说话人的语音识别结果的过程可以包括:

步骤 c1、基于语音识别模型的编码器模块,分别对目标说话人的提取语音特征和目标说话人的注册语音特征进行编码,以得到两个编码结果。

步骤 c2、基于语音识别模型的注意力模块,分别从两个编码结果中提取解码时刻解码所需的音频相关特征向量。

步骤 c3、基于语音识别模型的解码器模块,对分别从两个编码结果中提取的音频相关特征向量进行解码,得到解码时刻的识别结果。

需要说明的是,将目标说话人的提取语音特征输入语音识别模型,得到所述目标说话人的语音识别结果的过程,与训练阶段将指定说话人的提取语音特

征和指定说话人的注册语音特征输入语音识别模型,得到指定说话人的语音识别结果的实现过程类似,步骤 c1~步骤 c3 的具体实现过程可参见第二实施例中关于编码器模块、注意力模块和解码器模块的介绍,本实施例在此不做赘述。

5 经由上述第一实施例至第三实施例可知,本申请提供的语音识别方法具有如下几方面的优势:其一,本申请对目标说话人的注册语音提取多尺度声纹特征输入特征提取模型,增加特征提取模型输入信息的丰富程度,提升了特征提取模型的特征提取效果;其二,特征提取模型与语音识别模型进行联合训练,使得语音识别模型预测损失能够作用于特征提取模型,进而使得特征提取模型能够提取有利于语音识别的语音特征,从而能够提升语音识别结果的准确度;10 其三,将目标说话人的注册语音的语音特征作为语音识别模型的额外输入,以在特征提取模型提取的语音特征不佳时,能够辅助语音识别模型进行语音识别,从而得到较为准确的语音识别结果。综上,本申请提供的语音识别方法能够在复杂人声干扰的情况下,准确地识别出目标说话人的语音内容。

第四实施例

15 本申请实施例还提供了一种语音识别装置,下面对本申请实施例提供的语音识别装置进行描述,下文描述的语音识别装置与上文描述的语音识别方法可相互对应参照。

请参阅图 4,示出了本申请实施例提供的语音识别装置的结构示意图,可以包括:特征获取模块 401、特征提取模块 402 和语音识别模块 403。

20 特征获取模块 401,用于获取目标混合语音的语音特征以及目标说话人的说话人特征。

特征提取模块 402,用于以趋于目标语音特征为提取方向,根据所述目标混合语音的语音特征以及所述目标说话人的说话人特征,从所述目标混合语音的语音特征中提取所述目标说话人的语音特征,以得到所述目标说话人的提取25 语音特征,其中,所述目标语音特征为用于获得与所述目标说话人的真实语音内容一致的语音识别结果的语音特征。

语音识别模块 403,用于根据所述目标说话人的提取语音特征,获取所述目标说话人的语音识别结果。

可选的,特征获取模块 401 包括:语音特征获取模块和说话人特征获取模

块。

所述语音特征获取模块，用于获取目标混合语音的语音特征。

所述说话人特征获取模块，用于获取目标说话人的说话人特征。

5 可选的，所述说话人特征获取模块在获取目标说话人的说话人特征时，具体用于获取所述目标说话人的注册语音，对所述目标说话人的注册语音提取短时声纹特征和长时声纹特征，以得到多尺度声纹特征，作为所述目标说话人的说话人特征。

10 可选的，特征提取模块 402 具体用于利用预先建立的特征提取模型，以所述目标混合语音的语音特征以及所述目标说话人的说话人特征为依据，从所述目标混合语音的语音特征中提取所述目标说话人的语音特征。

其中，所述特征提取模型采用包含指定说话人的语音的训练混合语音的语音特征和指定说话人的说话人特征，以基于所述指定说话人的提取语音特征获取的语音识别结果为优化目标训练得到，所述指定说话人的提取语音特征为从所述训练混合语音的语音特征中提取的所述指定说话人的语音特征。

15 可选的，所述特征提取模型同时以所述指定说话人的提取语音特征和基于所述指定说话人的提取语音特征获取的语音识别结果为优化目标训练得到。

可选的，特征提取模块 402 可以包括：特征掩码确定子模块和语音特征提取子模块。

20 所述特征掩码确定子模块，用于将所述目标混合语音的语音特征以及所述目标说话人的说话人特征输入所述特征提取模型，得到所述目标说话人对应的特征掩码，其中，所述特征掩码能够表征对应说话人的语音特征在所述目标混合语音的语音特征中的占比。

