(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113643718 B
(45)授权公告日 2024.06.18

(21)申请号 202110934541.1
(22)申请日 2021.08.16
(65)同一申请的已公布的文献号
申请公布号 CN 113643718 A
(43)申请公布日 2021.11.12
(73)专利权人 贝壳找房(北京)科技有限公司
地址 100085 北京市海淀区西二旗西路2号院35号楼01层102-1
(72)发明人 解传栋 李先刚 邹伟 王健 常超 沈明
(74)专利代理机构 北京庚致知识产权代理事务所(特殊普通合伙) 11807
专利代理师 韩德凯
(51)Int.Cl.
G10L 25/48(2013.01)
G10L 25/30(2013.01)
G06F 18/23(2023.01)
(56)对比文件
CN 107331384 A,2017.11.07
CN 103971678 A,2014.08.06
审查员 梁吉

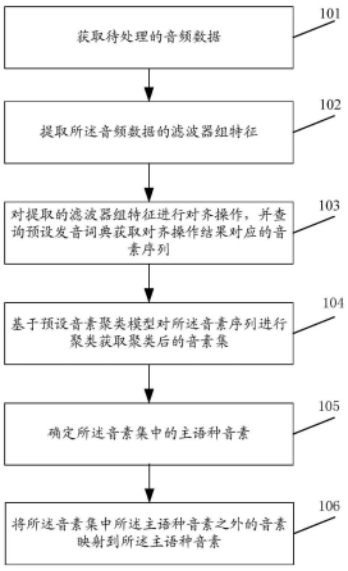
权利要求书2页 说明书9页 附图5页

(54)发明名称

音频数据处理方法和装置

(57)摘要

本公开实施例提供了一种音频数据处理方法和装置,所述方法包括:获取待处理的音频数据;提取所述音频数据的滤波器组特征;对提取的滤波器组特征进行对齐操作,并查询预设发音词典获取对齐操作结果对应的音素序列;确定所述音素集中的主语种音素;将所述音素集中所述主语种音素之外的音素映射到所述主语种音素。该方法能够在低成本前提下,提高音素映射的准确度。



CN 113643718 B

1.一种音频数据处理方法,其特征在于,所述方法包括:

获取待处理的音频数据;

提取所述音频数据的滤波器组特征;

对提取的滤波器组特征进行对齐操作,并查询预设发音词典获取对齐操作结果对应的音素序列;

基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集,其中,所述预设音素聚类模型根据音素的后验概率进行初步聚类,并根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并,直到合并后的音素集的数目为预设聚类数目;

确定所述音素集中的主语种音素;以及

将所述音素集中所述主语种音素之外的音素映射到所述主语种音素;

所述根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并,包括:

计算初步聚类的每个类的高斯对数似然度,以及两个类合并之后的高斯对数似然度;

根据每个类的高斯对数似然度,以及两个类合并之后的高斯对数似然度计算所述两个类的距离;

将距离最小的两个类对应的音素集合并。

2.根据权利要求1所述的方法,其特征在于,

通过高斯混合模型-隐马尔可夫模型对提取的滤波器组特征进行对齐操作。

3.根据权利要求1所述的方法,其特征在于,所述根据音素的后验概率进行初步聚类,包括:

基于深度神经网络-隐马尔可夫模型获得音素的后验概率;

根据所述音素的后验概率进行初步聚类。

4.根据权利要求1所述的方法,其特征在于,所述根据音素的后验概率进行初步聚类,包括:

基于音素的后验概率,计算音素之间的距离,将距离小于预设阈值的音素聚为一类。

5.根据权利要求1所述的方法,其特征在于,所述基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集之后,所述确定所述音素集中主语种音素之前,所述方法进一步包括:

将聚类后的音素集两两合并作为音素集组合;

获取每个音素集组合的高斯对数似然度,并按高斯对数似然度从大到小的顺序进行排列;

选择前预设聚类数目的音素集组合作为用于音素映射的音素集。

6.根据权利要求1至5中任一项所述的方法,其特征在于,所述确定所述音素集中的主语种音素,包括:

根据配置信息确定所述音素集中的主语种音素。

7.根据权利要求1至5中任一项所述的方法,其特征在于,所述确定所述音素集中的主语种音素,包括:

将音素集中后验概率最大的音素作为主语种音素。

8.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现权利要求1至7中任一项所述的方法。

9.一种计算机程序产品,包括计算机程序,其特征在于,该计算机程序被处理器执行时实现权利要求1至7中任一项所述的方法。

音频数据处理方法和装置

技术领域

[0001] 本公开实施例涉及一种音频数据处理方法和装置。

背景技术

[0002] 目前在语音领域,不同语种都有一套完整的发音体系,对应一套音素集。但是,实际应用中,经常会出现不同语种夹杂的情况,如中文夹杂英文、日文;即使同一语种中也会存在普通语言夹杂方言等情况。

[0003] 在实际应用中,需要将不同语种映射到同一种,如可以通过人工映射,也可以通过收集大量语音数据作为训练样本来实现映射模型的训练。

[0004] 在实现本公开的过程中,发明人发现通过人工映射方式实现成本比较高,且易出现错误;通过样本训练模型方式实现收集样本困难,导致模型映射准确度低。

发明内容

[0005] 有鉴于此,本公开实施例提供一种音频数据处理方法和装置,能够在低成本前提下,提高音素映射的准确度。

[0006] 为解决上述技术问题,本公开实施例的技术方案是这样实现的:

[0007] 在一个实施例中,提供了一种音频数据处理方法,所述方法包括:

[0008] 获取待处理的音频数据;

[0009] 提取所述音频数据的滤波器组特征;

[0010] 对提取的滤波器组特征进行对齐操作,并查询预设发音词典获取对齐操作结果对应的音素序列;

