文章编号:1001-5132（2007）03-0297-04

改善含噪语音说话人辨认系统性能的方法

韩春光，胡剑英，李　华

（宁波大红鹰职业技术学院，浙江 宁波 315175）

摘要：当对含噪语音进行说话人辨认时，系统的识别性能会明显变差，本文提出采用对倒谱参数非线性加权的方法，改善系统的噪声鲁棒性. 通过对多种加权窗的正识率比较，发现对 LPC 倒谱低阶参数加权提升，对美尔倒谱高阶参数的加权提升，均提高了系统的识别性能.

关键词：含噪语音；说话人辨认；倒谱参数；非线性加权

中图分类号：TN912.3　　文献标识码：A

说话人识别的基本原理实质就是语音信号的模式识别问题，关键是模式库构建[1]. 目前，影响说话人识别技术实用化的主要方面很多，其中应用环境的噪声以及传输信道的非线性效应等问题是人们研究的热点之一. 对此问题的研究文献很多，如采用维纳滤波改进带噪语音的语音质量，用于说话人识别系统的前端处理，对消除噪声的影响取得了不错的效果[2]；对电话录音的高斯滤波法也使系统的稳健性获得一定的改善[3]；语音提问方式的使用，也为克服电话信道的非线性效应提供了一种可能的方法[4]. 本文提出了用含噪语音倒谱参数非线性加权的方法[5,6]，即充分体现了特征参数中各阶分量的不同作用，而且算法简单，便于实现.

1　特征参数的选取

说话人识别是一种特殊的语音识别，其目的不是识别说话内容，而是识别说话人. 说话人识别注重从语音信号中提取个人特征，即提取语音信号中所包含的个性因素. 其关键的问题是：究竟用语音信号的哪些特征或特征变换描述说话人才有效且可靠？显然，这些特征应该具有区分性、稳定性和独立性. 描述说话人特征的参数很多，较常用和有效的特征量是 LPC 倒谱参数和 MFCC 倒谱参数.

1.1　LPC 倒谱(LPCC)

LPC倒谱是对语音信号的线性预测模型进行同态分析得到的参数，在实际应用中，为了避免复对数运算带来的相位卷绕问题，其实常使用倒谱作为特征参数. LPC倒谱参数可以由LPC系数按照递推公式直接推得，其递推过程如下[1]：

设对语音信号线性预测分析得到的声道模型系统传输函数为：

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

其冲击响应为 $h(n)$，则

$$H(z)=\sum_{n=0}^{\infty}h(n)z^{-n}. \tag{2}$$

根据倒谱定义，有

收稿日期：2007-04-08.　　宁波大学学报（理工版）网址：http://3xb.nbu.edu.cn
基金项目：浙江省教育厅科研基金（20050099）.
作者简介：韩春光（1966－），男，河北滦县人，副教授，主要研究方向：语音信号处理及通信与信息系统. E-mail: hancg66@163.com

$$\hat{H}(z)=\ln H(z)=\sum_{n=0}^{\infty}\hat{h}(n)z^{-n}. \tag{3}$$

将(1)式代入(3)式并在等式两边对 z^{-1} 求导，即

$$\frac{\partial}{\partial z^{-1}}\ln\left[\frac{1}{1-\sum_{k=1}^{p}a_k z^{-k}}\right]=\frac{\partial}{\partial z^{-1}}\sum_{n=1}^{\infty}\hat{h}(n)z^{-n}, \tag{4}$$

得到：

$$\sum_{n=1}^{\infty}n\hat{h}(n)z^{-n+1}=\frac{\sum_{k=1}^{p}ka_k z^{-k+1}}{1-\sum_{k=1}^{p}a_k z^{-k}}, \tag{5}$$

则：

$$\left(1-\sum_{k=1}^{p}a_k z^{-k}\right)\sum_{n=1}^{\infty}n\hat{h}(n)z^{-n+1}=\sum_{k=1}^{p}ka_k z^{-k+1}. \tag{6}$$

令(6)式左右两边相应系数相等，则得到 $h(n)$ 和 a_k 之间的递推关系如(7)式所示.

