(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
2018 年 12 月 20 日 (20.12.2018)

WIPO | PCT

(10) 国际公布号
WO 2018/227780 A1

(51) 国际专利分类号:
G10L 15/02 (2006.01) *G10L 15/14* (2006.01)

(21) 国际申请号: PCT/CN2017/100043

(22) 国际申请日: 2017 年 8 月 31 日 (31.08.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710445076.9 2017年6月12日 (12.06.2017) 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八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。

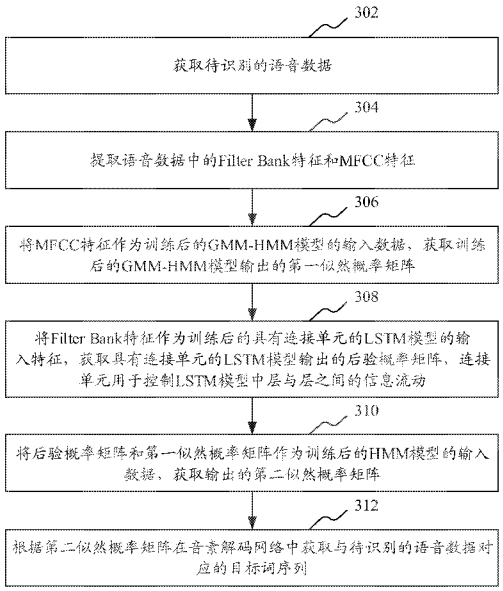
(72) 发明人: 梁浩 (LIANG, Hao); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 王健宗 (WANG, Jianzong); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 程宁 (CHENG, Ning); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 肖京 (XIAO, Jing); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区花城大道 85 号 3901 房, Guangdong 510623 (CN)。

(81) 指定国（除另有指明，要求每一种可提供的国家保护）: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,

(54) **Title:** SPEECH RECOGNITION METHOD AND DEVICE, COMPUTER DEVICE AND STORAGE MEDIUM

(54) **发明名称**: 语音识别方法、装置、计算机设备及存储介质



302 Acquire speech data to be recognized
304 Extracting a Filter Bank feature and a Mel-Frequency Cepstral Coefficient (MFCC) feature from the speech data
306 Using the MFCC feature as input data of a trained Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) model, acquire a first likelihood probability matrix outputted by the trained GMM-HMM model
308 Using the Filter Bank feature as an input feature of a trained long short-term memory (LSTM) model which has a connection unit, acquire a posterior probability matrix outputted by the LSTM model which has the connection unit, the connection unit being used for controlling the flow of information between layers in the LSTM model
310 Using the posterior probability matrix and the first likelihood probability matrix as input data of a trained HMM model, acquire an outputted second likelihood probability matrix
312 Acquire a corresponding target word sequence corresponding to the speech data to be recognized in a phoneme decoding network according to the second likelihood probability matrix

图 3

(57) **Abstract:** A speech recognition method, comprising: acquiring speech data to be recognized (302); extracting a Filter Bank feature and a Mel-Frequency Cepstral Coefficient (MFCC) feature from the speech data (304); using the MFCC feature as input data of a trained Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) model, acquiring a first likelihood probability matrix outputted by the trained GMM-HMM model (306); using the Filter Bank feature as an input feature of a trained long short-term memory (LSTM) model which has a connection unit, acquiring a posterior probability matrix outputted by the LSTM model which has the connection unit, the

WO 2018/227780 A1

[见续页]

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, 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))。

connection unit being used for controlling the flow of information between layers in the LSTM model (308); using the posterior probability matrix and the first likelihood probability matrix as input data of a trained HMM model, acquiring an outputted second likelihood probability matrix (310); and acquiring a corresponding target word sequence corresponding to the speech data to be recognized in a phoneme decoding network according to the second likelihood probability matrix (312).

(57) 摘要：一种语音识别方法，包括：获取待识别的语音数据（302）；提取语音数据中的Filter Bank特征和MFCC特征（304）；将MFCC特征作为训练后的GMM-HMM模型的输入数据，获取训练后的GMM-HMM模型输出的第一似然概率矩阵（306）；将Filter Bank特征作为训练后的具有连接单元的LSTM模型的输入特征，获取具有连接单元的LSTM模型输出的后验概率矩阵，连接单元用于控制LSTM模型中层与层之间的信息流动（308）；将后验概率矩阵和第一似然概率矩阵作为训练后的HMM模型的输入数据，获取输出的第二似然概率矩阵（310）；根据第二似然概率矩阵在音素解码网络中获取与待识别的语音数据对应的目标词序列（312）。

发明名称：语音识别方法、装置、计算机设备及存储介质

本申请要求于 2017 年 6 月 12 日提交中国专利局、申请号为 2017104450769、发明名称为“语音识别方法、装置、计算机设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机处理领域，特别是涉及一种语音识别方法、装置、计算机设备及存储介质。

背景技术

语音识别，也被称为自动语音识别(Automatic Speech Recognition, ASR)，其目标是让机器通过识别和理解，把语音信号变成文字，是现代人工智能发展的重要分支。语音识别技术的实现是自然语言处理的前提，并能有效推动声控交互相关领域的发展并极大方便人们的生活，如智能家居、语音输入。语音识别的准确度直接决定了技术应用的有效性。

传统的语音识别技术是基于 GMM-HMM（混合高斯模型和隐马尔科夫模型）进行声学模型的建立，近年来，随着深度学习技术的发展，基于 DNN-HMM(深度学习模型和隐马尔科夫模型)进行声学模型的建立相对于 GMM-HMM 在识别准确度上虽然有了很大的提升，但是还有待于进一步提高语音识别的准确度。

发明内容

根据本申请的各种实施例，提供一种语音识别方法、装置、计算机设备及存储介质。

一种语音识别方法，包括:

获取待识别的语音数据;

提取所述语音数据中的 Filter Bank 特征和 MFCC 特征;

将所述 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取所述训练后的 GMM-HMM 模型输出的第一似然概率矩阵;

将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动;

将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵; 及

根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

一种语音识别装置，包括:

获取模块，用于获取待识别的语音数据;

提取模块，用于提取所述语音数据中的 Filter Bank 特征和 MFCC 特征;

第一输出模块，用于将所述 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取所述训练后的 GMM-HMM 模型输出的第一似然概率矩阵;

后验概率矩阵输出模块，用于将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动;

第二输出模块，用于将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵; 及

解码模块，用于根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读

指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：

获取待识别的语音数据；

提取所述语音数据中的 Filter Bank 特征和 MFCC 特征；

将所述 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取所述训练后的 GMM-HMM 模型输出的第一似然概率矩阵；

将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动；

将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵；及

根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：获取待识别的语音数据；

提取所述语音数据中的 Filter Bank 特征和 MFCC 特征；

将所述 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取所述训练后的 GMM-HMM 模型输出的第一似然概率矩阵；

将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动；

将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵；及

根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

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

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为一个实施例中计算机设备的内部结构框图；

图 2 为一个实施例中语音识别的架构图；

图 3 为一个实施例中语音识别方法的流程图；

图 4 为一个实施例中通过具有连接单元 LSTM 模型获取后验概率矩阵的方法流程图；

图 5 为一个实施例中提取语音数据中的 Filter Bank 特征和 MFCC 特征的方法流程图；

图 6 为另一个实施例中通过具有连接单元 LSTM 模型获取后验概率矩阵的方法流程图；

图 7 为一个实施例中 GMM-HMM 模型和具有连接单元 LSTM 模型建立的方法流程图；

图 8 为另一个实施例中语音识别方法的流程图；

图 9 为一个实施例中语音识别装置的结构框图；

图 10 为一个实施例中后验概率矩阵输出模块的结构框图；

图 11 为另一个实施例中语音识别装置的结构框图；

图 12 为又一个实施例中语音识别装置的结构框图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及

实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

如图 1 所示，为一个实施例中计算机设备的内部结构示意图。该计算机设备可以是终端也可以是服务器。参照图 1，该计算机设备包括通过系统总线连接的处理器、非易失性存储介质、内存储器、网络接口、显示屏和输入装置。其中，该计算机设备的非易失性存储介质可存储操作系统和计算机可读指令，该计算机可读指令被执行时，可使得处理器执行一种语音识别方法。该计算机设备的处理器用于提供计算和控制能力，支撑整个计算机设备的运行。该内存储器中可储存有计算机可读指令，该计算机可读指令被处理器执行时，可使得处理器执行一种语音识别方法。计算机设备的网络接口用于进行网络通信。计算机设备的显示屏可以是液晶显示屏或者电子墨水显示屏，计算机设备的输入装置可以是显示屏上覆盖的触摸层，也可以是计算机设备外壳上设置的按键、轨迹球或触控板，还可以是外接的键盘、触控板或鼠标等。触摸层和显示屏构成触控屏。本领域技术人员可以理解，图 1 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

首先，介绍一下语音识别的框架，如图 2 所示，语音识别主要包括两个部分：声学模型和语言模型，然后结合字典就构成了语音识别的框架。语音识别的过程就是根据字典、声学模型和语言模型，将输入的语音特征序列转换为字符序列的过程。其中，声学模型的作用是得到语音特征与音素的映射，语言模型的作用是得到词与词、词与句子的映射，字典的作用是得到字词与音素之间的映射。具体语音识别的过程可以分为三步，第一步，把语音帧识别成音素状态，即进行语音帧和音素状态上的对齐。第二步是根据音素状态组合成音素。第三步是把音素组合成单词。其中，第一步是声学模型的作用，是重点也是难点，语音帧与音素状态的对齐结果越准确，就意味着语音识别的效果就会越好。其中，音素状态是比音素更细致的语音单位，通常一个音