[0011] 基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集;其中,所述预设音素聚类模型根据音素的后验概率进行初步聚类,并根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并;直到合并后的音素集的数目为预设聚类数目;

[0012] 确定所述音素集中的主语种音素;

[0013] 将所述音素集中所述主语种音素之外的音素映射到所述主语种音素。

[0014] 在另一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现所述音频数据处理方法的步骤。

[0015] 在另一个实施例中,提供了一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现所述音频数据处理方法。

[0016] 由上面的技术方案可见,上述实施例中获取待处理的音频数据对应的音素序列,基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集,并基于所述音素集将所述音频数据中主语种音素之外的音素映射到主语种音素。由于本公开实施例中的预设音素聚类模型是根据每个音素的后验概率进行初步聚类,再根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并,最终输出聚类结果;因此,本公开实施例中的预设音素聚类模型通过计算音素的后验概率和类的高斯对数似然度来确定映射关系,不需收集大

量样本进行训练获得,能够在低成本前提下,提高音素映射的准确度。

附图说明

[0017] 为了更清楚地说明本公开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公开的一些实施例,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他的附图。

[0018] 图1为本公开实施例一中音频数据处理流程示意图;

[0019] 图2为本公开实施例中预设音素聚类模型结构示意图;

[0020] 图3为本公开实施例中基于音素序列获取聚类后的音素集的流程示意图;

[0021] 图4为本公开实施例中基于高斯计算的混合聚类算法对初步聚类的音素集进行合并流程示意图;

[0022] 图5为本公开实施例二中音频数据处理流程示意图;

[0023] 图6为本公开实施例中音频数据处理装置结构示意图;

[0024] 图7为本发明实施例提供的电子设备的实体结构示意图。

具体实施方式

[0025] 下面将结合本公开实施例中的附图,对本公开实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅是本公开一部分实施例,而不是全部的实施例。基于本公开中的实施例,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他实施例,都属于本公开保护的范围。

[0026] 本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等(如果存在)是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本发明的实施例例如能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含。例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其他步骤或单元。

[0027] 下面以具体实施例对本发明的技术方案进行详细说明。下面几个具体实施例可以相互结合,对于相同或相似的概念或过程可能在某些实施例不再赘述。

[0028] 本公开实施例中提供一种音频数据处理方法,获取待处理的音频数据对应的音素序列,基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集,并将所述音素集中所述主语种音素之外的音素映射到所述主语种音素。由于本公开实施例中的预设音素聚类模型根据每个音素的后验概率进行初步聚类,再根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并;可见,本公开实施例中的预设音素聚类模型通过计算音素的后验概率和类的高斯对数似然度来确定映射关系,不需收集样本进行训练获得,因此本公开实施例提供的技术方案能够在低成本前提下,提高音素映射的准确度。

[0029] 下面结合附图,详细说明本公开实施例中实现音频数据处理过程。

[0030] 实施例一

[0031] 参见图1,图1为本公开实施例一中音频数据处理流程示意图。具体步骤为:

[0032] 步骤101,获取待处理的音频数据。

[0033] 步骤102,提取所述音频数据的滤波器组特征。

[0034] 滤波器组(Fbank)是需要语音特征参数提取方法之一,因其独特的基于倒谱的提取方式,更加的符合人类的听觉原理,是最为普遍、最有效的语音特征提取算法。

[0035] 可以基于Filter Bank算法提取音频信号的Fbank特征;Fbank特征提取方法相当于梅尔频率倒谱参数(Mel-Frequency Cepstral Coefficients,MFCC)去掉最后一步的离散余弦变换(有损变换),与MFCC特征相比,Fbank特征保留了更多的原始语音数据。

[0036] 本公开实施例对提取音频信号的Fbank特征的实现方式不进行限制。

[0037] 步骤103,对提取的滤波器组特征进行对齐操作,并查询预设发音词典获取对齐操作结果对应的音素序列。

[0038] 对滤波器组特征进行对齐操作后,获得的对齐操作结果为滤波器组特征对应的符号表示,如音素、文字等。

[0039] 本公开实施例中对提取的Fbank特征进行对齐操作可以但不限于通过高斯混合模型-隐马尔可夫模型(GMM-HMM)实现,其中,GMM全称为高斯混合模型(Gaussian Mixture Model),HMM全称为隐马尔可夫模型(Hidden Markov Model)。

[0040] 其中,GMM-HMM模型对输入的Fbank特征使用期望最大化算法(Expectation-maximization algorithm,EM)算法更新模型参数;基于更新后的参数重新对Fbank特征进行观察序列-状态序列的对齐;再次使用EM算法更新模型参数,进行对齐操作,直至收敛,获得收敛时的模型参数作为最终训练完成的GMM-HMM模型。

[0041] 训练好的GMM-HMM模型可用于本申请实施例中对滤波器组特征进行对齐操作。

[0042] 具体实现时,根据实际应用获取预设发音词典。

[0043] 使用对齐操作结果查询预设发音词典,获取对齐结果对应的音素序列。

[0044] 步骤104,基于预设音素聚类模型对所述音素序列进行聚类获取聚类后的音素集。

[0045] 其中,所述预设音素聚类模型所述预设音素聚类模型根据音素的后验概率进行初步聚类,并根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并;直到合并后的音素集的数目为预设聚类数目。

[0046] 结合图2,详细说明本公开实施例中预设音素聚类模型实现音素聚类过程,图2为本公开实施例中预设音素聚类模型结构示意图。

[0047] 参见图3,图3为本公开实施例中基于音素序列获取聚类后的音素集的流程示意图。具体步骤为:

[0048] 步骤301,获取音素序列中每个音素的后验概率。

[0049] 如图2所示可以基于深度神经网络(Deep Neural Networks,DNN)-隐马尔可夫模型(HMM)获得每个音素的后验概率,n个音素对应n个后验概率,第i个音素的后验概率为P_i,i=1…n,本公开实施例中具体实现时不限于通过该方式获取音素的后验概率。

