

(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 7 月 23 日 (23.07.2020)



(10) 国际公布号
WO 2020/147407 A1

- (51) 国际专利分类号:
G10L 15/04 (2013.01) **G10L 15/08** (2006.01)
- (21) 国际申请号: PCT/CN2019/118256
- (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:
201910038460.6 2019 年 1 月 16 日 (16.01.2019) 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 518000 (CN)。
- (72) 发明人: 吴欢(WU, Huan); 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

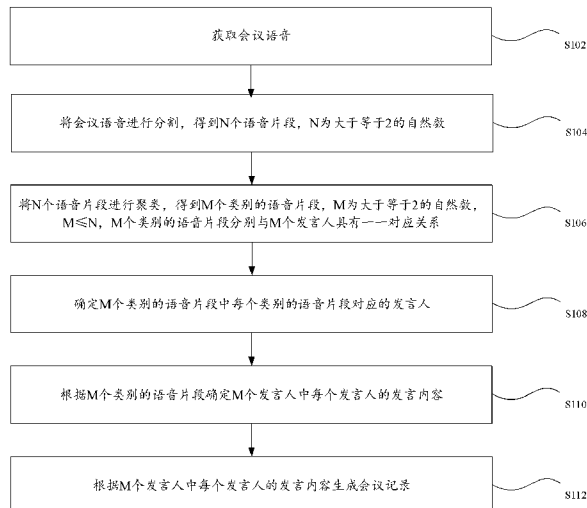
安金融中心 23 楼, Guangdong 518000 (CN)。田甜(TIAN, Tian); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 北京汇思诚业知识产权代理有限公司(UNI-INTEL PATENT AND TRADEMARK LAW FIRM); 中国北京市朝阳区建国门外大街永安东里甲 3 号通用国际中心 A 座 3 层, Beijing 100022 (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, 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,

(54) Title: CONFERENCE RECORD GENERATION METHOD AND APPARATUS, STORAGE MEDIUM AND COMPUTER DEVICE

(54) 发明名称: 一种会议记录生成方法、装置、存储介质及计算机设备



S102 ACQUIRE A CONFERENCE VOICE
S104 SEGMENT THE CONFERENCE VOICE TO OBTAIN N VOICE SEGMENTS, N BEING A NATURAL NUMBER EQUAL TO OR GREATER THAN TWO
S106 CLUSTER THE N VOICE SEGMENTS TO OBTAIN M CATEGORIES OF VOICE SEGMENTS, M BEING A NATURAL NUMBER EQUAL TO OR GREATER THAN TWO, $M \leq N$, THE M CATEGORIES OF VOICE SEGMENTS CORRESPONDING TO M SPEAKERS ON A ONE-TO-ONE BASIS
S108 DETERMINE THE SPEAKER CORRESPONDING TO EACH CATEGORY OF VOICE SEGMENTS AMONG THE M CATEGORIES OF VOICE SEGMENTS
S110 DETERMINE SPEECH CONTENT OF EACH OF THE M SPEAKERS ACCORDING TO THE M CATEGORIES OF VOICE SEGMENTS
S112 GENERATE A CONFERENCE RECORD ACCORDING TO THE SPEECH CONTENT OF EACH OF THE M SPEAKERS

图 1

(57) Abstract: A conference record generation method and apparatus, a storage medium and a computer device, relating to the field of artificial intelligence technology. The method comprises: acquiring a conference voice (S102); segmenting the conference voice to obtain N voice segments, N being a natural number equal to or greater than two (S104); clustering the N voice segments to obtain M categories of voice segments, M being a natural number equal to or greater than two, $M \leq N$, the M categories of voice segments corresponding to M speakers on a one-to-one basis (S106); determining the speaker corresponding to each category of voice segments among the M categories of voice segments (S108); determining speech content of each of the M speakers according to the M categories of voice segments (S110); and generating a conference record according to the speech content of each of the M speakers (S112). Said method is able to solve the problem in the prior art that manually recording a conference is time-consuming, laborious, and low-efficient.

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))。

(57) 摘要: 一种会议记录生成方法、装置、存储介质及计算机设备, 涉及人工智能技术领域, 该方法包括: 获取会议语音 (S102); 将会议语音进行分割, 得到N个语音片段, N为大于等于2的自然数 (S104); 将N个语音片段进行聚类, 得到M个类别的语音片段, M为大于等于2的自然数, $M \leq N$, M个类别的语音片段分别与M个发言人具有一一对应关系 (S106); 确定M个类别的语音片段中每个类别的语音片段对应的发言人 (S108); 根据M个类别的语音片段确定M个发言人中每个发言人的发言内容 (S110); 根据M个发言人中每个发言人的发言内容生成会议记录 (S112)。上述方法能够解决现有技术中人工整理会议记录费时费力、效率低的问题。

一种会议记录生成方法、装置、存储介质及计算机设备

本申请要求于 2019 年 01 月 16 日提交中国专利局、申请号为 201910038460.6、申请名称为“一种会议记录生成方法和装置”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】

本申请涉及人工智能技术领域，尤其涉及一种会议记录生成方法和装置。

【背景技术】

在会议过程中，由记录人员把会议的各个发言人的发言内容记录并整理，形成会议记录。当会议时间比较长，需要记录的内容比较多的时候，人工整理会议记录费时费力、效率低。

【申请内容】

有鉴于此，本申请实施例提供了一种会议记录生成方法和装置，用以解决现有技术中人工整理会议记录费时费力、效率低的问题。

一方面，本申请实施例提供了一种会议记录生成方法，所述方法包括：获取会议语音；将所述会议语音进行分割，得到 N 个语音片段， N 为大于等于 2 的自然数；将所述 N 个语音片段进行聚类，得到 M 个类别的语音片段， M 为大于等于 2 的自然数， $M \leq N$ ，所述 M 个类别的语音片段分别与 M 个发言人具有一一对应关系；确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人；根据所述 M 个类别的语音片段确定所述 M 个发言人中每个发言人的发言内容；根据所述 M 个发言人中每个发言人的发言内容生成会议记录。

一方面，本申请实施例提供了一种会议记录生成装置，所述装置包括：获取单元，用于获取会议语音；分割单元，用于将所述会议语音进行分割，得到 N 个语音片段， N 为大于等于 2 的自然数；聚类单元，用于将所述 N 个语音片段进行聚类，得到 M 个类别的语音片段， M 为大于等于 2 的自然数， $M \leq N$ ，所述 M 个类别的语音片段分别与 M 个发言人具有一一对应关系；第一确定单元，用于确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人；第二确定单元，用于根据所述 M 个类别的语音片段确定所述 M 个发言人中每个发言人的发言内容；生成单元，用于根据所述 M 个发言人中每个发言人的发言内容生成会议记录。

一方面，本申请实施例提供了一种存储介质，所述存储介质包括存储的程序，其中，在所述程序运行时控制所述存储介质所在设备执行上述的会议记录生成方法。

一方面，本申请实施例提供了一种计算机设备，包括存储器和处理器，所述存储器用于存储包括程序指令的信息，所述处理器用于控制程序指令的执行，所述程序指令被处理器加载并执行时实现上述的会议记录生成方法的步骤。