25 所述语音特征提取子模块，用于根据所述目标混合语音的语音特征和所述目标说话人对应的特征掩码，从所述目标混合语音的语音特征中提取所述目标说话人的语音特征。

可选的，语音识别模块 403，具体用于根据所述目标说话人的提取语音特征以及所述目标说话人的注册语音特征，获取所述目标说话人的语音识别结果；其中，所述目标说话人的注册语音特征为所述目标说话人的注册语音的语音特征。

可选的，语音识别模块 403，具体用于将至少包括所述目标说话人的提取语音特征的语音识别输入特征输入预先建立的语音识别模型，得到所述目标说话人的语音识别结果。

其中，所述语音识别模型与所述特征提取模型联合训练得到，所述语音识别模型采用所述指定说话人的提取语音特征，以基于所述指定说话人的提取语音特征获得的语音识别结果为优化目标训练得到。

可选的，语音识别模块 403 在将至少包括所述目标说话人的提取语音特征的语音识别输入特征输入预先建立的语音识别模型，得到所述目标说话人的语音识别结果时，具体用于：

10 基于所述语音识别模型的编码器模块，对所述语音识别输入特征进行编码，以得到编码结果；基于所述语音识别模型的注意力模块，从所述编码结果中提取解码时刻解码所需的音频相关特征向量；基于所述语音识别模型的解码器模块，对从所述编码结果中提取的所述音频相关特征向量进行解码，得到所述解码时刻的识别结果。

15 可选的，本申请实施例提供的语音识别装置还可以包括：模型训练模块。模型训练模块可以包括：提取语音特征获取模块、语音识别结果获取模块、参数更新模块。

所述提取语音特征获取模块，用于利用特征提取模型，从所述训练混合语音的语音特征中提取所述指定说话人的语音特征，以得到所述指定说话人的提取语音特征。

所述语音识别结果获取模块，用于利用语音识别模型和所述指定说话人的提取语音特征，获取所述指定说话人的语音识别结果。

25 所述模型更新模块，用于根据所述指定说话人的提取语音特征和所述指定说话人的语音识别结果对特征提取模型进行参数更新，并根据所述指定说话人的语音识别结果对语音识别模型进行参数更新。

可选的，模型更新模块可以包括：标注文本获取模块、标准语音特征获取模块、预测损失确定模块和参数更新模块。

所述训练混合语音对应有所述指定说话人的语音。

所述标准语音特征获取模块，用于获取获取所述指定说话人的语音的语音

特征作为所述指定说话人的标准语音特征。

所述标注文本获取模块,用于获取所述指定说话人的语音的标注文本。

所述预测损失确定模块,用于根据所述指定说话人的提取语音特征和所述指定说话人的标准语音特征确定第一预测损失,并根据所述指定说话人的语音识别结果和所述指定说话人的语音的标注文本,确定第二预测损失。

所述参数更新模块,用于根据所述第一预测损失和所述第二预测损失对特征提取模型进行参数更新,并根据所述第二预测损失对语音识别模型进行参数更新。

可选的,所述训练混合语音以及所述训练混合语音对应的所述指定说话人的语音从预先构建的训练数据集中获取,本申请实施例提供的语音识别装置还可以包括:训练数据集构建模块。

所述训练数据集构建模块用于:

获取多个说话人的多条语音,其中,每条语音为单一说话人的语音,每条语音具有标注文本;将所述多条语音中的部分语音或全部语音中的每条语音作为指定说话人的语音:将其它语音中其他说话人的一条或多条语音与该指定说话人的语音进行混合,以得到一条训练混合语音,将该指定说话人的语音与通过混合得到的训练混合语音作为一条训练数据;由获得的所有训练数据组成所述训练数据集。

本申请实施例提供的语音识别装置能够从目标混合语音的语音特征中提取出目标说话人的语音特征,进而能够根据提取出的目标说话人的语音特征,获得目标说话人的语音识别结果,由于本申请实施例在从目标混合语音的语音特征提取目标说话人的语音特征时,以趋于目标语音特征(用于获得与目标目标说话人的真实语音内容一致的语音识别结果的语音特征)为提取方向,因此,提取出的语音特征为目标语音特征或者趋近于目标语音特征的语音特征,基于该语音特征进行语音识别,能够获得较好的语音识别效果,即能够获得较为准确的语音识别结果,用户体验较好。

