基于 MMSE-MLSA 与感知滤波的语音增强算法

董 胡,马振中,赵 娜,刘 刚,童 欣
(长沙师范学院 信息科学与工程学院,湖南 长沙 410100)

摘 要:在语音通信过程中,纯净的语音信号可能受到各种不同类型的干扰噪声信号的影响,例如白噪声、色噪声等。针对常见语音增强算法在低信噪比的复杂噪声环境下语音增强后存在语音失真及残余噪声的问题,提出了一种结合改进对数谱幅度的最小均方误差(MMSE-MLSA)谱估计与感知滤波的语音增强算法。该算法采用 MMSE-MLSA 对含噪语音作初级谱估计增强处理,使用次级感知滤波器进一步掩蔽初级增强信号中的残余音乐噪声。仿真实验结果表明,在低信噪比的复杂噪声环境下,该算法能有效降低语音失真及去除残余音乐噪声,与另外两种语音增强算法比较,增强效果更加突出。

关键词:语音增强;最小均方误差;感知滤波;掩蔽阈值;谱估计
中图分类号:TN912.3 **文献标识码**:A **文章编号**:1673-629X(2019)08-0067-04
doi:10.3969/j.issn.1673-629X.2019.08.013

Speech Enhancement Algorithm Based on MMSE- MLSA and Perceptual Filtering

DONG Hu, MA Zhen-zhong, ZHAO Na, LIU Gang, TONG Xin
(School of Information Science and Engineering, Changsha Normal University, Changsha 410100, China)

Abstract: In the process of speech communication, pure speech signal may be affected by various types of interference noise signals, such as white noise, color noise, etc. In order to solve the problem of speech distortion and residual noise in speech enhancement algorithms under complex noise environment with low SNR, a speech enhancement algorithm based on minimum mean square error-modified log-spectral amplitude (MMSE-MLSA) spectrum estimation and perceptual filtering is proposed. The noisy speech is enhanced by MMSE-MLSA as primary spectrum estimation, then a secondary perceptual filter is used to mask the residual music noise after primary enhancement. The simulation shows that the proposed algorithm can effectively reduce speech distortion and remove residual music noise in complex noise environment with low SNR. Compared with the other two algorithms of speech enhancement, the enhancement effect of the method proposed is more prominent.

Key words: speech enhancement; minimum mean square error; perceptual filtering; masking threshold; spectral estimation

0 引 言

当前,常见的语音增强算法众多,诸如:谱减法、维纳滤波法、小波包去噪、MMSE-LSA 法等。谱减法及维纳滤波法总体来说计算量稍小,易实现,但也易出现音乐噪声[1-5]。小波包去噪法有较强的时频分析能力,适合非平稳信号处理,但阈值的设定是小波包去噪的关键点,阈值太大或太小都将影响去噪效果[6-8]。MMSE-LSA 算法的语音增强效果优于谱减法、维纳滤波法和小波包去噪法,但需要预测或假设语音频谱的分布,在低信噪比的复杂噪声环境下,其语音增强效果有待改善[9-10]。

针对上述语音增强算法所描述的问题,提出了一种改进对数谱幅度最小均方误差谱估计(MMSE-MLSA)与感知滤波结合的语音增强算法。该算法将降噪和噪声掩蔽进行单独处理,首先采用 MMSE-LSA 对含噪语音进行初级降噪,接着使用感知滤波器将初级降噪后残余噪声掩蔽掉。仿真实验结果表明,在低信噪比的复杂噪声环境下,与常见的谱减法及 MMSE-MLSA 相比较,该算法增强后的语音失真及残余音乐噪声更小,增强效果更明显。

收稿日期:2018-10-07 **修回日期**:2019-02-20 **网络出版时间**:2019-03-28
基金项目:国家自然科学基金(11474090,11774088);湖南省教育厅优秀青年项目(17B025);湖南省自然科学基金青年项目(2018JJ3557)
作者简介:董 胡(1982-),男,硕士,副教授,研究方向为信号处理及嵌入式设计。
网络出版地址:http://kns.cnki.net/kcms/detail/61.1450.TP.20190327.1633.068.html

