Journal of Computer Applications
计算机应用,2012,32(9):2542 -2544
ISSN 1001-9081
CODEN JYIIDU
2012-09-01
http://www.joca.cn

文章编号:1001 -9081(2012)09 -2542 -03
doi:10.3724/SP.J.1087.2012.02542

基于梅尔频率倒谱系数与翻转梅尔频率倒谱系数的说话人识别方法

胡峰松,张　璇*
(湖南大学 信息科学与工程学院,长沙 410082)
(* 通信作者电子邮箱 zhangxuan2007@163.com)

摘　要: 为提高说话人识别系统的识别率,提出了基于梅尔频率倒谱系数(MFCC)与翻转梅尔频率倒谱系数(IMFCC)为特征参数的特征提取新方法。该方法利用 Fisher 准则将 MFCC 和 IMFCC 相结合,构造了一种混合特征参数。实验结果表明,新的混合特征参数与 MFCC 相比,在纯净语音库及噪声环境中均具有较好的识别性能。

关键词: 说话人识别;梅尔频率倒谱系数;翻转梅尔频率倒谱系数;Fisher 准则;高斯混合模型
中图分类号: TN912.34　　**文献标志码:** A

Speaker recognition method based on Mel frequency cepstrum coefficient and inverted Mel frequency cepstrum coefficient

HU Feng-song, ZHANG Xuan*
(*College of Information Science and Engineering, Hunan University, Changsha Hunan* 410082, *China*)

Abstract: To improve the performance of speaker recognition system, a new method of feature extraction was proposed based on Mel Frequency Cepstrum Coefficient (MFCC) and Inverted MFCC (IMFCC). This method constructed a mixed feature by combining MFCC with IMFCC using Fisher criterion. The experimental results show that the mixed feature proposed in this paper has better recognition performance compared with MFCC not only in the pure voice database but also in the noisy environments.

Key words: speaker recognition; Mel Frequency Cepstrum Coefficient (MFCC); Inverted MFCC (IMFCC); Fisher criterion; Gaussian Mixture Model (GMM)

0　引言

说话人识别[1]是指根据说话人的声音识别说话人身份的技术,其基本的原理是将说话人的测试模型与训练好的模型进行匹配,从而来判断说话人的身份。随着计算机和信息技术的快速发展,以及对快速有效身份验证的迫切要求,基于生物特征的身份鉴别技术已成为研究热点。语音信号具有易于获取、传输和储存等特点,因此基于人类语言的说话人识别技术已成为生物认证技术的重要内容之一。

如何从说话人的语音信号中提取表征说话人的基本特征是说话人识别中最重要的问题之一。目前主流的说话人识别特征参数依然是梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC)[2]。近年来,提出了将 MFCC 与其他说话人特征组合后作为说话人识别系统的新方法,文献[3]利用 Fisher 准则将傅里叶分析和小波分析结合起来构造了一种新的特征参数,提高了系统的识别率;文献[4]将美尔倒谱系数及其差分与线性预测倒谱系数及其差分相结合作为识别的特征参数,并验证了其有效性;文献[5]用支持向量机分别以 MFCC 与翻转梅尔频率倒谱系数[6](Inverted MFCC, IMFCC)为特征单独执行分类,将其结果按多分类融合方法融合,实验证明其在一定程度上提高了识别率。

本文首先计算出 MFCC 参数和 IMFCC 参数,然后用 Fisher 准则[7-10]构造了一种混合特征参数,最后利用 TIMIT 和 NOIZEUS 语音库[11]进行实验的结果证明,这种混合参数有效地提高了说话人识别系统的识别率。

1　MFCC 的提取

MFCC 的分析基于人耳的听觉机理,具有较高的识别率和较好的鲁棒性。Mel 频率表达了一种常用的从语音频率到"感知频率"的对应关系。实际应用中,通常对 Mel 频率做如下近似:对 1 kHz 以下的语音信号采用线性频率;对 1 kHz 以上的语音信号采用对数频率。其转换关系如下:

$$F_{\mathrm{mel}}(f) = 2595 \lg(1 + f/700)$$

其中:频率 f 的单位是 Hz,梅尔频率 F_{mel} 的单位是 Mel。提取梅尔倒谱系数的过程如图 1 所示。

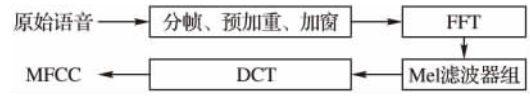

图 1　MFCC 的计算过程

2　IMFCC 的提取

传统的梅尔滤波器结构[12]如图 2(a)所示,在低频区域分布集中,高频部分分布稀疏,忽略了高频中的一些信息。翻

收稿日期:2012-03-13;修回日期:2012-06-06。

作者简介:胡峰松(1969 -),男,湖南长沙人,副教授,博士,主要研究方向:数字图像处理、说话人识别;　张璇(1988 -),女,湖南张家界人,硕士研究生,主要研究方向:说话人识别。

