

一种基于改进谱减法的语音去噪新方法

林 琴, 张道信, 吴小培

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

摘 要: 介绍了改进谱减法的原理及算法。提出在信噪比(SNR)较低的情况下, 根据语音短时能量和过零率, 判断在无声或有声期间是否偶然的噪声过高, 从而设定合适的参数降低噪声。实验证明该方法提高了输出信噪比, 改善了语音质量, 具有良好的语音去噪的效果。

关键词: 谱减法; 信噪比; 短时能量; 过零率; 语音去噪

中图分类号: TP18 **文献标识码:** A **文章编号:** 1673-629X(2007)07-0063-04

A New Speech Denoising Algorithm Based on Improved Spectral Subtraction

LIN Qin, ZHANG Dao-xin, WU Xiao-pe

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

Abstract: In this paper, the theory and algorithm of the improved spectral subtraction are presented. When the signal to noise ratio (SNR) is low, judge whether the speech contains the high occasional noise in the period unvoiced or voiced, according to the short-time energy and zero-cross ratio of the speech. Then the proper parameters are set to eliminate the noise. The experimental results show that the method enhances the signal to noise ratio (SNR), improves the speech quality and gets better denoising effect.

Key words: spectral subtraction; signal to noise ratio; short-time energy; zero-cross ratio; speech denoising

0 引言

人们在语音通信过程中不可避免地会受到来自周围环境、传输媒介引入的噪声的干扰。在各种语音处理的应用中, 由于背景噪声使得语音质量降低的现象非常普遍, 例如在移动通讯和语音识别中就希望语音信号有尽可能高的信噪比, 因此设法去除语音信号中的非相关噪声, 提高语音信号的信噪比, 改善语音质量, 即语音去噪技术就成为语音研究中的一个重要课题。

目前语音信号的去噪方法大体上有四类: 噪声对消法、谐波增强法、基于参数估计的语音再合成法和基于语音短时谱估计的增强算法^[1]。其中, 基于短时谱估计的谱相减法是一种发展较早且应用较为成熟的语音去噪算法, 该算法利用加性噪声与语音不相关的特

点, 在假设噪声是统计平稳的前提下, 用无语音间隙测算得到的噪声频谱估计值取代有语音期间噪声的频谱, 与含噪语音频谱相减, 从而获得语音频谱的估计值。谱减法具有算法简单、运算量小的特点, 便于实现快速处理, 往往能够获得较高的输出信噪比, 所以被广泛采用。该算法经典形式的不足之处是处理后会产生具有一定的节奏性起伏、听上去类似音乐的“音乐噪声”^[2]。针对这一问题人们提出了各种改进形式, 这些形式虽未将“音乐噪声”彻底滤除, 但的确在不同程度上使其有所削弱, 令听者乐于接受。

文中对谱减法经典形式及多种改进形式在理论上加以分析, 在此基础上, 提出了在信噪比较低的情况下一种简单的改进方案, 在改进谱减法的基础上, 根据语音的短时能量和过零率综合判断在无声或有声期间是否含有过高的偶然噪声, 这样设定一定的参数就可以去除剩余的偶然噪声。

1 谱减法的原理

1.1 理论分析

语音是非平稳随机过程, 但在 10 ~ 30ms 的分析

收稿日期: 2006-09-23
基金项目: 安徽省教育厅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: 林 琴(1982-), 女, 安徽六安人, 硕士研究生, 研究方向为智能信息处理; 张道信, 副教授, 硕士, 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。

帧内可近似看成是平稳的过程,因此语音的短时谱具有相对稳定特性^[3]。假设纯净语音信号 $s(n)$ 被加性噪声 $d(n)$ 所污染,形成带噪语音 $y(n)$,即 $y(n) = s(n) + d(n)$ 。短时谱处理时要将信号分帧处理,分帧时时域信号首先分成相叠的帧信号,帧长为 N ,帧移为 $N/2$,设观察到的一帧带噪语音信号为:

