(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
2017 年 5 月 11 日 (11.05.2017)

WIPO | PCT

(10) 国际公布号
WO 2017/076211 A1

(51) 国际专利分类号:
G10L 17/00 (2013.01)

(21) 国际申请号: PCT/CN2016/103490

(22) 国际申请日: 2016 年 10 月 27 日 (27.10.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510744743.4 2015 年 11 月 5 日 (05.11.2015) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 李晓辉 (LI, Xiaohui); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 李宏言 (LI, Hongyan); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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, JP, KE, KG, 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,

[见续页]

(54) Title: VOICE-BASED ROLE SEPARATION METHOD AND DEVICE

(54) 发明名称 : 基于语音的角色分离方法及装置

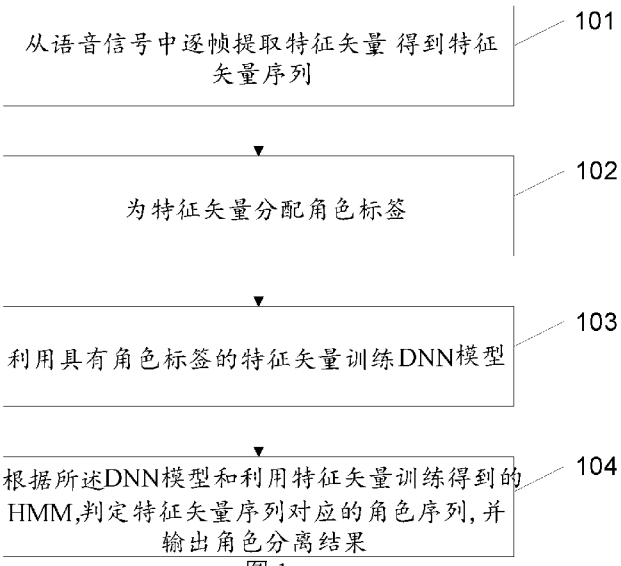

图 1

101 EXTRACT FEATURE VECTORS FROM A VOICE SIGNAL FRAME BY FRAME, SO AS TO OBTAIN A FEATURE VECTOR SEQUENCE
102 ALLOCATE ROLE LABELS TO THE FEATURE VECTORS
103 TRAIN, BY EMPLOYING THE FEATURE VECTORS HAVING THE ROLE LABELS, A DNN MODEL
104 DETERMINE, ACCORDING TO THE DNN MODEL AND AN HMM TRAINED BY USING THE FEATURE VECTORS, A ROLE SEQUENCE CORRESPONDING TO THE FEATURE VECTOR SEQUENCE, AND OUTPUT A ROLE SEPARATION RESULT

(57) Abstract: A voice-based role separation method and device. The method comprises: extracting feature vectors from a voice signal frame by frame, so as to obtain a feature vector sequence (101); allocating role labels to the feature vectors (102); training, by employing the feature vectors having the role labels, a deep neural network (DNN) model (103); and determining, according to the DNN model and a hidden Markov model (HMM) trained by using the feature vectors, a role sequence corresponding to the feature vector sequence, and outputting a role separation result (104), wherein the DNN model is configured to output, according to an inputted feature vector, probabilities corresponding to respective roles, and the HMM is configured to describe a transition relationship between the roles. Employing the DNN model having powerful feature extraction capability to establish a model of a speaker role, the method can better describe a role compared with the conventional GMM, and can generate more detailed and accurate description of a role, thereby providing a more accurate role separation result.

(57) 摘要:

[见续页]

WO 2017/076211 A1

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))。

一种基于语音的角色分离方法及装置，该方法包括：从语音信号中逐帧提取特征矢量，得到特征矢量序列（101）；为特征矢量分配角色标签（102）；利用具有角色标签的特征矢量训练深度神经网络 DNN 模型（103）；根据 DNN 模型和利用特征矢量训练得到的隐马尔科夫模型 HMM，判定特征矢量序列对应的角色序列，并输出角色分离结果（104）。其中，DNN 模型用于根据输入的特征矢量输出对应每个角色的概率，HMM 用于描述角色间的跳转关系。由于该方法采用了具有强大特征提取能力的 DNN 模型对说话人角色进行建模，比传统的 GMM 具有更为强大的刻画能力，对角色的刻画更加精细、准确，能够获得更为准确的角色分离结果。

基于语音的角色分离方法及装置

本申请要求2015年11月05日递交的申请号为201510744743.4、发明名称为“基于语音的角色分离方法及装置”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

本申请涉及语音识别领域，具体涉及一种基于语音的角色分离方法。本申请同时涉及一种基于语音的角色分离装置。

背景技术

语音是人类最自然的交流沟通方式，语音识别技术则是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的技术。语音识别是一门交叉学科，所涉及的领域包括：信号处理、模式识别、概率论和信息论、发声机理和听觉机理、人工智能等等。

在实际应用中，为了能够对语音信号作更为准确的分析，不仅需要进行语音识别，而且要判别出每段语音的说话人，因此很自然地出现了对语音按照角色进行分离的需求。在日常生活、会议以及电话对话等很多场景下，都存在对话语音，而通过对对话语音的角色分离，就可以判定哪部分语音是其中一个人说的，哪部分语音是另外一个人说的。在将对话语音按照角色分离之后，结合说话人识别、语音识别，会产生更为广阔的应用空间，例如，将客服中心的对话语音按照角色分离，然后进行语音识别就可以确定客服说了什么内容，客户说了什么内容，从而可以进行相应的客服质检或者进行客户潜在需求的挖掘。

现有技术中，通常采用GMM（Gaussian Mixture Model—高斯混合模型）和HMM（Hidden Markov Model—隐马尔科夫模型）进行对话语音的角色分离，即：对于每个角色使用GMM建模，对于不同角色之间的跳转采用HMM建模。由于GMM建模技术提出的时间比较早，而且其拟合任意函数的功能取决于混合高斯函数的个数，所以其对角色的刻画能力有一定的局限性，导致角色分离的准确率通常比较低，无法满足应用的需求。

发明内容

本申请实施例提供一种基于语音的角色分离方法和装置，以解决现有的基于GMM

和 HMM 的角色分离技术准确率比较低的问题。

本申请提供一种基于语音的角色分离方法，包括：

从语音信号中逐帧提取特征矢量，得到特征矢量序列；

为特征矢量分配角色标签；

利用具有角色标签的特征矢量训练深度神经网络 DNN 模型；

根据所述 DNN 模型和利用特征矢量训练得到的隐马尔科夫模型 HMM，判定特征矢量序列对应的角色序列，并输出角色分离结果；

其中，所述 DNN 模型用于根据输入的特征矢量输出对应每个角色的概率，HMM 用于描述角色间的跳转关系。

可选的，在所述从语音信号中逐帧提取特征矢量的步骤之后、在所述为特征矢量分配角色标签的步骤之前，执行下述操作：通过识别并剔除不包含语音内容的音频帧、将所述语音信号切分为语音段；

所述为特征矢量分配角色标签包括：为各语音段中的特征矢量分配角色标签；所述判定特征矢量序列对应的角色序列包括：判定各语音段所包含的特征矢量序列对应的角色序列。

可选的，所述为各语音段中的特征矢量分配角色标签包括：通过建立高斯混合模型 GMM 和 HMM，为各语音段中的特征矢量分配角色标签；其中所述 GMM 用于针对每个角色、根据输入的特征矢量输出该特征矢量对应于所述角色的概率；

所述根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定各语音段所包含的特征矢量序列对应的角色序列包括：根据所述 DNN 模型和为各语音段中的特征矢量分配角色标签所采用的 HMM，判定所述各语音段所包含的特征矢量序列对应的角色序列。

可选的，所述通过建立高斯混合模型 GMM 和 HMM，为各语音段中的特征矢量分配角色标签，包括：

按照预设的初始角色数量选择相应数量的语音段，并为每个语音段分别指定不同角色；

利用指定角色的语音段中的特征矢量，训练针对每个角色的 GMM 以及 HMM；

根据训练得到的 GMM 和 HMM 进行解码，获取输出各语音段所包含的特征矢量序列的概率值排序靠前的角色序列；

判断所述角色序列对应的概率值是否大于预设阈值；若是，按照所述角色序列为各语音段中的特征矢量分配角色标签。

可选的，当所述判断所述角色序列对应的概率值是否大于预设阈值的结果为否时，执行下述操作：

根据所述角色序列，为每个语音段指定对应的角色；

根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的 GMM 以及 HMM；

转到所述根据训练得到的 GMM 和 HMM 进行解码的步骤执行。

可选的，所述根据所述角色序列，为每个语音段指定对应的角色，包括：

针对每个语音段，将其中各特征矢量对应的角色的众数指定为所述语音段的角色。

可选的，所述根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的 GMM 以及 HMM，包括：在上一次训练得到的模型基础上采用增量方式训练所述 GMM 以及 HMM。

可选的，当所述判断所述角色序列对应的概率值是否大于预设阈值的结果为否时，执行下述操作：

判断在当前角色数量下训练 GMM 和 HMM 的次数是否小于预设的训练次数上限；

若是，执行所述根据所述角色序列为每个语音段指定对应的角色的步骤；

若否，执行下述操作：

调整角色数量，选择相应数量的语音段并为每个语音段分别指定不同角色；

并转到所述利用指定角色的语音段中的特征矢量，训练针对每个角色的 GMM 以及 HMM 的步骤执行。

可选的，当所述判断在当前角色数量下训练 GMM 和 HMM 的次数是否小于预设的训练次数上限的结果为否时，执行下述操作：

判断当前角色数量是否符合预设要求；若是，转到所述按照所述角色序列为各语音段中的特征矢量分配角色标签的步骤执行，若否，则执行所述调整角色数量的步骤。

可选的，所述预设的初始角色数量为 2，所述调整角色数量包括：为当前角色数量加 1。

可选的，所述从语音信号中逐帧提取特征矢量，得到特征矢量序列包括：

按照预先设定的帧长度对语音信号进行分帧处理，得到多个音频帧；

提取各音频帧的特征矢量，得到所述特征矢量序列。

可选的，所述提取各音频帧的特征矢量包括：提取 MFCC 特征、PLP 特征、或者 LPC 特征。

可选的，所述识别并剔除不包含语音内容的音频帧包括：采用 VAD 技术识别所述不包含语音内容的音频帧、并执行相应的剔除操作。

可选的，在采用 VAD 技术执行所述识别及剔除操作、并将所述语音信号切分为语音段之后，执行下述 VAD 平滑操作：

将时长小于预设阈值的语音段与相邻语音段合并。

可选的，所述利用具有角色标签的特征矢量训练深度神经网络 DNN 模型包括：采用反向传播算法训练所述 DNN 模型。

可选的，所述根据所述 DNN 模型和利用特征矢量训练得到的隐马尔科夫模型 HMM，判定特征矢量序列对应的角色序列，包括：根据所述 DNN 模型和 HMM 执行解码操作，获取输出所述特征矢量序列的概率值排序靠前的角色序列，并将所述角色序列作为与所述特征矢量序列对应的角色序列。

