

(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
2022 年 1 月 27 日 (27.01.2022)



(10) 国际公布号
WO 2022/016994 A1

- (51) 国际专利分类号:
G10L 15/22 (2006.01) *H04L 12/18* (2006.01)
G10L 17/02 (2013.01) *H04N 7/15* (2006.01)
- (21) 国际申请号: PCT/CN2021/096433
- (22) 国际申请日: 2021 年 5 月 27 日 (27.05.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202010718944.8 2020 年 7 月 23 日 (23.07.2020) 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 518000 (CN)。

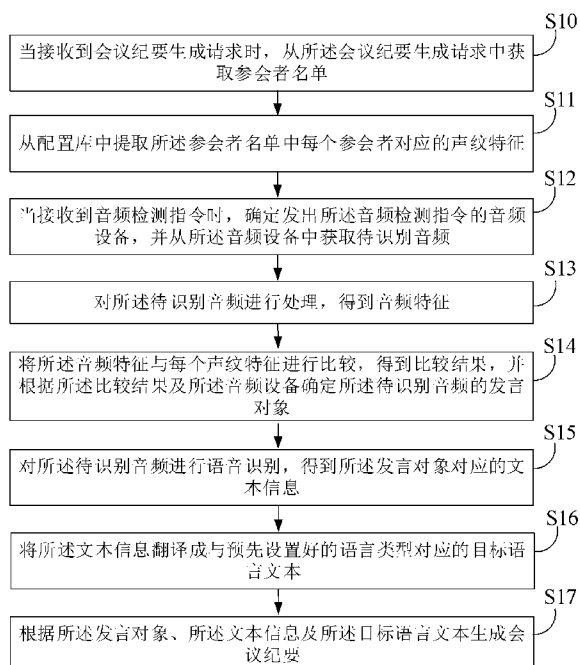
(72) 发明人: 刘璐 (LIU, Lu); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。彭涛 (PENG, Tao); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市赛恩倍吉知识产权代理有限公司 (SHENZHEN SCIENBIZIP INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市龙华新区龙观东路 83 号荣群大厦 9 楼, Guangdong 518109 (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,

(54) Title: AI RECOGNITION-BASED MEETING MINUTES GENERATION METHOD AND APPARATUS, DEVICE AND MEDIUM

(54) 发明名称: 基于 AI 识别的会议纪要生成方法、装置、设备及介质



- S10 When a meeting minutes generation request is received, acquire a list of participants from the meeting minutes generation request
- S11 From within a configuration database, extract voiceprint features corresponding to each participant in the list of participants
- S12 When an audio detection instruction is received, determine an audio device that issues the audio detection instruction, and acquire an audio to be recognized from within the audio device
- S13 Process the audio to obtain audio features
- S14 Compare the audio features with each of the voiceprint features to obtain comparison results, and determine a speaking subject of the audio according to the comparison results and the audio device
- S15 Perform voice recognition on the audio to obtain text information corresponding to the speaking subject
- S16 Translate the text information into target language text corresponding to a preset language type
- S17 Generate meeting minutes according to the speaking subject, the text information and the target language text

图 1

(57) Abstract: Provided are an artificial intelligence (AI) recognition-based meeting minutes generation method and apparatus, an electronic device and a computer-readable storage medium, which relate to the technical field of AI. The method comprises: when a meeting minutes generation request is received, acquiring a list of participants from the meeting minutes generation request (S10); from within a configuration database, extracting voiceprint features corresponding to each participant in the list of participants (S11); when an audio detection instruction is received, determining an audio device that issues the audio detection instruction, and acquiring an

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, 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))。

audio to be recognized from within the audio device (S12); in order to prevent a plurality of participants from speaking at the same time and causing voice overlap, processing the audio to obtain audio features (S13); comparing the audio features to the voiceprint features to obtain comparison results, and performing comprehensive analysis according to the comparison results and binding results of the audio device to accurately determine a speaking subject (S14); performing voice recognition on the audio to obtain text information (S15); translating the text information into target language text corresponding to a preset language type (S16); and generating meeting minutes according to the speaking subject, the text information and the target language text without needing to additionally provide translators (S17). The meeting minutes may also be stored on a blockchain.

(57) 摘要: 提供一种基于AI识别的会议纪要生成方法、装置、电子设备及计算机可读存储介质, 涉及人工智能技术领域。该方法包括: 当接收到会议纪要生成请求时, 从会议纪要生成请求中获取参会者名单 (S10); 从配置库中提取参会者名单中每个参会者对应的声纹特征 (S11); 当接收到音频检测指令时, 确定发出音频检测指令的音频设备, 并从音频设备中获取待识别音频 (S12); 避免多个参会者同时发言而导致语音重叠, 对待识别音频进行处理, 得到音频特征 (S13); 将音频特征与声纹特征进行比较, 得到比较结果, 并根据该比较结果及音频设备的绑定结果进行综合分析, 准确地确定发言对象 (S14); 对待识别音频进行语音识别, 得到文本信息 (S15); 将文本信息翻译成与预先设置好的语言类型对应的目标语言文本 (S16); 无需额外配置翻译人员, 根据发言对象、文本信息及目标语言文本生成会议纪要 (S17)。该会议纪要还可存储于区块链。

基于 AI 识别的会议纪要生成方法、装置、设备及介质

本申请要求于 2020 年 07 月 23 日提交中国专利局，申请号为 202010718944.8，发明名称为“基于 AI 识别的会议纪要生成方法、装置、设备及介质”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

本申请涉及数据处理技术领域，尤其涉及一种基于 AI 识别的会议纪要生成方法、装置、设备及介质。

背景技术

目前的远程会议是指两个或两个以上不同会议场所的个人或者群体通过传输线路及多媒体设备实现声音互传的会议，这种远程会议方式能够降低会议成本。

然而，发明人意识到，当会议中存在多位发言人员时，由于远程会议的局限性，会议记录人员只能根据发言人员的音色主观确定正在发表言论的对象，进而执行会议纪要的记录工作，通过主观确定发言对象的方式容易出现发言对象误判的情况，进而降低会议纪要的准确性。此外，当发言人员说话速度较快时，会议记录人员往往不能完整地记录会议过程中讨论的所有关键信息，导致会议纪要的关键信息的遗漏。另外，由于不同国家使用的语言有所不同，因此，当发言人员使用的语言不能为参会人员理解时，会议场所需要配备不同国家语言的会议记录人员做会议纪要，导致花费大量的人力资源。

发明内容

鉴于以上内容，有必要提供一种基于 AI 识别的会议纪要生成方法、装置、设备及介质，能够在没有额外配置翻译人员做会议纪要的场景下，快速完整地将待识别音频进行转换，避免会议纪要中的关键信息的遗漏，还能提高会议纪要的准确性。

本申请的第一方面提供一种基于 AI 识别的会议纪要生成方法，所述基于 AI 识别的会议纪要生成方法包括：

当接收到会议纪要生成请求时，从所述会议纪要生成请求中获取参会者名单；

从配置库中提取所述参会者名单中每个参会者对应的声纹特征；

当接收到音频检测指令时，确定发出所述音频检测指令的音频设备，并从所述音频设备中获取待识别音频；

对所述待识别音频进行处理，得到音频特征；

将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象；

对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息；

将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本；

根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

本申请的第二方面提供一种电子设备，所述电子设备包括处理器和存储器，所述处理器用于执行所述存储器中存储的计算机可读指令以实现以下步骤：

当接收到会议纪要生成请求时，从所述会议纪要生成请求中获取参会者名单；

从配置库中提取所述参会者名单中每个参会者对应的声纹特征；

当接收到音频检测指令时，确定发出所述音频检测指令的音频设备，并从所述音频设备中获取待识别音频；

对所述待识别音频进行处理，得到音频特征；

将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象；

对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息；

将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本；

根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

本申请的第三方面提供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至少一个计算机可读指令，所述至少一个计算机可读指令被处理器执行以实现以下步骤：