第五实施例

本申请实施例还提供了一种语音识别设备,请参阅图 5,示出了该语音识别设备的结构示意图,该语音识别设备可以包括:至少一个处理器 501,至少

一个通信接口 502, 至少一个存储器 503 和至少一个通信总线 504;

在本申请实施例中, 处理器 501、通信接口 502、存储器 503、通信总线 504 的数量为至少一个, 且处理器 501、通信接口 502、存储器 503 通过通信总线 504 完成相互间的通信;

5 处理器 501 可能是一个中央处理器 CPU, 或者是特定集成电路 ASIC (Application Specific Integrated Circuit), 或者是被配置成实施本发明实施例的一个或多个集成电路等;

存储器 503 可能包含高速 RAM 存储器, 也可能还包括非易失性存储器 (non-volatile memory) 等, 例如至少一个磁盘存储器;

10 其中, 存储器存储有程序, 处理器可调用存储器存储的程序, 所述程序用于:

获取目标混合语音的语音特征以及目标说话人的说话人特征;

以趋于目标语音特征为提取方向, 根据目标混合语音的语音特征以及目标说话人的说话人特征, 从目标混合语音的语音特征中提取目标说话人的语音特征, 以得到目标说话人的提取语音特征, 其中, 所述目标语音特征为用于获得与目标说话人的真实语音内容一致的语音识别结果的语音特征;

根据所述目标说话人的提取语音特征, 获取所述目标说话人的语音识别结果。

可选的, 所述程序的细化功能和扩展功能可参照上文描述。

20 **第六实施例**

本申请实施例还提供一种可读存储介质, 该可读存储介质可存储有适于处理器执行的程序, 所述程序用于:

获取目标混合语音的语音特征以及目标说话人的说话人特征;

25 以趋于目标语音特征为提取方向, 根据目标混合语音的语音特征以及目标说话人的说话人特征, 从目标混合语音的语音特征中提取目标说话人的语音特征, 以得到目标说话人的提取语音特征, 其中, 所述目标语音特征为用于获得与目标说话人的真实语音内容一致的语音识别结果的语音特征;

根据所述目标说话人的提取语音特征, 获取所述目标说话人的语音识别结果。

30 可选的, 所述程序的细化功能和扩展功能可参照上文描述。

最后，还需要说明的是，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。

本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。

对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业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公开的原理和新颖特点相一致的最宽的范围。

权 利 要 求 书

1、一种语音识别方法，其特征在于，包括：

获取目标混合语音的语音特征以及目标说话人的说话人特征；

5 以趋于目标语音特征为提取方向，根据所述目标混合语音的语音特征以及
所述目标说话人的说话人特征，从所述目标混合语音的语音特征中提取所述目
标说话人的语音特征，以得到所述目标说话人的提取语音特征，其中，所述目
标语音特征为用于获得与所述目标说话人的真实语音内容一致的语音识别结
果的语音特征；

10 根据所述目标说话人的提取语音特征，获取所述目标说话人的语音识别结
果。

2、根据权利要求 1 所述的语音识别方法，其特征在于，获取所述目标说
话人的说话人特征，包括：

获取所述目标说话人的注册语音；

15 对所述目标说话人的注册语音提取短时声纹特征和长时声纹特征，以得到
多尺度声纹特征，作为所述目标说话人的说话人特征。

3、根据权利要求 1 所述的语音识别方法，其特征在于，所述以趋于目标
语音特征为提取方向，根据所述目标混合语音的语音特征以及所述目标说话人
的说话人特征，从所述目标混合语音的语音特征中提取所述目标说话人的语音
特征，包括：

20 利用预先建立的特征提取模型，以所述目标混合语音的语音特征以及所述
目标说话人的说话人特征为依据，从所述目标混合语音的语音特征中提取所述
目标说话人的语音特征；

25 其中，所述特征提取模型采用包含指定说话人的语音的训练混合语音的语
音特征和所述指定说话人的说话人特征，以基于所述指定说话人的提取语音特
征获取的语音识别结果为优化目标训练得到，所述指定说话人的提取语音特征
为从所述训练混合语音的语音特征中提取的所述指定说话人的语音特征。

4、根据权利要求 3 所述的语音识别方法，其特征在于，所述特征提取模
型同时以所述指定说话人的提取语音特征和基于所述指定说话人的提取语音
特征获取的语音识别结果为优化目标训练得到。