$$\begin{cases}\hat{h}(1)=-a_1,\\ \hat{h}(n)=-a_n-\sum_{k=1}^{n-1}(1-k/n)a_k\hat{h}(n-k),1\leqslant n\leqslant p,\\ \hat{h}(n)=-\sum_{k=1}^{p}(1-k/n)a_k\hat{h}(n-k),n>p.\end{cases} \tag{7}$$

$h(n)$ 包含语音信号频谱中的包络信息，可以近似把 $h(n)$ 当作语音信号 $s(n)$ 的短时倒谱 $\hat{s}(n)$. 通过对 $h(n)$ 的分析，可分别估计出语音短时谱包络和声门激励参数，通常称 $h(n)$ 为 LPC 倒谱.

1.2 美尔(Mel)频谱

与LPC倒谱分析不同的是，Mel频率倒谱参数(MFCC)的分析着眼于人耳的听觉机理，依据听觉实验的结果来分析语音的频谱，使之更加符合人耳对频率高低的非线性心理感觉. 心理学研究发现，人类听觉系统对声音频率高低的感觉不是一个线性过程，因此提出了一个对频率主观心理感知的量—音高，在响度级为 40 dB情况下，我们把频率1 000 Hz的纯音定义为 1 000 MelHz. Stevens等人的研究显示，声音频率在 1 000 Hz以上时，随主观心理感知音高的线性增大与声音频率呈对数关系，即Mel频率刻度与声音频率呈对数关系，可近似表示如下[7]：

$$\mathrm{Mel}(f)=2\,595\lg(1+f/700), \tag{8}$$

式中，f 为频率；$\mathrm{Mel}(f)$ 为 Mel 频率. 同样是由心理学研究发现，在声压恒定的情况下，当噪声被限制在某个带宽内时，人耳感觉的主观响度是恒定的；而一旦噪声突破了这个带宽，则主观响度的变化便会被感知. 同样地，当声压恒定时，在这个带宽内的一个具有复杂包络的信号响度等价于在这个带宽中心频率位置的一个纯音的响度，而与信号本身的频率分布无关，但是当信号的带宽突破了此带宽时，其响度便不再等价，此带宽称为临界带. 根据上述关系式(8)以及临界带的划分，可将语音频域划分成一系列三角形的滤波器序列，即所谓 Mel 滤波器组. 每个滤波器的 Mel 频率刻度的带宽是恒定的，通常取带宽为 300 mels，间隔 150 mels，而频域的带宽则随频率增加而成对数增加，如图 1 所示. 图中的 m_i 是第 i 个滤波器输出的所有信号幅度加权求和后的对数谱能量，又称为 Mel 频谱.

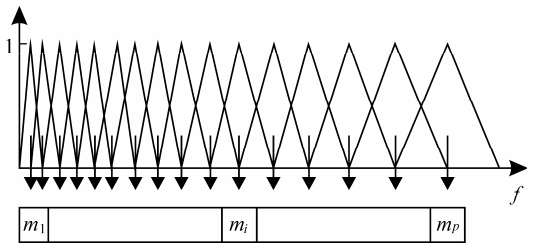

图 1 临界带滤波器组

1.3 美尔倒谱参数(MFCC)

在计算 MFCC 参数的过程中，首先对各帧语音信号作离散付氏变换，再将所得频谱的实频谱送入 Mel 滤波器组滤波，取每个临界带内所有信号频谱幅度加权和作为该临界带滤波器的输出，然后对所有滤波器输出作对数能量运算，形成该帧 Mel 频谱矢量 m_i，最后对 m_i 作离散余弦变换得 MFCC 参数，实现过程如图 2 所示.

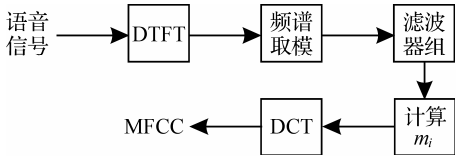

图 2 MFCC 参数实现

可见，MFCC 参数的具体计算过程也是一个同

态分析过程. 设语音信号为 $s(n)$，滤波器组频率特性为 $H_i(\omega)$，则图(2)中信号经滤波器组滤波后为：

$$m_i = \log\left(\sum_{k_i=0}^{K_i}|S(\omega)|H_i(\omega)\right), \tag{9}$$

式中，K_i是第i个滤波器覆盖的范围，对上式做离散余弦变换后得到MFCC参数 c_j ：

$$c_j = \sqrt{\frac{2}{p}}\sum_{i=1}^{p} m_i \cos(\frac{\pi j}{p}(i-0.5)), \tag{10}$$

其中， p 为滤波器组数； j 为 MFCC 参数的阶数.

