(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102324232 A
(43)申请公布日 2012.01.18

(21)申请号 201110267690.3

(22)申请日 2011.09.12

(71)申请人 辽宁工业大学
地址 121000 辽宁省锦州市古塔区士英街169号

(72)发明人 霍春宝 张健 赵立辉 刘春玲 张彩娟

(74)专利代理机构 锦州辽西专利事务所 21225
代理人 李辉

(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/06(2006.01)

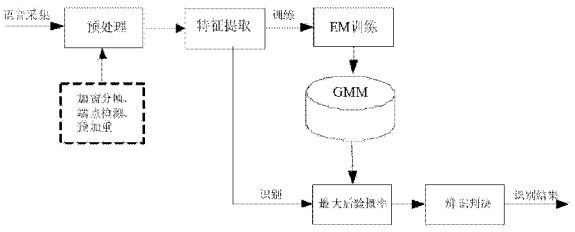
权利要求书 5 页 说明书 14 页 附图 4 页

(54)发明名称

基于高斯混合模型的声纹识别方法及系统

(57)摘要

一种基于高斯混合模型的声纹识别方法及系统，步骤如下：语音信号采集；语音信号预处理；语音信号特征参数提取：采用梅尔频率倒谱系数(MFCC)，MFCC的阶数通常取为12~16；模型训练：采用EM算法为说话人的语音信号特征参数训练高斯混合模型(GMM)，模型的参数初始化方法选用k-means算法；声纹辨识：将采集到的待识别语音信号特征参数与已建立的说话人语音模型进行比较，并根据最大后验概法进行判断，若对应的说话人模型使得待识别的话者语音特征向量X具有最大的后验概率，则识别出说话人。该方法采用了基于概率统计的高斯混合模型，能很好的反映说话人的语音在特征空间的分布，其概率密度函数比较常见，模型中的参数易于估计和训练，而且具有良好识别性能和抗噪能力。



CN 102324232 A

1. 一种基于高斯混合模型的声纹识别方法，其特征是具体步骤如下：

（1）、语音信号的采集：以程控交换综合实验箱的话机作为采集语音信号的终端设备，通过语音卡采集语音信号；

（2）、语音信号的预处理：通过计算机将提取的语音信号进行分帧加窗操作，在分帧过程中一帧包括 256 个采样点，帧移为 128 个采样点，所加的窗函数为汉明窗；端点检测，采用基于短时能量和短时过零率法相结合的端点检测法；预加重，加重系数的范围为 0.90~1.00；

（3）、语音信号特征参数提取：采用梅尔频率倒谱系数 (MFCC)，MFCC 的阶数通常取为 12~16；

（4）、模型训练：采用 EM 算法为说话人的语音信号特征参数训练高斯混合模型 (GMM)，模型的参数初始化方法选用 k-means 算法；

（5）、声纹辨识：通过将采集到的待识别语音信号特征参数与库中通过第 1 步骤 1、第 2 步骤、第 3 步骤已建立的说话人语音模型进行比较，并根据最大后验概法进行判断，若对应的说话人模型使得待识别的话者语音特征向量 X 具有最大的后验概率，则认为识别出说话人。

2. 根据权利要求 1 所述的基于高斯混合模型的声纹识别方法，其特征是语音信号特征参数提取步骤如下：

（1）将预处理后的语音信号进行短时傅里叶变换（DFT）得到其频谱 X(k)，语音信号的 DFT 公式为：

$$X(k)=\sum_{k=0}^{N-1}x(n)e^{-j2\pi k/N},0\le k<N \qquad (1)$$

其中，$x(n)$ 为输入的以帧为单位的语音信号，N 为傅里叶变换的点数，取 256；

（2）求频谱 $X(k)$ 的平方，即能量谱 $|X(k)|^2$，然后通过 Mel 频率滤波器对语音信号的频谱进行平滑，并消除谐波，凸显原先语音的共振峰；

Mel 频率滤波器是一组三角带通滤波器，中心频率为 $f(q)$，q =1，2，…，Q，Q 为三角带通滤波器的个数，Mel 滤波器 $H_q(K)$ 表示如下：

$$H_q(k)=\begin{cases}0 & k<f(q-1)\\ \dfrac{2[k-f(q-1)]}{[f(q+1)-f(q-1)][f(q)-f(q-1)]} & f(q-1)<k<f(q)\\ \dfrac{2[f(q+1)-k]}{[f(q+1)-f(q-1)][f(q+1)-f(q)]} & f(q)<k<f(q+1)\\ 0 & k>f(q+1)\end{cases} \qquad (2)$$

（3）对滤波器组输出的 Mel 频谱取对数：压缩语音频谱的动态范围；将频域中噪声的乘性成分转换成加性成分，对数 Mel 频谱 $S(q)$ 如下：

$$S(q)=\ln\left\{\sum_{k=0}^{N-1}|X(k)|^2 H_q(k)\right\} \quad (3)$$

（4）离散余弦变换（DCT）

将第3步骤获得的对数Mel频谱$S(q)$变换到时域，其结果为Mel频率倒谱系数（MFCC），第n个系数$C(n)$的计算如下式：

$$C(n)=\sum_{q=0}^{Q-1}S(q)\cos\{\frac{\pi n(q+0.5)}{Q}\},0\le n<L,0\le q<Q \quad (4)$$

其中，L为MFCC参数的阶数，Q为Mel滤波器的个数，L通常取12~16，Q取23~26，本发明依据实验情况取L=13，Q=25。

3. 根据权利要求1所述的基于高斯混合模型的声纹识别方法，其特征是模型训练时所采用的EM算法的具体步骤描述如下：

一个具有M阶混合分量的D维高斯混合模型（GMM）表示如下：

$$P(X/\lambda)=\sum_{i=1}^{M}w_i b_i(X) \quad (5)$$

式中，w_i是混合权重，$b_i(X)$是D维联合高斯概率分布，表示为：

$$b_i(X)=\frac{1}{2\pi^{D/2}|\sum_i|^{1/2}}\exp[-\frac{1}{2}(X-u_i)'\sum_i^{-1}(X-u_i)] \quad (6)$$

式中u_i是均值，$\sum_i$是协方差矩阵，完整的GMM用λ表述为：$\lambda=\{w_i,u_i,\sum_i\}$；

一组长度为T的训练矢量序列$X=\{X_1,X_2,\ldots X_T\}$的似然函数函数为$P(X|\lambda)$：

$$P(X/\lambda)=\prod_{t=1}^{T}P(X_t/\lambda) \quad (7)$$

为说话人建立GMM，就是通过EM算法训练模型的参数，实质上就是通过寻找一个模型参数λ^*，使$P(X|\lambda^*)\ge P(X|\lambda)$，然后再以新的$\lambda^*$为当前参数进行迭代，直到模型收敛为止，收敛条件$\lambda^*-\lambda<10^{-4}$，具体步骤如下：

第一步：GMM初始化：设定GMM的高斯分量的阶数M和初始模型$\lambda=\{w_i,u_i,\sum_i\}$；

第二步：E步，求期望：求解$Q(\lambda,\lambda^*)$函数

$Q(\lambda,\lambda^*)$是在X已知且给定λ^*的情况下，完成对数似然函数$\log P[(X,i)|\lambda]$对i求期望，即：

$$Q(\lambda,\lambda^*)=E\{\log P[(X,i)|\lambda]\} \quad (8)$$

整理得

$$Q(\lambda,\lambda^*)=\sum_{i=1}^{M}\sum_{t=1}^{T}\log(w_i)P(i|X_t,\lambda^*)+\sum_{i=1}^{M}\sum_{t=1}^{T}\log(P_i(X_t|\lambda_i))P(i|X_t,\lambda^*) \quad (9)$$

根据贝叶斯公式，求得训练数据在i的概率为：

$$P(i/X_t,\lambda)=\frac{w_i P_i(X_t)}{\sum_{j=1}^{M} w_j P_j(X_t)} \tag{10}$$

M 步，最大化：根据 $Q(\lambda,\lambda^*)$ 函数估计 $\lambda^*=\{w_i,u_i,\Sigma_i\}$ ；

首先计算 w_i，由于 w_i 存在约束条件 $\sum_{i=1}^{M} w_i=1$，故引入拉格朗日因子 α，并解如下方程：

$$\frac{\partial}{\partial w_i}[\sum_{i=1}^{M}\sum_{t=1}^{T}\log(w_i)P(i|X_t,\lambda^*)+\alpha(\sum_{i=1}^{M} w_i-1)]=0 \tag{11}$$

