文章编号:1001 - 9081(2001)04 - 0007 - 02

# 基于高斯混合模型的说话人确认系统

杨澄宇,赵 文,杨 鉴
(云南大学 信息与电子科学系,云南 昆明 650091)

摘 要:由于在人的话音频谱中,低频和较高频段含有较多说话人的个性信息,本文提出一种 LPC 倒谱的改进算法用于与文本无关的说话人识别。该改进算法通过话音频谱的各频段进行加权,突出说话人的个性信息,从而使说话人更易于区分。

关键词:说话人识别;LPC 谱加权倒谱;混合高斯模型
中图分类号:TP391.42 文献标识码:A

# SPEAKER IDENTIFICATION SYSTEM BASED ON GAUSSIAN MIXTURE MODELING

YANG Cheng-yu, ZHAO Wen, YANG Jian
*(Department of Information and Electronics, Yunnan University, Kunming Yunnan 650091, China)*

**Abstract:** We proceed a text - independent speaker identification system using Gaussian Mixture modeling. Since the lower and higher portion of the spectrum contain more reliable idiosyncrasy information on speaker than the middle portion, this paper presents a novel LPC ceptral algorithm for speaker identification task. By weighting the different portion in frequency domain, it can makes idiosyncrasy information be prominent and speaker recognition be easier.

**Key words:** speaker identification; LPC spectrum weighting ceptral; GMM

## 1 引言

说话人识别研究是传统语音识别的一个分支,可用于安全系统、多媒体数据库检索等领域。它又可分为说话人辨识(verification)和说话人确认(identification)两种不同的处理方法。给定一个说话人集合 S 和未知说话人 X 的一段话音,说话人辨识是指先判断 X 是否是 S 中的元素,如果是的话,再判断 X 是哪个元素。如果预先已知 X 属于 S,只需判断 X 是 S 中的哪个人,则称为说话人确认。另外,有的说话人识别系统通过规定训练和测试文本的内容来提高识别率,称为以文本有关的说话人识别。如果对文本内容不做要求,则称为以文本无关的说话人识别系统[1]。本文讨论的是与文本无关的说话人确认。

说话人识别系统大致包括两个主要部分,即话音特征参数提取部分和识别算法部分。本文用改进的 LPC 倒谱作特征参数,用概率模型—混合高斯模型(GMM)作识别算法。

混合高斯模型是一个状态的 *CHMM*,该模型用多个高斯分布的概率密度函数的组合来描述矢量在概率空间的分布情况。在说话人识别系统中,用混合高斯模型的参数描述说话人语音的特征矢量的概率分布。不同说话人将对应不同的混合高斯模型参数。

常用于说话人识别的语音特征参数主要有 LPC 倒谱和 Mel 倒谱,两者都属倒谱系数。在所有声道的音段参数中,基于 LPC 的倒谱系数最能有效反映说话人特征[2],本文将使用 LPC 倒谱。由于在人的语音频谱中,低频段和高频段含有较多的说话人个性信息[3],所以本文考虑在求 LPC 倒谱的过程中,对频谱各段加权,获得"LPC 谱加权的倒谱"参数,突出说话人的个性特征。

## 2 LPC 谱加权的倒谱

### 2.1 LPC 倒谱

(1) LPC 是线性预测分析的简写。按照线性预测分析,语音信号在时刻 n 的样点值 S(n),可以由过去的 P 个样点值的线性组合来近似。即

$$S(n) \approx a1S(n-1) + a1S(n-2) + \cdots + ap\,S(n-p)$$

上式可看成由过去的样本值的线性组合来预测当前样本的值。这里 a1、a2、a3…a p 是常数,称为 LPC 系数。按照语言信号产生模型,声道、声门激励及辐射的全部谱效应简化为一个时变的数字滤波器。其稳态系统函数为:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{1}$$

上式中,S(z)为系统输出,U(z)为系统的输入。清音的输入是随机噪声,浊音的输入是一定频率的脉冲,即声门波。G 为一常数,ai (i = 1,2…,p)为 LPC 系数。上式又称为 LPC 全极点模型。对于具有短时平稳性的语音段,该模型能对声道谱的

收稿日期:2000-10-23 基金项目:云南省自然科学基金资助(云南民族汉话语音数据库与语音识别研究,99F0016M)
杨澄宇(1970 - ),男,硕士,主要研究方向:语音识别与处理; 赵文(1976 - ),男,硕士,主要研究方向:语音识别与处理.

