ISSN 1009-3044
Computer Knowledge and Technology 电脑知识与技术
Vol.5 No.9, March 2009

基于GMM模型的说话人辨认系统

谢青松，潘 进，史永林，李国朋
（西安通信学院，陕西 西安 710106）

摘要：利用MATLAB软件，设计了一种基于GMM模型的与文本无关的说话人辨认系统。该系统包括语音活动检测、提取MFCC参数、训练GMM参数和判决辨认四部分。经过TIMIT数据库测试，该系统的性能良好。

关键词：说话人辨认；美尔频率倒谱系数；高斯混合模型

中图分类号：TN912　　**文献标识码**：A　　**文章编号**：1009-3044(2009)08-2186-02

A Speaker Identification System Based on GMM

XIE Qing-song, PAN Jin, SHI Yong-lin, LI Guo-peng
(Xi'an Communication Institute, Xi'an 710106, China)

Abstract: A text-independent speaker identification system based on GMM is designed by using MATLAB software. This system includes four parts: voice activity detection, abstracting MFCC parameters, training GMM parameters and identifying. The experiment based on TIMIT database shows that the performance of this system is good.

Key words: speaker identification; mel frequency cepstral coefficient; gaussian mixture model

1 引言

说话人识别是指通过对说话人语音信号的分析处理，自动确认说话人是否在所记录的话者集合中，并确定说话人是谁的过程。说话人识别技术按其识别任务可以分为两类：说话人辨认和说话人确认。前者用以判断某段语音是若干人中的哪一个人所说，是多者选一的问题；而后者用于确定某段语音是否是声称的某个说话人所说，是二选一的判定问题。根据识别对象的不同，说话人识别分为三类：文本有关、文本无关和文本提示型。其中，与文本无关的识别方法是当前说话人识别技术的研究重点。从现有的文献来看，在与文本无关的说话人辨认系统中，高斯混合模型（GMM）的性能最好[1]。本文在MATLAB环境下，设计了一个基于GMM模型的说话人辨认系统，实验结果验证该方法的有效性。

2 说话人辨认的系统结构

基于GMM的说话人辨认系统由4部分组成：语音活动检测（Voice Activity Detection，VAD）、提取MFCC参数、训练GMM参数和计算后验概率并判决辨认，如图1所示。

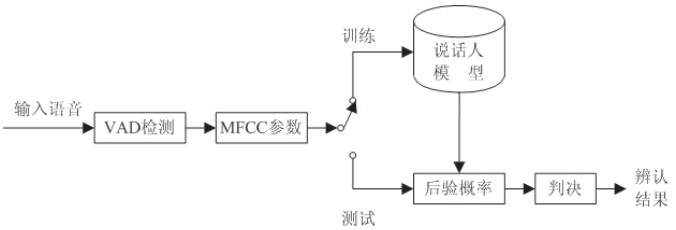

图1　说话人辨认系统框图

2.1 语音活动检测

语音活动检测用于去掉语音信号中的静音段，避免静音段对说话人辨认结果的影响。由于静音段的能量比有声段的能量小很多，可以借助信号的短时能量，从语音信号中检测出有声段和静音段。

VAD检测时，设定帧长为5ms，帧移为2.5ms，以每帧信号的均方根值（Root Mean Square，RMS）作为检测参数。当一帧信号的RMS值超过阈值0.03时（输入信号归一化后）判为有声段，否则判为静音段。在此基础上，进行VAD判决纠正。判决纠正的规则为：有声段的最小长度为20ms，有声段之间的最小距离为150ms。分别用于将静音段中高于阈值的孤立噪声段设定为静音段，将有声段之间低于阈值的清音段纠正为有声段。

2.2 提取MFCC参数

特征参数的提取是说话人辨认系统的重要组成部分。美尔频率倒谱系数（Mel Frequency Cepstral Coefficient，MFCC）是目前在说话人辨认领域使用最为广泛的一种特征参数。MFCC参数的计算过程如下[2]：

1）将语音信号中的有声段分帧，对每帧语音信号s(n)加Hamming窗，得到加窗后的语音x(n)；

2）对加窗语音x(n)进行离散傅里叶变换，取模的平方得到离散功率谱X(n)；

3）计算X(n)通过M个Mel滤波器后所得的功率谱值，即计算X(n)和Mel滤波器组的传递函数在各离散频率点上乘积之和，得到M个参数P_m,0≤m≤M；

4）计算P_m的自然对数，并进行离散余弦变换，即可得到MFCC参数。

2.3 训练GMM参数

高斯混合模型利用多维概率密度函数对语音信号进行建模，为每个说话人的语音建立一个GMM模型。在一个具有M个混合分量的K维GMM中，一个K维声学特征矢量o在该GMM下的概率密度函数为：

$$P(o \mid \lambda)=\sum_{i=1}^{M} P(o,i \mid \lambda)=\sum_{i=1}^{M} c_i P(o \mid i,\lambda) \tag{1}$$

其中，λ为GMM模型的参数集；i为隐状态号，也就是高斯分量的序号；C_i为第i个分量的混合权值，即隐状态i的先验概率，满足$\sum_{i=1}^{M} c_i = 1$

P(o|i,λ)为高斯混合分量，是P(o|q=i,λ)的简写形式，对应声学特征矢量o在隐状态i下的观察概率密度函数，一般用K维单高斯分布函数表示：

收稿日期：2009-01-10
作者简介：谢青松（1982-），男，四川苍溪人，助教，硕士，研究方向：语音信号处理。

$$P(o|i,\lambda)=N(o,\mu_i,\Sigma_i)=\frac{1}{(2\pi)^{K/2}|\Sigma_i|^{1/2}}\exp[-\frac{(o-\mu_i)^T\Sigma_i^{-1}(o-\mu_i)}{2}] \quad (2)$$

