

融合 SS、MFCC 和 PMC 技术的语音去噪方法

丁冬冬,余玉梅,江涛,庄丽,王米利,刘敬凤
(云南民族大学 数学与计算机科学学院,云南 昆明 650031)

摘要: 为提高语音识别系统在噪音情况下的识别率,提出了一种融合信号级去噪、参数级去噪、模型级去噪的方法. 首先用谱减法对带噪的语音信号进行去噪,再利用 Mel 倒谱系数(MFCC)对处理后的语音信号进行特征提取,最后经过并行模型结合处理法(PMC)处理得到较高识别率的语音信号.

关键词: 语音信号;谱减法;MFCC;PMC;信噪比

中图分类号: TN912.3 **文献标志码:** A **文章编号:** 1672-8513(2014)03-0232-03

A denoising method based on an integration of SS, MFCC and PMC technology

DING Dong-dong, SHE Yu-mei, JIANG Tao, ZHUANG Li, WANG Mi-li, LIU Jing-feng

(School of Mathematics and Computer Science, Yunnan University of Nationalities, Kunming 650031, China)

Abstract: In order to improve the recognition rate of the speech sound system under noisy circumstances, one special method is put forward in this paper, which integrates the methods of the signal denoising, parameter denoising and model denoising. This method first employs spectral subtraction to denoise the system and then extracts the features with the aid of MFCC. Finally, after PMCC-processing, the speech sound signal of a higher recognition rate is attained.

Key words: speech signal; spectral subtraction; MFCC; PMC; signal-to-noise rate

在现实应用中,噪声下语音识别的研究就变得越来越重要. 语音识别技术主要包括特征提取、模式匹配及模型训练等 3 个方面. 噪音处理有 3 种常用方法^[1-2]:信号级抗噪方法、特征参数级抗噪方法、模型级抗噪方法. 这 3 种方法的应用都针对某种特定环境下的语音去噪,有很大的局限性. 本文方法融合了这 3 种方法,有较强的适应环境的能力,具有更好的实用性.

1 常用去噪方法

1.1 信号级抗噪处理方法

信号级抗噪处理方法是从小带噪音的语音信号中提取较纯净的原始语音,也称为语音信号处理中的语音增强. 对于不同的噪音,必须采取不同的语音处理方法. 现在语音识别系统中运用比较多的有最小均方差法(MMSE)、谱减法(SS)及其改进形式、维纳滤波法、中值滤波法等等. 本文重点介绍谱减法.

在谱减法中,假定噪音是加性的,所以其思想是在频域上从带噪语音的功率谱中减去噪音的功率谱,从而得到比较纯净的语音频谱. 它的实现过程是先把语音信号经过快速 FFT 变换,然后平方得到语音幅度估计,将其相位恢复后再采用逆 FFT 变换恢复时域信号^[3]. 信号恢复的模型为:

$$y(t) = s(t) + n(t). \quad (1)$$

其中 $y(t)$ 表示含噪语音, $s(t)$ 表示纯语音信号, $n(t)$ 表示噪音信号.

对上式进行 FFT 变换得到

$$Y_w(w) = S_w(w) + N_w(w). \quad (2)$$

其功率谱有

$$|Y_w(w)|^2 = |S_w(w)|^2 + |N_w(w)|^2 + S_w(w)N_w^*(w) + S_w^*(w)N_w(w). \quad (3)$$

由于 $s(t)$ 和 $n(t)$ 相互独立,所有 $S_w(w)$ 与 $N_w(w)$ 也相互独立,而 $N_w(w)$ 为零均值的高斯分

收稿日期:2013-12-03.

基金项目:国家自然科学基金(61363022);云南省自然科学基金(2011FZ176);云南省计算机科学与技术专业核心课程教学团队项目.

作者简介:丁冬冬(1988-),男,硕士研究生. 主要研究方向:自然语言处理.

通信作者:余玉梅(1965-),女,教授,硕士生导师. 主要研究方向:人工智能.

布,所以原始语音的估值为

$$|\hat{S}_w(w)| = [|Y_w(w)|^2 - E[|N_w(w)|^2]]^{\frac{1}{2}} = [|Y_w(w)|^2 - \lambda_n(w)]^{\frac{1}{2}}. \quad (4)$$

其中 $\lambda_n(w)$ 是无语音时 $|N_w(w)|^2$ 的统计平均.

1.2 特征参数级抗噪处理方法

目前,大部分的语音识别系统使用的特征参数为倒谱类参数.如线性预测系数倒谱(LPCC)、Mel倒谱系数(MFCC)、共振峰、基音周期等.其中LPCC和MFCC的应用最为广泛.

Mel倒谱系数(MFCC)是基于人耳听觉特征的,主要思路是把语音信号的频谱通过滤波器转换成基于Mel频率尺度的非线性频谱,再对滤波器的输出进行对数和离散余弦变换,就可以得到MFCC系数,它与频率的关系可以近似表示为^[4]:

$$\text{mel}(f) = 2595 \lg(1 + \frac{f}{700}). \quad (5)$$

其中频率 f 的单位是 Hz. MFCC 分析是从人耳的听觉机理,依据实验结果来分析语音频谱,能够获得比较高的识别率和比较好的鲁棒性,具体提取步骤见文献[5-6].