当接收到会议纪要生成请求时，从所述会议纪要生成请求中获取参会者名单；

从配置库中提取所述参会者名单中每个参会者对应的声纹特征；

当接收到音频检测指令时，确定发出所述音频检测指令的音频设备，并从所述音频设备中获取待识别音频；

对所述待识别音频进行处理，得到音频特征；

将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象；

对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息；

将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本；

根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

本申请的第四方面提供一种基于 AI 识别的会议纪要生成装置，所述基于 AI 识别的会议纪要生成装置包括：

获取单元，用于当接收到会议纪要生成请求时，从所述会议纪要生成请求中获取参会者名单；

提取单元，用于从配置库中提取所述参会者名单中每个参会者对应的声纹特征；

确定单元，用于当接收到音频检测指令时，确定发出所述音频检测指令的音频设备，并从所述音频设备中获取待识别音频；

处理单元，用于对所述待识别音频进行处理，得到音频特征；

比较单元，用于将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象；

识别单元，用于对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息；

翻译单元，用于将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本；

生成单元，用于根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

由以上技术方案可以看出，本申请通过从所述音频设备中获取所述待识别音频，能够避免因多个参会者在同一时刻发表言论而导致所述待识别音频存在语音重叠，从而实现语音分离，为后续准确确定发言对象及准确转换文本信息作铺垫，同时，通过对声纹特征的比较结果及音频设备的绑定结果进行综合分析，能够准确地确定发言对象，从而提高会议纪要的准确性，此外，由于全程自动地将发言对象的音频转换为文本信息，而无需人为记录会议过程中所讨论的信息，能够快速完整地将待识别音频进行转换，从而避免会议纪要中的关键信息的遗漏，同时，本申请无需额外配置翻译人员，减少了人力资源的浪费。

附图说明

图 1 是本申请基于 AI 识别的会议纪要生成方法的较佳实施例的流程图。

图 2 是本申请基于 AI 识别的会议纪要生成装置的较佳实施例的功能模块图。

图 3 是本申请实现基于 AI 识别的会议纪要生成方法的较佳实施例的电子设备的结构示意图。

具体实施方式

为了使本申请的目的、技术方案和优点更加清楚，下面结合附图和具体实施例对本申请进行详细描述。

如图 1 所示，是本申请基于 AI 识别的会议纪要生成方法的较佳实施例的流程图。根据不同的需求，该流程图中步骤的顺序可以改变，某些步骤可以省略。

所述基于 AI 识别的会议纪要生成方法应用于一个或者多个电子设备中，所述电子设备是一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的设备，其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit, ASIC)、可编程门阵列(Field-Programmable Gate Array, FPGA)、数字处理器(Digital Signal Processor, DSP)、嵌入式设备等。

所述电子设备可以是任何一种可与用户进行人机交互的电子产品，例如，个人计算机、平板电脑、智能手机、个人数字助理(Personal Digital Assistant, PDA)、游戏机、交互式网络电视(Internet Protocol Television, IPTV)、智能穿戴式设备等。

所述电子设备还可以包括网络设备和/或用户设备。其中，所述网络设备包括，但不限于单个网络服务器、多个网络服务器组成的服务器组或基于云计算(Cloud Computing)的由大量主机或网络服务器构成的云。

所述电子设备所处的网络包括但不限于互联网、广域网、城域网、局域网、虚拟专用网络(Virtual Private Network, VPN)等。

在本申请的至少一个实施例中，本申请应用于人工智能领域，同时，本申请涉及声纹识别。

S10，当接收到会议纪要生成请求时，从所述会议纪要生成请求中获取参会者名单。

在本申请的至少一个实施例中，所述会议纪要生成请求所携带的信息包括，但不限于：所述参会者名单、会议主题等。

在本申请的至少一个实施例中，所述电子设备获取目标标签，进一步地，所述电子设备根据所述目标标签从所述会议纪要生成请求中提取与所述目标标签对应的信息，作为所述参会者名单。

通过上述实施方式，能够准确、快速地确定所述参会者名单。

S11，从配置库中提取所述参会者名单中每个参会者对应的声纹特征。

在本申请的至少一个实施例中，所述配置库中存储多个用户的声纹特征。所述声纹特征包括，但不限于：声学特征、词法特征、韵律特征等。

在本申请的至少一个实施例中，在从配置库中提取所述参会者名单中每个参会者对应的声纹特征之前，所述方法还包括：

所述电子设备获取多个用户的音频信息及所述多个用户中每个用户的身份识别码，进一步地，所述电子设备利用语音信号处理技术，对所述音频信息进行预处理，所述电子设备对预处理后的音频信息进行编码，得到所述音频信息的数字信号，更进一步地，所述电子设备从所述数字信号中提取声纹特征，所述电子设备以所述身份识别码与所述声纹特征建立映射关系并保存至所述配置库中。

通过所述语音信号处理技术，能够在复杂的语音环境中提取有效的语音信息，进而避免生成的数字信号中包括噪音，进而避免对所述声纹特征的干扰。

在本申请的至少一个实施例中，通过从所述配置库中提取所述声纹特征，进而将所述声纹特征与待识别音频进行比较，由于声纹特征不易被破解，因此能够提高安全性。

在本申请的至少一个实施例中，通过从配置库中提取每个参会者对应的声纹特征，避免了对所述配置库中的所有声纹特征进行一一比对，大大缩减了确定发言对象的范围，从而能

够快速准确地确定发言对象。

S12, 当接收到音频检测指令时, 确定发出所述音频检测指令的音频设备, 并从所述音频设备中获取待识别音频。

在本申请的至少一个实施例中, 当检测到任意用户在当前时刻发言时, 所述电子设备触发所述任意用户对应的音频设备生成所述音频检测指令, 所述音频检测指令携带的信息包括, 但不限于: 发出所述音频检测指令的音频设备的参数信息。

进一步地, 所述音频设备可以是麦克风, 也可以是话筒, 本申请对所述音频设备的选取不作限制。所述音频设备具有用于标识设备本身的参数信息, 参数信息可以包括各类软件信息及硬件信息。例如, 所述参数信息可以是媒体访问控制地址(Media Access Control Address, MAC)地址、设备序列号以及 IP 地址等。

在本申请的至少一个实施例中, 所述电子设备确定发出所述音频检测指令的音频设备包括:

所述电子设备解析所述音频检测指令, 得到参数信息, 进一步地, 所述电子设备根据所述参数信息确定所述音频设备。

在本申请的至少一个实施例中, 通过从所述音频设备中获取所述待识别音频, 能够避免因多个参会者在同一时刻发表言论而导致所述待识别音频存在语音重叠, 从而实现语音分离, 为后续准确确定发言对象及准确转换文本信息作铺垫。

S13, 对所述待识别音频进行处理, 得到音频特征。

在本申请的至少一个实施例中, 所述电子设备对所述待识别音频进行处理, 得到音频特征包括:

所述电子设备对所述待识别音频进行分块处理, 得到音频区域集, 进一步地, 所述电子设备提取所述音频区域集中每一帧对应的特征参数, 得到所述音频特征。

S14, 将所述音频特征与每个声纹特征进行比较, 得到比较结果, 并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象。

在本申请的至少一个实施例中, 在创建远程会议时, 所述电子设备预先将参会者与音频设备的参数信息进行绑定, 从而实现所述参会者与所述音频设备的关联关系。

在本申请的至少一个实施例中, 所述电子设备将所述音频特征与每个声纹特征进行比较, 得到比较结果, 并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象包括:

所述电子设备将所述音频特征与每个声纹特征进行比较, 得到比较结果, 所述比较结果为每个声纹特征与所述音频特征的相似度, 当检测到所述相似度中存在任意相似度大于阈值时, 所述电子设备将所述任意相似度中的最大相似度确定为目标相似度, 并将所述目标相似度对应的声纹特征确定为目标声纹特征, 进一步地, 所述电子设备将所述目标声纹特征对应的用户确定为第一用户, 所述电子设备获取与所述音频设备绑定的用户作为第二用户, 当检测到所述第一用户与所述第二用户为同一人时, 所述电子设备将所述第一用户确定为所述发言对象, 或者当检测到所述第一用户与所述第二用户不为同一人时, 所述电子设备启动摄像设备对当前使用所述音频设备的用户进行录像, 得到目标图像, 并将所述目标图像与人脸库中的预设图像进行匹配, 得到每张预设图像对应的用户的匹配度, 每张预设图像对应一个人员, 并获取每个人员的语言特征, 所述电子设备确定每个语言特征与所述音频特征的相似度, 所述电子设备采用预设权重对每个相似度与每个匹配度进行加权和运算, 得到每个人员的分值, 所述电子设备将分值最高的人员确定为所述发言对象。

