

基于非因果先验信噪比估计的语音增强改进算法^①

杨波, 王新房

(西安理工大学 自动化与信息工程学院, 西安 710048)

摘要: 为提高 MMSE-LSA 语音增强算法在低信噪比下的语音增强效果, 提出一种改进的 MMSE-LSA 算法。该算法采用非因果先验信噪比估计法来估计先验信噪比, 并引入无语音概率的思想, 对增益函数进行改进。实验结果表明, 相比传统 MMSE-LSA 算法, 改进算法能更好地抑制残留噪声, 提高语音的信噪比, 增强效果更好。

关键词: 语音增强; 非因果先验信噪比估计; 无语音概率; 增益函数

Improved Speech Enhancement Algorithm Based on Noncausal a Priori SNR Estimation

YANG Bo, WANG Xin-Fang

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: To improve the effect of MMSE-LSA speech enhancement algorithm at low signal-to-noise ratio(SNR), this paper proposes an improved MMSE-LSA algorithm, which uses noncausal estimation for a priori SNR, and introduces a knowledge of speech absence probability so as to modify gain function from MMSE-LSA. Experimental results show that the proposed algorithm is better than traditional MMSE-LSA algorithm in residual noise suppression and SNR improvement.

Key words: speech enhancement; noncausal a priori SNR estimation; speech absence probability; gain function

语音信号在通信过程中不可避免地要受到环境以及传输信道的影响, 导致音质恶化。语音增强是解决噪声污染的有效方法, 语音增强的目的是从复杂的背景噪声中尽可能提取纯的原始语音, 消除背景噪声, 改善语音质量, 提高语音的可懂度。目前常用的语音增强算法主要包括谱减法、自适应滤波法、基于语音产生模型语音增强算法、基于统计理论的最小均方误差估计以及基于人耳听觉掩蔽效应的语音增强算法等^[1]。其中文献[2]提出的一种短时对数谱最小均方误差估计(MMSE-LSA)语音增强算法, 能够有效的降低噪声对人耳听觉感知的影响, 被广泛研究和应用。

在基于 MMSE-LSA 的语音增强算法中, 先验信噪比是关键参数^[3], 先验信噪比估计准确与否将决定增强后语音的质量高低和残留的音乐噪声大小^[4]。如何准确而有效地得到先验信噪比的估计式, 保留信号中

的非平稳成分, 并且避免出现难以接受的残存噪声, 一直是语音增强研究的重点和难点问题。

本文提出了一种改进的 MMSE-LSA 算法, 该算法采用非因果先验信噪比估计法来估计先验信噪比, 通过计算语音有、无两个状态的概率, 对增益函数进行改进。与传统的 MMSE-LSA 算法相比, 改进的算法有更好的语音增强性能。

1 MMSE-LSA语音增强算法

假设 $x(n)$ 表示纯净语音信号, $d(n)$ 表示加性平稳噪声, 且两者互不相关。带噪语音信号表示为:

$$y(n) = x(n) + d(n) \quad (1)$$

设 $Y_k = R_k \exp(j\theta_k)$ 、 D_k 、 $X_k = A_k \exp(j\alpha_k)$ 分别表示信号 $y(n)$, $d(n)$, $x(n)$ 进行 FFT 变换后的第 k 个频谱分量, θ_k , α_k 分别为 Y_k 和 X_k 的相位。MMSE-LSA 语音增强算法的目的就是得到 A_k 的估计值 \hat{A}_k , 最小化

^① 收稿时间:2011-10-11;收到修改稿时间:2011-11-17

$E\{(\ln A_k - \ln \hat{A}_k)^2\}$ 得到 \hat{A}_k 的值为:

$$\hat{A}_k = \exp \{E[\ln A_k | Y_k]\} \quad (2)$$

根据文献[2]的方法, 可以通过先验信噪比和后验信噪比估计得到增益函数为:

$$G(\xi_k, \gamma_k) = \frac{\hat{A}_k}{R_k} = \frac{\xi_k}{\xi_k + 1} \exp \left\{ \frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt \right\} \quad (3)$$