30 5、根据权利要求 3 或 4 所述的语音识别方法，其特征在于，所述利用预

先建立的特征提取模型,以所述目标混合语音的语音特征以及所述目标说话人的说话人特征为依据,从所述目标混合语音的语音特征中提取所述目标说话人的语音特征,包括:

将所述目标混合语音的语音特征以及所述目标说话人的说话人特征输入
5 所述特征提取模型,得到所述目标说话人对应的特征掩码;

根据所述目标混合语音的语音特征和所述目标说话人对应的特征掩码,从所述目标混合语音的语音特征中提取所述目标说话人的语音特征。

6、根据权利要求 1 所述的语音识别方法,其特征在于,所述根据所述目标说话人的提取语音特征,获取所述目标说话人的语音识别结果,包括:

10 根据所述目标说话人的提取语音特征以及所述目标说话人的注册语音特征,获取所述目标说话人的语音识别结果;

其中,所述目标说话人的注册语音特征为所述目标说话人的注册语音的语音特征。

7、根据权利要求 3 或 4 所述的语音识别方法,其特征在于,所述根据所述目标说话人的提取语音特征,获取所述目标说话人的语音识别结果,包括:

将至少包括所述目标说话人的提取语音特征的语音识别输入特征输入预先建立的语音识别模型,得到所述目标说话人的语音识别结果;

所述语音识别模型与所述特征提取模型联合训练得到,所述语音识别模型采用所述指定说话人的提取语音特征,以基于所述指定说话人的提取语音特征
20 获得的语音识别结果为优化目标训练得到。

8、根据权利要求 7 所述的语音识别方法,其特征在于,将所述语音识别输入特征输入所述语音识别模型,得到所述目标说话人的语音识别结果,包括:

基于所述语音识别模型的编码器模块,对所述语音识别输入特征进行编码,以得到编码结果;

25 基于所述语音识别模型的注意力模块,从所述编码结果中提取解码时刻解码所需的音频相关特征向量;

基于所述语音识别模型的解码器模块,对从所述编码结果中提取的所述音频相关特征向量进行解码,得到所述解码时刻的识别结果。

9、根据权利要求 7 所述的语音识别方法,其特征在于,所述语音识别模型与
30 所述特征提取模型联合训练的过程包括:

利用特征提取模型,从所述训练混合语音的语音特征中提取所述指定说话人的语音特征,以得到所述指定说话人的提取语音特征;

利用语音识别模型和所述指定说话人的提取语音特征,获取所述指定说话人的语音识别结果;

- 5 根据所述指定说话人的提取语音特征和所述指定说话人的语音识别结果对特征提取模型进行参数更新,并根据所述指定说话人的语音识别结果对语音识别模型进行参数更新。

10、根据权利要求9所述的语音识别方法,其特征在于,所述训练混合语音对应有所述指定说话人的语音;

- 10 所述根据所述指定说话人的提取语音特征和所述指定说话人的语音识别结果对特征提取模型进行参数更新,并根据所述指定说话人的语音识别结果对语音识别模型进行参数更新,包括:

获取所述指定说话人的语音的标注文本,并获取所述指定说话人的语音的语音特征作为所述指定说话人的标准语音特征;

- 15 根据所述指定说话人的提取语音特征和所述指定说话人的标准语音特征确定第一预测损失,并根据所述指定说话人的语音识别结果和所述指定说话人的语音的标注文本,确定第二预测损失;

根据所述第一预测损失和所述第二预测损失对特征提取模型进行参数更新,并根据所述第二预测损失对语音识别模型进行参数更新。

- 20 11、根据权利要求10所述的语音识别方法,其特征在于,所述训练混合语音以及所述训练混合语音对应的所述指定说话人的语音从预先构建的训练数据集中获取;

所述训练数据集的构建过程包括:

- 25 获取多个说话人的多条语音,其中,每条语音为单一说话人的语音,每条语音具有标注文本;

将所述多条语音中的部分语音或全部语音中的每条语音作为指定说话人的语音:将其它语音中其他说话人的一条或多条语音与该指定说话人的语音进行混合,以得到一条训练混合语音,将该指定说话人的语音与通过混合得到的训练混合语音作为一条训练数据;