可选的，所述输出角色分离结果包括：根据特征矢量序列对应的角色序列，针对每个角色输出与其对应的特征矢量所属音频帧的起止时间信息。

可选的，所述选择相应数量的语音段，包括：选择时长满足预设要求的、所述数量的语音段。

相应的，本申请还提供一种基于语音的角色分离装置，包括：

特征提取单元，用于从语音信号中逐帧提取特征矢量，得到特征矢量序列；

标签分配单元，用于为特征矢量分配角色标签；

DNN 模型训练单元，用于利用具有角色标签的特征矢量训练 DNN 模型，其中所述 DNN 模型用于根据输入的特征矢量输出对应每个角色的概率；

角色判定单元，用于根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定特征矢量序列对应的角色序列并输出角色分离结果，其中所述 HMM 用于描述角色间的跳转关系。

可选的，所述装置还包括：

语音段切分单元，用于在所述特征提取单元提取特征矢量后、在触发所述标签分配单元工作之前，通过识别并剔除不包含语音内容的音频帧、将所述语音信号切分为语音段；

所述标签分配单元具体用于，为各语音段中的特征矢量分配角色标签；

所述角色判定单元具体用于，根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定各语音段所包含的特征矢量序列对应的角色序列并输出角色分离结果。

可选的，所述标签分配单元具体用于，通过建立 GMM 和 HMM，为各语音段中的特征矢量分配角色标签，其中所述 GMM 用于针对每个角色、根据输入的特征矢量输出该特征矢量对应于所述角色的概率；

所述角色判定单元具体用于，根据所述 DNN 模型和为各语音段中的特征矢量分配角色标签所采用的 HMM，判定所述各语音段所包含的特征矢量序列对应的角色序列。

可选的，所述标签分配单元包括：

初始角色指定子单元，用于按照预设的初始角色数量选择相应数量的语音段，并为每个语音段分别指定不同角色；

初始模型训练子单元，用于利用指定角色的语音段中的特征矢量，训练针对每个角色的 GMM 以及 HMM；

解码子单元，用于根据训练得到的 GMM 和 HMM 进行解码，获取输出各语音段所包含的特征矢量序列的概率值排序靠前的角色序列；

概率判断子单元，用于判断所述角色序列对应的概率值是否大于预设阈值；

标签分配子单元，用于当所述概率判断子单元的输出为是时，按照所述角色序列为各语音段中的特征矢量分配角色标签。

可选的，所述标签分配单元还包括：

逐语音段角色指定子单元，用于当所述概率判断子单元的输出为否时，根据所述角色序列，为每个语音段指定对应的角色；

模型更新训练子单元，用于根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的 GMM 以及 HMM，并触发所述解码子单元工作。

可选的，所述逐语音段角色指定子单元具体用于，针对每个语音段，将其中各特征矢量对应的角色的众数指定为所述语音段的角色。

可选的，所述模型更新训练子单元具体用于，在上一次训练得到的模型基础上采用增量方式训练所述 GMM 以及 HMM。

可选的，所述标签分配单元还包括：

训练次数判断子单元，用于当所述概率判断子单元的输出为否时，判断在当前角色数量下训练 GMM 和 HMM 的次数是否小于预设的训练次数上限，并在判断结果为是时，触发所述逐语音段角色指定子单元工作。

角色数量调整子单元，用于当所述训练次数判断子单元的输出为否时，调整角色数量，选择相应数量的语音段并为每个语音段分别指定不同角色，并触发所述初始模型训

练子单元工作。

可选的，所述标签分配单元还包括：

角色数量判断子单元，用于当所述训练次数判断子单元的输出为否时，判断当前角色数量是否符合预设要求，若符合则触发所述标签分配子单元工作，否则触发所述角色数量调整子单元工作。

可选的，所述特征提取单元包括：

分帧子单元，用于按照预先设定的帧长度对语音信号进行分帧处理，得到多个音频帧；

特征提取执行子单元，用于提取各音频帧的特征矢量，得到所述特征矢量序列。

可选的，所述特征提取执行子单元具体用于，提取各音频帧的 MFCC 特征、PLP 特征、或者 LPC 特征，得到所述特征矢量序列。

可选的，所述语音段切分单元具体用于，通过采用 VAD 技术识别并剔除所述不包含语音内容的音频帧、将所述语音信号切分为语音段。

可选的，所述装置还包括：

VAD 平滑单元，用于在所述语音段切分单元采用 VAD 技术切分语音段后，将时长小于预设阈值的语音段与相邻语音段合并。

可选的，所述 DNN 模型训练单元具体用于，采用反向传播算法训练所述 DNN 模型。

可选的，所述角色判定单元具体用于，根据所述 DNN 模型和 HMM 执行解码操作，获取输出所述特征矢量序列的概率值排序靠前的角色序列，并将所述角色序列作为与所述特征矢量序列对应的角色序列。

可选的，所述角色判定单元采用如下方式输出角色分离结果：根据特征矢量序列对应的角色序列，针对每个角色输出与其对应的特征矢量所属音频帧的起止时间信息。

可选的，所述初始角色指定子单元或所述角色数量调整子单元具体通过如下方式选择相应数量的语音段：选择时长满足预设要求的、所述数量的语音段。

与现有技术相比，本申请具有以下优点：

本申请提供的基于语音的角色分离方法，首先从语音信号中逐帧提取特征矢量序列，然后在为特征矢量分配角色标签的基础上训练 DNN 模型，并根据所述 DNN 模型以及利用特征矢量训练得到的 HMM，判定特征矢量序列对应的角色序列，从而得到角色分离结果。本申请提供的上述方法，由于采用了具有强大特征提取能力的 DNN 模型对说话人角色进行建模，比传统的 GMM 具有更为强大的刻画能力，对角色的刻画更加精细、

准确，因此能够获得更为准确的角色分离结果。

附图说明

图 1 是本申请的一种基于语音的角色分离方法的实施例的流程图；

图 2 是本申请实施例提供的从语音信号中提取特征矢量序列的处理流程图；

图 3 是本申请实施例提供的利用 GMM 和 HMM 为各语音段中的特征矢量分配角色标签的处理流程图；

图 4 是本申请实施例提供的语音段划分的示意图；

图 5 是本申请实施例提供的 DNN 网络的拓扑结构示意图；

图 6 是本申请的一种基于语音的角色分离装置的实施例的示意图。

具体实施方式

在下面的描述中阐述了很多具体细节以便于充分理解本申请。但是，本申请能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本申请内涵的情况下做类似推广，因此，本申请不受下面公开的具体实施的限制。

在本申请中，分别提供了一种基于语音的角色分离方法，以及一种基于语音的角色分离装置，在下面的实施例中逐一进行详细说明。为了便于理解，在对实施例进行描述之前，先对本申请的技术背景、技术方案、以及实施例的撰写方式作简要说明。

现有的应用于语音领域的角色分离技术通常采用 GMM（Gaussian mixture model—高斯混合模型）对角色进行建模、采用 HMM（Hidden Markov Model—隐马尔可夫模型）对角色之间的跳转进行建模。

所述 HMM 是统计模型，用来描述一个含有隐含未知参数的马尔可夫过程。隐马尔可夫模型是马尔可夫链的一种，它的状态（称为隐藏状态）不能直接观察到，但是与可观察到的观测向量是概率相关的，因此，HMM 是一个双重随机过程，包括两个部分：具有状态转移概率的马尔科夫链（通常用转移矩阵 A 描述），以及描述隐藏状态与观测向量之间的输出关系的随机过程（通常用混淆矩阵 B 描述，其中的每个元素为隐藏状态对应于观测向量的输出概率，也称为发射概率）。一个具有 N 个状态的 HMM 可以用三元组参数 $\lambda=\{\pi, A, B\}$表示，其中 π 为各状态的初始概率。

所述 GMM 可以简单理解为多个高斯密度函数的叠加，其核心思想是用多个高斯分布的概率密度函数的组合来描述特征矢量在概率空间的分布状况，采用该模型可以平滑

地近似任意形状的密度分布。其参数包括：各高斯分布的混合权重（mixing weight）、均值向量（mean vector）、协方差矩阵（covariance matrix）。

在现有的基于语音的角色分离应用中，通常对每个角色采用 GMM 建模，HMM 中的状态就是各个角色，观测向量是从语音信号中逐帧提取的特征矢量，各个状态输出特征矢量的发射概率由 GMM 决定（根据 GMM 可以获知混淆矩阵），而角色分离过程就是利用 GMM 和 HMM 确定与特征矢量序列对应的角色序列的过程。

由于 GMM 的函数拟合功能受限于所采用的高斯密度函数的个数，其本身的表达能力存在一定的局限性，导致现有的采用 GMM 和 HMM 进行角色分离的准确率比较低。针对这一问题，本申请的技术方案在为各语音帧的特征矢量预先分配角色标签的基础上，利用深度神经网络（DNN）决定 HMM 各状态的发射概率，并根据 DNN 和 HMM 判定与特征矢量序列对应的角色序列，由于 DNN 具有组合低层特征形成更加抽象的高层特征的强大能力，可以实现更为精准的角色刻画，因此能够获取更为准确的角色分离结果。