在本申请实施例中，将会议语音进行分割，得到 N 个语音片段，将 N 个语音片段进行聚类，得到 M 个类别的语音片段，确定每个类别的语音片段对应的发言人；根据 M 个类别的语音片段确定 M 个发言人的发言内容；根据各个发言人的发言内容，生成会议记录，解决了现有技术中人工整理会议记录费时费力、效率低的问题，达到了智能分析会议上的发言内容，高效整理出会议记录的效果。

【附图说明】

为了更清楚地说明本申请实施例的技术方案，下面将对实施例中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性的前提下，还可以根据这些附图获得其它的附图。

图 1 是根据本申请实施例一种可选的会议记录生成方法的流程图；

图 2 是根据本申请实施例一种可选的会议记录生成装置的示意图；

图 3 是本申请实施例提供的一种可选的计算机设备的示意图。

【具体实施方式】

为了更好的理解本申请的技术方案，下面结合附图对本申请实施例进行详细描述。

应当明确，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例，都属于本申请保护的范围。

在本申请实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本申请。在本申请实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。

应当理解，本文中使用的术语“和/或”仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，本文中

字符“/”，一般表示前后关联对象是一种“或”的关系。

图 1 是根据本申请实施例一种可选的会议记录生成方法的流程图，如图 1 所示，该方法包括：

步骤 S102，获取会议语音。

步骤 S104，将会议语音进行分割，得到 N 个语音片段，N 为大于等于 2 的自然数。

步骤 S106，将 N 个语音片段进行聚类，得到 M 个类别的语音片段，M 为大于等于 2 的自然数， $M \leq N$ ，M 个类别的语音片段分别与 M 个发言人具有一一对应关系。

步骤 S108，确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

步骤 S110，根据 M 个类别的语音片段确定 M 个发言人中每个发言人的发言内容。

步骤 S112，根据 M 个发言人中每个发言人的发言内容生成会议记录。

在本申请实施例中，会议可以包括两种情况：

第一种情况：所有参会人员都到场的情况。例如，某部门举行了一次会议，所有人聚在同一个会议室开会。

第二种情况：借助于某些应用软件进行开会的情况。例如，某部门举行了一次会议，一些人聚在同一个会议室，另一些人在外地出差，通过微信、QQ 或其他应用软件参加会议。再例如，某公司举行了一次会议，参会人员有 3 个，分别为北京的部门经理、上海的部门经理、深圳的部门经理，这些人的地理位置分别位于北京、上海、深圳，这三个人在不同的城市通过微信、QQ 或其他应用软件开会。

在本申请实施例中，会议语音可以为以上任意一种方式开会的过程中产生的语音。会议语音可以是在开会时现场录制的，例如，若干人聚在一起开会，其中一个人用手机、录音机、录音笔或者其他录音设备录制会议过程中产生的语音得到会议语音；会议语音也可以是通过录制通过即时通讯软件开会的过程中产生的语音得到的，例如，若干人通过微信/QQ 开会，其中一个人用手机、录音机、录音笔或者其他录音设备录制会议过程中产生的微信语音/QQ 语音得到会议语音。

在本申请实施例中，发言人指的是在会议过程中发言的人，发言人的数量小于或等于参会人员数量，如果所有参会人员都发言了，那么发言人的数量等于参会人员的数量；如果只有部分参会人员发言了，那么发言人的数量小于参会人员的数量。

下面举一个具体的例子对本申请实施例提供的会议记录生成方法进行说明。

例如，若干个人开会，录制会议过程中产生的语音，得到会议语音，假设会议语音为 20 分钟，将会议语音进行分割，例如得到 6000 ($N=6000$) 个语音片段，将这 6000 个语音片段进行聚类，得到 3 ($M=3$) 个类别的语音片段，其中，类别 1 包括 3000 个语音片段，分别为语音片段 $P(1, 1)$ 、语音片段 $P(1, 2)$ 、……、语音片段 $P(1, 3000)$ ，这 3000 个语音片段对应同一个发言人；类别 2 包括 1000 个语音片段，分别为语音片段 $P(2, 1)$ 、语音片段 $P(2, 2)$ 、……、语音片段 $P(2, 1000)$ ，这 1000 个语音片段对应同一个发言人；类别 3 包括 2000 个语音片段，分别为语音片段 $P(3, 1)$ 、语音片段 $P(3, 2)$ 、……、语音片段 $P(3, 2000)$ ，这 2000 个语音片段对应同一个发言人。然后，分别确定每个类别的语音片段对应的发言人，例如，假设确定出类别 1 包括的 3000 个语音片段对应发言人甲，类别 2 包括的 1000 个语音片段对应发言人乙，类别 3 包括的 2000 个语音片段对应发言人丙，如表 1 所示。根据语音片段 $P(1, 1)$ 、语音片段 $P(1, 2)$ 、……、语音片段 $P(1, 3000)$ 确定发言人甲的发言内容；根据语音片段 $P(2, 1)$ 、语音片段 $P(2, 2)$ 、……、语音片段 $P(2, 1000)$ 确定发言人乙的发言内容；根据语音片段 $P(3, 1)$ 、语音片段 $P(3, 2)$ 、……、语音片段 $P(3, 2000)$ 确定发言人丙的发言内容。根据发言人甲、发言人乙和发言人丙的发言内容生成会议记录。

表 1

发言人	发言人对 应的语音 片段的数 量	发言人对应的语音片段	发言人对 应的语音片段 所属的类别
甲	3000	语音片段 $P(1, 1)$ 、语音片段 $P(1, 2)$ 、……、 语音片段 $P(1, 3000)$	类别 1
乙	1000	语音片段 $P(2, 1)$ 、语音片段 $P(2, 2)$ 、……、 语音片段 $P(2, 1000)$	类别 2
丙	2000	语音片段 $P(3, 1)$ 、语音片段 $P(3, 2)$ 、……、 语音片段 $P(3, 2000)$	类别 3

在本申请实施例中，将会议语音进行分割，得到 N 个语音片段，将 N 个语音片段进行聚类，得到 M 个类别的语音片段，确定每个类别的语音片段对应的发言人；根据 M 个类别的语音片段确定 M 个发言人的发言内容；根据各个发言人的发言内容，生成会议记录，解决了现有技术中人工整理会议记录费时费力、效率低的问题，达到了智能分析会议上的发言内容，高效整理出会议记录的效果。

确定每个类别的语音片段对应的发言人，具体方法可以有多种，下面举出几种。

方法一：

从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段，得到 L 个文本片段，L 为自然数， $L \geq M$ ；向用户展示 L 个文本片段和发言人列表，发言人列表包括 M 个发言人中每个发言人的信息；接收匹配指令，匹配指令为用户发出的用于指示将 L 个文本片段中每个文本片段与发言人进行匹配的指令；根据匹配指令确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段，得到 L 个文本片段，具体地，可以从 M 个类别的语音片段中每个类别的语音片段中各随机选择至少一个语音片段转换成文本片段。

例如，从表 1 所示的 3 个类别的语音片段中每个类别的语音片段中各选择一个语音片段，从类别 1 中选出的语音片段为语音片段 P (1, 1)；从类别 2 中选出的语音片段为语音片段 P (2, 1)；从类别 3 中选出的语音片段为语音片段 P (3, 1)。将这 3 个语音片段分别进行转换，得到文本片段 F (1, 1)、文本片段 F (2, 1)、文本片段 F (3, 1)，这三个文本片段与上述三个语音片段之间的对应关系如表 2 所示。

