# 一种对混合说话人特征提取的新方法①

## A New Method to Extract Mixed Speaker Feature

毛 鹏 杨鼎才 (燕山大学 信息科学与工程学院 河北 秦皇岛 066004)

**摘 要:** 本文在使用基音周期和美尔倒谱系数(MFCC)计算特征参数的基础上利用主成分分析(PCA)和线性判别分析(LDA)相结合的方法，构造了一种新的混合特征参数。这种新的参数结合了基音周期和 MFCC 各自的特点，利用他们在说话人个人信息上的互补性，然后利用 PCA 和 LDA 相结合的方法提取特征，作为新的说话人特征。实验证明该方法具有更好地表征说话人特征的能力，能更好地识别说话人。

**关键词:** 说话人识别 基因周期 美尔倒谱系数 主成分分析 线性判别分析

## 1 引言

说话人辨认是通过对说话人语音信号的分析和特征提取，确定说话人是谁。在说话人辨认系统中，特征的选择直接影响着整个系统的识别性能，为了减少后续处理的复杂度、提高系统的识别率，人们希望能够尽可能的提取出一组维数小、鉴别能力强的特征矢量。在说话人识别系统中常用的语音特征参数主要有 LPC 倒谱系数(LPCC)、Mel 频率倒谱系数(MFCC)等[1,2]、基音周期和共振峰等。

在实际应用中，一般多选取 MFCC(Mel Frequency Cepstral Coefficients)作为特征矢量来使用。其原因是由于 MFCC 从人耳对频率高低的非线性心理感觉角度反映了语音短时幅度谱的特征，和传统的线性预测倒谱系数 LPCC 相比，其识别性能和抗噪性能具有明显的优势[3-5]。输出的 MFCC 从 Mel 标度频率(亦称感知频率)域提取出的输入语音信号的倒谱参数。这里，Mel 标度描述了人耳对频率感知的非线性特性，在一般的基于 MFCC 的算法设计中，三角滤波器组所包含的 Mel 滤波器的个数 N 以及组内各滤波器的中心频率是固定不变的．这种设计方法没有充分考虑到不同说话人的语音特征，不能有效根据计算得到的 MFCC 在不同说话人之间进行区分。

除了 MFCC 之外 ，基音频率 $f_p$ 也可以作为系统进行说话人识别时的特征来使用，它表征了说话人发浊音时声带振动产生的周期性，可以较好地刻画出不同人各自的声带特性，反映了声带激励源的特点。基音周期 T 为基音频率的倒数．若用一个基音周期内的采样点数表征基音周期，则 T 取决于基音频率 $f_p$ 和语音采样频率 $f_s$，即 $T = f_s / f_p$。基音容易被模仿，不宜单独使用，但可以和倒谱参数相结合。

为了提高说话人识别的准确性和鲁棒性，本文将 MFCC 和基音周期这两种特征进行组合，并且利用 PCA 和 LDA 相结合的方法，即通过 PCA 算法对初次参数进行降维、去相关，并在此基础上计算 LDA，得到更具有可分性的特征矢量。因为当样本数小于样本维数时，直接运用 LDA 算法会出现小样本问题 ，即 $S_W$ 奇异。所以用 PCA 来降低样本维数，解决小样本问题[9]。这样得到的特征矢量，与传统的简单混合特征值和单纯用 PCA 进行特征提取的方法相比，在相同维数时具有最大的类别区分度。

由于支持向量机(SVM)采用结构风险化(SRM)原理，兼顾训练误差和泛化能力，在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势。所以本文然后采用多类支持向量机[10]作为识别分类器。实验结果表明这种新的参数结合了基音周期和 MFCC 各自的优点，具有更好地表征说话人特征的能力，同时，通过 PCA 的降维也可以有效地减少系统运算量。

① 收稿时间:2008-10-12

## 2　说话人个性特征

### 2.1 基音周期

基音周期是语音信号的重要特征参数之一。王修信、徐国钰、胡维平[6]等人推导了一组双正交多小波滤波器，提出了多小波滤波器与自相关结合的基音周期检测方法，利用多小波滤波器从语音信号中提取低频信号，再使用自相关法检测语音信号基音周期。结果表明，该方法提取基音周期具有正确率较高、准确率较高和抗噪性较强的特点．在不同噪声环境下均优于自相关法、单小波与自相关法相结合的方法，尤其在较大噪声干扰下该方法具有明显的抗噪能力．不受语音信号非平稳特性的影响．可以有效地提取病态嗓音的基音周期。本文采用多小波滤波器与自相关结合的基音周期作为说话人特征之一。