素由 3 个音素状态构成。

如图 3 所示，在一个实施例中，提出了一种语音识别方法，该方法可应用于终端或服务器中，具体包括以下步骤：

步骤 302，获取待识别的语音数据。

在本实施例中，这里待识别的语音数据通常是通过交互应用获取到用户输入的音频数据，包括数字的音频和文字的音频。

步骤 304，提取语音数据中的 Filter Bank 特征和 MFCC 特征。

在本实施例中，Filter Bank(滤波器组)特征和 MFCC（Mel frequency cepstrum coefficient，梅尔倒谱系数）特征都是语音识别中用来表示语音特征的参数。其中，Filter Bank 用于深度学习模型，MFCC 用于混合高斯模型。在提取语音数据中的 FilterBank 特征和 MFCC 特征之前，一般需要对语音数据进行预处理。具体地，首先对输入的语音数据进行预加重处理，通过使用一个高通滤波器提升语音信号中的高频部分，使得频谱更平滑，然后将经过预加重处理的语音数据进行分帧加窗，从而将非平稳的语音信号转变为短时平稳的信号，接着通过端点检测，区分语音与噪声，并提取出有效的语音部分。为了提取语音数据中的 Filter Bank 特征和 MFCC 特征，首先，将经过预处理的语音数据进行快速傅里叶变换，从而将时域的语音信号转换为频域的能量谱进行分析，然后将能量谱通过一组梅尔尺度的三角滤波器组，突出语音的共振峰特征，之后计算每个滤波器组输出的对数能量，该滤波器组输出的特征就是 Filter Bank 特征。进一步的，将计算得到的对数能量经离散余弦变换得到 MFCC 系数，即 MFCC 特征。

步骤 306，将 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取训练后的 GMM-HMM 模型输出的第一似然概率矩阵。

在本实施例中，声学模型和语言模型共同实现对语音的识别。其中，声学模型的作用是用于识别语音帧与音素状态的对齐关系。GMM-HMM 模型属于声学模型的一部分，用于将语音帧与音素状态进行初步对齐。具体地，将提取的待识别的语音数据的 MFCC 特征作为训练后的 GMM-HMM 模型的输

入数据，然后获取该模型输出的似然概率矩阵，为了便于和后续进行区分，这里称为“第一似然概率矩阵”。似然概率矩阵表示的是语音帧与音素状态上的对齐关系，即根据计算得到的似然概率矩阵就可以得到语音帧与音素状态上的对齐关系，只不过，通过 GMM-HMM 训练得到的对齐关系并不十分准确，所以这里通过第一似然概率矩阵相当于对语音帧和音素状态进行了初步对齐。GMM 模型具体的计算公式如下：

$$f(\vec{x}|\vec{\mu},D)=\frac{1}{(2\pi)^{K/2}|D|^{1/2}}\exp(-\frac{1}{2}(x-\mu)^{T}D^{-1}(x-\mu))$$

其中，x 表示提取的语音特征（MFCC）向量，μ， D 分别为均值和方差矩阵，K 表示 MFCC 系数的阶数。

步骤 308，将 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取具有连接单元的 LSTM 模型输出的后验概率矩阵，连接单元用于控制 LSTM 模型中层与层之间的信息流动。

在本实施例中，LSTM 模型属于深度学习模型，也属于声学模型的一部分。具有连接单元的 LSTM 是在传统的 LSTM 模型的基础上提出的创新性的模型，该模型通过在传统的 LSTM 模型的层与层之间增加连接单元，通过该连接单元可以控制层与层之间的信息流动，所以通过该连接单元可以实现对有效信息的筛选，而且通过该连接单元可以使得 LSTM 模型训练的层次更深，层次越多，获得的特征表达越好，识别效果也就越好。所以具有连接单元的 LSTM 模型不但可以提高识别语音的速度，而且可以提高识别语音的准确度。具体地，连接单元是通过 sigmoid 函数来实现的，原理是将前一层的输出通过一个由 sigmoid 函数构成的门限来控制流入到后一层的信息，即输出作为后一层 LSTM 网络的输入。这个 sigmoid 函数的值是由前一层神经元节点的状态、前一层神经元节点的输出、后一层神经元节点的输入共同决定的。其中，神经元节点是负责神经网络模型的计算表达，每个节点包含一些计算关系，可以理解为一种计算公式，可以相同，也可以不同。每一层 LSTM 中的神经元结点的数量是根据输入的特征的帧数以及特征向量来决定的，比如，

如果输入是拼接了前后 5 帧，那么总共有 11 帧输入向量，而每一帧对应的特征向量是由提取的语音特征决定的，比如，如果提取的 Filter Bank 特征为 83 维的特征向量，那么相应的训练得到的 LSTM 模型中每一层的神经元节点为 11x83=913 个。

步骤 310，将后验概率矩阵和第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵。

在本实施例中，HMM（隐马尔科夫）模型是统计模型，它用来描述一个含有隐含未知参数的马尔科夫过程，作用是从可观察的参数中确定该过程中的隐含参数。HMM 模型中主要涉及 5 个参数，分别是 2 个状态集合和 3 个概率集合。其中，2 个状态集合分别为隐藏状态和观察状态，三个概率集合为初始矩阵，转移矩阵和混淆矩阵。其中，转移矩阵是训练得到的，也就是说，一旦 HMM 模型训练完成，该转移矩阵就确定了。在该实施例中，主要是采用可观察的语音特征（Filter Bank 特征）作为观察状态，来计算确定音素状态与语音帧的对应关系（即隐含状态）。如果想要确定音素状态与语音帧的对应关系，还需要确定两个参数，那就是初始矩阵和混淆矩阵。其中，通过具有连接单元的 LSTM 模型计算得到的后验概率矩阵就是 HMM 模型中需要确定的混淆矩阵，第一似然概率矩阵就是需要确定的初始矩阵。所以将后验概率矩阵和第一似然概率矩阵作为训练后的 HMM 模型的输入数据，就可以获取输出的第二似然概率矩阵。该第二似然概率矩阵表示的是音素状态与语音帧的最终的对齐关系。后续根据该确定的第二似然概率矩阵就可以在音素解码网络中获取与待识别的语音数据对应的目标词序列。

步骤 312，根据第二似然概率矩阵在音素解码网络中获取与待识别的语音数据对应的目标词序列。

在本实施例中，在语音识别过程中，包括两个部分，一个是声学模型，一个是语言模型。在语音识别前，首先需要根据训练后的声学模型和语言模型以及字典建一个音素级别的解码网络，根据搜索算法在该网络中寻找最佳的一条路径，其中，搜索算法可以采用维特比算法（Viterbi 算法）。这个路径

就是能够以最大概率输出与待识别语音数据对应的词串，这样就确定了这个语音数据中所包含的文字了。其中，音素解码级别的解码网络（即音素解码网络）是通过有限状态机(Finite State Transducer，FST)相关算法来完成的，如确定化算法 determination、最小化算法 minimization，通过将句子拆分成词、再将词拆分为音素（如汉语的声韵母、英文的音标），然后将音素和发音词典、语法等通过上述方法进行对齐计算，得到输出的音素解码网络。音素解码网络中包含了所有可能识别的路径表达，解码的过程就是根据输入的语音数据，对这个庞大网络进行路径的删减，得到一个或多个候选路径，存储在一种词网络的数据结构中，然后最后的识别就是对候选路径进行打分，分数最高的路径为识别结果。

在本实施例中，通过将混合高斯模型 GMM 和深度学习模型中的长短时递归神经网络 LSTM 结合进行语音识别，先采用 GMM-HMM 模型根据提取的 MFCC 特征计算得到第一似然概率矩阵，第一似然概率矩阵表示对语音数据在音素状态上对齐结果，然后再使用 LSTM 在之前初步对齐结果的基础上进行进一步的对齐，且该 LSTM 采用的是创新性的具有连接单元的 LSTM 模型，该模型通过在传统的 LSTM 模型的层与层之间增加了连接单元，该连接单元可以控制层与层之间信息流动，通过该连接单元可以实现对有效信息的筛选，不但可以提高识别的速度，而且可以提高识别的准确度。

如图 4 所示，在一个实施例中，所述连接单元采用 sigmoid 函数来实现；所述将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动，包括：

步骤 308a，将 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征。

步骤 308b，根据 LSTM 模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的 sigmoid 函数值；

步骤 308c，根据层与层之间的连接单元所对应的 sigmoid 函数值，输出与 Filter Bank 特征对应的后验概率矩阵。

在本实施例中，连接单元是采用 sigmoid 函数来实现的，在 LSTM 模型中通过 sigmoid 函数来控制层与层信息的流动，比如，控制是否流动以及流动多少。其中，sigmoid 函数对应的函数值的确定是由前一层神经元节点的状态、前一层神经元节点的输出和后一层神经元节点的输入来决定的。具体地，sigmoid 函数表示为：$\sigma(x=)+e^{-x}$，$d_t^{l+1}=\sigma(b^{l+1}+W_x^{l+1}X_t^{l+1}+W_c^{l+1}\odot c_t^{l+1}+W_l^{l+1}\odot c_t^l)$，其中，X 表示连接单元在该层的输入，t 表示 t 时刻，d 表示该连接单元的输出，l 表示该连接单元的前一层，l+1 表示该连接单元的下一层，b 表示偏置项，W 表示权重矩阵，其中，W_x 是与输入相关的权重矩阵，W_c 是与输出相关的权重矩阵，W_l 是与层次相关的权重矩阵，c 表示 LSTM 输出控制们的输出，LSTM 有三个门限控制，输入控制门、遗忘控制门、输出控制门，输出控制门的作用是控制该神经元节点的输出流动量。⊙ 是一种运算符，表示两个矩阵对应元素相乘。其中，偏置项 b 和权重矩阵 W 的值在模型完成训练后已经确定了，所以根据输入就可以确定层与层之间的信息流动了多少，在确定了层与层之间的信息流动，就可以获取输出的与 Filter Bank 特征对应的后验概率矩阵。

如图 5 所示，在一个实施例中，提取语音数据中的 Filter Bank 特征和 MFCC 特征的步骤 304 包括:

步骤 304A，将待识别的语音数据进行傅里叶变换转换为频域的能量谱。

在本实施例中，由于语音信号在时域上的变换通常都很难看出信号的特性，所以通常需要将它转换为频域上的能量分布来观察，不同的能量分布，代表不同语音的特性。所以需要将待识别的语音数据经过快速傅里叶变换以得到频谱上的能量分布。其中，是通过将每一帧语音信号进行快速傅里叶变换得到每一帧的频谱，对语音信号的频谱取模平方得到语音信号的功率谱(即能量谱)。