得到：

$$w_i=\frac{1}{T}\sum_{t=1}^{T}P(i|X_t,\lambda^*) \tag{12}$$

计算 u_i,Σ_i，因

$$\log[P(X|u_i,\Sigma_i)]=-\frac{D}{2}\log 2\pi-\frac{1}{2}\log|\Sigma_i|-\frac{1}{2}(X-u_i)^T\Sigma_i^{-1}(X-u_i) \tag{13}$$

上式中右边的第一项与参数 u_i,Σ_i 无关，故只需对 $Q'(\lambda,\lambda^*)$ 进行最大化：

$$Q'(\lambda,\lambda^*)=\sum_{i=1}^{M}\sum_{t=1}^{T}[-\frac{1}{2}\log|\Sigma_i|-\frac{1}{2}(X-u_i)^T\Sigma_i^{-1}(X-u_i)]P(i|X_t,\lambda^*) \tag{14}$$

对参数 u_i 求偏导可得：

$$\frac{\partial Q'(\lambda,\lambda^*)}{\partial u_i}=\sum_{t=1}^{T}\Sigma^{-1}(X_t-u_i)P(i|X_t,\lambda^*)=0 \tag{15}$$

整理得到

$$u_i=\frac{\sum_{t=1}^{T}P(i|X_t,\lambda^*)X_t}{\sum_{t=1}^{T}P(i|X_t,\lambda^*)} \tag{16}$$

对参数 Σ_i 求偏导可得

$$\frac{\partial Q'(\lambda,\lambda^*)}{\partial \Sigma_i}=\sum_{t=1}^{T}(\Sigma_i-(X_t-u_i)(X_t-u_i)^T P(i|X_t,\lambda^*)=0 \tag{17}$$

整理得到

$$\Sigma_i=\frac{\sum_{t=1}^{T}P(i|X_t,\lambda^*)(X_t-u_i)(X_t-u_i)^T}{\sum_{t=1}^{T}P(i|X_t,\lambda^*)} \tag{18}$$

第三步：EM 算法迭代 GMM

用 EM 算法迭代估计 GMM 的参数，当似然函数的值达到最大时停止迭代，即当 λ^* 值相对上次迭代时的 λ 值增幅小于设定的阈值（10^{-4}），则迭代终止，得到最终的模型参数：

混合权重 w_i^*：
$$w_i^* = \frac{1}{T}\sum_{t=1}^{T} P(i / X_t, \lambda) \tag{19}$$

均 值 u_i^*：
$$u_i^* = \frac{\sum_{t=1}^{T} P(i / X_t, \lambda) X_t}{\sum_{t=1}^{T} P(i / X_t, \lambda)} \tag{20}$$

方 差 $\sum_i^*$：
$$\sum_i^* = \frac{\sum_{t=1}^{T} P(i / X_t, \lambda)(X_t - u_i)^2}{\sum_{t=1}^{T} P(i / X_t, \lambda)} \tag{21}$$。

4. 根据权利要求 3 所述的基于高斯混合模型的声纹识别方法，其特征是在用 EM 算法训练 GMM 时，初始参数的选取采用改进的 k-means 算法，具体为：

设长度为 N 的 M 维特征矢量序列为：$X = \{X_1, X_2, \cdots, X_N\}$，其中第 $n(0 < n \le N)$ 个矢量可记为：$X_n = \{X_{n1}, X_{n2}, \ldots, X_{nM}\}$，，它可以被看作是语音信号中某一帧参数所组成的矢量；

说话人语音信号特征矢量的分布各不相同，其中第 m 维矢量的方差 S_m^2 为：

$$S_m^2 = \frac{1}{M}\sum_{m=1}^{M} (X_{nm} - \overline{X_n})^2 \tag{22}$$

式中，M 为特征矢量的维数

X_{nm} 为第 n 个矢量的第 m 维参数，$\overline{X_n}$ 为第 n 个矢量的平均值，第 m 维矢量的权值 π_m 为：

$$\pi_m = \frac{1}{S_m^2} \tag{23}$$

相应的基于方差的加权欧氏距离公式 $D(X_n, k)$ 为：

$$D(X_n, k) = \sqrt{\sum_{m=1}^{M} \pi_m (X_{nm} - C_{km})^2} \tag{24}$$

式中，X_{nm} 为待分类的特征矢量 X_n 中的第 m 个参数，C_{km} 为第 K 个类的聚类中心；

对于初始聚类中心的选取采用欧氏距离法，计算矢量集中矢量两两之间的距离，选择距离最大的两个矢量作为两个类的聚类中心，再从剩余的矢量集中选出到两个聚类中心距离最大的矢量作为另一个类的中心，如此反复直到选出 K 个聚类中心。

5. 根据权利要求 4 所述的基于高斯混合模型的声纹识别方法，其特征是改进的 K-means 聚类算法的具体步骤如下：

（1）从已有的 K 个聚类中心出发，利用公式 $D(X_n, k) = \sqrt{\sum_{m=1}^{M} \pi_m (X_{nm} - C_{km})^2}$，计算样本

集中的矢量与各个聚类中心的距离，把剩余矢量划分到离它距离最近的类中，形成初始聚类；

（2）按照步骤①的聚类，更新各个类的聚类中心；

（3）以新的聚类中心为参照点不断执行步骤①和②，直到聚类中心不再变化或变化微小时停止；

（4）得到初始 GMM 参数：

$$w_k = \frac{N_k}{T} \tag{25}$$

$$u_k = \frac{1}{N_k} \sum_{X_j \in C_k} X_j \tag{26}$$

$$\sum_{km} = \frac{1}{N_k} \sum_{X_j \in C_k} (X_{jm} - u_{km}) \tag{27}$$

其中，C_k 是第 k 个类的中心，X_j 是类 k 的第 j 个矢量，N_k 是类 k 中矢量总数。

6. 根据权利要求 2 所述的基于高斯混合模型的声纹识别方法，其特征是进行离散余弦变换时，L=13，Q=25。

7. 一种基于高斯混合模型的声纹识别系统，其特征是组成如下：

语音信号采集模块、语音信号预处理模块，语音信号特征参数提取模块，语音模型训练模块和声纹识别模块。

基于高斯混合模型的声纹识别方法及系统

技术领域

[0001] 本发明属于语音信号处理装置，涉及到一种用说话人的语音信号来辨识说话人身份的基于高斯混合模型的声纹识别方法及系统。

背景技术

[0002] 近年来，随着信息处理与人工智能技术的广泛应用，以及人们对快速有效身份验证的迫切要求，传统密码认证的身份识别已经逐渐失去了他的地位，而在生物识别领域中，基于说话人语音的身份识别技术却受到了越来越多的人的青睐。

[0003] 由于每个人的发音器官的生理差异以及后天形成的行为差异导致发音方式和说话习惯各不相同，因此用说话人的语音来识别身份成为可能。声纹识别除了具有不会遗忘、不需记忆、使用方便等优点外，还具有下列特性：首先，它的认证方式易于接受，使用的“密码”为声音，开口即得；其次，识别文本的内容可以随机，不易窃取，安全性能比较高；第三，识别使用的终端设备为麦克风或电话，成本低廉且易于和现有通信系统相结合。因此，声纹识别的应用前景非常广阔：在经济活动中，可以实现各银行的汇款、余额查询、转账等；在保密安全中，可以用指定的声音检查秘密场所的人员，其只响应特定说话人；在司法鉴定中，可以根据即时录音判断疑犯中作案者的真实身份；在生物医学中，可以使该系统只响应患者的命令，从而实现对使用者假肢的控制。

[0004] 声纹识别的关键技术主要是语音信号特征参数提取和模型匹配。语音信号特征参数大体可分为两类：一类是主要体现说话人发音器官生理特性的低层特征，如根据人耳对不同频率的语音信号的敏感程度提取的梅尔频率倒谱系数 (MFCC)，根据语音信号的全极点模型得到的线性预测倒谱系数 (LPCC) 等；另一类是主要体现说话人用语习惯、发音特点的高层特征，如反映说话人语音抑扬顿挫的韵律特征 (Prosodic Features)、反映说话人习惯用语中音素统计规律的音素特征 (Phone Features) 等。LPCC 是基于语音信号的发音模型建立的，容易受到假设模型的影响，高层特征虽然有些文献中使用，但识别率并不是很高。