[0050] 步骤302,基于音素的后验概率进行初步聚类。

[0051] 初步聚类时,基于音素的后验概率,计算音素之间的距离,距离小于预设阈值的两个音素聚为一类,以此类推,按照距离的不同,分成多类,作为初步聚类后的多个类别。

[0052] 本申请实施例中对基于音素的后验概率进行初步聚类的方式不进行限制,在具体

实现时可以采用层次聚类(Hierarchical Clustering)算法实现初步聚类,也可以通过K平均(K-means)算法进行初步聚类。

[0053] 如图2所示,初步聚类获取的聚类类别的个数为m,$S_i^1, i = 1 \ldots m$,每个类别中包含1个或多个音素。

[0054] 步骤303,基于高斯计算的混合聚类算法对初步聚类的音素集进行合并,直到合并后的音素集的数目为预设聚类数目。

[0055] 基于高斯计算的混合聚类算法对初步聚类的音素集进行合并,即合并(距离)最近的对(Merge Closest Pair)。

[0056] 参见图4,图4为本公开实施例中基于高斯计算的混合聚类算法对初步聚类的音素集进行合并流程示意图。具体步骤为:

[0057] 步骤401,计算初步聚类的每个类的高斯对数似然度,以及两个类合并之后的高斯对数似然度。

[0058] 假设第k类别有n_k个音素,第k类别中的音素服从参数为$\emptyset_k$的高斯分布,因此,第k类别的高斯对数似然度可以用如下公式表示:

[0059] $$L_k = \sum_{i=1}^{n_k} \log\left[G(x_i;\phi_k)\right]$$

[0060] 其中,L_k表示第k类别对应的高斯对数似然度,n_k表示第k类别对应的音素集中音素的个数,x_i为第k类别对应的音素中第i个音素,$G(x_i;\varphi_k)$表示第k类别对应的音素集中第i个音素服从参数为φ_k的高斯分布。

[0061] 两个类别合并后,如第k和第j类别合并后的高斯对数似然度可以通过如下公式进行计算:

[0062] $$L_{k+j} = \sum_{i=1}^{n_k+n_k} \log\left[G(x_i;\phi_{k+j})\right]$$

[0063] 其中,L_{k+j}表示第k类对应的高斯对数似然度,n_k表示第k类别对应的音素集中音素的个数,n_j表示第j类别对应的音素集中音素的个数,x_i为第k类别和第j类别合并后对应的音素中第i个音素,$G(x_i;\varphi_{k+j})$表示第k类别和第j类别合并后对应音素集中第i个音素服从参数为φ_{k+j}的高斯分布。

[0064] 步骤402,根据每个类的高斯对数似然度,以及两个类合并之后的高斯对数似然度计算所述两个类的距离。

[0065] 通过下述公式计算两个类别的距离(相似度):

[0066] $\Delta = L_{k+j} - (L_k + L_j)$

[0067] 步骤403,将距离最小的两个类对应的音素集合并。

[0068] 不断执行步骤401到步骤403,直到合并后的音素集的数目为预设聚类数目,如图2所示假设预设聚类数目为I,则合并后的音素集为$C_1 \ldots C_I$时,结束音素集的合并,完成预设音素聚类模型的聚类。

[0069] 步骤304,将合并后的预设聚类数目的音素集作为输出的音素集。

[0070] 步骤105,确定所述音素集中的主语种音素。

[0071] 本公开实施例中针对聚类后的每个音素集会确定一个主语种音素,具体实现时可以通过如下实现方式实现,但不限于下述两种实现方式:

[0072] 第一种:

[0073] 根据配置信息确定每个音素集中的主语种音素。

[0074] 在配置信息中指定作为主语种音素的音素。

[0075] 如英语和汉语夹杂的音频数据,根据实际应用可以设置英语对应的音素作为主语种音素,也可以设置汉语对应的音素作为主语种音素。

[0076] 第二种:

[0077] 将音素集中后验概率最大的音素作为主语种音素。

[0078] 针对一个音素集中的所有音素,确定后验概率最大的音素作为主语种音素。

[0079] 步骤106,将所述音素集中所述主语种音素之外的音素映射到所述主语种音素。

[0080] 本公开实施例中获取待处理的音频数据对应的音素序列,基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集,并将所述音素集中所述主语种音素之外的音素映射到所述主语种音素。由于本公开实施例中的预设音素聚类模型是根据每个音素的后验概率进行初步聚类,再根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并,最终输出聚类结果;因此,本公开实施例中的预设音素聚类模型通过计算音素的后验概率和类的高斯对数似然度来确定映射关系,不需收集大量样本进行训练获得,能够在低成本前提下,提高音素映射的准确度。

[0081] 实施例二

[0082] 参见图5,图5为本公开实施例二中音频数据处理流程示意图。具体步骤为:

[0083] 步骤501,获取待处理的音频数据。

[0084] 步骤502,提取所述音频数据的滤波器组特征。

[0085] Fbank是需要语音特征参数提取方法之一,因其独特的基于倒谱的提取方式,更加的符合人类的听觉原理,是最为普遍、最有效的语音特征提取算法。

[0086] 可以基于Filter Bank算法提取音频信号的Fbank特征;Fbank特征提取方法相当于MFCC去掉最后一步的离散余弦变换(有损变换),与MFCC特征相比,Fbank特征保留了更多的原始语音数据。

