

# 基于改进 MFCC 和短时能量的咳嗽音身份识别

赵青, 成谢锋, 朱冬梅

(南京邮电大学 电子科学与工程学院, 江苏 南京 210003)

**摘要:**介绍了基于咳嗽音信号的身份识别方法。针对咳嗽音信号的短时突发特点,提出了一种适合咳嗽音的改进 MFCC 特征参数 MFCC\_N,然后将 MFCC\_N 与短时能量(E)作为组合特征参数应用于身份识别系统中。在 MATLAB 7.0 平台上实现基于 GMM 的咳嗽音身份识别系统,分别提取咳嗽音信号的 MFCC、MFCC+ $\Delta$ MFCC 和 MFCC\_N+E 作为识别参数进行对比实验。实验结果表明,采用提出的组合特征参数 MFCC\_N+E 进行身份识别是可行有效的,与传统参数 MFCC、MFCC+ $\Delta$ MFCC 相比,采用特征参数 MFCC\_N+E 的识别系统具有较高的识别率和较低的计算复杂度。

**关键词:**咳嗽音;特征参数;短时能量;高斯混合模型;识别

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2012)06-0082-03

## Cough Sound Identification Based on Improved MFCC and Short-time Energy

ZHAO Qing, CHENG Xie-feng, ZHU Dong-mei

(Department of Electronic Science and Engineering, Nanjing University of Posts and Telecommunications,  
Nanjing 210003, China)

**Abstract:** Introduce an identification method based on cough sounds. Concerning the short burst characteristic of cough sounds, a new feature parameter MFCC\_N is proposed, then put MFCC\_N and short-time energy together as a new series of coefficients. In order to verify the validity of the new coefficient, an identification system based on GMM is established. Respectively extract MFCC, MFCC+ $\Delta$ MFCC and MFCC\_N+E as recognition parameter to compare. The experiments were implemented in MATLAB 7.0 environment. The results showed that the system performance with new parameter was feasible and could obtain higher recognition rate and low computational complexity than that with MFCC and MFCC+ $\Delta$ MFCC.

**Key words:** cough sound; feature parameter; short-time energy; GMM; recognition

## 0 引言

近年来,生物特征识别成为国际上的研究热点之一。目前,国内外研究领域中常见的生物特征有指纹、掌纹、虹膜、视网膜、静脉、步态、耳廓、语音以及手部、面部模式等,但是这些特征中有些容易被人为的伪造和窃取。为了提高识别的精度和安全性,目前已有不少学者将目光投向人体内部的信息,如 DNA、心音信号等<sup>[1,2]</sup>。文中提出了一种基于咳嗽音信号的生物特征识别方法。咳嗽音信号是人体重要的生理信号之一,含有每个人独特的生理特征,因此,这些特征也可以用来对人的身份进行辨认和确认。生物特征识别最关键的是特征提取。文中在深入研究咳嗽音时域和频

域特征后,提出将传统的 MFCC 和动态 MFCC 加权相加,再与时域特征短时能量组合,作为新的特征参数。

## 1 特征参数提取

特征提取本质上起降维作用,用较少的维数表征每个人的特征。文中特征提取的任务是从每个人的咳嗽音信号中获得一组能够描述其个性特征的参数。在提取特征参数之前,需要对原始样本进行预处理,使之能更加精确地提取特征。

### 1.1 预处理

预处理包括预加重、端点检测、分帧加窗。

(1)预加重:预加重的目的是提升咳嗽音信号的高频部分,使信号频谱变得平坦,以利于进行频谱分析或声道参数分析<sup>[3]</sup>。预加重可以采用 6dB/oct 的数字滤波器来实现,文中采用的数字滤波器的传递函数为: $H(z) = 1 - \mu z^{-1}$ ,其中  $\mu$  为预加重系数,取值 0.97。

(2)端点检测:端点检测就是准确地找出咳嗽音

收稿日期:2011-11-10;修回日期:2012-02-13

基金项目:南京邮电大学基金项目(NY207139)