[0005] 针对各种语音信号特征参数而提出的模型匹配方法主要有动态时间规整 (DTW) 法、矢量量化 (VQ) 法、高斯混合模型 (GMM) 法、人工神经网络 (ANN) 法等。其中 DTW 模型依赖于参数的时间顺序，实时性能较差，适合基于孤立字（词）的说话人识别；在 VQ 模型中，聚类的矢量仅用一个中心来表示，并且各个码本对距离的贡献相等，因此在语音信号很短的情况下，识别率会急剧下降。在 ANN 模型中，对最佳模型拓扑结构的设计的训练算法并不一定能保证收敛，而且会存在过学习的问题。GMM 是在说话人的语音信号中提取出反映说话人个性的特征参数，并以此为基础根据概率统计特性建立相应的数学模型，从而有效的反映说话人的语音信号特征参数在特征空间的分布。而且其概率密度函数比较常见，模型中的参数易于估计和训练。但是在传统基于 GMM 的声纹识别中，模型初始参数的选取比较随机，这严重影响了系统的识别率。

发明内容

[0006] 本发明要解决的技术问题是提出一种基于高斯混合模型的声纹识别方法及系统。该方法采用了基于概率统计的高斯混合模型，能很好的反映说话人的语音在特征空间的分布，其概率密度函数比较常见，模型中的参数易于估计和训练，而且具有良好识别性能和抗噪能力。

[0007] 一种基于高斯混合模型的声纹识别方法，具体步骤如下：

1、语音信号的采集：以程控交换综合实验箱的话机作为采集语音信号的终端设备，通过语音卡采集语音信号；

2、语音信号的预处理：通过计算机将提取的语音信号进行分帧加窗操作，在分帧过程中一帧包括256个采样点，帧移为128个采样点，所加的窗函数为汉明窗；端点检测，采用基于短时能量和短时过零率法相结合的端点检测法；预加重，加重系数的范围为0.90~1.00；

3、语音信号特征参数提取：采用梅尔频率倒谱系数(MFCC)，MFCC的阶数通常取为12~16；

4、模型训练：采用EM算法为说话人的语音信号特征参数训练高斯混合模型(GMM)，模型的参数初始化方法选用k-means算法；

5、声纹辨识：通过将采集到的待识别语音信号特征参数与库中通过上述步骤1、2、3、4已建立的说话人语音模型进行比较，并根据最大后验概法进行判断，若对应的说话人模型使得待识别的话者语音特征向量X具有最大的后验概率，则认为识别出说话人。

[0008] 上述的语音信号特征参数提取步骤如下：

(1) 将预处理后的语音信号进行短时傅里叶变换(DFT)得到其频谱X(k)，语音信号的DFT公式为：

$$X(k)=\sum_{n=0}^{N-1}x(n)e^{-j2\pi nk/N},0\le k<N \qquad (1)$$

其中，$x(n)$ 为输入的以帧为单位的语音信号，N为傅里叶变换的点数，取256；

(2) 求频谱 $X(k)$ 的平方，即能量谱 $|X(k)|^2$，然后通过Mel频率滤波器对语音信号的频谱进行平滑，并消除谐波，凸显原先语音的共振峰；

Mel频率滤波器是一组三角带通滤波器，中心频率为 $f(q)$，q =1,2，…,Q，Q为三角带通滤波器的个数，Mel滤波器 $H_q(K)$ 表示如下：

$$H_q(k)=\begin{cases}0 & k<f(q-1)\\ \dfrac{2[k-f(q-1)]}{[f(q+1)-f(q-1)][f(q)-f(q-1)]} & f(q-1)<k<f(q)\\ \dfrac{2[f(q+1)-k]}{[f(q+1)-f(q-1)][f(q+1)-f(q)]} & f(q)<k<f(q+1)\\ 0 & k>f(q+1)\end{cases} \qquad (2)$$

(3)对滤波器组输出的Mel频谱取对数：压缩语音频谱的动态范围；将频域中噪声的乘性成分转换成加性成分，对数Mel频谱 $S(q)$ 如下：

$$S(q)=\ln\left\{\sum_{k=0}^{N-1}|X(k)|^2 H_q(k)\right\} \quad (3)$$

(4) 离散余弦变换 (DCT)

将步骤(3)获得的对数 Mel 频谱 $S(q)$ 变换到时域,其结果为 Mel 频率倒谱系数 (MFCC),第 n 个系数 $C(n)$ 的计算如下式 :

$$C(n)=\sum_{q=0}^{Q-1}S(q)\cos\{\frac{\pi n(q+0.5)}{Q}\},0\le n<L,0\le q<Q \quad (4)$$

其中,L 为 MFCC 参数的阶数,Q 为 Mel 滤波器的个数,L 通常取 12~16,Q 取 23~26,本发明依据实验情况取 L=13,Q=25 ;

上述的模型训练时所采用的 EM 算法的具体步骤描述如下 :

一个具有 M 阶混合分量的 D 维高斯混合模型 (GMM) 表示如下 :

$$P(X/\lambda)=\sum_{i=1}^{M}w_i b_i(X) \quad (5)$$

式中,w_i 是混合权重,$b_i(X)$ 是 D 维联合高斯概率分布,表示为 :

$$b_i(X)=\frac{1}{2\pi^{D/2}|\Sigma_i|^{1/2}}\exp[-\frac{1}{2}(X-u_i)'\Sigma_i^{-1}(X-u_i)] \quad (6)$$

式中 u_i 是均值,Σ_i 是协方差矩阵,完整的 GMM 用 λ 表述为 :$\lambda=\{w_i,u_i,\Sigma_i\}$;

一组长度为 T 的训练矢量序列 $X=\{X_1,X_2,\ldots X_T\}$ 的似然函数函数为 $P(X|\lambda)$:

$$P(X/\lambda)=\prod_{t=1}^{T}P(X_t/\lambda) \quad (7)$$

为说话人建立 GMM,就是通过 EM 算法训练模型的参数,实质上就是通过寻找一个模型参数 λ^*,使 $P(X|\lambda^*)\ge P(X|\lambda)$,然后再以新的 λ^* 为当前参数进行迭代,直到模型收敛为止,收敛条件 $\lambda^*-\lambda<10^{-4}$,具体步骤如下 :

(1) GMM 初始化 :设定 GMM 的高斯分量的阶数 M 和初始模型 $\lambda=\{w_i,u_i,\Sigma_i\}$;

(2) E 步,求期望 :求解 $Q(\lambda,\lambda^*)$ 函数

$Q(\lambda,\lambda^*)$ 是在 X 已知且给定 λ^* 的情况下,完成对数似然函数 $\log P[(X,i)|\lambda]$ 对 i 求期望,即 :

$$Q(\lambda,\lambda^*)=E\{\log P[(X,i)|\lambda]\} \quad (8)$$

整理得

$$Q(\lambda,\lambda^*)=\sum_{i=1}^{M}\sum_{t=1}^{T}\log(w_i)P(i|X_t,\lambda^*)+\sum_{i=1}^{M}\sum_{t=1}^{T}\log(P_i(X_t|\lambda_i))P(i|X_t,\lambda^*) \quad (9)$$

根据贝叶斯公式,求得训练数据在 i 的概率为 :

$$P(i/X_t,\lambda)=\frac{w_i P_i(X_t)}{\sum_{j=1}^{M} w_j P_j(X_t)} \tag{10}$$

M 步，最大化：根据 $Q(\lambda,\lambda^*)$ 函数估计 $\lambda^*=\{w_i,u_i,\Sigma_i\}$；

首先计算 w_i，由于 w_i 存在约束条件 $\sum_{i=1}^{M} w_i=1$，故引入拉格朗日因子 α，并解如下方程：

$$\frac{\partial}{\partial w_i}[\sum_{i=1}^{M}\sum_{t=1}^{T}\log(w_i)P(i|X_t,\lambda^*)+\alpha(\sum_{i=1}^{M} w_i-1)]=0 \tag{11}$$

