

(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



(10) 国际公布号

WO 2020/147256 A1

(43) 国际公布日

2020 年 7 月 23 日 (23.07.2020)

(51) 国际专利分类号:

G06F 16/61 (2019.01)

(21) 国际申请号:

PCT/CN2019/091098

(22) 国际申请日:

2019 年 6 月 13 日 (13.06.2019)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201910038369.4 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: 胡燕(HU, Yan); 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)。徐媛(XU, Yuan); 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); 中国北京市海淀区高粱桥斜街 59 号中坤大厦 603 室, Beijing 100044 (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, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: CONFERENCE CONTENT DISTINGUISHING METHOD AND APPARATUS, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 会议内容区分方法、装置、计算机设备及存储介质

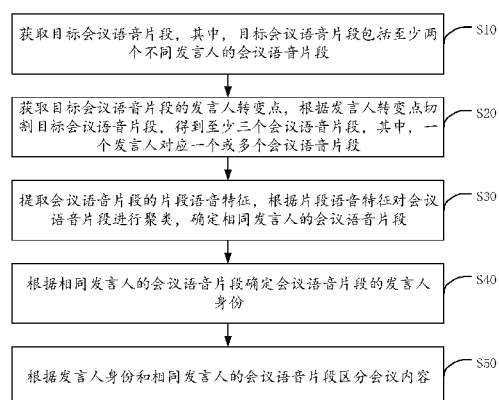


图 1

- S10 Obtain a target conference voice segment, wherein the target conference voice segment comprises conference voice segments of at least two different speakers
- S20 Obtain a speaker transition point of the target conference voice segment, and cut the target conference voice segment according to the speaker transition point to obtain at least three conference voice segments, wherein one speaker corresponds to one or more conference voice segments
- S30 Extract segment voice features of the conference voice segments, cluster the conference voice segments according to the segment voice features, and determine conference voice segments of the same speaker
- S40 Determine the speaker identity of the conference voice segments according to the conference voice segments of the same speaker
- S50 Distinguish conference contents according to the speaker identities and the conference voice segments of the same speaker

(57) Abstract: A conference content distinguishing method and apparatus, and a computer device and a storage medium, relating to the field of artificial intelligence. The conference content distinguishing method comprises: obtaining a target conference voice segment, wherein the target conference voice segment comprises conference voice segments of at least two different speakers (S10); obtaining a speaker transition point of the target conference voice segment, and cutting the target conference voice segment according to the speaker transition point to obtain at least three conference voice segments, wherein one speaker corresponds to one or more conference voice segments (S20); extracting segment voice features of the conference voice segments, clustering the conference voice segments according to the segment voice features, and determining conference voice segments of the same speaker (S30); determining the speaker identity of the conference voice segments according to the conference voice segments of the same speaker (S40); and distinguishing conference contents according to the speaker identities and the conference voice segments of the same speaker (S50). By adopting the

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))。

conference content distinguishing method, the conference content can be efficiently distinguished.

(57) 摘要: 一种会议内容区分方法、装置、计算机设备及存储介质, 涉及人工智能领域。该会议内容区分方法包括: 获取目标会议语音片段, 其中, 目标会议语音片段包括至少两个不同发言人的会议语音片段 (S10); 获取目标会议语音片段的发言人转变点, 根据发言人转变点切割目标会议语音片段, 得到至少三个会议语音片段, 其中, 一个发言人对应一个或多个会议语音片段 (S20); 提取会议语音片段的片段语音特征, 根据片段语音特征对会议语音片段进行聚类, 确定相同发言人的会议语音片段 (S30); 根据相同发言人的会议语音片段确定会议语音片段的发言人身份 (S40); 根据发言人身份和相同发言人的会议语音片段区分会议内容 (S50)。采用该会议内容区分方法能够对会议内容进行高效区分。

会议内容区分方法、装置、计算机设备及存储介质

本申请以 2019 年 1 月 16 日提交的申请号为 201910038369.4，名称为“会议内容区分方法、装置、计算机设备及存储介质”的中国发明专利申请为基础，并要求其优先权。

【技术领域】

本申请涉及人工智能领域，尤其涉及一种会议内容区分方法、装置、计算机设备及存储介质。

【背景技术】

对会议内容进行高效整理一直是个难题。目前，大多数会议内容都是通过人工进行整理的，也有少部分采用语音识别技术，由机器识别发言人的语音并转换为文字记录。但是，机器只能单纯实现从语音到文字上的转换，并不能够对会议内容进行区分、整理。

【发明内容】

有鉴于此，本申请实施例提供了一种会议内容区分方法、装置、计算机设备及存储介质，用以解决难以对会议内容进行高效区分的问题。

第一方面，本申请实施例提供了一种会议内容区分方法，包括：

获取目标会议语音片段，其中，所述目标会议语音片段包括至少两个不同发言人的会议语音片段；

获取所述目标会议语音片段的发言人转变点，根据所述发言人转变点切割所述目标会议语音片段，得到至少三个会议语音片段，其中，一个所述发言人对应一个或多个所述会议语音片段；

提取所述会议语音片段的片段语音特征，根据所述片段语音特征对所述会议语音片段进行聚类，确定相同发言人的会议语音片段；

根据所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份；

根据所述发言人身份和所述相同发言人的会议语音片段区分会议内容。

第二方面，本申请实施例提供了一种会议内容区分装置，包括：

目标片段获取模块，用于获取目标会议语音片段，其中，所述目标会议语音片段包括至少两个不同发言人的会议语音片段；

会议语音片段获取模块，用于获取所述目标会议语音片段的发言人转变点，根据所述发言人转变点切割所述目标会议语音片段，得到至少三个会议语音片段，其中，一个所述发言人对一个或多个所述会议语音片段；

相同发言人语音片段确定模块，用于提取所述会议语音片段的片段语音特征，根据所述片段语音特征对所述会议语音片段进行聚类，确定相同发言人的会议语音片段；

发言人身份确定模块，用于根据所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份；

区分模块，用于根据所述发言人身份和所述相同发言人的会议语音片段区分会议内容。

第三方面，一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现上述会议内容区分方法的步骤。

第四方面，本申请实施例提供了一种计算机非易失性可读存储介质，包括：计算机可执行指令，当所述计算机可执行指令被运行时，用以执行第一方面任一项所述的会议内容区分方法。