作者简介:赵青(1984-),女,河南开封人,硕士研究生,研究方向为智能信息系统与应用;成谢锋,教授,硕士生导师,从事智能信息、智能仪器方面的研究工作。

信号的起始点和终止点,这是咳嗽音识别中的一个重要环节。有效的端点检测不仅能消除无声段的噪声干扰,而且可以缩短咳嗽音信号的处理时间。

(3) 分帧加窗:咳嗽音信号是非平稳信号,加窗的目的是把它分为一个个短时段,每个短时段称为一帧。在这样一段短时间(10~30ms)内,咳嗽音信号可以看成是平稳信号。为尽可能减少信息的丢失,相邻帧之间重叠半帧。

## 1.2 传统 MFCC 参数的提取

人耳对不同频率声音信号有不同的感知能力,在 1000Hz 以下呈近似线性关系,在 1000Hz 以上成对数关系,这就使得人耳对低频信号比对高频信号更为敏感<sup>[4]</sup>。MFCC 是基于 Mel 频率域提出来的倒谱系数,更符合人耳的听觉特性<sup>[5]</sup>,MFCC 参数强调声音信号的低频信息。正常人的咳嗽音信号频率集中在 200~500Hz。因此,提取咳嗽音的 MFCC 参数能更好地突出咳嗽音的特征,有利于识别,同时消除高频噪声的干扰。

Mel 频率和线性频率的转换关系是:

$$f_{\text{mel}} = 2595 \log_{10} \left( 1 + \frac{f}{700} \right) \quad (1)$$

MFCC 的计算过程<sup>[6]</sup>如图 1 所示:

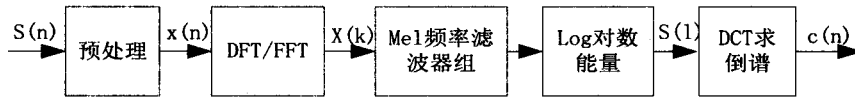


图 1 MFCC 的提取过程

具体步骤如下:

(1) 原始咳嗽音信号  $S(n)$  通过预加重、端点检测、分帧加窗等处理后,得到咳嗽音的单帧时域信号  $x(n)$ 。

(2) 将每帧时域信号  $x(n)$  经过离散傅立叶变换 (DFT) 后得到线性频谱  $X(k)$ ,转换公式为:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi n k}{N}} \quad (0 \leq n, k \leq N-1) \quad (2)$$

(3) 将线性频谱  $X(k)$  通过 Mel 频率滤波器组,并经过对数能量的处理,得到对数频谱  $S(l)$ 。

(4) 将上述对数频谱  $S(l)$  经过离散余弦变换 (DCT) 变换到倒谱域,就得到单帧咳嗽音信号的 MFCC 系数  $c(n)$ :

$$c(n) = \sum_{l=0}^{L-1} S(l) \cos \left( \frac{\pi n (l + 0.5)}{L} \right) \quad (0 \leq l \leq L) \quad (3)$$

提取了 MFCC 参数后,可利用公式(4)提取其差分特征参数。差分特征参数用来描述咳嗽音的动态特征。

其中,  $d_t$  是第  $t$  个一阶差分倒谱系数,  $y$  是倒谱系数的维数,  $\delta$  为一阶导数的时间差,其取值 1 或 2,  $1 \leq \theta$

$\leq \delta$ ,  $c_t$  是第  $t$  个倒谱系数<sup>[7]</sup>。

$$d_t = \begin{cases} c_{t+1} - c_t & t < \delta \\ c_t - c_{t+1} & t \geq y - \delta \\ \frac{\sum_{\theta=1}^{\delta} \theta (c_{t+\theta} - c_{t-\theta})}{2 \sum_{\theta=1}^{\delta} \theta^2} & \text{其他} \end{cases} \quad (4)$$

## 1.3 对传统 MFCC 的改进

传统的 MFCC 只反映了咳嗽音信号的静态特性而没有考虑其动态特性,很难达到可靠的性能要求。为了描述咳嗽音信号的动态特性进而提高系统性能,比较流行的方法是组合传统 MFCC 及其一阶差分<sup>[8,9]</sup>。然而,这种方法不但增加了特征矢量的维数,而且增加了识别系统的计算复杂度。

针对以上问题,文中提出了一种改进的 MFCC 算法,设为 MFCC\_N,公式为:

$$\text{MFCC\_N} = \alpha \cdot \text{MFCC} + \beta \cdot \Delta \text{MFCC} + \gamma \cdot \Delta^2 \text{MFCC} \quad (5)$$