1 语音增强算法原理

图 1 显示了文中算法的原理。含噪语音信号先通过改进对数谱幅度最小均方误差谱估计作增强处理，然后使用感知滤波器掩蔽上一级增强信号中的残余噪声，最终获得增强后的语音。从图 1 可知，整个增强算法分为四个部分：MMSE-MLSA 谱估计、噪声估计、掩蔽阈值估计和感知滤波器。

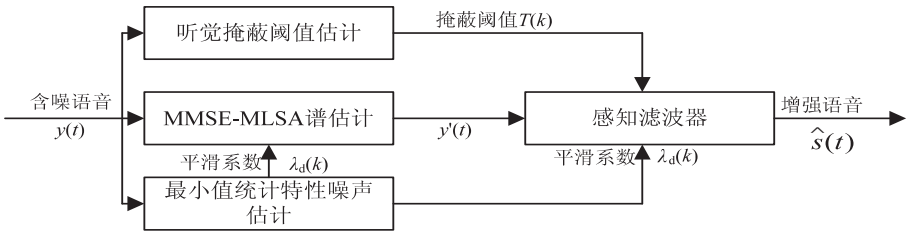

图 1 MMSE-MLSA 语音增强算法原理

1.1 MMSE-MLSA 谱估计

设 $s(t)$ 为纯净语音信号，$n(t)$ 为噪声信号，$y(t)$ 为含噪语音信号，若仅考虑加性噪声，有如下表达式[11-12]：

$$y(t) = s(t) + n(t) \tag{1}$$

令 $Y(k)$、$S(k)$、$N(k)$ 分别表示 $y(t)$、$s(t)$、$n(t)$ 作 FFT 变换后所对应的第 k 个频谱幅度，并假设语音与噪声统计是独立的，则有：

$$Y(k) = S(k) + N(k) \tag{2}$$

假定语音增强的谱增益函数为 $G_S(k)$，估计的纯净语音幅度谱为 $\hat{S}(k)$，有：

$$\hat{S}(k) = G_S(k)Y(k) \tag{3}$$

相比于 MMSE 估计法[13]，MLSA-MMSE 估计法更适合人耳的听觉特性，能更好地抑制噪声，故文中语音增强算法初级选择 MLSA-MMSE 估计法。对 MLSA-MMSE 估计法的谱增益函数 $G_S(k)$ 作如下定义：

$$G_S(k) = \frac{\xi(k)}{1+\xi(k)}\exp\left\{\frac{1}{2}\int_{v(k)}^{\infty}\frac{e^{-t}}{t}dt\right\} \tag{4}$$

其中，$\xi(k)$ 为先验信噪比；$\gamma(k)$ 为后验信噪比，则有[14]：

$$\xi(k) = \eta\frac{\lambda_x^2(k)}{\lambda_n(k)} + (1-\eta)\max[0,\gamma(k)-1] \tag{5}$$

$$\gamma(k) = Y^2(k)/\lambda_n(k) \tag{6}$$

$$v(k) = \xi(k)\gamma(k)/(1+\xi(k)) \tag{7}$$

假设 $H_0(k)$ 和 $H_1(k)$ 分别表示语音缺失和存在，并且假设对于语音和噪声短时傅里叶变换系数的复高斯分布，信号的条件概率密度作如下定义：

$$p(Y(k)\mid H_0(k)) = \frac{1}{\pi\lambda_n(k)}\exp\left\{-\frac{|Y(k)|^2}{\lambda_n(k)}\right\} \tag{8}$$

$$p(Y(k)\mid H_1(k)) = \frac{1}{\pi(\lambda_x(k)+\lambda_n(k))}\exp\left\{-\frac{|Y(k)|^2}{\lambda_x(k)+\lambda_n(k)}\right\} \tag{9}$$

其中，η 是权重因子（$0 \leqslant \eta \leqslant 1$），这里取 $\eta = 0.98$。$\lambda_x(k) \triangleq E[|X(k)|^2H_1(k)]$、$\lambda_n(k) \triangleq E[|N(k)|^2]$ 分别是语音和噪声第 k 个谱分量的数学期望。根据贝叶斯规则，语音存在的概率条件 $p(k) \triangleq P(H_1(k)\mid Y(k))$ 可表示如下[15]：

$$p(k) = \left\{1+\frac{q(k)}{1-q(k)}(1+\xi(k))\exp(-v(k))\right\}^{-1} \tag{10}$$