本申请的技术方案，首先为从语音信号中提取的特征矢量分配角色标签，此时分配的角色标签通常并不是很准确，但是可以为后续执行有监督的学习过程提供参考，在此基础上训练得到的 DNN 模型能够更为准确地刻画角色，从而使角色分离结果更为准确。在具体实施本申请的技术方案时可以采用基于统计的算法或者采用分类器等方式，实现所述分配角色标签的功能，在本申请提供的下述实施例中采用了根据 GMM 和 HMM 为特征矢量分配角色标签的实施方式。

下面，对本申请的实施例进行详细说明。请参考图 1，其为本申请的一种基于语音的角色分离方法的实施例的流程图。所述方法包括如下步骤：

步骤 101、从语音信号中逐帧提取特征矢量，得到特征矢量序列。

待进行角色分离的语音信号通常是时域信号，本步骤通过分帧和提取特征矢量两个处理过程，获取能够表征所述语音信号的特征矢量序列，下面结合附图 2 做进一步说明。

步骤 101-1、按照预先设定的帧长度对语音信号进行分帧处理，得到多个音频帧。

在具体实施时，可以根据需求预先设定帧长度，例如可以设置为 10ms、或者 15ms 等，然后根据所述帧长度对时域的语音信号进行逐帧切分，从而将语音信号切分为多个音频帧。根据所采用的切分策略的不同，相邻音频帧可以不存在交叠、也可以是有交叠的。

步骤 101-2、提取各音频帧的特征矢量，得到所述特征矢量序列。

将时域的语音信号切分为多个音频帧后，可以逐帧提取能够表征语音信号的特征矢

量。由于语音信号在时域上的描述能力相对较弱，通常可以针对每个音频帧进行傅里叶变换，然后提取频域特征作为音频帧的特征矢量，例如，可以提取 MFCC（Mel Frequency Cepstrum Coefficient—梅尔频率倒谱系数）特征、PLP(Perceptual Linear Predictive—感知线性预测)特征、或者 LPC（Linear Predictive Coding—线性预测编码）特征等。

下面以提取某一音频帧的 MFCC 特征为例，对特征矢量的提取过程作进一步描述。首先将音频帧的时域信号通过 FFT（Fast Fourier Transformation—快速傅氏变换）得到对应的频谱信息，将所述频谱信息通过 Mel 滤波器组得到 Mel 频谱，在 Mel 频谱上进行倒谱分析，其核心一般是采用 DCT（Discrete Cosine Transform—离散余弦变换）进行逆变换，然后取预设的 N 个系数（例如 N=12 或者 38），则得到了所述音频帧的特征矢量：MFCC 特征。对每个音频帧都采用上述方式进行处理，可以得到表征所述语音信号的一系列特征矢量，即本申请所述的特征矢量序列。

步骤 102、为特征矢量分配角色标签。

本实施例通过建立 GMM 和 HMM 为特征矢量序列中的特征矢量分配角色标签。考虑到在一段语音信号中除了包含对应于各角色的语音信号外，可能还包含没有语音内容的部分，例如：由于倾听、思考等原因造成的静音部分。由于这些部分不包含角色的信息，为了提高角色分离的准确性，可以预先从语音信号中识别并剔除这样的音频帧。

基于上述考虑，本实施例在为特征矢量分配角色标签之前先剔除不包含语音内容的音频帧、并进行语音段的划分，然后在此基础上为各语音段中的特征矢量分配角色标签，所述分配角色标签包括：进行角色的初始划分，在初始划分的基础上迭代训练 GMM 和 HMM，如果训练得到的模型不满足预设要求，则调整角色数量然后重新训练 GMM 和 HMM，直至训练得到的模型满足预设要求，则根据该模型为各语音段中的特征矢量分配角色标签。下面结合附图 3 对上述处理过程进行详细说明。

步骤 102-1、通过识别并剔除不包含语音内容的音频帧、将所述语音信号切分为语音段。

现有技术通常采用声学切分方式，即：根据已有的模型从语音信号中分离出比如“音乐段”、“语音段”、“静音段”等。这种方式需要提前训练好各种音频段对应的声学模型，比如“音乐段”对应的声学模型，基于该声学模型，能够从语音信号中分离出该声学模型对应的音频段。

优选地，本申请的技术方案可以采用 VAD（Voice Activity Detection—语音活动检测）技术识别不包含语音内容的部分，这样，相对于采用声学切分方式的技术，可以不需要

提前训练不同音频段对应的声学模型，适应性更强。例如，可以通过计算音频帧的能量特征、过零率等识别音频帧是否为静音帧，对于存在环境噪声且比较强的情况下，可以综合采用上述多种手段、或者通过建立噪声模型进行识别。

识别出不包含语音内容的音频帧后，一方面可以将这部分音频帧从语音信号中剔除，以提高角色分离的准确性；另一方面通过对不包含语音内容的音频帧的识别，相当于识别出了每段有效语音（包含语音内容）的起点和终点，因此可以在此基础上进行语音段的划分。

请参见附图 4，其为本实施例提供的语音段划分的示意图，在该图中通过 VAD 技术检测出在时间 t_2 与 t_3 之间、以及 t_4 与 t_5 之间的各音频帧为静音帧，本步骤可以从语音信号中剔除这部分静音帧，并相应划分出 3 个语音段：位于 t_1 与 t_2 之间的语音段 1（seg1）、位于 t_3 和 t_4 之间的语音段 2（seg2）、以及位于 t_5 和 t_6 之间的语音段 3（seg3），每个语音段都包含若干个音频帧，每个音频帧都有对应的特征矢量。在划分语音段的基础上，可以粗略地进行角色分配，便于为后续的训练提供一个合理的起点。

优选的，在采用 VAD 技术进行上述处理后，还可以执行 VAD 平滑操作。这主要是考虑到人类实际发声的情况，真实的语音段的持续时间不会太短，如果执行上述 VAD 操作后，得到的某些语音段的持续时间小于预先设置的阈值（例如，语音段长度为 30ms，而预先设置的阈值为 100ms），则可以将这样的语音段与相邻语音段合并，形成较长的语音段。进行 VAD 平滑处理后得到的语音段划分更接近于真实的情况，有助于提高角色分离的准确性。

本步骤通过 VAD 技术将语音信号划分成若干个语音段，后续步骤 102-2 至 102-11 的任务则是利用 GMM 和 HMM 为各语音段中的特征矢量分配角色标签。

步骤 102-2、按照预设的初始角色数量选择相应数量的语音段，并为每个语音段分别指定不同角色。

本步骤可以随机地从已经划分好的语音段中选择与初始角色数量相同的语音段，考虑到所选语音段要用于进行 GMM 和 HMM 的初始训练，如果时长比较短则可用于训练的数据较少，时长太长则包含一个以上角色的可能性就会增加，这两种情况都不利于进行初始训练，因此本实施例提供一种优选实施方式，即：根据初始角色数量选择时长满足预设要求的语音段、并为每个语音段分别指定不同的角色。

本实施例中预设的初始角色数量为 2，预设的选择语音段的要求为：时长在 2s 至 4s 之间，因此本步骤从已经划分好的语音段中选择满足上述要求的 2 个语音段，并为每个

语音段分别指定不同的角色。仍以图 4 所示的语音段划分为例，seg1 和 seg2 各自满足上述的时长要求，因此可以选择 seg1 和 seg2 这两个语音段，并为 seg1 指定角色 1（s1），为 seg2 指定角色 2（s2）。

步骤 102-3、利用指定角色的语音段中的特征矢量，训练针对每个角色的 GMM 以及 HMM。

本步骤根据指定角色的语音段包含的特征矢量，训练针对每个角色的 GMM、以及描述角色间的跳转关系的 HMM，本步骤是在特定角色数量下进行的初始训练。仍以图 4 所示的语音段划分为例，在初始角色数量下，seg1 中包含的特征矢量用于训练角色 1 的 GMM（gmm1），seg2 中包含的特征矢量用于训练角色 2 的 GMM（gmm2），如果在该角色数量下训练得到的 GMM 和 HMM 不满足要求，则可以调整角色数量并重复转到本步骤，根据调整后的角色数量执行相应的初始训练。

针对每个角色训练 GMM 以及 HMM 的过程，也就是在给定观测序列（即：各语音段包含的特征矢量序列，也即训练样本）的基础上学习与 HMM 相关的各个参数的过程，所述各个参数包括：HMM 的转移矩阵 A、每个角色对应的 GMM 的均值向量、协方差矩阵等参数。在具体实施时可以采用 Baum-Welch 算法进行训练，先根据训练样本估计各参数的初始值，根据训练样本和各参数的初始值，估计在时刻 t 处于某一状态 s_j 的后验概率$\gamma t(s_j)$，然后根据计算得到的后验概率更新 HMM 的各个参数，再根据训练样本和更新后的各参数再次估计后验概率$\gamma t(s_j)$.....，反复迭代执行上述过程，直到找到一组 HMM 参数使得输出观测序列的概率最大化。得到满足上述要求的参数后，则在特定角色数量下的 GMM 和 HMM 初始训练完毕。

步骤 102-4、根据训练得到的 GMM 和 HMM 进行解码，获取输出各语音段所包含的特征矢量序列的概率值排序靠前的角色序列。

在步骤 102-1 中已经将语音信号划分为若干个语音段，每个语音段中的每个音频帧都有对应的特征矢量，共同组成本步骤所述的特征矢量序列。本步骤在给定所述特征矢量序列、以及已训练好的 GMM 和 HMM 的基础上，找到所述特征矢量序列可能从属的 HMM 状态序列，即：角色序列。

本步骤完成的功能是通常所述的 HMM 解码过程，根据所述特征矢量序列，搜索输出该特征矢量序列的概率值排序靠前的角色序列，作为优选实施方式，通常可以选择最大概率值对应的角色序列，即最可能输出所述特征矢量序列的角色序列，也称为最优的隐状态序列。

在具体实施时，可以采用穷举搜索的方法，计算每个可能的角色序列输出所述特征矢量序列的概率值，并从中选择最大值。为了提高计算效率，作为优选实施方式，可以采用维特比（Viterbi）算法，利用 HMM 的转移概率在时间上的不变性来降低计算的复杂度，并在搜索得到输出所述特征矢量序列的最大概率值后，根据搜索过程记录的信息进行回溯，获取对应的角色序列。

