

(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
2020 年 12 月 10 日 (10.12.2020)



(10) 国际公布号
WO 2020/244153 A1

(51) 国际专利分类号:

G10L 15/02 (2006.01) *G10L 15/18* (2013.01)
G10L 15/26 (2006.01) *G10L 15/04* (2013.01)
G10L 15/30 (2013.01) *G10L 15/05* (2013.01)
G10L 15/28 (2013.01) *G10L 17/00* (2013.01)

(21) 国际申请号: PCT/CN2019/118267

(22) 国际申请日: 2019 年 11 月 14 日 (14.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

201910494580.7 2019 年 6 月 5 日 (05.06.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福

安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

(72) 发明人: 陈家荣(CHEN, Jiarong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司(ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路 6 号 4501 房(部位: 自编 01-03 和 08-12 单元)(仅限办公用途), Guangdong 510623 (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,

(54) Title: CONFERENCE VOICE DATA PROCESSING METHOD AND APPARATUS, COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 会议语音数据处理方法、装置、计算机设备和存储介质



图 2

(57) Abstract: A conference voice data processing method and apparatus, a computer device and a storage medium. The method comprises: receiving a conference recording request sent by a terminal, sending a voice recording instruction to the terminal according to the conference recording request, and acquiring voice data uploaded by the terminal; performing feature extraction on the voice data to obtain multiple pieces of voice feature information; performing, by means of a preset voiceprint recognition model, voiceprint recognition on the voice feature information in order to obtain multiple voice fragments and corresponding voiceprint identifiers, and

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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, 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))。

converting the multiple voice fragments into corresponding text information; analyzing, by means of a preset semantic analysis model, key words and correction information in the text information, correcting the text information according to the correction information, and adding corresponding key words to corrected text information; and generating, according to a conference theme and the text information having the key words added thereto, corresponding conference report data according to a preset means.

(57) 摘要: 一种会议语音数据处理方法、装置、计算机设备和存储介质, 所述方法包括: 接收终端发送的会议记录请求, 根据会议记录请求向终端发送录音指令, 获取终端上传的语音数据; 对语音数据进行特征提取, 得到多个语音特征信息; 通过预设的声纹识别模型对语音特征信息进行声纹识别, 得到多个语音片段和对应的声纹标识, 并将多个语音片段转换为对应的文本信息; 通过预设的语义分析模型分析出文本信息中的关键字和校正信息, 根据校正信息对文本信息进行校正, 并对校正后的文本信息添加对应的关键字; 根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

会议语音数据处理方法、装置、计算机设备和存储介质

相关申请的交叉引用：

本申请要求于 2019 年 06 月 05 日提交至中国专利局，申请号为 2019104945807，申请名称为“会议语音数据处理方法、装置、计算机设备和存储介质”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

本申请涉及一种会议语音数据处理方法、装置、计算机设备和存储介质。

背景技术

随着经济和互联网的迅速发展，商业社会和商务模式也在迅速发展，各种商务会议的召开需求也不断增大，会议中通常包含有价值的会议信息需要记录下来。

然而，传统的会议中是通过人工记录会议纪要，会议记录效率较低。随着人工智能技术的迅速发展，出现了一些从音视频会议中提取会议纪要的方式，这种方式通常是从会议音视频中提取语音信息，再通过语音识别转换为对应的文本信息进行存储。但这种方式在时间较长会议内容较多的情况下，生成的文本信息内容较多较繁琐，无法区分具体的会议内容来自哪个发言人，导致会议记录的效率较低，会议语音数据的识别准确率也较低。因此，如何有效提高会议语音数据的识别准确率成为目前需要解决的技术问题。

发明内容

根据本申请公开的各种实施例，提供一种会议语音数据处理方法、装置、计算机设备和存储介质。

一种会议语音数据处理方法包括：

接收终端发送的会议记录请求，根据所述会议记录请求向所述终端发送录音指令，所述会议记录请求携带了会议主题；

获取所述终端根据所述录音指令上传的语音数据；

对所述语音数据进行特征提取，得到多个语音特征信息；

将所述多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音

片段和对应的声纹标识，根据所述声纹标识将所述多个语音片段转换为对应的文本信息；

将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信息，利用所述关键字和校正信息生成分析结果；

根据所述分析结果中的校正信息对所述文本信息进行校正，对校正后的文本信息添加对应的关键字；及

根据所述会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

一种会议语音数据处理装置包括：

请求接收模块，用于接收终端发送的会议记录请求，根据所述会议记录请求向所述终端发送录音指令，所述会议记录请求携带了会议主题；

数据获取模块，用于获取所述终端根据所述录音指令上传的语音数据；

特征提取模块，用于对所述语音数据进行特征提取，得到多个语音特征信息；

声纹识别模块，用于将所述多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据所述声纹标识将所述多个语音片段转换为对应的文本信息；

语义分析模块，用于将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信息，利用所述关键字和校正信息生成分析结果；

文本校正模块，用于根据所述分析结果中的校正信息对所述文本信息进行校正，对校正后的文本信息添加对应的关键字；

会议报告生成模块，用于根据所述会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

一种计算机设备，包括存储器及一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

接收终端发送的会议记录请求，根据所述会议记录请求向所述终端发送录音指令，所述会议记录请求携带了会议主题；

获取所述终端根据所述录音指令上传的语音数据；

对所述语音数据进行特征提取，得到多个语音特征信息；

将所述多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据所述声纹标识将所述多个语音片段转换为对应的文本信息；

将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信

息，利用所述关键字和校正信息生成分析结果；

根据所述分析结果中的校正信息对所述文本信息进行校正，对校正后的文本信息添加对应的关键字；及

根据所述会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

接收终端发送的会议记录请求，根据所述会议记录请求向所述终端发送录音指令，所述会议记录请求携带了会议主题；

获取所述终端根据所述录音指令上传的语音数据；

对所述语音数据进行特征提取，得到多个语音特征信息；

将所述多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据所述声纹标识将所述多个语音片段转换为对应的文本信息；

将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信息，利用所述关键字和校正信息生成分析结果；

根据所述分析结果中的校正信息对所述文本信息进行校正，对校正后的文本信息添加对应的关键字；及

根据所述会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

下面将对实施例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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为根据一个或多个实施例中会议语音数据处理方法的应用场景图。