其中,  $\Delta \text{MFCC}$  是一阶差分,  $\Delta^2 \text{MFCC}$  是二阶差分,  $\alpha, \beta, \gamma$  是权重,由于它们所起的作用不同,所以限制  $0 < \gamma < \beta < \alpha < 1$ , 且  $\alpha + \beta + \gamma = 1$ 。在公式(5)中, MFCC 描述声道特性,  $\Delta \text{MFCC}$  代表动态特性,  $\Delta^2 \text{MFCC}$  作为平衡因子。改进的方法大大降低了特征参数矩阵的维数,对于  $T$  帧  $S$  维的咳嗽音信号,普通方法提

取到的特征参数矩阵大小为  $T \times (2 \times S)$ , 而用文中的提取算法得到的特征参数矩阵大小为  $T \times S$ 。MFCC\_N 与传统 MFCC 有相同的维数,相同的幅度变化趋势,但同时比传统 MFCC 有更好的性能, MFCC\_N 不仅反映了声道特性,而且更好地逼近了咳嗽音信号的动态变化特性。

## 1.4 短时能量

咳嗽音信号具有突发性、短暂性特征,信号能量随时间变化比较明显。短时能量给出了反映这些幅度变化的一个合适的描述方法。短时能量的定义为:

$$E_n = \sum_{m=n-N+1}^n [x(m)w(n-m)]^2 \quad (6)$$

其中,  $x(m)$  为待分析的咳嗽音信号,  $w(n)$  为窗函数,  $N$  为窗长。

## 1.5 组合特征参数

文中将改进的 MFCC 参数 MFCC\_N 与短时能量组合作为咳嗽音的特征矢量。短时能量是咳嗽音信号一个重要的时域特征, MFCC\_N 是基于人听觉的特征,二者相关性不大,组合之后从不同角度更好地反映了咳嗽音信号的特征。另外,短时能量是标量值且计算

简单,运算复杂度低,与 MFCC\_N 组合后不会明显增加特征矢量的维数。

文中 MFCC 取 16 维,则 MFCC\_N 也是 16 维,将每帧的短时能量  $E$  作为组合特征矢量的第 17 维。

## 2 高斯混合模型(GMM)

GMM 是基于参数估计的概率统计模型,它的优点是可以平滑地逼近任意形状的概率密度函数,并且是易于处理、相当稳定的参考模型<sup>[10]</sup>。它用多个高斯分布的线性组合来描述特征矢量在特征空间的分布,每个人的咳嗽音特征在特征空间都有特定的分布,即每个人的咳嗽音特征参数都对应一个 GMM。用训练样本中提取的咳嗽音特征参数训练 GMM 模型,最终得到每个人的 GMM 模型参数  $\lambda$ ,表示为如下三元式:  $\lambda = \{ \omega_i, \mu_i, \sum_i \}$ ,  $i=1,2,\cdots,M$ 。其中,  $\omega_i$  是混合分量的权值;  $\mu_i$  是均值矢量;  $\sum_i$  是协方差矩阵;  $M$  是混合阶数。GMM 模型中的参数是由不同人的特征参数分布决定的,因此表征了不同人的身份。

为不同人的咳嗽音建立 GMM 模型,实际上就是通过训练样本的特征矢量来估计 GMM 的参数  $\lambda$ 。最常用的估计算法是最大似然准则(ML),通过 EM 算法迭代来实现<sup>[11]</sup>。对于给定的特征矢量序列  $X = \{X_t\}$ ,  $t=1,2,\cdots,T$ ,  $T$  是特征矢量的帧数,则对数似然度可以定义为<sup>[12]</sup>:

$$L(X/\lambda) = \frac{1}{T} \sum_{t=1}^T \log P(X_t/\lambda) \tag{7}$$

在  $W$  个不同人的咳嗽音闭集中,识别的目的是在这  $W$  个不同人的咳嗽音中找到属于  $K$  的咳嗽音,  $K \in \{1,2,\cdots,W\}$ ,通过求取最大的后验概率  $P(\lambda_k/X)$  得到  $K$  的模型  $\lambda_k$ 。GMM1,GMM2, $\cdots$ ,GMM $W$  分别为 1,2, $\cdots$ , $W$  个人咳嗽音的训练模型。基于 GMM 的咳嗽音身份识别模型如图 2 所示:

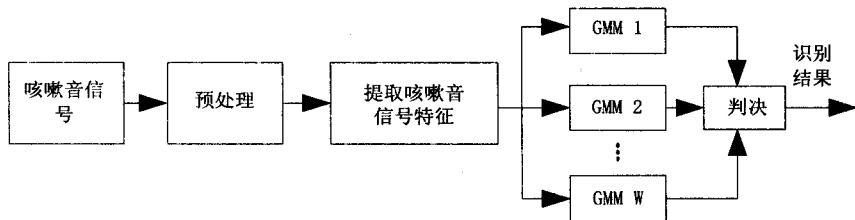


图 2 基于 GMM 的咳嗽音识别

## 3 实验结果与分析

实验中所用的咳嗽音信号来自 20 个正常的人,男女各半,都在平静的状态下采集。采用的采集软件为 Cool Edit Pro,采样频率设置为 8kHz,精度为 16 位,单声道,存储为 wav 格式。每人采集三组数据,两组用于

训练,一组用于测试。实验是以 Matlab 7.0 软件为开发平台实现的。实验分别采用 MFCC, MFCC+ $\Delta$ MFCC, MFCC\_N+E 为特征参数,维数分别是 16、32、17。其中  $\alpha$ 、 $\beta$ 、 $\gamma$  的值由多次实验分析得出,最终选出最适合本实验中咳嗽音特征的  $\alpha$ 、 $\beta$ 、 $\gamma$  值( $\alpha=1/2$ ,  $\beta=1/3$ ,  $\gamma=1/6$ )。

实验一,本实验研究高斯混合阶数对识别性能的影响,从而选择合适的混合阶数。实验中所用的 GMM 阶数分别为 4、8 和 12。表 1 给出了不同特征在不同混合阶数下的识别性能。

表 1 不同高斯混合阶数下各个特征的识别率(%)

特征参数	阶数 M		
	4	8	12
MFCC(16 维)	73.27	82.10	85.52
MFCC+ $\Delta$ MFCC (32 维)	74.82	83.48	87.40
MFCC_N+E (17 维)	75.86	84.55	88.53

实验二,为了得到较高的识别率,同时要有较低的计算复杂度,本实验对不同阶数、不同特征参数进行了复杂度研究。复杂度的评判准则是识别时间,单位为秒。实验结果见表 2。

表 2 不同特征参数的复杂度对比

阶数 M	特征参数		
	MFCC (16 维)	MFCC+ $\Delta$ MFCC (32 维)	MFCC_N+ E (17 维)
4	0.62	0.93	0.71
8	0.85	1.46	1.07
12	1.03	1.65	1.24

由表 1 可以看出随着 GMM 阶数的增加识别率逐渐上升。另外,对于每一个相同的 GMM 阶数(4、8 或 12),采用组合特征参数 MFCC\_N+E 比单独使用 MFCC 的识别率提高 3%,同时比 MFCC+ $\Delta$ MFCC 的识别率高 1.2%。由表 2 可以得到阶数越高,计算复杂度越大。特征参数 MFCC+ $\Delta$ MFCC 的复杂度大约是 MFCC\_N+E 的 1.3 倍。综上所述,采用组合特征参数 MFCC\_N+E 进行咳嗽音身份识别是可行且有效的,在提高识别率的同时节约了识别时间。

## 4 结束语

文中提出了一种新的生物特征识别方法—咳嗽音身份识别方法。通过分析咳嗽音的特点,提出了一种

(下转第 88 页)

表 2 实验二识别结果

	训练样本数	测试样本数	识别率
PCA 方法	250	390	79.74%
直方图均衡化	250	390	87.69%
文中方法	250	390	97.69%

## 5 结束语

不同光照条件一直是影响人脸识别性能的一个很重要的因素,文中从分类的角度出发提出了一种解决不同光照条件的人脸识别方法,该方法对人脸图像的灰度图进行分析并分类,然后对不同的类别分别进行主成分分析得到不同的投影子空间进行人脸识别,这样可以在很大程度上消除不同光照条件对人脸识别的影响。实验结果表明该方法能有效地解决不同光照条件对人脸识别的影响。

### 参考文献:

- [1] 赵振勇,王保华,王力,等.人脸图像的特征提取[J].计算机技术与发展,2007,17(5):221-224.
- [2] Zhao W,Chellappa R,Rosenfeld A,et al. Face Recognition: A Literature Survey[J]. ACM Computing Surveys,2003,35(4):399-458.
- [3] Moses Y,Ullman S. Limitations of Non-model-based Recognition Schemes[C]//Proc. of 2nd European Computer Vision

Conference. [s. l.]:Springer-Verlag,1992:820-828.