得到：

$$w_i=\frac{1}{T}\sum_{t=1}^{T}P(i|X_t,\lambda^*) \tag{12}$$

计算 u_i,Σ_i，因

$$\log[P(X|u_i,\Sigma_i)]=-\frac{D}{2}\log 2\pi-\frac{1}{2}\log|\Sigma_i|-\frac{1}{2}(X-u_i)^T\Sigma_i^{-1}(X-u_i) \tag{13}$$

上式中右边的第一项与参数 u_i,Σ_i 无关，故只需对 $Q'(\lambda,\lambda^*)$ 进行最大化：

$$Q'(\lambda,\lambda^*)=\sum_{i=1}^{M}\sum_{t=1}^{T}[-\frac{1}{2}\log|\Sigma_i|-\frac{1}{2}(X-u_i)^T\Sigma_i^{-1}(X-u_i)]P(i|X_t,\lambda^*) \tag{14}$$

对参数 u_i 求偏导可得：

$$\frac{\partial Q'(\lambda,\lambda^*)}{\partial u_i}=\sum_{t=1}^{T}\Sigma^{-1}(X_t-u_i)P(i|X_t,\lambda^*)=0 \tag{15}$$

整理得到

$$u_i=\frac{\sum_{t=1}^{T}P(i|X_t,\lambda^*)X_t}{\sum_{t=1}^{T}P(i|X_t,\lambda^*)} \tag{16}$$

对参数 Σ_i 求偏导可得

$$\frac{\partial Q'(\lambda,\lambda^*)}{\partial \Sigma_i}=\sum_{t=1}^{T}(\Sigma_i-(X_t-u_i)(X_t-u_i)^T P(i|X_t,\lambda^*)=0 \tag{17}$$

整理得到

$$\Sigma_i=\frac{\sum_{t=1}^{T}P(i|X_t,\lambda^*)(X_t-u_i)(X_t-u_i)^T}{\sum_{t=1}^{T}P(i|X_t,\lambda^*)} \tag{18}$$

（3）EM 算法迭代 GMM

用 EM 算法迭代估计 GMM 的参数，当似然函数的值达到最大时停止迭代，即当 λ^* 值相对上次迭代时的 λ 值增幅小于设定的阈值（10^{-4}），则迭代终止，得到最终的模型参数：

混合权重 w_i^*：
$$w_i^* = \frac{1}{T}\sum_{t=1}^{T} P(i / X_t, \lambda) \tag{19}$$

均值 u_i^*：
$$u_i^* = \frac{\sum_{t=1}^{T} P(i / X_t, \lambda) X_t}{\sum_{t=1}^{T} P(i / X_t, \lambda)} \tag{20}$$

方差 $\sum_i^*$：
$$\sum_i^* = \frac{\sum_{t=1}^{T} P(i / X_t, \lambda)(X_t - u_i)^2}{\sum_{t=1}^{T} P(i / X_t, \lambda)} \tag{21}$$

上述在用 EM 算法训练 GMM 时，初始参数的选取采用改进的 k-means 算法，具体为：

设长度为 N 的 M 维特征矢量序列为：$X = \{X_1, X_2, \cdots, X_N\}$，其中第 $n(0 < n \le N)$ 个矢量可记为：$X_n = \{X_{n1}, X_{n2}, \ldots, X_{nM}\}$，，它可以被看作是语音信号中某一帧参数所组成的矢量；

说话人语音信号特征矢量的分布各不相同，其中第 m 维矢量的方差 S_m^2 为：

$$S_m^2 = \frac{1}{M}\sum_{m=1}^{M} (X_{nm} - \overline{X_n})^2 \tag{22}$$

式中，M 为特征矢量的维数。X_{nm} 为第 n 个矢量的第 m 维参数，$\overline{X_n}$ 为第 n 个矢量的平均值，第 m 维矢量的权值 π_m 为：

$$\pi_m = \frac{1}{S_m^2} \tag{23}$$

相应的基于方差的加权欧氏距离公式 $D(X_n, k)$ 为：

$$D(X_n, k) = \sqrt{\sum_{m=1}^{M} \pi_m \cdot (X_{nm} - C_{km})^2} \tag{24}$$

式中，X_{nm} 为待分类的特征矢量 X_n 中的第 m 个参数，C_{km} 为第 K 个类的聚类中心；

对于初始聚类中心的选取采用欧氏距离法，计算矢量集中矢量两两之间的距离，选择距离最大的两个矢量作为两个类的聚类中心，再从剩余的矢量集中选出到两个聚类中心距离最大的矢量作为另一个类的中心，如此反复直到选出 K 个聚类中心。

[0009] 上述的改进的 K-means 聚类算法的具体步骤如下：

① 从已有的 K 个聚类中心出发，利用公式 $D(X_n, k) = \sqrt{\sum_{m=1}^{M} \pi_m \cdot (X_{nm} - C_{km})^2}$，计算样本集中的矢量与各个聚类中心的距离，把剩余矢量划分到离它距离最近的类中，形成初始聚类；

② 按照步骤①的聚类，更新各个类的聚类中心；

③ 以新的聚类中心为参照点不断执行步骤①和②，直到聚类中心不再变化或变化微小时停止；

④ 得到初始 GMM 参数：

$$w_k = \frac{N_k}{T} \tag{25}$$

$$u_k = \frac{1}{N_k} \sum_{x_j \in C_k} X_j \tag{26}$$

$$\Sigma_{km} = \frac{1}{N_k} \sum_{x_j \in C_k} (X_{jm} - u_{km}) \tag{27}$$

其中，C_k 是第 k 个类的中心，X_j 是类 k 的第 j 个矢量，N_k 是类 k 中矢量总数。

[0010] 上述进行离散余弦变换时，根据实验确定为 L=13，Q=25。

[0011] 一种基于高斯混合模型的声纹识别系统，组成如下：

语音信号采集模块、语音信号预处理模块，语音信号特征参数提取模块，语音模型训练模块和声纹识别模块。

[0012] 本发明与现有技术相比的有益效果是：

采用改进的 GMM 作为说话人语音信号的模型，通过语音卡采集语音信号，利用语音信号处理技术对采集到的语音信号进行预处理，然后提取语音信号特征参数，利用高斯混合模型对得到的语音信号特征参数建立语音模型从而构建一个说话人识别系统。采用 MFCC 参数，具有良好识别性能和抗噪能力且能充分模拟人耳感知能力；采用高斯混合模型，更具有灵活性，其概率密度函数比较常见，模型中的参数易于估计和训练，而且具有良好识别性能。

附图说明

[0013] 图 1 是本发明的系统框图；

图 2 是本发明的主流程图；

图 3 是 EM 算法训练 GMM 流程图；

图 4 是 k-means 聚类算法初始 GMM 参数流程图；

图 5 是基于高斯混合模型的声纹识别人机交互界面。

具体实施方式

[0014] 如附图 1 所示，该基于高斯混合模型的声纹识别系统，组成如下：

语音信号采集模块、语音信号预处理模块，语音信号特征参数提取模块，语音模型训练模块和声纹识别模块。

[0015] 如图 2- 图 4 所示，该基于高斯混合模型的声纹识别方法的具体步骤如下：

1、语音信号的采集

语音信号的采集是将原始的语音模拟信号转换为数字信号，设置通道号、采样频率，本发明采用杭州三汇公司生产的 SHT-8B/PCI 型语音卡进行语音信号的采集，通道号为 2（语音卡默认通道号为 2），采样频率为 8KHz（语音卡默认采样频率）。识别的终端设备为程控交换综合实验箱的电话机，且程控交换实验箱的的交换方式为空分交换，话路为甲二路（共四路：甲一路，甲二路，乙一路，乙二路，本发明随机选取甲二路，对实验结果无影响）。

[0016] 2、语音信号的预处理

（1）加窗分帧

语音信号的时变特性决定对其进行处理必须在一小段语音上进行，因此要对其进行分帧处理，同时为了保证语音信号不会因为分帧而导致信息的丢失，帧与帧之间要保证一定的重叠，即帧移，帧移与帧长的比值一般在 0~1/2 之间。本发明中使用的帧长为 256 个采样点，帧移为 128 个采样点。窗函数 $w(n)$ 采用平滑特性较好的汉明窗函数，如下所示：

$$w(n)=\begin{cases}0.54-0.46\cos(2\pi n/N-1), & 0\le n\le N-1\\ 0, & n=else\end{cases} \tag{28}$$