步骤 304B，将频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，

计算得到待识别语音数据的 Filter Bank 特征。

在本实施例中，为了得到待识别语音数据的 Filter Bank 特征，需要将得到的频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算每个三角滤波器组输出的对数能量，即得到待识别语音数据的 Filter Bank 特征。其中，Filter Bank 特征也是通过将每一帧语音信号对应的能量谱作为梅尔尺度的三角滤波器组的输入特征，然后得到每一帧语音信号对应的 Filter Bank 特征。

步骤 304C，将 Filter Bank 特征经过离散余弦变换得到待识别语音数据的 MFCC 特征。

在本实施例中，为了得到待识别语音数据的 MFCC 特征，还需要将经过滤波器组输出的对数能量进行离散余弦变换得到相应的 MFCC 特征。通过将每一帧语音信号对应的 Filter Bank 特征经过离散余弦变换得到每一帧语音信号对应的 MFCC 特征。其中，Filter Bank 特征与 MFCC 特征的区别在于，Filter Bank 特征在不同特征维度之间存在数据相关性，而 MFCC 特征则是采用离散余弦变换去除 Filter Bank 特征的数据相关性所得到的特征。

如图 6 所示，在一个实施例中，将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动的步骤 308 包括：

步骤 308A，获取待识别语音数据中每一帧语音数据对应的 Filter Bank 特征并按照时间排序。

在本实施例中，在提取待识别语音数据中的 Filter Bank 特征时是通过先将语音数据进行分帧处理，然后提取每一帧语音数据对应的 Filter Bank 特征，并按照时间的先后顺序排序，即按照待识别语音数据中每一帧出现的先后顺序将对应的每一帧的 Filter Bank 特征进行排序。

步骤 308B，将每一帧语音数据的 Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的 Filter Bank 特征作为训练后的具有连接单元的 LSTM

模型的输入特征，通过连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率。

在本实施例中，深度学习模型的输入采用的是多帧特征，相对于传统的只有单帧输入的混合高斯模型更有优势，因为通过拼接前后语音帧有利于获取到上下文相关信息对当前的影响。所以一般是将每一帧语音数据和与每一帧语音数据的前后预设帧数的 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征。比如，将当前帧和该当前帧的前后 5 帧进行拼接，共 11 帧数据作为训练后的具有连接单元的 LSTM 模型的输入特征，这 11 帧语音特征序列通过具有连接单元的 LSTM 中的各个结点，输出该帧语音数据对应的音素状态上的后验概率。

步骤 308C，根据每一帧语音数据对应的后验概率确定待识别语音数据对应的后验概率矩阵。

在本实施例中，当获取到每一帧语音数据对应的后验概率后就确定待识别语音数据对应的后验概率矩阵。后验概率矩阵是有一个个后验概率组成的。由于通过具有连接单元 LSTM 模型既可以包含有时间维度的信息，又可以包含有层次纬度的信息，所以相对于之前只有时间维度信息的传统模型，该模型能更好的得到待识别语音数据对应的后验概率矩阵。

如图 7 所示，在一个实施例中，在获取待识别的语音数据的步骤之前还包括：步骤 301，GMM-HMM 模型的建立和具有连接单元 LSTM 模型的建立。具体包括：

步骤 301A，采用训练语料库对 GMM-HMM 进行训练，通过不断的迭代训练确定 GMM-HMM 模型对应的方差和均值，根据方差和均值生成训练后的 GMM-HMM 模型。

在本实施例中，GMM-HMM 声学模型的建立依次采用了单音素训练以及三音素进行训练，其中，三音素训练考虑了当前音素的前后相关音素影响，能够得到更加准确的对齐效果，也就能产生更好的识别结果。根据特征和作

用的不用，三音素训练一般采用基于 delta+delta-delta 特征的三音素训练，线性判别分析+最大似然线性特征转换的三音素训练。具体地，首先对输入的训练预料库中的语音特征进行归一化，默认对方差进行归一化。语音特征归一化是为了消除电话信道等卷积噪声在特征提取计算造成的偏差。然后利用少量特征数据快速得到一个初始化的 GMM-HMM 模型，然后通过不断的迭代训练确定混合高斯模型 GMM-HMM 对应的方差和均值，一旦方差和均值确定，那么相应的 GMM-HMM 的模型就相应的确定了。

步骤 301B，根据训练语料库中提取的 MFCC 特征，采用训练后的 GMM-HMM 模型获取到训练语料库对应的似然概率矩阵。

在本实施例中，采用训练预料库中的语音数据进行训练，提取训练语料库中语音的 MFCC 特征，然后作为上述训练后的 GMM-HMM 模型的输入特征，获取到输出的训练语料库中语音对应的似然概率矩阵。似然概率矩阵代表的是语音帧与音素状态上的对齐关系，通过训练后的 GMM-HMM 输出似然概率矩阵目的是将其作为后续训练深度学习模型的初始对齐关系，便于后续深度学习模型能够得到更好的深度学习的结果。

步骤 301C，根据训练预料库中提取的 Filter Bank 特征和似然概率矩阵对具有连接单元的 LSTM 模型进行训练，确定与具有连接单元 LSTM 模型对应的权重矩阵和偏置矩阵，根据权重矩阵和偏置矩阵生成训练后的具有连接单元的 LSTM 模型。

在本实施例中，将上述通过 GMM-HMM 计算得到的对齐结果（即似然概率矩阵）和原始语音特征一起作为具有连接单元 LSTM 模型的输入特征进行训练，其中，这里的原始语音特征采用的 Filter Bank 特征，相对于 MFCC 特征，Filter Bank 特征具有数据相关性，所以具有更好的语音特征表达。通过对具有连接单元 LSTM 模型进行训练，确定每一层 LSTM 对应的权重矩阵和偏置矩阵。具体地，具有连接单元 LSTM 也属于深度神经网络模型中的一种，神经网络层一般分为三类：输入层、隐藏层和输出层，其中，隐含层有多层。训练具有连接单元 LSTM 模型的目的就是为了确定每一层中所有的权

重矩阵和偏置矩阵以及相应的层数，训练的算法可以采用前向传播算法、维特比算法等现有的算法，这里不对具体的训练算法进行限定。

如图 8 所示，在一个实施例中，提出了一种语音识别方法，该方法包括以下步骤：

步骤 802，获取待识别的语音数据。

步骤 804，提取语音数据中的 Filter Bank 特征和 MFCC 特征。

步骤 806，将 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取训练后的 GMM-HMM 模型输出的第一似然概率矩阵。

步骤 808，将 Filter Bank 特征和第一似然概率矩阵作为训练后的 DNN-HMM 模型的输入数据，获取训练后的 DNN-HMM 模型输出的第二似然概率矩阵。

步骤 810，将 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取具有连接单元的 LSTM 模型输出的后验概率矩阵，连接单元用于控制 LSTM 模型中层与层之间的信息流动。

步骤 812，将后验概率矩阵和第二似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第三似然概率矩阵。

步骤 814，根据第三似然概率矩阵在音素解码网络中获取与待识别的语音数据对应的目标词序列。

在本实施例中，为了能得到更准确的识别效果，在通过训练后的 GMM-HMM 模型得到初步对齐结果（第一似然概率矩阵），再经过训练后的 DNN-HMM 进行进一步的对齐，能够获取更好的对齐效果。由于深度神经网络模型比传统的混合高斯模型能得到更好的语音特征表达，因此使用深度神经网络模型做进一步强制对齐能进一步提高准确率。然后将该进一步对齐的结果（第二似然概率矩阵）代入具有创新型的具有连接单元 LSTM-HMM 模型，可以获取到最后的对齐结果（第三似然概率矩阵）。需要说明的是，这里的对齐结果是指语音帧与音素状态的对齐关系。上述不管是混合高斯模型还是深度学习模型等都是声学模型的一部分，而声学模型的作用就是获取语音

帧与音素状态的对齐关系，便于后续结合语言模型在音素解码网络中获取与待识别语音数据对应的目标词序列。

如图 9 所示，在一个实施例中，提出了一种语音识别装置，该装置包括：

获取模块 902，用于获取待识别的语音数据。

提取模块 904，用于提取语音数据中的 Filter Bank 特征和 MFCC 特征。

第一输出模块 906，用于将 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取训练后的 GMM-HMM 模型输出的第一似然概率矩阵。

后验概率矩阵输出模块 908，将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动。

第二输出模块 910，用于将后验概率矩阵和第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵。

解码模块 912，用于根据第二似然概率矩阵在音素解码网络中获取与待识别的语音数据对应的目标词序列。

在一个实施例中，提取模块还用于将待识别的语音数据进行傅里叶变换转换为频域的能量谱，将频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算得到待识别语音数据的 Filter Bank 特征，将 Filter Bank 特征经过离散余弦变换得到待识别语音数据的 MFCC 特征。

在一个实施例中，连接单元采用 sigmoid 函数来实现；所述后验概率矩阵输出模块 908 还用于将所述 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征；根据所述 LSTM 模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的 sigmoid 函数值；根据所述层与层之间的连接单元所对应的 sigmoid 函数值，输出与所述 Filter Bank 特征对应的后验概率矩阵。

如图 10 所示，在一个实施例中，后验概率矩阵输出模块 908 包括：

排序模块 908A，用于获取待识别语音数据中每一帧语音数据对应的 Filter Bank 特征并按照时间排序。

后验概率输出模块 908B，用于将每一帧语音数据的 Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征，通过所述连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率。

确定模块 908C，用于根据每一帧语音数据对应的后验概率确定待识别语音数据对应的后验概率矩阵。

如图 11 所示，在一个实施例中，上述语音识别装置还包括：

GMM-HMM 模型训练模块 914，用于采用训练语料库对 GMM-HMM 模型进行训练，通过不断的迭代训练确定 GMM-HMM 模型对应的方差和均值，根据方差和均值生成训练后的 GMM-HMM 模型。

似然概率矩阵获取模块 916，用于根据训练语料库中提取的 MFCC 特征，采用训练后的 GMM-HMM 模型获取到训练语料库对应的似然概率矩阵。

LSTM 模型训练模块 918，用于根据训练预料库中提取的 Filter Bank 特征和似然概率矩阵对具有连接单元 LSTM 模型进行训练，确定与具有连接单元 LSTM 模型对应的权重矩阵和偏置矩阵，根据权重矩阵和偏置矩阵生成训练后的具有连接单元 LSTM 模型。

