

# 基于改进谱减方法的语音增强研究

丁 伟, 吴小培

(安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039)

**摘 要:** 由于噪声的影响导致语音信号的质量降低, 因此需要对语音信号进行语音增强。语音增强是语音信号处理的前沿领域, 其主要目标是从带噪语音中提取纯净的原始语音信号。介绍了实现语音增强方法的原理, 利用实验仿真了传统谱减法和改进谱减方法, 改进法通过对带噪信号进行参数调整, 然后进行频域谱减, 实验结果表明改进方法对语音增强效果明显好于传统方法。此外, 对传统谱减法和改进谱减法的信噪比分别进行了计算, 结果表明改进谱减方法的信噪比相对传统谱减方法有很大提高。

**关键词:** 语音增强; 改进谱相减法; 信噪比

**中图分类号:** TP18

**文献标识码:** A

**文章编号:** 1673-629X(2008)09-0098-03

## Implementation of Speech Enhancement Based on Improved Spectral Subtraction

DING Wei, WU Xiao-pei

(Ministry of Education Key Lab. of Intelligent Computing and Signal Processing,  
Anhui University, Hefei 230039, China)

**Abstract:** Because of noise, quality of speech signal is degraded, speech enhancement makes very important. Speech enhancement is a long-standing problem in speech processing. Its primary aim is to extract pure speech signal under noise background. In this paper, the fundamental theory of spectral subtraction for noise removal is introduced, then enframe-based spectral subtraction method is proposed. The improved method revises parameters through noise signal. The experiment results show that the improved spectral subtraction method can achieve higher SNR than the traditional method does.

**Key words:** speech enhancement; spectral subtraction; signal to noise ratio

### 0 引言

在实际环境下的语音不可避免地要受到噪声的影响, 从而降低语音信号的质量, 因此需要对语音信号进行语音增强<sup>[1]</sup>。语音增强的主要目标是从带噪语音信号中提取尽可能纯净的原始语音或原始语音参数。目前的语音增强方法有<sup>[2]</sup>: 谱相减法、噪声对消法、基于语音参数模型的方法、基于状态转移的 HMM 方法、神经网络法、小波去噪法和信号子空间法等。这些方法在一定程度上解决了带噪语音的去噪问题。

文中主要研究谱减法在语音增强领域的应用, 对比了语音信号在直接经过 FFT 变换进行谱减和经过

加窗分帧处理过后的谱减方法, 对比结果显示经过加窗分帧后再进行语音增强效果明显较好。

### 1 谱减法理论分析

传统谱减法的基本思想如图 1 所示<sup>[3]</sup>。基本谱减法假设噪声与短时平稳的语音信号相互独立。在此条件下, 把带噪语音和噪声进行 FFT 变换, 然后在变换域进行相减得到去噪后的语音, 也就是从带噪语音的功率谱中减去噪声功率谱, 从而得到较为纯净的语音信号频谱, 再经过 IFFT 获得去噪的语音信号, 从而达到去噪效果。

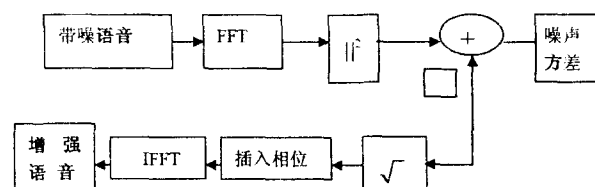


图 1 谱减法原理图

收稿日期: 2007-12-26

基金项目: 安徽省自然科学基金资助项目(2002kj003); 安徽大学“模式识别”创新基金资助项目

作者简介: 丁 伟(1983-), 男, 安徽淮北人, 硕士研究生, 研究方向为智能信息处理; 吴小培, 教授, 博士, 博士生导师, 研究方向为智能信息处理。

如果设  $s(t)$  为纯净语音信号,  $n(t)$  为噪声信号,  $y(t)$  为带噪语音信号, 则有:

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

用  $Y(\omega)$ 、 $S(\omega)$ 、 $N(\omega)$  分别表示  $y(t)$ 、 $s(t)$ 、 $n(t)$  的傅里叶变换, 则由式(1)可得:

$$Y(\omega) = S(\omega) + N^*(\omega) \quad (2)$$

由式(2)可得:

$$|Y(\omega)|^2 = |S(\omega)|^2 + |N(\omega)|^2 + 2R_e\{S(\omega)N^*(\omega)\} \quad (3)$$

由式(3)可得:

$$E(|Y(\omega)|^2) = E(|S(\omega)|^2) + E(|N(\omega)|^2) + 2E\{R_e\{S(\omega)N^*(\omega)\}\} \quad (4)$$