式中 N 为窗口长度，本发明为 256 个点。

[0017] （2）端点检测

本发明采用基于短时能量和短时平均过零率相结合的端点检测法对语音信号进行端点检测，从而判断语音信号的起始点和终止点。短时能量检测浊音，过零率检测清音。假设 $x(\cdot)$ 为语音信号，$w(\cdot)$ 为汉明窗函数，则定义短时能量 E_n 为

$$E_n=\sum_{m=-\infty}^{\infty}[x(m)w(n-m)]^2=\sum_{m=-\infty}^{\infty}x^2(m)h(n-m)=x^2(n)*h(n) \tag{29}$$

式中，$w(n-m)$ 为窗函数，$h(n)=w^2(n)$，E_n 表示语音信号的第 n 个点开始加窗函数时的短时能量。

[0018] 短时平均过零率 Z_N 为：

$$\begin{aligned}Z_N&=\frac{1}{2}\sum_{m=-\infty}^{\infty}\left|\mathrm{sgn}[x(m)-\mathrm{sgn}[x(m-1)]\right|w(n-m)\\&=\frac{1}{2}\sum_{m=n}^{n+N-1}\left|\mathrm{sgn}[x_w(w)-\mathrm{sgn}[x_w(m-1)]\right|\end{aligned} \tag{30}$$

式中，N 是窗函数长度，$\mathrm{sgn}(\cdot)$ 是符号函数，$\mathrm{sgn}[x(n)]=\begin{cases}1, & x(n)\ge 0\\ -1, & x(n)<0\end{cases}$

（3）预加重

由于语音信号的平均功率谱受到声门激励和口鼻辐射的影响，高频端大约在 8000Hz 以上按 6dB/ 倍程跌落，为此要进行预加重处理以提升语音信号的高频部分，使信号的频谱变得平坦。预加重用 6dB/ 倍程的具有提升高频特性的数字滤波器来实现，它一般是一阶的数字滤波器 $H(z)$，即

$$H(z)=1-uz^{-1} \tag{31}$$

其中 u 取值在 0.90~1.00 之间系统的识别率最高，本发明取 u=0.97。

[0019]

3、语音信号特征参数提取

语音信号特征参数提取就是从说话人的语音信号中提取出能够反映说话人个性的参数，具体过程如下：

(1)将预处理后的语音信号进行短时傅里叶变换(DFT)得到其频谱 X(k)。语音信号的 DFT 公式为：

$$X(k)=\sum_{k=0}^{N-1}x(n)e^{-j2\pi k/N},0\le k<N \tag{32}$$

其中，$x(n)$ 为输入的以帧为单位的语音信号，N 为傅里叶变换的点数，取 256。

[0020] (2)求频谱 $X(k)$ 的平方，即能量谱 $|X(k)|^2$，然后将它们通过 Mel 滤波器，以实现对语音信号的频谱进行平滑，并消除谐波，凸显原先语音的共振峰。

[0021] Mel 频率滤波器是一组三角带通滤波器，中心频率为 $f(q)$，q =1，2，…，Q，Q 为三角带通滤波器的个数，Mel 滤波器 $H_q(K)$ 表示如下：

$$H_q(k)=\begin{cases}0 & k<f(q-1)\\ \dfrac{2[k-f(q-1)]}{[f(q+1)-f(q-1)][f(q)-f(q-1)]} & f(q-1)<k<f(q)\\ \dfrac{2[f(q+1)-k]}{[f(q+1)-f(q-1)][f(q+1)-f(q)]} & f(q)<k<f(q+1)\\ 0 & k>f(q+1)\end{cases} \tag{33}$$

(3)对滤波器组的输出取对数：压缩语音频谱的动态范围；将频域中的噪声的乘性成分转换成加性成分，得到的对数 Mel 频谱 $S(q)$ 如下：

$$S(q)=\ln\left\{\sum_{k=0}^{N-1}|X(k)|^2H_q(k)\right\} \tag{34}$$

(4)离散余弦变换 (DCT)

将步骤(3)获得的 Mel 频谱 $S(q)$ 变换到时域，其结果就是 Mel 频率倒谱系数 (MFCC)。第 n 个系数 $C(n)$ 的计算如下式：

$$C(n)=\sum_{q=0}^{Q-1}S(q)\cos\left\{\frac{\pi n(q+0.5)}{Q}\right\},0\le n<L,0\le q<Q \tag{35}$$

其中，L 为 MFCC 的阶数，Q 为 Mel 滤波器的个数，二者取值常依据实验情况来定。本发明取 L=13，Q=25。

[0022] 4、模型训练

(1)基本原理

基于高斯混合模型(GMM)的声纹识别的基本原理是为说话人集合中的每一个说话人的

语音信号建立一个模型，模型的参数由说话人语音信号特征参数的空间分布决定。不同说话人的语音信号特征参数的统计分布不同，因此通过比较不同的说话人的 GMM，就可以判别出不同的说话人。

[0023] GMM 本质上是一种多维概率密度函数的线性加权组合。一个具有 M 阶混合分量的 D 维 GMM 表示如下：

$$P(X/\lambda)=\sum_{i=1}^{M}w_i b_i(X) \tag{36}$$

式中，w_i 是混合权重，代表每个高斯分布的幅度大小，且 $\sum_{i=1}^{M}w_i=1$。$b_i(X)$ 是 D 维的联合高斯概率分布，表示为：

$$b_i(X)=\frac{1}{2\pi^{D/2}|\Sigma_i|^{1/2}}\exp[-\frac{1}{2}(X-u_i)'\Sigma_i^{-1}(X-u_i)] \tag{37}$$

上式中 u_i 是均值，代表每个高斯分布的位置，Σ_i 是协方差矩阵，代表高斯分布的范围。完整的 GMM 用 λ 表述为：$\lambda=\{w_i,u_i,\Sigma_i\}$。在 GMM 中，协方差矩阵 Σ_i 是对角矩阵。

[0024] （2）模型训练

在发明中，为说话人建立 GMM，实际上就是通过 EM 算法训练估计模型的参数。对于一组长度为 T 的训练矢量序列 $X=\{X_1,X_2,\ldots X_T\}$，它的似然函数为 $P(X|\lambda)$：

$$P(X/\lambda)=\prod_{t=1}^{T}P(X_t/\lambda) \tag{38}$$

训练的目的就是找到一组模型参数 λ^*，使 $P(X|\lambda^*)\geq P(X|\lambda)$，然后再以新的 λ^* 为当前参数进行迭代，直到模型收敛为止（收敛条件 $\lambda^*-\lambda<10^{-4}$）。具体步骤如下：

1）GMM 初始化：设定 GMM 的高斯分量的阶数 M 和初始模型 $\lambda=\{w_i,u_i,\Sigma_i\}$；

2）E 步，求期望：求解 $Q(\lambda,\lambda^*)$ 函数

$Q(\lambda,\lambda^*)$ 是在 X 已知且给定 λ^* 的情况下，完成对数似然函数 $\log P[(X,i)|\lambda]$ 对 i 求期望，即：

$$Q(\lambda,\lambda^*)=E\{\log P[(X,i)|\lambda]\} \tag{39}$$

整理得

$$Q(\lambda,\lambda^*)=\sum_{i=1}^{M}\sum_{t=1}^{T}\log(w_i)P(i|X_t,\lambda^*)+\sum_{i=1}^{M}\sum_{t=1}^{T}\log(P_i(X_t|\lambda_i))P(i|X_t,\lambda^*) \tag{40}$$

根据贝叶斯公式，求得训练数据在 i 的概率为：

$$P(i/X_t,\lambda)=\frac{w_i P_i(X_t)}{\sum_{j=1}^{M}w_j P_j(X_t)} \tag{41}$$

M 步，最大化：根据 $Q(\lambda,\lambda^*)$ 函数估计 $\lambda^*=\{w_i,u_i,\Sigma_i\}$；

首先计算 w_i，由于 w_i 存在约束条件 $\sum_{i=1}^{M} w_i = 1$，故引入拉格朗日因子 α，并解如下方程：

$$\frac{\partial}{\partial w_i}[\sum_{i=1}^{M}\sum_{t=1}^{T}\log(w_i)P(i|X_t,\lambda^*)+\alpha(\sum_{i=1}^{M} w_i - 1)] = 0 \quad (42)$$