表 2

语音片段	语音片段转化得到的文本片段
语音片段 P (1, 1)	文本片段 F (1, 1)
语音片段 P (2, 1)	文本片段 F (2, 1)
语音片段 P (3, 1)	文本片段 F (3, 1)

向用户展示这 3 (L=3) 个文本片段和发言人列表，发言人列表包括 3 个发言人中每个发言人的信息。发言人的信息可以包括发言人的姓名、职位等。

用户可以是会议的主持人，可以是其他参会人员。

用户看到这 3 个文本片段后，即可知道与文本片段相对应的是哪位参

会人员的发言。例如，假设有一个文本片段的内容是：“大家好，我是今天会议的主持人。”当用户看到这个文本片段后，即可知道这个文本片段对应的是会议主持人的发言。用户可发出匹配指令，匹配指令为用于指示将3个文本片段中每个文本片段与发言人进行匹配的指令，例如，匹配指令指示按照表3将文本片段与发言人进行匹配。

表 3

文本片段	文本片段对应的发言人
文本片段 F (1, 1)	甲
文本片段 F (2, 1)	乙
文本片段 F (3, 1)	丙

由于文本片段 F (1, 1) 是类别 1 中的语音片段转换得到的，而类别 1 中的所有语音片段对应的是同一个发言人，因此，文本片段 F (1, 1) 对应的发言人甲即为类别 1 中的所有语音片段对应的发言人，即，类别 1 中的所有语音片段都是发言人甲发出的；同理，由于文本片段 F (2, 1) 是类别 2 中的语音片段转换得到的，类别 2 中的所有语音片段都是发言人乙发出的；同理，由于文本片段 F (3, 1) 是类别 3 中的语音片段转换得到的，类别 3 中的所有语音片段都是发言人丙发出的，语音片段与发言人之间的对应关系如表 4 所示。

表 4

语音片段	语音片段对应的发言人
语音片段 P (1, 1)、语音片段 P (1, 2)、.....、语音片段 P (1, 3000)	甲
语音片段 P (2, 1)、语音片段 P (2, 2)、.....、语音片段 P (2, 1000)	乙
语音片段 P (3, 1)、语音片段 P (3, 2)、.....、语音片段 P (3, 2000)	丙

方法二：

从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段，得到 Z 个语音片段，Z 为自然数， $Z \geq M$ ；向用户播放选择出的 Z 个语音片段并展示发言人列表，发言人列表包括 M 个发言人中每个发言人的信息；接收匹配指令，匹配指令为用户发出的用于指示将 Z 个语音片段中每个语音片段与发言人进行匹配的指令；根据匹配指令确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语

音片段，具体地，可以从 M 个类别的语音片段中每个类别的语音片段中各随机选择至少一个语音片段。

假设从类别 1 中随机选择了 2 个语音片段，分别为语音片段 F (1, 32)、语音片段 F (1, 450)；从类别 2 中随机选择了 2 个语音片段，分别为语音片段 F (2, 100)、语音片段 F (2, 400)；从类别 3 中随机选择了 2 个语音片段，分别为语音片段 F (3, 900)、语音片段 F (3, 600)。

向用户播放这 6 ($Z=6$) 个语音片段，用户听到这 6 个语音片段后，根据声音的音色能够轻松识别每个语音片段是哪位参会人员的发言。用户可发出匹配指令，匹配指令为用于指示将 6 个语音片段中每个语音片段与发言人进行匹配的指令，匹配方式如表 5 所示。

表 5

语音片段	语音片段对应的发言人
语音片段 F (1, 32)、语音片段 F (1, 450)	甲
语音片段 F (2, 100)、语音片段 F (2, 400)	乙
语音片段 F (3, 900)、语音片段 F (3, 600)	丙

由于语音片段 F (1, 32)、语音片段 F (1, 450) 是类别 1 中的语音片段，而类别 1 中的所有语音片段对应的是同一个发言人，因此，语音片段 F (1, 32)、语音片段 F (1, 450) 对应的发言人甲即为类别 1 中的所有语音片段对应的发言人，即，类别 1 中的所有语音片段都是发言人甲发出的；同理，由于语音片段 F (2, 100)、语音片段 F (2, 400) 是类别 2 中的语音片段，类别 2 中的所有语音片段都是发言人乙发出的；同理，由于语音片段 F (3, 900)、语音片段 F (3, 600) 是类别 3 中的语音片段，类别 3 中的所有语音片段都是发言人丙发出的，语音片段与发言人之间的对应关系如表 4 所示。

在本申请实施例中，通过根据聚类算法将同一个发言人对应的语音片段聚类到一起，然后从每个类别中随机选择一个或多个语音片段，向用户播放选择的语音片段，请用户将语音片段与发言人进行对应；或者将选择出的语音片段转换成文本片段，向用户展示文本片段，请用户将文本片段与发言人进行对应，非常简单方便，不需要事先知道发言人的声纹特征或其他声音相关的特征。

将 N 个语音片段进行聚类的具体过程如下：

S1：从 N 个语音片段中随机选择 M 个语音片段，将选择的 M 个语音片段作为 M 个类别的聚类中心；S2：对于剩余的 $N-M$ 个语音片段中的第 i 个语音片段，计算第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离，并将第 i 个语音片段归类到与第 i 个语音片段距离最近的聚类中心对应的类别中，i 依次取 1 至 $N-M$ 之间的自然数；S3：当 M 个语音片段归类完成之后，根据 M 个类别中每个类别包括的语音片段重新计算 M

个类别的聚类中心，并更新 M 个类别的聚类中心，循环执行 S2 和 S3，直到 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内。

在本申请实施例中，可采用 K-means 算法对语音片段进行聚类。 M 即为发言人的数量，该数量可以由会议主持人或其他参会人员提供。

K-means 算法是很典型的基于距离的聚类算法，采用距离作为相似性的评价指标，即认为两个对象的距离越近，其相似度就越大。该算法认为簇是由距离靠近的对象组成的，因此把得到紧凑且独立的簇作为最终目标。初始类聚类中心点的选取对聚类结果具有较大的影响，因为在该算法第一步中是随机的选取任意 k （在本申请实施例中， $k=M$ ）个对象作为初始聚类的中心，初始地代表一个簇。该算法在每次迭代中对数据集中剩余的每个对象，根据其与各个簇中心的距离将每个对象重新赋给最近的簇。当考察完所有数据对象后，一次迭代运算完成，新的聚类中心被计算出来。如果在一次迭代前后，新的质心与原质心相等或小于指定阈值，算法结束。在本申请实施例中，当 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内时，循环结束，得到聚类结果。

上述步骤 S2：对于剩余的 $N-M$ 个语音片段中的第 i 个语音片段，计算第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离，可以通过声纹特征进行计算，具体过程可以是：提取第 i 个语音片段（待聚类的语音片段）的声纹特征；提取 M 个聚类中心中的每个聚类中心的声纹特征；将第 i 个语音片段的声纹特征与 M 个聚类中心中的每个聚类中心的声纹特征进行相似度计算，将计算出的相似度作为第 i 个语音片段与聚类中心之间的距离。