转梅尔滤波器的结构刚好与传统梅滤波器相反，在低频区域分布稀疏，在高频区域分布集中(其结构如图 2(b) 所示) 。翻转后的梅尔滤波器响应为:

$$\hat{H}_i(k) = H_{P-i+1}[N/2 - k + 1]$$

其中: N 为快速傅里叶变换(Fast Fourier Transform，FFT) 的采样点数，P 为滤波器的个数，$H_i(k)$ 为梅尔滤波器组的响应。

$$H_i(k) = \begin{cases} \dfrac{2(k - f[i-1])}{(f[i+1] - f[i-1])(f[i] - f[i-1])}, & f[i-1] < k \leqslant f[i] \\ \dfrac{2(f[i+1] - k)}{(f[i+1] - f[i-1])(f[i+1] - f[i])}, & f[i] < k < f[i+1] \\ 0, & k \leqslant f[i-1] \text{ 或 } k \geqslant f[i+1] \end{cases}$$

提取 IMFCC 的过程与提取 MFCC 一样，唯一不同的就是经过预处理和 FFT 的信号是通过翻转梅尔滤波器组滤波，然后对其进行离散余弦变换(Discrete Cosine Transform，DCT) 后就得到 IMFCC。

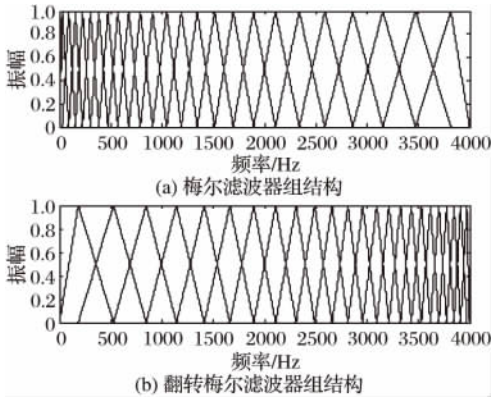

图 2　两种滤波器结构

3　Fisher 准则

以上两种语音参数关系互补，可结合两种特征参数的优势来获得更好的识别结果。但是直接将它们进行叠加，会增加特征的维数，从而增加了训练和识别时的计算量; 还有特征各维的区分度不同，有些特征可能属于冗余信息或者对识别性能有干扰的信息，直接叠加反而会影响整体的识别性能。

在模式识别中，一个参数的类别可分离性[11] 可用 Fisher 准则来测定:

$$r_{\text{Fisher}} = \frac{\sigma_{\text{Between}}}{\sigma_{\text{Within}}}$$

其中: r_{Fisher} 为特征参数的 Fisher 比，某个参数对训练集样本的 Fisher 比越大，则这个参数的类别区分度越好。σ_{Within} 是这个特征的对于各个类的类内散度(方差) 之和，在有 M 种类 w_j $(j = 1,2,\cdots,M)$，各类的样本数为 n_j 的情况下，第 t 个参量的 σ_{Within} 的计算公式如下:

$$\sigma_{\text{Within}} = \sum_{j=1}^{M}\left[\frac{1}{n_j}\sum_{c\in w_j}(c_t^{(j)} - m_t^{(j)})^2\right]$$

其中: $c_t^{(j)}$ 表示第 t 个参量在第 j 类上的取值，$m_t^{(j)}$ 表示第 t 个参量在第 j 类上的均值。σ_{Between} 是这个特征的对于各个类的类间散度(方差) 之和，计算公式如下:

$$\sigma_{\text{Between}} = \sum_{j=1}^{M}(m_t^{(j)} - m_t)^2$$

其中 m_t 表示第 t 个参量在所有类上的均值。

图 3 ~ 4 给出了求得的 12 维 MFCC 和 12 维 IMFCC 的维数和 Fisher 比的关系。可看出，特征值各维的贡献量是不同的。

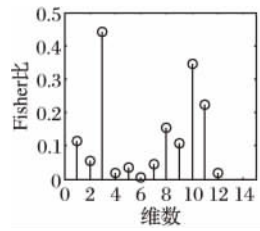

图 3　MFCC 参数各维 Fisher 比

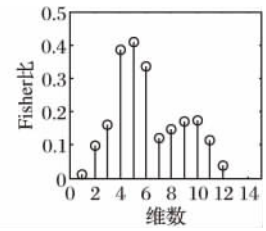

图 4　IMFCC 参数各维 Fisher 比

因此，本文分别从 MFCC 和 IMFCC 中选出 Fisher 比最大的 6 组组成最终用于识别的特征矢量。与特征矢量简单叠加相比，减少了特征的维数，并且所选取的特征参数之间的冗余信息少，识别效果更好。新的混合特征的算法提取流程如图 5 所示。其提取算法如下:

1) 预处理。

预处理包括分帧、预加重、加窗。

2) FFT。

将加窗后的声音信号进行 FFT，将声音信号由时域变换到频域，得到声音信号的离散功率谱 $X(t)$ 。

3) 滤波器组滤波。

将 $X(t)$ 分别通过 Mel 滤波器组和翻转 Mel 滤波器组，对其进行滤波处理。

4) DCT。

对经过滤波之后的信号进行 DCT，就分别得到 MFCC 和 IMFCC。

5) Fisher 选择。

分别从 MFCC 和 IMFCC 中选出 Fisher 比最大的 6 组组合起来，其结果就是混合特征参数。

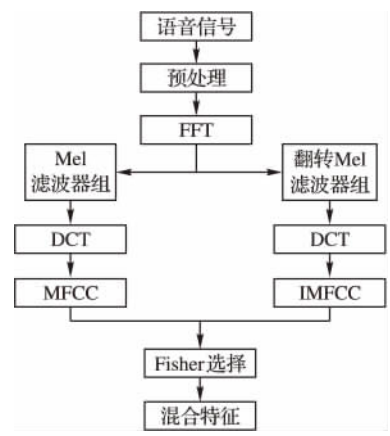

图 5　混合特征的计算过程

4　实验

4.1　实验设计

为了验证本文提出的混合特征参数的有效性及其在噪声环境[13] 中的识别能力，分别用 TIMIT 和 NOIZEUS 语音库进行了两个仿真实验。

实验 1 测试混合特征参数的有效性，采用不含噪声的 TIMIT 语音库中的 dr1(男 28 个，女 18 个)部分作为数据集进行实验。

实验 2 测试混合特征参数在噪声环境中的识别能力，采用 NOIZEUS 语音库，分别在 Airport noise、Babble noise、Car noise、Train Station noise 条件下进行实验。

在实验中首先对语音信号进行预处理，然后对每一帧语音分别提取 MFCC、IMFCC 及其混合特征参数。最后采用高斯混合模型 GMM[14] 作为分类器进行识别，混合度为 16。

4.2 实验结果及分析

实验 1 的结果如表 1 所示。可看出，以 IMFCC 单独作为特征参数没有以 MFCC 的识别率好，但通过本文方法，用 Fisher 准则进行参数选择后的混合特征参数高于 MFCC 的识别率。实验证明，本文提出的混合特征参数用于说话人识别是有效的。

表 1 TIMIT 数据库识别结果 %

特征	测试 1	测试 2
MFCC	93.48	91.30
IMFCC	91.30	89.13
混合特征	95.65	93.48

实验 2 的结果如图 6 所示，由图可知，在噪声环境下，虽然 MFCC 的识别率均高于 IMFCC，但都低于本文提出的混合特征的识别率。同时，还可看出，随着信噪比的增大，识别率也均有所提高。实验结果证明，本文提出的混合特征参数的识别率高于 MFCC。