如图 12 所示，在一个实施例中，提出了一种语音识别装置，该装置包括：

获取模块 1202，用于获取待识别的语音数据。

提取模块 1204，用于提取语音数据中的 Filter Bank 特征和 MFCC 特征。

第一输出模块 1206，用于将 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取训练后的 GMM-HMM 模型输出的第一似然概率矩阵。

第二输出模块 1208，将 Filter Bank 特征和第一似然概率矩阵作为训练后的 DNN-HMM 模型的输入数据，获取训练后 DNN-HMM 输出的第二似然概率矩阵。

后验概率矩阵输出模块 1210，用于将所述 Filter Bank 特征作为训练后的

具有连接单元的LSTM模型的输入特征，获取所述具有连接单元的LSTM模型输出的后验概率矩阵，所述连接单元用于控制所述LSTM模型中层与层之间的信息流动。

第三输出模块1212，用于将后验概率矩阵和第二似然概率矩阵作为训练后的HMM模型的输入数据，获取输出的第三似然概率矩阵。

解码模块1214，用于根据第三似然概率矩阵在音素解码网络中获取与待识别的语音数据对应的目标词序列。

上述语言识别装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。其中，网络接口可以是以太网卡或无线网卡等。上述各模块可以硬件形式内嵌于或独立于服务器中的处理器中，也可以以软件形式存储于服务器中的存储器中，以便于处理器调用执行以上各个模块对应的操作。该处理器可以为中央处理单元（CPU)、微处理器、单片机等。

上述语言识别装置可以实现为一种计算机程序的形式，计算机程序可以在如图1所示的计算机设备上运行。

在一个实施例中，提出一种计算机设备，计算机设备的内部结构可对应于如图1所示的结构，即该计算机设备既可以是服务器也可以是终端，其包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现以下步骤：获取待识别的语音数据，提取所述语音数据中的Filter Bank特征和MFCC特征，将所述MFCC特征作为训练后的GMM-HMM模型的输入数据，获取所述训练后的GMM-HMM模型输出的第一似然概率矩阵，将所述Filter Bank特征作为训练后的具有连接单元的LSTM模型的输入特征，获取所述具有连接单元的LSTM模型输出的后验概率矩阵，所述连接单元用于控制所述LSTM模型中层与层之间的信息流动，将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的HMM模型的输入数据，获取输出的第二似然概率矩阵，根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目

标词序列。

在一个实施例中，所述连接单元采用 sigmoid 函数来实现；所述处理器所执行的所述将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动，包括：将所述 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征；根据所述 LSTM 模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的 sigmoid 函数值，根据所述层与层之间的连接单元所对应的 sigmoid 函数值，输出与所述 Filter Bank 特征对应的后验概率矩阵。

在一个实施例中，所述处理器所执行的所述提取所述语音数据中的 Filter Bank 特征和 MFCC 特征，包括：将所述待识别的语音数据进行傅里叶变换转换为频域的能量谱；将所述频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算得到待识别语音数据的 Filter Bank 特征；将所述 Filter Bank 特征经过离散余弦变换得到待识别语音数据的 MFCC 特征。

在一个实施例中，所述处理器所执行的将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动的步骤包括：获取待识别语音数据中每一帧语音数据对应的 Filter Bank 特征并按照时间排序；将每一帧语音数据的 Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征，通过所述连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率；根据所述每一帧语音数据对应的后验概率确定所述待识别语音数据对应的后验概率矩阵。

在一个实施例中，在所述获取待识别的语音数据的步骤之前，所述处理器执行所述计算机程序是还用于实现以下步骤：采用训练语料库对

GMM-HMM 模型进行训练，通过不断的迭代训练确定所述 GMM-HMM 模型对应的方差和均值；根据所述方差和均值生成训练后的 GMM-HMM 模型；根据所述训练语料库中提取的 MFCC 特征，采用训练后的 GMM-HMM 模型获取到所述训练语料库对应的似然概率矩阵；根据所述训练预料库中提取的 Filter Bank 特征和所述似然概率矩阵对所述具有连接单元的 LSTM 模型进行训练，确定与所述具有连接单元的 LSTM 模型对应的权重矩阵和偏置矩阵；根据所述权重矩阵和偏置矩阵生成训练后的具有连接单元的 LSTM 模型。

在一个实施例中，提出了一种计算机可读存储介质，其上存储有计算机指令，该指令被处理器执行时实现以下步骤：获取待识别的语音数据；提取所述语音数据中的 Filter Bank 特征和 MFCC 特征；将所述 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取所述训练后的 GMM-HMM 模型输出的第一似然概率矩阵；将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动；将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵；根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

在一个实施例中，所述连接单元采用 sigmoid 函数来实现；所述处理器所执行的所述将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动，包括：将所述 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征；根据所述 LSTM 模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的 sigmoid 函数值；根据所述层与层之间的连接单元所对应的 sigmoid 函数值，输出与所述 Filter Bank 特征对应的后验概率矩阵。

在一个实施例中，所述处理器所执行的所述提取所述语音数据中的 Filter Bank 特征和 MFCC 特征，包括：将所述待识别的语音数据进行傅里叶变换转换为频域的能量谱；将所述频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算得到待识别语音数据的 Filter Bank 特征；将所述 Filter Bank 特征经过离散余弦变换得到待识别语音数据的 MFCC 特征。

在一个实施例中，所述处理器所执行的将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动的步骤包括：获取待识别语音数据中每一帧语音数据对应的 Filter Bank 特征并按照时间排序；将每一帧语音数据的 Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征，通过所述连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率；根据所述每一帧语音数据对应的后验概率确定所述待识别语音数据对应的后验概率矩阵。

在一个实施例中，在所述获取待识别的语音数据的步骤之前，所述处理器执行所述计算机程序是还用于实现以下步骤：采用训练语料库对 GMM-HMM 模型进行训练，通过不断的迭代训练确定所述 GMM-HMM 模型对应的方差和均值；根据所述方差和均值生成训练后的 GMM-HMM 模型；根据所述训练语料库中提取的 MFCC 特征，采用训练后的 GMM-HMM 模型获取到所述训练语料库对应的似然概率矩阵；根据所述训练预料库中提取的 Filter Bank 特征和所述似然概率矩阵对所述具有连接单元的 LSTM 模型进行训练，确定与所述具有连接单元的 LSTM 模型对应的权重矩阵和偏置矩阵；根据所述权重矩阵和偏置矩阵生成训练后的具有连接单元的 LSTM 模型。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，该计算机程序可存储

于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，前述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory，ROM）等非易失性存储介质等。

以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

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

权利要求书

1、一种语音识别方法，包括:

获取待识别的语音数据;

提取所述语音数据中的 Filter Bank 特征和 MFCC 特征;

将所述 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取所述训练后的 GMM-HMM 模型输出的第一似然概率矩阵;

将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动;

将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵；及

根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

2、根据权利要求 1 所述的方法，其特征在于，所述连接单元采用 sigmoid 函数来实现;所述将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动，包括:

将所述 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征;

根据所述 LSTM 模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的 sigmoid 函数值;

根据所述层与层之间的连接单元所对应的 sigmoid 函数值，输出与所述 Filter Bank 特征对应的后验概率矩阵。

3、根据权利要求 1 所述的方法，其特征在于，所述提取所述语音数据中的 Filter Bank 特征和 MFCC 特征的步骤包括:

将所述待识别的语音数据进行傅里叶变换转换为频域的能量谱;

将所述频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算得

到待识别语音数据的 Filter Bank 特征;

将所述 Filter Bank 特征经过离散余弦变换得到待识别语音数据的 MFCC 特征。

4、根据权利要求 1 所述的方法，其特征在于，将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动的步骤包括:

获取待识别语音数据中每一帧语音数据对应的 Filter Bank 特征并按照时间排序；将每一帧语音数据的 Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征，通过所述连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率；根据所述每一帧语音数据对应的后验概率确定所述待识别语音数据对应的后验概率矩阵。

5、根据权利要求 1 所述的方法，其特征在于，在所述获取待识别的语音数据的步骤之前还包括:

采用训练语料库对 GMM-HMM 模型进行训练，通过不断的迭代训练确定所述 GMM-HMM 模型对应的方差和均值;

根据所述方差和均值生成训练后的 GMM-HMM 模型;

根据所述训练语料库中提取的 MFCC 特征，采用训练后的 GMM-HMM 模型获取到所述训练语料库对应的似然概率矩阵;

根据所述训练预料库中提取的 Filter Bank 特征和所述似然概率矩阵对所述具有连接单元的 LSTM 模型进行训练，确定与所述具有连接单元的 LSTM 模型对应的权重矩阵和偏置矩阵;

根据所述权重矩阵和偏置矩阵生成训练后的具有连接单元的 LSTM 模型。

6、一种语音识别装置，包括:

获取模块，用于获取待识别的语音数据；

提取模块，用于提取所述语音数据中的 Filter Bank 特征和 MFCC 特征；

第一输出模块，用于将所述 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取所述训练后的 GMM-HMM 模型输出的第一似然概率矩阵；

后验概率矩阵输出模块，用于将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动；

第二输出模块，用于将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵；及

解码模块，用于根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

7、根据权利要求 6 所述的装置，其特征在于，所述连接单元采用 sigmoid 函数来实现；所述后验概率矩阵输出模块还用于将所述 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征；根据所述 LSTM 模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的 sigmoid 函数值；根据所述层与层之间的连接单元所对应的 sigmoid 函数值，输出与所述 Filter Bank 特征对应的后验概率矩阵。

8、根据权利要求 6 所述的装置，其特征在于，所述提取模块包括：

转换模块，用于将所述待识别的语音数据进行傅里叶变换转换为频域的能量谱；

计算模块，用于将所述频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算得到待识别语音数据的 Filter Bank 特征；

变换模块，用于将所述 Filter Bank 特征经过离散余弦变换得到待识别语音数据的 MFCC 特征。

9、根据权利要求 6 所述的装置，其特征在于，所述后验概率矩阵输出模块包括：

排序模块，用于获取待识别语音数据中每一帧语音数据对应的Filter Bank特征并按照时间排序；

后验概率输出模块，用于将每一帧语音数据的Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的Filter Bank特征作为所述训练后的具有连接单元的LSTM模型的输入特征，通过所述连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率；