图 2 为根据一个或多个实施例中会议语音数据处理方法的流程示意图。

图 3 为根据一个或多个实施例中通过声纹识别模型对语音特征信息进行声纹识别步骤的流程示意图。

图 4 为根据一个或多个实施例中通过语义分析模型对文本信息进行分析步骤的流程示意图。

图 5 为根据一个或多个实施例中会议语音数据处理装置的框图。

图 6 为根据一个或多个实施例中计算机设备的框图。

具体实施方式

为了使本申请的技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请提供的会议语音数据处理方法，可以应用于如图 1 所示的应用环境中。终端 102 通过网络与服务器 104 进行通信。终端 102 可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备，服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群来实现。用户可以通过对应的终端 102 向服务器 104 发送会议记录请求，以对会议过程中的语音进行记录，会议记录请求携带了会议主题。服务器 104 接收到终端 102 发送的会议记录请求后，根据会议记录请求向终端 102 发送录音指令，终端 102 则根据录音指令录制会议过程中的语音数据并上传。服务器 104 则获取终端 102 根据录音指令上传的语音数据，对语音数据进行特征提取，得到多个语音特征信息后，获取预设的声纹识别模型，通过声纹识别模型对所述语音特征信息进行声纹识别，由此能够有效得到多个语音片段和有效识别出对应的声纹标识，服务器 104 进而根据声纹标识将多个语音片段转换为对应的文本信息。服务器 104 进一步获取预设的语义分析模型，通过语义分析模型对文本信息进行上下文语义分析，由此能够准确有效地分析出文本信息中的关键字和校正信息，并根据校正信息对文本信息进行校正，并对校正后的文本信息添加对应的关键字，进而根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

在其中一个实施例中，如图 2 所示，提供了一种会议语音数据处理方法，以该方法应用于图 1 中的服务器为例进行说明，包括以下步骤：

步骤 202，接收终端发送的会议记录请求，根据会议记录请求向终端发送录音指令，会议记录请求携带了会议主题。

步骤 204，获取终端根据录音指令上传的语音数据。

用户可以预先在应用程序上进行注册，并采集每个用户的语音进行声纹识别以验证身份。服务器则利用多个用户注册后的声纹信息和用户标识生成声纹模型库。用户在召开会议的时候，可以通过终端记录会议过程中的会议语音信息。终端可以向服务器发起会议记录请求，会议记录请求中携带了会议关键字。终端可以为一个，也可以为多个。服务器接收到终

端发送的会议记录请求后，则向终端发送录音指令，终端进而根据录音指令进行录音，并将录制的语音数据实时上传至服务器。

步骤 206，对语音数据进行特征提取，得到多个语音特征信息。

服务器接收到终端上传的语音数据后，对语音信号进行预处理，例如服务器可以获取语音数据中的语音信号，并对语音数据中的语音信号进行噪声抑制等预处理，得到预处理后的语音信号。服务器进一步对预处理后的语音信号数据进行特征提取，并对特征提取后的语音信号进行语音端点检测，服务器则根据语音端点将语音数据切分为多个语音特征信息。

步骤 208，将多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据声纹标识将多个语音片段转换为对应的文本信息。

服务器进一步获取预设的声纹识别模型，将预处理后的语音信号数据输入至预先训练的声纹识别模型中，通过声纹识别模型计算多个语音特征信息对应的特征参数序列，根据特征参数序列的相似度将语音信号切分为多个语音片段，服务器并在声纹模型库中进行匹配，得到对应匹配的声纹标识，由此能够利用声纹识别模型识别得到多个语音片段和对应的声纹标识。服务器进而根据识别后的声纹标识将语音信号转换为对应的文本信息。

步骤 210，将文本信息输入至已训练的语义分析模型，分析文本信息中的关键字和校正信息，利用关键字和校正信息生成分析结果。

服务器将语音信号转换为对应的文本信息后，进一步获取预设的语义分析模型，语义分析模型可以是预先利用大量语料数据训练得到的语义分析模型。服务器将识别出的文本信息输入至已训练的语义分析模型中，通过语义分析模型对识别出的文本信息根据上下文语义进行分析，分析出存在歧义或不清楚的文本信息，进而分析出校正信息。校正信息可以包括有误文本、替换文本和对应的文本位置。服务器同时还通过语义分析模型对文本信息进行上下文分析，并识别出文本信息中出现频率较高的关键字，得到分析出的关键字，进而得到包含关键字和校正信息的分析结果。

步骤 212，根据分析结果中的校正信息对文本信息进行校正，对校正后的文本信息添加对应的关键字。

步骤 214，根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

服务器通过语义分析模型对文本信息进行分析并得到分析结果后，根据校正信息对文本信息的上下文存在歧义或不清楚文本进行调整和校正。服务器并根据分析出的关键字，在文本信息对应的位置添加关键字。服务器还可以根据关键字对应的预设字符对文本信息进行调整。

整，得到归纳后的文本信息。

当服务器将会议过程中的所有语音数据转换为对应的文本信息后，服务器则进一步根据会议主题获取预设的会议报告模板，并根据会议报告目标和添加关键字后的文本信息按照预设方式生成对应的会议报告数据。服务器通过对会议中的语音数据进行语音识别和声纹识别，能够准确有效地识别出会议中的各个发言人和对应的语音，服务器根据用户标识将语音数据转换为对应的文本信息，并按照预设方式将文本信息生成对应的会议报告数据，由此能够有效地生成对应的会议报告数据。

上述会议语音数据处理方法中，服务器接收到终端上传的语音数据后，对语音数据进行特征提取，得到多个语音特征信息后，获取预设的声纹识别模型，通过声纹识别模型对所述语音特征信息进行声纹识别，由此能够有效得到多个语音片段和有效识别出对应的声纹标识，服务器进而根据声纹标识将多个语音片段转换为对应的文本信息。服务器进一步获取预设的语义分析模型，通过语义分析模型对文本信息进行上下文语义分析，由此能够准确有效地分析出文本信息中的关键字和校正信息，并根据校正信息对文本信息进行校正，并对校正后的文本信息添加对应的关键字，进而根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据，由此能够有效地生成对应的会议报告数据，进而能够有效地提高会议语音数据的处理效率和识别准确率。

在其中一个实施例中，对语音数据进行特征提取，得到多个语音特征信息的步骤包括：获取语音数据的语音信号，对语音信号进行分帧加窗，提取出对应的声学特征和频谱特征；对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；将声学特征向量和频谱特征向量输入至已训练的语音端点检测模型，通过语音端点检测模型检测语音信号的多个起始点和终止点；根据语音信号的多个起始点和终止点将语音数据切分为多个语音特征信息。

服务器接收到终端上传的语音数据后，对语音数据进行声学特征提取。具体地，服务器提取语音数据中的语音信号，终端上传的语音数据中的语音信号通常为带有噪声的带噪语音信号。服务器获取语音信号后，对语音信号进行加窗分帧，提取出对应的声学特征和频谱特征，并对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量。

服务器进一步获取预设的语音端点检测模型，语音端点检测模型可以是预先已训练得到的模型。服务器将声学特征向量和频谱特征向量输入至语音端点检测模型，通过语音端点检测模型对输入的声学特征向量和频谱特征向量进行分类，可以得到声学特征向量和频谱特征向量对应的决策值。当得到的决策值为预设的第一阈值时，对声学特征向量或频谱特征向量