1.3 模型级抗噪处理方法

模型级抗噪处理方法有2类,第1类是用于测试具有相同环境的少数数据做模型且快速适应,即自适应方法;第2类是直接在识别模型中增加对环境噪声的处理.目前模型级抗噪处理方法中,效果比较好的是并行模型结合处理法(PMC).PMC是基于模型的噪声补偿算法,通过引入噪声的统计知识,调整用纯语音训练出的模型参数,也就是隐马尔科夫模型(HMM)各个状态的概率密度输出函数的均值和方差,使其反映识别时实际遇到的带噪语音的统计特征.其过程图如图1^[7-10].

图1中先对HMM的参数进行IDCT(逆离散余弦变换),从倒谱域转换到对数谱域,而此时它们在对数谱域输出的概率密度函数很接近高斯分布,或者叠加的高斯分布,所以这样就可以降低复杂度.合并后模型的参数再通过DCT(离散余弦变换)转换到倒谱域用于识别.

2 改进的去噪方法

以上3种方法对噪声处理有一定的效果,但是自身都有不足.其中信号级去噪很难区分清辅音和宽带噪声等;参数级去噪的缺点是噪声的时变性很强,很难直接去除噪声等;模型级去噪仅限于对噪声模型的自适应等.由于这些方法自身的缺陷,所以单一的去噪方法对于语音的去噪效果并不能达到实际要求的效果.

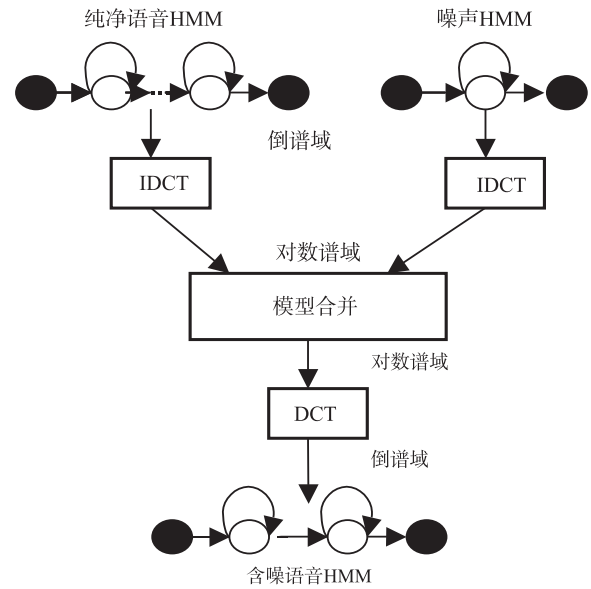


图1 PMC原理图

鉴于这3种方法的互补,本文提出一种改进的去噪方法,对这3种方法进行综合应用,具体操作如下:

1) 首先对带噪语音进行信号级去噪处理,本文选择其中的谱减法.因为谱减法的算法简单、运算量小,能够实现快速处理,适应性比较强.普通的谱减法去噪会出现残留而呈现出“音乐噪声”^[11],所以在此方法基础上进行了改进.通过频谱相减时给噪声谱乘上一个大于1的参数,这样频谱相减时减去的值比估计的噪声谱多,如式(6).

$$|S_w(w)| = [|Y_w(w)|^\alpha - \beta \lambda_n(w)]^{\frac{1}{\alpha}}. \quad (6)$$

其中 α 和 β 是调节参数,适当调节它们可以达到较好的去噪效果.

2) 再对去噪后的语音进行特征参数级去噪处理,本文选择MFCC方法.

对谱减法处理后的带噪语音先进行MFCC特征提取,经过FFT变换后得到Mel滤波器组,滤波器组的个数为 M ,三角滤波器的中心频率是 $f(m)$,它们之间的间隔随 m 的增加而增加.

三角函数滤波器函数为:

$$H_m = \begin{cases} 0, & k < f(m-1); \\ \frac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \leq k < f(m); \\ \frac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) \leq k \leq f(m+1); \\ 0, & k > f(m+1). \end{cases} \quad (7)$$

其中: $0 < m \leq M$; $f(m) = (\frac{N}{F_s})f_{\text{mel}}^{-1}(f_{\text{mel}}(f_l) + m \frac{f_{\text{mel}}(f_h) - f_{\text{mel}}(f_l)}{M+1})$, f_l 和 f_h 是滤波器应用范围的

最低频率和最高频率, N 是 FFT 长度, F_s 是采样率, $f_{\text{mel}}^{-1}()$ 是 $f_{\text{mel}}()$ 的逆函数. 再对经过 Mel 频率滤波器的信号去对数, 然后进行 DCT 变换, 就可以得到 MFCC 的特征参数.

3) 最后再对去噪后的语音进行模型级去噪处理, 本文选择 PMC 技术. 因为 PMC 也是假设噪音和纯净语音线性叠加, 而且此方法能够适用于非平稳特点的噪声, 并且可以在不重新进行训练的条件下接近在噪声背景下训练得到的语音模型的性能.