步骤 102-5、判断所述角色序列对应的概率值是否大于预设阈值，若是，执行步骤 102-6，否则转到步骤 102-7 执行。

如果步骤 102-4 通过解码过程获取的角色序列对应的概率值大于预先设定的阈值，例如：0.5，通常可以认为目前的 GMM 和 HMM 已经稳定，可以执行步骤 102-6 为各语音段中的特征矢量分配角色标签（后续步骤 104 可以利用所述已稳定的 HMM 判定特征矢量序列对应的角色序列），否则转到步骤 102-7 判断是否继续进行迭代训练。

步骤 102-6、按照所述角色序列为各语音段中的特征矢量分配角色标签。

由于目前的 GMM 和 HMM 已经稳定，因此可以按照步骤 102-4 通过解码获取的角色序列为各语音段中的特征矢量分配角色标签。在具体实施时，由于所述角色序列中的每个角色与各语音段中的每个特征矢量是一一对应的，因此可以根据该一一对应关系为每个特征矢量分配角色标签。至此，各语音段中的特征矢量都有了各自的角色标签，步骤 102 执行完毕，可以继续执行步骤 103 。

步骤 102-7、判断在当前角色数量下训练 GMM 和 HMM 的次数是否小于预设的训练次数上限；若是，则执行步骤 102-8，否则转到步骤 102-10 执行。

执行到本步骤，说明目前训练得到的 GMM 和 HMM 还没有稳定，需要继续进行迭代训练。考虑到在训练过程所采用的当前角色数量与实际角色数量（所述语音信号涉及的真实角色数量）不一致的情况下，GMM 和 HMM 即使经过多次迭代训练也可能无法满足要求（解码操作所获取的角色序列对应的概率值始终不满足大于预设阈值的条件），为了避免出现无意义的循环迭代过程，可以预先设置在每种角色数量下训练 GMM 和 HMM 的训练次数上限。如果本步骤判断出在当前角色数量下的训练次数小于所述上限，则继续执行步骤 102-8 为各语音段指定角色以便继续进行迭代训练，否则说明目前采用的角色数量可能与实际情况不一致，因此可以转到步骤 102-10 判断是否需要调整角色数量。

步骤 102-8、根据所述角色序列为每个语音段指定对应的角色。

在步骤 102-4 中已经通过解码获取了角色序列，由于角色序列中的每个角色与各语

音段中的特征矢量是一一对应的，因此可以获知各语音段中每个特征矢量对应的角色。本步骤针对语音信号中的每个语音段，通过计算其中各特征矢量对应的角色的众数，为所述语音段指定角色。例如：某语音段包含 10 个音频帧，也即包含 10 个特征矢量，其中 8 个特征矢量对应角色 1（s1），2 个特征矢量对应于角色 2（s2），那么所述语音段中各特征矢量对应的角色的众数为角色 1（s1），因此将角色 1（s1）指定为所述语音段的角色。

步骤 102-9、根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的 GMM 以及 HMM，并转到步骤 102-4 继续执行。

在步骤 102-8 为每个语音段指定角色的基础上，可以训练针对每个角色的 GMM 以及 HMM。仍以图 4 所示的语音段划分为例，如果步骤 102-8 将 seg1 和 seg3 指定为角色 1（s1），seg2 指定为角色 2（s2），那么 seg1 和 seg3 包含的特征矢量可以用于训练角色 1 的 GMM（gmm1），seg2 中包含的特征矢量用于训练角色 2 的 GMM（gmm2）。GMM 和 HMM 的训练方法请参见步骤 102-3 的相关文字，此处不再赘述。

在具体实施中，本技术方案通常为迭代训练过程，为了提高训练效率，本步骤可以在上一次训练得到的 GMM 和 HMM 的基础上采用增量方式训练新的 GMM 和 HMM，即在上一次训练得到的参数基础上，利用目前的样本数据，继续调整各个参数，从而可以提高训练速度。

完成上述训练过程，得到新的 GMM 和 HMM 后，可以转到步骤 102-4 执行，根据新的模型进行解码以及执行后续的操作。

步骤 102-10、判断当前角色数量是否符合预设要求；若是，转到步骤 102-6 执行，否则继续执行步骤 102-11。

执行到本步骤，通常说明在当前角色数量下训练得到的 GMM 和 HMM 并未稳定、而且训练次数已经等于或者超过了预设的训练次数上限，在这种情况下可以判断当前角色数量是否符合预设要求，若符合，则说明可以停止角色分离过程，转到步骤 102-6 进行角色标签的分配，否则继续执行步骤 102-11 进行角色数量的调整。

步骤 102-11、调整角色数量，选择相应数量的语音段并为每个语音段分别指定不同角色；并转到步骤 102-3 继续执行。

例如，当前角色数量为 2，对角色数量的预设要求为“角色数量等于 4”，步骤 102-10 判定当前角色数量尚未符合预设要求，这种情况下，可以执行本步骤进行角色数量的调整，例如：为当前角色数量加 1，即将当前角色数量更新为 3。

根据调整后的角色数量，从语音信号包含的各个语音段中选择相应数量的语音段，并为所选每个语音段分别指定不同的角色。其中对所选语音段的时长要求，可以参见步骤 102-2 中的相关文字，此处不再赘述。

仍以图 4 所示的语音段划分为例，如果当前角色数量从 2 增加为 3，并且 seg1、seg2 和 seg3 都满足选择语音段的时长要求，那么本步骤可以选择这 3 个语音段，并为 seg1 指定角色 1（s1），为 seg2 指定角色 2（s2），为 seg3 指定角色 3（s3）。

完成上述调整角色数量以及选择语音段的操作后，可以转到步骤 102-3 针对调整后的角色数量初始训练 GMM 和 HMM。

步骤 103、利用具有角色标签的特征矢量训练 DNN 模型。

此时，已经为各语音段中的特征矢量分配了角色标签，在此基础上，本步骤以具有角色标签的特征矢量为样本训练 DNN 模型，所述 DNN 模型用于根据输入的特征矢量输出对应每个角色的概率。为了便于理解，先对 DNN 作简要说明。

DNN（Deep Neural Networks—深度神经网络）通常指包括 1 个输入层、3 个以上隐含层（也可以包含 7 个、9 个、甚至更多的隐含层）、以及 1 个输出层的神经网络。每个隐含层都能够提取一定的特征，并将本层的输出作为下一层的输入，通过逐层提取特征，将低层特征形成更加抽象的高层特征，从而能够实现对物体或者种类的识别。

请参见图 5，其为 DNN 网络的拓扑结构示意图，图中的 DNN 网络总共有 n 层，每层有多个神经元，不同层之间全连接；每层都有自己的激励函数 f（例如 Sigmoid 函数）。输入为特征矢量 v，第 i 层到第 i+1 层的转移矩阵为 $w_{i(i+1)}$，第 i+1 层的偏置矢量为 $b_{(i+1)}$，第 i 层的输出为 out_i，第 i+1 的输入为 in_{i+1}，计算过程为：

$$in_{i+1} = out_i * w_{i(i+1)} + b_{(i+1)}$$

$$out_{i+1} = f(in_{i+1})$$

由此可见 DNN 模型的参数包括层间的转移矩阵 w 和每一层的偏置矢量 b 等，训练 DNN 模型的主要任务就是确定上述参数。在实际应用中通常采用 BP（Back-propagation—反向传播）算法进行训练，训练过程是一个有监督的学习过程：输入信号为带有标签的特征矢量，分层向前传播，到达输出层后再逐层反向传播，通过梯度下降法调整各层的参数以使网络的实际输出不断接近期望输出。对于每层有上千神经元的 DNN 网络来说，其参数的数量可能是百万级的甚至更多，完成上述训练过程获取的 DNN 模型，通常具有非常强大的特征提取能力以及识别能力。

在本实施例中，DNN 模型用于根据输入的特征矢量输出对应每个角色的概率，因此

DNN 模型的输出层可以采用分类器（例如 Softmax）作为激活函数，在步骤 102 完成预先分配角色标签的处理后，如果角色标签涉及的角色数量为 n，那么 DNN 模型的输出层可以包括 n 个节点，分别对应于 n 个角色，针对输入的特征矢量每个节点输出该特征矢量对应所属角色的概率值。

本步骤以带有角色标签的特征矢量作为样本，对构建的上述 DNN 模型进行有监督训练。在具体实施时，可以直接采用上述 BP 算法进行训练，考虑到单纯采用 BP 算法训练有可能出现陷入局部极小值的情况、导致最终得到的模型无法满足应用的需求，因此本实施例采用预训练（pre-training）与 BP 算法相结合的方式进行 DNN 模型的训练。

预训练通常采用非监督贪心逐层训练算法，先采用非监督方式训练含有一个隐层的网络，然后保留训练好的参数，使网络层数加 1，接着训练含两个隐层的网络.....以此类推，直到含有最大隐层的网络。这样逐层训练完之后，以该无监督训练过程学习到的参数值作为初始值，再采用传统 BP 算法进行有监督的训练，最终得到 DNN 模型。

由于经过 pre-training 得到的初始分布比纯 BP 算法采用的随机初始参数更接近于最终的收敛值，相当于使后续的有监督训练过程有一个好的起点，因此训练得到的 DNN 模型通常不会陷入局部极小值，能够获得较高的识别率。

步骤 104、根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定特征矢量序列对应的角色序列，并输出角色分离结果。

由于所述 DNN 模型用于根据输入的特征矢量输出对应每个角色的概率，同时根据特征矢量序列的角色标签的分布情况可以获知对应每个角色的先验概率，而每个特征矢量的先验概率通常也是固定的，因此依据贝叶斯定理，根据 DNN 模型的输出以及上述先验概率可以获知每个角色输出相应特征矢量的概率，也即可以采用步骤 103 训练好的 DNN 模型决定 HMM 各状态的发射概率。

所述 HMM 可以是在采用上述 DNN 模型决定 HMM 发射概率的基础上，用特征矢量序列训练得到的。考虑到在步骤 102 为特征矢量分配角色标签时所采用的 HMM 对各角色之间的跳转关系的描述已基本稳定，可以不再进行额外的训练，因此本实施例直接采用该 HMM，并用训练得到的 DNN 模型替换 GMM，即：由 DNN 模型决定 HMM 各状态的发射概率。