包络提供很好的近似。

(2) 倒谱是对一帧短时话音进行同态解卷得到的一组时域值。因为一个线性时不变系统的激励信号和系统的冲击响应是按卷积方式结合起来的,语音信号作为线性时不变系统的输出,通过对其进行解卷积处理,使话音中包含的激励源和声道冲击响应分离开,从而更清晰表示出话音包含的各种特征。同态解卷是解卷中的一种[4],其过程如图所示:

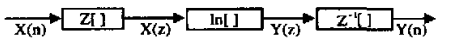


图 1 同态解卷

由图 1 可见,对具有卷积关系的信号 X1(n)和 X2(n),经同态解卷过程后,对应的时域输出信号 Y1(n),Y2(n)之间是相加的关系。Y(n) = Y1(n) + Y2(n)即称为 X(n)的倒谱。

(3) LPC 倒谱是指由 LPC 系数推导出的倒谱系数。在图 2 所示倒谱的定义中,第一步骤是算出 X(n)的 Z 变换 X(z)。而 LPC 系数与 X(n)的 Z 变换 X(z)有式(1)的对应关系。因此,X(z)可由 LPC 系数求出,然后再按上图的步骤则可求出倒谱系数。这样的倒谱系数就叫 LPC 倒谱系数。

### 2.2 LPC 谱加权的倒谱

由参考文献[3]知,在人的语音中,说话人声门的特征分布在中/低频带内。摩擦音的特征主要分布在高频段,另外,高频段还包含说话人声道全长的信息及声带的各种交叠方式(cross - modes of the vocal tract)的信息。所以对高频段和低频段进行增强,有助于说话人识别任务。另外,中频段内往往能量较高,而包含的主要是与话语内容有关的信息。在与文本无关的任务中,对中频段的能量适当减弱,将有利于识别性能的提高。图 2 给出了 LPC 谱加权倒谱的计算过程。在用 LPC 系数求出该段语音的频谱后,对各频段分别加权,最后再转化为频域加权的 LPC 倒谱。

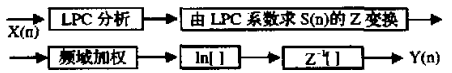


图 2 LPC 谱加权的倒谱的计算过程

## 3 混合高斯模型(GMM)

GMM 目前常被用于"与文本无关的说话人识别",具有良好的识别性能[5]。

定义:一个混合高斯模型的概率密度函数是由 M 个高斯概率密度函数加权求和而得到的,如下式所示:

$$P(x \mid \lambda) = \sum_{i=1}^{M} C_i b_i(x) \tag{2}$$

这里,P(x|λ)是某个 GMM 模型(说话人)的观测值 x 的概率密度函数。x 是 d 维的随机矢量,$C_i$, i = 1,2,…,M 是高斯分布混合时的权重值。$b_i(x)$, i = 1,2,…,M 是第 i 个单个高斯分布概率密度函数,如下式:

$$bi(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \mid \sum_i \mid^{\frac{1}{2}}} exp\{-\frac{1}{2}(x-\mu_i)^t \sum_i^{-1}(x-\mu_i)\} \tag{3}$$

上式中 $\mu_i$ 是均值矢量,$\Sigma_i$ 是协方差矩阵。另外,混合权重值满足:

$$\sum_{i=1}^{M} C_i = 1 \tag{4}$$

λ 为混合高斯模型的参数,由各单个高斯模型的均值矢量、协方差矩阵、和混和权重值共同组成,如下式:

$$\lambda = \{C_i, \mu_i, \sum_i\} \quad i = 1,2\cdots\cdots,M$$

在本实验中,每个说话人用一个混合高斯模型来参数化,即求出一个模型参数 λ 去对应一个说话人。

判别准则:给定 M 个说话人 $Y_1, Y_2, \cdots\cdots, Y_M$ 的高斯混和模型的参数集 $\{\lambda_i\}$ (i = 1,2,……,M)及测试语音的特征矢量集 $X = x_1, x_2, \cdots, x_T$ 时,可算出 X 属于第 j 个模型 $\lambda_j$ 的概率,即后验概率,见下式:

$$P(X \mid \lambda_j) = P(X_1 \mid \lambda_j) P(X_2 \mid \lambda_j) \cdots P(X_T \mid \lambda_j) \tag{5}$$

为了避免运算溢出,上式等号两边取对数得:

$$lnP(X \mid \lambda_j) = lnP(X_1 \mid \lambda_j) + lnP(X_2 \mid \lambda_j) + \cdots + lnP(X_T \mid \lambda_j) \tag{6}$$