添加语音标签。其中，第一阈值可以是一个范围值。当得到的决策值为预设的第二阈值时，对声学特征向量或频谱特征向量添加非语音标签。进而可以得到添加语音标签的声学特征向量和添加语音标签的频谱特征向量。对添加语音标签的声学特征向量和添加语音标签的频谱特征向量进行解析，得到添加语音标签的语音信号，根据语音信号的时序和添加语音标签确定语音信号的多个起始点和终止点，进而根据语音信号的多个起始点和终止点将语音数据切分为多个语音特征信息。通过利用语音端点检测模型对语音信号进行端点检测和分类，从而能够准确地识别出带噪语音信号中语音信号和非语音信号，进而能够有效提取出语音数据中的语音特征信息。

在其中一个实施例中，该方法还包括：获取多个语音样本数据，语音样本数据包括已标注样本数据和未标注样本数据，利用已标注样本数据生成训练集，利用未标注样本数据生成验证集；将训练集中的语音样本数据输入至预设的声纹识别模型中进行训练，得到初始声纹识别模型；将验证集中的语音样本数据输入至初始声纹识别模型中进行训练和验证；及直到验证集中的满足预设匹配度值的语音样本数据的数量达到目标阈值时，停止训练，得到训练完成的声纹识别模型。

服务器在获取预设的声纹识别模型之前，需要预先构建出声纹识别模型。具体地，服务器可以首先获取大量的语音样本数据，语音样本数据包括已标注样本数据和未标注样本数据，已标注样本数据为已经预先标注声纹标识的语音样本数据。服务器则将语音样本数据分为训练集和验证集，具体地，服务器利用已标注样本数据生成训练集，训练集中则为已经标注的语音样本数据，服务器并利用未标注样本数据生成验证集。服务器则将训练集中的语音样本数据输入至预设的声纹识别模型中进行训练，得到训练后的初始声纹识别模型，服务器进而将验证集中的语音样本数据输入至初始的声纹识别模型中进行持续训练和验证。直到验证集中的满足预设匹配度值的语音样本数据的数量达到预设阈值时，则停止训练，进而得到训练完成的声纹识别模型。服务器进一步将声纹模型库中用户的声纹输入至训练完成的声纹识别模型中，由此能够有效地构建出识别准确率较高的声纹识别模型。

在其中一个实施例中，如图3所示，通过声纹识别模型对语音特征信息进行声纹识别的步骤，具体包括以下内容：

步骤302，通过声纹识别模型计算多个语音特征信息中的特征参数序列。

步骤304，根据特征参数序列对语音特征信息进行切分，得到多个语音片段。

步骤306，计算多个语音片段的特征参数序列的相似度，将相似度达到预设阈值的多个语音片段进行归类。

步骤 308，根据归类后的特征参数序列在声纹模型库中进行匹配，对归类后的语音片段添加相匹配的声纹标识。

服务器对语音信号数据进行特征提取得到多个语音特征信息后，则获取预设的声纹识别模型，声纹识别模型可以是服务器预先利用大量语音样本数据进行训练得到的模型。服务器将多个语音特征信息输入至声纹识别模型中，通过声纹识别模型对语音特征信息进行声纹识别。具体地，服务器通过声纹识别模型计算出每个语音特征信息中的特征参数序列，每个语音特征序列中可能包括不同特征参数序列的语音片段。服务器则根据不同的特征参数序列将语音特征信息切分为多个语音片段，并将特征参数序列相同的语音片段进行归类，例如，可以将特征参数序列相同的多个语音片段归为同一类。

服务器进而根据归类后的特征参数序列在声纹模型库中进行匹配，声纹模型库中预先存储了多个发言人对应的声纹特征和对应的声纹标识。服务器通过将多个语音片段的特征参数序列与声纹模型库中的声纹特征进行匹配，提取出匹配度最高的特征参数序列对应的声纹标识，进而能够有效地匹配出每个语音片段所对应的声纹标识，服务器则对每一个语音片段添加识别出的对应的声纹标识，由此能够有效地识别出语音数据中各个发言人对应的语音片段信息。

通过声纹识别模型计算出每个语音特征片段对应的特征参数序列，并将特征参数序列与预设的声纹模型库中的特征参数序列进行比对，提取出匹配度最高的特征参数序列的用户对应的声纹标识，并将用户标识添加至对应的语音特征片段中，由此能够准确有效地识别出会议中的各个发言人和对应的语音。

服务器将多个语音片段转换为对应的文本信息后，进一步获取预设的语义分析模型，通过语义分析模型对文本信息进行上下文语义分析，得到包括关键字和校正信息的分析结果。服务器则根据校正信息对文本信息进行校正，并对校正后的文本信息添加对应的关键字，进而根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。由此能够有效地生成对应的会议报告数据，进而能够有效地提高会议语音数据的处理效率和识别准确率。

在其中一个实施例中，如图 4 所示，通过语义分析模型对文本信息进行上下文语义分析的步骤，具体包括以下内容：

步骤 402，通过语义分析模型对所述文本信息进行上下文分析，得到多个文本的语义和词频。

步骤 404，根据多个文本的语义和词频识别出文本信息中的关键字。

步骤 406，根据多个文本的语义和关键字分析文本信息中的待校正文本和对应的文本位置，根据多个文本的语义确定待校正文本对应的校正文本。

步骤 408，根据待校正文本和校正文本以及对应的文本位置生成校正信息，利用关键字和校正信息生成文本信息的分析结果。

服务器接收到终端上传的语音数据后，对语音数据进行特征提取，得到多个语音特征信息后，获取预设的声纹识别模型，通过声纹识别模型对所述语音特征信息进行声纹识别，由此能够有效得到多个语音片段和有效识别出对应的声纹标识，服务器进而根据声纹标识将多个语音片段转换为对应的文本信息。

服务器得到语音数据对应的文本信息后，进一步获取预设的语义分析模型，通过语义分析模型对文本信息进行上下文语义分析。具体地，语义分析模型可以是基于神经网络的机器学习模型。服务器通过语义分析模型对文本信息进行上下文分析，分析多个文本的语义，并识别出文本信息中出现频率较高的文本，根据多个文本的语义和词频得到分析出的关键字。其中，词频可以是某个词或短语在文本信息中出现的频率。

服务器进一步通过语义分析模型根据多个文本的语义和关键字分析出存在歧义或不清楚的文本信息，进而分析出文本信息中的待校正文本和对应的文本位置，根据多个文本的语义确定待校正文本对应的校正文本。服务器根据待校正文本和校正文本以及对应的文本位置生成校正信息。服务器则利用关键字和校正信息对文本信息进行校正，并对校正后的文本信息添加对应的关键字，进而根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据，由此能够有效地生成对应的会议报告数据，进而能够有效地提高会议语音数据的处理效率和识别准确率。