上述技术方案中的一个技术方案具有如下有益效果：

在本申请实施例中，首先将获取的目标会议语音片段根据发言人转变点进行切割，得到至少三个会议室语音片段，能够将包括至少两个不同发言人的会议语音片段的目标会议语音片段实现合理切割，使得每个得到的会议室语音片段来自一位发言人；然后提取会议语音片段的片段语音特征，根据片段语音特征所表达出的相似性对会议室语音片段聚类，根据聚类的结果确定相同发言人的会议语音片段，将会议语音片段按类别进行区分；最后根据相同发言人的会议语音片段确定每一个会议语音片段对应的发言人身份，从而根据发言人身份和相同发言人的会议语音片段确定会议内容中各个会议语音片段具体的所属情况，实现会议室内容的高效区分。

【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性的前提下，还可以根据这些附图获得其它的附图。

图 1 是本申请一实施例中基于会议内容区分方法的一流程图；

图 2 是本申请一实施例中基于会议内容区分装置的一示意图；

图 3 是本申请一实施例中计算机设备的一示意图。

【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做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相同的字段，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。

应当理解，尽管在本申请实施例中可能采用术语第一、第二、第三等来描述预设范围等，但这些预设范围不应限于这些术语。这些术语仅用来将预设范围彼此区分开。例如，在不脱离本申请实施例范围的情况下，第一预设范围也可以被称为第二预设范围，类似地，第二预设范围也可以被称为第一预设范围。

取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”或“响应于检测”。类似地，取决于语境，短语“如果确定”或“如果检测（陈述的条件或事件）”可以被解释成为“当确定时”或“响应于确定”或“当检测（陈述的条件或事件）时”或“响应于检测（陈述的条件或事件）”。

图 1 示出本实施例中会议内容区分方法的一流程图。该会议内容区分方法可应用在会议内容区分系统中，在进行现场会议、网络会议时均可采用该系统对会议内容进行高效的区分，该会议内容区分系统具体可应用在计算机设备上。其中，该计算机设备是可与用户进行人机交互的设备，包括但不限于电脑、智能手机和平板等设备。如图 1 所示，该会议内容区分方法包括如下步骤：

S10: 获取目标会议语音片段, 其中, 目标会议语音片段包括至少两个不同发言人的会议语音片段。

可以理解地, 目标会议语音片段中包括至少两个不同发言人的会议语音片段, 需要说明的是, 这里的包括至少两个不同发言人的会议语音片段在切割时至少能得到三个会议语音片段, 否则没有必要对会议内容进行区分。表示该目标会议语音片段是由至少两个不同发言人发出的会议语音片段组成的, 该目标会议语音片段是一个混杂的语音片段, 本方案的一个目的为将混杂不同发言人的会议语音片段进行区分, 确定目标会议语音片段中不同发言人各自对应的会议语音片段。

在一实施例中, 步骤 S10 中, 获取目标会议语音片段, 具体包括:

S11: 获取原始会议语音片段。

在一实施例中, 原始会议语音片段是指采用录音设备在会议上录取到的语音信息, 其中, 会议可以是参会人员在现场参与的现场会议, 也可以是参会人员通过网络在线上参与的线上会议, 当然, 也可以是参会人员在现场参与的借助网络建立的线上会议, 在此不对会议的举办形式进行限定。

可以理解地, 在采用会议内容区分系统时, 当会议的举办形式为现场会议时, 将通过与计算机设备连接的录音设备或者内嵌的录音设备采集不同发言人在会议上的发言, 在采集过程中是按时间连续进行采集, 因此会议期间的静默时段(无人发言的时段)也同样会采集下来。在会议期间录取到的语音信息为原始会议语音片段, 该原始会议语音片段包括不同时间不同发言人发出的会议语音片段, 也包括无人发言的静默片段。

可以理解地, 在会议的举办形式为线上会议时, 如采用微信群组的方式进行的线上会议, 将采用移动设备的录音模块采集语音信息, 获取原始会议语音片段。参会人员在线上会议后往往忘记部分会议内容, 从而无法达到良好的会议效果, 因此, 可以对原始会议语音片段进行处理, 区分会议内容, 让参会人员可以随时回顾所需的会议内容。

S12: 采用静音检测去除原始会议语音片段中的静默片段, 得到目标会议语音片段。

其中, 静音检测是指对原始会议语音片段中无人发言的静默片段的检测, 静默片段是指没有发言人进行发言的语音片段。在一实施例中, 可以采用语音端点检测(Voice Activity Detection, 简称 VAD) 的技术实现, 包括采用帧幅度、帧能量、短时过零率和深度神经网络等方式实现。从而准确去除原始通话片段中的静默语音片段, 将原始会议语音片段中发言人发言时的会议语音片段保留下来, 可以排除原始会议语音片段中的静默片段的干扰, 为提高区分会议内容的效率和准确率提供重要的技术基础。

特别地，在会议的举办形式为线上会议时，可以采用设置检测语音信息的短时能量值的方式去除静默片段，此时可以通过判断短时能量值是否大于预设阈值，直接获取目标会议语音片段。

S20: 获取目标会议语音片段的发言人转变点，根据发言人转变点切割目标会议语音片段，得到至少三个会议语音片段，其中，一个发言人对应一个或多个会议语音片段。

可以理解地，发言人转变点是指在目标会议语音片段中不同发言人的会议语音片段的交界点，如发言人 A 的会议语音片段和发言人 B 的会议语音片段在目标会议语音片段上相邻，那么两者之间的交界点即发言人转变点。

在一实施例中，获取目标会议语音片段的发言人转变点，具体地，基于贝叶斯信息准则和似然比，检测目标会议语音片段上的发言人转变点，其中，贝叶斯信息准则(Bayesian information criterion, 简称 BIC) 是在不完全情报下，对部分未知的状态用主观概率估计，然后用贝叶斯公式对发生概率进行修正，最后再利用期望值和修正概率做出最优决策。似然比(likelihood ratio, 简称 LR) 是反映真实性的一种指标。可以理解地，基于贝叶斯信息准则和似然比可以对目标会议语音片段上的特征点进行比较，基于贝叶斯信息准则计算目标会议语音片段上的特征点之间的似然比，从而判断特征点是否为发言人转变点。

在得到发言人转变点后，根据发言人转变点切割目标会议语音片段，得到至少三个会议语音片段，其中，一个发言人对应一个或多个会议语音片段。本实施例中，切割得到的会议语音片段对应的是某一个发言人的某次发言，也即是说，一个切割得到的会议语音片段属于某一个发言人，而不能同时属于多个发言人。可以理解地，本实施例中提及的会议语音片段为根据发言人转变点切割的、对应某一发言人的某次发言的语音片段。