由于每个人与发音有关的解剖学结构不同，并且受社会经济状况、教育水平、出生地等影响，不同人的声纹特征不完全相同。本申请实施例中提取的声纹特征可以为韵律特征。音色、音强、音高等，总称为语音的韵律特征，又称超音段特征。音强显示语音的重音、轻音等强弱变化，音高表现语音的字调与语调。

在本申请实施例中，提取语音片段的声纹特征，通过声纹特征对语音片段进行聚类，将声纹特征相似度高的语音片段聚在一起，作为同一个发言人发出的语音片段，在这个过程中，并不需要预先知道发言人的声纹特征，更不需要预先存储发言人的声纹特征，保护了发言人的隐私，安全性高，用户体验好。

可选地，将会议语音进行分割，得到 N 个语音片段，包括：确定会议语音中的静音片段；去除会议语音中的静音片段；根据静音片段对去除静音片段后的会议语音进行分割，得到 W 个长语音片段， W 为大于等于 2 的自然数， $W < N$ ；提取 W 个长语音片段中每一个长语音片段的声学特征；对 W 个长语音片段中每一个长语音片段的声学特征进行相对熵分析；根据相对熵分析的结果对 W 个长语音片段进行切分，得到 N 个语音片段。

可选地，对每一个长语音段的声学特征进行相对熵分析；根据相对熵分析的结果对长语音段进行切分，包括：对长语音段进行分帧，得到长语音段的语音帧，提取语音帧的声学特征，对声学特征进行相对熵分析，确定相对熵的最大值处，判断长语音段的时长是否大于预设时长；如果长语音段的时长大于预设时长，在相对熵的最大值处对长语音段进行切分。

在概率论或信息论中，相对熵 (relative entropy)，又称 KL 散度 (Kullback-Leibler divergence)，是描述两个概率分布 P 和 Q 差异的一种方法。它是非对称的，这意味着 $D(P||Q) \neq D(Q||P)$ 。特别的，在信息论中， $D(P||Q)$ 表示当用概率分布 Q 来拟合真实分布 P 时，产生的信息损耗，其中 P 表示真实分布， Q 表示 P 的拟合分布。

对一个离散随机变量的两个概率分布 P 和 Q 来说，它们的 KL 散度定义为： $D(P||Q) = \sum P(i) \ln P(i)/Q(i)$ ，对于连续的随机变量，定义类似。

图 2 是根据本申请实施例一种可选的会议记录生成装置的示意图，该装置用于执行上述会议记录生成方法，如图 2 所示，该装置包括：获取单元 10、分割单元 20、聚类单元 30、第一确定单元 40、第二确定单元 50、生成单元 60。

获取单元 10，用于获取会议语音。

分割单元 20，用于将会议语音进行分割，得到 N 个语音片段， N 为大于等于 2 的自然数。

聚类单元 30，用于将 N 个语音片段进行聚类，得到 M 个类别的语音片段， M 为大于等于 2 的自然数， $M \leq N$ ， M 个类别的语音片段分别与 M 个发言人具有一一对应关系。

第一确定单元 40，用于确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

第二确定单元 50，用于根据 M 个类别的语音片段确定 M 个发言人中每个发言人的发言内容。

生成单元 60，用于根据 M 个发言人中每个发言人的发言内容生成会议记录。

在本申请实施例中，将会议语音进行分割，得到 N 个语音片段，将 N 个语音片段进行聚类，得到 M 个类别的语音片段，确定每个类别的语音片段对应的发言人；根据 M 个类别的语音片段确定 M 个发言人的发言内容；根据各个发言人的发言内容，生成会议记录，解决了现有技术中人工整理会议记录费时费力、效率低的问题，达到了智能分析会议上的发言内容，高效整理出会议记录的效果。

可选地，第一确定单元 40 包括：第一选择子单元、第一展示子单元、第一接收子单元、第一确定子单元。第一选择子单元，用于从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段，得到 L 个文本片段， L 为自然数， $L \geq M$ 。第一展示子单元，用

于向用户展示 L 个文本片段和发言人列表，发言人列表包括 M 个发言人中每个发言人的信息。第一接收子单元，用于接收匹配指令，匹配指令为用户发出的用于指示将 L 个文本片段中每个文本片段与发言人进行匹配的指令。第一确定子单元，用于根据匹配指令确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

可选地，第一确定单元 40 包括：第二选择子单元、第二展示子单元、第二接收子单元、第二确定子单元。第二选择子单元，用于从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段，得到 Z 个语音片段， Z 为自然数， $Z \geq M$ 。第二展示子单元，用于向用户播放选择出的 Z 个语音片段并展示发言人列表，发言人列表包括 M 个发言人中每个发言人的信息。第二接收子单元，用于接收匹配指令，匹配指令为用户发出的用于指示将 Z 个语音片段中每个语音片段与发言人进行匹配的指令。第二确定子单元，用于根据匹配指令确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

可选地，聚类单元用于执行以下步骤：S1：从 N 个语音片段中随机选择 M 个语音片段，将选择的 M 个语音片段作为 M 个类别的聚类中心。S2：对于剩余的 $N-M$ 个语音片段中的第 i 个语音片段，计算第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离，并将第 i 个语音片段归类到与第 i 个语音片段距离最近的聚类中心对应的类别中， i 依次取 1 至 $N-M$ 之间的自然数。S3：当 M 个语音片段归类完成之后，根据 M 个类别中每个类别包括的语音片段重新计算 M 个类别的聚类中心，并更新 M 个类别的聚类中心。循环执行 S2 和 S3，直到 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内。

可选地，分割单元 20 包括：第三确定子单元、去除子单元、分割子单元、提取子单元、相对熵分析子单元、切分子单元。第三确定子单元，用于确定会议语音中的静音片段。去除子单元，用于去除会议语音中的静音片段。分割子单元，用于根据静音片段对去除静音片段后的会议语音进行分割，得到 W 个长语音片段， W 为大于等于 2 的自然数， $W < N$ 。提取子单元，用于提取 W 个长语音片段中每一个长语音片段的声学特征。相对熵分析子单元，用于对 W 个长语音片段中每一个长语音片段的声学特征进行相对熵分析。切分子单元，用于根据相对熵分析的结果对 W 个长语音片段进行切分，得到 N 个语音片段。

一方面，本申请实施例提供了一种存储介质，存储介质包括存储的程序，其中，在程序运行时控制存储介质所在设备执行以下步骤：获取会议语音；将会议语音进行分割，得到 N 个语音片段， N 为大于等于 2 的自然数；将 N 个语音片段进行聚类，得到 M 个类别的语音片段， M 为大于等于 2 的自然数， $M \leq N$ ， M 个类别的语音片段分别与 M 个发言人具有一一对应关系；确定 M 个类别的语音片段中每个类别的语音片段对应的发言人；根据 M 个类别的语音片段确定 M 个发言人中每个发言人的发言内容；