确定模块，用于根据所述每一帧语音数据对应的后验概率确定所述待识别语音数据对应的后验概率矩阵。

10、根据权利要求6所述的装置，其特征在于，所述装置还包括：

GMM-HMM模型训练模块，用于采用训练语料库对GMM-HMM进行训练，通过不断的迭代训练确定所述 GMM-HMM 模型对应的方差和均值，根据所述方差和均值生成训练后的GMM-HMM模型；

似然概率矩阵获取模块，用于根据所述训练语料库中提取的MFCC特征，采用训练后的GMM-HMM模型获取到所述训练语料库对应的似然概率矩阵；

LSTM模型训练模块，用于根据所述训练预料库中提取的Filter Bank特征和所述似然概率矩阵对所述具有连接单元的LSTM模型进行训练，确定与所述具有连接单元的LSTM模型对应的权重矩阵和偏置矩阵，根据所述权重矩阵和偏置矩阵生成训练后的具有连接单元的LSTM模型。

11、一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：

获取待识别的语音数据；

提取所述语音数据中的Filter Bank特征和MFCC特征；

将所述MFCC特征作为训练后的GMM-HMM模型的输入数据，获取所述训练后的GMM-HMM模型输出的第一似然概率矩阵；

将所述Filter Bank特征作为训练后的具有连接单元的LSTM模型的输入特征，获取所述具有连接单元的LSTM模型输出的后验概率矩阵，所述连接

单元用于控制所述 LSTM 模型中层与层之间的信息流动;

将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵；及

根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

12、根据权利要求 11 所述的计算机设备，其特征在于，所述连接单元采用 sigmoid 函数来实现;

所述处理器所执行的所述将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动，包括: 将所述 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征; 根据所述 LSTM 模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的 sigmoid 函数值，根据所述层与层之间的连接单元所对应的 sigmoid 函数值，输出与所述 Filter Bank 特征对应的后验概率矩阵。

13、根据权利要求 11 所述的计算机设备，其特征在于，所述处理器所执行的所述提取所述语音数据中的 Filter Bank 特征和 MFCC 特征，包括:

将所述待识别的语音数据进行傅里叶变换转换为频域的能量谱;

将所述频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算得到待识别语音数据的 Filter Bank 特征;

将所述 Filter Bank 特征经过离散余弦变换得到待识别语音数据的 MFCC 特征。

14、根据权利要求 11 所述的计算机设备，其特征在于，所述处理器所执行的将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动的步骤包括:

获取待识别语音数据中每一帧语音数据对应的 Filter Bank 特征并按照时间排序；

将每一帧语音数据的 Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征，通过所述连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率；

根据所述每一帧语音数据对应的后验概率确定所述待识别语音数据对应的后验概率矩阵。

15、根据权利要求 11 所述的计算机设备，其特征在于，在所述获取待识别的语音数据的步骤之前，所述处理器执行所述计算机可读指令时还用于实现以下步骤：

采用训练语料库对 GMM-HMM 模型进行训练，通过不断的迭代训练确定所述 GMM-HMM 模型对应的方差和均值；

根据所述方差和均值生成训练后的 GMM-HMM 模型；根据所述训练语料库中提取的 MFCC 特征，采用训练后的 GMM-HMM 模型获取到所述训练语料库对应的似然概率矩阵；

根据所述训练预料库中提取的 Filter Bank 特征和所述似然概率矩阵对所述具有连接单元的 LSTM 模型进行训练，确定与所述具有连接单元的 LSTM 模型对应的权重矩阵和偏置矩阵；

根据所述权重矩阵和偏置矩阵生成训练后的具有连接单元的 LSTM 模型。

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

获取待识别的语音数据；

提取所述语音数据中的 Filter Bank 特征和 MFCC 特征；

将所述 MFCC 特征作为训练后的 GMM-HMM 模型的输入数据，获取所

述训练后的 GMM-HMM 模型输出的第一似然概率矩阵;

将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动;

将所述后验概率矩阵和所述第一似然概率矩阵作为训练后的 HMM 模型的输入数据，获取输出的第二似然概率矩阵；及

根据所述第二似然概率矩阵在音素解码网络中获取与所述待识别的语音数据对应的目标词序列。

17、根据权利要求 16 所述的存储介质，其特征在于，所述连接单元采用 sigmoid 函数来实现;

所述处理器所执行的所述将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动，包括:

将所述 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征;

根据所述 LSTM 模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的 sigmoid 函数值，根据所述层与层之间的连接单元所对应的 sigmoid 函数值，输出与所述 Filter Bank 特征对应的后验概率矩阵。

18、根据权利要求 16 所述的存储介质，其特征在于，所述处理器所执行的所述提取所述语音数据中的 Filter Bank 特征和 MFCC 特征，包括:

将所述待识别的语音数据进行傅里叶变换转换为频域的能量谱;

将所述频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算得到待识别语音数据的 Filter Bank 特征;

将所述 Filter Bank 特征经过离散余弦变换得到待识别语音数据的 MFCC

特征。

19、根据权利要求 16 所述的存储介质，其特征在于，所述处理器所执行的将所述 Filter Bank 特征作为训练后的具有连接单元的 LSTM 模型的输入特征，获取所述具有连接单元的 LSTM 模型输出的后验概率矩阵，所述连接单元用于控制所述 LSTM 模型中层与层之间的信息流动的步骤包括：

获取待识别语音数据中每一帧语音数据对应的 Filter Bank 特征并按照时间排序；

将每一帧语音数据的 Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的 Filter Bank 特征作为所述训练后的具有连接单元的 LSTM 模型的输入特征，通过所述连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率；

根据所述每一帧语音数据对应的后验概率确定所述待识别语音数据对应的后验概率矩阵。

20、根据权利要求 16 所述的存储介质，其特征在于，在所述获取待识别的语音数据的步骤之前，所述处理器执行所述计算机可读指令时还用于实现以下步骤：

采用训练语料库对 GMM-HMM 模型进行训练，通过不断的迭代训练确定所述 GMM-HMM 模型对应的方差和均值；

根据所述方差和均值生成训练后的 GMM-HMM 模型；

根据所述训练语料库中提取的 MFCC 特征，采用训练后的 GMM-HMM 模型获取到所述训练语料库对应的似然概率矩阵；

根据所述训练预料库中提取的 Filter Bank 特征和所述似然概率矩阵对所述具有连接单元的 LSTM 模型进行训练，确定与所述具有连接单元的 LSTM 模型对应的权重矩阵和偏置矩阵；

