

基于非平稳噪声估计的改进谱减语音增强算法

孙晋松, 田 岚

SUN Jin-song, TIAN Lan

山东大学 信息科学与工程学院, 济南 250100

School of Information Science and Engineering, Shandong University, Jinan 250100, China

E-mail: sunjinsongd@163.com

SUN Jin-song, TIAN Lan. Improved spectral subtraction algorithm applied to speech enhancement based on non-stationary noise estimation. *Computer Engineering and Applications*, 2010, 46(5): 120-122.

Abstract: The traditional spectral subtraction can't suppress the non-stationary noise in the realistic voice communication effectively. In order to improve the performance in eliminating the non-stationary noise, a noise estimation algorithm is proposed, which divides the power spectrum of the noisy speech into several subbands according to the Bark frequencies and tracks the noise by the improved minima controlled recursive averaging method in every individual band. The noise estimation approach can estimate the noise power spectrum more accurately while reducing the computation. Then the spectral subtraction method using the noise estimation approach is presented and compared with the traditional spectral subtraction. The result shows that the proposed method can suppress the noise better while reducing the musical noise more efficiently.

Key words: spectral subtraction; speech enhancement; non-stationary noise; Bark bands

摘 要: 传统的谱减法无法有效地抑制实际语音通信中的非平稳噪声, 为了进一步提高谱减法的去噪性能, 提出了一种改进的噪声估计算法, 首先将带噪语音的功率谱按照 Bark 频率进行子带划分, 然后分别在每个子带内利用改进的最小统计量控制递归平均方法跟踪噪声的变化, 从而在准确估计非平稳噪声的功率谱的同时减少计算量。将该算法应用到谱减法中, 并与传统的增强型谱减法进行对比实验, 实验结果表明: 改进的谱减法能够更好地去除各种非平稳噪声, 而且能够有效抑制“音乐噪声”, 使得增强后的语音具有更好的音质。

关键词: 谱减法; 语音增强; 非平稳噪声; Bark 子带

DOI: 10.3778/j.issn.1002-8331.2010.05.036 文章编号: 1002-8331(2010)05-0120-03 文献标识码: A 中图分类号: TN912.35

1 引言

在语音通信中, 语音信号会受到背景噪声的干扰, 从而降低通信系统的性能, 因此需要对语音信号进行增强处理, 抑制背景噪声, 恢复纯净语音, 从而提高语音的感知质量, 缓解听觉疲劳。目前许多语音增强算法可以用于抑制噪声^[1-3], 但是由于谱减法计算简单, 性能良好, 因此在实际中被广泛应用。在谱减语音增强算法中, 噪声功率谱的准确估计对整个增强系统的性能起到至关重要的作用。传统的谱减法中, 噪声通常被认为是平稳的, 其功率谱在发声前与发声期间基本没有变化, 因此噪声的功率谱可以通过在无声段进行更新得到, 而在语音段则保持不变^[4]。但是在实际的语音通信中, 背景噪声常常为非平稳噪声, 噪声的功率谱随时间有较大的变化, 而且语音段往往持续较长的时间, 因此如果仅仅在无声段进行噪声的功率谱估计, 则会使得噪声的功率谱不能得到及时更新, 继而造成估计的噪声功率谱与实际的噪声功率谱之间较大的偏差, 从而不能有效地去除噪声。

为了准确地估计噪声的功率谱, 从而更有效地抑制背景噪声, Rangachari 等人提出了一种改进的最小统计量控制递归平

均方法^[5]用于估计非平稳噪声的功率谱, 其使用了一种非线性局部最小值跟踪方法以及与频率相关的阈值设定, 从而能够更准确地估计噪声的功率谱, 但是由于该算法需要对信号功率谱的每个频率成分进行统计分析, 因此其计算量比较大。在该算法的基础上, 根据人耳的听觉感知特点将带噪语音信号的功率谱按照 Bark 临界带划分为一系列的子带, 然后分别在每一个子带内估计噪声的功率谱, 在有效地估计非平稳噪声功率谱的基础上减少了运算量从而更好地满足实时处理的要求。将该噪声估计算法应用到谱减法中, 以期进一步提高语音增强系统的噪声抑制能力, 改善增强后语音的听觉质量。