在其中一个实施例中，根据校正信息对文本信息进行校正的步骤包括：根据校正信息确定文本信息中的待校正位置；利用校正信息中的校正文本对待校正位置的待校正文本进行替换，并删除待校正文本，得到校正后的文本信息。

服务器将语音数据转换为对应的文本信息后，进一步获取预设的语义分析模型，通过语义分析模型对文本信息进行上下文语义分析。具体地，语义分析模型可以是基于神经网络的机器学习模型。服务器通过语义分析模型对文本信息进行上下文分析，分析多个文本的语义，并识别出文本信息中出现频率较高的文本，根据多个文本的语义和词频得到分析出的关键字。

服务器进一步通过语义分析模型根据多个文本的语义和关键字分析出存在歧义或不清楚的文本信息，进而分析出文本信息中的校正信息，校正信息中可以包括待校正文本和校正文本以及对应的文本位置。服务器则利用关键字和校正信息对文本信息进行校正，具体地，服

务器根据校正信息确定文本信息中的待校正位置，利用校正信息中的校正文本对待校正位置的待校正文本进行替换，并删除待校正文本，从而将待校正文本修改为校正文本，进而得到校正后的文本信息。

服务器并对校正后的文本信息添加对应的关键字，具体地，服务器根据识别出的关键字，将关键字添加至关键字所对应的文本区域。服务器进而根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据，由此能够有效地生成对应的会议报告数据。通过语义分析模型分析多个文本的语义和词频，由此分析出关键字和校正信息，并根据关键字和校正信息对文本信息进行调整和校正，进而能够有效地提高会议语音数据的识别准确率。

在其中一个实施例中，该方法还包括：接收到终端发送的查询请求，查询请求携带关键字；根据关键字获取与关键字相关联的会议文本内容；将文本内容按照预设方式发送至终端，并进行显示。

服务器将会议过程中的语音数据生成对应的会议报告数据后，将会以报告数据进行存储。用户可以利用关键字查询会议报告数据中对应的会议文本内容。具体地，用户可以通过对应的用户终端向服务器发送查询请求，查询请求中携带了会议主题和关键字。其中，关键字也可以包括用户标识，用户标识与声纹标识可以相一致。服务器接收到用户终端发送的查询请求后，根据会议主题和关键字从数据库中获取会议报告数据中与该关键字相关联的会议文本内容。并将会议文本内容按照预设方式发送至用户终端，例如可以对文本内容进行突出显示。由此可以有效地使用户快速便捷地了解到用户所需的会议内容。

应该理解的是，虽然图 2-4 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 2-4 中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

在其中一个实施例中，如图 5 所示，提供了一种会议语音数据处理装置，包括：数据获取模块 502、特征提取模块 504、声纹识别模块 506、语义分析模块 508、文本校正模块 510 和会议报告生成模块 512，其中：

数据获取模块 502，用于接收终端发送的会议记录请求，根据会议记录请求向终端发送录音指令，会议记录请求携带了会议主题；获取终端根据录音指令上传的语音数据；

特征提取模块 504，用于对语音数据进行特征提取，得到多个语音特征信息；

声纹识别模块 506，用于将多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据声纹标识将多个语音片段转换为对应的文本信息；

语义分析模块 508，用于将所述文本信息输入至已训练的语义分析模型，分析文本信息中的关键字和校正信息，利用关键字和校正信息生成分析结果；

文本校正模块 510，用于根据分析结果中的校正信息对文本信息进行校正，对校正后的文本信息添加对应的关键字；

会议报告生成模块 512，用于根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

在其中一个实施例中，特征提取模块 504 还用于获取语音数据的语音信号，对语音信号进行分帧加窗，提取出对应的声学特征和频谱特征；对声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；将声学特征向量和频谱特征向量输入至已训练的语音端点检测模型，通过语音端点检测模型检测语音信号的多个起始点和终止点；根据语音信号的多个起始点和终止点将语音数据切分为多个语音特征信息。

在其中一个实施例中，该装置还包括声纹识别模型训练模块，用于获取多个语音样本数据，语音样本数据包括已标注样本数据和未标注样本数据，利用已标注样本数据生成训练集，利用未标注样本数据生成验证集；将训练集中的语音样本数据输入至预设的声纹识别模型中进行训练，得到初始声纹识别模型；将验证集中的语音样本数据输入至初始声纹识别模型中进行训练和验证；及直到验证集中的满足预设匹配度值的语音样本数据的数量达到目标阈值时，停止训练，得到训练完成的声纹识别模型。

在其中一个实施例中，声纹识别模块 506 还用于通过声纹识别模型计算多个语音特征信息中的特征参数序列；根据特征参数序列对语音特征信息进行切分，得到多个语音片段；计算多个语音片段的特征参数序列的相似度，将相似度达到预设阈值的多个语音片段进行归类；根据归类后的特征参数序列在声纹模型库中进行匹配，对归类后的语音片段添加相匹配的声纹标识。

在其中一个实施例中，语义分析模块 508 还用于通过语义分析模型对文本信息进行上下文分析，得到多个文本的语义和词频；根据多个文本的语义和词频识别出文本信息中的关键字；根据多个文本的语义和关键字分析文本信息中的待校正文本和对应的文本位置，根据多个文本的语义确定待校正文本对应的校正文本；及根据待校正文本和校正文本以及对应的文本位置生成校正信息，利用关键字和校正信息生成文本信息的分析结果。

在其中一个实施例中，文本校正模块 510 还用于根据校正信息确定文本信息中的待校正位置；及利用校正信息中的校正文本对待校正位置的待校正文本进行替换，并删除待校正文本，得到校正后的文本信息。

在其中一个实施例中，该装置还包括查询模块，用于接收到终端发送的查询请求，查询请求携带会议主题和关键字；根据会议主题和关键字获取与关键字相关联的会议文本内容；将文本内容按照预设方式发送至所述终端，并按照预设方式进行显示。

关于会议语音数据处理装置的具体限定可以参见上文中对于会议语音数据处理方法的限定，在此不再赘述。上述会议语音数据处理置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 6 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储语音数据、语音特征信息、声纹模型库、文本信息以及会议报告数据等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现本申请任意一个实施例中提供的会议语音数据处理方法的步骤。

本领域技术人员可以理解，图 6 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

一种计算机设备，包括存储器及一个或多个处理器，存储器中储存有计算机可读指令，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