得到：

$$w_i = \frac{1}{T}\sum_{t=1}^{T} P(i|X_t,\lambda^*) \quad (43)$$

计算 u_i, Σ_i，因

$$\log[P(X|u_i,\Sigma_i)] = -\frac{D}{2}\log 2\pi - \frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(X-u_i)^T\Sigma_i^{-1}(X-u_i) \quad (44)$$

上式中右边的第一项与参数 u_i, Σ_i 无关，故只需对 $Q'(\lambda,\lambda^*)$ 进行最大化：

$$Q'(\lambda,\lambda^*) = \sum_{i=1}^{M}\sum_{t=1}^{T}[-\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(X-u_i)^T\Sigma_i^{-1}(X-u_i)]P(i|X_t,\lambda^*) \quad (45)$$

对参数 u_i 求偏导可得：

$$\frac{\partial Q'(\lambda,\lambda^*)}{\partial u_i} = \sum_{t=1}^{T}\Sigma^{-1}(X_t-u_i)P(i|X_t,\lambda^*) = 0 \quad (46)$$

整理后得到

$$u_i = \frac{\sum_{t=1}^{T} P(i|X_t,\lambda^*)X_t}{\sum_{t=1}^{T} P(i|X_t,\lambda^*)} \quad (47)$$

对参数 Σ_i 求偏导可得

$$\frac{\partial Q'(\lambda,\lambda^*)}{\partial \Sigma_i} = \sum_{t=1}^{T}(\Sigma_i - (X_t-u_i)(X_t-u_i)^T P(i|X_t,\lambda^*) = 0 \quad (48)$$

整理后得到

$$\Sigma_i = \frac{\sum_{t=1}^{T} P(i|X_t,\lambda^*)(X_t-u_i)(X_t-u_i)^T}{\sum_{t=1}^{T} P(i|X_t,\lambda^*)} \quad (49)$$

3）EM 算法迭代估计 GMM

用 EM 算法反复迭代估计 GMM 的参数，当似然函数的值达到最大时迭代停止，即当 λ^* 值相对上次迭代时的 λ 值增幅小于设定的阈值（10^{-4}），则迭代终止，得到最终的模型参数：

混合权重 w_i^*：

$$w_i^* = \frac{1}{T}\sum_{t=1}^{T} P(i/X_t,\lambda) \quad (50)$$

均 值 u_i^*：

$$u_i^* = \frac{\sum_{t=1}^{T} P(i/X_t,\lambda)X_t}{\sum_{t=1}^{T} P(i/X_t,\lambda)} \tag{51}$$

方 差 $\sum_i^*$：

$$\sum_i^* = \frac{\sum_{t=1}^{T} P(i/X_t,\lambda)(X_t-u_i)^2}{\sum_{t=1}^{T} P(i/X_t,\lambda)} \tag{52}$$

EM 算法训练 GMM 的流程图见附图 2。

[0025] 在用 EM 算法训练 GMM 时，初始参数的选取本发明采用改进的 k-means 算法。

[0026] 设长度为 N 的 M 维特征矢量序列为：$X=\{X_1,X_2,\cdots,X_N\}$，其中第 $n(0<n\le N)$ 个矢量可记为：$X_n=\{X_{n1},X_{n2},\ldots,X_{nM}\}$，它可以被看作是语音信号中某一帧参数所组成的矢量。

[0027] 说话人语音信号特征矢量的分布各不相同，其中第 m 维矢量的方差 S_m^2 为：

$$S_m^2 = \frac{1}{M}\sum_{m=1}^{M}(X_{nm}-\overline{X_n})^2 \tag{53}$$

式中，M 为特征矢量的维数。X_{nm} 为第 n 个矢量的第 m 维参数，$\overline{X_n}$ 为第 n 个矢量的平均值，第 m 维矢量的权值 π_m 为：

$$\pi_m = \frac{1}{S_m^2} \tag{54}$$

相应的基于方差的加权欧氏距离公式 $D(X_n,k)$ 为：

$$D(X_n,k)=\sqrt{\sum_{m=1}^{M}\pi_m\,(X_{nm}-C_{km})^2} \tag{55}$$

式中，X_{nm} 为待分类的特征矢量 X_n 中的第 m 维参数，C_{km} 为第 K 个类的聚类中心。

[0028] 对于初始聚类中心的选取采用欧氏距离法，计算矢量集中对象两两之间的距离，选择距离最大的两个矢量作为两个类的聚类中心，再从剩余的矢量集中选出到两个聚类中心距离最大的矢量作为另一个类的中心，如此反复直到选出 K 个聚类中心。

[0029] 改进的 K-means 聚类算法的具体步骤如下：

① 从已有的 K 个聚类中心出发，利用公式 $D(X_n,k)=\sqrt{\sum_{m=1}^{M}\pi_m\,(X_{nm}-C_{km})^2}$，计算样本集中的矢量与各个聚类中心的距离，把剩余样本矢量划分到离它距离最近的类中，形成初始聚类；

② 按照步骤①的聚类，更新各个类的聚类中心；

③ 以新的聚类中心为参照点不断执行步骤②和③，直到聚类中心不再变化或变化微小时停止；

④ 得到初始 GMM 参数：

$$w_k = \frac{N_k}{T} \tag{56}$$

$$u_k = \frac{1}{N_k} \sum_{X_j \in C_k} X_j \tag{57}$$

$$\Sigma_{km} = \frac{1}{N_k} \sum_{X_j \in C_k} (X_{jm} - u_{km}) \tag{58}$$

其中，C_k 是第 k 个类的中心，X_j 是类 k 的第 j 个矢量，N_k 是类 k 中矢量总数。

[0030] k-means 聚类算法初始化 GMM 参数的流程图见附图 3。

[0031] 5、声纹辨识

对于一个声纹识别系统，若有 N 个说话人，其对应的 M 阶的 GMM 分别为 $\lambda_1, \lambda_2, \ldots, \lambda_N$。在辨识阶段，给定一个待识别的语音样本的特征矢量序列 $X = \{X_1, X_2, \ldots, X_T\}$，则这段语音属于第 n 个说话人的最大后验概率为：

$$P(\lambda_n | X) = \frac{P(X | \lambda_n) P(\lambda_n)}{P(X)} = \frac{P(X | \lambda_n) P(\lambda_n)}{\sum_{i=1}^{M} P(X | \lambda_i) P(\lambda_i)} \tag{59}$$

式中 $P(X)$ 为所有说话人条件下特征矢量序列 X 的概率密度，$P(X | \lambda_n)$ 为特征矢量序列 X 是第 n 个人产生的条件概率，且有

$$P(X | \lambda_n) = \prod_{t=1}^{T} P(X_t | \lambda_n) \tag{60}$$

$P(\lambda_n)$ 为第 n 个人说话的先验概率，假定该语音信号出自封闭集里的每个人的可能性相等，则有：

$$P(\lambda_n) = \frac{1}{N} \quad 1 \le n \le N \tag{61}$$

对于一个确定的观察矢量序列 X，$P(X)$ 是一个确定的常数值，对所有的话者来说都相等，因此求取后验概率的最大值可以通过求取 $P(X | \lambda_n)$ 获得，识别结果为：

$$n^* = \arg\max P(X | \lambda_n) \tag{62}$$

n^* 为识别出的说话人，即判决结果。

[0032] 在实际应用中，常采用对数似然函数：

$$L(X|\lambda_n)=\log P(X|\lambda_n)=\sum_{t=1}^{T}\log P(X_t|\lambda_n) \quad (63)$$

因此最终的识别结果为：

$$n^*=\arg\max L(X|\lambda_n) \quad (64)$$

本系统属于闭集识别，也就是说所有待识别的说话人都属于已知的说话人集合。说话人识别的人机交互界面如附图 4 所示。在声纹识别系统的人机交互界面中，“语音卡状态显示”列表视图显示当前语音卡可用的语音通道号及通道状态；“语音样本库”列表视图显示当前语音样本库中的说话人样本数目及说话人姓名。“声纹识别参数设置”一栏显示语音采集所要设置的参数，包括：训练时长（默认 23s），测试时长（默认 15s）以及候选人个数（默认 1）。