对于上述处理过的带噪语音进行估计求解, 下面是最大似然估计方程式:

$$\hat{H}_R^{\text{cep}} = \arg_{H_R^{\text{cep}}} \min \left[\left| \mu^{\text{cep}}(H_R^{\text{cep}}) - 1/T \sum_{\tau=1}^T \hat{s}^{\text{cep}}(\tau) \right| \right]. \quad (8)$$

其中, $\mu^{\text{cep}}(H_R^{\text{cep}})$ 是带噪模型的均值参数.

然后用 PMC 技术与干净的语音模型进行合成, 得到一个干净语音模型, 利用这个模型进行识别, 能够得到比较好的识别率. 过程中 PMC 的处理可以表示为:

$$\lambda_{p-s}^{\text{cep}} = DCT \{ \lg [\exp(IDCT(\lambda_s^{\text{cep}} + g \cdot \lambda_R^{\text{cep}}))] \}, \quad (9)$$

其中 λ 是模型参数, $p-s$ 是伪干净语音数据, g 是一个加权控制因子. 处理的参数是基于高斯分布的均值和方差矢量, 参数的变换方法采用对数正态近似、对数求和近似的方法.

3 实验结果

实验中采用的语音材料来自海天瑞声科技有限公司的数据库, 选择了其中 10 人的 100 个词语录音, 实验添加的噪音信号为高斯噪音. 表 1 中的不同信噪比(-5 dB, 0 dB, 5 dB, 10 dB, 15 dB, 20 dB) 都是由纯净的语音信号和噪音信号线性相加而成的. 然后对噪声语音信号采取 Hamming 窗进行分帧, 每帧时间为 25 ms, 叠加的时间为 15 ms. 再利用 Matlab 编程对这几种不同信噪比的语音信号进行 SS、MFCC、PMC 处理, 几种联合算法及本文算法实验结果如表 1.

由表 1 可以看出, 在信噪比较低时, 系统的识别率比较低, 在信噪比较高时, 语音增强能有比较好的鲁棒性, 识别率比较高. 相对而言, 在不同信噪比的情况下, 单一的去噪方法及这几种联合算法识别率都不是很高, 而改进的算法能够提高系统的识别率. 因此改进的方法能够强于任何一种单独算法及以上的联合算法.

表 1 几种方法在不同信噪比下的识别率 %

方法	-5 dB	0 dB	5 dB	10 dB	15 dB	20 dB
SS	17	36	51	81	91	95
MFCC	24	68	70	75	82	87
PMC	21	43	84	85	86	89
SS - MFCC	46	70	72	80	83	89
SS - PMC	40	68	80	86	89	92
本文方法	61	83	92	95	97	98

4 结语

语音去噪一直是语音识别研究中一个重要的难题, 信号级抗噪方法、特征参数级抗噪方法、模型级抗噪方法的应用都对于某种特定环境下的语音去噪, 有很大的局限性. 本文提出了一种结合信号级去噪方法、参数级去噪方法、模型级去噪方法的综合方法, 实验仿真表明, 本文方法能够有效地提高系统的识别率.

参考文献:

- [1] 杨大利, 徐明星, 吴文虎. 噪音环境下的语音识别研究[J]. 计算机工程与应用, 2003, 39(20): 1-4.
- [2] 刘菁华. 一种改进的语音识别抗噪算法[J]. 华侨大学学报: 自然科学版, 2009, 30(1): 117-118.
- [3] 肖全宝, 徐晨, 宋广为, 等. 用于语音识别的基于高谱分辨率的谱减法[J]. 广西师范大学学报: 自然科学版, 2006, 24(4): 26-29.
- [4] 李泽, 崔宣, 马雨廷, 等. MFCC 和 LPCC 特征参数在说话人识别中的研究[J]. 河南工程学院学报: 自然科学版, 2010, 22(2): 51-55.
- [5] 王华朋, 杨洪臣. 声纹识别特征 MFCC 的提取方法研究[J]. 中国人民公安大学学报: 自然科学版, 2008, 14(1): 28-30.
- [6] 刘顺兰, 窦园园, 应娜. 噪声背景下语音识别特征参数选择研究[J]. 杭州电子科技大学学报, 2011, 31(4): 73-76.
- [7] 金连斌, 丁庆海, 陈显治. PMC 在噪声环境下的语音识别中的应用[J]. 解放军理工大学学报: 自然科学版, 2001, 2(2): 42-45.
- [8] 丁沛, 曹志刚. 融合语音增强与后续补偿的抗噪声语音识别方法[J]. 清华大学学报: 自然科学版, 2003, 43(7): 919-922.
- [9] JANG J S R. Audio signal processing and recognition[EB/OL]. (2008-01-23) [2013-09-20]. <http://neural.cs.nthu.edu.tw/jang/books/audioSignalProcessing/>.
- [10] 宗成庆. 统计自然语音处理[M]. 北京: 清华大学出版社, 2010.
- [11] 职振华, 马建芬. 改进的谱减法在语音增强中的应用[J]. 电声技术, 2008, 32(2): 46-48.