2 谱减法原理

假设背景噪声为加性噪声, 且与语音信号不相关。一帧带噪语音信号可以表示为:

$$y(\lambda, n) = s(\lambda, n) + d(\lambda, n), 0 \leq n \leq N-1 \quad (1)$$

其中 λ 为帧号, n 表示第 n 个样点, N 表示一帧语音信号的长度。 $y(\lambda, n)$ 为带噪语音信号, $s(\lambda, n)$ 为纯净的语音信号, $d(\lambda, n)$ 为噪声。设 $Y(\lambda, w), S(\lambda, w), D(\lambda, w)$ 分别为 $y(\lambda, n), s(\lambda, n),$

作者简介: 孙晋松(1985-), 硕士, 主要研究领域为语音信号处理; 田岚(1965-), 教授, 主要研究领域为语音信号处理, 电子耳蜗等。

收稿日期: 2008-08-27 **修回日期:** 2008-11-10

$d(\lambda, n)$ 的短时频谱, 则其对应的功率谱之间存在如下关系:

$$|Y(\lambda, w)|^2 = |S(\lambda, w)|^2 + |D(\lambda, w)|^2 \quad (2)$$

在传统的谱减法中, 噪声功率谱 $|D(\lambda, w)|^2$ 通过在无声段利用式(3)得到:

$$|\tilde{D}(\lambda, w)|^2 = \rho|\tilde{D}(\lambda-1, w)|^2 + (1-\rho)|Y(\lambda, w)|^2 \quad (3)$$

式中 $|\tilde{D}(\lambda, w)|^2$ 为噪声功率谱的估计值, ρ 为常数, 通常取值为0.75~0.95。继而可以得到纯净语音信号的幅度谱估计值为:

$$|\tilde{S}(\lambda, w)| = G(\lambda, w)|Y(\lambda, w)| \quad (4)$$

其中 $G(\lambda, w)$ 的表达式如式(5)所示:

$$G(\lambda, w) = \begin{cases} [1 - \alpha(\frac{|\tilde{D}(\lambda, w)|}{|Y(\lambda, w)|})^{\gamma_1, \gamma_2}] & \text{如果 } (\frac{|\tilde{D}(\lambda, w)|}{|Y(\lambda, w)|})^{\gamma_1} < \frac{1}{\alpha + \beta} \\ [\beta(\frac{|\tilde{D}(\lambda, w)|}{|Y(\lambda, w)|})^{\gamma_1, \gamma_2}] & \text{其他} \end{cases} \quad (5)$$

其中 $\alpha(\alpha > 1)$ 称为过减因子, α 越大则残留的噪声越少, 但同时也增加了语音信号的失真。 $\beta(0 < \beta < 1)$ 为谱平滑因子, 其作用是通过人为地引入噪声, 从而达到掩蔽音乐噪声的目的。考虑到人耳对语音信号的相位不敏感, 因此将带噪语音的相位与 $|\tilde{S}(\lambda, w)|$ 相乘后再按照重叠相加的方法即可得到增强后的语音。

3 噪声功率谱估计算法

首先对带噪语音信号加窗处理, 然后进行傅里叶变换得到其短时频谱 $Y(\lambda, w)$, 继而得到其功率谱为 $P_y(\lambda, w) = |Y(\lambda, w)|^2$ 。将带噪语音的短时功率谱按照 Bark 子带进行划分, 这是因为 Bark 子带的划分方式与人耳的听觉感知特性是一致的。对带噪语音的 Bark 子带划分如式(6)所示:

$$P(\lambda, k) = \sum_{w_{lk} \leq w \leq w_{mk}} P_y(\lambda, w) \quad (6)$$

式中 k 为 Bark 频带的标号, 其最大值与信号的采样频率有关, 当信号以 16 kHz 采样时, 其频率范围对应 21 个 Bark 带, 即 k 的最大值为 21。 w_{lk} 和 w_{mk} 为第 k 个 Bark 带的上下两个截止频率。