其中, $v_k = \frac{\xi_k}{\xi_k + 1} \gamma_k$; ξ_k 和 γ_k 分别表示带噪语音信号的先验信噪比和后验信噪比值。 ξ_k 和 γ_k 分别定义为:

$$\xi_k = \lambda_x(k) / \lambda_d(k) \quad (4)$$

$$\gamma_k = R_k^2 / \lambda_d(k) \quad (5)$$

其中, $\lambda_x(k) = E[X_k^2]$; $\lambda_d(k) = E[D_k^2]$ 。

从式(3)可以看出, 这个增益函数只与先验信噪比 ξ_k 和后验信噪比 γ_k 有关, 先验信噪比 ξ_k 对最终的增益函数值的确定有很大影响。传统 MMSE-LSA 算法采用 Ephraim 和 Malah 提出的直接判决方法^[2]来估计先验信噪比。其计算公式如下:

$$\xi_k(l) = \alpha \frac{\hat{A}_k^2(l-1)}{\lambda_d(k, l-1)} + (1-\alpha) \max[\gamma_k(l) - 1, 0] \quad (6)$$

其中, l 为当前帧数; $l-1$ 为前一帧数; α 为平滑系数。

2 改进的MMSE-LSA语音增强算法

本文主要从先验信噪比估计和增益函数这两方面对传统的 MMSE-LSA 算法进行改进。由式(6)可知, 直接判决方法只利用了带噪语音信号的当前帧和当前帧之前的数据, 并没有利用到后面帧的数据, 据此文献[5]提出非因果先验信噪比估计方法, 该方法利用当前帧之前、当前帧和当前帧之后的数据来联合估计先验信噪比, 由于利用当前帧之后的数据, 因此, 估计值更准确, 更接近实际先验信噪比值。另外, 传统 MMSE-LSA 算法下的增益函数没有考虑到在噪声环境下语音出现的不确定性, 据此文献[6]提出基于概率的语音判决方法, 该方法很好地利用语音有、无的概率, 对增益函数进行修正, 得到最佳增益函数。

2.1 非因果先验信噪比估计方法

据文献[5]可知, 先验信噪比 $\xi_k(l)$ 可通过 $\hat{\xi}_{l|l+L}(k, l)$ 估计值得到, $\hat{\xi}_{l|l+L}(k, l)$ 定义为:

$$\hat{\xi}_{l|l+L}(k, l) = \frac{\hat{\xi}'_{l|l+L}(k, l)}{1 + \hat{\xi}'_{l|l+L}(k, l)} \left(1 + \frac{\hat{\xi}'_{l|l+L}(k, l) \gamma(k, l)}{1 + \hat{\xi}'_{l|l+L}(k, l)} \right) \quad (7)$$

其中, $\hat{\xi}_{l|l+L}(k, l)$ 表示在第 $0 \sim l+L$ 帧带噪语音信号条件下所估计出来的先验信噪比; $\hat{\xi}'_{l|l+L}(k, l)$ 表示在除当前帧之外第 $0 \sim l+L$ 帧带噪语音信号条件下所估计出来的先验信噪比。

$$\hat{\xi}'_{l|l+L}(k, l) = \max \left\{ \mu \hat{A}^2(k, l-1) / \lambda_d(k, l-1) + (1-\mu) [\mu' \hat{\xi}_{l-1|l+L-1}(k, l-1) + (1-\mu') \hat{\xi}_{l|l+L}(k, l)], \xi_{\min} \right\} \quad (8)$$

其中, $\hat{\xi}_{l|l+L}(k, l)$ 表示在除当前帧之外第 $l \sim l+L$ 帧带噪语音信号下所估计出来的先验信噪比。 μ 和 μ' 为平滑因子; ξ_{\min} 表示估计的先验信噪比最小阈值。

$$\hat{\xi}_{l|l+L}(k, l) = \max \left\{ \frac{1}{L} \sum_{n=1}^L \gamma(k, l+n) - \beta, 0 \right\} \quad (9)$$

其中, $\beta (\beta \geq 1)$ 表示过减因子。

通过以上过程就可以得到非因果先验信噪比估计方法下的先验信噪比 $\xi_k(l)$ 。

2.2 改进最佳增益函数

传统的 MMSE-LSA 算法对语音谱的估计是在假定语音出现条件下进行估计的, 这显然与实际情况不符。考虑到在噪声环境下语音出现的不确定性, 假设 H_0^k, H_1^k 分别表示无语音和有语音两种状态, 并且认为语音和噪声的短时傅立叶变换的系数符合高斯复分布。两种状态下带噪语音信号 Y_k 表示为:

$$\begin{cases} H_0^k : Y_k = D_k \\ H_1^k : Y_k = X_k + D_k \end{cases} \quad (10)$$

据文献[6]中的方法, 式(2)改写为:

$$\hat{A}_k = \exp \left\{ E[\ln A_k | Y_k, H_1^k] P(H_1^k | Y_k) + E[\ln A_k | Y_k, H_0^k] P(H_0^k | Y_k) \right\} \quad (11)$$