[0033] 如图 5 所示，以下结合实例进行具体说明：假设语音样本库中预先存了 100 个人的语音，当张 XX 拨通电话时，其声音如何识别的过程。

[0034] 1、若张 XX 不属于已知的语音样本库

（1）语音信号的采集：以程控交换综合实验箱的话机作为采集语音的终端设备，通过语音卡采集语音；

首先，设置需要采集的训练语音的“训练时长”参数（范围：10-39s），然后在姓名编辑框中添加说话人的姓名“张 XX”，点击“添加说话人”按钮。添加完成后点击“确定”，然后拨通程控交换综合实验箱的电话（号码：8700），接通后，语音卡通道 2（默认为通道 2）的状态更新为“录音中”，此时语音卡就可以进行采集语音。采集的语音达到预定的训练时长，电话会自动挂断；

（2）语音信号的预处理：通过计算机和 VC 软件结合对提取的语音信号进行分帧加窗操作，在分帧过程中一帧包括 256 个采样点，帧移为 128 个采样点，所加的窗函数为汉明窗；端点检测，采用基于短时能量和短时过零率法相结合的检测法；预加重，加重系数的值为 0.97；

（3）提取特征参数：利用计算机与 VC 软件结合提取 13 阶的 MFCC 参数；

（4）模型训练：选用 k-means 算法对模型参数进行初始化，然后采用 EM 算法为说话人的语音信号特征参数训练高斯混合模型（GMM）；

（5）说话人识别

首先，设置需要采集的测试语音的“测试时长”参数（范围：5-20s），拨通程控交换综合实验箱的电话（号码：8700），利用语音卡 （通道为 2）采集语音。采集的语音达到预定的测试时长，电话会自动挂断；

然后软件禁止“进行说话人辨识”按钮使用，对说话人的语音进行步骤（2）、（3）的操作，最后将提取的待测试的说话人的语音与库中的语音模型进行比较，点击 “进行说话人辨识”按钮，选择要显示的候选人数（范围 1-3），若对应的说话人模型使得待识别的话者语音特征向量 X 具有最大的后验概率，则认为识别出说话人，同时在“说话人辨识”视图列表上显示辨识结果“张 XX”和识别度。

[0035] 2、若张 XX 属于已知的语音样本库

若张 XX 属于已知的语音样本库则直接进行说话人辨识：首先，设置需要采集的测试语

音的“测试时长”参数（范围：5-20s），拨通程控交换综合实验箱的电话（号码：8700），利用语音卡（通道为 2）采集语音。采集的语音达到预定的测试时长，电话会自动挂断；

然后软件禁止“进行说话人辨识”按钮使用，对说话人的语音进行步骤（2）、（3）的操作，最后将提取的待测试的说话人的语音与库中的语音模型进行比较，点击“进行说话人辨识”按钮，选择要显示的候选人数（范围 1-3），若对应的说话人模型使得待识别的话者语音特征向量 X 具有最大的后验概率，则认为识别出说话人，同时在“说话人辨识”视图列表上显示辨识结果“张 XX”和识别度。