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

其中, $s(i)$ 为一帧纯净语音, $d(i)$ 为一帧平稳加性高斯噪声,且有 $s(i)$ 与 $d(i)$ 相互独立,为了避免分帧时的截断效应,应在变换到频域之前对 $y(i)$ 加窗处理^[4]。

1.2 原理

谱减法的经典形式即在频域将带噪语音的功率谱减去噪声的功率谱得到语音功率谱估计,开方后就得到语音幅度估计,将其相位恢复后再采用逆傅里叶变换恢复时域信号。考虑到人耳对相位失真的感知不敏感,相位恢复时所采用的相位是带噪语音的相位信息。谱减法的基本原理如图 1 所示,设 $y(i)$ 的傅里叶系数为 Y_k , $s(i)$ 的傅里叶系数为 S_k , $d(i)$ 的傅里叶系数为 N_k ,则如式(2)所示:

$$|Y_k|^2 = |S_k|^2 + |N_k|^2 + S_k N_k^* + S_k^* N_k \quad (2)$$

由于 $s(i)$ 和 $d(i)$ 相互独立, N_k 满足高斯分布且均值为零,所以有:

$$E[|Y_k|^2] = E[|S_k|^2] + E[|N_k|^2] \quad (3)$$

对于一个分析帧内的短时平稳过程,有: $|Y_k|^2 = |S_k|^2 + \lambda_n(k)$,其中 $\lambda_n(k)$ 是无话时 $|N_k|^2$ 的统计平均。此时,原始语音的估计值如式(4)所示:

$$|S_k|^2 = [|Y_k|^2 - E[|N_k|^2]]^{1/2} = [|Y_k|^2 - \lambda_n(k)]^{1/2} \quad (4)$$

$|S_k|$ 是增强后语音信号的幅度,图 1 是功率谱相减法的经典形式^[4]。

在谱减法中,采用的是用噪声的统计均值代替当前帧的噪声,因而在当前帧噪声的一些随机谱峰处,经过谱减后就剩下残余的谱峰,而在当前帧的噪声谱谷处,由于采用了半波整流,残余噪声谱成份被设成接近于零的值,这样经过增强后的残余噪声在频域上就成为一些离散的谱峰,相应地在时域就呈现出一些类似正弦信号的叠加,所以就会呈现出乐音的特性,即“音乐噪声”。

1.3 改进的谱相减法算法

针对基本形式的谱减法的残留噪声过多且残余“音乐噪声”过强,众多研究者提出了各种谱相减法的

改进形式。提出较早的有:在估计噪声谱时用无语音频谱的最大值取代基本算法中的均值进行处理^[5];谱相减时给噪声谱乘上一个大于 1 的参数,使剩余噪声能量减至更低等^[6]。这些方案均在不同程度上削弱了“音乐噪声”对语音的影响。

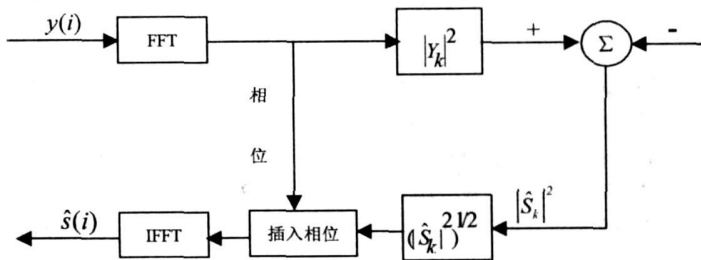


图 1 谱减法基本原理框图

Robert D. Preuss 提出了取各无语音频谱的最大值,替代经典形式中取各无语音频谱均值,作为估计噪声谱。如式(5), (6)所示:

$$|Y_k| = |S_k| + |N_k| \quad (5)$$

$$|S_k| = |Y_k| - \max(|N_k|) \quad (6)$$

通过这种改进即可使噪声分量得到进一步抑制,以达到减少“音乐噪声”的目的^[9]。但试听时发现个别位置语音损失较大。

Berouti 提出的重要改进方案是通过谱相减时给噪声谱乘上一个大于 1 的参数,使在频谱减法时减去的值比估计的噪声谱更多,即:

$$|S_k| = [|Y_k|^\alpha - \beta \lambda_n(k)]^{1/\alpha} \quad (7)$$

这里引入了两个参数 α 和 β 用来进行调节^[9]。显然,当 $\alpha = 2, \beta = 1$ 时就是普通谱相减法。通过适当调节这两个参数可使去噪效果达到比较好的增强效果。但过多增加去噪程度会使增强后的语音失真增大。

2 文中算法

在输入信噪比较低的情况下,改进的谱减法也很难去除偶然的高噪声。文中在改进谱减法的基础上,根据语音的短时能量和过零率^[7]综合判断在无声或有声期间是否含有偶然噪声过高,由此设定合适的参数降低偶然噪声,从而提高语音质量和输出信噪比。设 $\hat{s}(i)$ 为谱减法输出的某一帧增强语音信号,短时能量为:

$$\text{amp} = \sum_{i=1}^N s^2(i) \quad (8)$$

短时过零率为:

$$\text{ZCR} = \frac{1}{2} \sum_{n=1}^N |\text{sgn}[\hat{s}(i)] - \text{sgn}[\hat{s}(i-1)]| \quad (9)$$

式中 ZCR 表示一帧信号样本改变符号的次数即过零率。其中 $\text{sgn}[\cdot]$ 是符号函数。

$$\text{sgn}[s(i)] = \begin{cases} 1 & s(i) \geq 0 \\ -1 & s(i) < 0 \end{cases} \quad (10)$$

设:
$$FF = \log(\text{amp}/\text{ZCR}), h1 = \alpha_1 * \max(FF) + \alpha_2 * \min(FF) \quad (11)$$

$$h2 = \beta_1 * \max(FF) + \beta_2 * \min(FF), h3 = \text{mean}(FF) \quad (12)$$

式中: $\alpha_1 + \alpha_2 = 1, \beta_1 + \beta_2 = 1 (\alpha_1, \alpha_2, \beta_1, \beta_2 > 0), h1, h2, h3$ 为门限值。根据 FF 来判断浊音和清音中是否含有剩余白噪声, 无声段中是否含有剩余白噪声, 然后设定合适的参数来降低噪声。因为 amp 和 ZCR 可以判断清音和浊音、有声和无声。浊音的 amp 比清音大, 清音具有较高的 ZCR 。

3 实验结果分析及评价

语音信号是一典型的非平稳随机信号, 用文中方法对实际语音信号进行了仿真研究。实验中, 所用语音是在安静的环境下, 利用话筒录得的一段语音信号。内容为“走进了经济生活, 走出了自己的风格”。经 8kHz 采样, PCM 编码, 量化为数字信号, 单声道, 数据长 28000 点。噪声为高斯白噪声。按一定比例混合生成三种不同的信噪比(-5dB, 0dB, 5dB)的带噪语音。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, 用 Matlab7 仿真实现。

如图 2, 图 3 所示, 给出不同输入 SNR 下语音去噪的仿真结果。不难看出, 经过去噪后的语音质量和输出 SNR 在一定程度上都有所改善, 文中算法的去噪效果明显优于改进算法, 特别是在低信噪比的情况下, 效果更为明显。