2　倒谱参数加权

考虑到各阶特征参数在语音识别中的贡献大小不同，以及在噪声环境下，高阶MFCC参数对说话人识别贡献大的特点[5]，经Matlab仿真实验，分别对LPC倒谱和MFCC参数进行多种加权窗口对比(如图 3、图 4 所示)，发现对LPC倒谱参数加半余弦窗(wc)，对MFCC参数加半正弦窗(ws)，均改善了系统性能.

3　实验

3.1　实验系统

本文采用的说话人辨认系统的组成系统如图 5 所示. 在对语音信号分帧处理时，对 LPCC 取每帧 320 个点(40 ms)，帧移 160 点(20 ms)；计算 MFCC 时取每帧 256 个点(32 ms)，帧移 128 点(16 ms). 预加重系数为 0.95，并均加哈明窗处理.

3.2　实验条件

本实验共使用了 60 人的语音，其中包括 24 个男声，36 个女声. 在室内普通噪声环境下，使用笔记本电脑录制. 统计分析结果，该语音库中各种语音的平均信噪比约为 25 dB. 实验中，在每个说话人的语音库中随机取 16 个短语，长约 33～42 s，用于训练；再另外随机取 6 个短语，长约 13～17 s，用于测试识别. 计算倒谱参数时，LPC 系数取 16 阶，LPCC 取 20 维；Mel 滤波器组个数取 20，MFCC 参数取 12 阶.

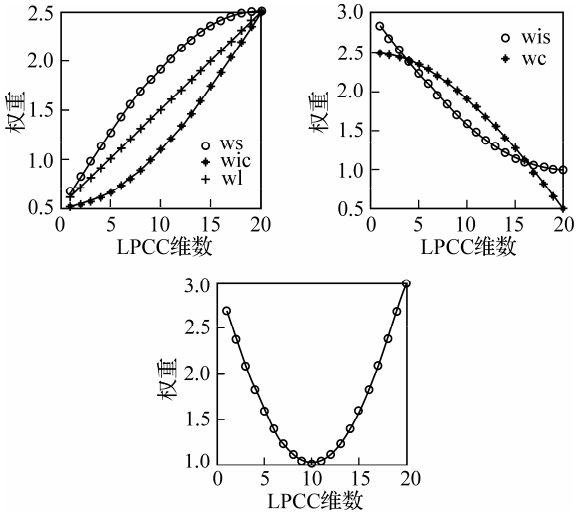

图 3　LPC 倒谱参数加权窗

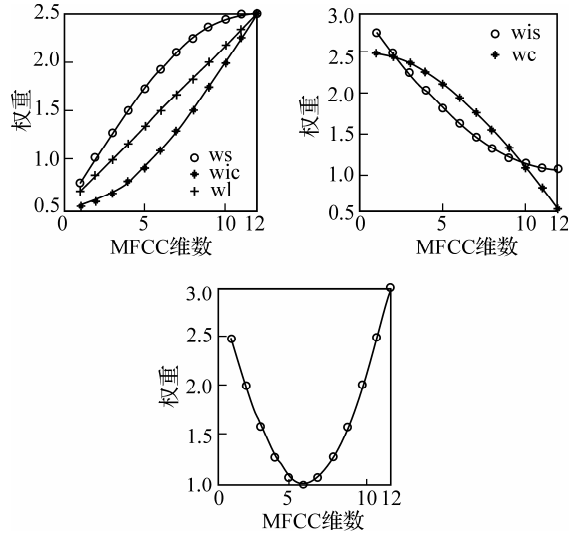

图 4　MFCC 参数加权窗

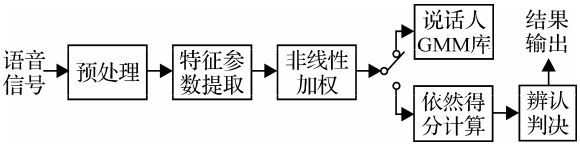

图 5　说话人辨认系统

高斯混合模型 GMM 的初始权值取各单高斯分布权重均等，取混合数的倒数即 1/nmix；协方差矩阵取满秩矩阵，各元素初始值均设定为 1；各高斯分布的初始均值由程序按正态分布随机给定. 用 EM 算法对 GMM 模型进行训练，最大迭代次数为 10.