- 30 由获得的所有训练数据组成所述训练数据集。

12、一种语音识别装置，其特征在于，包括：特征获取模块、特征提取模块和语音识别模块；

所述特征获取模块，用于获取目标混合语音的语音特征以及目标说话人的说话人特征；

5 所述特征提取模块，用于以趋于目标语音特征为提取方向，根据所述目标混合语音的语音特征以及所述目标说话人的说话人特征，从所述目标混合语音的语音特征中提取所述目标说话人的语音特征，以得到所述目标说话人的提取语音特征，其中，所述目标语音特征为用于获得与所述目标说话人的真实语音内容一致的语音识别结果的语音特征；

10 所述语音识别模块，用于根据所述目标说话人的提取语音特征，获取所述目标说话人的语音识别结果。

13、根据权利要求 12 所述的语音识别装置，其特征在于，所述特征获取模块包括：说话人特征获取模块；

15 所述说话人特征获取模块，用于获取所述目标说话人的注册语音，对所述目标说话人的注册语音提取短时声纹特征和长时声纹特征，以得到多尺度声纹特征，作为所述目标说话人的说话人特征。

14、根据权利要求 12 所述的语音识别装置，其特征在于，所述特征提取模块具体用于利用预先建立的特征提取模型，以所述目标混合语音的语音特征以及所述目标说话人的说话人特征为依据，从所述目标混合语音的语音特征中
20 提取所述目标说话人的语音特征；

其中，所述特征提取模型采用包含指定说话人的语音的训练混合语音的语音特征和指定说话人的说话人特征，以基于所述指定说话人的提取语音特征获取的语音识别结果为优化目标训练得到，所述指定说话人的提取语音特征为从所述训练混合语音的语音特征中提取的所述指定说话人的语音特征。

25 15、根据权利要求 12 所述的语音识别装置，其特征在于，所述语音识别模块，具体用于根据所述目标说话人的提取语音特征以及所述目标说话人的注册语音特征，获取所述目标说话人的语音识别结果；

其中，所述目标说话人的注册语音特征为所述目标说话人的注册语音的语音特征。

30 16、根据权利要求 14 所述的语音识别装置，其特征在于，所述语音识别

模块,用于将至少包括所述目标说话人的提取语音特征的语音识别输入特征输入预先建立的语音识别模型,得到所述目标说话人的语音识别结果;

其中,所述语音识别模型与所述特征提取模型联合训练得到,所述语音识别模型采用所述指定说话人的提取语音特征,以基于所述指定说话人的提取语

5 音特征获得的语音识别结果为优化目标训练得到。

17、一种语音识别设备,其特征在于,包括:存储器和处理器;

所述存储器,用于存储程序;

所述处理器,用于执行所述程序,实现如权利要求 1~11 中任一项所述的语音识别方法的各个步骤。

10 18、一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时,实现如权利要求 1~11 中任一项所述的语音识别方法的各个步骤。