根据 M 个发言人中每个发言人的发言内容生成会议记录。

可选地，在程序运行时控制存储介质所在设备还执行以下步骤：从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段，得到 L 个文本片段，L 为自然数， $L \geq M$ ；向用户展示 L 个文本片段和发言人列表，发言人列表包括 M 个发言人中每个发言人的信息；接收匹配指令，匹配指令为用户发出的用于指示将 L 个文本片段中每个文本片段与发言人进行匹配的指令；根据匹配指令确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

可选地，在程序运行时控制存储介质所在设备还执行以下步骤：从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段，得到 Z 个语音片段，Z 为自然数， $Z \geq M$ ；向用户播放选择出的 Z 个语音片段并展示发言人列表，发言人列表包括 M 个发言人中每个发言人的信息；接收匹配指令，匹配指令为用户发出的用于指示将 Z 个语音片段中每个语音片段与发言人进行匹配的指令；根据匹配指令确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

可选地，在程序运行时控制存储介质所在设备还执行以下步骤：S1：从 N 个语音片段中随机选择 M 个语音片段，将选择的 M 个语音片段作为 M 个类别的聚类中心；S2：对于剩余的 N-M 个语音片段中的第 i 个语音片段，计算第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离，并将第 i 个语音片段归类到与第 i 个语音片段距离最近的聚类中心对应的类别中，i 依次取 1 至 N-M 之间的自然数；S3：当 M 个语音片段归类完成之后，根据 M 个类别中每个类别包括的语音片段重新计算 M 个类别的聚类中心，并更新 M 个类别的聚类中心，循环执行 S2 和 S3，直到 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内。

可选地，在程序运行时控制存储介质所在设备还执行以下步骤：确定会议语音中的静音片段；去除会议语音中的静音片段；根据静音片段对去除静音片段后的会议语音进行分割，得到 W 个长语音片段，W 为大于等于 2 的自然数， $W < N$ ；提取 W 个长语音片段中每一个长语音片段的声学特征；对 W 个长语音片段中每一个长语音片段的声学特征进行相对熵分析；根据相对熵分析的结果对 W 个长语音片段进行切分，得到 N 个语音片段。

一方面，本申请实施例提供了一种计算机设备，包括存储器和处理器，存储器用于存储包括程序指令的信息，处理器用于控制程序指令的执行，程序指令被处理器加载并执行时实现以下步骤：获取会议语音；将会议语音进行分割，得到 N 个语音片段，N 为大于等于 2 的自然数；将 N 个语音片段进行聚类，得到 M 个类别的语音片段，M 为大于等于 2 的自然数， $M \leq N$ ，M 个类别的语音片段分别与 M 个发言人具有一一对应关系；确定 M 个类别的语音片段中每个类别的语音片段对应的发言人；根据 M 个类别的语音片段确定 M 个发言人中每个发言人的发言内容；根据 M 个发言

人中每个发言人的发言内容生成会议记录。

可选地，程序指令被处理器加载并执行时还实现以下步骤：从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段，得到 L 个文本片段， L 为自然数， $L \geq M$ ；向用户展示 L 个文本片段和发言人列表，发言人列表包括 M 个发言人中每个发言人的信息；接收匹配指令，匹配指令为用户发出的用于指示将 L 个文本片段中每个文本片段与发言人进行匹配的指令；根据匹配指令确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

可选地，程序指令被处理器加载并执行时还实现以下步骤：从 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段，得到 Z 个语音片段， Z 为自然数， $Z \geq M$ ；向用户播放选择出的 Z 个语音片段并展示发言人列表，发言人列表包括 M 个发言人中每个发言人的信息；接收匹配指令，匹配指令为用户发出的用于指示将 Z 个语音片段中每个语音片段与发言人进行匹配的指令；根据匹配指令确定 M 个类别的语音片段中每个类别的语音片段对应的发言人。

可选地，程序指令被处理器加载并执行时还实现以下步骤：S1：从 N 个语音片段中随机选择 M 个语音片段，将选择的 M 个语音片段作为 M 个类别的聚类中心；S2：对于剩余的 $N-M$ 个语音片段中的第 i 个语音片段，计算第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离，并将第 i 个语音片段归类到与第 i 个语音片段距离最近的聚类中心对应的类别中， i 依次取 1 至 $N-M$ 之间的自然数；S3：当 M 个语音片段归类完成之后，根据 M 个类别中每个类别包括的语音片段重新计算 M 个类别的聚类中心，并更新 M 个类别的聚类中心，循环执行 S2 和 S3，直到 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内。

可选地，程序指令被处理器加载并执行时还实现以下步骤：确定会议语音中的静音片段；去除会议语音中的静音片段；根据静音片段对去除静音片段后的会议语音进行分割，得到 W 个长语音片段， W 为大于等于 2 的自然数， $W < N$ ；提取 W 个长语音片段中每一个长语音片段的声学特征；对 W 个长语音片段中每一个长语音片段的声学特征进行相对熵分析；根据相对熵分析的结果对 W 个长语音片段进行切分，得到 N 个语音片段。

图 3 是本申请实施例提供的一种计算机设备的示意图。如图 3 所示，该实施例的计算机设备 50 包括：处理器 51、存储器 52 以及存储在存储器 52 中并可在处理器 51 上运行的计算机程序 53，该计算机程序 53 被处理器 51 执行时实现实施例中的会议记录生成方法，为避免重复，此处不一一赘述。或者，该计算机程序被处理器 51 执行时实现实施例中会议记录生成装置中各模型/单元的功能，为避免重复，此处不一一赘述。

计算机设备 50 可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。计算机设备可包括，但不仅限于，处理器 51、存储

器 52。本领域技术人员可以理解，图 3 仅仅是计算机设备 50 的示例，并不构成对计算机设备 50 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，例如计算机设备还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 51 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。

存储器 52 可以是计算机设备 50 的内部存储单元，例如计算机设备 50 的硬盘或内存。存储器 52 也可以是计算机设备 50 的外部存储设备，例如计算机设备 50 上配备的插接式硬盘，智能存储卡 (Smart Media Card, SMC)，安全数字 (Secure Digital, SD) 卡，闪存卡 (Flash Card) 等。进一步地，存储器 52 还可以既包括计算机设备 50 的内部存储单元也包括外部存储设备。存储器 52 用于存储计算机程序以及计算机设备所需的其它程序和数据。存储器 52 还可以用于暂时地存储已经输出或者将要输出的数据。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

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，所述单元的划分，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，例如，多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元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单元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。

以上所述仅为本申请的较佳实施例而已，并不用以限制本申请，凡在本申请的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本申请保护的范围之内。

权 利 要 求 书

1、一种会议记录生成方法，其特征在于，所述方法包括：

获取会议语音；

将所述会议语音进行分割，得到 N 个语音片段， N 为大于等于 2 的自然数；

将所述 N 个语音片段进行聚类，得到 M 个类别的语音片段， M 为大于等于 2 的自然数， $M \leq N$ ，所述 M 个类别的语音片段分别与 M 个发言人具有一一对应关系；

确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人；

根据所述 M 个类别的语音片段确定所述 M 个发言人中每个发言人的发言内容；

根据所述 M 个发言人中每个发言人的发言内容生成会议记录。

2、根据权利要求 1 所述的方法，其特征在于，所述确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人，包括：

从所述 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段，得到 L 个文本片段， L 为自然数， $L \geq M$ ；