通过对声纹特征的比较结果及音频设备的绑定结果进行综合分析, 能够准确地确定发言对象, 从而提高会议纪要的准确性。

S15, 对所述待识别音频进行语音识别, 得到所述发言对象对应的文本信息。

在本申请的至少一个实施例中, 所述电子设备对所述待识别音频进行语音识别, 得到所述发言对象对应的文本信息包括:

所述电子设备利用移动窗函数对所述待识别音频按照预设时长进行分割, 得到至少一帧

语音, 所述电子设备对每帧语音进行预处理, 并计算每帧预处理后的语音的能量值, 当检测到预处理后的语音中连续预设帧的能量值小于配置值, 及接下来的语音中连续所述预设帧的能量值大于所述配置值时, 所述电子设备将能量值开始增大的语音帧确定为语音的前端点, 当检测到预处理后的语音中连续所述预设帧的能量值大于所述配置值, 及接下来的语音中连续所述预设帧的能量值小于所述配置值时, 所述电子设备将能量值开始减小的语音帧确定为语音的后端点, 所述电子设备删除所述前端点至所述后端点之外的语音帧, 得到目标音频, 所述电子设备对所述目标音频进行分块处理, 得到目标语音帧, 进一步地, 所述电子设备将所述目标语音帧转换为所述文本信息。

其中, 所述预设时长可以根据需要进行调整, 一般情况下, 所述预设时长设置为 20ms。所述预处理包括去除直流处理和加窗处理。进一步地, 所述预设帧可以根据语音应用场景不同设置为不同的值, 所述预设帧的具体值可以根据实际需求进行调整, 本申请不作限制。更进一步地, 所述配置值可以根据静音时的平均能量值确定, 本申请也不作限制。

通过对所述待识别音频进行处理, 能够对所述待识别音频中的静音部分和噪声部分进行消除, 不仅节省所述待识别音频的存储空间, 还提高所述待识别音频的转换效率, 同时还能准确地将所述待识别音频转换为所述文本信息。

由于全程自动地将发言对象的音频转换为文本信息, 而无需人为记录会议过程中所讨论的信息, 能够快速完整地将待识别音频进行转换, 从而避免会议纪要中的关键信息的遗漏。

S16, 将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本。

在本申请的至少一个实施例中, 所述电子设备根据所述语言类型确定目标翻译模块, 进一步地, 所述电子设备通过所述目标翻译模块将所述文本信息翻译成所述目标语言文本。

例如, 待识别音频 A 的内容是汉语的“我是主持人”, 所述电子设备将其转换成文本信息“我是主持人”。若预先设置好的语言类型为英文类型, 所述电子设备根据所述英文类型确定目标翻译模块为中译英模块, 所述中译英模块将文本信息“我是主持人”翻译成目标语言文本, 即: “I'm the host”。

在本申请的至少一个实施例中, 所述方法还包括:

所述电子设备将所述目标语言文本转换为目标语言音频, 并播放所述目标语言音频。

通过上述实施方式, 当发言人员使用的语言不能为参会人员理解时, 无需额外配置翻译人员, 参会人员即可理解发言人员表述的含义。

S17, 根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

在本申请的至少一个实施例中, 所述音频设备中存储所述待识别音频及所述识别音频对应的发言时间。

在本申请的至少一个实施例中, 所述电子设备根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要包括:

所述电子设备获取预先配置的会议纪要模版, 并获取所述文本信息的发言时间, 所述电子设备依据所述发言时间从先至后的顺序, 将所述发言对象、所述文本信息及所述目标语言文本录入所述会议纪要模板, 得到会议记录, 所述电子设备对所述会议记录进行分词处理, 得到分词, 所述电子设备采用编码算法将所述分词转换为第一词向量, 并将预设词组转换为第二词向量, 所述电子设备利用余弦距离公式, 计算所述第一词向量与所述第二词向量的距离, 所述电子设备将距离值大于预设值的第一词向量确定为目标词向量, 并将所述目标词向量对应的分词确定为目标词组, 所述电子设备在所述会议记录中标识所述目标词组, 得到所述会议纪要。

具体地, 所述电子设备在所述会议记录中标识所述目标词组的方式包括以下一种或者多种方式的组合:

(1) 所述电子设备对所述目标词组进行加粗标识。

(2) 所述电子设备对所述目标词组进行高亮标识。

在本申请的至少一个实施例中, 在根据所述发言对象、所述文本信息及所述目标语言文

本生成会议纪要后，所述方法还包括：

所述电子设备获取所述会议纪要生成请求的会议主题，所述电子设备根据所述会议主题及所述会议纪要生成提示信息，所述电子设备采用对称加密技术加密所述提示信息，得到密文，所述电子设备获取所述参会者名单中每个参会者的终端设备，并将所述密文发送至所述终端设备。

通过加密所述提示信息，能够提高所述提示信息的安全性，避免所述会议纪要被篡改，进而将所述密文发送至每个参会者的终端设备，能够提醒每个参会者及时查收所述会议纪要。

需要说明的是，为了提高安全性及隐私性，生成的会议纪要可存储于区块链。

由以上技术方案可以看出，本申请通过从所述音频设备中获取所述待识别音频，能够避免因多个参会者在同一时刻发表言论而导致所述待识别音频存在语音重叠，从而实现语音分离，为后续准确确定发言对象及准确转换文本信息作铺垫，同时，通过对声纹特征的比较结果及音频设备的绑定结果进行综合分析，能够准确地确定发言对象，从而提高会议纪要的准确性，此外，由于全程自动地将发言对象的音频转换为文本信息，而无需人为记录会议过程中所讨论的信息，能够快速完整地将待识别音频进行转换，从而避免会议纪要中的关键信息的遗漏，同时，本申请无需额外配置翻译人员，减少了人力资源的浪费。

如图2所示，是本申请基于AI识别的会议纪要生成装置的较佳实施例的功能模块图。所述基于AI识别的会议纪要生成装置11包括获取单元110、提取单元111、确定单元112、处理单元113、比较单元114、识别单元115、翻译单元116、生成单元117、预处理单元118、编码单元119、建立单元120、加密单元121及发送单元122。本申请所称的模块/单元是指一种能够被处理器13所获取，并且能够完成固定功能的一系列计算机程序段，其存储在存储器12中。在本实施例中，关于各模块/单元的功能将在后续的实施例中详述。

当接收到会议纪要生成请求时，获取单元110从所述会议纪要生成请求中获取参会者名单。

在本申请的至少一个实施例中，所述会议纪要生成请求所携带的信息包括，但不限于：所述参会者名单、会议主题等。

在本申请的至少一个实施例中，所述获取单元110获取目标标签，进一步地，所述获取单元110根据所述目标标签从所述会议纪要生成请求中提取与所述目标标签对应的信息，作为所述参会者名单。

通过上述实施方式，能够准确、快速地确定所述参会者名单。

提取单元111从配置库中提取所述参会者名单中每个参会者对应的声纹特征。

在本申请的至少一个实施例中，所述配置库中存储多个用户的声纹特征。所述声纹特征包括，但不限于：声学特征、词法特征、韵律特征等。

在本申请的至少一个实施例中，在从配置库中提取所述参会者名单中每个参会者对应的声纹特征之前，所述获取单元110获取多个用户的音频信息及所述多个用户中每个用户的身份识别码，进一步地，预处理单元118利用语音信号处理技术，对所述音频信息进行预处理，编码单元119对预处理后的音频信息进行编码，得到所述音频信息的数字信号，更进一步地，所述提取单元111从所述数字信号中提取声纹特征，建立单元120以所述身份识别码与所述声纹特征建立映射关系并保存至所述配置库中。