在本实施例中，能够将包括至少两个不同发言人的会议语音片段的目标会议语音片段实现合理切割，使得每个得到的会议室语音片段只来自一位发言人。

S30: 提取会议语音片段的片段语音特征，根据片段语音特征对会议语音片段进行聚类，确定相同发言人的会议语音片段。

其中，片段语音特征是从会议语音片段提取的、代表会议语音片段的语音特征。

在一实施例中，通过提取会议语音片段的片段语音特征，可以按照片段语音特征间的相似程度来判断会议片段语音分别是由那几个发言人发出的。具体地，根据片段语音特征对会议语音片段进行聚类，按片段语音特征将会议语音片段归成不同的类别，每一个类别实际上对应的就是一个发言人。

在本实施例中，可以确定相同发言人的会议语音片段，将会议语音片段按类别进行区分，

使得来自相同发言人的会议语音片段归成同一类。

进一步地, 在步骤 S30 中, 具体包括:

S311: 通过预先训练的通用背景模型和高斯混合模型从会议语音片段中提取 i-vector 特征作为片段语音特征;

具体地, 从会议语音片段中提取的特征可以是 i-vector 特征, i-vector 特征是指基于通用背景模型 (Universal Background Model, 简称 UBM), 从高斯混合模型 (Gaussian mixture model, GMM) 均值超矢量中提取的一个紧凑的特征矢量, i-vector 特征除了包含说话人的身份信息外, 还包括关于声道, 话筒, 说话方式, 语音等信息, 可以全面地体现声音的声纹特征, 采用该 i-vector 特征进行聚类得到的结果更为准确, 能够提高聚类结果的准确性。

S312: 采用预先训练的双协方差概率线性判别模型对 i-vector 特征建模, 得到会议语音片段的特征表达模型;

其中, 在片段语音特征识别中, 双协方差概率线性判别分析模型用来从 i-vector 中提取说话人信息, 可以对片段语音特征进行比对和区分。双协方差概率线性判别分析模型假设 i-vector 是由另外两个参数提取的: 一个发言人的向量 y 和一个剩余向量 ϵ , 剩余向量 ϵ 表示与发言人无关的项。采用预先训练好的双协方差概率线性判别分析模型对每个片段语音特征进行建模, 能够在确定相同发言人的会议语音片段时, 达到更精确的聚类效果。

在建模前: 设在一个会议发言过程中, 发言人的总数有 S 个。将会议语音片段提取的 i-vector 表示为 $\Phi = \{\phi_1, \dots, \phi_M\}$ 。对于每一个会议语音片段 $m = 1, \dots, M$, 定义一个维度为 $S \times 1$ 的指示向量 i_m , 如果发言人 s 在会议语音片段 m 中说话了, 则 i_m 中的元素 $i_{ms} = 1$, 如果发言人 s 在会议语音片段 m 中没说话, i_m 中的元素 $i_{ms} = 0$ 。令 $I = \{i_1, \dots, i_M\}$ 为一个给出的关于会议语音片段的指示向量集合。假设事件为发言人 s 在一个片段中说话, 则给该时间赋上一个先验概率 $\pi_s = \frac{1}{S}$ 。对于每个发言人 s 的样本 $y_s \in N(y; \mu, \Lambda^{-1})$, 即每个发言人 s 的样本服从均值为 μ , 协方差为 Λ^{-1} 的正态分布, 对于每一个会议语音片段, 服从于多项式分布 $Mult(\Pi)$ 的样本 i_m , 其中 $\Pi = (\pi_1, \dots, \pi_S)$ 。

有了上述建模的前提条件, 特征表达模型的表达式为: $\phi_m = y_k + \epsilon_m$, 其中, ϕ_m 表示第 m 个会议语音片段提取的 i-vector 特征, y 表示会议语音片段中的与发言人关联向量, 为了和上述 y_s 中的 s 做区分, 令 k 为使 $i_{mk} = 1$ 的索引, i_m 表示与会议语音片段相关的指示向量, $\epsilon_m \in N(\epsilon; \bar{0}, L^{-1})$, $N(\epsilon; \bar{0}, L^{-1})$ 表示第 m 个会议语音片段的发言人无关向量 ϵ 服从均值为 0, 协方差为 L^{-1} 的高斯分布。双协方差概率线性判别分析模型中的双协方差即分别来自 y_k 和 ϵ_m 。可以理解

地，建模的过程即计算每一个会议语音片段在双协方差概率线性判别分析模型中的特征表示。通过建立每一个会议语音片段的特征表达模型，可以利用该特征表达模型确定相同发言人的会议语音片段。

S313：采用特征表达模型对会议语音片段进行聚类，确定相同发言人的会议语音片段。

在一实施例中，采用特征表达模型对会议语音片段进行聚类具体采用的可以是变分贝叶斯算法，其中，变分贝叶斯算法（Variational Bayes，简称 VB）是提供一种局部最优，但具有确定解的近似后验方法。

本实施例中，根据特征表达模型和变分贝叶斯算法获取会议语音片段的后验概率和发言人的后验概率，并对会议语音片段的后验概率和发言人的后验概率进行更新，得到发言人在会议语音片段中发过言的后验概率，从而确定相同发言人的会议语音片段。

进一步地，在步骤 S30 之前，还包括：

S321：初始化会议语音片段的后验概率中发言人的个数，将会议语音片段的后验概率中每个不同发言人作为一对。

在一实施例中，初始化会议语音片段的后验概率中发言人的个数具体可以是初始化为 3 个发言人。

S322：计算每一对发言人之间的距离，得到距离最远的两个发言人。

其中，在双协方差概率线性判别分析模型中，可以采用余弦相似度和/或似然比分数作为衡量距离的标准。

S323：重复预设次数的初始化会议语音片段的后验概率中发言人的个数，将会议语音片段的后验概率中每个不同发言人作为一对和计算每一对发言人之间的距离，得到距离最远的两个发言人的步骤，得到在预设次数的步骤中距离最远的两个发言人，并将在预设次数的步骤中距离最远的两个发言人作为变分贝叶斯计算的起点。

可以理解地，本步骤为重复预设次数（如 20 次）的步骤 S321-S322，再将所有预设次数的步骤中距离最远的两个发言人作为变分贝叶斯计算的起点。

步骤 S321-S323 中是对变分贝叶斯算法的优化步骤，可以提高变分贝叶斯算法在采用最大期望算法进行迭代时得到的运算结果更加准确，并最终根据准确地得到发言人在会议语音中说过话的后验概率，从而更好地对会议语音片段按发言人进行区分。