[0087] 本公开实施例对提取音频信号的Fbank特征的实现方式不进行限制。

[0088] 步骤503,对提取的滤波器组特征进行对齐操作,并查询预设发音词典获取对齐操作结果对应的音素序列。

[0089] 本公开实施例中对提取的Fbank特征进行对齐操作可以但不限于通过高斯混合模型(GMM)-隐马尔可夫(HMM)模型实现。

[0090] GMM-HMM模型对输入的Fbank特征进行初始化对齐;使用EM算法更新模型参数;基于更新后的参数重新对音频进行(观察序列-状态序列)的对齐;再次使用EM算法更新模型参数,进行对齐操作,直至收敛,完成GMM-HMM模型训练。

[0091] 具体实现时,根据实际应用获取预设发音词典。

[0092] 使用对齐操作结果查询预设发音词典,获取对齐结果对应的音素序列。

[0093] 步骤504,基于预设音素聚类模型对所述音素序列进行聚类获取聚类后的音素集。

[0094] 其中,所述预设音素聚类模型所述预设音素聚类模型根据音素的后验概率进行初

步聚类,并根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并;直到合并后的音素集的数目为预设聚类数目。

[0095] 预设音素聚类模型实现输入音素序列输出聚类后的音素集的过程如下:

[0096] 第一步、获取音素序列中每个音素的后验概率。

[0097] 可以基于DNN-HMM模型获得每个音素的后验概率,本公开实施例中具体实现时不限于通过该方式获取音素的后验概率。

[0098] 第二步、基于音素的后验概率进行初步聚类。

[0099] 初步聚类时,基于音素的后验概率,计算音素之间的距离,距离小于预设阈值的两个音素聚为一类,以此类推,按照距离的不同,分成多类,作为初步聚类后的多个类别。

[0100] 本申请实施例中对基于音素的后验概率进行初步聚类的方式不进行限制,在具体实现时可以采用层次聚类(Hierarchical Clustering)算法实现初步聚类,也可以通过K平均(K-means)算法进行初步聚类。

[0101] 第三步、基于高斯计算的混合聚类算法对初步聚类的音素集进行合并,直到合并后的音素集的数目为预设聚类数目。

[0102] 本步骤中基于高斯计算的混合聚类算法对初步聚类的音素集进行合并的具体实现可以为如下方式:

[0103] 计算初步聚类的每个类的高斯对数似然度,以及两个类合并之后的高斯对数似然度。

[0104] 假设第k类别有n_k个音素,第k类别中的音素服从参数为ϕ_k的高斯分布,因此,第k类别的高斯对数似然度可以用如下公式表示:

[0105] $$L_k=\sum_{i=1}^{n_k}\log\left[G(x_i;\phi_k)\right]$$

[0106] 其中,L_k表示第k类别对应的高斯对数似然度,n_k表示第k类别对应的音素集中音素的个数,x_i为第k类别对应的音素中第i个音素,$G(x_i;\varphi_k)$表示第k类别对应音素集中的第i个音素服从参数为φ_k的高斯分布。

[0107] 两个类别合并后,如第k和第j类别合并后的高斯对数似然度可以通过如下公式进行计算:

[0108] $$L_{k+j}=\sum_{i=1}^{n_k+n_k}\log\left[G(x_i;\phi_{k+j})\right]$$

[0109] 其中,L_{k+j}表示第k类对应的高斯对数似然度,n_k表示第k类别对应的音素集中音素的个数,n_j表示第j类别对应的音素集中音素的个数,x_i为第k类别和第j类别合并后对应的音素中第i个音素,$G(x_i;\varphi_{k+j})$表示第k类别和第j类别合并后对应音素集中第i个音素服从参数为φ_{k+j}的高斯分布。

[0110] 然后根据每个类的高斯对数似然度,以及两个类合并之后的高斯对数似然度计算所述两个类的距离。

[0111] 通过下述公式计算两个类别的距离(相似度):

[0112] $\Delta=L_{k+j}-(L_k+L_j)$

[0113] 最后,将距离最小的两个类对应的音素集合并。

[0114] 不断执行合并操作,直到合并后的音素集的数目为预设聚类数目。

[0115] 第四步、将合并后的预设聚类数目的音素集作为输出的音素集。

[0116] 步骤505,将聚类后的音素集两两合并作为音素集组合。

[0117] 如聚类后的音素集有I个,则将聚类后的音素集两两合并,则共有2^{I-1}个音素集组合。

[0118] 步骤506,获取每个音素集组合的高斯对数似然度,并按高斯对数似然度从大到小的顺序进行排列。

[0119] 根据每个音素集组合内的所有音素,计算该音素集组合的高斯对数似然度。

[0120] 步骤507,选择前预设聚类数目的音素集组合作为用于音素映射的音素集。

[0121] 步骤508,确定所述音素集中的主语种音素。

[0122] 本公开实施例中针对聚类后的每个音素集会确定一个主语种音素,具体实现时可以通过如下实现方式实现,但不限于下述两种实现方式:

[0123] 第一种:

[0124] 根据配置信息确定每个音素集中的主语种音素。

[0125] 在配置信息中指定作为主语种音素的音素。

[0126] 如英语和汉语夹杂的音频数据,根据实际应用可以设置英语对应的音素作为主语种音素,也可以设置汉语对应的音素作为主语种音素。

[0127] 第二种:

[0128] 将音素集中后验概率最大的音素作为主语种音素。

[0129] 针对一个音素集中的所有音素,确定后验概率最大的音素作为主语种音素。

[0130] 步骤509,将所述音素集中所述主语种音素之外的音素映射到所述主语种音素。

[0131] 本公开实施例中获取待处理的音频数据对应的音素序列,基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集,再对聚类后的音素集进行两两合并,进一步选择高斯对数似然度大的预设聚类数目的音素集组合作为本公开实施例中用于进行音素映射的音素集,并将所述音素集中所述主语种音素之外的音素映射到所述主语种音素。能进一步提高音素映射的准确性;并且由于本公开实施例中的预设音素聚类模型是根据每个音素的后验概率进行初步聚类,再根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并,最终输出聚类结果;因此,本公开实施例中的预设音素聚类模型通过计算音素的后验概率和类的高斯对数似然度来确定映射关系,不需收集大量样本进行训练获得,能够在低成本前提下,提高音素映射的准确度。