### 2.2 美尔倒谱系数

美尔倒谱系数不同于 LPC，它不是从声道模型入手进行分析，它的产生建立在人耳对声音频率的非线性感知基础之上。

MFCC 是在频谱上采用滤波器组的方法计算出来的，这组滤波器在频率 Mel 坐标上是等带宽的。这是因为人类在对 1kHz 以下的声音频率范围的感知遵循近似线性关 系；对 1kHz 以上的声音频率范围的感知遵循在对数频率坐标上的近似线性关系。MFCC 弱化了语音频谱高频成分，对噪声具有适应性，是鲁棒说话人系统中常用的一种特征参量。

语音信号经过预加重、分帧加窗等预处理后，分别求得每一帧的基音周期和 MFCC 系数，然后把他们合并成新的组合特征。

## 3　结合PCA和LDA进行特征降维

MFCC 和基音周期对基本音素的表征能力各有所长，但如果直接将它们进行叠加，特征的维数即增加了一倍，这样就增加了训练和识别时的计算量，不利于系统的实时运行。我们结合 PCA 和 LDA 进行降维。

### 3.1 主分量分析进行提取

由于实际上众多观测数据之间不可避免的存在相关、冗余、噪声等影响，必然导致计算量庞大以及造成分析的不准确。 语音的 mel 倒谱系数(p 维向量)之间也存在一定的相关和噪声。而 PCA 方法就是利用 p 元实际观测数据构造出元综合观测变量，一般 $m<p$，使得 m 元综合变量既尽可能多的反映原来 p 元数据提供的信息，同时 m 元综合变量之间又相互独立，可以有效的减少计算量，还可以提高分析精度。记 m 元综合变量为 $y_1, y_2 ... y_m$ 由 p 元实际观测数据线性表示[7]，即

$$\begin{cases} y_1 = u_{11}x_1 + u_{12}x_2 + ...u_{1p}x_p \\ y_2 = u_{21}x_1 + u_{22}x_2 + ...u_{2p}x_p \\ \cdots \\ y_m = u_{m1}x_1 + u_{m2}x_2 + ...u_{mp}x_p \end{cases} \quad (1)$$

式中，$u_{ij}$ 由以下原则确定：

① $y_i$ 与 $y_j$ 互不相($i \neq j; i, j = 1, 2, ..., m$)

② $y_1, y_2, ...y_m$ 分别称为原实际观测的 p 元变量的第 $1, 2, ..., m$ 个主成员，其中 $y_1$ 的方差在总方差中占的比例最大，其他依次递减。实际应用中一般选取前 m 个方差最大的成分，即所谓的主成分。可以看出，主成分分析其实就是构造原来实际观测变量的一系列线性组合，使得各个线性组合在互不相关的前提下尽可能多的反映原始数据的信息(用方差来衡量信息量)。用 PCA 对混合特征进行特征提取，既可以去噪、去相关，还可以把特征维数(m)降到样本数以下，解决了 LDA 算法会出现的小样本问题。

### 3.2 线性判别分析进行再提取

如果想更准确的识别说话人，应该使训练和测试特征更具可分性。LDA 方法又称应用线性辨别分析方法[8]。它的目标就是从高维特征空间里提取出最具有判别能力的低维特征。这些特征能帮助将同一个类别的所有样本聚集在一起，不同类别的样本尽量分开,即选择使类间和类内离散度的比值最大的特征(Fisher 准则)。LDA 方法定义类内离散度矩阵 $S_W$ 和类间离散度矩阵 $S_B$ 如下：

我们在这里提出根据特征参数的类别可分离性来对它们排序，由此选出那些可分离性最优的特征参数，达到了降维的目的并得到较优的识别性能，能更好的表征说话人的特征，具有很好的分类能力。