前面已经假设  $s(t)$  和  $n(t)$  是相互独立的, 所以  $S(\omega)$  与  $N(\omega)$  也独立。而  $N(\omega)$  为零均值的高斯分布, 故  $E\{R_e\{S(\omega)N^*(\omega)\}\}$  等于 0。所以有<sup>[5]</sup>:

$$E(|Y(\omega)|^2) = E(|S(\omega)|^2) + E(|N(\omega)|^2) \quad (5)$$

前面假设语音信号是短时平稳, 因此需要对输入语音信号加窗后再予以处理, 所以由式(5)可得出:

$$|Y_t(\omega)|^2 = |S_t(\omega)|^2 + |N_t(\omega)|^2 \quad (6)$$

( $t$  表示加窗分帧后的第  $t$  帧)

由于平稳噪声的功率谱在发音前和发音期间可以认为基本没有变化, 这样可以通过发音前的所谓“寂静段”(认为在这一段里没有语音只有噪声)来估计噪声的功率谱  $|N(\omega)|^2$ <sup>[4]</sup>, 从而有:

$$|S_t(\omega)|^2 = |Y_t(\omega)|^2 - |N_t(\omega)|^2 \quad (7)$$

由此得到原始语音的估计值:

$$|S_t(\omega)| = [|Y_t(\omega)|^2 - |N_t(\omega)|^2]^{1/2} \quad (8)$$

频域处理过程中只考虑了功率谱的变换, 而最后 IFFT 变换中需要借助相位谱来恢复降噪后的语音时域信号。根据人耳对相位变换不敏感这一特点, 可以用原带噪语音信号  $y(t)$  的相位谱来代替估计之后的语音信号的相位谱, 从而可以得到降噪后的语音时域信号<sup>[5]</sup>。此外, 谱减法的物理意义可以通过下式得出:

$$G_k = (1 - 1/r_k)^{1/2}$$

其中  $G_k$  是第  $k$  个频谱分量的增益函数,  $r_k$  是后验信噪比 ( $r_k = |Y_k|^2 / |N_k|^2$ ), 当信噪比高时, 含语音的可能性很大, 衰减小, 反之, 含语音的可能性很小, 衰减大。

## 2 改进谱减法

“音乐噪声”的来源是: 噪声的功率谱随机变化范围很宽, 在频域中的最大、最小值往往相差几个数量级。因而, 在用传统谱减法去噪后留下较大的功率谱

分量残余, 它们在频谱上呈现出随机出现的尖峰, 而在听觉上形成有规律的噪声<sup>[6]</sup>。

为了有效地减少宽带和“音乐噪声”, 可以对谱减法进行改进(如图 2 所示), 我们知道噪声和语音的能量分布有很大差别, 噪声能量分布比较均匀, 而语音主要集中于某些频率或频段上, 因此可以幅度较高的时帧处去噪时减去  $n * E[|N_k|^2]$ , 可更好地突出语音功率谱, 这时可以把公式(8)灵活地变为:

$$|S_t(\omega)| = [|Y_t(\omega)|^m - \alpha * |N_t(\omega)|^m]^{1/m}$$

其中  $\alpha, m$  是两个参数, 通过不断调整  $\alpha, m$  的值来取得更好的去噪效果<sup>[7]</sup>。

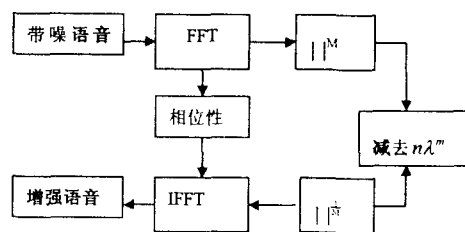


图2 改进后原理图

## 3 实验仿真

实验是自己录制的带噪语音“大学”, 经 8kHz 采样, PCM 编码, 16bit 量化为数字信号, 单声道, 数据长 12672 点。把带噪语音通过长度为 256 点(32ms)的汉明窗形成长度为 256 点的语音帧, 相邻两帧之间重叠 128 点, 然后对带噪语音逐帧进行去噪处理。在实验中取两组  $\alpha, m$  参数分别为  $\alpha = 1, m = 2$  和  $\alpha = 0.5, m = 5$ , 其中第一组数据就是传统谱减法, 从实验仿真结果来看, 改进后效果明显好于传统谱减法。