根据 bayes 判别规则,从 M 个说话人集合中确认未知说话人得公式如下:

$$i^* = arg(\max_{1 \le i \le N}(lnP(X \mid \lambda_i))) \tag{7}$$

## 4 说话人确认系统的实现及实验结果

### 4.1 说话人确认系统实践

本实验共录制了 30 人的话音,18 个男声和 12 个女声。话音在普通实验室环境下录制,麦克风的音量在半刻度以上。在录音开始的 1 ~ 2 秒录的是室内噪音。录音人按照日常说话习惯说话,没有特殊要求,内容不限,本实验的话音中有方言还有英语。每个人分别录两段话音,一段长一分钟,用于训练说话人概率模型,另一段话音长 30 秒,用作测试语音。话音经 11025Hz 采样及 16Bit 量化后检测端点并分为 20ms 的短时帧。

### 4.2 系统框图

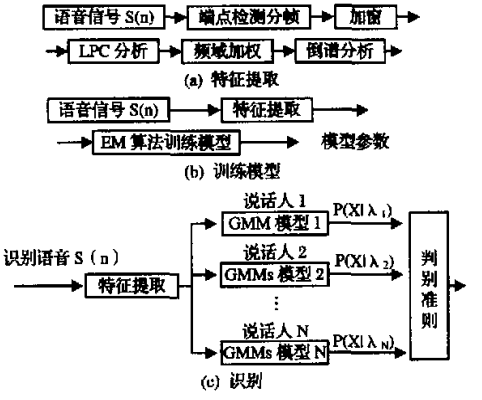


图 3 系统框图

### 4.3 LPC 谱加权曲线图

选加权系数有很多方法,本次实验用的是一组经验值,原则上,应适当增强高频和低频部分的幅值,降低中频部分的幅值。如图 4,图中曲线为 0 ~ 2π 的余弦曲线的近似。$f_l$, $f_h$ 为低、中、高频的分界。实验中,该组权值可使识别性能得到明显改善。

(下转第 11 页)

（上接第 8 页）

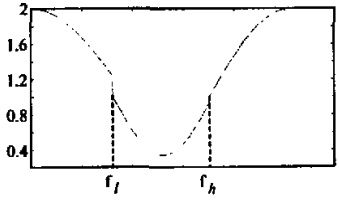


图 4 频谱加权曲线

### 4.4 实验结果

说话人的人数对识别结果的影响：

| | 10 人 | 20 人 | 30 人 |
|---|---|---|---|
| 使用 LPC 倒谱 | 100% | 90% | 80% |
| 使用频域加权 LPC 倒谱 | 100% | 100% | 100% |

当说话人集合的人数增多时，实验表明，改进后的方法取得较好的识别率。

测试话音长度对识别率的影响：

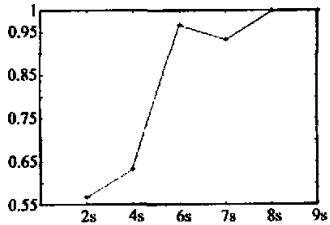


图 5 正识率曲线



图 5 为正识率和测试语音音长之间的关系曲线。横轴表示测试语音音长而纵轴为正识率。这里，训练语音长 20s，说话人集合为 30 人。由图可见，当语音长度超过 6s 时，正识率显著提高。当音长超过 8s 时，正识率可到 100%。

## 5 结论

由以上实验可知，使用改进后的特征参数可以有效的提高正识率，而且所需测试话音的长度、训练话音的长度均不长，有利于实际应用。另外，实验表明，GMM 可对随机变量的概率分布情况提供较好的近似。用于分类和模式识别时，该方法简单而有效。但是，当若干个模式的概率分布较接近时，将降低正确判识的比率。这时对各模式的概率分布进行预先调整，突出模式的特征可有效提高正识率。

### 参考文献

[1] 牟晓隆，胡起秀，吴文虎．与文本无关的复合策略说话人辨识系统[J]．清华大学学报(自然科学版)，1997，37(3)：16－19．

[2] 李蓝华．倒谱参数基音信息有效结合进行说话人辨识[J]．信号处理，2000，(1)：16－19．

[3] Qiguang Lin, En－Ee Jan, ChiWei che, et al.. Selective Use Of The Speech Spectrum And A VQGMM Method For Speaker Identification [A]. ICSLP 96[C]. CDROM. Volumn 4. SuP1L2.

[4] 拉宾纳，R.W.谢弗语音信号数字处理[M]，北京：科学出版社，1984．

[5] K. Markov, S. Nakagawa. Text－Independent speaker identification on *TIMIT database*[A]. *Proceedings, Acous. Soc. Jap.* [C], *1995. 83－84.*