S40：根据相同发言人的会议语音片段确定会议语音片段的发言人身份。

可以理解地，在步骤 S30 中已将相同发言人的会议语音片段区分开来，但是无法确定发言人身份。本实施例中，根据相同发言人的会议语音片段，可以通过会议语音片段的片段语音特

征确定会议语音片段的发言人身份，从而实现在发言人身份层面上的会议语音片段区分。

进一步地，在步骤 S40 中，根据相同发言人的会议语音片段确定会议语音片段的发言人身份，包括：

S411：在每一相同发言人的会议语音片段中各获取预设个数的会议语音片段，并进行展示。

在一实施例中，对于已经聚类好的会议语音片段，可以从相同发言人的会议语音片段中随机抽取预设个数的会议语音片段，只需要保证每个相同发言人的会议语音片段至少有一个被抽出来就可以，例如：共有 3 个发言人 A、B 和 C，发言人 A 共有 5 个会议语音片段，发言人 B 共有 10 个会议语音片段，发言人 C 共有 20 个会议语音片段，这时可以分别抽取发言人 A、B 和 C 各两个会议语音片段，获取的会议语音片段个数可以预先设置好，至少要保证每个发言人有一个会议语音片段。在获取预设个数的会议语音片段后，将其进行展示，具体地，展示可以是展示给参会人员，且这些参会人员是知道会议上发言人的发言人身份的。展示的方式可以是采用播放音频的方式进行展示，参会人员可以根据展示的会议语音片段确定展示的会议语音片段所对应的发言人身份。

可以理解地，采用该方式无需预先存储发言人的发言人身份。特别是在不清楚哪些参会人员会发言的情况下，采用存储发言人的发言人身份的方式需要预先采集所有参会人员的声纹特征，从而确定发言人身份。这种方式会明显提高工作量，需要对声纹识别模型进行预训练的操作，并且，不是每个参会人员都有时间参与发言人身份的录入，在大型会议、现场会议的场景下并不合适。采用本实施例展示会议语音片段的方式更为灵活和高效。

S412：响应于展示，获取发言人身份确认指令，根据发言人身份确认指令确认预设个数的会议语音片段的发言人身份，得到第一确认结果。

其中，发言人身份确认指令为确认发言人身份的指令。

在一实施例中，会议内容区分系统获取用户输入的发言人身份确认指令，确认预设个数的会议语音片段的发言人身份。可以理解地，会议内容区分系统在展示每个相同发言人至少一个的会议语音片段后根据用户（参会人员）的发言人身份确认指令确认展示的会议语音片段所对应的发言人身份。

S413：根据第一确认结果和相同发言人的会议语音片段确定会议语音片段的发言人身份。

可以理解地，第一确认结果代表的是展示的会议语音片段，由于已采用聚类的方式将相同的会议语音片段归成同一类，因此，可以根据第一确认结果和相同发言人的会议语音片段直接确定会议语音片段的发言人身份，可以快速确定所有会议语音片段的发言人身份。

进一步地，在步骤 S40 中，根据相同发言人的会议语音片段确定会议语音片段的发言人身

份，还包括：

S421：在每一相同发言人的会议语音片段中各获取预设个数的会议语音片段，并输入到预先训练的声纹识别模型中。

在一实施例中，与步骤 S411 相比，不将每一相同发言人的会议语音片段中各获取预设个数的会议语音片段进行展示，而是采用自动识别的方式，将获取的预设个数的会议语音片段输入到声纹识别模型中，以让声纹识别模型自动识别会议语音片段的发言人身份。

S422：通过声纹识别模型识别预设个数的会议语音片段，确认预设个数的会议语音片段的发言人身份，得到第二确认结果。

在一实施例中，采用声纹识别模型自动识别预设个数的会议语音片段的发言人身份。可以理解地，声纹识别模型是预先训练好的，需要预先录入参会人员的发言人身份（如将一个预先录入的声纹特征与对应的发言人身份进行绑定）。采用声纹识别可以在会议内容区分系统直接确认预设个数的会议语音片段的发言人身份，无需与用户进行信息交互。可以理解地，在会议形式为线上形式，如若干人参与微信群组的会议时，用户可以在一次录入个人的发言人身份后，在每次会议时都可以自动确认预设个数的会议语音片段的发言人身份，无需每次都采用与用户进行信息交互的形式确认发言人身份。采用声纹识别模型识别更适用与人数不多的线上会议，能够实现全自动的发言人身份确认。

S423：根据第二确认结果和相同发言人的会议语音片段确定会议语音片段的发言人身份。

可以理解地，第二确认结果代表的是展示的会议语音片段，由于已采用聚类的方式将相同的会议语音片段归成同一类，因此，可以根据第二确认结果和相同发言人的会议语音片段直接确定会议语音片段的发言人身份，可以快速确定所有会议语音片段的发言人身份。

S50：根据发言人身份和相同发言人的会议语音片段区分会议内容。

在一实施例中，会议内容为不同发言人在会议上的发言，该不同发言人在会议上的发言以不同发言人的会议语音片段表示。因此，在知道发言人身份和相同发言人的会议语音片段的情况下，可以确定会议上哪个发言人说了哪些话，也就实现了区分会议内容的目的。

具体地，在步骤 S50 中，包括将相同发言人的会议语音片段按发言人身份输入到语音转文本模型中，得到不同发言人的会议内容，从而实现了会议内容的区分。

进一步地，在步骤 S50 之后，还包括：

采用预先训练的神经网络模型和神经语音模型对会议内容进行分析，生成会议纪要和/或执行列表。

可以理解地，神经网络模型和神经语音模型是根据大量的会议纪要和/或执行列表训

练得到的，学习了会议纪要和/或执行列表的深层特征，可以对会议内容进行深层分析，根据输入的会议内容生成会议纪要和/或执行列表。采用该生成会议纪要和/或执行列表的方式无需人工进行整理，能够提高整理会议内容的效率。

在本申请实施例中，首先将获取的目标会议语音片段根据发言人转变点进行切割，得到至少三个会议室语音片段，能够将包括至少两个不同发言人的会议语音片段的目标会议语音片段实现合理切割，使得每个得到的会议室语音片段来自一位发言人；然后提取会议语音片段的片段语音特征，根据片段语音特征所表达出的相似性对会议室语音片段聚类，根据聚类的结果确定相同发言人的会议语音片段，将会议语音片段按类别进行区分；最后根据相同发言人的会议语音片段确定每一个会议语音片段对应的发言人身份，从而根据发言人身份和相同发言人的会议语音片段确定会议内容中各个会议语音片段具体的所属情况，实现会议室内容的高效区分。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不对本申请实施例的实施过程构成任何限定。