通过所述语音信号处理技术，能够在复杂的语音环境中提取有效的语音信息，进而避免生成的数字信号中包括噪音，进而避免对所述声纹特征的干扰。

在本申请的至少一个实施例中，通过从所述配置库中提取所述声纹特征，进而将所述声纹特征与待识别音频进行比较，由于声纹特征不易被破解，因此能够提高安全性。

在本申请的至少一个实施例中，通过从配置库中提取每个参会者对应的声纹特征，避免了对所述配置库中的所有声纹特征进行一一比对，大大缩减了确定发言对象的范围，从而能够快速准确地确定发言对象。

当接收到音频检测指令时，确定单元 112 确定发出所述音频检测指令的音频设备，并从所述音频设备中获取待识别音频。

在本申请的至少一个实施例中，当检测到任意用户在当前时刻发言时，触发所述任意用户对应的音频设备生成所述音频检测指令，所述音频检测指令携带的信息包括，但不限于：发出所述音频检测指令的音频设备的参数信息。

进一步地，所述音频设备可以是麦克风，也可以是话筒，本申请对所述音频设备的选取不作限制。所述音频设备具有用于标识设备本身的参数信息，所述参数信息可以包括各类软件信息及硬件信息。例如，所述参数信息可以是媒体访问控制地址(Media Access Control Address, MAC)地址、设备序列号以及 IP 地址等。

在本申请的至少一个实施例中，所述确定单元 112 确定发出所述音频检测指令的音频设备包括：

所述确定单元 112 解析所述音频检测指令，得到参数信息，进一步地，所述确定单元 112 根据所述参数信息确定所述音频设备。

在本申请的至少一个实施例中，通过从所述音频设备中获取所述待识别音频，能够避免因多个参会者在同一时刻发表言论而导致所述待识别音频存在语音重叠，从而实现语音分离，为后续准确确定发言对象及准确转换文本信息作铺垫。

处理单元 113 对所述待识别音频进行处理，得到音频特征。

在本申请的至少一个实施例中，所述处理单元 113 对所述待识别音频进行处理，得到音频特征包括：

所述处理单元 113 对所述待识别音频进行分块处理，得到音频区域集，进一步地，所述处理单元 113 提取所述音频区域集中每一帧对应的特征参数，得到所述音频特征。

比较单元 114 将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象。

在本申请的至少一个实施例中，在创建远程会议时，预先将参会者与音频设备的参数信息进行绑定，从而实现所述参会者与所述音频设备的关联关系。

在本申请的至少一个实施例中，所述比较单元 114 将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象包括：

所述比较单元 114 将所述音频特征与每个声纹特征进行比较，得到比较结果，所述比较结果为每个声纹特征与所述音频特征的相似度，当检测到所述相似度中存在任意相似度大于阈值时，所述比较单元 114 将所述任意相似度中的最大相似度确定为目标相似度，并将所述目标相似度对应的声纹特征确定为目标声纹特征，进一步地，所述比较单元 114 将所述目标声纹特征对应的用户确定为第一用户，所述比较单元 114 获取与所述音频设备绑定的用户作为第二用户，当检测到所述第一用户与所述第二用户为同一人时，所述比较单元 114 将所述第一用户确定为所述发言对象，或者当检测到所述第一用户与所述第二用户不为同一人时，所述比较单元 114 启动摄像设备对当前使用所述音频设备的用户进行录像，得到目标图像，并将所述目标图像与人脸库中的预设图像进行匹配，得到每张预设图像对应的用户的匹配度，每张预设图像对应一个人员，并获取每个人员的语言特征，所述比较单元 114 确定每个语言特征与所述音频特征的相似度，所述比较单元 114 采用预设权重对每个相似度与每个匹配度进行加权和运算，得到每个人员的分值，所述比较单元 114 将分值最高的人员确定为所述发言对象。

通过对声纹特征的比较结果及音频设备的绑定结果进行综合分析，能够准确地确定发言对象，从而提高会议纪要的准确性。

识别单元 115 对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息。

在本申请的至少一个实施例中，所述识别单元 115 对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息包括：

所述识别单元 115 利用移动窗函数对所述待识别音频按照预设时长进行分割, 得到至少一帧语音, 所述识别单元 115 对每帧语音进行预处理, 并计算每帧预处理后的语音的能量值, 当检测到预处理后的语音中连续预设帧的能量值小于配置值, 及接下来的语音中连续所述预设帧的能量值大于所述配置值时, 所述识别单元 115 将能量值开始增大的语音帧确定为语音的前端点, 当检测到预处理后的语音中连续所述预设帧的能量值大于所述配置值, 及接下来的语音中连续所述预设帧的能量值小于所述配置值时, 所述识别单元 115 将能量值开始减小的语音帧确定为语音的后端点, 所述识别单元 115 删除所述前端点至所述后端点之外的语音帧, 得到目标音频, 所述识别单元 115 对所述目标音频进行分块处理, 得到目标语音帧, 进一步地, 所述识别单元 115 将所述目标语音帧转换为所述文本信息。

其中, 所述预设时长可以根据需要进行调整, 一般情况下, 所述预设时长设置为 20ms。所述预处理包括去除直流处理和加窗处理。进一步地, 所述预设帧可以根据语音应用场景不同设置为不同的值, 所述预设帧的具体值可以根据实际需求进行调整, 本申请不作限制。更进一步地, 所述配置值可以根据静音时的平均能量值确定, 本申请也不作限制。

通过对所述待识别音频进行处理, 能够对所述待识别音频中的静音部分和噪声部分进行消除, 不仅节省所述待识别音频的存储空间, 还提高所述待识别音频的转换效率, 同时还能准确地将所述待识别音频转换为所述文本信息。

由于全程自动地将发言对象的音频转换为文本信息, 而无需人为记录会议过程中所讨论的信息, 能够快速完整地将待识别音频进行转换, 从而避免会议纪要中的关键信息的遗漏。

翻译单元 116 将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本。

在本申请的至少一个实施例中, 所述翻译单元 116 根据所述语言类型确定目标翻译模块, 进一步地, 所述翻译单元 116 通过所述目标翻译模块将所述文本信息翻译成所述目标语言文本。

例如, 待识别音频 A 的内容是汉语的“我是主持人”, 所述翻译单元 116 将其转换成文本信息“我是主持人”。若预先设置好的语言类型为英文类型, 所述翻译单元 116 根据所述英文类型确定目标翻译模块为中译英模块, 所述中译英模块将文本信息“我是主持人”翻译成目标语言文本, 即: “I'm the host”。

在本申请的至少一个实施例中, 将所述目标语言文本转换为目标语言音频, 并播放所述目标语言音频。

通过上述实施方式, 当发言人员使用的语言不能为参会人员理解时, 无需额外配置翻译人员, 参会人员即可理解发言人员表述的含义。

生成单元 117 根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

在本申请的至少一个实施例中, 所述音频设备中存储所述待识别音频及所述识别音频对应的发言时间。

在本申请的至少一个实施例中, 所述生成单元 117 根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要包括:

所述生成单元 117 获取预先配置的会议纪要模版, 并获取所述文本信息的发言时间, 所述生成单元 117 依据所述发言时间从先至后的顺序, 将所述发言对象、所述文本信息及所述目标语言文本录入所述会议纪要模板, 得到会议记录, 所述生成单元 117 对所述会议记录进行分词处理, 得到分词, 所述生成单元 117 采用编码算法将所述分词转换为第一词向量, 并将预设词组转换为第二词向量, 所述生成单元 117 利用余弦距离公式, 计算所述第一词向量与所述第二词向量的距离, 所述生成单元 117 将距离值大于预设值的第一词向量确定为目标词向量, 并将所述目标词向量对应的分词确定为目标词组, 所述生成单元 117 在所述会议记录中标识所述目标词组, 得到所述会议纪要。

具体地, 所述生成单元 117 在所述会议记录中标识所述目标词组的方式包括以下一种或者多种方式的组合:

(1) 所述生成单元 117 对所述目标词组进行加粗标识。

(2) 所述生成单元 117 对所述目标词组进行高亮标识。

在本申请的至少一个实施例中，在根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要后，所述获取单元 110 获取所述会议纪要生成请求的会议主题，所述生成单元 117 根据所述会议主题及所述会议纪要生成提示信息，加密单元 121 采用对称加密技术加密所述提示信息，得到密文，所述获取单元 110 获取所述参会者名单中每个参会者的终端设备，发送单元 122 将所述密文发送至所述终端设备。

通过加密所述提示信息，能够提高所述提示信息的安全性，避免所述会议纪要被篡改，进而将所述密文发送至每个参会者的终端设备，能够提醒每个参会者及时查收所述会议纪要。

需要说明的是，为了提高安全性及隐私性，生成的会议纪要可存储于区块链。

由以上技术方案可以看出，本申请通过从所述音频设备中获取所述待识别音频，能够避免因多个参会者在同一时刻发表言论而导致所述待识别音频存在语音重叠，从而实现语音分离，为后续准确确定发言对象及准确转换文本信息作铺垫，同时，通过对声纹特征的比较结果及音频设备的绑定结果进行综合分析，能够准确地确定发言对象，从而提高会议纪要的准确性，此外，由于全程自动地将发言对象的音频转换为文本信息，而无需人为记录会议过程中所讨论的信息，能够快速完整地将待识别音频进行转换，从而避免会议纪要中的关键信息的遗漏，同时，本申请无需额外配置翻译人员，减少了人力资源的浪费。

如图 3 所示，是本申请实现基于 AI 识别的会议纪要生成方法的较佳实施例的电子设备的结构示意图。

在本申请的一个实施例中，所述电子设备 1 包括，但不限于，存储器 12、处理器 13，以及存储在所述存储器 12 中并可在所述处理器 13 上运行的计算机程序，例如基于 AI 识别的会议纪要生成程序。

本领域技术人员可以理解，所述示意图仅仅是电子设备 1 的示例，并不构成对电子设备 1 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如所述电子设备 1 还可以包括输入输出设备、网络接入设备、总线等。

所述处理器 13 可以是中央处理单元(Central Processing Unit, CPU)，还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现场可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。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，所述处理器 13 是所述电子设备 1 的运算核心和控制中心，利用各种接口和线路连接整个电子设备 1 的各个部分，及获取所述电子设备 1 的操作系统以及安装的各类应用程序、程序代码等。

所述处理器 13 获取所述电子设备 1 的操作系统以及安装的各类应用程序。所述处理器 13 获取所述应用程序以实现上述各个基于 AI 识别的会议纪要生成方法实施例中的步骤，例如如图 1 所示的步骤。

示例性的，所述计算机程序可以被分割成一个或多个模块/单元，所述一个或者多个模块/单元被存储在所述存储器 12 中，并由所述处理器 13 获取，以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机可读指令段，该指令段用于描述所述计算机程序在所述电子设备 1 中的获取过程。例如，所述计算机程序可以被分割成获取单元 110、提取单元 111、确定单元 112、处理单元 113、比较单元 114、识别单元 115、翻译单元 116、生成单元 117、预处理单元 118、编码单元 119、建立单元 120、加密单元 121 及发送单元 122。

所述存储器 12 可用于存储所述计算机程序和/或模块，所述处理器 13 通过运行或获取存储在所述存储器 12 内的计算机程序和/或模块，以及调用存储在存储器 12 内的数据，实现所述电子设备 1 的各种功能。所述存储器 12 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（比如声音播放功能、图像播放功能等）等；存储数据区可存储根据电子设备的使用所创建的数据等。此外，存储器 12 可以包括

非易失性存储器，也可以包括易失性存储器，例如硬盘、内存、插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）、至少一个磁盘存储器件、闪存器件、或其他存储器件。

所述存储器 12 可以是电子设备 1 的外部存储器和/或内部存储器。进一步地，所述存储器 12 可以是具有实物形式的存储器，如内存条、TF 卡（Trans-flash Card）等等。

所述电子设备 1 集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读存储介质中，所述计算机可读存储介质可以是非易失性的存储介质，也可以是易失性的存储介质。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，也可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一计算机可读存储介质中，该计算机程序在被处理器获取时，可实现上述各个方法实施例的步骤。

其中，所述计算机程序包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可获取文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM, Read-Only Memory）、随机存取存储器。

进一步地，计算机可用存储介质可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序等；存储数据区可存储根据区块链节点的使用所创建的数据等。

本申请所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain)，本质上是一个去中心化的数据库，是一串使用密码学方法相关联产生的数据块，每一个数据块中包含了一批次网络交易的信息，用于验证其信息的有效性（防伪）和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

结合图 1，所述电子设备 1 中的所述存储器 12 存储多个指令以实现一种基于 AI 识别的会议纪要生成方法，所述处理器 13 可获取所述多个指令从而实现：当接收到会议纪要生成请求时，从所述会议纪要生成请求中获取参会者名单；从配置库中提取所述参会者名单中每个参会者对应的声纹特征；当接收到音频检测指令时，确定发出所述音频检测指令的音频设备，并从所述音频设备中获取待识别音频；对所述待识别音频进行处理，得到音频特征；将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象；对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息；将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本；根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

具体地，所述处理器 13 对上述指令的具体实现方法可参考图 1 对应实施例中相关步骤的描述，在此不赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。

所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块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。

另外，在本申请各个实施例中的各功能模块可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能模块的形式实现。

因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本申请的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本申请内。不应将权利要求中的任何附关联图标记视为限制所涉及

的权利要求。

此外，显然“包括”一词不排除其他单元或步骤，单数不排除复数。本申请中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一、第二等词语用来表示名称，而并不表示任何特定的顺序。

最后应说明的是，以上实施例仅用以说明本申请的技术方案而非限制，尽管参照较佳实施例对本申请进行了详细说明，本领域的普通技术人员应当理解，可以对本申请的技术方案进行修改或等同替换，而不脱离本申请技术方案的精神和范围。

权 利 要 求 书

1. 一种基于 AI 识别的会议纪要生成方法, 其中, 所述基于 AI 识别的会议纪要生成方法包括:

当接收到会议纪要生成请求时, 从所述会议纪要生成请求中获取参会者名单;

从配置库中提取所述参会者名单中每个参会者对应的声纹特征;

当接收到音频检测指令时, 确定发出所述音频检测指令的音频设备, 并从所述音频设备中获取待识别音频;

对所述待识别音频进行处理, 得到音频特征;

将所述音频特征与每个声纹特征进行比较, 得到比较结果, 并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象;

对所述待识别音频进行语音识别, 得到所述发言对象对应的文本信息;

将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本;

根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

2. 根据权利要求 1 所述的基于 AI 识别的会议纪要生成方法, 其中, 在从配置库中提取所述参会者名单中每个参会者对应的声纹特征之前, 所述基于 AI 识别的会议纪要生成方法还包括:

获取多个用户的音频信息及所述多个用户中每个用户的身份识别码;

利用语音信号处理技术, 对所述音频信息进行预处理;

对预处理后的音频信息进行编码, 得到所述音频信息的数字信号;

从所述数字信号中提取声纹特征;

以所述身份识别码与所述声纹特征建立映射关系并保存至所述配置库中。

3. 根据权利要求 1 所述的基于 AI 识别的会议纪要生成方法, 其中, 所述确定发出所述音频检测指令的音频设备包括:

解析所述音频检测指令, 得到参数信息;

根据所述参数信息确定所述音频设备。

4. 根据权利要求 1 所述的基于 AI 识别的会议纪要生成方法, 其中, 所述对所述待识别音频进行处理, 得到音频特征包括:

对所述待识别音频进行分块处理, 得到音频区域集;

提取所述音频区域集中每一帧对应的特征参数, 得到所述音频特征。

5. 根据权利要求 1 所述的基于 AI 识别的会议纪要生成方法, 其中, 所述将所述音频特征与每个声纹特征进行比较, 得到比较结果, 并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象包括:

将所述音频特征与每个声纹特征进行比较, 得到比较结果, 所述比较结果为每个声纹特征与所述音频特征的相似度;