其中，μ_i为均值矢量；Σ_i为协方差矩阵。 因此式（1）可以理解为，M阶GMM是用M个单高斯分布的线性组合来描述的。即GMM参数集λ可由各混合分量的权重，均值和协方差矩阵组成，表示为如下三元组的形式：$\lambda=\{C_i,\mu_i,\Sigma_i\},i=1,\cdots,M$。当协方差矩阵$\Sigma i=diag\{\sigma^2_{i0},\sigma^2_{i1},\cdots,\sigma^2_{iK-1}\}$时算法简单，并且性能也很好，此时：

$$P(o|i,\lambda)=\prod_{k=0}^{K-1}\frac{1}{\sqrt{2\pi}\sigma_{ik}}\exp[-\frac{(o_k-\mu_{ik})^2}{2\sigma_{ik}^2}] \quad (3)$$

其中，o_k和μ_{ik}分别为矢量o和矢量μ_i的第k个分量，σ^2_{ik}(k=0,1,…,K-1)为GMM第i个分量所对应的特征矢量的第k维分量的方差。

假设可用的训练特征矢量序列为$O\{o_1,o_2,\cdots,o_T\}$，则高斯混合模型的似然函数可以表示为 $P(O|\lambda)=\prod_{t=1}^{T}P(o_t|\lambda)$ 。训练的目的就是找到一组参数λ，使似然概率$P(O|\lambda)$最大。训练时，首先采用K均值聚类算法初始化GMM参数，然后采用EM算法[3,4]通过迭代估计新的GMM参数。设训练数据落在假定的隐状态 的概率可以表为 $P(q_t=i|o_t,\lambda)=\frac{c_iP(o_t|i,\lambda)}{P(o_t|\lambda)}$ ，则迭代公式如下：

$$\overline{c_i}=\frac{1}{T}\sum_{i=1}^{T}P(q_t=i|o_t,\lambda) \quad (4)$$

$$\overline{\mu_i}=\frac{\sum_{i=1}^{T}P(q_t=i|o_t,\lambda)o_t}{\sum_{i=1}^{T}P(q_t=i|o_t,\lambda)} \quad (5)$$

$$\overline{\sigma_{ik}^2}=\frac{\sum_{i=1}^{T}P(q_t=i|o_t,\lambda)(o_{ik}-\overline{\mu_{ik}})^2}{\sum_{i=1}^{T}P(q_t=i|o_t,\lambda)} \quad (6)$$

2.4 说话者辨认

对于有N个说话人的说话者辨认系统，每个说话人用一个GMM模型来代表，记为λ_1，$\lambda_2,\cdots,\lambda_N$。在辨认说话人时，计算测试语音的特征矢量序列对于每个GMM模型的对数似然度得分，得分最高的模型对应的说话人即为测试序列的说话人。即识别目标函数为：

$$n^*=\arg\max_{1\le n\le N}\ln P(O|\lambda_n)=\arg\max_{1\le n\le N}\sum_{t=1}^{T}\ln P(o_t|\lambda_n) \quad (7)$$

3 实验结果

实验所用的语音取自TIMIT数据库，从该数据库中随机抽取20个人的数据。每个人有10条语音，8条用于训练GMM参数，2条用于测试。在计算MFCC参数时，设定帧长为20ms，帧移为20ms；Mel滤波器的个数为24，分别去掉第0、1、2、22、23阶分量，得到19阶的MFCC参数。

本系统采用的性能参数如下[4]：设待测试的特征矢量序列为$O=\{o_1,o_2,\cdots,o_L\}$，每次测试时从左至右依次取T个分量，每次抽取的间隔为1个分量，则总的测试特征矢量个数为L-T+1。若正确辨认的特征矢量个数为CRN，则辨认率SIR为：

$$SIR=\frac{CRN}{L-T+1}\times 100\% \quad (8)$$

在基于GMM的说话者辨认系统中，影响辨认率的主要参数有高斯混合分量的个数M和测试特征矢量的分量个数T。为此，本文在M与T取不同值的情况下进行了测试。

表1给出了在M与T取不同值时的辨认率。从表1中可以看出：当固定M时，辨认率随T的增加而增加；当T一定时，辨认率随M的增加而提高；在参数M=32、T=200时，辨认率达到最大值。

表1 基于GMM的说话人辨认系统的辨认率

M \ T	50	100	150	200
2	40.48%	52.26%	68.65%	76.47%
4	52.72%	69.50%	75.09%	84.05%
8	60.26%	75.11%	78.85%	89.31%
16	67.12%	85.58%	90.60%	93.82%
32	74.89%	90.07%	94.62%	96.92%

4 结束语

本文在MATLAB环境下，实现了一个基于GMM模型的说话人辨认系统。该系统以MFCC参数为特征矢量，为每个说话者建立一个GMM参数集，实验结果表明该方法的有效性。对于含有20个说话人的测试集，当参数M=32、T=200时，辨认率可达96.92%。

参考文献：

[1] Reynolds D A. An overview of automatic speaker recognition technology[C].IEEE International conference on acoustics speech and signal processing,2002(4):4072-4075.
[2] 何强,何英.MTLAB扩展编程[M].北京:清华大学出版社,2002:336-338.
[3] 韩纪庆,张磊,郑铁然.语音信号处理[M].北京:清华大学出版社,2004:273-277.
[4] Reynolds D A, Rose R C. Robust text independent speaker identification using guassian mixture speaker models[J].IEEE Trans on speech and audio processing,1995,3(1):72-83.