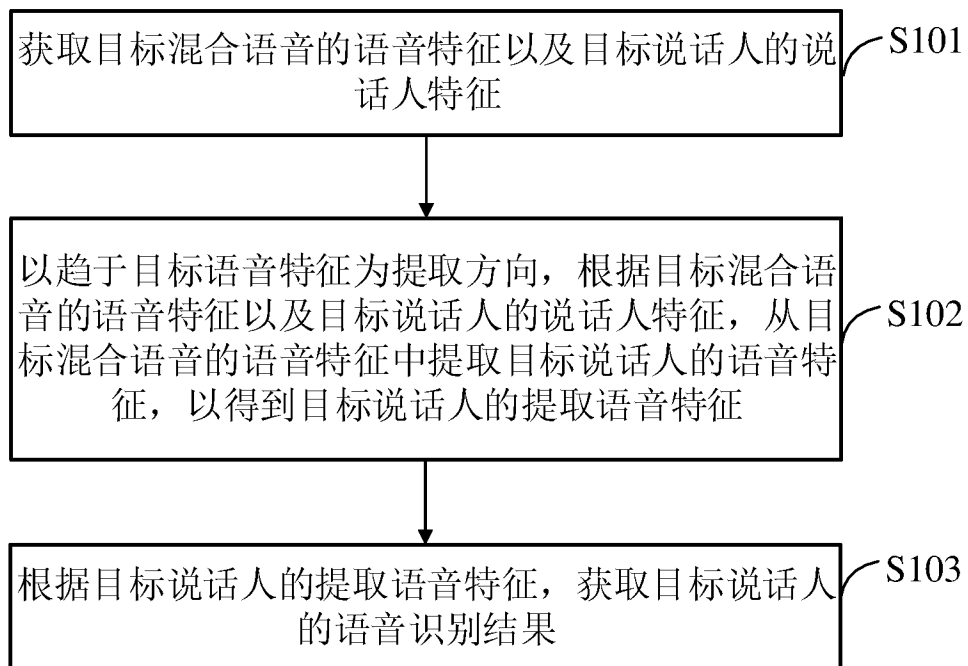


图 1

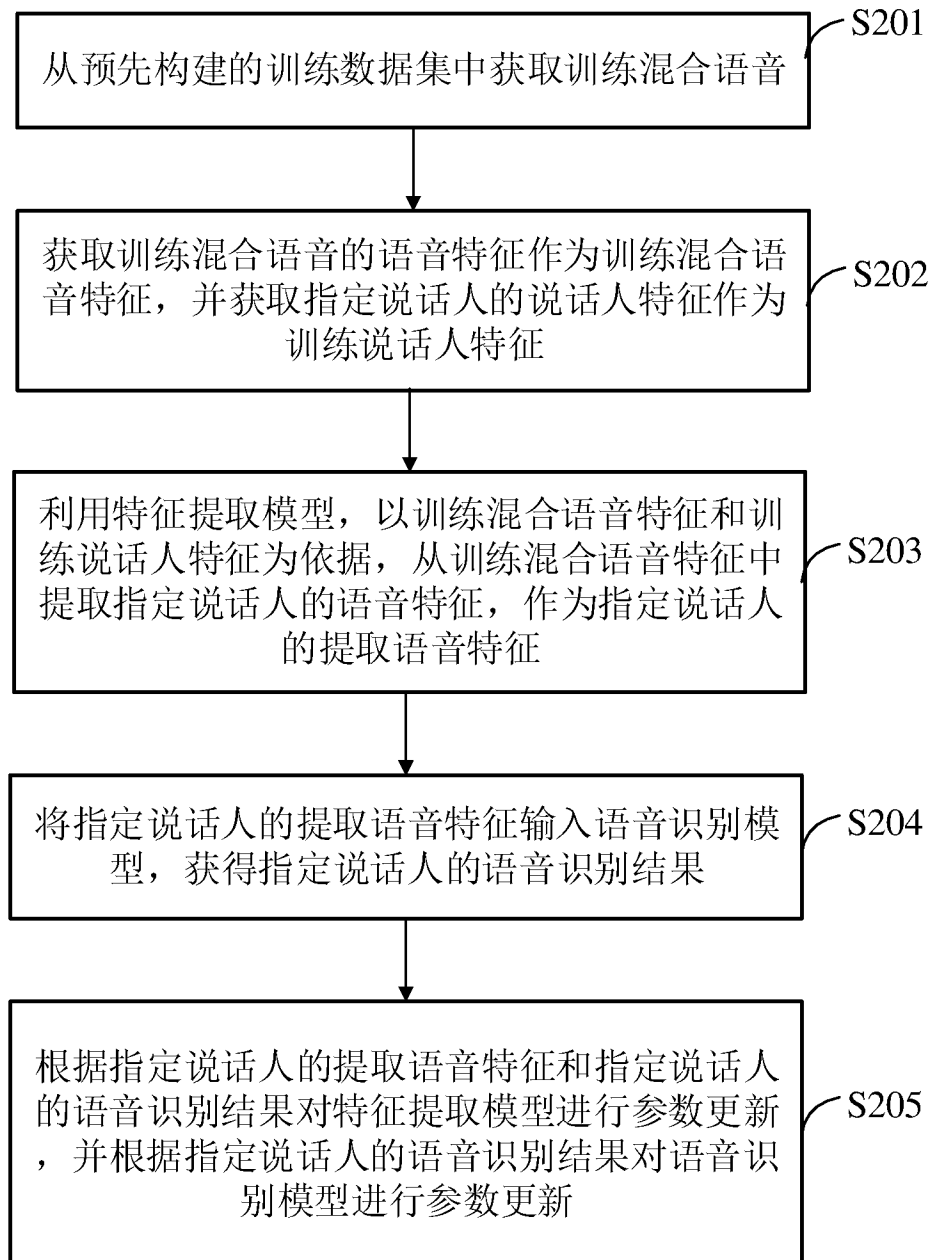


图 2

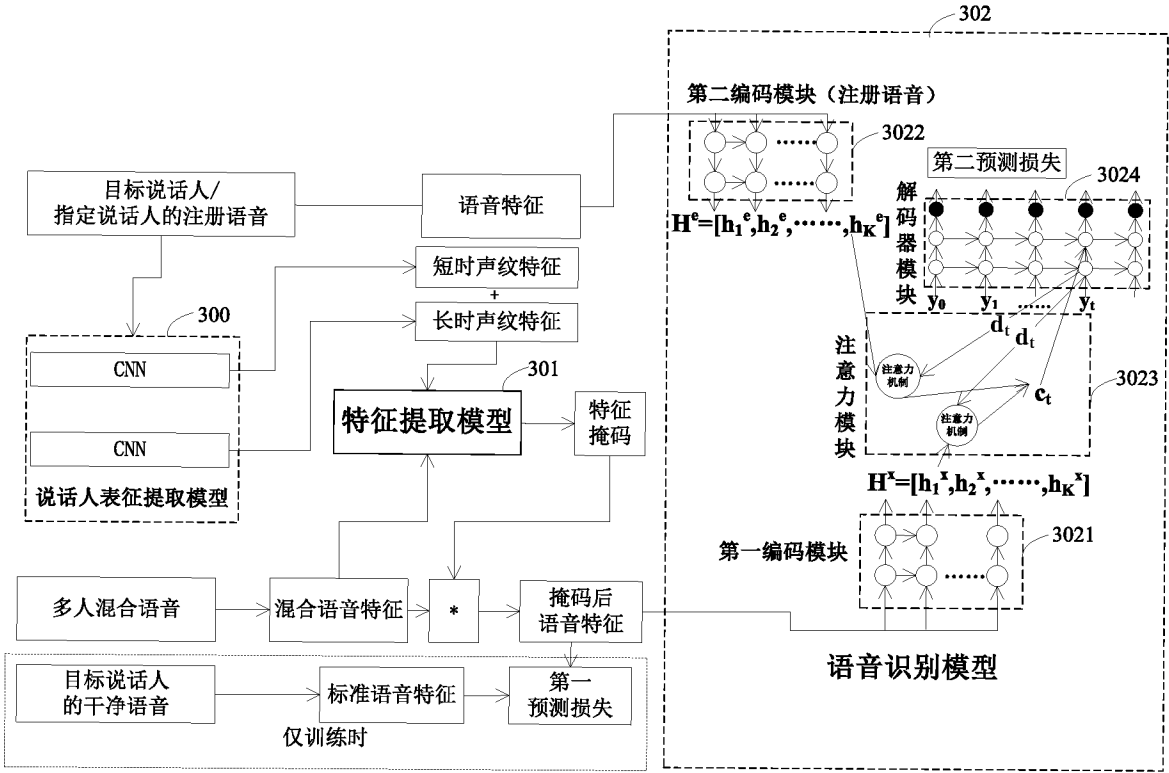


图 3

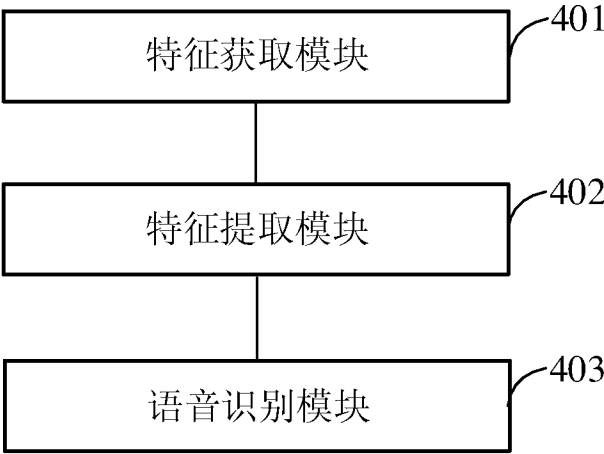


图 4

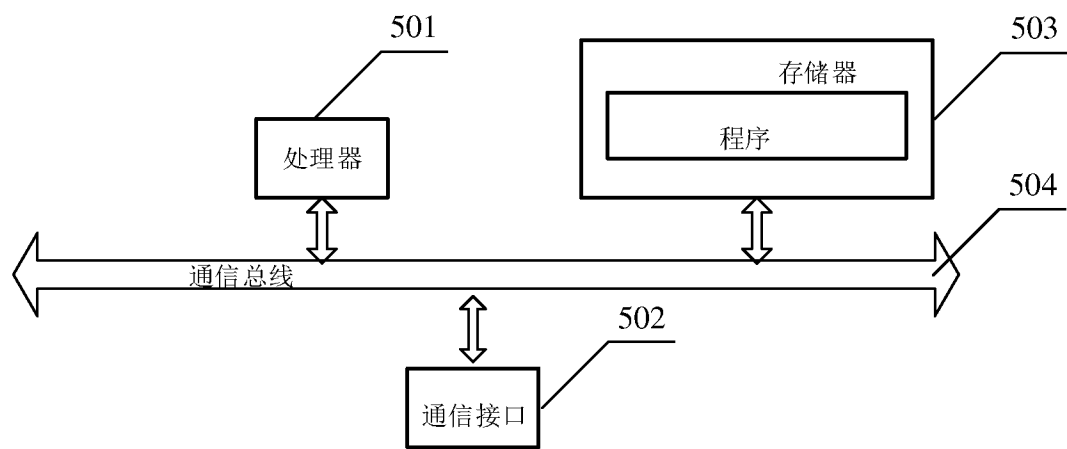


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/129733

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/02(2013.01)i; G10L 17/04(2013.01)i; G10L 25/24(2013.01)i; G10L 25/30(2013.01)i; G10L 25/51(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNTXT; CNKI; VEN; USTXT; WOTXT; EPTXT; IEEE: 科大讯飞, 语音识别, 目标说话人, 混合语音, 特征, 内容, 一致, 模型, 训练, speech, recognition, mixed speech, target voice feature, voice content, target speaker, model

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 110827853 A (GUANGZHOU GUOYIN INTELLIGENT TECHNOLOGY CO., LTD.) 21 February 2020 (2020-02-21) description, paragraphs [0003]-[0032]	1-18
A	CN 110288989 A (ANHUI XINGBO YUANSHI INFORMATION TECHNOLOGY CO., LTD.) 27 September 2019 (2019-09-27) entire document	1-18