向用户展示所述 L 个文本片段和发言人列表，所述发言人列表包括所述 M 个发言人中每个发言人的信息；

接收匹配指令，所述匹配指令为所述用户发出的用于指示将所述 L 个文本片段中每个文本片段与发言人进行匹配的指令；

根据所述匹配指令确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人。

3、根据权利要求 1 所述的方法，其特征在于，所述确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人，包括：

从所述 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段，得到 Z 个语音片段， Z 为自然数， $Z \geq M$ ；

向用户播放选择出的所述 Z 个语音片段并展示发言人列表，所述发言人列表包括所述 M 个发言人中每个发言人的信息；

接收匹配指令，所述匹配指令为所述用户发出的用于指示将所述 Z 个语音片段中每个语音片段与发言人进行匹配的指令；

根据所述匹配指令确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人。

4、根据权利要求 1 所述的方法，其特征在于，所述将所述 N 个语音片段进行聚类，包括：

S1：从所述 N 个语音片段中随机选择 M 个语音片段，将选择的 M 个语音片段作为 M 个类别的聚类中心；

S2: 对于剩余的 $N-M$ 个语音片段中的第 i 个语音片段, 计算所述第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离, 并将所述第 i 个语音片段归类到与所述第 i 个语音片段距离最近的聚类中心对应的类别中, i 依次取 1 至 $N-M$ 之间的自然数;

S3: 当所述 M 个语音片段归类完成之后, 根据所述 M 个类别中每个类别包括的语音片段重新计算所述 M 个类别的聚类中心, 并更新所述 M 个类别的聚类中心,

循环执行 S2 和 S3, 直到所述 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内。

5、根据权利要求 1 至 4 任一项所述的方法, 其特征在于, 所述将所述会议语音进行分割, 得到 N 个语音片段, 包括:

确定所述会议语音中的静音片段;

去除所述会议语音中的静音片段;

根据所述静音片段对去除所述静音片段后的会议语音进行分割, 得到 W 个长语音片段, W 为大于等于 2 的自然数, $W < N$;

提取所述 W 个长语音片段中每一个长语音片段的声学特征;

对所述 W 个长语音片段中每一个长语音片段的声学特征进行相对熵分析;

根据相对熵分析的结果对所述 W 个长语音片段进行切分, 得到所述 N 个语音片段。

6、一种会议记录生成装置, 其特征在于, 所述装置包括:

获取单元, 用于获取会议语音;

分割单元, 用于将所述会议语音进行分割, 得到 N 个语音片段, N 为大于等于 2 的自然数;

聚类单元, 用于将所述 N 个语音片段进行聚类, 得到 M 个类别的语音片段, M 为大于等于 2 的自然数, $M \leq N$, 所述 M 个类别的语音片段分别与 M 个发言人具有一一对应关系;

第一确定单元, 用于确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人;

第二确定单元, 用于根据所述 M 个类别的语音片段确定所述 M 个发言人中每个发言人的发言内容;

生成单元, 用于根据所述 M 个发言人中每个发言人的发言内容生成会议记录。

7、根据权利要求 6 所述的装置, 其特征在于, 所述第一确定单元包括:

第一选择子单元, 用于从所述 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段, 得到 L 个文本片段, L 为自然数, $L \geq M$;

第一展示子单元,用于向用户展示所述 L 个文本片段和发言人列表,所述发言人列表包括所述 M 个发言人中每个发言人的信息;

第一接收子单元,用于接收匹配指令,所述匹配指令为所述用户发出的用于指示将所述 L 个文本片段中每个文本片段与发言人进行匹配的指令;

第一确定子单元,用于根据所述匹配指令确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人。

8、根据权利要求 6 所述的装置,其特征在于,所述第一确定单元包括:

第二选择子单元,用于从所述 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段,得到 Z 个语音片段, Z 为自然数, $Z \geq M$;

第二展示子单元,用于向用户播放选择出的所述 Z 个语音片段并展示发言人列表,所述发言人列表包括所述 M 个发言人中每个发言人的信息;

第二接收子单元,用于接收匹配指令,所述匹配指令为所述用户发出的用于指示将所述 Z 个语音片段中每个语音片段与发言人进行匹配的指令;

第二确定子单元,用于根据所述匹配指令确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人。

9、根据权利要求 6 所述的装置,其特征在于,所述聚类单元用于执行以下步骤:

S1:从所述 N 个语音片段中随机选择 M 个语音片段,将选择的 M 个语音片段作为 M 个类别的聚类中心;

S2:对于剩余的 N-M 个语音片段中的第 i 个语音片段,计算所述第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离,并将所述第 i 个语音片段归类到与所述第 i 个语音片段距离最近的聚类中心对应的类别中, i 依次取 1 至 N-M 之间的自然数;

S3:当所述 M 个语音片段归类完成之后,根据所述 M 个类别中每个类别包括的语音片段重新计算所述 M 个类别的聚类中心,并更新所述 M 个类别的聚类中心,

循环执行 S2 和 S3,直到所述 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内。

10、根据权利要求 6 至 9 任一项所述的装置,其特征在于,所述分割单元包括:

第三确定子单元,用于确定所述会议语音中的静音片段;

去除子单元,用于去除所述会议语音中的静音片段;

分割子单元,用于根据所述静音片段对去除所述静音片段后的会议语音进行分割,得到 W 个长语音片段, W 为大于等于 2 的自然数, $W < N$;

提取子单元，用于提取所述 W 个长语音片段中每一个长语音片段的声学特征；

相对熵分析子单元，用于对所述 W 个长语音片段中每一个长语音片段的声学特征进行相对熵分析；

切分子单元，用于根据相对熵分析的结果对所述 W 个长语音片段进行切分，得到所述 N 个语音片段。

11、一种存储介质，其特征在于，所述存储介质包括存储的程序，其中，在所述程序运行时控制所述存储介质所在设备执行以下步骤：

获取会议语音；

将所述会议语音进行分割，得到 N 个语音片段，N 为大于等于 2 的自然数；

将所述 N 个语音片段进行聚类，得到 M 个类别的语音片段，M 为大于等于 2 的自然数， $M \leq N$ ，所述 M 个类别的语音片段分别与 M 个发言人具有一一对应关系；

确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人；

根据所述 M 个类别的语音片段确定所述 M 个发言人中每个发言人的发言内容；

根据所述 M 个发言人中每个发言人的发言内容生成会议记录。

12、根据权利要求 11 所述的存储介质，其特征在于，在所述程序运行时控制所述存储介质所在设备执行所述确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人，包括：

从所述 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段，得到 L 个文本片段，L 为自然数， $L \geq M$ ；

向用户展示所述 L 个文本片段和发言人列表，所述发言人列表包括所述 M 个发言人中每个发言人的信息；

接收匹配指令，所述匹配指令为所述用户发出的用于指示将所述 L 个文本片段中每个文本片段与发言人进行匹配的指令；

根据所述匹配指令确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人。

13、根据权利要求 11 所述的存储介质，其特征在于，在所述程序运行时控制所述存储介质所在设备执行所述确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人，包括：

从所述 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段，得到 Z 个语音片段，Z 为自然数， $Z \geq M$ ；

