

文章编号:1001-5132(2007)01-0051-04

基于小波包分析的特征参数提取

胡文吉, 王让定

(宁波大学 纵横智能软件研究所, 浙江 宁波 315211)

摘要: 在分析 MFCC 提取原理的基础上, 结合小波包分析理论, 得到新的特征参数. 提出了一种新的特征参数提取方法, 用动态时间规整法, 分别测试了 MFCC 的识别率和新的特征参数的识别率. 研究证明新的特征参数不仅具有较高的识别率, 而且有一定的抗噪声能力.

关键词: 小波包; 美尔倒谱系数; 特征参数; 动态时间规整

中图分类号: TP309

文献标识码: A

随着计算机处理能力的大幅度提高, 语音识别越来越受到人们的关注. 提取语音特征参数是语音识别的重要步骤, 特征参数的好坏直接影响识别效果. 在过去几十年的研究中, 人们提出了许多特征参数, 如基音周期^[1]、共振峰频率、线性预测系数(LPC)、线谱对(LSP)、倒谱系数等^[2], 目前使用最为广泛的特征参数是基于全声道全极点模型的线性预测倒谱系数(LPCC)和基于人耳听觉模型的美尔倒谱系数(MFCC)^[3].

虽然MFCC得到了更广泛的应用, 但MFCC也有其自身的缺点, 由于MFCC是基于语音信号短时平稳的假设, 在短时傅里叶变换的基础上提取的^[4]. 实际上语音信号是一种典型的非平稳信号, 它的频谱特性随时间而改变. 近年来, 小波变换已经成功地应用于语音编码, 在语音识别中, 用小波变换来提取特征参数还处于研究阶段, 大部分研究只是提取直接小波系数用作识别^[5]. 研究表明: 直接将小波系数作为特征参数, 其识别率较低, 只有 80% 左右, 但具有较好的噪声鲁棒性. 而小波包分析是

小波分析的推广, 它能为信号提供一种更精细的分析方法, 将频带进行多层次划分.

本文在文献[6,7]基础上, 借鉴MFCC的提取过程, 结合小波包对语音频带的多层次划分, 并根据人耳听觉特性, 选择相应频带, 提出了一种基于小波包的特征参数. 在识别算法上, 用动态时间规整(DTW)法^[8], 分别测试了MFCC的识别率和新的特征参数的识别率.

1 特征参数提取

1.1 MFCC 的提取

MFCC 基于人耳的听觉特性, 将频谱转化为基于 Mel 频标的非线性频谱, 然后转换到倒谱域上. Mel 频率 F_{Mel} 与实际频率 f_{actual} 的关系可由公式(1)表示:

$$F_{\text{Mel}} = 2595 \lg(1 + f_{\text{actual}} / 700). \quad (1)$$

MFCC 定义为语音信号经过快速傅里叶变换后所得的加窗短时信号的实倒谱. 一般 MFCC 的

计算过程为:(1)对预处理后的每一帧语音信号进行快速傅里叶变换得到其频谱;(2)求它的频谱幅度的平方,即能量谱,并用1组三角形滤波器在频域对能量谱进行带通滤波,中心频率从 $0 \sim F_{\text{Mel}}/2$ 间按Mel频率分布;(3)对滤波器组的输出取对数,再进行离散余弦变换,得到MFCC参数.MFCC的具体计算过程如图1所示.

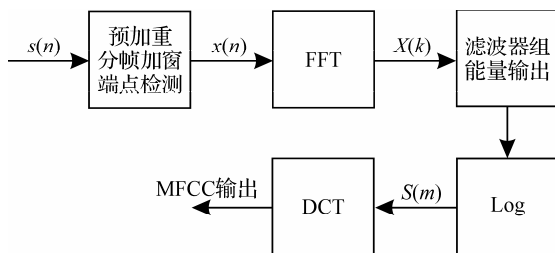


图1 MFCC计算过程

1.2 新特征参数的提取

小波变换被称为数学的显微镜^[9],具有分辨率可变、实现简单和无平稳性要求等诸多优点,它通过有限个基函数在尺度-频率域上对信号进行分析,在控制分辨率的同时,保留了时域信息,因此在时变信号的处理上受到了极大关注.并且小波变换在各分析频段的恒 Q (品质因数)特性与人耳听觉对信号的加工特点相一致,这一良好的特性为利用小波变换提取语音特征参数奠定了基础.

小波变换中的多分辨分析其最终目的是力求构造一个在频率上高度逼近 $L^2(R)$ 空间的正交小波基,这些频率和分辨率不同的正交小波基相当于带宽各异的带通滤波器.小波包分析就是在多分辨分析的基础上,对高频信号部分进一步分解,这样根据被分析信号的特性,从而提取感兴趣频带的信息,为信号提供了1种更为精细的分析方法.