当检测到所述相似度中存在任意相似度大于阈值时, 将所述任意相似度中的最大相似度确定为目标相似度, 并将所述目标相似度对应的声纹特征确定为目标声纹特征;

将所述目标声纹特征对应的用户确定为第一用户;

获取与所述音频设备绑定的用户作为第二用户;

当检测到所述第一用户与所述第二用户为同一人时, 将所述第一用户确定为所述发言对象; 或者

当检测到所述第一用户与所述第二用户不为同一人时, 启动摄像设备对当前使用所述音频设备的用户进行录像, 得到目标图像, 并将所述目标图像与人脸库中的预设图像进行匹配, 得到每张预设图像的匹配度, 每张预设图像对应一个人员, 并获取每个人员的语言特征, 确定每个语言特征与所述音频特征的相似度, 采用预设权重对每个相似度与每个匹配度进行加权和运算, 得到每个人员的分值, 将分值最高的人员确定为所述发

言对象。

6. 根据权利要求1所述的基于AI识别的会议纪要生成方法,其中,所述对所述待识别音频进行语音识别,得到所述发言对象对应的文本信息包括:

利用移动窗函数对所述待识别音频按照预设时长进行分割,得到至少一帧语音;

对每帧语音进行预处理,并计算每帧预处理后的语音的能量值;

当检测到预处理后的语音中连续预设帧的能量值小于配置值,及接下来的语音中连续所述预设帧的能量值大于所述配置值时,将能量值开始增大的语音帧确定为语音的前端点;

当检测到预处理后的语音中连续所述预设帧的能量值大于所述配置值,及接下来的语音中连续所述预设帧的能量值小于所述配置值时,将能量值开始减小的语音帧确定为语音的后端点;

删除所述前端点至所述后端点之外的语音帧,得到目标音频;

对所述目标音频进行分块处理,得到目标语音帧;

将所述目标语音帧转换为所述文本信息。

7. 根据权利要求1所述的基于AI识别的会议纪要生成方法,其中,所述根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要包括:

获取预先配置的会议纪要模版,并获取所述文本信息的发言时间;

依据所述发言时间从先至后的顺序,将所述发言对象、所述文本信息及所述目标语言文本录入所述会议纪要模板,得到会议记录;

对所述会议记录进行分词处理,得到分词;

采用编码算法将所述分词转换为第一词向量,并将预设词组转换为第二词向量;

利用余弦距离公式,计算所述第一词向量与所述第二词向量的距离;

将距离值大于预设值的第一词向量确定为目标词向量,并将所述目标词向量对应的分词确定为目标词组;

在所述会议记录中标识所述目标词组,得到所述会议纪要。

8. 一种基于AI识别的会议纪要生成装置,其中,所述基于AI识别的会议纪要生成装置包括:

获取单元,用于当接收到会议纪要生成请求时,从所述会议纪要生成请求中获取参会者名单;

提取单元,用于从配置库中提取所述参会者名单中每个参会者对应的声纹特征;

确定单元,用于当接收到音频检测指令时,确定发出所述音频检测指令的音频设备,并从所述音频设备中获取待识别音频;

处理单元,用于对所述待识别音频进行处理,得到音频特征;

比较单元,用于将所述音频特征与每个声纹特征进行比较,得到比较结果,并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象;

识别单元,用于对所述待识别音频进行语音识别,得到所述发言对象对应的文本信息;

翻译单元,用于将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本;

生成单元,用于根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

9. 一种电子设备,其中,所述电子设备包括处理器和存储器,所述处理器用于执行存储器中存储的至少一个计算机可读指令以实现以下步骤:

当接收到会议纪要生成请求时,从所述会议纪要生成请求中获取参会者名单;

从配置库中提取所述参会者名单中每个参会者对应的声纹特征;

当接收到音频检测指令时,确定发出所述音频检测指令的音频设备,并从所述音频

设备中获取待识别音频；

对所述待识别音频进行处理，得到音频特征；

将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象；

对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息；

将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本；

根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

10. 根据权利要求9所述的电子设备，其中，在从配置库中提取所述参会者名单中每个参会者对应的声纹特征之前，所述处理器执行所述至少一个计算机可读指令还用以实现以下步骤：

获取多个用户的音频信息及所述多个用户中每个用户的身份识别码；

利用语音信号处理技术，对所述音频信息进行预处理；

对预处理后的音频信息进行编码，得到所述音频信息的数字信号；

从所述数字信号中提取声纹特征；

以所述身份识别码与所述声纹特征建立映射关系并保存至所述配置库中。

11. 根据权利要求9所述的电子设备，其中，在所述确定发出所述音频检测指令的音频设备时，所述处理器执行所述至少一个计算机可读指令以实现以下步骤：

解析所述音频检测指令，得到参数信息；

根据所述参数信息确定所述音频设备。

12. 根据权利要求9所述的电子设备，其中，在所述对所述待识别音频进行处理，得到音频特征时，所述处理器执行所述至少一个计算机可读指令以实现以下步骤：

对所述待识别音频进行分块处理，得到音频区域集；

提取所述音频区域集中每一帧对应的特征参数，得到所述音频特征。

13. 根据权利要求9所述的电子设备，其中，在所述将所述音频特征与每个声纹特征进行比较，得到比较结果，并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象时，所述处理器执行所述至少一个计算机可读指令以实现以下步骤：

将所述音频特征与每个声纹特征进行比较，得到比较结果，所述比较结果为每个声纹特征与所述音频特征的相似度；

当检测到所述相似度中存在任意相似度大于阈值时，将所述任意相似度中的最大相似度确定为目标相似度，并将所述目标相似度对应的声纹特征确定为目标声纹特征；

将所述目标声纹特征对应的用户确定为第一用户；

获取与所述音频设备绑定的用户作为第二用户；

当检测到所述第一用户与所述第二用户为同一人时，将所述第一用户确定为所述发言对象；或者

当检测到所述第一用户与所述第二用户不为同一人时，启动摄像设备对当前使用所述音频设备的用户进行录像，得到目标图像，并将所述目标图像与人脸库中的预设图像进行匹配，得到每张预设图像的匹配度，每张预设图像对应一个人员，并获取每个人员的语言特征，确定每个语言特征与所述音频特征的相似度，采用预设权重对每个相似度与每个匹配度进行加权和运算，得到每个人员的分值，将分值最高的人员确定为所述发言对象。

14. 根据权利要求9所述的电子设备，其中，在所述对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息时，所述处理器执行所述至少一个计算机可读指令以实现以下步骤：

利用移动窗函数对所述待识别音频按照预设时长进行分割，得到至少一帧语音；

对每帧语音进行预处理，并计算每帧预处理后的语音的能量值；

当检测到预处理后的语音中连续预设帧的能量值小于配置值，及接下来的语音中连续所述预设帧的能量值大于所述配置值时，将能量值开始增大的语音帧确定为语音的前端点；

当检测到预处理后的语音中连续所述预设帧的能量值大于所述配置值, 及接下来的语音中连续所述预设帧的能量值小于所述配置值时, 将能量值开始减小的语音帧确定为语音的后端点;

删除所述前端点至所述后端点之外的语音帧, 得到目标音频;

对所述目标音频进行分块处理, 得到目标语音帧;

将所述目标语音帧转换为所述文本信息。

15. 一种计算机可读存储介质, 其中, 所述计算机可读存储介质存储有至少一个计算机可读指令, 所述至少一个计算机可读指令被处理器执行时实现以下步骤:

当接收到会议纪要生成请求时, 从所述会议纪要生成请求中获取参会者名单;

从配置库中提取所述参会者名单中每个参会者对应的声纹特征;

当接收到音频检测指令时, 确定发出所述音频检测指令的音频设备, 并从所述音频设备中获取待识别音频;

对所述待识别音频进行处理, 得到音频特征;

将所述音频特征与每个声纹特征进行比较, 得到比较结果, 并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象;

对所述待识别音频进行语音识别, 得到所述发言对象对应的文本信息;

将所述文本信息翻译成与预先设置好的语言类型对应的目标语言文本;