其中，$q(k) \triangleq P(H_0(k))$ 表示非语音的先验概率。

令 $A = |S|$ 代表语音谱幅度，谱增益函数 $G_{\min}$ 作如下定义：

$$\exp\{[\log A(k)\mid Y(k),H_0(k)]\} = G_{\min}\cdot|Y(k)| \tag{11}$$

$\hat{G}_S(k)$ 最终定义如下：

$$\hat{G}_S(k) = \{G_s(k)\}^{p(k)}\cdot G_{\min}^{1-p(k)} \tag{12}$$

1.2 噪声估计

作为语音增强算法中的重要组成部分，如果噪声估计过高，则弱语音将被消除，增强后的语音将出现失真；如果估计过低，则增强后的语音将残留过多的噪声。基于最小值统计特性，估计算法能使估计的噪声较好地跟踪噪声改变。所以，在该算法中，噪声估计选择最小值统计特性算法。

1.3 听觉掩蔽阈值估计

听觉掩蔽是听觉系统的一个心理声学特性，在音频编码中应用广泛。通过模拟人耳的频率选择特性和掩蔽特性来计算掩蔽阈值。在对掩蔽阈值作计算之前，语音谱需作粗略估计。其中，语音谱的粗略值可通过下式进行估计：

$$\tilde{S}_p(k) = \max(Y_p(k) - 2\hat{N}(k),\varepsilon) \tag{13}$$

利用语音谱的粗略估计值 $\tilde{S}_p(k)$，结合 Johnston 模型计算听觉掩蔽阈值。

1.4 感知滤波器

含噪信号经初级增强后，存在一定的残留噪声，其

可以被人耳的听觉掩蔽特性掩盖而不被完全去除。如果它被完全去掉,则可能降低语音的可懂度,导致语音失真。因此,基于听觉掩蔽效应的感知过滤器被用作过滤处理。

假设增强后的语音信号 $\hat{S}(k) = G(k) \times S(k)$,其中 $G(k)$ 为感知滤波器,定义如下:

$$|G(k)|^2 \times |N(k)|^2 \leqslant T(k) \tag{14}$$

其中, $T(k)$ 为掩蔽阈值。

感知滤波器模型定义如下:

$$G(k) = \min\left[\theta \times \sqrt{\frac{T(k)}{|N(k)|^2}}, 1\right] \tag{15}$$

其中, $0 < \theta < 1$。通过实验取 $\theta = 0.8$。

2　实验与结果分析

实验用的语音数据采样率为 16 kHz,帧长为 512,重叠 1/2,每一帧添加 Hanming 窗。实验用的噪声来自 Noisex-92 数据库中的白噪声、factory 噪声和 M109 坦克噪声。将上述噪声信号和纯净语音信号混合成 10 dB、5 dB、0 dB、-5 dB 的含噪语音信号。

分别采用谱减算法、MMSE-LSA 算法及文中提出的算法对含噪(M109)语音作增强处理,结果如图 2 所示。从图 2 可知,对于语音信号中语音幅值较弱的部分,谱减算法和 MMSE-LSA 算法的增强效果都不佳,尤其是谱减算法,几乎完全丢失了语音幅值较弱的信号;而文中提出的算法不仅能较好地去除含噪语音中的 M109 噪声,同时能较好地恢复出原来语音幅值较弱的部分。

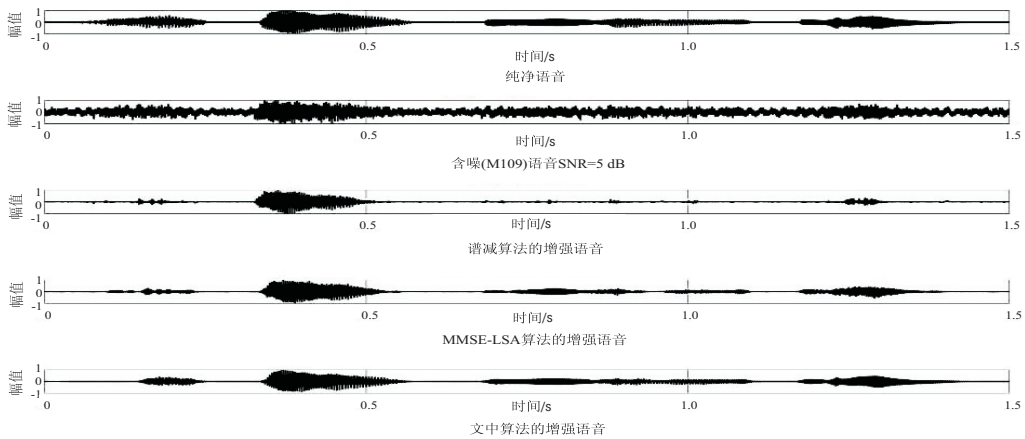

图 2　含噪(M109)语音 SNR=-5 dB 的语音增强结果

2.1　分段信噪比(SEGSNR)