根据所述权重矩阵和偏置矩阵生成训练后的具有连接单元的 LSTM 模型。

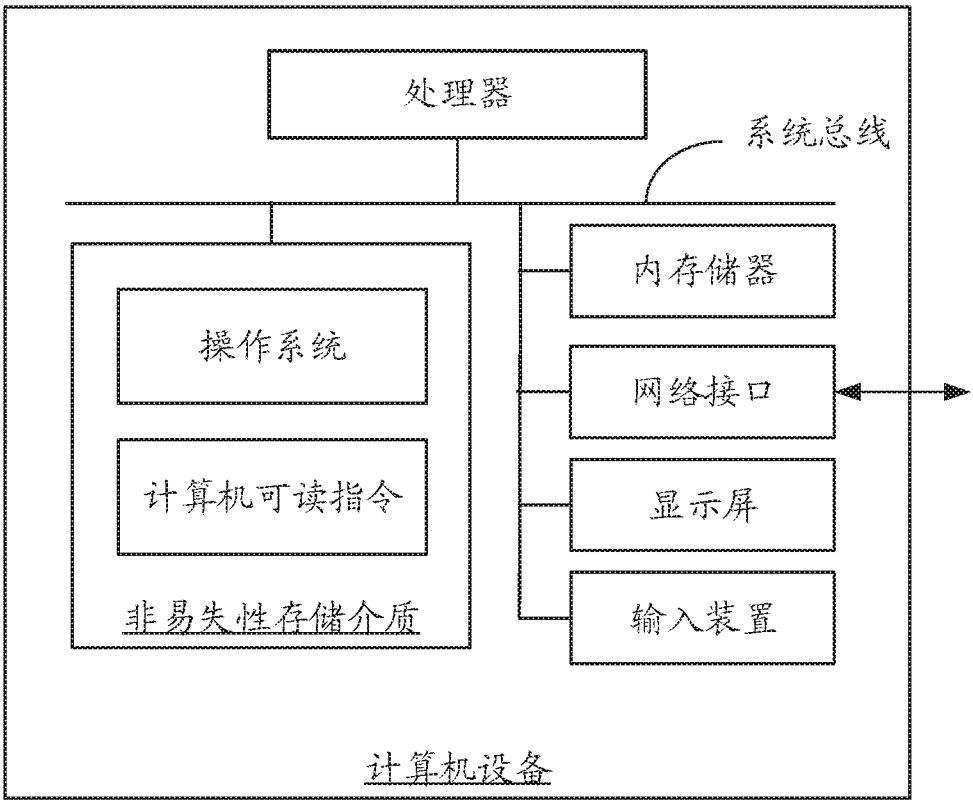

处理器
系统总线
内存储器
操作系统
网络接口
计算机可读指令
显示屏
非易失性存储介质
输入装置
计算机设备

图 1

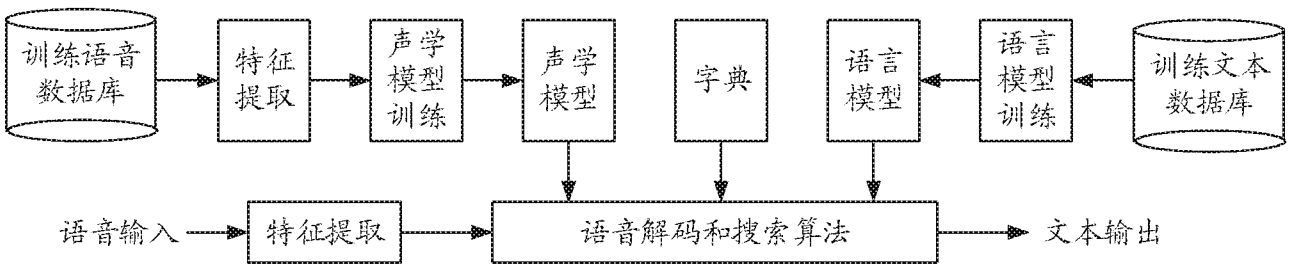

训练语音数据库
特征提取
声学模型训练
声学模型
字典
语言模型
语言模型训练
训练文本数据库
语音输入
特征提取
语音解码和搜索算法
文本输出

图 2

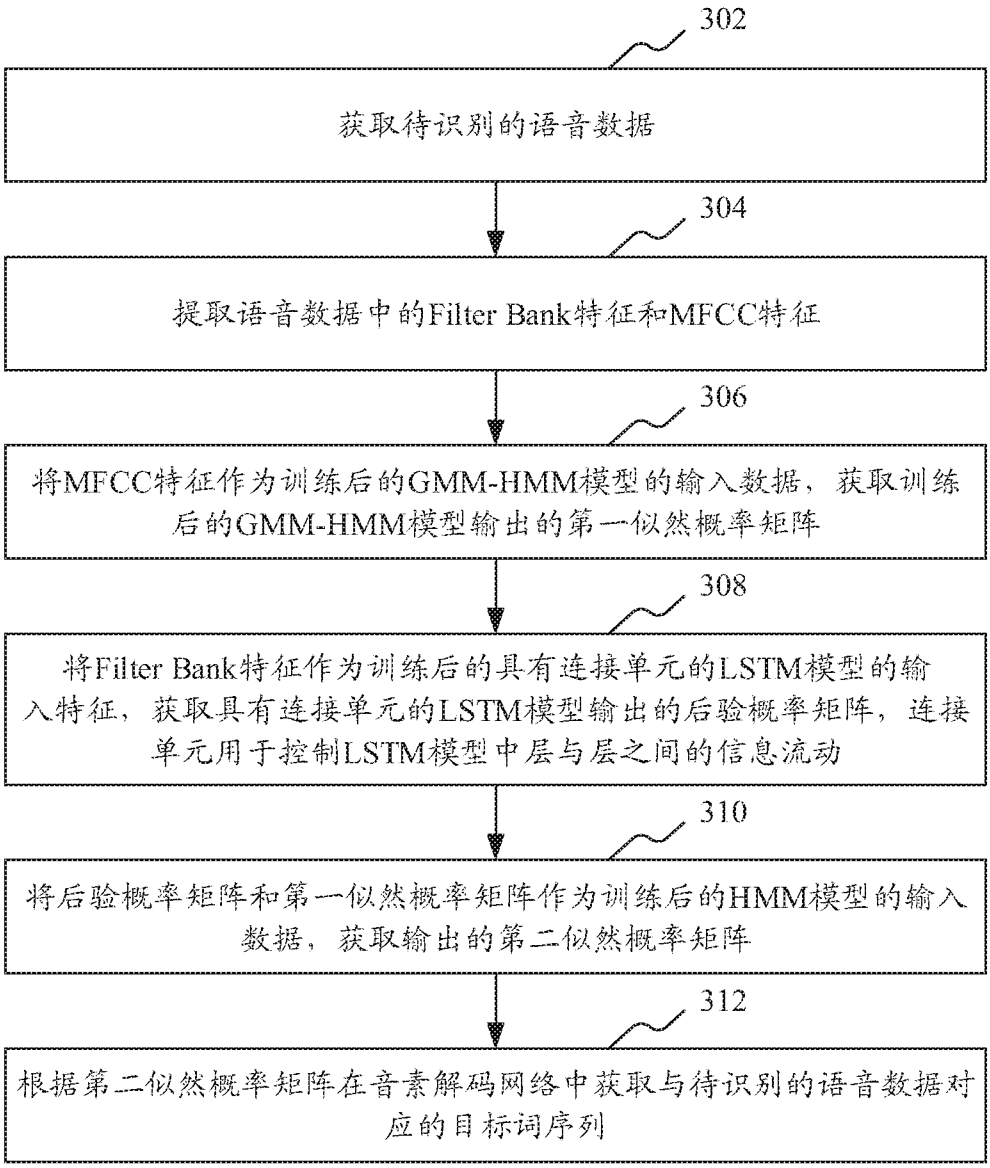

302
获取待识别的语音数据
304
提取语音数据中的Filter Bank特征和MFCC特征
306
将MFCC特征作为训练后的GMM-HMM模型的输入数据，获取训练后的GMM-HMM模型输出的第一似然概率矩阵
308
将Filter Bank特征作为训练后的具有连接单元的LSTM模型的输入特征，获取具有连接单元的LSTM模型输出的后验概率矩阵，连接单元用于控制LSTM模型中层与层之间的信息流动
310
将后验概率矩阵和第一似然概率矩阵作为训练后的HMM模型的输入数据，获取输出的第二似然概率矩阵
312
根据第二似然概率矩阵在音素解码网络中获取与待识别的语音数据对应的目标词序列

图 3

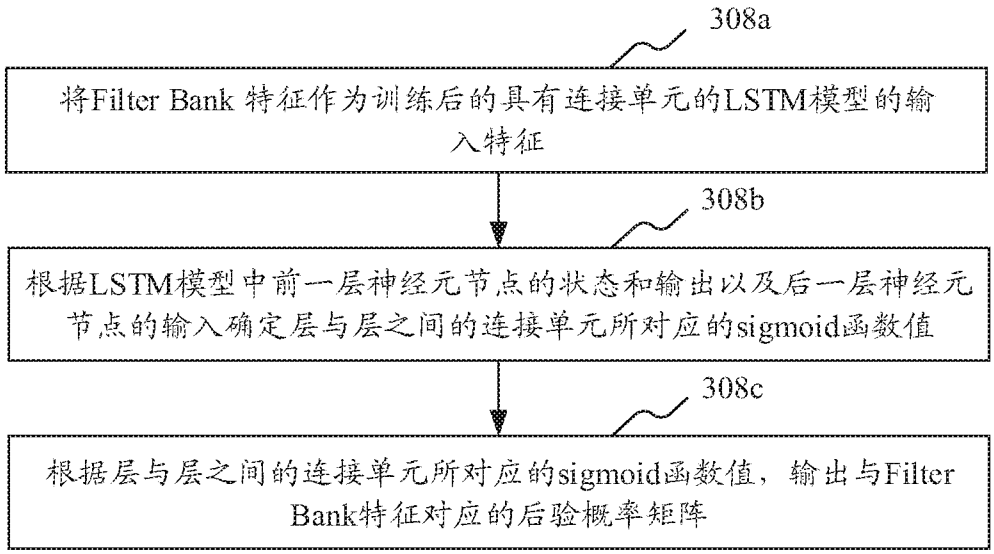

308a
将Filter Bank 特征作为训练后的具有连接单元的LSTM模型的输入特征
308b
根据LSTM模型中前一层神经元节点的状态和输出以及后一层神经元节点的输入确定层与层之间的连接单元所对应的sigmoid函数值
308c
根据层与层之间的连接单元所对应的sigmoid函数值，输出与Filter Bank特征对应的后验概率矩阵

图 4

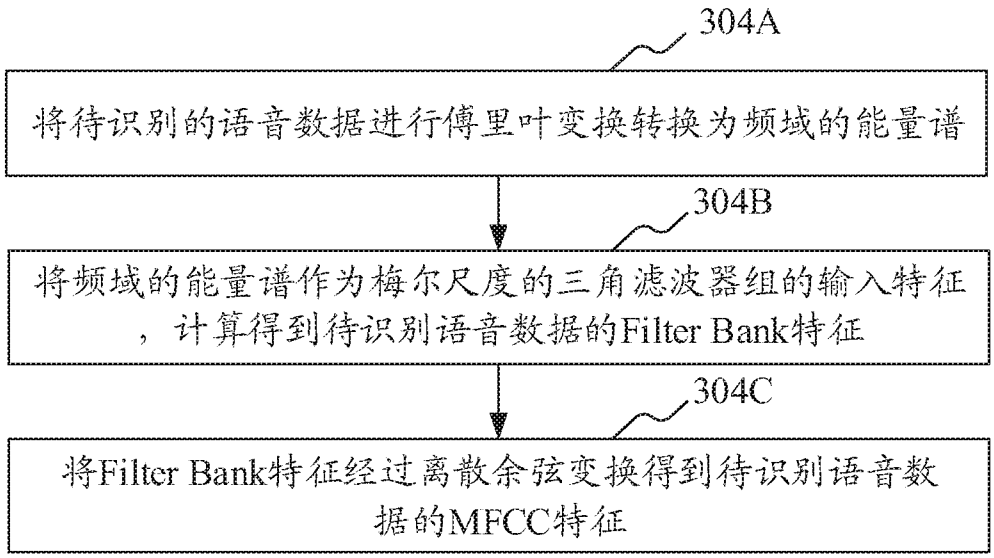

304A
将待识别的语音数据进行傅里叶变换转换为频域的能量谱
304B
将频域的能量谱作为梅尔尺度的三角滤波器组的输入特征，计算得到待识别语音数据的Filter Bank特征
304C
将Filter Bank特征经过离散余弦变换得到待识别语音数据的MFCC特征