接收终端发送的会议记录请求，根据会议记录请求向终端发送录音指令，会议记录请求携带了会议主题；

获取终端根据录音指令上传的语音数据；

对语音数据进行特征提取，得到多个语音特征信息；

将多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据声纹标识将多个语音片段转换为对应的文本信息；

将文本信息输入至已训练的语义分析模型，分析文本信息中的关键字和校正信息，利用关键字和校正信息生成分析结果；

根据分析结果中的校正信息对文本信息进行校正，对校正后的文本信息添加对应的关键字；及

根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

接收终端发送的会议记录请求，根据会议记录请求向终端发送录音指令，会议记录请求携带了会议主题；

获取终端根据录音指令上传的语音数据；

对语音数据进行特征提取，得到多个语音特征信息；

将多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据声纹标识将多个语音片段转换为对应的文本信息；

将文本信息输入至已训练的语义分析模型，分析文本信息中的关键字和校正信息，利用关键字和校正信息生成分析结果；

根据分析结果中的校正信息对文本信息进行校正，对校正后的文本信息添加对应的关键字；及

根据会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程ROM（PROM）、电可编程ROM（EPROM）、电可擦除可编程ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM（SRAM）、动态RAM（DRAM）、同步DRAM（SDRAM）、双数据率SDRAM（DDRSDRAM）、增强型SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接RAM（RDRAM）、直接存储器总线动态RAM（DRDRAM）、以及存储器总线动态RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各

个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种会议语音数据处理方法，所述方法包括：

接收终端发送的会议记录请求，根据所述会议记录请求向所述终端发送录音指令，所述会议记录请求携带了会议主题；

获取所述终端根据所述录音指令上传的语音数据；

对所述语音数据进行特征提取，得到多个语音特征信息；

将所述多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据所述声纹标识将所述多个语音片段转换为对应的文本信息；

将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信息，利用所述关键字和校正信息生成分析结果；

根据所述分析结果中的校正信息对所述文本信息进行校正，对校正后的文本信息添加对应的关键字；及

根据所述会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

2、根据权利要求1所述的方法，其特征在于，所述对所述语音数据进行特征提取，得到多个语音特征信息的步骤包括：

获取所述语音数据的语音信号，对所述语音信号进行分帧加窗，提取出对应的声学特征和频谱特征；

对所述声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；

将所述声学特征向量和频谱特征向量输入至已训练的语音端点检测模型，通过所述语音端点检测模型检测所述语音信号的多个起始点和终止点；及

根据所述语音信号的多个起始点和终止点将所述语音数据切分为多个语音特征信息。

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：

获取多个语音样本数据，所述语音样本数据包括已标注样本数据和未标注样本数据，利用所述已标注样本数据生成训练集，利用所述未标注样本数据生成验证集；

将训练集中的语音样本数据输入至预设的声纹识别模型中进行训练，得到初始声纹识别模型；

将所述验证集中的语音样本数据输入至所述初始声纹识别模型中进行训练和验证；及

直到所述验证集中的满足预设匹配度值的语音样本数据的数量达到目标阈值时，停止训练，得到训练完成的声纹识别模型。

4、根据权利要求1所述的方法，其特征在于，所述将所述多个语音特征信息输入至已训

训练的声纹识别模型中进行声纹识别的步骤包括：

通过所述声纹识别模型计算多个语音特征信息中的特征参数序列；

根据所述特征参数序列对所述语音特征信息进行切分，得到多个语音片段；

计算所述多个语音片段的特征参数序列的相似度，将所述相似度达到预设阈值的多个语音片段进行归类；及

根据归类后的特征参数序列在声纹模型库中进行匹配，对归类后的语音片段添加相匹配的声纹标识。

5、根据权利要求1所述的方法，其特征在于，所述将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信息的步骤包括：

通过所述语义分析模型对所述文本信息进行上下文分析，得到多个文本的语义和词频；

根据所述多个文本的语义和词频识别出文本信息中的关键字；

根据所述多个文本的语义和关键字分析文本信息中的待校正文本和对应的文本位置，根据所述多个文本的语义确定待校正文本对应的校正文本；及

根据所述待校正文本和校正文本以及对应的文本位置生成校正信息，利用所述关键字和所述校正信息生成文本信息的分析结果。

6、根据权利要求5所述的方法，其特征在于，所述根据所述校正信息对所述文本信息进行校正的步骤包括：

根据所述校正信息确定所述文本信息中的待校正位置；及

利用所述校正信息中的校正文本对所述待校正位置的待校正文本进行替换，并删除所述待校正文本，得到校正后的文本信息。

7、根据权利要求1至6任意一项所述的方法，其特征在于，所述方法还包括：

接收到终端发送的查询请求，所述查询请求携带会议主题和关键字；

根据所述会议主题和关键字获取与所述关键字相关联的会议文本内容；及

将所述文本内容按照预设方式发送至所述终端，按照预设方式进行显示。

8、一种会议语音数据处理装置，所述装置包括：

请求接收模块，用于接收终端发送的会议记录请求，根据所述会议记录请求向所述终端发送录音指令，所述会议记录请求携带了会议主题；

数据获取模块，用于获取所述终端根据所述录音指令上传的语音数据；

特征提取模块，用于对所述语音数据进行特征提取，得到多个语音特征信息；

声纹识别模块，用于将所述多个语音特征信息输入至已训练的声纹识别模型中进行声纹

识别，得到多个语音片段和对应的声纹标识，根据所述声纹标识将所述多个语音片段转换为对应的文本信息；

语义分析模块，用于将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信息，利用所述关键字和校正信息生成分析结果；

文本校正模块，用于根据所述分析结果中的校正信息对所述文本信息进行校正，并对校正后的文本信息添加对应的关键字；及

会议报告生成模块，用于根据所述会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

9、根据权利要求 8 所述的装置，其特征在于，所述特征提取模块还用于获取所述语音数据的语音信号，对所述语音信号进行分帧加窗，提取出对应的声学特征和频谱特征；对所述声学特征和频谱特征进行转换，得到对应的声学特征向量和频谱特征向量；将所述声学特征向量和频谱特征向量输入至已训练的语音端点检测模型，通过所述语音端点检测模型检测所述语音信号的多个起始点和终止点；及根据所述语音信号的多个起始点和终止点将所述语音数据切分为多个语音特征信息。

10、根据权利要求 8 所述的装置，其特征在于，所述装置还包括声纹识别模型训练模块，用于获取多个语音样本数据，所述语音样本数据包括已标注样本数据和未标注样本数据，利用所述已标注样本数据生成训练集，利用所述未标注样本数据生成验证集；将训练集中的语音样本数据输入至预设的声纹识别模型中进行训练，得到初始声纹识别模型；将所述验证集中的语音样本数据输入至所述初始声纹识别模型中进行训练和验证；及直到所述验证集中的满足预设匹配度值的语音样本数据的数量达到目标阈值时，停止训练，得到训练完成的声纹识别模型。