其中, $P(H_1^k | Y_k)$ 和 $P(H_0^k | Y_k)$ 分别代表有语音的后验概率和无语音的后验概率, 并分别用 p_k 和 $1-p_k$ 表示; p_k 是语音存在的概率, 定义为:

$$p_k = \frac{1 - q_k}{1 - q_k + q_k (1 + \xi_k) \exp(-v_k)} \quad (12)$$

其中, q_k 是带噪语音信号在第 k 个频谱分量无语音时的先验概率, 其定义如下:

$$q_k = 1 - P_{local} P_{global} P_{frame} \quad (13)$$

其中, P_{local} 和 P_{global} 分别表示当前帧的局部频域窗的语音存在概率和全局频域窗的语音存在概率; P_{frame} 表示整个当前帧的语音存在概率。基于前面二元假设模型: 有语音和无语音时, 分别有:

$$\begin{aligned} \exp\{E[\ln A_k | Y_k, H_1^k]\} &= G_{H_1}(\xi_k, \gamma_k) R_k \\ \exp\{E[\ln A_k | Y_k, H_0^k]\} &= G_{\min} R_k \end{aligned} \quad (14)$$

其中, G_{\min} 是无语音时的最小增益函数; $G_{H_1}(\xi_k, \gamma_k)$ 表示语音存在条件下的增益函数, 类似于 MMSE-LSA, 计算公式推导为:

$$G_{H_1}(\xi_k, \gamma_k) = \frac{\xi_k}{\xi_k + 1} \exp\left\{\frac{1}{2} \int_{\nu_k}^{\infty} \frac{e^{-t}}{t} dt\right\} \quad (15)$$

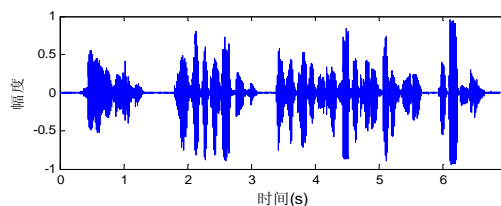
因此, 最后的增益函数可以表示为:

$$G(\xi_k, \gamma_k) = G_{H_1}(\xi_k, \gamma_k)^{p_k} G_{\min}^{1-p_k} \quad (16)$$

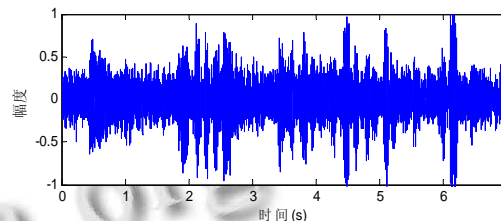
3 实验结果及评价

实验中纯净语音是在安静环境下录制的语音数据, 噪声是取自于 Noisex-92 噪音库中的白噪声(White)、坦克噪声(M109)和战斗机机舱噪声(F16), 合成信噪比为-5dB、0dB、5dB、10dB 的带噪语音, 对带噪语音进行 8kHz 采样, 16 位线性量化。实验中帧长为 256 点, 帧移 128 点, 加窗处理采用 Hamming 窗。

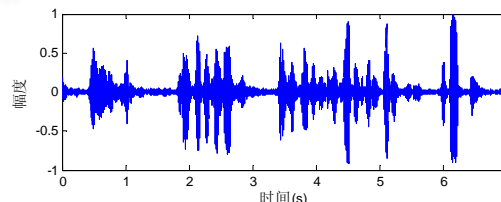
以信噪比为 0 dB, M109 噪声污染的语音数据降噪结果为例, 图 1(a)为纯净语音, 图 1(b)为带噪语音, 图 1(c)是 MMSE-LSA 增强后的语音, 图 1(d)改进算法增强后的语音。从图 1 中可以看出, 传统 MMSE-LSA 算法有一定的降噪能力, 但处理后的带噪语音仍然有明显的残余噪声。改进算法处理后的带噪语音, 降噪效果更好, 残余噪声减弱, 经主观试听, 处理后带噪语音的可懂度、自然度均得到了明显的改善, 更接近原始语音。