根据所述发言对象、所述文本信息及所述目标语言文本生成会议纪要。

16. 根据权利要求 15 所述的存储介质, 其中, 在从配置库中提取所述参会者名单中每个参会者对应的声纹特征之前, 所述至少一个计算机可读指令被处理器执行还用以实现以下步骤:

获取多个用户的音频信息及所述多个用户中每个用户的身份识别码;

利用语音信号处理技术, 对所述音频信息进行预处理;

对预处理后的音频信息进行编码, 得到所述音频信息的数字信号;

从所述数字信号中提取声纹特征;

以所述身份识别码与所述声纹特征建立映射关系并保存至所述配置库中。

17. 根据权利要求 15 所述的存储介质, 其中, 在所述确定发出所述音频检测指令的音频设备时, 所述至少一个计算机可读指令被处理器执行以实现以下步骤:

解析所述音频检测指令, 得到参数信息;

根据所述参数信息确定所述音频设备。

18. 根据权利要求 15 所述的存储介质, 其中, 在所述对所述待识别音频进行处理, 得到音频特征时, 所述至少一个计算机可读指令被处理器执行以实现以下步骤:

对所述待识别音频进行分块处理, 得到音频区域集;

提取所述音频区域集中每一帧对应的特征参数, 得到所述音频特征。

19. 根据权利要求 15 所述的存储介质, 其中, 在所述将所述音频特征与每个声纹特征进行比较, 得到比较结果, 并根据所述比较结果及所述音频设备确定所述待识别音频的发言对象时, 所述至少一个计算机可读指令被处理器执行时以实现以下步骤:

将所述音频特征与每个声纹特征进行比较, 得到比较结果, 所述比较结果为每个声纹特征与所述音频特征的相似度;

当检测到所述相似度中存在任意相似度大于阈值时, 将所述任意相似度中的最大相似度确定为目标相似度, 并将所述目标相似度对应的声纹特征确定为目标声纹特征;

将所述目标声纹特征对应的用户确定为第一用户;

获取与所述音频设备绑定的用户作为第二用户;

当检测到所述第一用户与所述第二用户为同一人时, 将所述第一用户确定为所述发言对象; 或者

当检测到所述第一用户与所述第二用户不为同一人时, 启动摄像设备对当前使用所述音

频设备的用户进行录像，得到目标图像，并将所述目标图像与人脸库中的预设图像进行匹配，得到每张预设图像的匹配度，每张预设图像对应一个人员，并获取每个人员的语言特征，确定每个语言特征与所述音频特征的相似度，采用预设权重对每个相似度与每个匹配度进行加权和运算，得到每个人员的分值，将分值最高的人员确定为所述发言对象。

20. 根据权利要求 15 所述的存储介质，其中，在所述对所述待识别音频进行语音识别，得到所述发言对象对应的文本信息时，所述至少一个计算机可读指令被处理器执行以实现以下步骤：

利用移动窗函数对所述待识别音频按照预设时长进行分割，得到至少一帧语音；

对每帧语音进行预处理，并计算每帧预处理后的语音的能量值；

当检测到预处理后的语音中连续预设帧的能量值小于配置值，及接下来的语音中连续所述预设帧的能量值大于所述配置值时，将能量值开始增大的语音帧确定为语音的前端点；

当检测到预处理后的语音中连续所述预设帧的能量值大于所述配置值，及接下来的语音中连续所述预设帧的能量值小于所述配置值时，将能量值开始减小的语音帧确定为语音的后端点；