11、根据权利要求 8 所述的装置，其特征在于，所述声纹识别模块还用于通过所述声纹识别模型计算多个语音特征信息中的特征参数序列；根据所述特征参数序列对所述语音特征信息进行切分，得到多个语音片段；计算所述多个语音片段的特征参数序列的相似度，将所述相似度达到预设阈值的多个语音片段进行归类；及根据归类后的特征参数序列在声纹模型库中进行匹配，对归类后的语音片段添加相匹配的声纹标识。

12、根据权利要求 8 所述的装置，其特征在于，所述语义分析模块还用于通过所述语义分析模型对所述文本信息进行上下文分析，得到多个文本的语义和词频；根据所述多个文本的语义和词频识别出文本信息中的关键字；根据所述多个文本的语义和关键字分析文本信息中待校正文本和对应的文本位置，根据所述多个文本的语义确定待校正文本对应的校正文本；

及根据所述待校正文本和校正文本以及对应的文本位置生成校正信息，利用所述关键字和所述校正信息生成文本信息的分析结果。

13、根据权利要求 12 所述的装置，其特征在于，所述语义分析模块还用于根据所述校正信息确定所述文本信息中的待校正位置；及利用所述校正信息中的校正文本对所述待校正位置的待校正文本进行替换，并删除所述待校正文本，得到校正后的文本信息。

14、根据权利要求 8 至 13 任一项所述的装置，其特征在于，所述装置还包括查询模块，用于接收到终端发送的查询请求，所述查询请求携带会议主题和关键字；根据所述会议主题和关键字获取与所述关键字相关联的会议文本内容；及将所述文本内容按照预设方式发送至所述终端，按照预设方式进行显示。

15、一种计算机设备，包括存储器及一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

接收终端发送的会议记录请求，根据所述会议记录请求向所述终端发送录音指令，所述会议记录请求携带了会议主题；

获取所述终端根据所述录音指令上传的语音数据；

对所述语音数据进行特征提取，得到多个语音特征信息；

将所述多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据所述声纹标识将所述多个语音片段转换为对应的文本信息；

将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信息，利用所述关键字和校正信息生成分析结果；

根据所述分析结果中的校正信息对所述文本信息进行校正，对校正后的文本信息添加对应的关键字；及

根据所述会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

16、根据权利要求 15 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时还执行以下步骤：通过所述声纹识别模型计算多个语音特征信息中的特征参数序列；根据所述特征参数序列对所述语音特征信息进行切分，得到多个语音片段；计算所述多个语音片段的特征参数序列的相似度，将所述相似度达到预设阈值的多个语音片段进行归类；及根据归类后的特征参数序列在声纹模型库中进行匹配，对归类后的语音片段添加相匹配的声纹标识。

17、根据权利要求 15 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读

指令时还执行以下步骤：通过所述语义分析模型对所述文本信息进行上下文分析，得到多个文本的语义和词频；根据所述多个文本的语义和词频识别出文本信息中的关键字；根据所述多个文本的语义和关键字分析文本信息中的待校正文本和对应的文本位置，根据所述多个文本的语义确定待校正文本对应的校正文本；及根据所述待校正文本和校正文本以及对应的文本位置生成校正信息，利用所述关键字和所述校正信息生成文本信息的分析结果。

18、一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

接收终端发送的会议记录请求，根据所述会议记录请求向所述终端发送录音指令，所述会议记录请求携带了会议主题；

获取所述终端根据所述录音指令上传的语音数据；

对所述语音数据进行特征提取，得到多个语音特征信息；

将所述多个语音特征信息输入至已训练的声纹识别模型中进行声纹识别，得到多个语音片段和对应的声纹标识，根据所述声纹标识将所述多个语音片段转换为对应的文本信息；

将所述文本信息输入至已训练的语义分析模型，分析所述文本信息中的关键字和校正信息，利用所述关键字和校正信息生成分析结果；

根据所述分析结果中的校正信息对所述文本信息进行校正，对校正后的文本信息添加对应的关键字；及

根据所述会议主题和添加关键字的文本信息按照预设方式生成对应的会议报告数据。

19、根据权利要求 18 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：通过所述声纹识别模型计算多个语音特征信息中的特征参数序列；根据所述特征参数序列对所述语音特征信息进行切分，得到多个语音片段；计算所述多个语音片段的特征参数序列的相似度，将所述相似度达到预设阈值的多个语音片段进行归类；及根据归类后的特征参数序列在声纹模型库中进行匹配，对归类后的语音片段添加相匹配的声纹标识。

20、根据权利要求 18 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：通过所述语义分析模型对所述文本信息进行上下文分析，得到多个文本的语义和词频；根据所述多个文本的语义和词频识别出文本信息中的关键字；根据所述多个文本的语义和关键字分析文本信息中的待校正文本和对应的文本位置，根据所述多个文本的语义确定待校正文本对应的校正文本；及根据所述待校正文本和校正文本以及对应的文本位置生成校正信息，利用所述关键字和所述校正信息生成文本信息的分析结果。