图 5

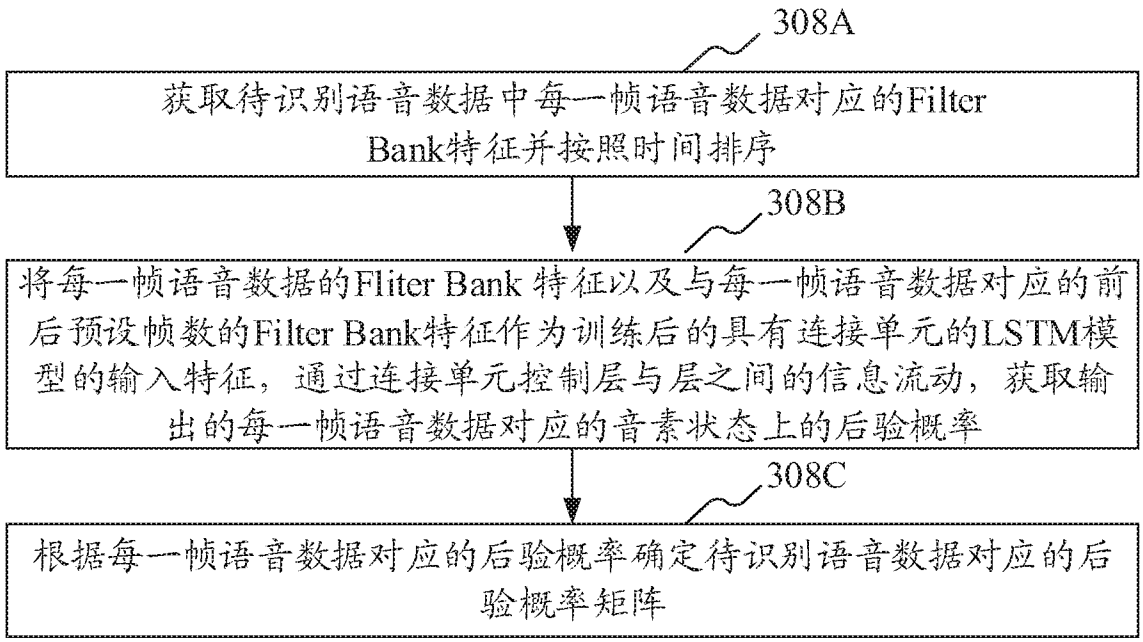

308A
获取待识别语音数据中每一帧语音数据对应的Filter Bank特征并按照时间排序
308B
将每一帧语音数据的Fliter Bank 特征以及与每一帧语音数据对应的前后预设帧数的Filter Bank特征作为训练后的具有连接单元的LSTM模型的输入特征，通过连接单元控制层与层之间的信息流动，获取输出的每一帧语音数据对应的音素状态上的后验概率
308C
根据每一帧语音数据对应的后验概率确定待识别语音数据对应的后验概率矩阵

图 6

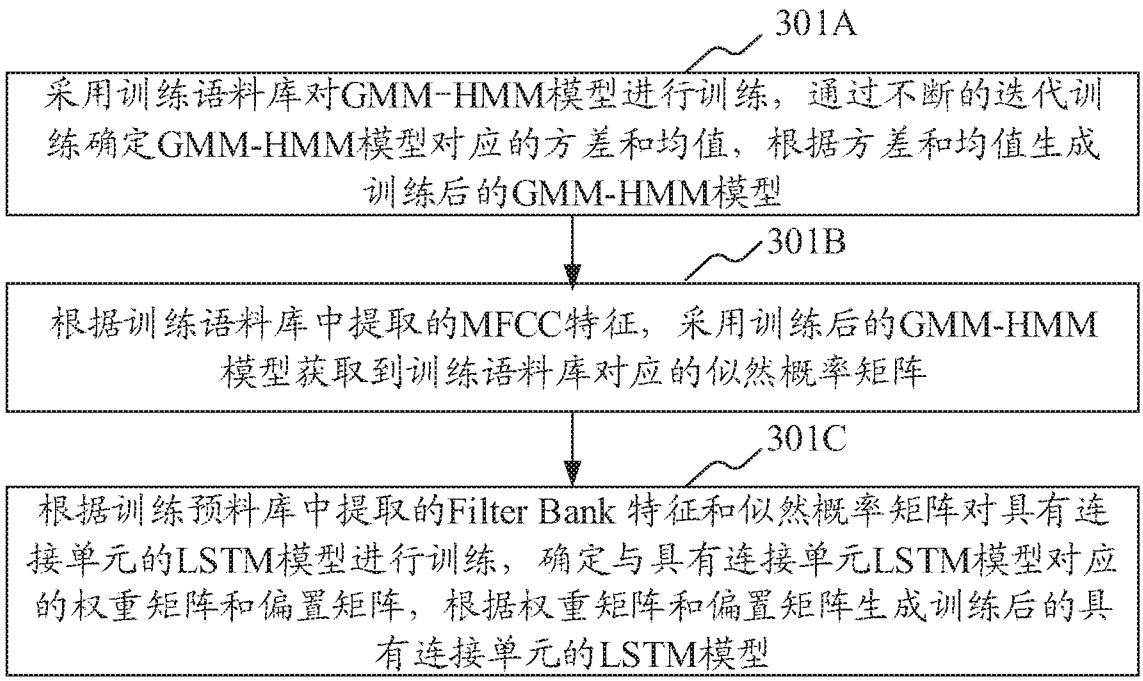

301A
采用训练语料库对GMM-HMM模型进行训练，通过不断的迭代训练确定GMM-HMM模型对应的方差和均值，根据方差和均值生成训练后的GMM-HMM模型
301B
根据训练语料库中提取的MFCC特征，采用训练后的GMM-HMM模型获取到训练语料库对应的似然概率矩阵
301C
根据训练预料库中提取的Filter Bank 特征和似然概率矩阵对具有连接单元的LSTM模型进行训练，确定与具有连接单元LSTM模型对应的权重矩阵和偏置矩阵，根据权重矩阵和偏置矩阵生成训练后的具有连接单元的LSTM模型

图 7

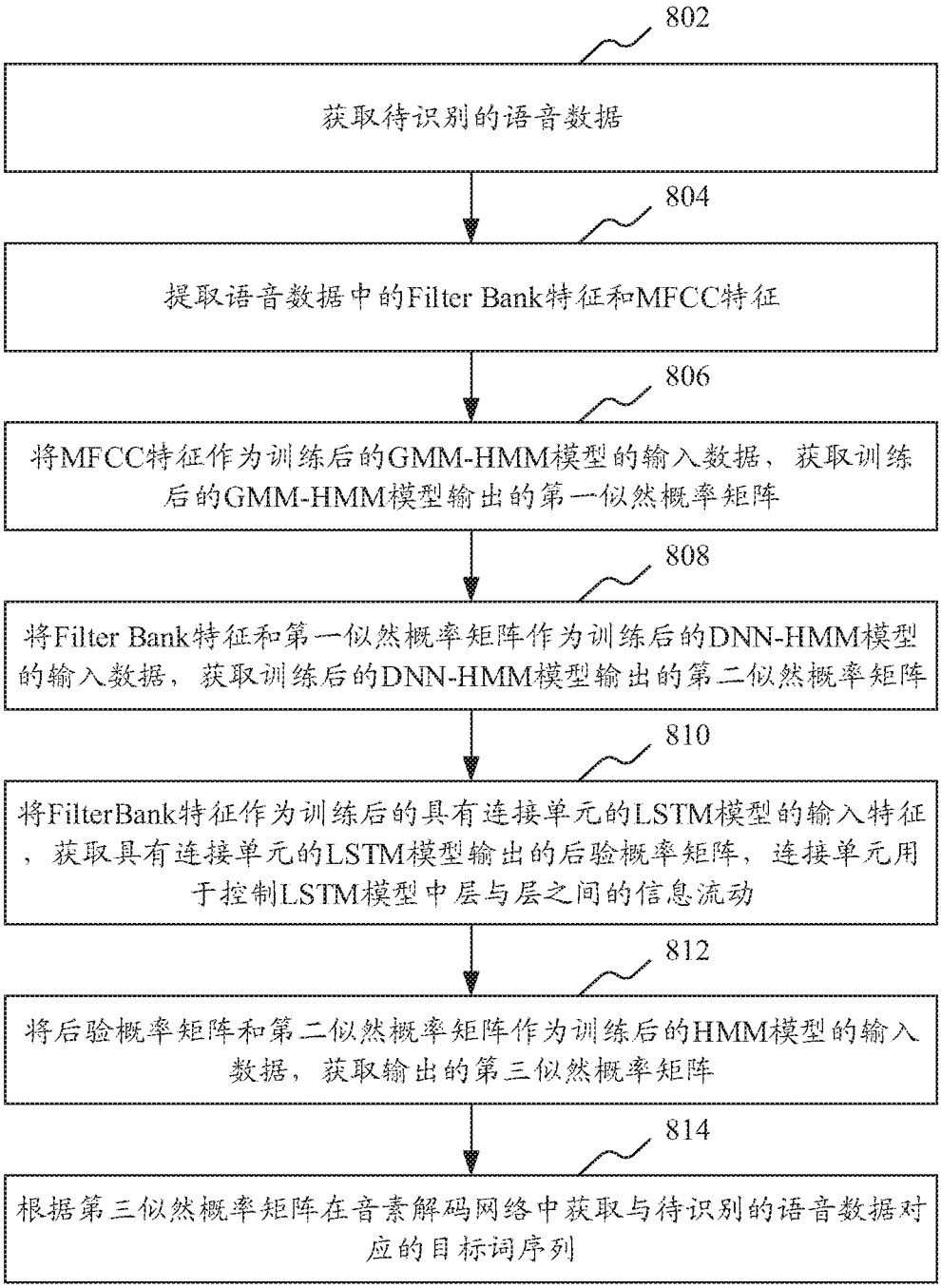

802
获取待识别的语音数据
804
提取语音数据中的Filter Bank特征和MFCC特征
806
将MFCC特征作为训练后的GMM-HMM模型的输入数据，获取训练后的GMM-HMM模型输出的第一似然概率矩阵
808
将Filter Bank特征和第一似然概率矩阵作为训练后的DNN-HMM模型的输入数据，获取训练后的DNN-HMM模型输出的第二似然概率矩阵
810
将FilterBank特征作为训练后的具有连接单元的LSTM模型的输入特征，获取具有连接单元的LSTM模型输出的后验概率矩阵，连接单元用于控制LSTM模型中层与层之间的信息流动
812
将后验概率矩阵和第二似然概率矩阵作为训练后的HMM模型的输入数据，获取输出的第三似然概率矩阵
814
根据第三似然概率矩阵在音素解码网络中获取与待识别的语音数据对应的目标词序列