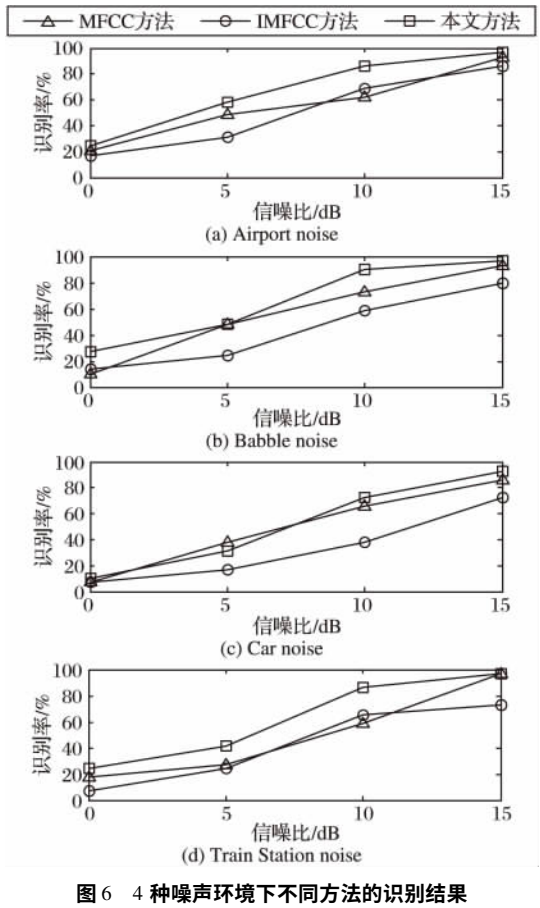

图 6 4 种噪声环境下不同方法的识别结果

5 结语

本文分析了梅尔滤波器组和翻转梅尔滤波器组的结构，根据两者的结构特点，利用 Fisher 准则，构造了一种混合特征参数。仿真实验结果表明，本文提出的混合特征参数与 MFCC 和 IMFCC 相比，在纯净语音及噪声环境下均具有较好的识别性能。

参考文献:

[1] CAMBELL J P. Speaker recognition: a tutorial [J]. Proceedings of the IEEE, 1997, 185(9): 1437 – 1462.

[2] DAVIS S B, MERMELSTEIN P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1980, 28(4): 357 – 365.

[3] 汪峥，连翰，王建军. 说话人识别中特征参数提取的一种新方法[J]. 复旦学报: 自然科学版，2005，44(1): 197 – 200.

[4] 于明，袁玉倩，董浩，等. 一种基于 MFCC 和 LPCC 的文本相关说话人识别方法[J]. 计算机应用，2006，26(4): 883 – 885.

[5] QIAN ZHEN, LIU LI-YAN, LI XUE-YAO. Speaker identification based on MFCC and IMFCC [C]// ICISE: Proceedings of 2009 the 1st International Conference on Information Science and Engineering. Piscataway, NJ: IEEE Press, 2009: 5416 – 5419.

[6] 刘丽岩. 基于 MFCC 与 IMFCC 的说话人识别研究[D]. 哈尔滨: 哈尔滨工程大学，2008.

[7] FISHER R A. The use of multiple measurements in taxonomic problems [J]. Annals of Eugenics, 1936, 7(1): 179 – 188.

[8] ZHU JIAN-WEI, SUN SHUI-FA, DAN ZHI-PING, *et al*. MFCC extraction based on f-ratio and correlated distance criterion in speaker recognition[C]// MINES '09: Proceedings of the 2009 International Conference on Multimedia Information Networking and Security. Washington, DC: IEEE Computer Society, 2009: 329 – 333.

[9] RGOUTAM S, SANDIPAN C, SUMAN S. An f-ratio based optimization technique for automatic speaker recognition system [C]// Proceedings of the IEEE INDICON 2004 India Annual Conference. Piscataway, NJ: IEEE Press, 2005: 352 – 355.

[10] 廖余. 基于混合特征和高斯混合模型的说话人识别研究 [D]. 昆明: 昆明理工大学，2009.

[11] HU YI, LOIZOU P C. Subjective evaluation and comparison of speech enhancement algorithms [J]. Speech Communication, 2007, 49(7/8): 588 – 601.

[12] SANDIPAN C, ANINDYA R, SOURAV M, *et al*. Capturing complementary information via reversed filter bank and parallel implementation with MFCC for improved text-independent speaker identification[C]// Proceedings of the International Conference on Computing: Theory and Applications. Piscataway, NJ: IEEE Press, 2007: 463 – 467.

[13] KUMAR P, JAKHANWAL N, CHANDRA M. Text dependent speaker identification in noisy environment [C]// Proceedings of 2011 International Conference on Devices and Communications (IC-DeCom). Piscataway, NJ: IEEE Press, 2011: 1 – 4.

[14] REYNOLDS D, ROSE R. Robust text-independent speaker identification using Gaussian mixture speaker models [J]. IEEE Transactions Speech and Audio Processing, 1995, 3(1): 72 – 83.