[0132] 基于同样的发明构思,本公开实施例中还提供一种音频数据处理装置。参见图6,图6为本公开实施例中音频数据处理装置结构示意图。所述装置包括:第一获取单元601、提取单元602、对齐单元603、查询单元604、第二获取单元605、确定单元606和处理单元607;

[0133] 第一获取单元601,用于获取待处理的音频数据;

[0134] 提取单元602,用于提取第一获取单元601获取的音频数据的滤波器组特征;

[0135] 对齐单元603,用于对提取单元602提取的滤波器组特征进行对齐操作;

[0136] 查询单元604,用于查询预设发音词典获取对齐单元603获取的对齐操作结果对应的音素序列;

[0137] 第二获取单元605,用于基于预设音素聚类模型获取查询单元604获取的音素序列对应的聚类后的音素集;其中,所述预设音素聚类模型根据每个音素的后验概率进行初步聚类,根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并;直到合并后的音素集的数目为预设聚类数目;

[0138] 确定单元606,用于确定所述音素集中的主语种音素;

[0139] 处理单元607,用于将第二获取单元605获取的音素集中所述确定单元606确定的主语种音素之外的音素映射到所述主语种音素。

[0140] 在另一个实施例中,

[0141] 对齐单元603,具体用于通过高斯混合模型-隐马尔可夫模型对提取的滤波器组特征进行对齐操作。

[0142] 在另一个实施例中,

[0143] 所述预设音素聚类模型根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并时,包括:计算初步聚类的每个类的高斯对数似然度,以及两个类合并之后的高斯对数似然度;根据每个类的高斯对数似然度,以及两个类合并之后的高斯对数似然度计算所述两个类的距离;将距离最小的两个类对应的音素集合并。

[0144] 在另一个实施例中,

[0145] 所述预设音素聚类模型根据音素的后验概率进行初步聚类时,基于深度神经网络-隐马尔可夫模型获得音素的后验概率;根据所述音素的后验概率进行初步聚类。

[0146] 在另一个实施例中,

[0147] 所述预设音素聚类模型根据音素的后验概率进行初步聚类时,基于音素的后验概率,计算音素之间的距离,将距离小于预设阈值的音素聚为一类。

[0148] 处理单元607,进一步用于在第二获取单元605基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集之后,所述确定单元606确定所述音素集中的主语种音素之前,将聚类后的音素集两两合并作为音素集组合;获取每个音素集组合的高斯对数似然度,并按高斯对数似然度从大到小的顺序进行排列;选择前预设聚类数目的音素集组合作为用于音素映射的音素集。

[0149] 在另一个实施例中,

[0150] 确定单元606,具体用于根据配置信息确定每个音素集中的主语种音素。

[0151] 在另一个实施例中,

[0152] 确定单元606,具体用于将音素集中后验概率最大的音素作为主语种音素。

[0153] 上述实施例的单元可以集成于一体,也可以分离部署;可以合并为一个单元,也可以进一步拆分成多个子单元。

[0154] 在另一个实施例中,还提供一种电子设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现所述音频数据处理方法的步骤。

[0155] 在另一个实施例中,还提供一种计算机可读存储介质,其上存储有计算机指令,所述指令被处理器执行时可实现所述音频数据处理方法中的步骤。

[0156] 在另一个实施例中,还提供了一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现所述音频数据处理方法。

[0157] 图7为本发明实施例提供的电子设备的实体结构示意图。如图7所示,该电子设备可以包括:处理器(Processor)710、通信接口(Communications Interface)720、存储器(Memory)730和通信总线740,其中,处理器710,通信接口720,存储器730通过通信总线740完成相互间的通信。处理器710可以调用存储器730中的逻辑指令,以执行如下方法:

[0158] 获取待处理的音频数据;

[0159] 提取所述音频数据的滤波器组特征;

[0160] 对提取的滤波器组特征进行对齐操作,并查询预设发音词典获取对齐操作结果对应的音素序列;

[0161] 基于预设音素聚类模型获取所述音素序列对应的聚类后的音素集;其中,所述预设音素聚类模型根据音素的后验概率进行初步聚类,并根据聚类后的每一类的高斯对数似然度对初步聚类后的音素集进行合并;直到合并后的音素集的数目为预设聚类数目;

[0162] 确定所述音素集中的主语种音素;

[0163] 将所述音素集中所述主语种音素之外的音素映射到所述主语种音素。

[0164] 此外,上述的存储器730中的逻辑指令可以通过软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0165] 以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0166] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在计算机可读存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行各个实施例或者实施例的某些部分所述的方法。