MFCC参数提取过程是把信号的频谱通过一系列的带通滤波器,求得每个滤波器输出的对数能量,再进行DCT变换得到的.从滤波器的角度看,小波包分析就是信号通过一系列不同频带范围的带通滤波器,从而获得信号在各频带内的信息.MEL滤波器组与小波包分析的作用在某种意义上可以互换,用小波包分析来代替FFT与MEL滤波

器组,然后再进行DCT变换,从而得到一组新的语音特征参数.

由于人耳对语音信号的感知是非线性的,MEL滤波器组在频带上的分布及带宽类似于临界频带的划分,这对选取小波包分析后的结点频带具有指导意义.因此,信号在经过小波包分析后,应选取出与MEL滤波器组频带相一致的范围内的的小波包系数.新的特征参数的提取计算可由图2表示.

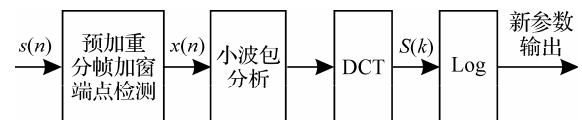


图2 新参数的计算过程

(1) 对输入的语音信号进行预处理.

预加重:用以提升高频部分,使信号的频谱变得平坦,预加重通常使用一阶数字滤波器实现,即

$$H(z) = 1 - \mu z^{-1}, \quad (2)$$

式中, μ 的典型值为0.94.

分帧、加窗:这里的分帧是为了提高特征参数提取的精度,而不是为了满足平稳假设的条件.

$$w(n) = 0.54 - 0.46 \cos 2\pi n / (N-1), \quad (3)$$

$$0 \leq n \leq N-1.$$

端点检测(短时能量与过零率):去除静音段,提取有用帧,减少计算量.

(2) 对各帧信号进行小波包分解,得到各子带系数 $WPX_{k,m}$,其中 $WPX_{k,m}$ 为第 i 个子带的第 m 个小波系数.

(3) 计算所选结点频带内信号的对数能量:

$$S_k = \ln \frac{\sum_{m=1}^{N_k} [WPX_{k,m}]^2}{N_k}, \quad k = 1, 2, \dots, K, \quad (4)$$

其中: k 为子带序号; N_k 为第 k 个子带中小波包分解系数的个数; K 为所选频带的数目,在本文的实验中, $K = 24$.

(4) 将上述 S_k 经DCT得到新的特征参数.

$$c(i) = \sum_{k=1}^K S_k \cos \left[\frac{\pi i (k-1/2)}{k} \right], \quad (5)$$

$$1 \leq k \leq K, i = 1, 2, \dots, p,$$

其中, p 是参数的维数, 由小波包分解后选择的结点数决定.

从以上新的特征参数的提取过程可以看出, 本文提出的新型特征参数是结合了 MFCC 的提取思想与小波包分析而产生的, 根据临界带宽的划分, 用相应的小波包分析结点频带内的信号能量来取代 MFCC 提取时 MEL 滤波器输出的信号能量, 因此免去了对信号的 FFT 变换及 MEL 滤波器的设计, 而用小波包分析并选取适当的小波包结点来完成. 从整体上来看, 新的特征参数与 MFCC 的提取很相似, 前端与后端处理基本一致, 只是中间有所差异, 但其中间参数的性质是一致的, 因而与 MFCC 具有相同的属性, 均可归于信号的倒谱系数. 而其中的小波包分析只是提取过程中的一步, 得出的小波包系数只是 1 个中间参量, 因此新的特征参数与直接小波系数完全不同.

2 实验结果

第一组实验的语音样本均在相对安静的环境下取得, 用 11 025 Hz 采样, 16 bit 量化, 包含 20 个不同长短的词(词组), 每个词(词组)朗读多遍, 直到得到一对一致性较好的特征矢量序列. 最终得到的模板是在一致性较好的特征矢量序列对沿

动态时间规整(DTW)的路径上求平均, 建立模板. 然后再录 10 组同样的语音用做识别.

为了比较 MFCC 与新的特征参数的性能, 又进行了 2 组实验. 第一组实验提取语音的 MFCC, 根据临界带宽的划分, 选 Mel 滤波器的个数为 24, 语音帧长 256 个采样点, 帧移 80 个采样点; 第二组实验用小波包分析提取特征参数, 这里小波函数用 db2, 语音帧长也是 256 个采样点, 帧移 80 个采样点. 实验中, 进行六级小波包分解, 选择 24 个小波包结点. 表 1 显示选取的小波包结点及其对应的频率范围. 第一组 MFCC 识别率为 81.5%, 而新的特征参数识别率为 94.0%.