在本实施例中，步骤 102-1 进行了语音段的切分，本步骤根据所述 DNN 模型和预先分配角色标签时所采用的 HMM，判定各语音段所包含的特征矢量序列对应的角色序列。

根据特征矢量序列确定角色序列的过程是通常所述的解码问题，可以根据所述 DNN

模型和 HMM 执行解码操作，获取输出所述特征矢量序列的概率值排序靠前（例如概率值最大）的角色序列，并将所述角色序列作为与所述特征矢量序列对应的角色序列。具体说明请参见步骤 102-4 中的相关文字，此处不再赘述。

通过解码过程得到与各语音段所包含的特征矢量序列对应的角色序列后，则可以输出相应的角色分离结果。由于角色序列中的每个角色与特征矢量是一一对应的，而每个特征矢量对应的音频帧都有各自的时间起止点，因此本步骤可以针对每个角色输出与其对应的特征矢量所属音频帧的起止时间信息。

至此，通过步骤 101 至步骤 104，对本申请提供的基于语音的角色分离方法的具体实施方式进行了详细的说明。需要说明的是，本实施例在步骤 102 为特征矢量预先分配角色标签的过程中采用了自顶向下、逐渐增加角色数量的方式。在其他实施方式中，也可以采用自底向上、逐渐减少角色数量的方式：最初可以将切分得到的每个语音段分别指定给不同的角色，然后训练针对每个角色的 GMM 和 HMM，如果通过迭代训练得到的 GMM 和 HMM 在执行解码操作后得到的概率值始终不大于预设阈值，那么在调整角色数量时，可以通过评估每个角色的 GMM 彼此之间的相似度（例如计算 KL 散度），将相似度满足预设要求的 GMM 对应的语音段进行合并，并相应减少角色数量，重复迭代执行上述过程，直到 HMM 通过解码得到的概率值大于预设阈值或者角色数量符合预设要求，则停止迭代过程，并根据解码得到的角色序列为各语音段中的特征矢量分配角色标签。

综上所述，本申请提供的基于语音的角色分离方法，由于采用具有强大特征提取能力的 DNN 模型对角色进行建模，比传统的 GMM 具有更为强大的刻画能力，对角色的刻画更加精细、准确，因此能够获得更为准确的角色分离结果。本申请的技术方案不仅可以应用于对客服中心、会议语音等对话语音进行角色分离的场景中，还可以应用于其它需要对语音信号中的角色进行分离的场景中，只要所述语音信号中包含两个或者两个以上角色，就都可以采用本申请的技术方案，并取得相应的有益效果。

在上述的实施例中，提供了一种基于语音的角色分离方法，与之相对应的，本申请还提供一种基于语音的角色分离装置。请参看图 6，其为本申请的一种基于语音的角色分离装置的实施例示意图。由于装置实施例基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。下述描述的装置实施例仅仅是示意性的。

本实施例的一种基于语音的角色分离装置，包括：特征提取单元 601，用于从语音

信号中逐帧提取特征矢量，得到特征矢量序列；标签分配单元 602，用于为特征矢量分配角色标签；DNN 模型训练单元 603，用于利用具有角色标签的特征矢量训练 DNN 模型，其中所述 DNN 模型用于根据输入的特征矢量输出对应每个角色的概率；角色判定单元 604，用于根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定特征矢量序列对应的角色序列并输出角色分离结果，其中所述 HMM 用于描述角色间的跳转关系。

可选的，所述装置还包括：

语音段切分单元，用于在所述特征提取单元提取特征矢量后、在触发所述标签分配单元工作之前，通过识别并剔除不包含语音内容的音频帧、将所述语音信号切分为语音段；

所述标签分配单元具体用于，为各语音段中的特征矢量分配角色标签；

所述角色判定单元具体用于，根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定各语音段所包含的特征矢量序列对应的角色序列并输出角色分离结果。

可选的，所述标签分配单元具体用于，通过建立 GMM 和 HMM，为各语音段中的特征矢量预先分配角色标签，其中所述 GMM 用于针对每个角色、根据输入的特征矢量输出该特征矢量对应于所述角色的概率；

所述角色判定单元具体用于，根据所述 DNN 模型和为各语音段中的特征矢量分配角色标签所采用的 HMM，判定所述各语音段所包含的特征矢量序列对应的角色序列。

可选的，所述标签分配单元包括：

初始角色指定子单元，用于按照预设的初始角色数量选择相应数量的语音段，并为每个语音段分别指定不同角色；

初始模型训练子单元，用于利用指定角色的语音段中的特征矢量，训练针对每个角色的 GMM 以及 HMM；

解码子单元，用于根据训练得到的 GMM 和 HMM 进行解码，获取输出各语音段所包含的特征矢量序列的概率值排序靠前的角色序列；

概率判断子单元，用于判断所述角色序列对应的概率值是否大于预设阈值；

标签分配子单元，用于当所述概率判断子单元的输出为是时，按照所述角色序列为各语音段中的特征矢量分配角色标签。

可选的，所述标签分配单元还包括：

逐语音段角色指定子单元，用于当所述概率判断子单元的输出为否时，根据所述角色序列，为每个语音段指定对应的角色；

模型更新训练子单元，用于根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的 GMM 以及 HMM，并触发所述解码子单元工作。

可选的，所述逐语音段角色指定子单元具体用于，针对每个语音段，将其中各特征矢量对应的角色的众数指定为所述语音段的角色。

可选的，所述模型更新训练子单元具体用于，在上一次训练得到的模型基础上采用增量方式训练所述 GMM 以及 HMM。

可选的，所述标签分配单元还包括：

训练次数判断子单元，用于当所述概率判断子单元的输出为否时，判断在当前角色数量下训练 GMM 和 HMM 的次数是否小于预设的训练次数上限，并在判断结果为是时，触发所述逐语音段角色指定子单元工作。

角色数量调整子单元，用于当所述训练次数判断子单元的输出为否时，调整角色数量，选择相应数量的语音段并为每个语音段分别指定不同角色，并触发所述初始模型训练子单元工作。

可选的，所述标签分配单元还包括：

角色数量判断子单元，用于当所述训练次数判断子单元的输出为否时，判断当前角色数量是否符合预设要求，若符合则触发所述标签分配子单元工作，否则触发所述角色数量调整子单元工作。

可选的，所述特征提取单元包括：

分帧子单元，用于按照预先设定的帧长度对语音信号进行分帧处理，得到多个音频帧；

特征提取执行子单元，用于提取各音频帧的特征矢量，得到所述特征矢量序列。

可选的，所述特征提取执行子单元具体用于，提取各音频帧的 MFCC 特征、PLP 特征、或者 LPC 特征，得到所述特征矢量序列。

可选的，所述语音段切分单元具体用于，通过采用 VAD 技术识别并剔除所述不包含语音内容的音频帧、将所述语音信号切分为语音段。

可选的，所述装置还包括：

VAD 平滑单元，用于在所述语音段切分单元采用 VAD 技术切分语音段后，将时长小于预设阈值的语音段与相邻语音段合并。

可选的，所述 DNN 模型训练单元具体用于，采用反向传播算法训练所述 DNN 模型。

可选的，所述角色判定单元具体用于，根据所述 DNN 模型和 HMM 执行解码操作，

获取输出所述特征矢量序列的概率值排序靠前的角色序列，并将所述角色序列作为与所述特征矢量序列对应的角色序列。

可选的，所述角色判定单元采用如下方式输出角色分离结果：根据特征矢量序列对应的角色序列，针对每个角色输出与其对应的特征矢量所属音频帧的起止时间信息。

可选的，所述初始角色指定子单元或所述角色数量调整子单元具体通过如下方式选择相应数量的语音段：选择时长满足预设要求的、所述数量的语音段。

本申请虽然以较佳实施例公开如上，但其并不是用来限定本申请，任何本领域技术人员在不脱离本申请的精神和范围内，都可以做出可能的变动和修改，因此本申请的保护范围应当以本申请权利要求所界定的范围为准。

在一个典型的配置中，计算设备包括一个或多个处理器 (CPU)、 输入/输出接口、网络接口和内存。

内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器 (RAM) 和/或非易失性内存等形式，如只读存储器 (ROM) 或闪存(flash RAM)。内存是计算机可读介质的示例。

1、计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、 程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、 其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非暂存电脑可读媒体 (transitory media)，如调制的数据信号和载波。

2、本领域技术人员应明白，本申请的实施例可提供为方法、系统或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

权 利 要 求 书

1、一种基于语音的角色分离方法，其特征在于，包括：

从语音信号中逐帧提取特征矢量，得到特征矢量序列；

为特征矢量分配角色标签；

利用具有角色标签的特征矢量训练深度神经网络 DNN 模型；

根据所述 DNN 模型和利用特征矢量训练得到的隐马尔科夫模型 HMM，判定特征矢量序列对应的角色序列，并输出角色分离结果；

其中，所述 DNN 模型用于根据输入的特征矢量输出对应每个角色的概率，HMM 用于描述角色间的跳转关系。

2、根据权利要求 1 所述的基于语音的角色分离方法，其特征在于，在所述从语音信号中逐帧提取特征矢量的步骤之后、在所述为特征矢量分配角色标签的步骤之前，执行下述操作：通过识别并剔除不包含语音内容的音频帧、将所述语音信号切分为语音段；

所述为特征矢量分配角色标签包括：为各语音段中的特征矢量分配角色标签；所述判定特征矢量序列对应的角色序列包括：判定各语音段所包含的特征矢量序列对应的角色序列。

3、根据权利要求 2 所述的基于语音的角色分离方法，其特征在于，所述为各语音段中的特征矢量分配角色标签包括：通过建立高斯混合模型 GMM 和 HMM，为各语音段中的特征矢量分配角色标签；其中所述 GMM 用于针对每个角色、根据输入的特征矢量输出该特征矢量对应于所述角色的概率；

所述根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定各语音段所包含的特征矢量序列对应的角色序列包括：根据所述 DNN 模型和为各语音段中的特征矢量分配角色标签所采用的 HMM，判定所述各语音段所包含的特征矢量序列对应的角色序列。