利用 SEGSNR 的提高量来衡量噪声的衰减量:

$$\mathrm{SEGSNR} = \frac{10}{L}\sum_{i=0}^{L-1} \log_{10} \frac{\sum_{k=0}^{N-1} s^2(k,i)}{\sum_{k=0}^{N-1} [\hat{s}(k,i) - s(k,i)]^2} \tag{16}$$

其中, L 表示帧数; N 表示帧采样点。

通常 SEGSNR 越大,表示信号中包含的噪声和语音失真越小,相应波形越接近纯净语音。

对一定信噪比的含噪语音分别采用谱减法、MMSE-LSA 和文中算法进行语音增强仿真测试,结果如图 3 所示。可以看出,文中提出的语音增强算法 SEGSNR 提高量最大。

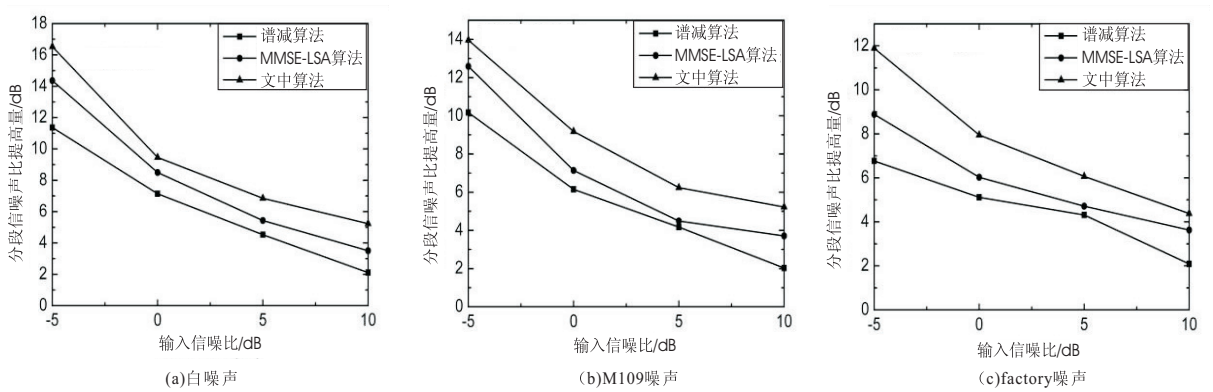

图 3　各种算法增强后 SEGSNR 的提高量

2.2 MOS 得分

MOS 得分测试由 10 名本专业学生(男女各 5 人)进行语音试听,由试听者对原始语音和增强后语音作对照测听,给出主观得分,结果如图 4 所示。

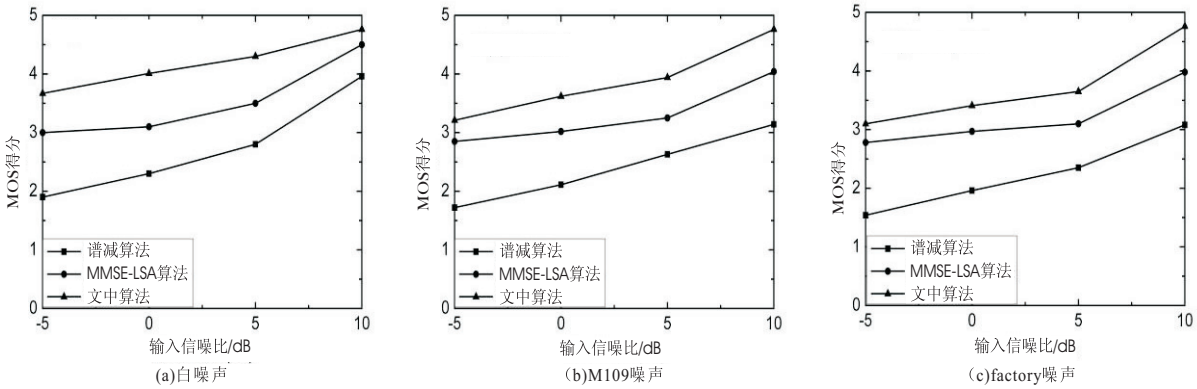

图 4 各种算法在不同 SNR 下的 MOS 得分

从图 4 可知,文中算法的 MOS 得分最高,MMSE-LSA 次之,谱减法增强后语音中存在更多的残余音乐噪声,且主观听觉较差,因此增强后的得分最低。而文中算法对增强后的信号中的噪声作掩蔽处理,因此主观评价较高,虽然存在少量的背景噪声,但音乐噪音的减少更明显,主观听觉更好,分数更高。

3 结束语

文中提出了基于 MMSE-MLSA 与感知滤波的语音增强算法。语音增强算法分为两级,初级采用 MMSE-MLSA 对含噪语音作谱估计增强处理,去除含噪语音中的大部分噪声。针对初级语音增强中存在的残余噪声,次级使用感知滤波器对初级增强后的信号进行感知滤波,进一步去除信号中的残余音乐噪声。仿真实验结果表明,在低信噪比的复杂噪声环境下,与谱减算法及 MMSE-LSA 算法相比较,该算法能有效降低语音失真及去除残余音乐噪声,语音增强效果更明显。

参考文献:

[1] 陈 琪,郭 英,张 群,等. 基于听觉感知的 LSA-MMSE 改进型语音增强方法[J]. 信号处理,2008,24(6):1037-1040.

[2] 张 勇,刘 轶. 非平稳噪声环境下结合听觉掩蔽的语音增强[J]. 计算机工程与设计,2015,36(5):1279-1284.

[3] 董 胡. 低信噪比环境下改进的语音端点检测算法[J]. 计算机技术与发展,2016,26(3):71-74.

[4] 殷 明,孔冉冉. 基于可调 Q-因子小波变换的语音增强算法[J]. 计算机应用研究,2014,31(11):3316-3319.

[5] LU C T,LEI C L,SHEN J H,et al. Noise reduction using spectral-subtraction algorithm with over-subtraction and spectral-reservation factors adapted by harmonic properties [J]. Noise Control Engineering Journal,2017,65(6):509-521.