[0167] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明保护的范围之内。

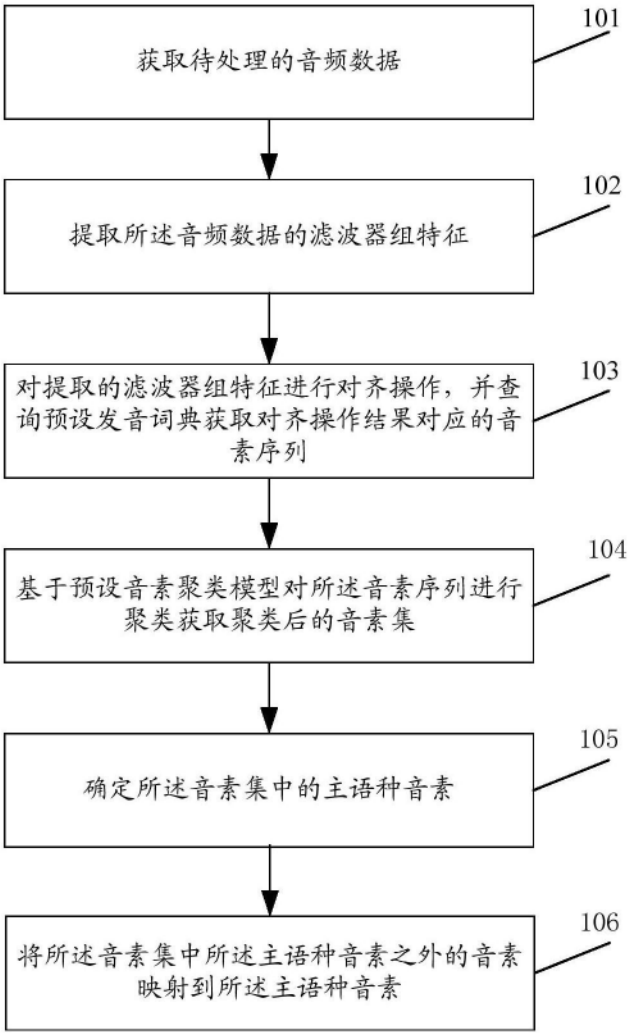

获取待处理的音频数据
101
提取所述音频数据的滤波器组特征
102
对提取的滤波器组特征进行对齐操作，并查询预设发音词典获取对齐操作结果对应的音素序列
103
基于预设音素聚类模型对所述音素序列进行聚类获取聚类后的音素集
104
确定所述音素集中的主语种音素
105
将所述音素集中所述主语种音素之外的音素映射到所述主语种音素
106

图1

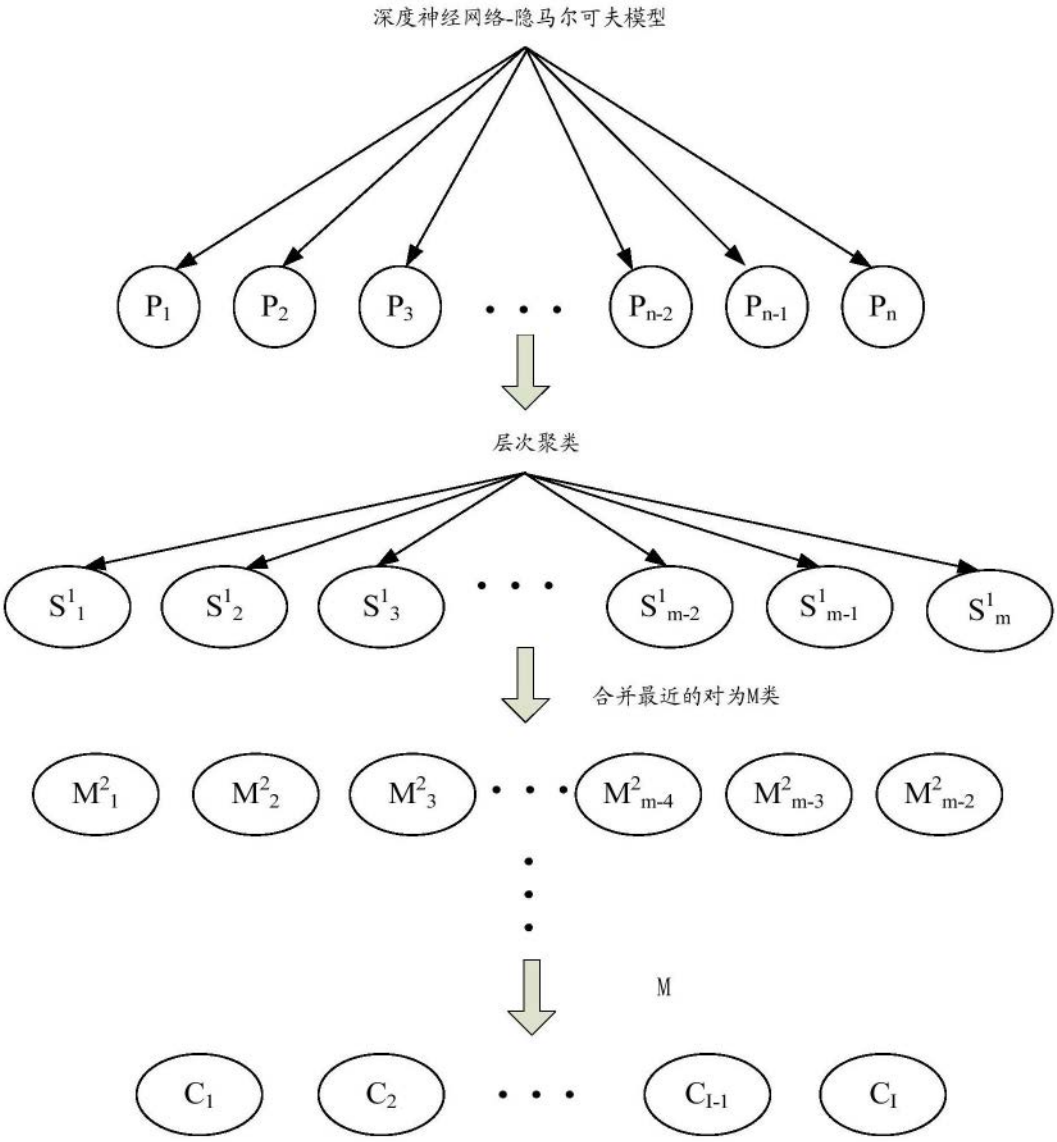

深度神经网络-隐马尔可夫模型
P1
P2
P3
Pn-2
Pn-1
Pn
层次聚类
S11
S12
S13
S1m-2
S1m-1
S1m
合并最近的对为M类
M21
M22
M23
M2m-4
M2m-3
M2m-2
M
C1
C2
CI-1
CI