4、根据权利要求 3 所述的基于语音的角色分离方法，其特征在于，所述通过建立高斯混合模型 GMM 和 HMM，为各语音段中的特征矢量分配角色标签，包括：

按照预设的初始角色数量选择相应数量的语音段，并为每个语音段分别指定不同角色；

利用指定角色的语音段中的特征矢量，训练针对每个角色的 GMM 以及 HMM；

根据训练得到的 GMM 和 HMM 进行解码，获取输出各语音段所包含的特征矢量序列的概率值排序靠前的角色序列；

判断所述角色序列对应的概率值是否大于预设阈值；若是，按照所述角色序列为各语音段中的特征矢量分配角色标签。

5、根据权利要求 4 所述的基于语音的角色分离方法，其特征在于，当所述判断所述角色序列对应的概率值是否大于预设阈值的结果为否时，执行下述操作：

根据所述角色序列，为每个语音段指定对应的角色；

根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的 GMM 以及 HMM；

转到所述根据训练得到的 GMM 和 HMM 进行解码的步骤执行。

6、根据权利要求 5 所述的基于语音的角色分离方法，其特征在于，所述根据所述角色序列，为每个语音段指定对应的角色，包括：

针对每个语音段，将其中各特征矢量对应的角色的众数指定为所述语音段的角色。

7、根据权利要求 5 所述的基于语音的角色分离方法，其特征在于，所述根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的 GMM 以及 HMM，包括：

在上一次训练得到的模型基础上采用增量方式训练所述 GMM 以及 HMM。

8、根据权利要求 5 所述的基于语音的角色分离方法，其特征在于，当所述判断所述角色序列对应的概率值是否大于预设阈值的结果为否时，执行下述操作：

判断在当前角色数量下训练 GMM 和 HMM 的次数是否小于预设的训练次数上限；

若是，执行所述根据所述角色序列为每个语音段指定对应的角色的步骤；

若否，执行下述操作：

调整角色数量，选择相应数量的语音段并为每个语音段分别指定不同角色；

并转到所述利用指定角色的语音段中的特征矢量，训练针对每个角色的 GMM 以及 HMM 的步骤执行。

9、根据权利要求 8 所述的基于语音的角色分离方法，其特征在于，当所述判断在当前角色数量下训练 GMM 和 HMM 的次数是否小于预设的训练次数上限的结果为否时，执行下述操作：

判断当前角色数量是否符合预设要求；若是，转到所述按照所述角色序列为各语音段中的特征矢量分配角色标签的步骤执行，若否，则执行所述调整角色数量的步骤。

10、根据权利要求 8 所述的基于语音的角色分离方法，其特征在于，所述预设的初始角色数量为 2，所述调整角色数量包括：为当前角色数量加 1。

11、根据权利要求 1 所述的基于语音的角色分离方法，其特征在于，所述从语音信

号中逐帧提取特征矢量，得到特征矢量序列包括：

按照预先设定的帧长度对语音信号进行分帧处理，得到多个音频帧；

提取各音频帧的特征矢量，得到所述特征矢量序列。

12、根据权利要求 11 所述的基于语音的角色分离方法，其特征在于，所述提取各音频帧的特征矢量包括：提取 MFCC 特征、PLP 特征、或者 LPC 特征。

13、根据权利要求 2 所述的基于语音的角色分离方法，其特征在于，所述识别并剔除不包含语音内容的音频帧包括：采用 VAD 技术识别所述不包含语音内容的音频帧、并执行相应的剔除操作。

14、根据权利要求 13 所述的基于语音的角色分离方法，其特征在于，在采用 VAD 技术执行所述识别及剔除操作、并将所述语音信号切分为语音段之后，执行下述 VAD 平滑操作：

将时长小于预设阈值的语音段与相邻语音段合并。

15、根据权利要求 1 所述的基于语音的角色分离方法，其特征在于，所述利用具有角色标签的特征矢量训练深度神经网络 DNN 模型包括：采用反向传播算法训练所述 DNN 模型。

16、根据权利要求 1 所述的基于语音的角色分离方法，其特征在于，所述根据所述 DNN 模型和利用特征矢量训练得到的隐马尔科夫模型 HMM，判定特征矢量序列对应的角色序列，包括：根据所述 DNN 模型和 HMM 执行解码操作，获取输出所述特征矢量序列的概率值排序靠前的角色序列，并将所述角色序列作为与所述特征矢量序列对应的角色序列。

17、根据权利要求 1 所述的基于语音的角色分离方法，其特征在于，所述输出角色分离结果包括：根据特征矢量序列对应的角色序列，针对每个角色输出与其对应的特征矢量所属音频帧的起止时间信息。

18、根据权利要求 4 或 8 所述的基于语音的角色分离方法，其特征在于，所述选择相应数量的语音段，包括：选择时长满足预设要求的、所述数量的语音段。

19、一种基于语音的角色分离装置，其特征在于，包括：

特征提取单元，用于从语音信号中逐帧提取特征矢量，得到特征矢量序列；

标签分配单元，用于为特征矢量分配角色标签；

DNN 模型训练单元，用于利用具有角色标签的特征矢量训练 DNN 模型，其中所述 DNN 模型用于根据输入的特征矢量输出对应每个角色的概率；

角色判定单元，用于根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定特征矢量序列对应的角色序列并输出角色分离结果，其中所述 HMM 用于描述角色间的跳转关系。

20、根据权利要求 19 所述的基于语音的角色分离装置，其特征在于，还包括：

语音段切分单元，用于在所述特征提取单元提取特征矢量后、在触发所述标签分配单元工作之前，通过识别并剔除不包含语音内容的音频帧、将所述语音信号切分为语音段；

所述标签分配单元具体用于，为各语音段中的特征矢量分配角色标签；

所述角色判定单元具体用于，根据所述 DNN 模型和利用特征矢量训练得到的 HMM，判定各语音段所包含的特征矢量序列对应的角色序列并输出角色分离结果。

21、根据权利要求 20 所述的基于语音的角色分离装置，其特征在于，所述标签分配单元具体用于，通过建立 GMM 和 HMM，为各语音段中的特征矢量分配角色标签，其中所述 GMM 用于针对每个角色、根据输入的特征矢量输出该特征矢量对应于所述角色的概率；

所述角色判定单元具体用于，根据所述 DNN 模型和为各语音段中的特征矢量分配角色标签所采用的 HMM，判定所述各语音段所包含的特征矢量序列对应的角色序列。

22、根据权利要求 21 所述的基于语音的角色分离装置，其特征在于，所述标签分配单元包括：

初始角色指定子单元，用于按照预设的初始角色数量选择相应数量的语音段，并为每个语音段分别指定不同角色；

初始模型训练子单元，用于利用指定角色的语音段中的特征矢量，训练针对每个角色的 GMM 以及 HMM；

解码子单元，用于根据训练得到的 GMM 和 HMM 进行解码，获取输出各语音段所包含的特征矢量序列的概率值排序靠前的角色序列；

概率判断子单元，用于判断所述角色序列对应的概率值是否大于预设阈值；

标签分配子单元，用于当所述概率判断子单元的输出为是时，按照所述角色序列为各语音段中的特征矢量分配角色标签。

23、根据权利要求 22 所述的基于语音的角色分离装置，其特征在于，所述标签分配单元还包括：

逐语音段角色指定子单元，用于当所述概率判断子单元的输出为否时，根据所述角

色序列，为每个语音段指定对应的角色；

模型更新训练子单元，用于根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的 GMM 以及 HMM，并触发所述解码子单元工作。

24、根据权利要求 23 所述的基于语音的角色分离装置，其特征在于，所述逐语音段角色指定子单元具体用于，针对每个语音段，将其中各特征矢量对应的角色的众数指定为所述语音段的角色。

25、根据权利要求 23 所述的基于语音的角色分离装置，其特征在于，所述模型更新训练子单元具体用于，在上一次训练得到的模型基础上采用增量方式训练所述 GMM 以及 HMM。

26、根据权利要求 23 所述的基于语音的角色分离装置，其特征在于，所述标签分配单元还包括：

训练次数判断子单元，用于当所述概率判断子单元的输出为否时，判断在当前角色数量下训练 GMM 和 HMM 的次数是否小于预设的训练次数上限，并在判断结果为是时，触发所述逐语音段角色指定子单元工作；

角色数量调整子单元，用于当所述训练次数判断子单元的输出为否时，调整角色数量，选择相应数量的语音段并为每个语音段分别指定不同角色，并触发所述初始模型训练子单元工作。

27、根据权利要求 26 所述的基于语音的角色分离装置，其特征在于，所述标签分配单元还包括：

角色数量判断子单元，用于当所述训练次数判断子单元的输出为否时，判断当前角色数量是否符合预设要求，若符合则触发所述标签分配子单元工作，否则触发所述角色数量调整子单元工作。

28、根据权利要求 19 所述的基于语音的角色分离装置，其特征在于，所述特征提取单元包括：

分帧子单元，用于按照预先设定的帧长度对语音信号进行分帧处理，得到多个音频帧；

特征提取执行子单元，用于提取各音频帧的特征矢量，得到所述特征矢量序列。

29、根据权利要求 28 所述的基于语音的角色分离装置，其特征在于，所述特征提取执行子单元具体用于，提取各音频帧的 MFCC 特征、PLP 特征、或者 LPC 特征，得到所述特征矢量序列。

30、根据权利要求 20 所述的基于语音的角色分离装置，其特征在于，所述语音段切分单元具体用于，通过采用 VAD 技术识别并剔除所述不包含语音内容的音频帧、将所述语音信号切分为语音段。

31、根据权利要求 30 所述的基于语音的角色分离装置，其特征在于，还包括：

VAD 平滑单元，用于在所述语音段切分单元采用 VAD 技术切分语音段后，将时长小于预设阈值的语音段与相邻语音段合并。

32、根据权利要求 19 所述的基于语音的角色分离装置，其特征在于，所述 DNN 模型训练单元具体用于，采用反向传播算法训练所述 DNN 模型。

33、根据权利要求 19 所述的基于语音的角色分离装置，其特征在于，所述角色判定单元具体用于，根据所述 DNN 模型和 HMM 执行解码操作，获取输出所述特征矢量序列的概率值排序靠前的角色序列，并将所述角色序列作为与所述特征矢量序列对应的角色序列。