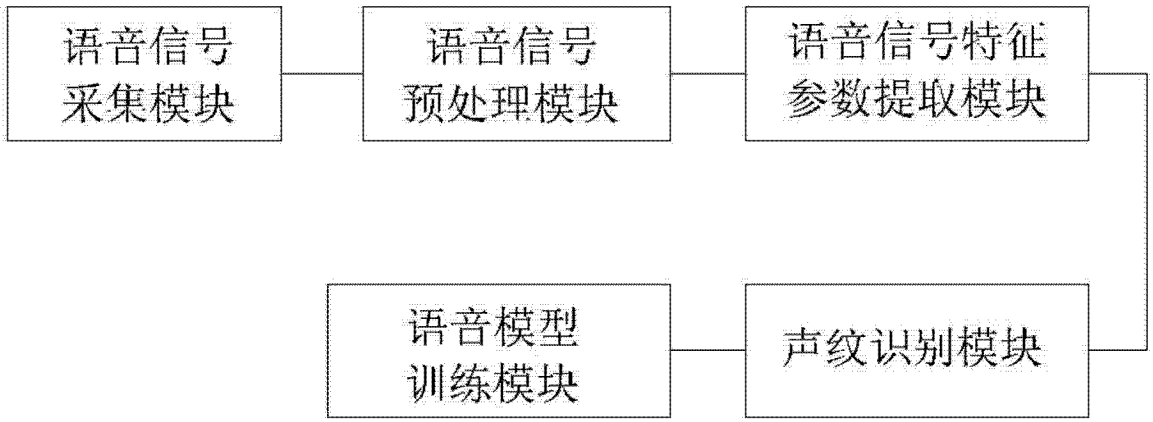


图 1

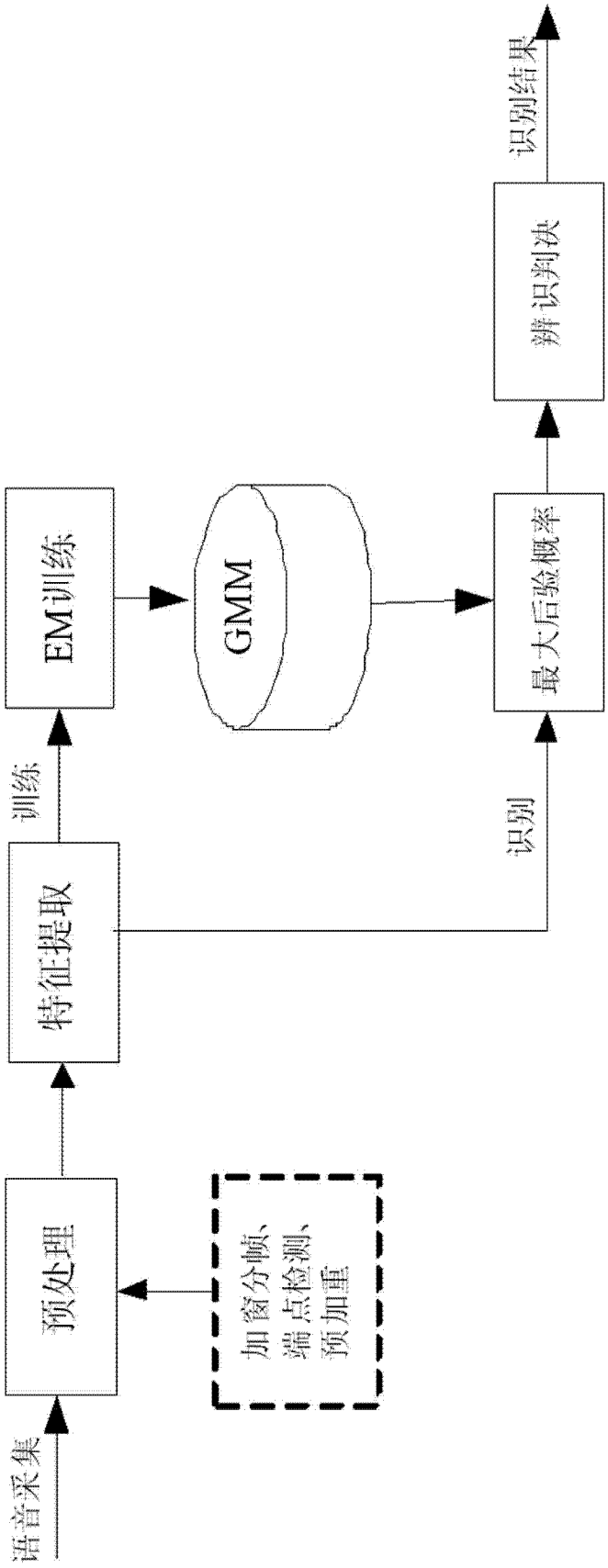


图 2

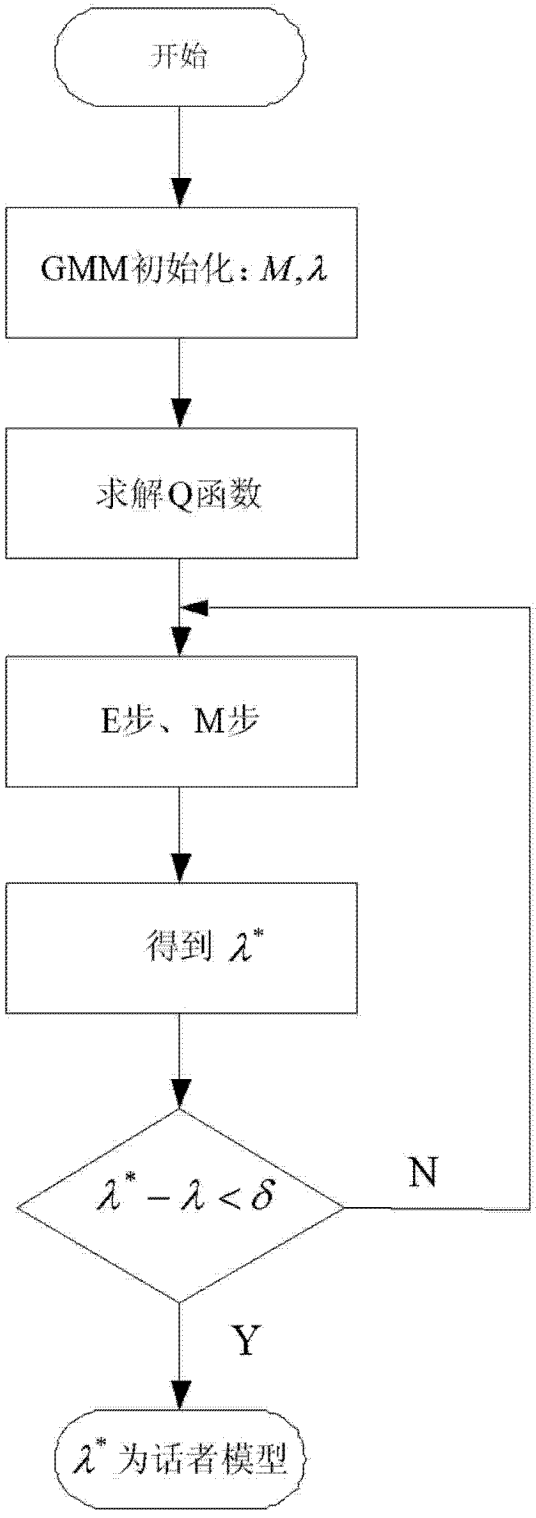

开始
GMM初始化：M,λ
求解Q函数
E步、M步
得到 λ^*
$\lambda^* - \lambda < \delta$
N
Y
λ^* 为话者模型

图 3

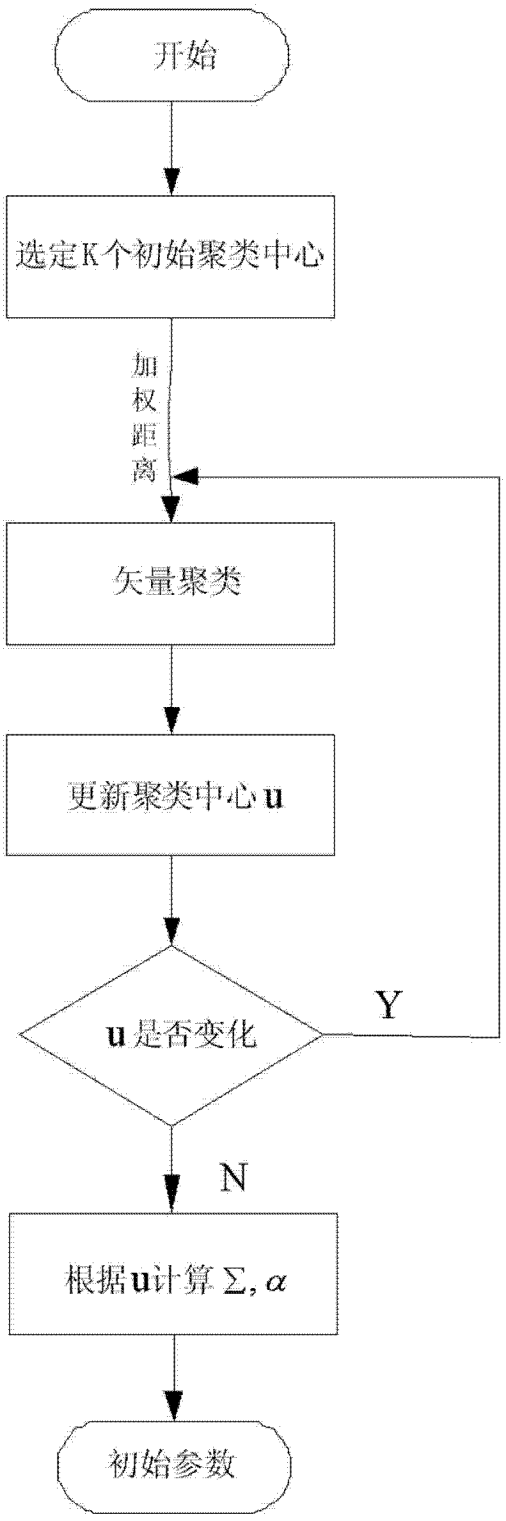

开始
选定K个初始聚类中心
加权距离
矢量聚类
更新聚类中心 u
u 是否变化
Y
N
根据u计算 Σ, α
初始参数

图 4

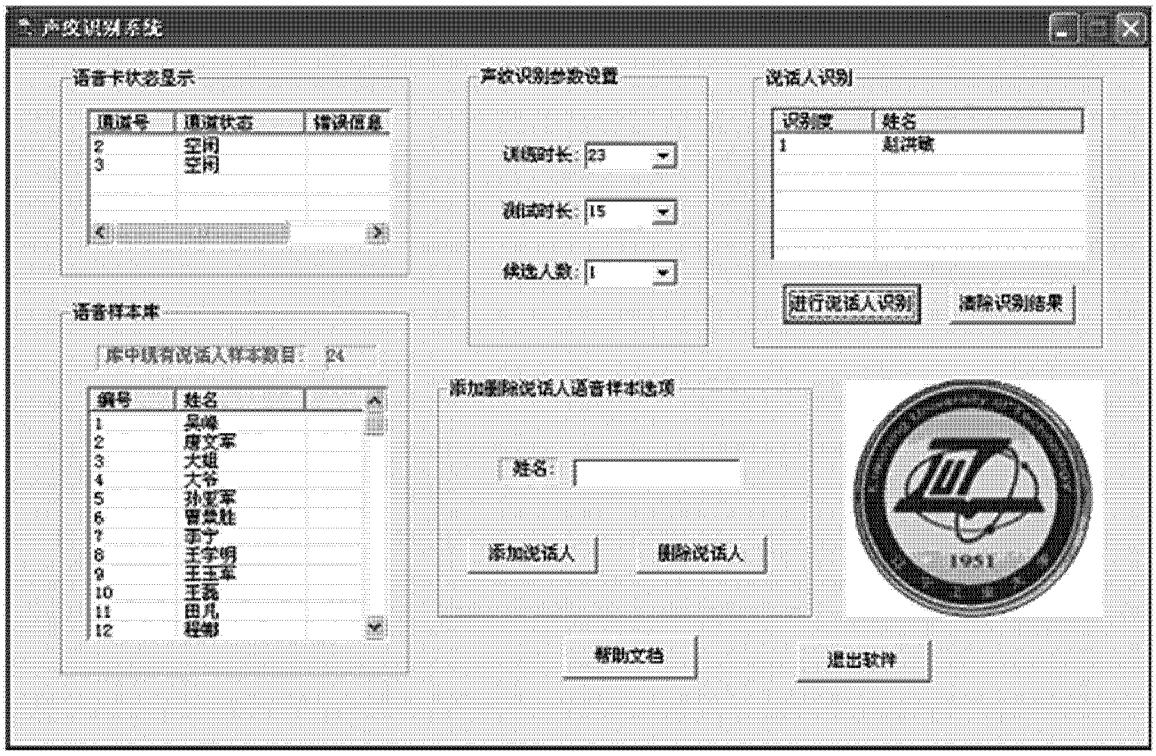

声纹识别系统
语音卡状态显示
通道号
通道状态
错误信息
2
空闲
3
空闲
语音样本库
库中现有说话人样本数目
24
编号
姓名
1
吴峰
2
康文军
3
大姐
4
大爷
5
孙军军
6
曹楚魁
7
李宁
8
王宇明
9
王玉军
10
王磊
11
田凡
12
程娜
声纹识别参数设置
训练时长:
23
测试时长:
15
候选人数:
1
添加删除说话人语音样本选项
姓名:
添加说话人
删除说话人
帮助文档
说话人识别
识别度
姓名
1
赵洪斌
进行说话人识别
清除识别结果
退出软件

图 5