获得各个子带的信号能量后, 首先利用式(7)对各个频带内的能量进行一阶平滑处理, 得到平滑后的频带能量 $P'(\lambda, k)$:

$$P'(\lambda, k) = \eta P'(\lambda-1, k) + (1-\eta)P(\lambda, k) \quad (7)$$

然后根据公式(8)获得信号能量的局部最小值 $P_{\min}(\lambda, k)$:

如果 $P_{\min}(\lambda-1, k) < P'(\lambda, k)$, 则

$$P_{\min}(\lambda, k) = \gamma P_{\min}(\lambda-1, k) + \frac{1-\gamma}{1-\delta} (P'(\lambda, k) - \delta P'(\lambda-1, k))$$

否则

$$P_{\min}(\lambda, k) = P'(\lambda, k) \quad (8)$$

如果一帧带噪语音的 Bark 频带内主要为噪声, 则其能量 $P'(\lambda, k)$ 与对应的局部能量最小值 $P_{\min}(\lambda, k)$ 相差较小; 而如果频带内主要为语音信号的频率成分时, $P'(\lambda, k)$ 则与 $P_{\min}(\lambda, k)$ 具有较大的差别。因此, 可以通过计算 $P'(\lambda, k)$ 与局部最小能量 $P_{\min}(\lambda, k)$ 的比值来确定该时间帧各个频带内主要为语音信号还是噪声。设

$$S_r(\lambda, k) = \frac{P'(\lambda, k)}{P_{\min}(\lambda, k)} \quad (9)$$

若 $S_r(\lambda, k)$ 大于门限值 Thr 时, 则认为第 k 个频带内主要是以语音成分为主否则认为主要是噪声成分, 并且同时得到估计的语音帧出现概率 $p(\lambda, k)$, 即

如果 $S_r(\lambda, k) > Thr$ 时, 则

$$p(\lambda, k) = \alpha_p p(\lambda-1, k) + (1-\alpha_p)$$

否则

$$p(\lambda, k) = \alpha_p p(\lambda-1, k) \quad (10)$$

其中 α_p 为平滑算子。则估计的噪声功率谱为:

$$D(\lambda, w) = \alpha_s(\lambda, k)D(\lambda-1, w) + (1-\alpha_s(\lambda, k))P_y(\lambda, w) \quad (11)$$

其中 $\alpha_s(\lambda, k) = \alpha_s + (1-\alpha_s)p(\lambda, k)$ 。

图 1 所示为 $f=1500$ Hz 时真实的噪声功率谱与使用该算法估计得到的噪声功率谱的对比, 其中噪声类型为工厂噪声, 平均信噪比为 0 dB, 细线代表真实的噪声功率谱, 粗线代表估计的噪声功率谱。从图 1 中可以看出, 提出的噪声估计算法能够有效地跟踪非平稳噪声的功率谱的变化。

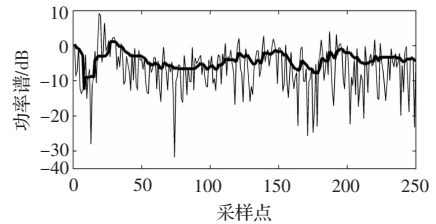


图 1 真实噪声功率谱与估计噪声功率谱的对比图

4 改进的谱减法的设计

将改进的噪声估计算法应用到谱减法中得到的语音增强原理图如图 2 所示。

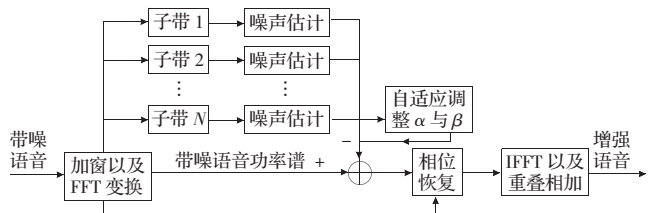


图 2 改进谱减法原理图

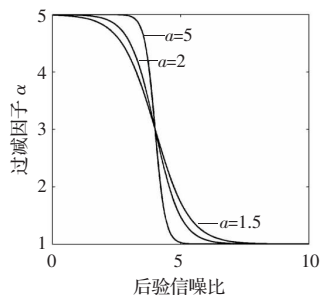
在传统的增强谱减法中, 过减因子 α 与谱平滑因子 β 均为常数, α 与平均信噪比 SNR 通过式(12)关联^[6]。