A	CN 111128197 A (BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 08 May 2020 (2020-05-08) entire document	1-18
A	CN 109874096 A (TIANJIN UNIVERSITY) 11 June 2019 (2019-06-11) entire document	1-18
A	CN 111433847 A (UBTECH ROBOTICS CORP.) 17 July 2020 (2020-07-17) entire document	1-18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

05 May 2022

Date of mailing of the international search report

20 May 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/129733

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110827853	A	21 February 2020	None			
CN	110288989	A	27 September 2019	None			
CN	111128197	A	08 May 2020	None			
CN	109874096	A	11 June 2019	None			
CN	111433847	A	17 July 2020	US	2021201890	A1	01 July 2021
				WO	2021134520	A1	08 July 2021

国际检索报告

国际申请号

PCT/CN2021/129733

A. 主题的分类 G10L 17/02(2013.01)i; G10L 17/04(2013.01)i; G10L 25/24(2013.01)i; G10L 25/30(2013.01)i; G10L 25/51(2013.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G10L 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNTXT;CNKI;VEN;USTXT;WOTXT;EPTXT;IEEE:科大讯飞, 语音识别, 目标说话人, 混合语音, 特征, 内容, 一致, 模型, 训练, speech, recognition, mixed speech, target voice feature, voice content, target speaker, model		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 110827853 A (广州国音智能科技有限公司) 2020年2月21日 (2020 - 02 - 21) 说明书第[0003]-[0032]段	1-18
A	CN 110288989 A (安徽兴博远信息科技有限公司) 2019年9月27日 (2019 - 09 - 27) 全文	1-18
A	CN 111128197 A (北京邮电大学) 2020年5月8日 (2020 - 05 - 08) 全文	1-18
A	CN 109874096 A (天津大学) 2019年6月11日 (2019 - 06 - 11) 全文	1-18
A	CN 111433847 A (深圳市优必选科技股份有限公司) 2020年7月17日 (2020 - 07 - 17) 全文	1-18
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2022年5月5日		国际检索报告邮寄日期 2022年5月20日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 张瀛 电话号码 (86-512) 88997329

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/129733

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110827853	A	2020年2月21日	无	
CN	110288989	A	2019年9月27日	无	
CN	111128197	A	2020年5月8日	无	
CN	109874096	A	2019年6月11日	无	
CN	111433847	A	2020年7月17日	US 2021201890 A1	2021年7月1日
				WO 2021134520 A1	2021年7月8日