衡量说话人特征参数有效性的 Fisher 比[5]公式如下：

$$S_W = \sum_{i=1}^{C} \sum_{j=1}^{N_i} P_i (x_{ij} - m_i)(x_{ij} - m_i)^T \quad (2)$$

$$S_B = \sum_{i=1}^{C} P_i (m_i - m)(m_i - m)^T \quad (3)$$

其中,N 表示样本总数，包含 C 类模式，$N_i$ $(i=1,2,...,N)$表示第 i 类样本的数量。$m_i$ 表示各类模式的均值，m 表示总样本均值。C 类模式表示为：

$$x_i=\{x_{i1},x_{i2}...x_{iN_i}\},x_{ij}(1\le i\le C;1\le j\le N_i)$$

LDA 的目的就是要寻找变换 W，当 $S_W$ 非奇异时，使得 Fisher 准则最大：

$$JW=\arg\max\frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|} \quad (4)$$

这里的 $W_i\ (1\le i\le m)$ 就是满足如下等式的解：

$$S_B W_i=\lambda S_W W_i \quad (5)$$

我们用变换矩阵对 PCA 提取后的混合特征(维)特征变换，经再次提取后得到更具有判别能力的低维特征。然后将其送入分类器 SVM 进行识别。

## 4 实验结果

本实验环境是基于 Matlab7.1，陆正波开发的 OSU-SVM 工具箱，Intel Celeron 处理器、CPU2.8GHz、RAM 512MB。

实验的训练数据和测试数据采用 PKU－SRSC 语音数据库中的窄带语音，包括 3 男 3 女，录音内容为 10 个数字串 2 遍。每个人有 40 次录音，前 20 次语音与后 20 次录音时间相隔大约一周时间，每次录音大约 5s 左右。

本实验语音预处理采用 8KHz 采样，8bit 量化。预加重系数采用 0.9375。分帧汉明窗长 256 点(32ms)，帧移 60 点(7.5ms)。进行端点检测滤出静音段，只保留有声段。

采用基音周期和 MFCC 及一阶差分，以及对混合特征降维后三种特征分别用 SVM 识别的识别率进行比较。

识别结果比较如表 1 所示：

表 1 不同特征的识别率

| 维数 | 基音(%) | MFCC 及一阶差分(%) | 混合特征(%) |
|---|---|---|---|
| 20 | 53.24 | 94.18 | 98.76 |
| 24 | 50.28 | 90.97 | 98.86 |
| 28 | 51.16 | 89.89 | 97.45 |
| 30 | 54.57 | 84.36 | 98.44 |
| 34 | 55.63 | 82.85 | 98.36 |

## 5 结论

本文提出了基于核线性判别分析的说话人特征提取方法，分类器采用高斯核，惩罚因子和参数采用网格搜索，对北京大学 PKU-SRSC 语音库进行了仿真实验。从实验数据看出，它所得到的特征参数有效地结合了基音周期和 MFCC 各自的优点。和其他特征参数相比，具有更好的分类能力和稳定性 ,而且不增加训练和识别过程的计算量，具有很好的实用性，证明了我们提出的方法的有效性。

## 参考文献


1 赵力.语音信号处理.北京:机械工业出版社, 2003:51－65.

2 张军英.说话人识别的现代方法与技术.西安:西北大学出版社, 1994:15－26.

3 Hung WW，Wang HC. On the Use of Weighted Filter Bank Analysis for the Derivation of Robust MFCCs. IEEE Signal Processing Letters, 2001,8(3): 70－73.

4 Molau S, Pitz M, Schluter R, Ney H. Computing Mel-Frequency Cepstral Coefficients on the Power Spectrum. Proc of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City,USA, 2001(1):73－76.

5 Shao Y，Liu BZ，Li ZG. A Speaker Recognition System Using MFCC Features and Weighted Vector Quantization. Computer Engineering and Applications, 2002,38 (5):127－128.

6 王修信,徐国钰,胡维平,梁冬冬,卢小春,王强.一种多小波滤波器在基音周期提取中的应用.计算机工程与应用, 2007:191－193.

7 范金城.数据分析.北京:科学出版社, 2002:141－153.

8 Liu C, Wechsler H. Enhanced fisher linear discriminant models for face recognition. Proceedings of Fourteenth International Conference on Pattern Recognition, Australia, 1998,2: 368－372.

9 何国辉,甘俊英.PCA-LDA 算法在性别鉴别中的应用.计算机工程, 2006(9):208－210.

10 李红莲,焦瑞莉,范京.支持向量机多类分类方法的精度分析.北京机械工业学院学报, 2008(2):32－35.