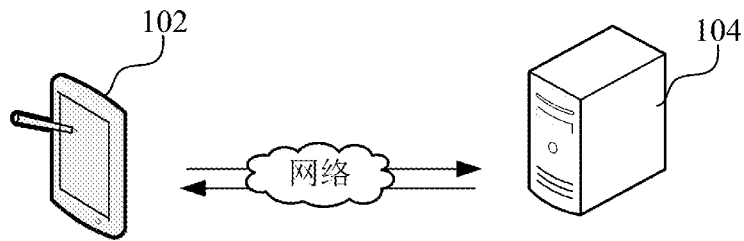


图 1

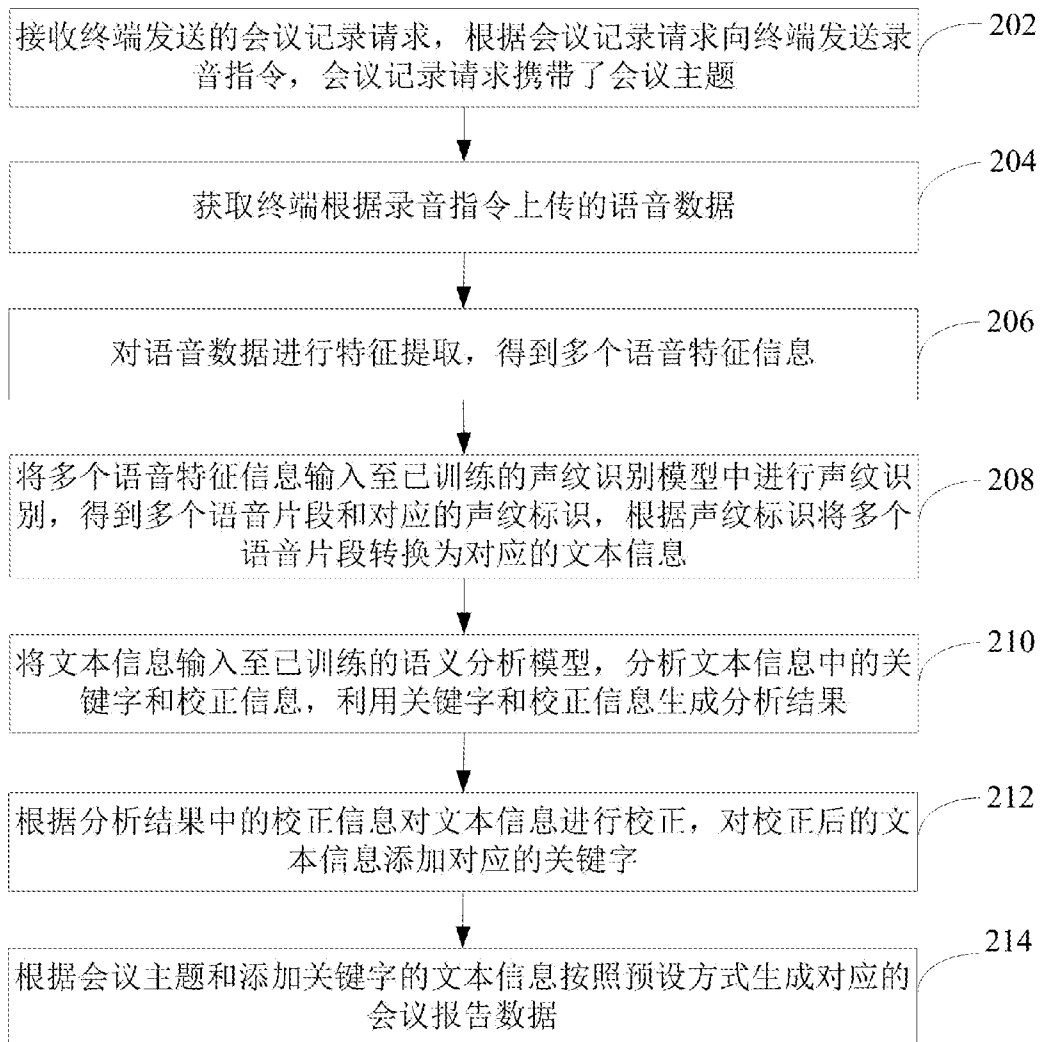


图 2

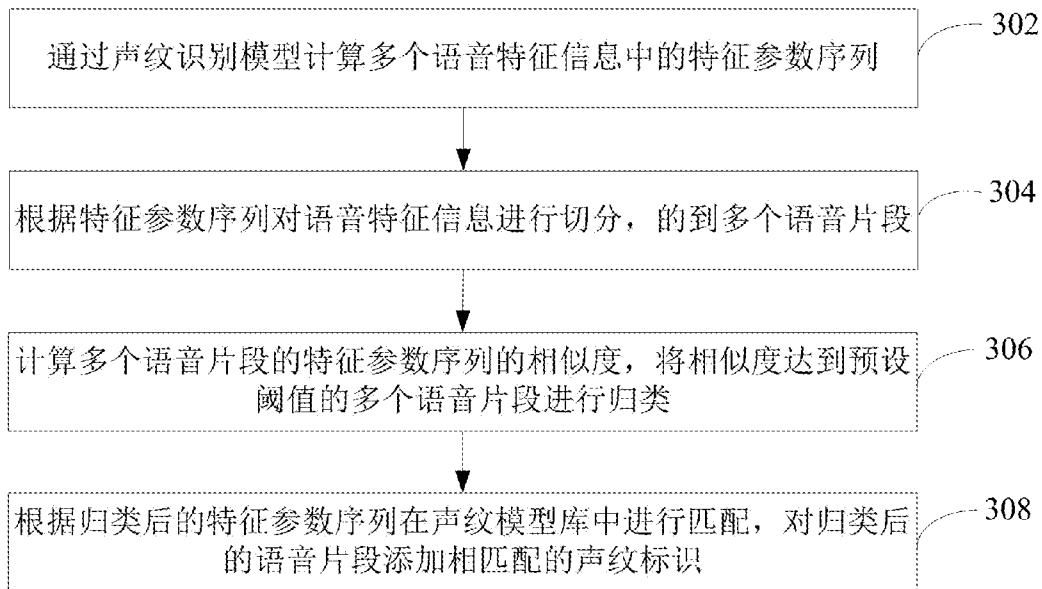


图 3

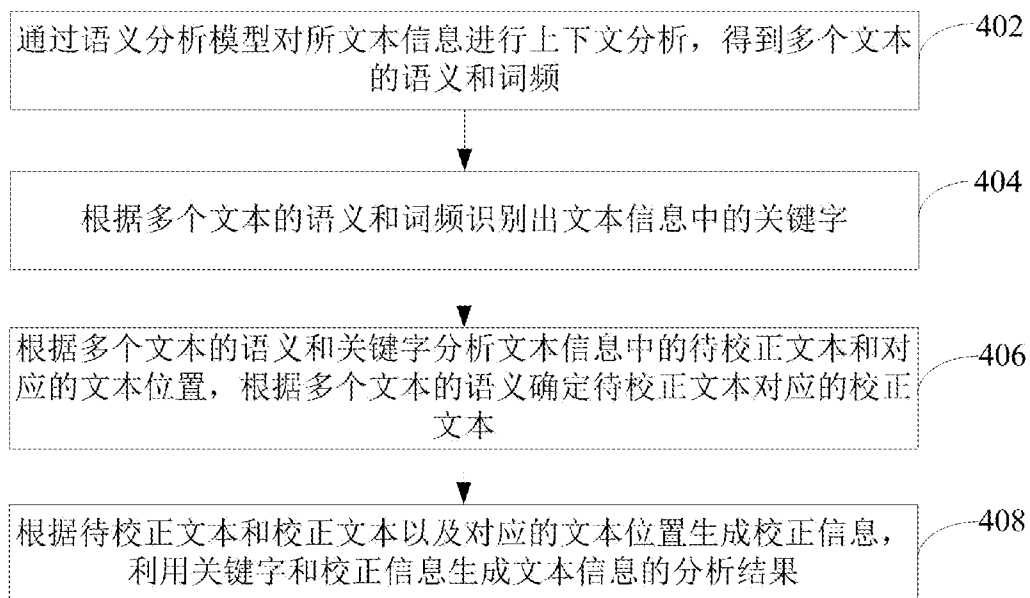


图 4

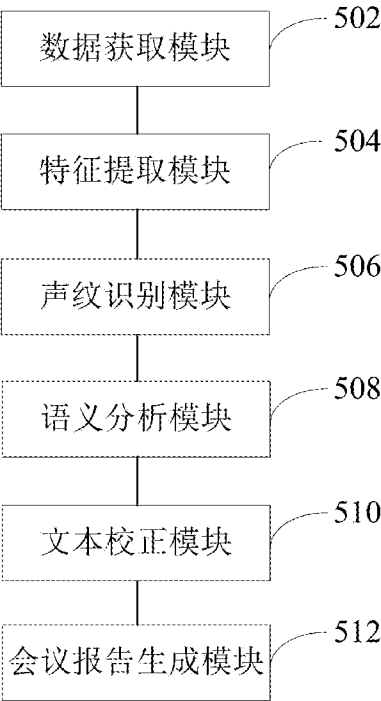


图 5

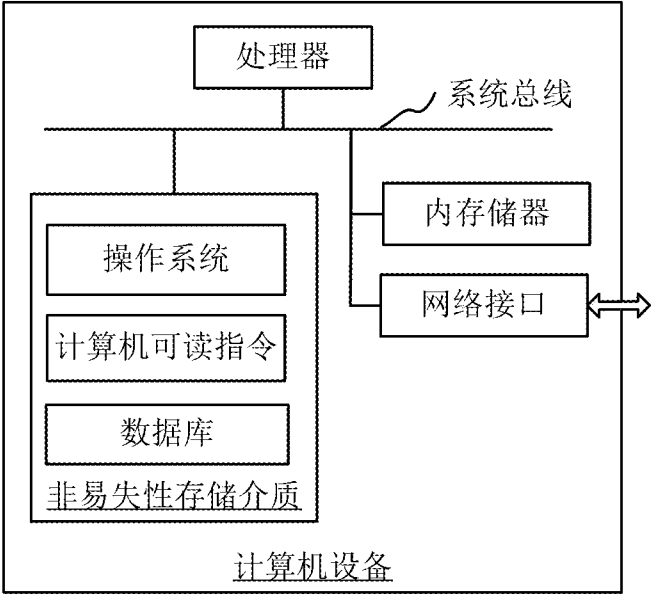


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/118267

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02(2006.01)i; G10L 15/26(2006.01)i; G10L 15/30(2013.01)i; G10L 15/28(2013.01)i; G10L 15/18(2013.01)i;
G10L 15/04(2013.01)i; G10L 15/05(2013.01)i; G10L 17/00(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)

G10L15; G10L17

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)

CNABS; CNTXT; VEN; USTXT; WOTXT; EPTXT; CNKI; IEEE: 语音, 声音, 发言, 讲话, 声纹, 会议, 主题, 记录, 关键字, 校正, 语义, 特征提取, 分段, 端点, 检测, voice, sound, speech, voiceprint, meeting, conference, minutes, theme, key word?, correct+, semantic, feature extract+, section, fragment, endpoint?, detect+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110322872 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 11 October 2019 (2019-10-11) description, paragraphs [0037]-[0097]	1-20
Y	CN 109388701 A (ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) 26 February 2019 (2019-02-26) description, paragraphs [0060]-[0152]	1-20
Y	CN 1264887 A (TSINGHUA UNIVERSITY) 30 August 2000 (2000-08-30) description, page 4, line 1 to page 5, line 4	1-20
A	CN 104252864 A (INTERNATIONAL BUSINESS MACHINES CORPORATION) 31 December 2014 (2014-12-31) entire document	1-20



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

03 February 2020

Date of mailing of the international search report

03 March 2020

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/CN2019/118267

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110322872	A	11 October 2019	None			
CN	109388701	A	26 February 2019	None			
CN	1264887	A	30 August 2000	CN	1141696	C	10 March 2004
CN	104252864	A	31 December 2014	CN	104252864	B	10 April 2018
				US	2015006170	A1	01 January 2015

国际检索报告

国际申请号

PCT/CN2019/118267