传统谱减法实现的语音增强如图 3 所示。

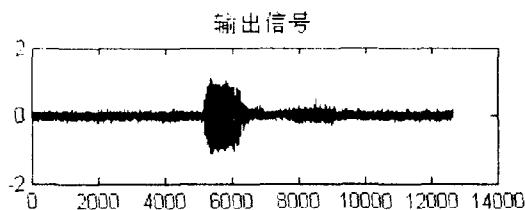


图3 基本谱减法去噪效果图  
经过改进后的语音增强如图 4 所示。

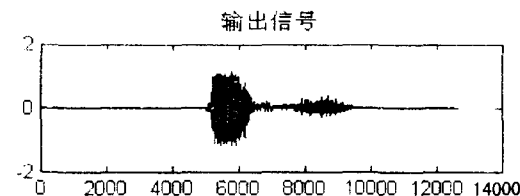


图4 改进谱减法去噪效果图  
仿真结果表明经过改进后语音信号明显接近于纯

净语音信号,降噪效果明显好于传统谱减方法,此外在信号频谱方面也进行了对比,如图 5 所示。

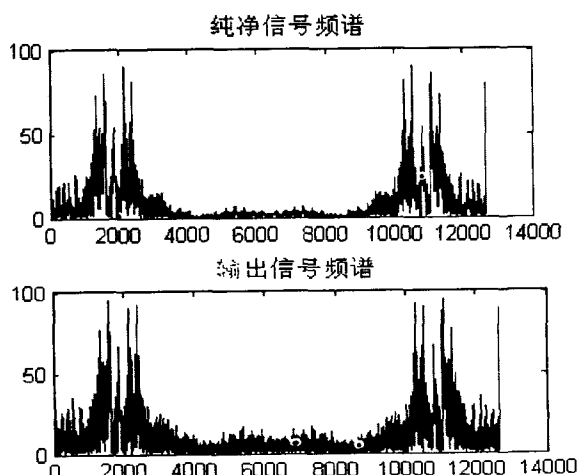


图 5 基本谱减法去噪语音信号频谱对比图

上图是传统谱减法的带噪频谱图和去噪后的频谱图,经过改进后的频谱图如图 6 所示。

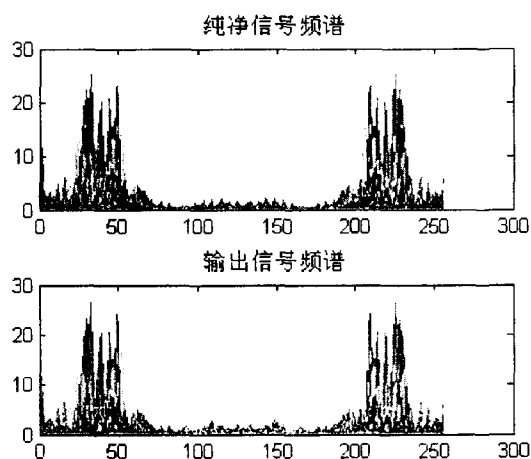


图 6 改进谱减法去噪语音信号频谱对比图

由图表明改进的方法在频谱图上也明显好于传统谱减方法。语音去噪效果的客观评价是以语音系统的输入信号和输出信号之间的误差大小来判别语音质量的好坏。信噪比(SNR)<sup>[8]</sup>是衡量针对宽带噪声失真的语音去噪效果的常规方法。假设  $y(n)$  为带噪语音信号,  $s(n)$  为其中的纯净语音信号,  $\hat{s}(n)$  为去噪后的语音信号,  $L$  为语音信号的长度。

输入信噪比为:

$$SNR_{in} = 10 \lg \frac{\sum_{n=0}^L s^2(n)}{\sum_{n=0}^L [y(n) - s(n)]^2}$$

输出信噪比为:

$$SNR_{out} = 10 \lg \frac{\sum_{n=0}^L s^2(n)}{\sum_{n=0}^L [\hat{s}(n) - s(n)]^2}$$

本实验仿真中的输入输出的信噪比如表 1 所示。

表 1 信噪比实验结果比较

|                          |         |
|--------------------------|---------|
| 输入信噪比 $SNR_{in}$ (dB)    | 2.9220  |
| 传统谱减法输出 $SNR_{out}$ (dB) | 6.4456  |
| 改进方法输出 $SNR_{out}$ (dB)  | 13.5733 |

从上表中可以看出,改进谱减方法的信噪比明显高于传统谱减方法。

#### 4 结束语

语音增强是语音信号处理的前沿领域,也是语音识别和语音合成等方向的基础,应用谱减方法进行语音增强,也只是语音增强一个应用,此外还有自适应滤波以及小波变换等用于语音增强方法,有待于进一步研究。

#### 参考文献:

- [1] 陶智,葛良. 基于谱减法语音增强和噪声消除的研究[J]. 苏州大学学报, 2002, 18(3): 58-61.
- [2] 王振力,张雄伟,白志强. 语音增强新方法的研究[J]. 南京邮电大学学报, 2007, 27(2): 10-14.
- [3] 易克初,田赋,付强. 语音信号处理[M]. 北京:国防工业出版社, 2000.
- [4] 赵力. 语音信号处理[M]. 北京:机械工业出版社, 2003.
- [5] 钱国青,赵鹤鸣. 基于改进谱减算法的语音增强新方法[J]. 计算机工程与应用, 2005, 41(35): 42-43.
- [6] Bollsf. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans on Acoust, Speech, Signal Processing, 1997, 227(2): 113-120.
- [7] Sim B L, Tong Y C, Chang J S, et al. A parametric formulation of the generalized spectral subtraction method[J]. IEEE Trans on Speech and Audio Processing, 1998, 6(4): 328-337.
- [8] WEI Wei, CHEN Yan-ph. Speech enhancement by spectral component selection[C] // Proceedings of ICSP. [s. l.]: [s. n.], 2000: 674-678.

中国计算机学会会刊、中国科技核心期刊:  
《计算机技术与发展》欢迎订阅,邮发代号:52-127。