图2

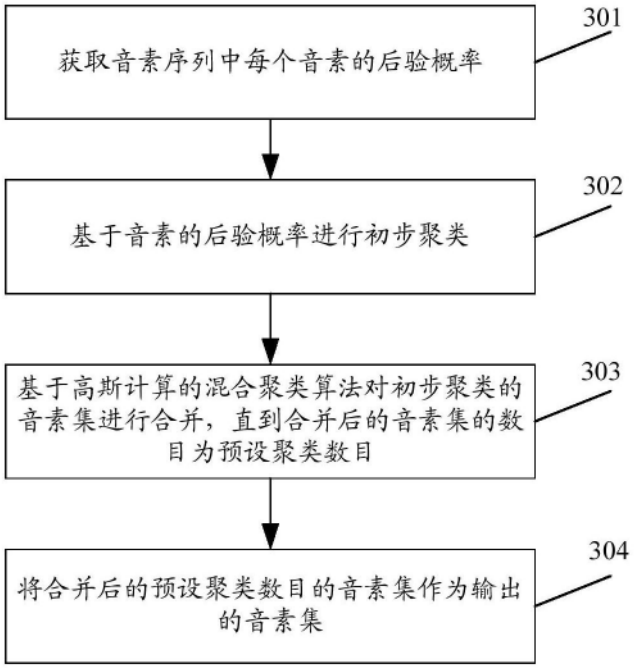

获取音素序列中每个音素的后验概率
301
基于音素的后验概率进行初步聚类
302
基于高斯计算的混合聚类算法对初步聚类的音素集进行合并，直到合并后的音素集的数目为预设聚类数目
303
将合并后的预设聚类数目的音素集作为输出的音素集
304

图3

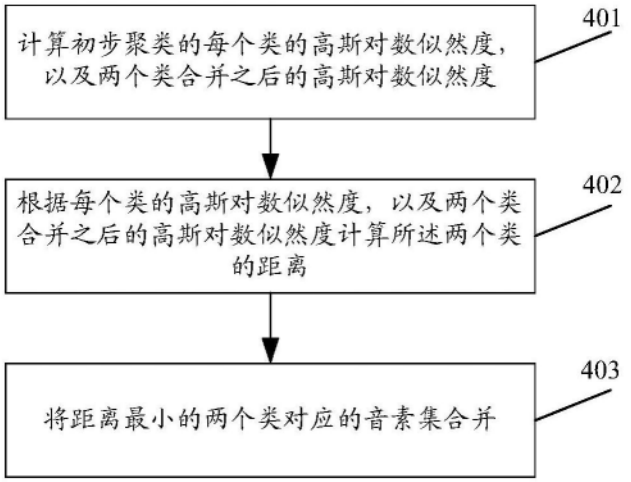

计算初步聚类的每个类的高斯对数似然度，以及两个类合并之后的高斯对数似然度
401
根据每个类的高斯对数似然度，以及两个类合并之后的高斯对数似然度计算所述两个类的距离
402
将距离最小的两个类对应的音素集合并
403