基于实施例中所提供的会议内容区分方法，本申请实施例进一步给出实现上述方法实施例中各步骤及方法的装置实施例。

图2示出与实施例中会议内容区分方法一一对应的会议内容区分装置的原理框图。如图2所示，该会议内容区分装置包括目标片段获取模块10、会议语音片段获取模块20、相同发言人语音片段确定模块30、发言人身份确定模块40和区分模块50。其中，目标片段获取模块10、会议语音片段获取模块20、相同发言人语音片段确定模块30、发言人身份确定模块40和区分模块50的实现功能与实施例中会议内容区分方法对应的步骤一一对应，为避免赘述，本实施例不一一详述。

目标片段获取模块10，用于获取目标会议语音片段，其中，目标会议语音片段包括至少两个不同发言人的会议语音片段。

会议语音片段获取模块20，用于获取目标会议语音片段的发言人转变点，根据发言人转变点切割目标会议语音片段，得到至少三个会议语音片段，其中，一个发言人对应一个或多个会议语音片段。

相同发言人语音片段确定模块30，用于提取会议语音片段的片段语音特征，根据片段语音特征对会议语音片段进行聚类，确定相同发言人的会议语音片段。

发言人身份确定模块40，用于根据相同发言人的会议语音片段确定会议语音片段的发言人身份。

区分模块50，用于根据发言人身份和相同发言人的会议语音片段区分会议内容。

可选地，相同发言人语音片段确定模块 30 包括片段语音特征提取单元、特征表达模型获取单元和相同发言人语音片段获取单元。

片段语音特征提取单元，用于通过预先训练的通用背景模型和高斯混合模型从会议语音片段中提取 i-vector 特征作为片段语音特征。

特征表达模型获取单元，用于采用预先训练的双协方差概率线性判别模型对 i-vector 特征建模，得到会议语音片段的特征表达模型。

相同发言人语音片段获取单元，用于采用特征表达模型对会议语音片段进行聚类，确定相同发言人的会议语音片段。

可选地，发言人身份确定模块 40 包括展示单元、第一确认结果获取单元和第一发言人身份确定单元。

展示单元，用于在每一相同发言人的会议语音片段中各获取预设个数的会议语音片段，并进行展示。

第一确认结果获取单元，用于响应于展示，获取发言人身份确认指令，根据发言人身份确认指令确认预设个数的会议语音片段的发言人身份，得到第一确认结果。

第一发言人身份确定单元，用于根据第一确认结果和相同发言人的会议语音片段确定会议语音片段的发言人身份。

可选地，发言人身份确定模块 40 还包括输入单元、第二确认结果获取单元和第二发言人身份确定单元。

输入单元，用于在每一相同发言人的会议语音片段中各获取预设个数的会议语音片段，并输入到预先训练的声纹识别模型中。

第二确认结果获取单元，用于通过声纹识别模型识别预设个数的会议语音片段，确认预设个数的会议语音片段的发言人身份，得到第二确认结果。

第二发言人身份确定单元，用于根据第二确认结果和相同发言人的会议语音片段确定会议语音片段的发言人身份。

可选地，区分模块 50 具体用于将相同发言人的会议语音片段按发言人身份输入到语音转文本模型中，得到不同发言人的会议内容。

可选地，会议内容区分装置还包括生成单元，用于采用预先训练的神经网络模型和神经语音模型对会议内容进行分析，生成会议纪要和/或执行列表。

可选地，目标片段获取模块 10 包括原始会议语音片段获取单元和目标会议语音片段获取单元。

原始会议语音片段获取单元，用于获取原始会议语音片段。

目标会议语音片段获取单元，用于采用静音检测去除原始会议语音片段中的静默片段，得到目标会议语音片段。

在本申请实施例中，首先将获取的目标会议语音片段根据发言人转变点进行切割，得到至少三个会议室语音片段，能够将包括至少两个不同发言人的会议语音片段的目标会议语音片段实现合理切割，使得每个得到的会议室语音片段来自一位发言人；然后提取会议语音片段的片段语音特征，根据片段语音特征所表达出的相似性对会议室语音片段聚类，根据聚类的结果确定相同发言人的会议语音片段，将会议语音片段按类别进行区分；最后根据相同发言人的会议语音片段确定每一个会议语音片段对应的发言人身份，从而根据发言人身份和相同发言人的会议语音片段确定会议内容中各个会议语音片段具体的所属情况，实现会议室内容的高效区分。

本实施例提供一计算机非易失性可读存储介质，该计算机非易失性可读存储介质上存储有计算机可读指令，该计算机可读指令被处理器执行时实现实施例中会议内容区分方法，为避免重复，此处不一一赘述。或者，该计算机可读指令被处理器执行时实现实施例中会议内容区分装置中各模块/单元的功能，为避免重复，此处不一一赘述。

图3是本申请一实施例提供的计算机设备的示意图。如图3所示，该实施例的计算机设备60包括：处理器61、存储器62以及存储在存储器62中并可在处理器61上运行的计算机可读指令63，该计算机可读指令63被处理器61执行时实现实施例中的会议内容区分方法，为避免重复，此处不一一赘述。或者，该计算机可读指令63被处理器61执行时实现实施例中会议内容区分装置中各模型/单元的功能，为避免重复，此处不一一赘述。

计算机设备60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60可包括，但不仅限于，处理器61、存储器62。本领域技术人员可以理解，图3仅仅是计算机设备60的示例，并不构成对计算机设备60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61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 62 可以是计算机设备 60 的内部存储单元,例如计算机设备 60 的硬盘或内存。存储器 62 也可以是计算机设备 60 的外部存储设备,例如计算机设备 60 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 62 还可以既包括计算机设备 60 的内部存储单元也包括外部存储设备。存储器 62 用于存储计算机可读指令以及计算机设备所需的其他程序和数据。存储器 62 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,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将装置的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。

以上实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围,均应包含在本申请的保护范围之内。

权 利 要 求 书

1、一种会议内容区分方法，其特征在于，所述方法包括：

获取目标会议语音片段，其中，所述目标会议语音片段包括至少两个不同发言人的会议语音片段；

获取所述目标会议语音片段的发言人转变点，根据所述发言人转变点切割所述目标会议语音片段，得到至少三个会议语音片段，其中，一个所述发言人对应一个或多个所述会议语音片段；