- [4] Wei S,Robust S L. Face recognition under lighting variations [C]//Proceedings of 17th International Conference on Pattern Recognition. [s. l.]:[s. n.],2004:354-357.
- [5] Miao J,Yin B C,Chen X C. A Hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-center Template[J]. Pattern Recognition,1999,32(7):1237-1248.
- [6] Buhmann J,Lades M,Eeckman F. A silicon retina object recognition[R]. Institut fur Informatik,Universitat Bonn,1993.
- [7] 时书剑,马燕.基于 Gabor 滤波和 KPCA 的人脸识别方法[J].计算机技术与发展,2010,20(4):51-53.
- [8] Ramamoorthi R,Hanrahan P. On the relationship between radiance and irradiance:determining the illumination from images of a convex Lambertian object[J]. Journal of the Optical Society of America A,2001,18(10):2448-2459.
- [9] Georgiades A S,Belhumeur P N,Kriegman D J. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence,2011,23(6):643-660.
- [10] 张文增,陈强,都东,等.直线检测的灰度积分投影方法[J].清华大学学报,2005,45(11):1446-1449.
- [11] 尹飞,冯大政.基于 PCA 算法的人脸识别[J].计算机技术与发展,2008,18(10):31-33.
- [12] Turk M,Pentland A. Eigenfaces for Recognition[J]. Journal of Cognitive Neuroscience,1991,3(1):71-86.

(上接第 84 页)

改进的 MFCC 参数 MFCC\_N 与短时能量相结合的方法。该组合参数综合考虑了咳嗽音的静态特性、动态特性和时域特性,对咳嗽音信号特征描述得更准确、更全面。经多次实验验证,组合参数 MFCC\_N+E 比传统参数 MFCC、MFCC+ $\Delta$ MFCC 的识别率高,而且维数少,计算复杂度相对较低。在文中的研究基础上,下一步将扩大咳嗽音样本,继续研究基于咳嗽音的身份识别,同时设法提高系统识别率。

### 参考文献:

- [1] 成谢锋,马勇,陶冶薇,等.基于数据融合的三段式心音身份识别技术[J].仪器仪表学报,2010,31(8):1712-1720.
- [2] 宋明明,成谢锋,王厚大. HILBERT 包络法在心音身份识别中的应用[J].微型机与应用,2010,29(2):78-82.
- [3] 王炳锡,屈丹,彭焯.实用语音识别基础[M].北京:国防工业出版社,2005:26-27.
- [4] 魏丹芳,李应.基于 MFCC 和加权动态特征组合的环境音分类[J].计算机与数字工程,2010,38(2):7-10.
- [5] Twang Y,Li B,Jiang X Q,et al. Speaker Recognition Based on Dynamic MFCC Parameters[C]//International Conference

on Image Analysis and Signal Processing. [s. l.]:[s. n.],2009:406-409.

- [6] 邵央,刘丙哲,李宗葛.基于 MFCC 和加权矢量量化的说话人识别系统[J].计算机工程与应用,2002(5):127-128.
- [7] 郭春霞,裘雪红.基于 MFCC 的说话人识别系统[J].电子科技,2005(11):53-56.
- [8] Lai Y P,Siu M H,Mark B. Joint Optimization of the Frequency Domain and Time-domain Transformation in Deriving Generalized Static and Dynamic MFCCs[J]. IEEE Signal Processing Letters,2006,13(11):707-710.
- [9] Wang C,Miao Z J,Meng X. Differential MFCC and Vector Quantization Used for Real-time Speaker Recognition System[J]. Congress on Image and Signal Processing,2008(5):319-323.
- [10] 陈芬菲.基于 GMM 的说话人识别系统[J].微处理机,2006(8):76-79.
- [11] Dempster A,Laird N,Rubin D. Maximum likelihood from incomplete data via the EM algorithm[J]. J Royal Stat Soc,1977,39(1):1-38.
- [12] 赵力.语音信号处理[M].北京:机械工业出版社,2003:249-250.