$$\alpha = \begin{cases} 4.75 & SNR < -5 \text{ dB} \\ 4 - 0.15 \times SNR & -5 \text{ dB} \leq SNR < 20 \text{ dB} \\ 1 & 20 \text{ dB} \leq SNR \end{cases} \quad (12)$$

但是在不同的时间帧不同的频率点上, 信噪比是不相同的, 因此采用固定的 α 和 β 是不合理的。在以语音信号为主的频率点上, 应该采用较小的 α , 尽可能地减少语音信号的失真, 而在以噪声为主的频率点上, 采用较大的 α , 才能够更好地去除背景噪声, 而 β 的选择与 α 正好相反。因此, 为了自适应地调整过减因子, 将采用式(13)来实现 α 与后验信噪比 SNR_{post} 之间的关联。

$$\alpha = 1 + \frac{4}{1 + e^{\frac{a(SNR_{post} - b)}{a(SNR_{post} - b)}}} \quad (13)$$

其中 $SNR_{post} = \tilde{P}_d(\lambda, w) / P_y(\lambda, w)$, α 的取值范围为 (1, 5), a 则决定了 α 随 SNR_{post} 变化的路径, 而 b 则决定了 α 取中值时对应的后验信噪比。当 $b=4, a=1.5, 2, 5$ 时的函数图像如图 3 所示, 从图 3 中可以看出, 随着 a 的增大, 过减因子 α 随 SNR_{post} 的变化路径越陡峭。在实验中, 选择 $a=2$ 作为 α 与后验信噪比关联函数的参数。对应的频谱阶因子可以按照式 $\beta=0.001/\alpha$ 来设定。

图3 α 与后验信噪比关系图

5 实验结果

实验中所用到的纯净语音信号和噪声的采样频率均为 16 kHz, 采样大小为 16 bit, 其中纯净语音信号为男声所说“你要到哪里去”, 噪声采用的是 Noisex-92 噪声库中的汽车噪声、工厂噪声和 F16 战斗机噪声。对带噪语音信号进行分帧处理, 每一帧数据的长度为 32 ms, 半帧重叠。进行噪声估计时采用的参数设置如下: $\rho=0.85, \eta=0.7, \gamma=0.998, \delta=0.8, \alpha_p=0.8, \alpha_d=$

$0.85, Thr=$

2	1 Hz $\leq f \leq$ 1 kHz
3	1 kHz $< f \leq$ 3 kHz
5	其他

。将一段纯净的语音信号分别叠

加上不同信噪比的三种噪声, 并分别采用使用传统增强谱减法 and 提出的改进谱减法对带噪语音进行处理。当原始纯净语音叠加信噪比 SNR=5 dB 的 F16 战斗机噪声时, 两种方法的处理结果如图 4 所示。

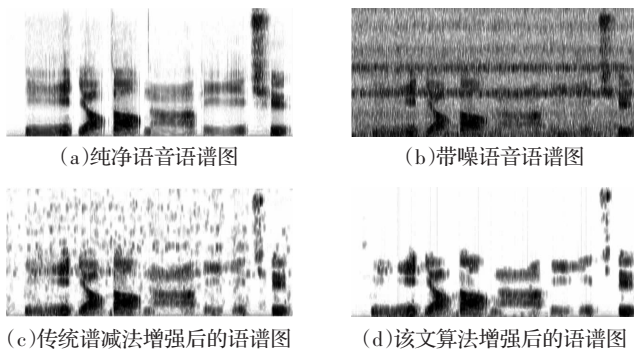


图4 两种语音增强算法的效果图

从语谱图中可以看出, 传统谱减法并没有完全去除背景噪声, 这样进行谱减后, 增强后的语音中会出现“音乐噪声”, 从而影响合成语音的听觉效果; 而由于该文采用的噪声估计算法能够更准确地估计噪声的功率谱, 因此经过该文算法得到的增强语音能够更好地抑制背景噪声, 恢复原始纯净语音信号。此外,