提取所述会议语音片段的片段语音特征，根据所述片段语音特征对所述会议语音片段进行聚类，确定相同发言人的会议语音片段；

根据所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份；

根据所述发言人身份和所述相同发言人的会议语音片段区分会议内容。

2、根据权利要求1所述的方法，其特征在于，所述提取所述会议语音片段的片段语音特征，根据所述片段语音特征对所述会议语音片段进行聚类，确定相同发言人的会议语音片段，包括：

通过预先训练的通用背景模型和高斯混合模型从所述会议语音片段中提取 i-vector 特征作为所述片段语音特征；

采用预先训练的双协方差概率线性判别模型对所述 i-vector 特征建模，得到会议语音片段的特征表达模型；

采用所述特征表达模型对所述会议语音片段进行聚类，确定所述相同发言人的会议语音片段。

3、根据权利要求1所述的方法，其特征在于，所述根据所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份，包括：

在每一所述相同发言人的会议语音片段中各获取预设个数的会议语音片段，并进行展示；

响应于所述展示，获取发言人身份确认指令，根据所述发言人身份确认指令确认所述预设个数的会议语音片段的发言人身份，得到第一确认结果；

根据所述第一确认结果和所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份。

4、根据权利要求1所述的方法，其特征在于，所述根据所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份，还包括：

在每一所述相同发言人的会议语音片段中各获取预设个数的会议语音片段,并输入到预先训练的声纹识别模型中;

通过所述声纹识别模型识别所述预设个数的会议语音片段,确认所述预设个数的会议语音片段的发言人身份,得到第二确认结果;

根据所述第二确认结果和所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份。

5、根据权利要求1所述的方法,其特征在于,所述根据所述发言人身份和所述相同发言人的会议语音片段区分会议内容,包括:

将所述相同发言人的会议语音片段按所述发言人身份输入到语音转文本模型中,得到不同发言人的会议内容。

6、根据权利要求1-5任意一项所述的方法,其特征在于,在所述根据所述发言人身份和所述相同发言人的会议语音片段区分会议内容之后,还包括:

采用预先训练的神经网络模型和神经语音模型对所述会议内容进行分析,生成会议纪要和/或执行列表。

7、根据权利要求1所述的方法,其特征在于,所述获取目标会议语音片段,包括:

获取原始会议语音片段;

采用静音检测去除所述原始会议语音片段中的静默片段,得到所述目标会议语音片段。

8、一种会议内容区分装置,其特征在于,所述装置包括:

目标片段获取模块,用于获取目标会议语音片段,其中,所述目标会议语音片段包括至少两个不同发言人的会议语音片段;

会议语音片段获取模块,用于获取所述目标会议语音片段的发言人转变点,根据所述发言人转变点切割所述目标会议语音片段,得到至少三个会议语音片段,其中,一个所述发言人对一个或多个所述会议语音片段;

相同发言人语音片段确定模块,用于提取所述会议语音片段的片段语音特征,根据所述片段语音特征对所述会议语音片段进行聚类,确定相同发言人的会议语音片段;

发言人身份确定模块,用于根据所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份;

区分模块,用于根据所述发言人身份和所述相同发言人的会议语音片段区分会议内容。

9、根据权利要求8所述的装置,其特征在于,所述发言人身份确定模块包括展示单元、第一确认结果获取单元和第一发言人身份确定单元:

展示单元, 用于在每一所述相同发言人的会议语音片段中各获取预设个数的会议语音片段, 并进行展示;

第一确认结果获取单元, 用于响应于所述展示, 获取发言人身份确认指令, 根据所述发言人身份确认指令确认所述预设个数的会议语音片段的发言人身份, 得到第一确认结果;

第一发言人身份确定单元, 用于根据所述第一确认结果和所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份。

10、根据权利要求 8 所述的装置, 其特征在于, 所述发言人身份确定模块还包括输入单元、第二确认结果获取单元和第二发言人身份确定单元:

输入单元, 用于在每一所述相同发言人的会议语音片段中各获取预设个数的会议语音片段, 并输入到预先训练的声纹识别模型中;

第二确认结果获取单元, 用于通过所述声纹识别模型识别所述预设个数的会议语音片段, 确认所述预设个数的会议语音片段的发言人身份, 得到第二确认结果;

第二发言人身份确定单元, 用于根据所述第二确认结果和所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份。

11、一种计算机设备, 包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令, 其特征在于, 所述处理器执行所述计算机可读指令时实现如下步骤:

获取目标会议语音片段, 其中, 所述目标会议语音片段包括至少两个不同发言人的会议语音片段;

获取所述目标会议语音片段的发言人转变点, 根据所述发言人转变点切割所述目标会议语音片段, 得到至少三个会议语音片段, 其中, 一个所述发言人对应一个或多个所述会议语音片段;

提取所述会议语音片段的片段语音特征, 根据所述片段语音特征对所述会议语音片段进行聚类, 确定相同发言人的会议语音片段;

根据所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份;

根据所述发言人身份和所述相同发言人的会议语音片段区分会议内容。

12、根据权利要求 11 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, 所述处理器还实现如下步骤:

通过预先训练的通用背景模型和高斯混合模型从所述会议语音片段中提取 i-vector 特征作为所述片段语音特征;

采用预先训练的双协方差概率线性判别模型对所述 i-vector 特征建模, 得到会议语音片段

的特征表达模型；

采用所述特征表达模型对所述会议语音片段进行聚类，确定所述相同发言人的会议语音片段。

13、根据权利要求 11 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，所述处理器还实现如下步骤：

在每一所述相同发言人的会议语音片段中各获取预设个数的会议语音片段，并进行展示；

响应于所述展示，获取发言人身份确认指令，根据所述发言人身份确认指令确认所述预设个数的会议语音片段的发言人身份，得到第一确认结果；

根据所述第一确认结果和所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份。

14、根据权利要求 11 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，所述处理器还实现如下步骤：

在每一所述相同发言人的会议语音片段中各获取预设个数的会议语音片段，并输入到预先训练的声纹识别模型中；

通过所述声纹识别模型识别所述预设个数的会议语音片段，确认所述预设个数的会议语音片段的发言人身份，得到第二确认结果；

根据所述第二确认结果和所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份。

15、根据权利要求 11 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，所述处理器还实现如下步骤：

将所述相同发言人的会议语音片段按所述发言人身份输入到语音转文本模型中，得到不同发言人的会议内容。

16、一种计算机非易失性可读存储介质，所述计算机非易失性可读存储介质存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现如下步骤：获取目标会议语音片段，其中，所述目标会议语音片段包括至少两个不同发言人的会议语音片段；