[6] SANAM T F, SHAHNAZ C. Noisy speech enhancement based on an adaptive threshold and a modified hard thresholding function in wavelet packet domain[J]. Digital Signal Processing,2013,23(3):941-951.

[7] 任永梅,张雪英,贾海蓉. 一种新阈值函数的小波包语音增强算法[J]. 计算机应用研究,2013,30(1): 114-116.

[8] 田玉静,左红伟,董玉民,等. Bark 子带小波包自适应阈值语音去噪方法[J]. 计算机应用,2010,30(11):3111-3114.

[9] HONG K K,ROSE R C. Cepstrum-domain model combination based on decomposition of speech and noise using MMSE-LSA for ASR in noisy environments[J]. IEEE Transactions on Audio, Speech, & Language Processing, 2009,17(4):704-713.

[10] NARIMANI G,MARTIN P A,TAYLOR D P. Spectral analysis of fractionally-spaced MMSE equalizers and stability of the LMS algorithm[J]. IEEE Transactions on Communications,2018,66(4):1675-1688.

[11] WEI Jie, WANG Ming, YANG Hongxiao. MMSE-LSA based wavelet threshold denoising algorithm for low SNR speech[C]//IEEE international conference on digital signal processing. Beijing,China:IEEE,2017:253-256.

[12] HSIEH H J,CHEN B,HUNG J W. Histogram equalization of contextual statistics of speech features for robust speech recognition[J]. Multimedia Tools & Applications, 2015, 74(17):6769-6795.

[13] 余建潮,张瑞林. 改进增益函数的 MMSE 语音增强算法[J]. 计算机工程与设计,2010,31(14):3287-3289.

[14] LV T,ZHANG H Y,YAN C H. Double mode surveillance system based on remote audio/video signals acquisition[J]. Applied Acoustics,2018,129:316-321.

[15] EPHRAIM Y,MALAH D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator [J]. IEEE Transactions on Acoustics,Speech,& Signal Processing,1984,32(6):1109-1121.