向用户播放选择出的所述 Z 个语音片段并展示发言人列表，所述发言人列表包括所述 M 个发言人中每个发言人的信息；

接收匹配指令，所述匹配指令为所述用户发出的用于指示将所述 Z 个语音片段中每个语音片段与发言人进行匹配的指令；

根据所述匹配指令确定所述 M 个类别的语音片段中每个类别的语音

片段对应的发言人。

14、根据权利要求 11 所述的存储介质，其特征在于，在所述程序运行时控制所述存储介质所在设备执行所述将所述 N 个语音片段进行聚类，包括：

S1：从所述 N 个语音片段中随机选择 M 个语音片段，将选择的 M 个语音片段作为 M 个类别的聚类中心；

S2：对于剩余的 N-M 个语音片段中的第 i 个语音片段，计算所述第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离，并将所述第 i 个语音片段归类到与所述第 i 个语音片段距离最近的聚类中心对应的类别中，i 依次取 1 至 N-M 之间的自然数；

S3：当所述 M 个语音片段归类完成之后，根据所述 M 个类别中每个类别包括的语音片段重新计算所述 M 个类别的聚类中心，并更新所述 M 个类别的聚类中心，

循环执行 S2 和 S3，直到所述 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内。

15、根据权利要求 11 至 14 任一项所述的存储介质，其特征在于，在所述程序运行时控制所述存储介质所在设备执行所述将所述会议语音进行分割，得到 N 个语音片段，包括：

确定所述会议语音中的静音片段；

去除所述会议语音中的静音片段；

根据所述静音片段对去除所述静音片段后的会议语音进行分割，得到 W 个长语音片段，W 为大于等于 2 的自然数， $W < N$ ；

提取所述 W 个长语音片段中每一个长语音片段的声学特征；

对所述 W 个长语音片段中每一个长语音片段的声学特征进行相对熵分析；

根据相对熵分析的结果对所述 W 个长语音片段进行切分，得到所述 N 个语音片段。

16、一种计算机设备，包括存储器和处理器，所述存储器用于存储包括程序指令的信息，所述处理器用于控制程序指令的执行，其特征在于，所述程序指令被处理器加载并执行时实现以下步骤：

获取会议语音；

将所述会议语音进行分割，得到 N 个语音片段，N 为大于等于 2 的自然数；

将所述 N 个语音片段进行聚类，得到 M 个类别的语音片段，M 为大于等于 2 的自然数， $M \leq N$ ，所述 M 个类别的语音片段分别与 M 个发言人具有一一对应关系；

确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人；

根据所述 M 个类别的语音片段确定所述 M 个发言人中每个发言人的发言内容；

根据所述 M 个发言人中每个发言人的发言内容生成会议记录。

17、根据权利要求 16 所述的计算机设备，其特征在于，所述程序指令被处理器加载并执行时实现所述确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人，包括：

从所述 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段转换成文本片段，得到 L 个文本片段，L 为自然数， $L \geq M$ ；

向用户展示所述 L 个文本片段和发言人列表，所述发言人列表包括所述 M 个发言人中每个发言人的信息；

接收匹配指令，所述匹配指令为所述用户发出的用于指示将所述 L 个文本片段中每个文本片段与发言人进行匹配的指令；

根据所述匹配指令确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人。

18、根据权利要求 16 所述的计算机设备，其特征在于，所述程序指令被处理器加载并执行时实现所述确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人，包括：

从所述 M 个类别的语音片段中每个类别的语音片段中各选择至少一个语音片段，得到 Z 个语音片段，Z 为自然数， $Z \geq M$ ；

向用户播放选择出的所述 Z 个语音片段并展示发言人列表，所述发言人列表包括所述 M 个发言人中每个发言人的信息；

接收匹配指令，所述匹配指令为所述用户发出的用于指示将所述 Z 个语音片段中每个语音片段与发言人进行匹配的指令；

根据所述匹配指令确定所述 M 个类别的语音片段中每个类别的语音片段对应的发言人。

19、根据权利要求 16 所述的计算机设备，其特征在于，所述程序指令被处理器加载并执行时实现所述将所述 N 个语音片段进行聚类，包括：

S1：从所述 N 个语音片段中随机选择 M 个语音片段，将选择的 M 个语音片段作为 M 个类别的聚类中心；

S2：对于剩余的 N-M 个语音片段中的第 i 个语音片段，计算所述第 i 个语音片段与 M 个聚类中心中每个聚类中心之间的距离，并将所述第 i 个语音片段归类到与所述第 i 个语音片段距离最近的聚类中心对应的类别中，i 依次取 1 至 N-M 之间的自然数；

S3：当所述 M 个语音片段归类完成之后，根据所述 M 个类别中每个类别包括的语音片段重新计算所述 M 个类别的聚类中心，并更新所述 M 个类别的聚类中心，

循环执行 S2 和 S3，直到所述 M 个类别中每个类别的相邻两次聚类中心的距离在预设距离之内。

20、根据权利要求 16 至 19 任一项所述的计算机设备，其特征在于，所述程序指令被处理器加载并执行时实现所述将所述会议语音进行分割，得到 N 个语音片段，包括：