获取所述目标会议语音片段的发言人转变点，根据所述发言人转变点切割所述目标会议语音片段，得到至少三个会议语音片段，其中，一个所述发言人对应一个或多个所述会议语音片段；

提取所述会议语音片段的片段语音特征，根据所述片段语音特征对所述会议语音片段进行聚类，确定相同发言人的会议语音片段；

根据所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份；

根据所述发言人身份和所述相同发言人的会议语音片段区分会议内容。

17、根据权利要求 16 所述的计算机非易失性可读存储介质，其特征在于，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还实现如下步骤：

通过预先训练的通用背景模型和高斯混合模型从所述会议语音片段中提取 i-vector 特征作为所述片段语音特征；

采用预先训练的双协方差概率线性判别模型对所述 i-vector 特征建模，得到会议语音片段的特征表达模型；

采用所述特征表达模型对所述会议语音片段进行聚类，确定所述相同发言人的会议语音片段。

18、根据权利要求 16 所述的计算机非易失性可读存储介质，其特征在于，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还实现如下步骤：

在每一所述相同发言人的会议语音片段中各获取预设个数的会议语音片段，并进行展示；

响应于所述展示，获取发言人身份确认指令，根据所述发言人身份确认指令确认所述预设个数的会议语音片段的发言人身份，得到第一确认结果；

根据所述第一确认结果和所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份。

19、根据权利要求 16 所述的计算机非易失性可读存储介质，其特征在于，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还实现如下步骤：

在每一所述相同发言人的会议语音片段中各获取预设个数的会议语音片段，并输入到预先训练的声纹识别模型中；

通过所述声纹识别模型识别所述预设个数的会议语音片段，确认所述预设个数的会议语音片段的发言人身份，得到第二确认结果；

根据所述第二确认结果和所述相同发言人的会议语音片段确定所述会议语音片段的发言人身份。

20、根据权利要求 16 所述的计算机非易失性可读存储介质，其特征在于，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还实现如下步骤：