在实验中还采用分段信噪比作为客观衡量带噪语音增强效果的方法, 并将改进谱减法与传统的增强谱减法进行对比, 结果如表 1 所示。其中增强前的信噪比用平均信噪比表示, 而增强后的信噪比为分段信噪比。从表 1 中可以看出, 在多种噪声环境下, 改进谱减法的性能都明显优于传统的增强谱减法, 主观试听改进谱减法增强后的语音, “音乐噪声”基本已经消除, 具有更好的音质。

表1 两种谱减法的增强效果对比表

对比算法	汽车噪声				F16 战斗机噪声				工厂噪声				dB
	-3	0	3	5	-3	0	3	5	-3	0	3	5	
传统增强谱减法	3.94	4.85	5.64	6.21	1.18	2.56	3.63	4.42	0.39	1.45	2.94	3.97	
改进谱减法	4.52	5.20	6.17	6.53	1.57	3.06	4.24	4.95	0.62	1.67	3.11	4.35	

6 结论

提出了一种改进的噪声估计算法, 其将带噪语音的功率谱进行 Bark 子带划分, 并按照改进的最小统计量控制递归平均方法更新每个子带内的噪声功率谱, 该算法能够有效地跟踪各种非平稳噪声的功率谱, 同时减少了计算量。将该算法应用到谱减法中, 更好地满足了实时处理的要求, 与传统的增强谱减法相比, 改进的谱减法在多种噪声环境下均能更好地去除噪声, 恢复原始纯净语音信号, 而且增强后的语音能够更有效地抑制“音乐噪声”, 从而改善了主观听觉质量。将该算法置于语音通信系统中, 能够进一步提高通信系统的性能。

参考文献:

- [1] Cohen I. Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator[J]. IEEE Signal Processing Letters, 2002, 9(4): 113-116.
- [2] Jabloun F, Champagne B. Incorporating the human hearing properties in the signal subspace approach for speech enhancement[J]. IEEE Trans Speech Audio Process, 2003, 11(6): 700-708.
- [3] Chen S H, Wang J F. Speech enhancement using perceptual wavelet packet decomposition and teager energy operator[J]. Journal of VLSI Signal Processing, 2004, 36: 125-139.
- [4] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2003: 280-282.
- [5] Sundararajan R, Philipos C L. A noise-estimation algorithm for highly non-stationary environment[J]. Speech Communication, 2006, 48(2): 220-231.
- [6] Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise[C]. IEEE ICASSP, Washington DC, 1979: 208-211.
- [7] 路[M]. 北京: 清华大学出版社, 1998.
- [8] 苗传江. HNC(概念层次网络)理论导论[M]. 北京: 清华大学出版社, 2005.
- [9] 黄曾阳. 语言概念空间的基本定理和数学物理表达式[M]. 北京: 海洋出版社, 2004.
- [10] 张全, 吴晨, 韦向峰. 汉语句间成分共享类型及分布研究[J]. 计算机科学, 2007(1): 166-169.
- [11] 贾宁, 张全. 基于句间关系的汉语语义块省略恢复[J]. 中文信息学报, 2008(6): 33-37.
- [12] Hobbs J R. Pronoun resolution[R]. City College, New York, 1976.
- [13] Yang X F, Su J, Tan C L. Kernel-based pronoun resolution with structured syntactic knowledge[C]. // Coling-ACL2006, 2006: 41-48.
- [14] 周国栋, 孔芳, 朱巧明. 指代消解: 国内外研究现状及趋势[C]. // 内容计算的研究与应用前沿——第九届全国计算语言学学术会议论文集. 北京: 清华大学出版社, 2007: 264-269.
- [15] 黄曾阳. HNC(概念层次网络)理论——计算机理解自然语言的新思

(上接 27 页)