34、根据权利要求 19 所述的基于语音的角色分离装置，其特征在于，所述角色判定单元采用如下方式输出角色分离结果：根据特征矢量序列对应的角色序列，针对每个角色输出与其对应的特征矢量所属音频帧的起止时间信息。

35、根据权利要求 22 或 26 所述的基于语音的角色分离装置，其特征在于，所述初始角色指定子单元或所述角色数量调整子单元具体通过如下方式选择相应数量的语音段：选择时长满足预设要求的、所述数量的语音段。

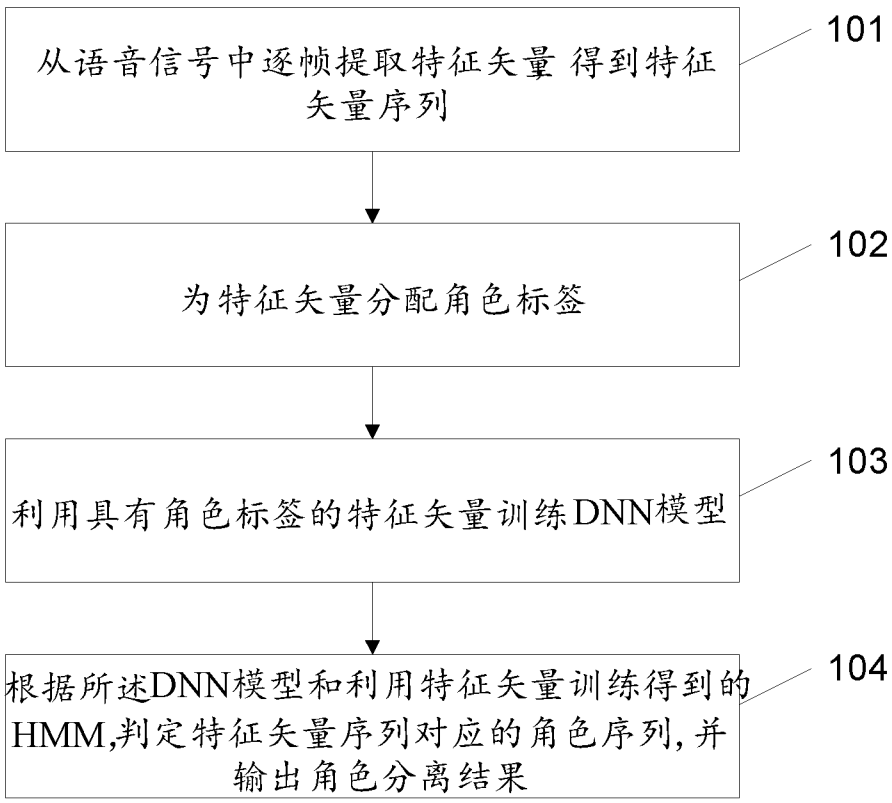
从语音信号中逐帧提取特征矢量 得到特征矢量序列
101
为特征矢量分配角色标签
102
利用具有角色标签的特征矢量训练DNN模型
103
根据所述DNN模型和利用特征矢量训练得到的HMM,判定特征矢量序列对应的角色序列,并输出角色分离结果
104

图 1

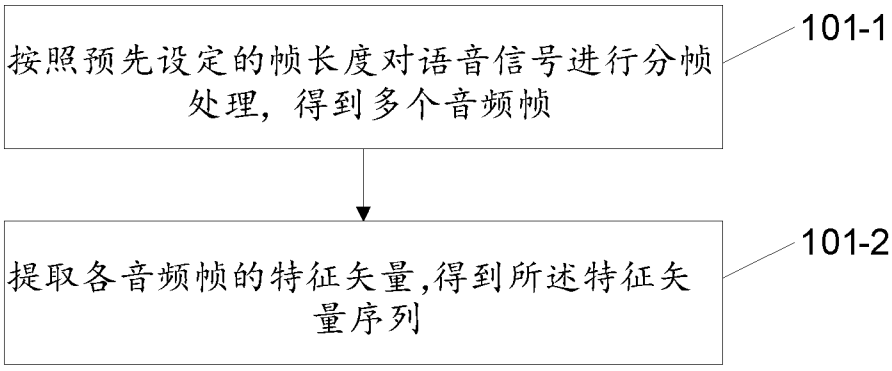
按照预先设定的帧长度对语音信号进行分帧处理，得到多个音频帧
101-1
提取各音频帧的特征矢量,得到所述特征矢量序列
101-2

图 2

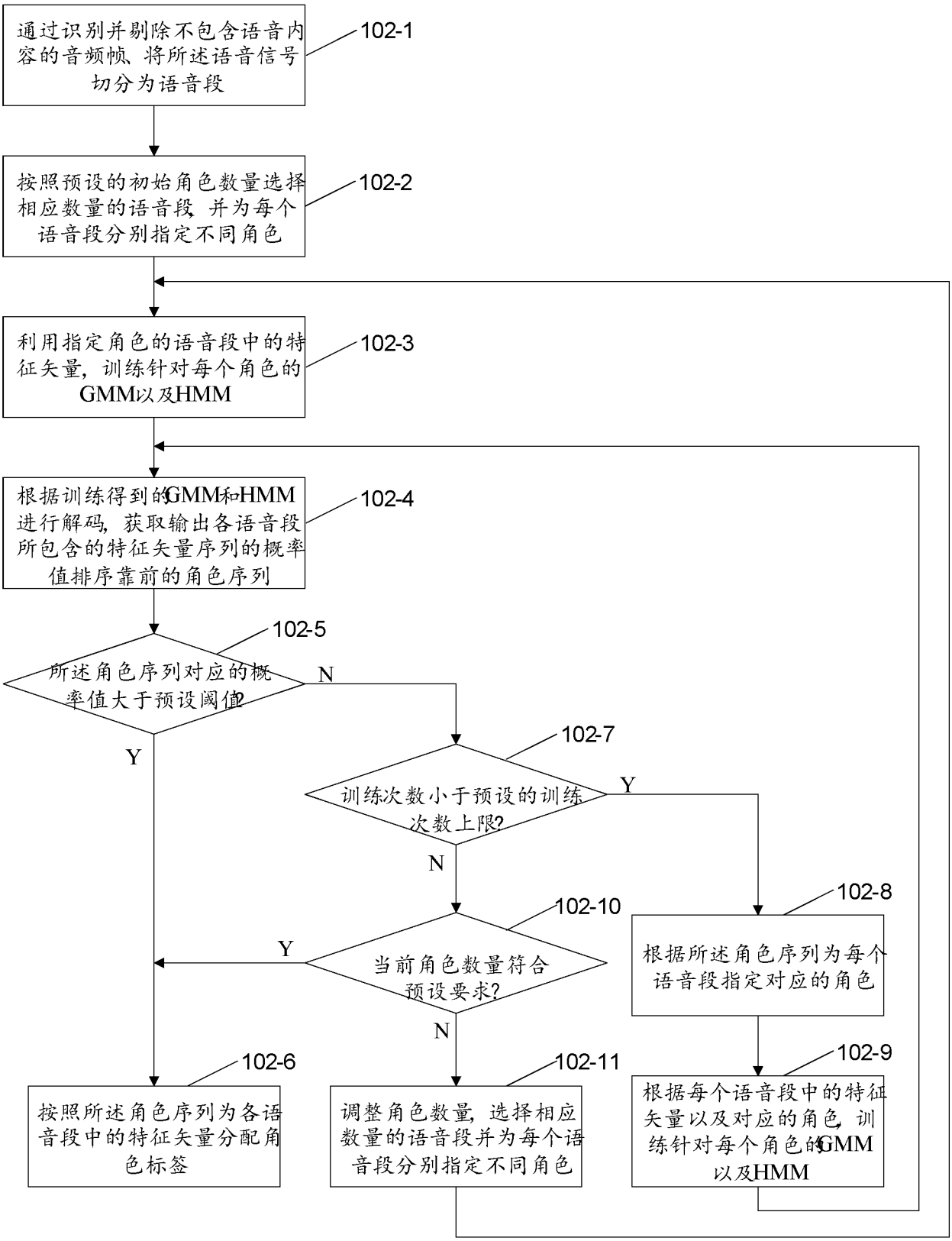
通过识别并剔除不包含语音内容的音频帧，将所述语音信号切分为语音段
102-1
按照预设的初始角色数量选择相应数量的语音段，并为每个语音段分别指定不同角色
102-2
利用指定角色的语音段中的特征矢量，训练针对每个角色的GMM以及HMM
102-3
根据训练得到的GMM和HMM进行解码，获取输出各语音段所包含的特征矢量序列的概率值排序靠前的角色序列
102-4
所述角色序列对应的概率值大于预设阈值?
102-5
N
Y
训练次数小于预设的训练次数上限?
102-7
Y
N
当前角色数量符合预设要求?
102-10
Y
N
根据所述角色序列为每个语音段指定对应的角色
102-8
按照所述角色序列为各语音段中的特征矢量分配角色标签
102-6
调整角色数量，选择相应数量的语音段并为每个语音段分别指定不同角色
102-11
根据每个语音段中的特征矢量以及对应的角色，训练针对每个角色的GMM以及HMM
102-9