将所述相同发言人的会议语音片段按所述发言人身份输入到语音转文本模型中，得到不同发言人的会议内容。

1/2

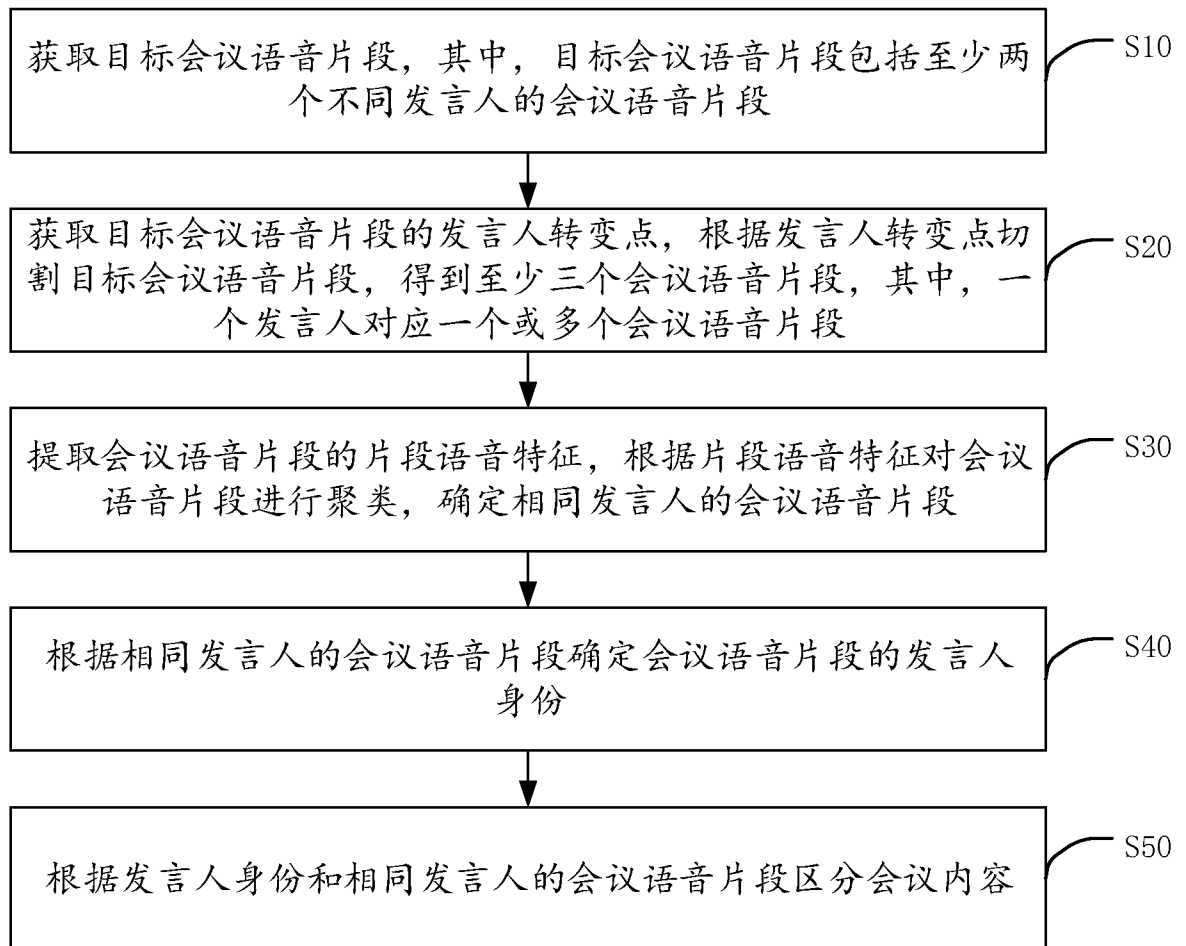


图 1

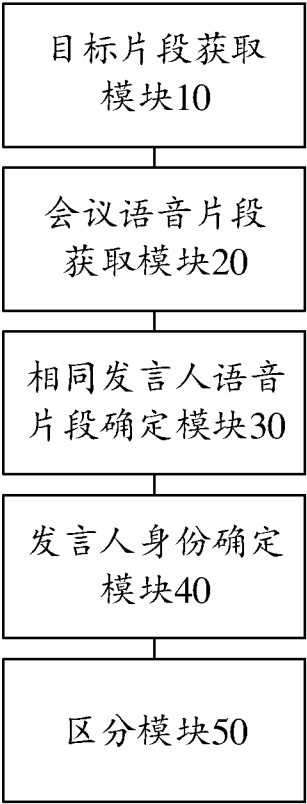


图 2

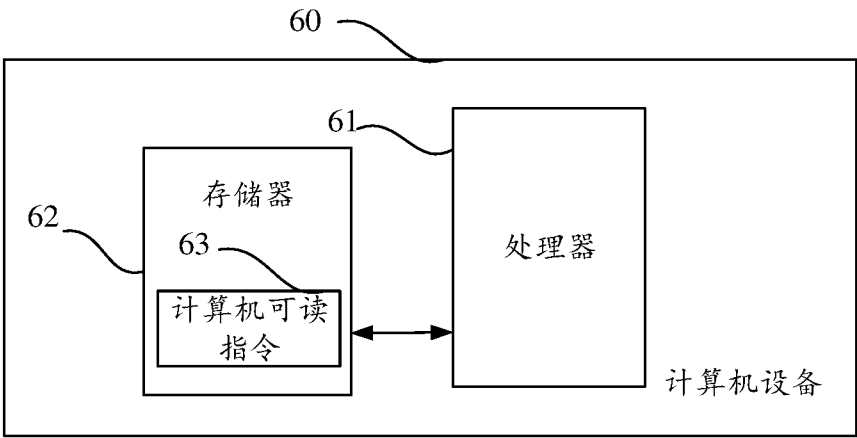


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/091098

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/61(2019.01)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)

G06F

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)

EPODOC, WPI, CNPAT, CNKI, IEEE: 会议, 说话, 通话, 识别, 区分, 分离, 声纹, 聚类, 分类, 类别, 背景模型, 高斯混合模型, 语音片段, 协方差, 身份, 认证, 片段, 分段, VAD, i-vector, BIC, LR, UBM, meeting, conference, voice, speaker, characteristic, classification, Bayesian, VB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 107689225 A (START COMPUTER EQUIPMENT CO., LTD.) 13 February 2018 (2018-02-13) description, paragraphs [0006]-[0008]	1-20
Y	CN 107545898 A (TSINGHUA UNIVERSITY et al.) 05 January 2018 (2018-01-05) description, paragraphs [0041] and [0042]	1-20
A	CN 108986826 A (PING AN LIFE INSURANCE COMPANY OF CHINA) 11 December 2018 (2018-12-11) entire document	1-20
A	CN 108766445 A (AI SPEECH LTD. et al.) 06 November 2018 (2018-11-06) entire document	1-20
A	US 2004214558 A1 (BELLSOUTH INTELLECTUAL PROPERTY CORPORATION) 28 October 2004 (2004-10-28) 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

19 September 2019

Date of mailing of the international search report

16 October 2019

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/CN2019/091098

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107689225	A	13 February 2018	None			
CN	107545898	A	05 January 2018	None			
CN	108986826	A	11 December 2018	None			
CN	108766445	A	06 November 2018	None			
US	2004214558	A1	28 October 2004	US	2008152113	A1	26 June 2008
				US	2004131167	A1	08 July 2004
				WO	03055188	A1	03 July 2003
				AU	2002357250	A1	09 July 2003

国际检索报告

国际申请号

PCT/CN2019/091098

A. 主题的分类 G06F 16/61 (2019.01) 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) G06F 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)) EP0DOC, WPI, CNPAT, CNKI, IEEE: 会议, 说话, 通话, 识别, 区分, 分离, 声纹, 聚类, 分类, 类别, 背景模型, 高斯混合模型, 语音片段, 协方差, 身份, 认证, 片段, 分段, VAD, i-vector, BIC, LR, UBM, meeting, conference, voice, speaker, characteristic, classification, Bayesian, VB																				
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>Y</td> <td>CN 107689225 A (福建实达电脑设备有限公司) 2018年 2月 13日 (2018 - 02 - 13) 说明书第[0006]-[0008]段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 107545898 A (清华大学 等) 2018年 1月 5日 (2018 - 01 - 05) 说明书第[0041]-[0042]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>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</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108766445 A (苏州思必驰信息科技有限公司 等) 2018年 11月 6日 (2018 - 11 - 06) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2004214558 A1 (BELLSOUTH INTELLECTUAL PROPERTY CORPORATION) 2004年 10月 28日 (2004 - 10 - 28) 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	Y	CN 107689225 A (福建实达电脑设备有限公司) 2018年 2月 13日 (2018 - 02 - 13) 说明书第[0006]-[0008]段	1-20	Y	CN 107545898 A (清华大学 等) 2018年 1月 5日 (2018 - 01 - 05) 说明书第[0041]-[0042]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 108766445 A (苏州思必驰信息科技有限公司 等) 2018年 11月 6日 (2018 - 11 - 06) 全文	1-20	A	US 2004214558 A1 (BELLSOUTH INTELLECTUAL PROPERTY CORPORATION) 2004年 10月 28日 (2004 - 10 - 28) 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																		
Y	CN 107689225 A (福建实达电脑设备有限公司) 2018年 2月 13日 (2018 - 02 - 13) 说明书第[0006]-[0008]段	1-20																		
Y	CN 107545898 A (清华大学 等) 2018年 1月 5日 (2018 - 01 - 05) 说明书第[0041]-[0042]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 108766445 A (苏州思必驰信息科技有限公司 等) 2018年 11月 6日 (2018 - 11 - 06) 全文	1-20																		
A	US 2004214558 A1 (BELLSOUTH INTELLECTUAL PROPERTY CORPORATION) 2004年 10月 28日 (2004 - 10 - 28) 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 2019年 9月 19日		国际检索报告邮寄日期 2019年 10月 16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 高民芳 电话号码 86-(10)-53961520																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/091098

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	107689225	A	2018年 2月 13日	无	
CN	107545898	A	2018年 1月 5日	无	
CN	108986826	A	2018年 12月 11日	无	
CN	108766445	A	2018年 11月 6日	无	
US	2004214558	A1	2004年 10月 28日	US 2008152113 A1	2008年 6月 26日
				US 2004131167 A1	2004年 7月 8日
				WO 03055188 A1	2003年 7月 3日
				AU 2002357250 A1	2003年 7月 9日