确定所述会议语音中的静音片段；

去除所述会议语音中的静音片段；

根据所述静音片段对去除所述静音片段后的会议语音进行分割，得到W个长语音片段，W为大于等于2的自然数， $W < N$ ；

提取所述W个长语音片段中每一个长语音片段的声学特征；

对所述W个长语音片段中每一个长语音片段的声学特征进行相对熵分析；

根据相对熵分析的结果对所述W个长语音片段进行切分，得到所述N个语音片段。

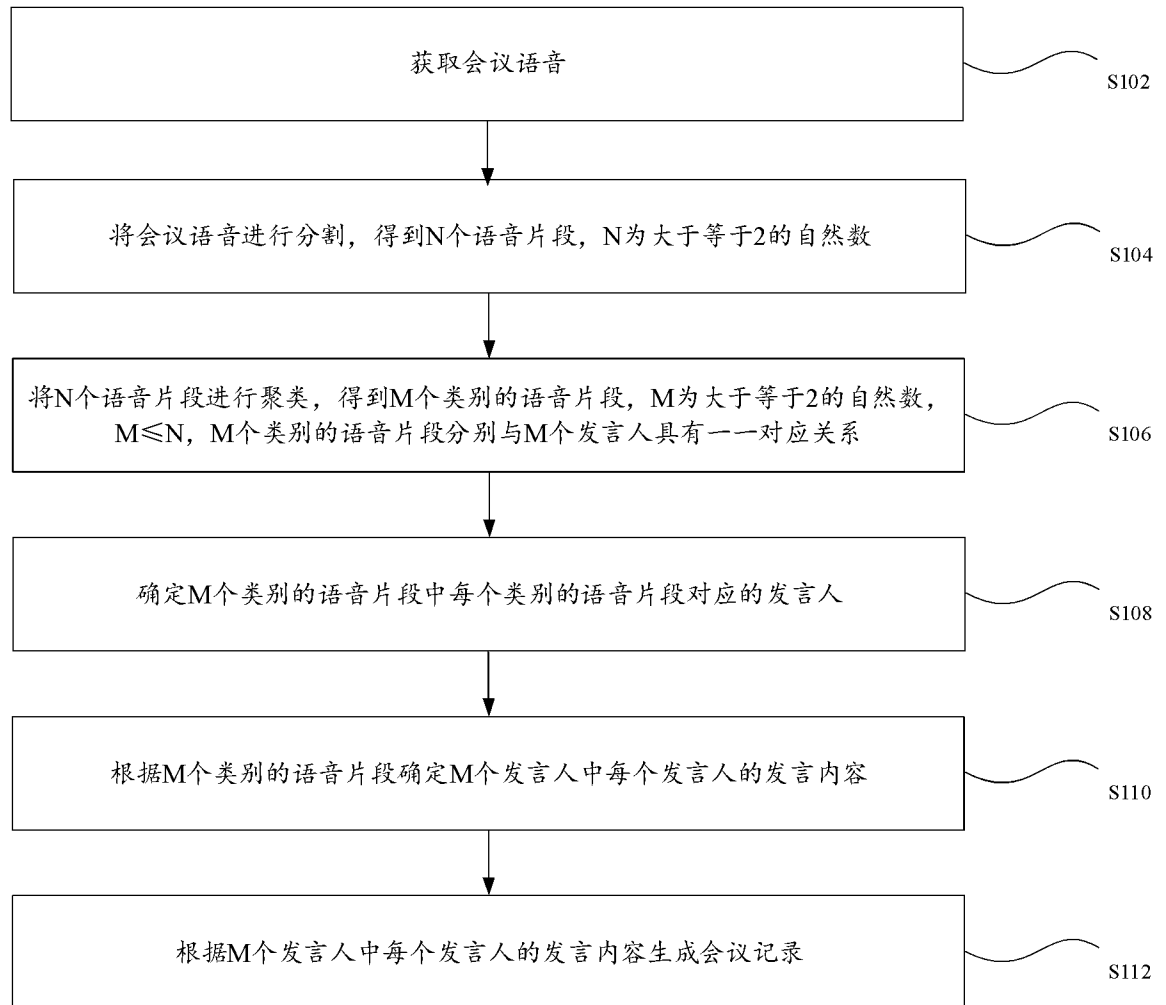


图 1

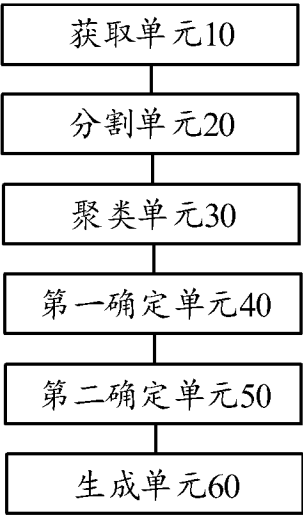


图 2

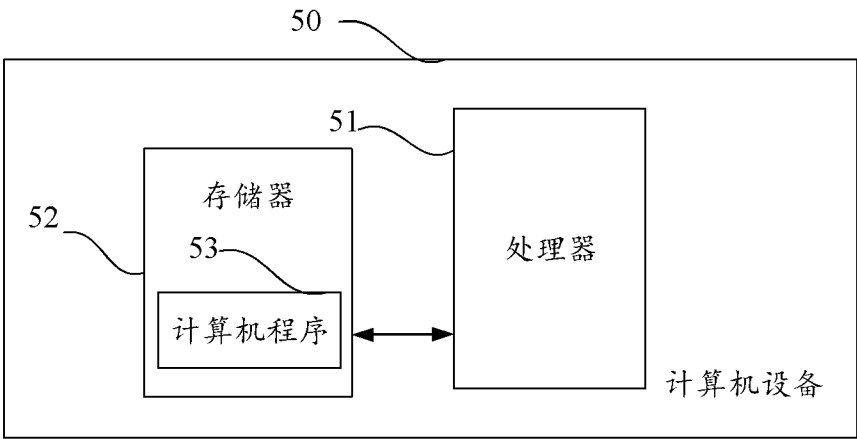


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/118256

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/04(2013.01)i; G10L 15/08(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

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)

CNPAT, WPI, EPODOC, IEEE, CNKI: 会议, 语音, 声音, 分割, 片段, 聚类, 记录, 发言人, 类别, 静音, conference, speech, voice, segmentation, segment, cluster, record, speaker, category, mute

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109767757 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 17 May 2019 (2019-05-17) claims 1-10, and description, paragraphs [0004]-[0015]	1-20
X	CN 103530432 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 22 January 2014 (2014-01-22) description, paragraphs [0006]-[0032]	1-20
A	CN 105810207 A (FUTAIHUA INDUSTRIAL (SHENZHEN) CO., LTD. et al.) 27 July 2016 (2016-07-27) entire document	1-20
A	CN 103559882 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 05 February 2014 (2014-02-05) entire document	1-20
A	US 2018075860 A1 (NUANCE COMMUNICATIONS, INC.) 15 March 2018 (2018-03-15) 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

11 February 2020

Date of mailing of the international search report

17 February 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/CN2019/118256

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109767757	A	17 May 2019	None			
CN	103530432	A	22 January 2014	None			
CN	105810207	A	27 July 2016	None			
CN	103559882	A	05 February 2014	CN	103559882	B	10 August 2016
US	2018075860	A1	15 March 2018	US	10424317	B2	24 September 2019
				EP	3513404	A1	24 July 2019
				CN	110178178	A	27 August 2019
				WO	2018053077	A1	22 March 2018

国际检索报告

国际申请号

PCT/CN2019/118256

A. 主题的分类 G10L 15/04(2013.01)i; G10L 15/08(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 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)) CNPAT, WPI, EPDOC, IEEE, CNKI: 会议, 语音, 声音, 分割, 片段, 聚类, 记录, 发言人, 类别, 静音, conference, speech, voice, segmentation, segment, cluster, record, speaker, category, mute		
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 109767757 A (平安科技深圳有限公司) 2019年 5月 17日 (2019 - 05 - 17) 权利要求1-10, 说明书第[0004]-[0015]段	1-20
X	CN 103530432 A (华南理工大学) 2014年 1月 22日 (2014 - 01 - 22) 说明书第[0006]-[0032]段	1-20
A	CN 105810207 A (富泰华工业深圳有限公司 等) 2016年 7月 27日 (2016 - 07 - 27) 全文	1-20
A	CN 103559882 A (华南理工大学) 2014年 2月 5日 (2014 - 02 - 05) 全文	1-20
A	US 2018075860 A1 (NUANCE COMMUNICATIONS, INC.) 2018年 3月 15日 (2018 - 03 - 15) 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 11日		国际检索报告邮寄日期 2020年 2月 17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 石爽 电话号码 86-(10)-53961413

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/118256

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	109767757	A	2019年 5月 17日	无			
CN	103530432	A	2014年 1月 22日	无			
CN	105810207	A	2016年 7月 27日	无			
CN	103559882	A	2014年 2月 5日	CN	103559882	B	2016年 8月 10日
US	2018075860	A1	2018年 3月 15日	US	10424317	B2	2019年 9月 24日
				EP	3513404	A1	2019年 7月 24日
				CN	110178178	A	2019年 8月 27日
				WO	2018053077	A1	2018年 3月 22日