第二组实验中的语音样本有一定的噪声干扰, 识别方法与第一次实验相同. 第二组 MFCC 识别率为 81.5%, 而新的特征参数的识别率为 86.5%.

由以上实验可以得出以下结论:

(1) 在相对安静的环境下, MFCC 和新的特征参数的识别率比较接近, 新的特征参数的识别率略高于 MFCC 的识别率;

(2) 在有一定噪声干扰的环境下, 新的特征参数的识别率高于 MFCC, 说明用小波包提取的特征参数具有一定的抗噪声能力;

(3) 从计算量上来讲, 用小波包分析代替了 FFT 变换与 MEL 滤波器组的设计, 因此新的参数

表 1 小波包分解结点选取

小波包结点	频率范围 / Hz	中心频率 / Hz	小波包结点	频率范围 / Hz	中心频率 / Hz
[6,0]	0 ~ 86	43	[5,8]	1 367 ~ 1 548	1 462
[6,1]	86 ~ 172	129	[5,9]	1 548 ~ 1 720	1 634
[6,2]	172 ~ 258	215	[5,10]	1 720 ~ 1 892	1 806
[6,3]	258 ~ 344	301	[5,11]	1 892 ~ 2 064	1 978
[6,4]	344 ~ 430	387	[4,6]	2 064 ~ 2 408	2 236
[6,5]	430 ~ 516	473	[4,7]	2 408 ~ 2 752	2 580
[6,6]	516 ~ 602	559	[4,8]	2 752 ~ 3 096	2 924
[6,7]	602 ~ 688	645	[4,9]	3 096 ~ 3 440	3 268
[5,4]	688 ~ 860	774	[4,10]	3 440 ~ 3 784	3 612
[5,5]	860 ~ 1 032	946	[4,11]	3 784 ~ 4 128	3 956
[5,6]	1 032 ~ 1 204	1 118	[3,6]	4 128 ~ 4 816	4 472
[5,7]	1 204 ~ 1 367	1 290	[3,7]	4 816 ~ 5 504	5 160

的计算量与 MFCC 相当. 在此选取新参数与 MFCC 都是 12 维的参数, 因此最终两者的存储量是相同的.

由以上分析得出, 经过小波包分析提取的特征参数, 其性能还是比较理想的.

3 结论

本文借鉴 MFCC 提取的方法, 用小波包分析来替代快速傅里叶变换和三角带通滤波器组, 来提取语音的特征参数, 并用动态时间规整法进行语音识别. 实验证明: 新的特征参数具有较高的识别率, 并具有一定的抗噪声能力.

参考文献:

- [1] 赵瑞珍, 宋国乡. 基音检测的小波快速算法[J]. 电子科技, 1998, 43(1):16-19.
- [2] Hong Kook Kim, Rose R C. Cepstrum-domain model combination based on decomposition of speech and noise

for noisy speech recognition[J]. Acoustics, Speech, and Signal Processing, 2002(1):209-212.

- [3] 王让定, 柴佩琪. 语音倒谱特征的研究[J]. 计算机工程, 2003, 29(13):31-33.
- [4] 武妍, 金明曦, 王洪波. 基于 KL - 小波包分析的文本无关的说话人识别[J]. 计算机工程与应用, 2005(4):26-28.
- [5] 赵铮, 侯伯亨. 基于小波变换说话人识别技术的研究[J]. 西安电子科技大学学报, 2000, 27(4):437-441.
- [6] Engin Avci, Zuhtu Hakan Akpolat. Speech recognition using a wavelet packet adaptive network based fuzzy inference system[J]. Expert Systems with Applications, 2006, 31(3):495-503.
- [7] 汪铮, 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, 2005, 44(1):197-200.
- [8] 刘敬伟, 徐美芝, 郑忠国, 等. 基于 DTW 的语音识别和说话人识别的特征选择[J]. 模式识别与人工智能, 2005(1):50-54.
- [9] 飞思科技. 小波分析理论与MATLAB7 实现[M]. 北京: 电子工业出版社, 2005.

Extraction of Feature Coefficient Based on Wavelet Packet Analysis

HU Wen-ji, WANG Rang-ding

(CKC Software Lab, Ningbo University, Ningbo 315211, China)

Abstract: In the paper, a method of new feature factor extraction is proposed. Based on the MFCC extraction theory and combining with wavelet packet analysis, a new feature factor is obtained. Using DTW, we have tested the speech recognition rate of MFCC and the new feature factor. The experiment suggests that the new feature factor achieves not only high recognition rate, but also anti-noise capacity.

Key words: wavelet packet; Mel frequency cepstral coefficient; feature coefficient; dynamic time warping

CLC number: TP309

Document code: A

(责任编辑 章践立)