图4

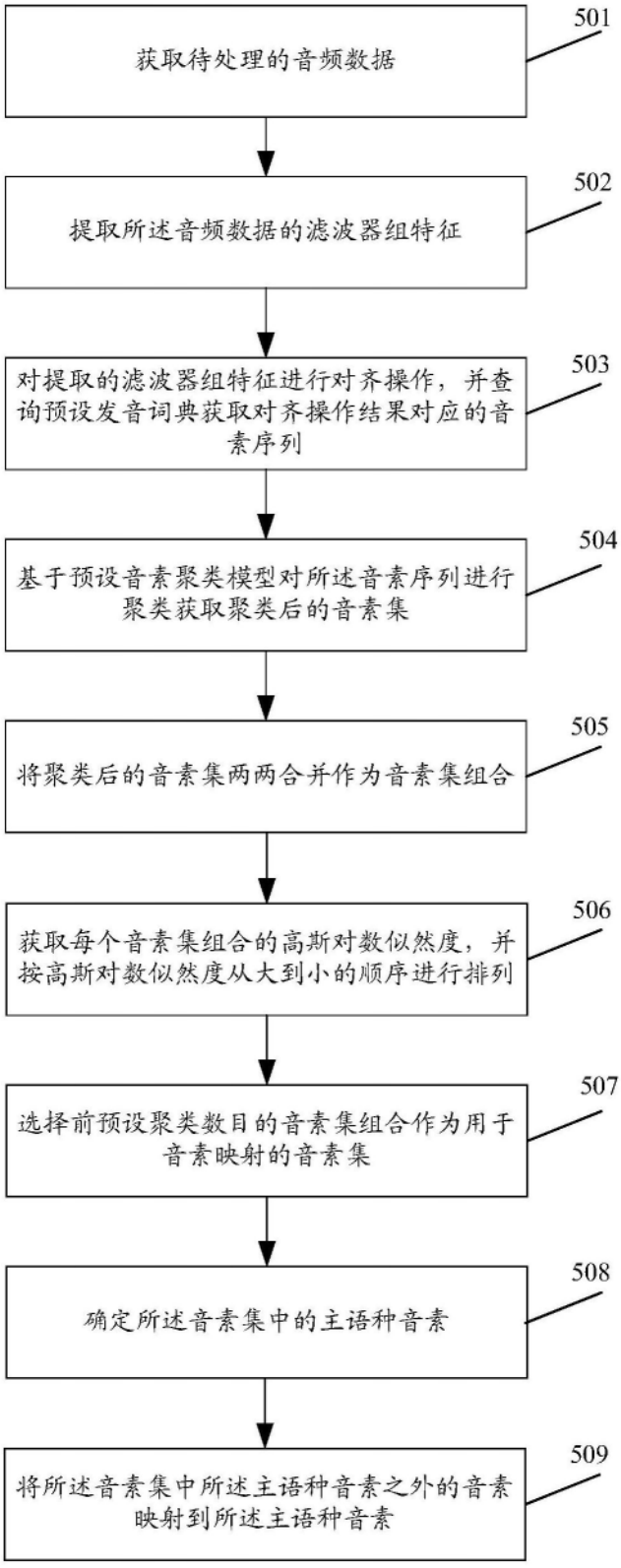

501
获取待处理的音频数据
502
提取所述音频数据的滤波器组特征
503
对提取的滤波器组特征进行对齐操作，并查询预设发音词典获取对齐操作结果对应的音素序列
504
基于预设音素聚类模型对所述音素序列进行聚类获取聚类后的音素集
505
将聚类后的音素集两两合并作为音素集组合
506
获取每个音素集组合的高斯对数似然度，并按高斯对数似然度从大到小的顺序进行排列
507
选择前预设聚类数目的音素集组合作为用于音素映射的音素集
508
确定所述音素集中的主语种音素
509
将所述音素集中所述主语种音素之外的音素映射到所述主语种音素

图5

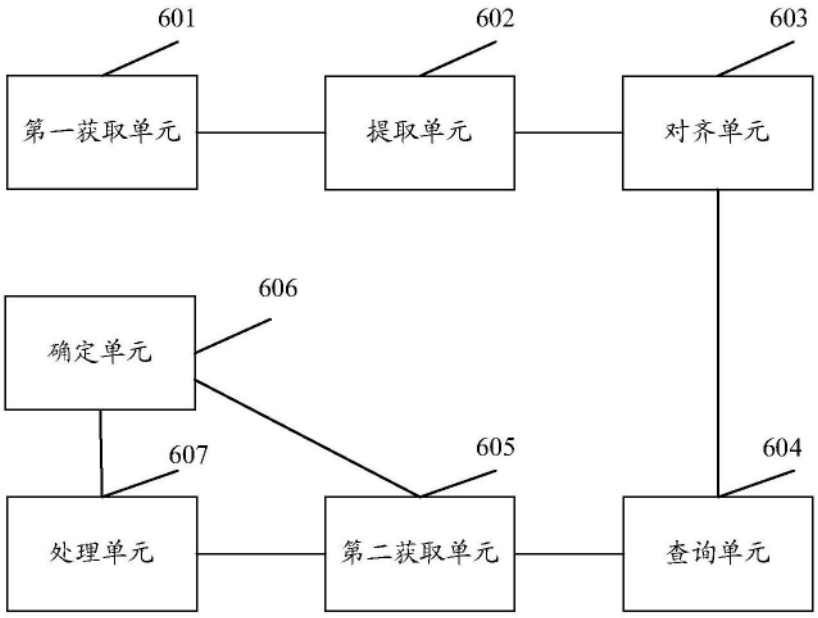

601
602
603
第一获取单元
提取单元
对齐单元
606
确定单元
607
605
604
处理单元
第二获取单元
查询单元

图6

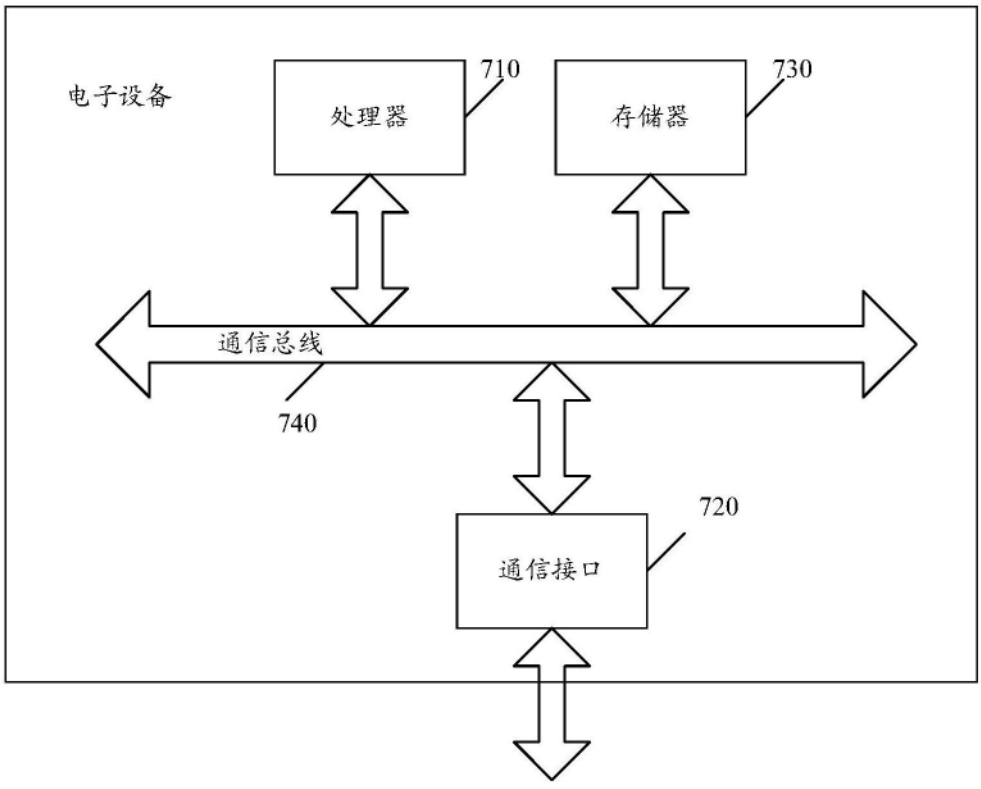

电子设备
710
处理器
730
存储器
通信总线
740
720
通信接口

图7