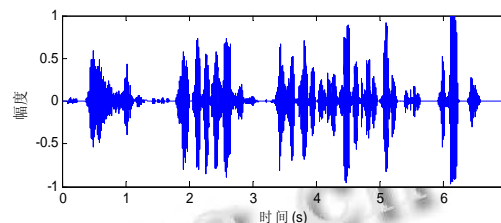
(a) 纯净语音



(b) 带噪语音



(c) MMSE-LSA 增强后的语音



(d) 改进算法增强后的语音

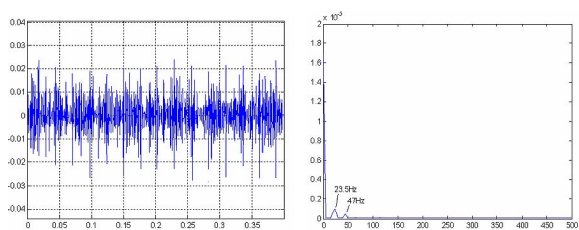
图 1 两种方法增强后的语音时域波形图 (输入信噪比 0dB, 噪声为 M109)

表 1 列出了不同信噪比下各算法增强后分段信噪比的增加量。由表 1 可以看出, 改进算法的增强效果优于传统 MMSE-LSA 算法, 噪声被更好地消除, 分段信噪比的增加量得到进一步提高。

表 1 各种增强算法分段信噪比增加量对比表 (单位: dB)

增强算法	White Noise				M109 Noise				F16 Noise			
	-5	0	5	10	-5	0	5	10	-5	0	5	10
MMSE-LSA	5.29	4.61	3.63	2.74	5.26	4.70	4.14	3.60	4.81	3.88	2.91	2.30
改进算法	7.22	6.45	5.53	4.50	7.02	6.63	5.90	5.17	6.82	5.88	4.99	4.09

(下转第 176 页)



(a)小波消噪振动信号 (b)循环自相关函数幅值切片

图9 经过小波预处理振动信号及循环谱切片图

图9(a)为振动信号经过小波重构后的信号,可以看出,重构信号与原始信号波形已经有了很大的差异,信号中的冲击特征也得到了很大程度的强化,并且,原始信号中原来比较明显的冲击点仍然能够保留得很完整。图9(b)为小波重构后的信号在循环频率等于23.5Hz处的切片谱,在切片谱中可以很明显的看出23.5Hz和其2倍频47Hz。

通过以上试验对比,当滚动轴承信号发生早期故障时,由于背景信号的干扰,测得的振动信号循环平稳性大大降低,如果直接对其进行循环平稳分析,效

果往往不太理想,采用连续小波变换与重构,能够有效地强化信号中的冲击特征,从而增强信号的循环平稳性,在此基础上再进行二阶循环统计量分析,能够比较有效地对滚动轴承的早期故障进行诊断。

参考文献

- 1 樊永生. 机械设备诊断的现代信号处理方法. 北京: 国防工业出版社, 2009.
- 2 盛兆顺, 尹琦玲. 设备状态监测与故障诊断技术及应用. 北京: 化学工业出版社, 2003.
- 3 何正嘉, 等. 机械设备非平稳信号的故障诊断原理及应用. 北京: 高等教育出版社, 2001.
- 4 何正嘉, 等. 现代信号处理及工程应用. 西安: 西安交通大学出版社, 2007.
- 5 余红英. 齿轮振动信号分解及其在故障诊断中的应用. 振动、测试与诊断, 2005, (2): 109-113.
- 6 樊永生, 郑钢铁. 振动信号检测技术研究及其在故障诊断中的应用. 应用力学学报, 2006, 23(3): 388-392.

(上接第202页)

4 结语

本文首先介绍了传统MMSE-LSA语音增强算法,并在此基础上提出了一种改进的算法。该算法通过非因果先验信噪比估计法来估计先验信噪比,较好地弥补了传统MMSE-LSA算法中先验信噪比估计的不足;同时,考虑到在噪声环境下语音出现的不确定性,通过计算语音有、无两个状态的概率,对增益函数进行修正,得到最佳增益函数。实验结果表明,相比传统MMSE-LSA算法,改进算法处理后的增强语音在残余噪声、信噪比和语音失真方面有很大改善。

参考文献

- 1 Loizou P. Speech Enhancement: Theory and Practice. Boca Raton, FL: CRC Press, 2007.

- 2 Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. on Acoustics Speech and Signal Processing, 1985, 33(2): 443-445.
- 3 杨秋成, 范炜玮. 基于先验信噪比估计的语音增强方法. 信号处理, 2008, 24(2): 329-332.
- 4 郑成诗, 周崑, 李晓东. 基于联合语音出现概率的先验信噪比估计算法. 电子与信息学报, 2008, 30(7): 1680-1683.
- 5 Cohen I. On the Decision-directed Estimation Approach of Ephraim and Malah. IEEE Int'l Conf. on Acoustics Speech and Signal Processing, ICASSP'04. 2004, (1): 293-296.
- 6 Cohen I. Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. IEEE Signal Processing Letters, 2002, 9(4): 113-116.