删除所述前端点至所述后端点之外的语音帧，得到目标音频；

对所述目标音频进行分块处理，得到目标语音帧；

将所述目标语音帧转换为所述文本信息。

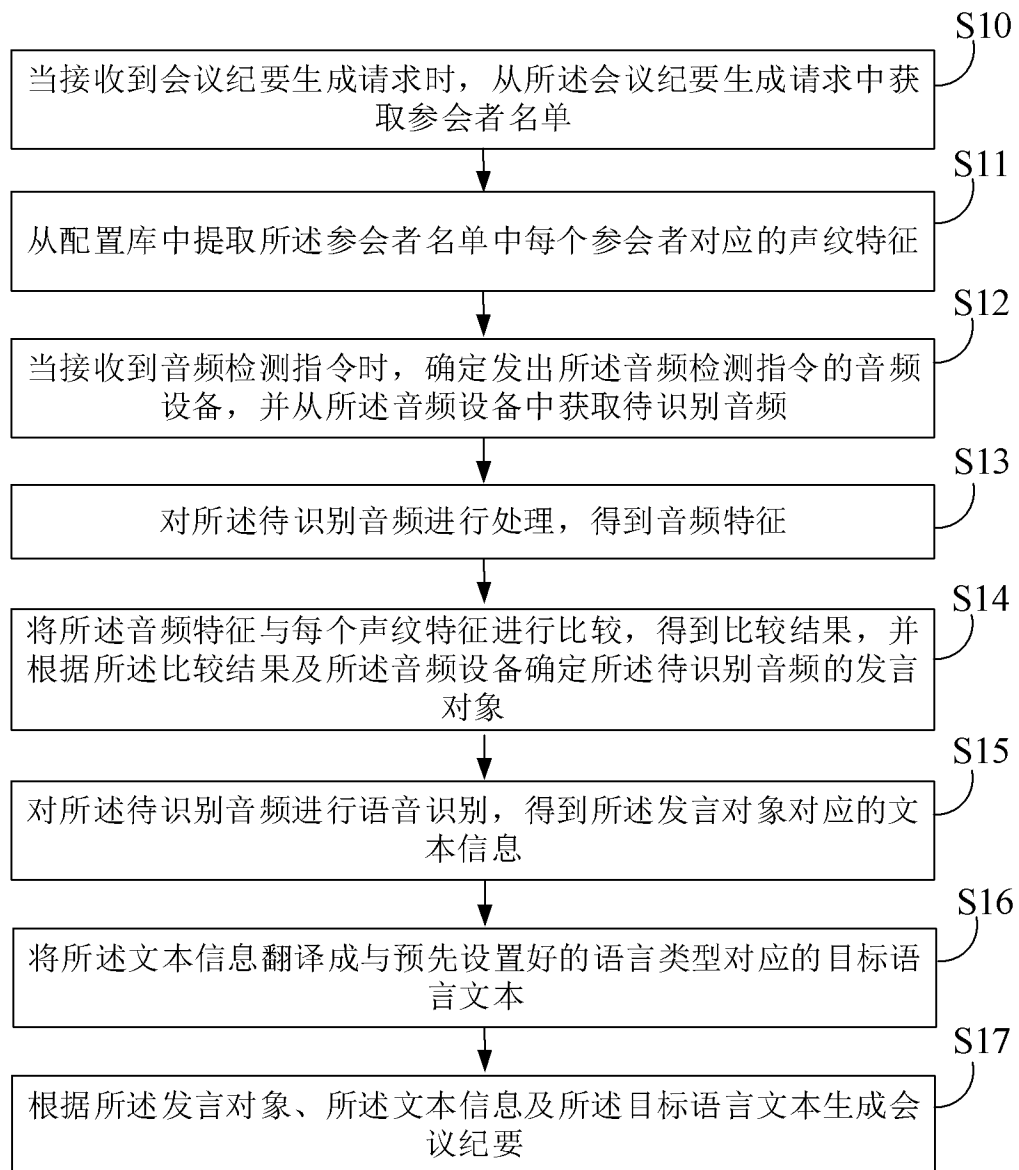


图 1

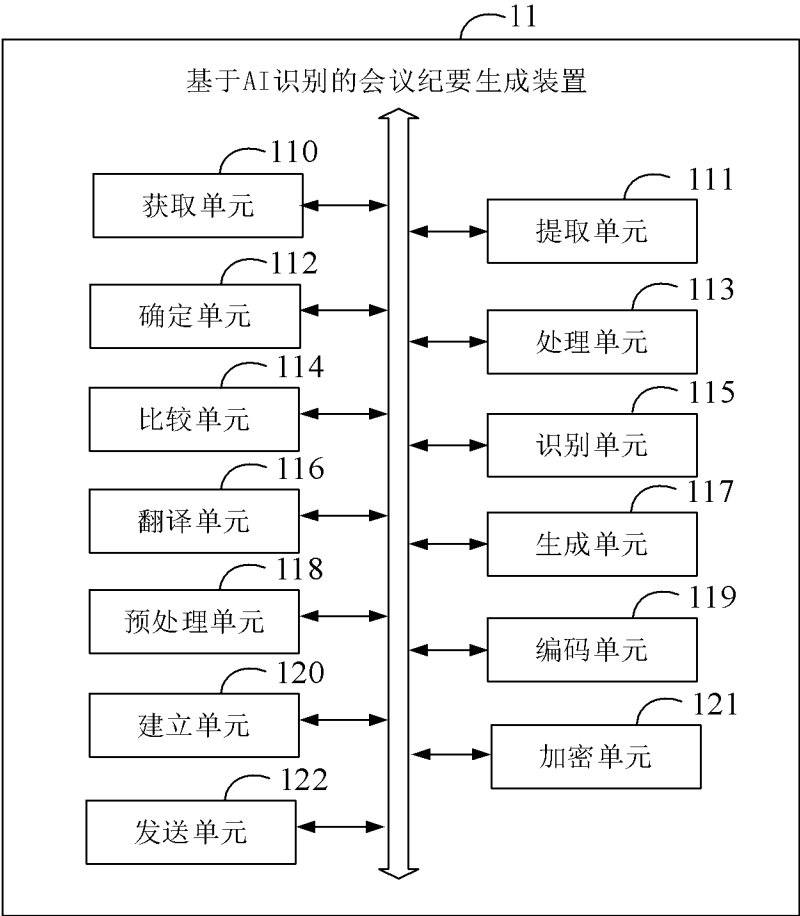


图 2

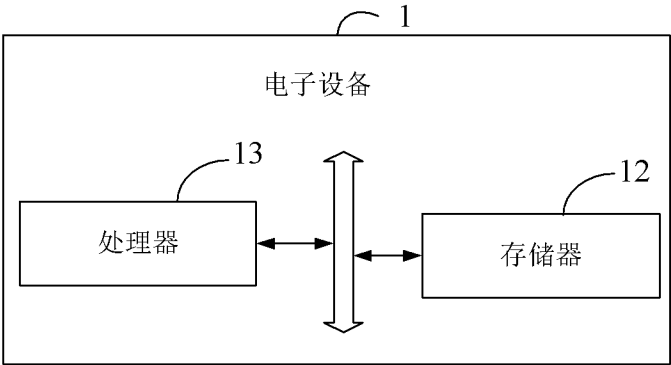


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/096433

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/22(2006.01)i; G10L 17/02(2013.01)i; H04L 12/18(2006.01)i; H04N 7/15(2006.01)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 15; G10L 17; H04L 12/18+; H04N 7; H04M 3/56+

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, CNKI, VEN, USTXT, WOTXT, EPTXT: 会议, 开会, 参会, 与会, 纪要, 记录, 转写, 翻译, 声纹, 设备, 麦克风, 话筒, 电脑, 终端, IP, MAC, 绑定, 映射, 关联, 匹配, 对应, 面部, 图像, 摄像, 语种, 语言; meeting?, conference, record+, minut+, transcript+, translat+, voice print?, voiceprint?, language, match+, participant?, speaker

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111883123 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 03 November 2020 (2020-11-03) description paragraphs [0005]-[0217], claims 1-10, figures 1-3	1-20
Y	CN 109388701 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 26 February 2019 (2019-02-26) description paragraphs [0053]-[0150]	1-20
Y	CN 109560941 A (SHENZHEN WATER WORLD CO., LTD.) 02 April 2019 (2019-04-02) description, paragraphs [0057]-[0099]	1-20
Y	CN 105632498 A (TOSHIBA CORPORATION) 01 June 2016 (2016-06-01) description, paragraphs [0007]-[0017]	1-20
A	CN 108986826 A (PING AN LIFE INSURANCE COMPANY OF CHINA, LTD.) 11 December 2018 (2018-12-11) entire document	1-20
A	CN 106782545 A (GUANGZHOU CVTE ELECTRONIC TECHNOLOGY COMPANY LIMITED et al.) 31 May 2017 (2017-05-31) 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

22 August 2021

Date of mailing of the international search report

30 August 2021

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

International application No.

PCT/CN2021/096433**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 110049270 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 23 July 2019 (2019-07-23) entire document	1-20
A	CN 109309804 A (WUHU XINGTU ROBOT TECH CO., LTD.) 05 February 2019 (2019-02-05) entire document	1-20
A	US 2019273767 A1 (NELSON STEVEN et al.) 05 September 2019 (2019-09-05) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/096433

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111883123	A	03 November 2020	None			
CN	109388701	A	26 February 2019	None			
CN	109560941	A	02 April 2019	None			
CN	105632498	A	01 June 2016	None			
CN	108986826	A	11 December 2018	None			
CN	106782545	A	31 May 2017	CN	106782545	B	16 July 2019
				WO	2018107605	A1	21 June 2018
CN	110049270	A	23 July 2019	None			
CN	109309804	A	05 February 2019	None			
US	2019273767	A1	05 September 2019	US	10757148	B2	25 August 2020

国际检索报告

国际申请号

PCT/CN2021/096433

A. 主题的分类

G10L 15/22(2006.01)i; G10L 17/02(2013.01)i; H04L 12/18(2006.01)i; H04N 7/15(2006.01)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 15; G10L 17; H04L 12/18+; H04N 7; H04M 3/56+

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, CNKI, VEN, USTXT, WOTXT, EPTXT: 会议, 开会, 参会, 与会, 纪要, 记录, 转写, 翻译, 声纹, 设备, 麦克, 话筒, 电脑, 终端, IP, MAC, 绑定, 映射, 关联, 匹配, 对应, 面部, 图像, 摄像, 语种, 语言;
meeting?, conference, record+, minut+, transcript+, translat+, voice print?, voiceprint?, language,
match+, participant?, speaker

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
PX	CN 111883123 A (平安科技深圳有限公司) 2020年 11月 3日 (2020 - 11 - 03) 说明书第[0005]-[0217]段, 权利要求1-10, 附图1-3	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[0053]-[0150]段	1-20
Y	CN 109560941 A (深圳市沃特沃德股份有限公司) 2019年 4月 2日 (2019 - 04 - 02) 说明书[0057]-[0099]段	1-20
Y	CN 105632498 A (株式会社 东芝) 2016年 6月 1日 (2016 - 06 - 01) 说明书[0007]-[0017]段	1-20
A	CN 108986826 A (中国平安人寿保险股份有限公司) 2018年 12月 11日 (2018 - 12 - 11) 全文	1-20
A	CN 106782545 A (广州视源电子科技股份有限公司 等) 2017年 5月 31日 (2017 - 05 - 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

2021年 8月 22日

国际检索报告邮寄日期

2021年 8月 30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

杨士林

电话号码 86-(010)-62085717

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 110049270 A (平安科技深圳有限公司) 2019年 7月 23日 (2019 - 07 - 23) 全文	1-20
A	CN 109309804 A (芜湖星途机器人科技有限公司) 2019年 2月 5日 (2019 - 02 - 05) 全文	1-20
A	US 2019273767 A1 (NELSON STEVEN 等) 2019年 9月 5日 (2019 - 09 - 05) 全文	1-20

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/096433

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	111883123	A	2020年 11月 3日	无	
CN	109388701	A	2019年 2月 26日	无	
CN	109560941	A	2019年 4月 2日	无	
CN	105632498	A	2016年 6月 1日	无	
CN	108986826	A	2018年 12月 11日	无	
CN	106782545	A	2017年 5月 31日	CN 106782545 B	2019年 7月 16日
				WO 2018107605 A1	2018年 6月 21日
CN	110049270	A	2019年 7月 23日	无	
CN	109309804	A	2019年 2月 5日	无	
US	2019273767	A1	2019年 9月 5日	US 10757148 B2	2020年 8月 25日