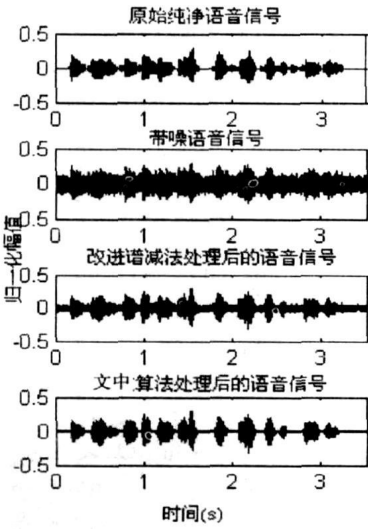


图 2 输入 SNR=0dB 的实验结果

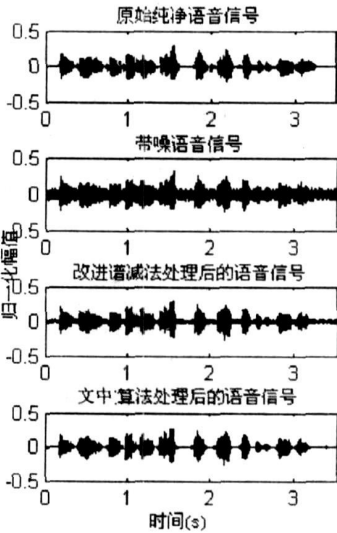


图 3 输入 SNR=5dB 的实验结果

如图 4, 图 5 所示, 给出了不同输入 SNR 下语音去噪前后的宽带语谱图伪彩色显示图。其横坐标表示时间, 纵坐标表示频率, 而每像素的灰度值大小反映相应时刻和相应频率的信号能量密度, 这种时频图称为语谱图 (Spectrogram)^[8]。从语谱图上不仅能看出任一时时刻发音器官的共振峰特征, 而且可以看出语音的基音频率, 是否清音、爆破音等。语谱分析在语音识别、合成及编码中很有意义。

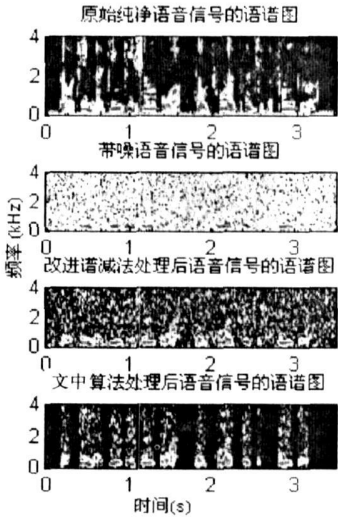


图 4 输入 SNR=0dB 的实验结果

语谱图中的花纹有横杠、乱纹和竖直条等。横杠是与时间轴平行的几条带纹, 它们是共振峰。从横杠对应的频率和宽度可以确定相应的共振峰频率和带宽。在一个语音段的语谱图中, 有没有横杠出现是判断它是否是浊音的标志。竖直条是语谱图中出现与时间轴垂直的一条窄条。每个竖直条相当于一个基音, 条纹的起点相当于声门脉冲的起点, 条纹之间的距离表示基音周期。条纹越密表示基音频率越高。实验结

果表明,用文中算法去噪后的语音,其语谱图能清楚地看到浊音共振峰时变过程,特别是时间分辨率高。

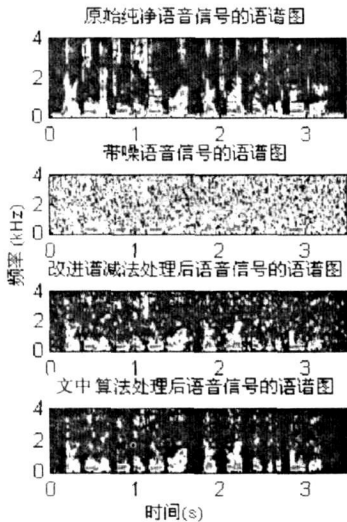


图 5 输入 SNR=5dB 的实验结果

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)^[9]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,则输入信噪比为:

$$SNR_{in} = 10\lg \frac{\sum_{n=0}^L s^2(n)}{\sum_{n=0}^L [y(n) - s(n)]^2} \tag{13}$$