图 8

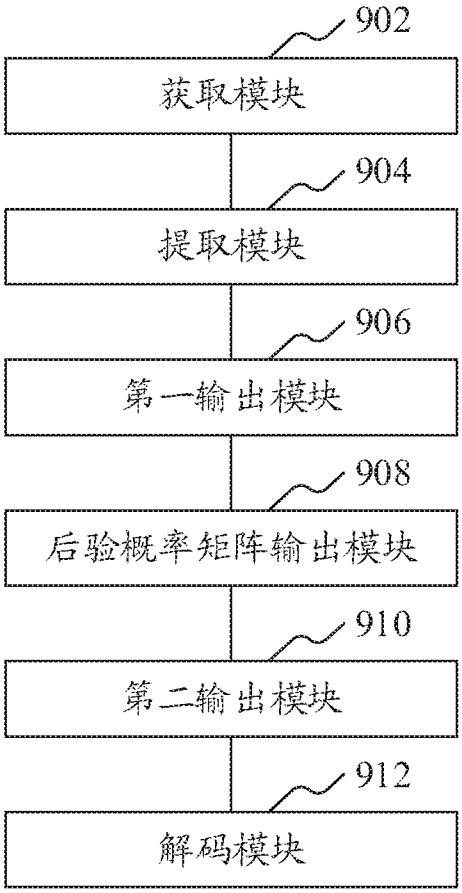

902
获取模块
904
提取模块
906
第一输出模块
908
后验概率矩阵输出模块
910
第二输出模块
912
解码模块

图 9

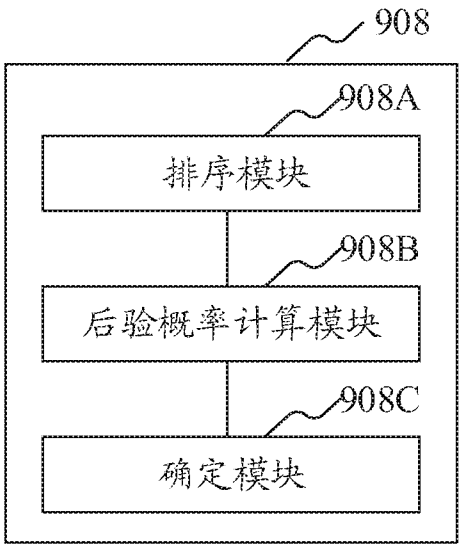

908
908A
排序模块
908B
后验概率计算模块
908C
确定模块

图 10

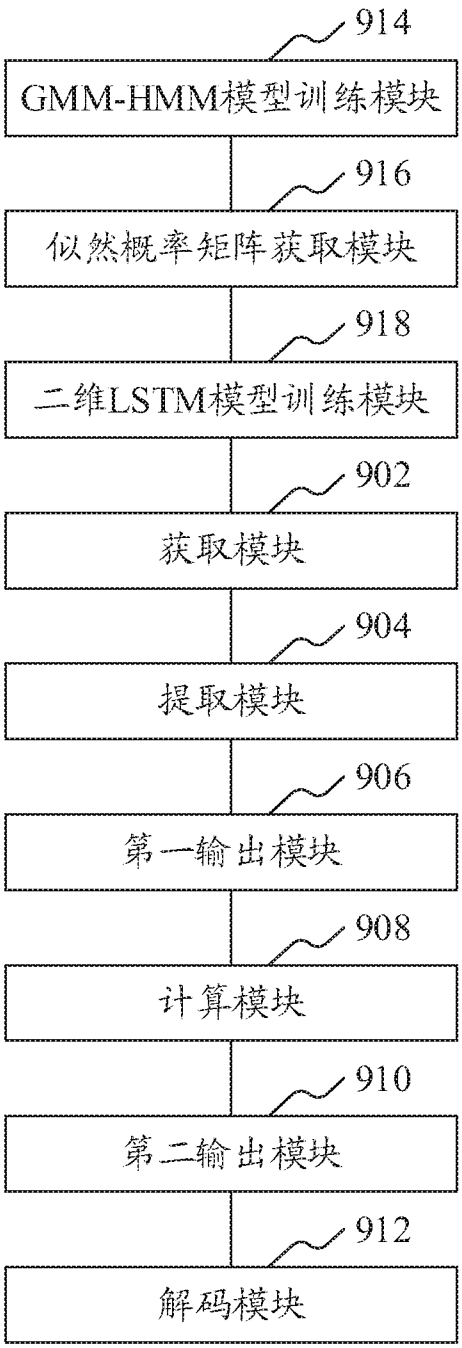

914
GMM-HMM模型训练模块
916
似然概率矩阵获取模块
918
二维LSTM模型训练模块
902
获取模块
904
提取模块
906
第一输出模块
908
计算模块
910
第二输出模块
912
解码模块

图 11

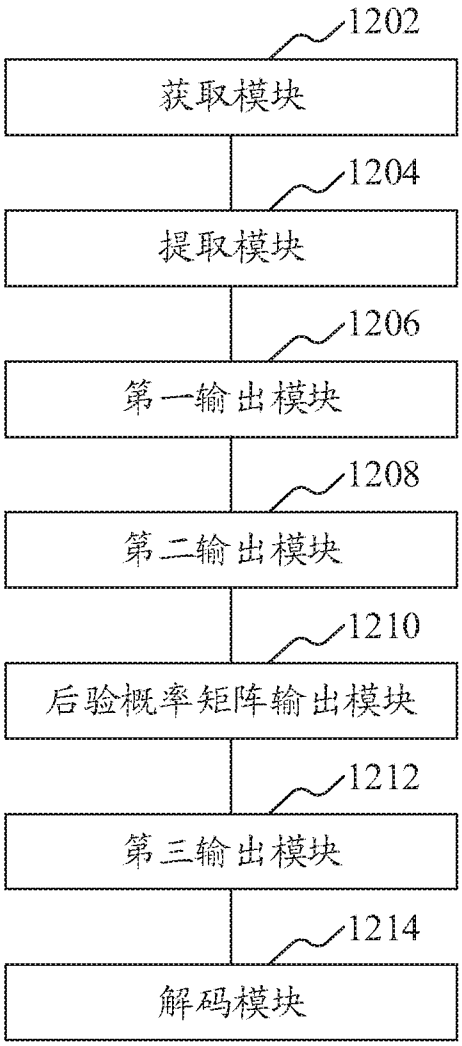

1202
获取模块
1204
提取模块
1206
第一输出模块
1208
第二输出模块
1210
后验概率矩阵输出模块
1212
第三输出模块
1214
解码模块

图 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/100043

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02 (2006.01) i; G10L 15/14 (2006.01) i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 15/-

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)
WPI, EPODOC, CNPAT, CNKI: 平安科技, 梁浩, 语音, 识别, 滤波器组, 长短期记忆网络, 长短时记忆网络, 长短时递归神经网络, 长短期递归神经网络, 梅尔, 美尔, 混合高斯模型, 高斯混合模型, 隐马尔科夫模型, 马尔科夫, 马尔可夫, 似然概率, 后验概率; speech, voice, recogni+, identif+, distinguish+, filter bank, filterbank, filter-bank, LSTM, long short-term memory, long short term memory, MFCC, mel frequency, mel-frequency, GMM, Gaussian mixed model, HMM, hidden Markov model, likelihood, probability, posterior

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105976812 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 28 September 2016 (28.09.2016), description, paragraphs [0004]-[0136], and figures 1-7	1-20
A	CN 106328122 A (SHENZHEN VISION TECHNOLOGY CO., LTD.), 11 January 2017 (11.01.2017), entire document	1-20
A	CN 105513591 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 20 April 2016 (20.04.2016), entire document	1-20
A	CN 105679316 A (SHENZHEN WEIFU ROBOT TECHNOLOGY CO., LTD.), 15 June 2016 (15.06.2016), entire document	1-20
A	CN 105810192 A (SPREADTRUM COMMUNICATIONS (SHANGHAI) CO., LTD.), 27 July 2016 (27.07.2016), entire document	1-20
A	US 2017161256 A1 (MITSUBISHI ELECTRIC RESEARCH LABORATORIES INC.), 08 June 2017 (08.06.2017), 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	Date of mailing of the international search report
01 March 2018	15 March 2018
Name and mailing address of the ISA 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 CHEN, Honghong Telephone No. (86-10) 53962642

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/100043

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105976812 A	28 September 2016	None	
CN 106328122 A	11 January 2017	None	
CN 105513591 A	20 April 2016	None	
CN 105679316 A	15 June 2016	None	
CN 105810192 A	27 July 2016	None	
US 2017161256 A1	08 June 2017	US 9842106 B2	12 December 2017
		WO 2017094911 A1	08 June 2017

国际检索报告

国际申请号 PCT/CN2017/100043

A. 主题的分类

G10L 15/02(2006.01)i；G10L 15/14(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)

G10L15/-

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））

WPI, EPODOC, CNPAT, CNKI: 平安科技,梁浩,语音,识别,滤波器组,长短期记忆网络,长短时记忆网络,长短时递归神经网络,长短期递归神经网络,梅尔,美尔,混合高斯模型,高斯混合模型,隐马尔科夫模型,马尔科夫,马尔可夫,似然概率,后验概率; speech, voice, recogni+, identif+, distinguish+, filter bank, filterbank, filter-bank, LSTM, long short-term memory, long short term memory, MFCC, mel frequency, mel-frequency, GMM, Gaussian mixed model, HMM, hidden Markov model, likelihood, probability, posterior

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 105976812 A (腾讯科技深圳有限公司) 2016年 9月 28日 (2016-09-28) 说明书第[0004]-[0136]段、附图1-7	1-20
A	CN 106328122 A (深圳市唯特视科技有限公司) 2017年 1月 11日 (2017-01-11) 全文	1-20
A	CN 105513591 A (百度在线网络技术北京有限公司) 2016年 4月 20日 (2016-04-20) 全文	1-20
A	CN 105679316 A (深圳微服机器人科技有限公司) 2016年 6月 15日 (2016-06-15) 全文	1-20
A	CN 105810192 A (展讯通信上海有限公司) 2016年 7月 27日 (2016-07-27) 全文	1-20
A	US 2017161256 A1 (MITSUBISHI ELECTRIC RESEARCH LABORATORIES INC.) 2017年 6月 8日 (2017-06-08) 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	国际检索报告邮寄日期
2018年 3月 1日	2018年 3月 15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 陈红红 电话号码 (86-10)53962642

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/100043

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	105976812	A	2016年 9月 28日	无			
CN	106328122	A	2017年 1月 11日	无			
CN	105513591	A	2016年 4月 20日	无			
CN	105679316	A	2016年 6月 15日	无			
CN	105810192	A	2016年 7月 27日	无			
US	2017161256	A1	2017年 6月 8日	US	9842106	B2	2017年 12月 12日
				WO	2017094911	A1	2017年 6月 8日