图 3

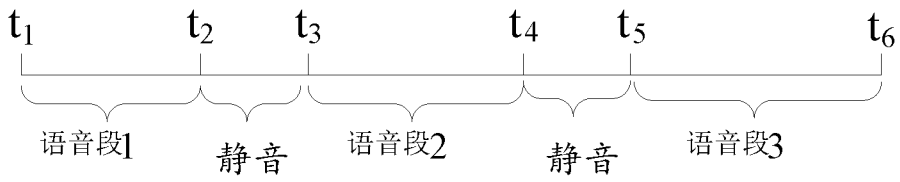

图 4

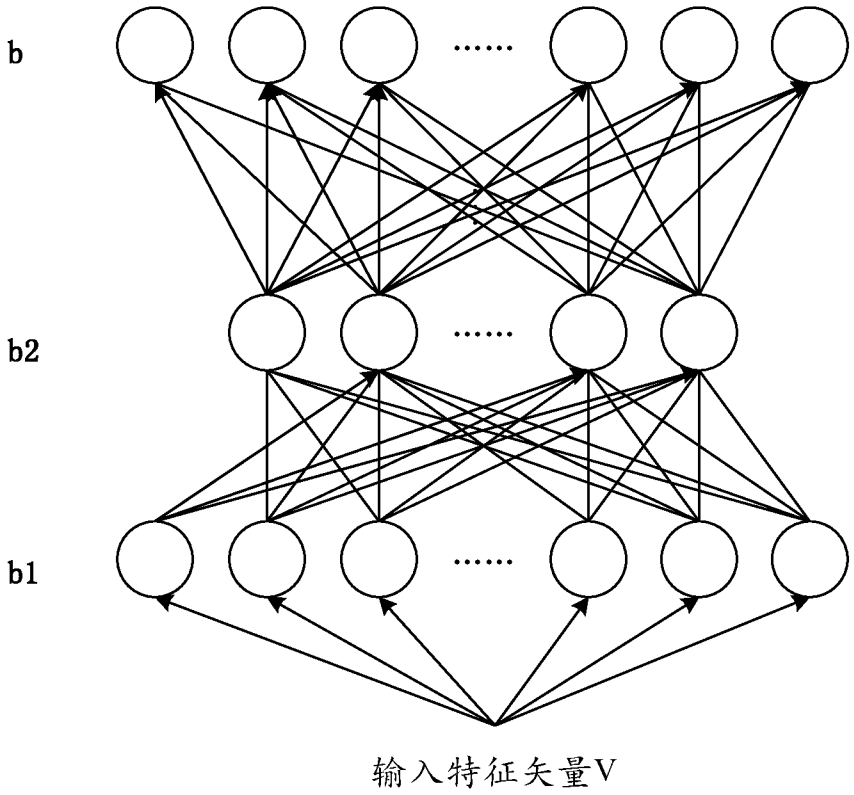

图 5

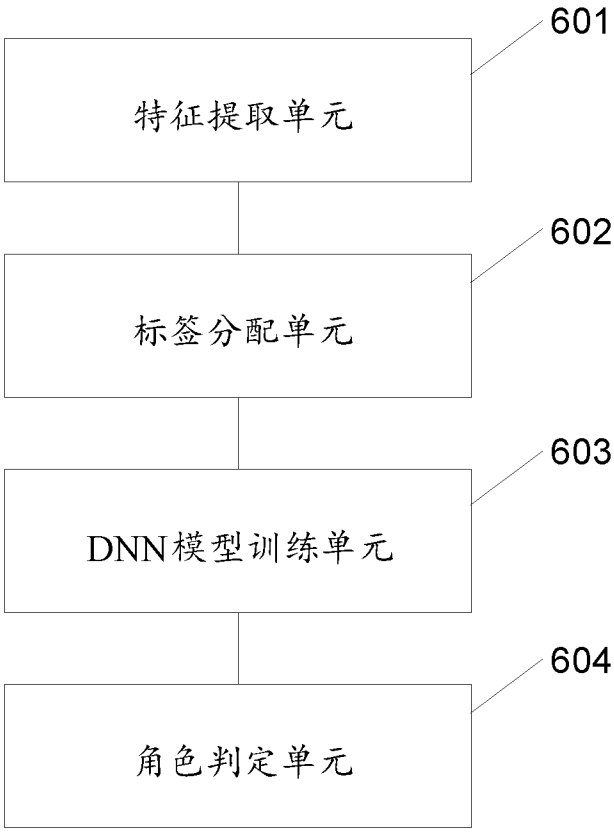

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/103490

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/00 (2013.01) i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

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)

CNABS, CNTXT, VEN, CNKI, BAIDU SCHOLAR: role, affirm, neural network, hidden markov model, gaussian mixture, speaker, separat+, verificat+, recognit+, identificat+, voiceprint, DNN, HMM, GMM

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 103971690 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 06 August 2014 (06.08.2014), description, paragraphs [0027] and [0038]-[0051], and figure 1	1-35
Y	Miao, Yajie et al. "Improving Low-Resource CD-DNN-HMM Using Dropout and Multilingual DNN Training.", PROCEEDINGS OF INTERSPEECH., 29 August 2013 (29.08.2013), pages 2237-2241	1-35
A	CN 104732978 A (SHANGHAI JIAO TONG UNIVERSITY et al.), 24 June 2015 (24.06.2015), the whole document	1-35
A	CN 1366295 A (PANASONIC CORPORATION), 28 August 2002 (28.08.2002), the whole document	1-35
A	CN 103531199 A (FUZHOU UNIVERSITY), 22 January 2014 (22.01.2014), the whole document	1-35
A	CN 104751227 A (ANHUI USTC IFLYTEK CO., LTD.), 01 July 2015 (01.07.2015), the whole document	1-35

☒ 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	Date of mailing of the international search report
22 January 2017 (22.01.2017)	**07 February 2017 (07.02.2017)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **YOU, Xiaomei** Telephone No.: (86-10) **62089539**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/103490

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104157290 A (DALIAN UNIVERSITY OF TECHNOLOGY), 19 November 2014 (19.11.2014), the whole document	1-35
A	CN 101814159 A (YU, Hua), 25 August 2010 (25.08.2010), the whole document	1-35
A	CN 104143327 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 12 November 2014 (12.11.2014), the whole document	1-35

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/103490

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103971690 A	06 August 2014	TW 201430830 A	01 August 2014
		WO 2014114116 A1	31 July 2014
		TW I527023 B	21 March 2016
		US 20160358610 A1	08 December 2016
		US 9502038 B2	22 November 2016
		US 2014214417 A1	31 July 2014
CN 104732978 A	24 June 2015	None	
CN 1366295 A	28 August 2002	JP 2002082694 A	22 March 2002
		EP 1178467 B1	09 March 2005
		DE 60109240 D1	14 April 2005
		CN 1253851 C	26 April 2006
		ES 2239650 T3	01 October 2005
		EP 1178467 A1	06 February 2002
		DE 60109240 T2	16 February 2006
		US 6697778 B1	24 February 2004
CN 103531199 A	22 January 2014	CN 103531199 B	09 March 2016
CN 104751227 A	01 July 2015	None	
CN 104157290 A	19 November 2014	None	
CN 101814159 A	25 August 2010	CN 101814159 B	24 July 2013
CN 104143327 A	12 November 2014	TW I530940 B	21 April 2016
		TW 201503106 A	16 January 2015
		WO 2015003436 A1	15 January 2015
		CN 104143327 B	09 December 2015
		US 2015019214 A1	15 January 2015
		US 9508347 B2	29 November 2016
		HK 1199672 A0	10 July 2015

国际检索报告

国际申请号
PCT/CN2016/103490

A. 主题的分类

G10L 17/00(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

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）)

CNABS,CNTXT,VEN,CNKI,百度学术：角色，分离，分开，辨识，说话人，声纹，确认，验证，认证，识别，深度神经网络，隐马尔科夫，隐马尔可夫，高斯混合，speaker,separat+,verificat+,recognit+,identificat+,voiceprint,DNN,HMM,GMM

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
Y	CN 103971690 A (腾讯科技深圳有限公司) 2014年 8月 6日 (2014-08-06) 说明书第[0027]，[0038]-[0051]段，图1	1-35
Y	Yajie Miao等. "Improving Low-Resource CD-DNN-HMM using Dropout and Multilingual DNN Training." Proceedings of INTERSPEECH., 2013年 8月 29日 (2013-08-29), 第2237-2241页.	1-35
A	CN 104732978 A (上海交通大学 等) 2015年 6月 24日 (2015-06-24) 全文	1-35
A	CN 1366295 A (松下电器产业株式会社) 2002年 8月 28日 (2002-08-28) 全文	1-35
A	CN 103531199 A (福州大学) 2014年 1月 22日 (2014-01-22) 全文	1-35
A	CN 104751227 A (安徽科大讯飞信息科技股份有限公司) 2015年 7月 1日 (2015-07-01) 全文	1-35

☑ 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	国际检索报告邮寄日期
2017年 1月 22日	2017年 2月 7日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(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 游晓梅 电话号码 (86-10)62089539

表 PCT/ISA/210 (第2页) (2009年7月)

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 104157290 A (大连理工大学) 2014年 11月 19日 (2014-11-19) 全文	1-35
A	CN 101814159 A (余华) 2010年 8月 25日 (2010-08-25) 全文	1-35
A	CN 104143327 A (腾讯科技深圳有限公司) 2014年 11月 12日 (2014-11-12) 全文	1-35

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/103490

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	103971690	A	2014年 8月 6日	TW	201430830	A	2014年 8月 1日
				WO	2014114116	A1	2014年 7月 31日
				TW	I527023	B	2016年 3月 21日
				US	20160358610	A1	2016年 12月 8日
				US	9502038	B2	2016年 11月 22日
				US	2014214417	A1	2014年 7月 31日
CN	104732978	A	2015年 6月 24日		无		
CN	1366295	A	2002年 8月 28日	JP	2002082694	A	2002年 3月 22日
				EP	1178467	B1	2005年 3月 9日
				DE	60109240	D1	2005年 4月 14日
				CN	1253851	C	2006年 4月 26日
				ES	2239650	T3	2005年 10月 1日
				EP	1178467	A1	2002年 2月 6日
				DE	60109240	T2	2006年 2月 16日
				US	6697778	B1	2004年 2月 24日
CN	103531199	A	2014年 1月 22日	CN	103531199	B	2016年 3月 9日
CN	104751227	A	2015年 7月 1日		无		
CN	104157290	A	2014年 11月 19日		无		
CN	101814159	A	2010年 8月 25日	CN	101814159	B	2013年 7月 24日
CN	104143327	A	2014年 11月 12日	TW	I530940	B	2016年 4月 21日
				TW	201503106	A	2015年 1月 16日
				WO	2015003436	A1	2015年 1月 15日
				CN	104143327	B	2015年 12月 9日
				US	2015019214	A1	2015年 1月 15日
				US	9508347	B2	2016年 11月 29日
				HK	1199672	A0	2015年 7月 10日