输出信噪比为:

$$SNR_{out} = 10\lg \frac{\sum_{n=0}^L s^2(n)}{\sum_{n=0}^L [\hat{s}(n) - s(n)]^2} \tag{14}$$

表 1 语音去噪效果实验结果比较

| 输入信噪比 SNR _{in} (dB) | -5 | 0 | 5 |
|------------------------------|--------|--------|---------|
| 改进谱减法 SNR _{out} | 1.5480 | 3.8227 | 8.4300 |
| 文中方法 SNR _{out} | 5.2653 | 7.6118 | 10.8180 |

表 1 给出了在语音信号输入信噪比分别为-5dB,

0dB 和 5dB 的情况下,由文中方法得出的实验结果输出信噪比的比较。

4 结 论

与直接采用改进的谱减法进行语音去噪相比,文中提出的语音去噪方法使得噪声得到很明显的抑制,提高了语音质量。实验表明,特别是在输入信噪比较低的情况下,采用文中算法,更好地提高了带噪语音的输出信噪比,明显削弱了谱减法所带来的“音乐”噪声,抑制加性噪声效果最好、运算量小、系统简单。但语音的可懂度却受到一定的损失,略微存在失真,针对这个不足还在不断地研究中。

参考文献:

[1] 杨行峻,迟惠生. 语音信号与数字处理[M]. 北京: 电子工业出版社, 1995: 391—400.

[2] 曹晓琳,张素莉,吴平,等. 基于 MATLAB 的谱相减语音增强算法的研究[J]. 计算机仿真, 2006, 23(3): 278—283.

[3] 易克出,田斌,付强. 语音信号处理[M]. 北京: 国防工业出版社, 2000.

[4] 金学骥. 语音增强算法的研究与实现[D]. 杭州: 浙江大学, 2005: 12—13.

[5] Preuss R D. A Frequency Domain Noise Canceling Preprocessor for Narrowband Speech Communications Systems[C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. [s. l.]: [s. n.], 1979: 212—215.

[6] Benouiti M. Enhancement of Speech Corrupted by Acoustic Noise[C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. [s. l.]: [s. n.], 1979: 208—211.

[7] Soo Kim Nam, Chang Joon-Hyuk. Voice activity detection based on complex laplacian model[J]. Electronics Letters, 2003 (7): 632—633.

[8] 李富强,万红,黄俊杰. 基于 MATLAB 的语谱图显示与分析[J]. 微计算机信息, 2005, 21(30): 172—174.

[9] 王振力,张雄伟,郑翔等. 一种新的子域语音增强方法[J]. 信号处理, 2006, 22(3): 327—328.

(上接第 62 页)

[4] Das G, Lin K, Mannila H, et al. Rule Discovery from Time Series[C]// In: Proceeding of the 3rd International Conference of Knowledge Discovery and Data Mining. Gregory P, California: [s. n.], 1998: 16—22.

[5] Mark L, Klein Y, Kandel A. Knowledge Discovery in Time Series Databases[C]// In: IEEE Transaction on Systems, Man, and Cybernetics—Part B: Cybernetics. Gregory P, California: [s. n.], 2001: 20—30.

[6] Smyth P, Goodman R M. Rule Induction Using Information

Theory[C]// In: Gregory P, William J. Knowledge Discovery in Databases. Cambridge: MIT Press, 1991: 159—176.

[7] Agrawal R, Mannila H, Srikant R, et al. Fast Discover of Association Rules[C]// In: Fayyad M, Piatetshy—Shapiro G, Smyth P. Advance in Knowledge Discovery and Data Mining. Menlo Park, California: AAAI/ MIT Press, 1996: 307—328.

[8] Liu B, Hsu W, Ma Y. Integrating Classification and Association Rule Mining[C]// in Proc. of the 4th Int’l Conf. on Knowledge Discovery and Data Mining. New York, NY: [s. n.], 1998.