3.3 实验结果

实验中通过编程对每帧语音引入零均值的高斯加性白噪声，使语音信号的信噪比由 25 dB 下降到 15 dB 和 5 dB 的水平. 结果显示：对 LPC 倒谱参数加半余弦窗，提升其低阶系数比重时，在噪声环境下系统识别率明显改善，而对 MFCC 参数加半正弦窗，提升其高阶参数时，同样显著改善了噪声环境下系统识别性能，如表 1 所示.

表 1 系统识别率比较

SNR	25 dB	15 dB	5 dB
LPCC / %	81.7	65.0	43.3
LPCC+wc / %	83.3	70.0	55.0
MFCC / %	80.0	61.7	31.7
MFCC+ws / %	85.0	70.0	45.0

由表 1 看到，对于语音信号信噪比为 15 dB 和 5 dB 时，加权后 LPCC 参数的系统识别率比加权前分别提高了 5%和 11.7%，而 MFCC 系数则分别提高了 8.3%和 13.3%，效果都非常明显.

4 总结

通过倒谱加权对倒谱中不同成分提升或抑制是改善特征性能的有效方法. 半余弦加权窗对 LPCC 参数低阶成分提升后，改善了噪声环境中说话人识别系统性能，说明 LPCC 低阶参数的鲁棒性能较好. LPCC 低阶参数反映的是语音短时谱的包络信息，主要包含语音共振峰和声门激励信息，因此，这部分参数包含更多的说话人个性特征. 实验结果也说明了这一点. 半正弦加权窗对 MFCC 高阶参数成分的提升，同样改善了系统的性能，这说明 MFCC 参数高阶成分的鲁棒性能较好，实验结果显示，增加高阶参数在特征量中的比重有利于改善含噪语音辨认系统的识别性能.

参考文献：

[1] 易克初, 田斌, 付强. 语音信号处理[M]. 北京: 国防工业出版社, 2001.
[2] 白俊梅, 张世磊, 张树武, 等. 噪声环境下的鲁棒性说话人识别[J]. 中文信息学报, 2006, 20(1):91-97.
[3] 周静芳, 陈一宁, 李科, 等. 基于高斯语音滤波的稳健文本无关说话人识别[J]. 计算机工程, 2005, 31(2):179-181.
[4] 朱民雄, 闻新, 黄健群, 等. 计算机语音技术[M]. 北京: 北京航空航天大学出版社, 2002.
[5] Rabiner L, Juang B H. 语音识别基本原理[M]. 影印版. 北京: 清华大学出版社, 1999.
[6] 许雯, 董林, 田家斌. 一种改进的高斯混合模型算法[J]. 信息工程大学学报, 2005, 6(2):65-67.
[7] Young S, Evermann G, Gales M, et al. The HTK book (for HTK version 3.4)[EB/OL]. [2006-12-06]. http://htk.eng.cam.ac.uk/.

A Method to Improve Performance of Speaker Identification System with Noise

HAN Chun-guang, HU Jian-ying, LI Hua

(Ningbo Dahongying Vocational Technical College, Ningbo 315175, China)

Abstract: The performance of the speaker recognition system tends to be deteriorated in the noisy speech cirsumstances. In this paper, a nonlinear weighting method is proposed for selecting the cepstral coefficient in an effort to improve the system robustness to noise. Having compared with a variety of recognition rates obtained in weighting windows, it is found that the system recognition performance is improved along with the increase of weighting of the low-order cepstral term of LPCC and the high-order cepstral term of MFCC.

Key words: the speech with noise; speaker identification; cepstral coefficient; nonlinear weighting

CLC number: TN912.3 **Document code:** A

（责任编辑 章践立）