A. 主题的分类 G10L 15/02(2006.01)i; G10L 15/26(2006.01)i; G10L 15/30(2013.01)i; G10L 15/28(2013.01)i; G10L 15/18(2013.01)i; G10L 15/04(2013.01)i; G10L 15/05(2013.01)i; G10L 17/00(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) G10L15; G10L17 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)) CNABS;CNTXT;VEN;USTXT;WOTXT;EPTXT;CNKI;IEEE:语音, 声音, 发言, 讲话, 声纹, 会议, 主题, 记录, 关键字, 校正, 语义, 特征提取, 分段, 端点, 检测, voice, sound, speech, voiceprint, meeting, conference, minutes, theme, key word?, correct+, semantic, feature extract+, section, fragment, endpoint?, detect+																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110322872 A (平安科技深圳有限公司) 2019年 10月 11日 (2019 - 10 - 11) 说明书第[0037]-[0097]段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 109388701 A (深圳壹账通智能科技有限公司) 2019年 2月 26日 (2019 - 02 - 26) 说明书第[0060]-[0152]段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 1264887 A (清华大学) 2000年 8月 30日 (2000 - 08 - 30) 说明书第4页第1行-第5页第4行</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 104252864 A (国际商业机器公司) 2014年 12月 31日 (2014 - 12 - 31) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110322872 A (平安科技深圳有限公司) 2019年 10月 11日 (2019 - 10 - 11) 说明书第[0037]-[0097]段	1-20	Y	CN 109388701 A (深圳壹账通智能科技有限公司) 2019年 2月 26日 (2019 - 02 - 26) 说明书第[0060]-[0152]段	1-20	Y	CN 1264887 A (清华大学) 2000年 8月 30日 (2000 - 08 - 30) 说明书第4页第1行-第5页第4行	1-20	A	CN 104252864 A (国际商业机器公司) 2014年 12月 31日 (2014 - 12 - 31) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
PX	CN 110322872 A (平安科技深圳有限公司) 2019年 10月 11日 (2019 - 10 - 11) 说明书第[0037]-[0097]段	1-20															
Y	CN 109388701 A (深圳壹账通智能科技有限公司) 2019年 2月 26日 (2019 - 02 - 26) 说明书第[0060]-[0152]段	1-20															
Y	CN 1264887 A (清华大学) 2000年 8月 30日 (2000 - 08 - 30) 说明书第4页第1行-第5页第4行	1-20															
A	CN 104252864 A (国际商业机器公司) 2014年 12月 31日 (2014 - 12 - 31) 全文	1-20															
☐ 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 2020年 2月 3日		国际检索报告邮寄日期 2020年 3月 3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		受权官员 王莹 电话号码 86-(0512)-88997264															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/118267

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	110322872	A	2019年 10月 11日	无	
CN	109388701	A	2019年 2月 26日	无	
CN	1264887	A	2000年 8月 30日	CN 1141696	C 2004年 3月 10日
CN	104252864	A	2014年 12月 31日	CN 104252864	B 2018年 4月 10日
				US 2015006170	A1 2015年 1月 1日