

基于掩蔽效应的改进型自适应语音增强算法

龚亮¹ 张艳萍¹

摘要

在分析经典自适应滤波算法的基础上提出了一种有效的消除噪声且减小语音失真的语音增强方法. 为了进一步提高语音增强效果, 根据人耳掩蔽效应把语音信号中关键频率段的阈值作为自适应滤波算法的动态系数, 根据此系数估算误差函数. 在 Matlab 上对此算法输入不同信噪比的信号, 仿真结果表明: 输出的信噪比明显高于传统滤波算法, 且残留噪声较小.

关键词

语音增强; 听觉掩蔽效应; 自适应滤波

中图分类号 TN912.3

文献标志码 A

收稿日期 2010-07-10

作者简介

龚亮, 男, 硕士生, 主要研究方向为数字信号处理. richardgung@163.com

张艳萍, 女, 博士, 教授, 研究领域为水声通信、自适应信号处理、嵌入式应用系统. zypgjc@163.com

0 引言

Introduction

为了增强语音系统的实际应用和处理能力, 人们在过去的几十年中提出了许多语音增强算法. 所谓语音增强就是对带噪语音进行处理, 改善语音质量, 提高语音的可懂度、清晰度和舒适度, 使人易于接受.

常用的基于短时谱估计的算法有谱减法、维纳滤波法和最小均方误差估计的方法等. 谱减法^[1]在输入信噪比较高时, 能较好地抑制背景噪声, 但信噪比较低时性能下降, 残留音乐噪声较大. 维纳滤波算法与谱减法的出发点不同, 但形式上差别不大, 在信噪比降低时, 增强后的语音会发生突变, 其清晰度与可懂度均随之下降, 音乐噪声成分也有所增加. 基于最小均方准则的 LMS(Least Mean Square) 算法^[2]是将最小均方误差准则应用于频域, 该算法能抑制音乐噪声, 但在低信噪比时音乐噪声和语音失真现象依然存在^[3]. 为了更好地提高语音增强效果, Virag^[4]利用人耳掩蔽效应获得谱相减的参数, 抑制噪声, 减小语音失真. 听觉掩蔽效应的基本原理就是能量较高的信号可以使能量较低的信号不易被察觉, 因此应用听觉掩蔽效应进行语音增强时, 语音信号能够掩蔽与其同时进入人耳的一部分能量较小的噪声信号. 本文利用人耳的听觉掩蔽效应对最小均方算法进行改进, 取得了较好的噪声抑制效果. 仿真结果证明: 增强后语音的信噪比和试听质量得到了提高.

1 LMS 算法基本原理

The basic principles of LMS algorithm

最小均方(LMS)算法的工作原理如图1所示.

假定语音信号服从拉普拉斯分布, 噪声服从高斯分布^[5], 输入信号矢量 $x(n) = [x(n) \ x(n-1) \ \cdots \ x(n-L)]^T$, $y(n)$ 为输出信号, $d(n)$ 为参考信号, $e(n)$ 为 $y(n)$ 与 $d(n)$ 的误差信号. 由零时刻的滤波器滤波系数矢量估值 $w(n)$, $x(n)$, $d(n)$ 计算得到误差信号

$$e(n) = d(n) - x(n)^T w(n). \quad (1)$$

经典 LMS 算法迭代公式^[6]为

$$w(n+1) = w(n) + 2\mu e(n) x(n). \quad (2)$$

将式(1)代入式(2)得到权系数的更新估值 $w(n+1)$, 如果时间

¹ 南京信息工程大学 电子与信息工程学院, 南京 210044

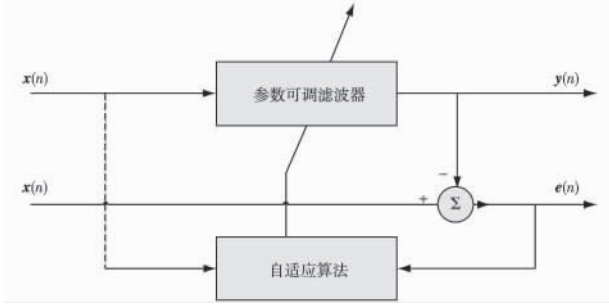


图1 自适应滤波器原理框

Fig.1 Block diagram of adaptive filter

指数 n 增加 1, 就重复计算步骤, 一直到达稳态为止, 就得到权系数矢量 $w(n) = [w_0(n), w_1(n), \dots, w_L(n)]$. 其中 μ 为步长参数, 必须满足:

$$0 < \mu < 2 / \lambda_{\max}. \quad (3)$$

λ_{\max} 为 $x(n)$ 自相关矩阵的最大特征值, μ 控制收敛速度和稳态误差.

2 人耳听觉掩蔽效应

The masking properties of human auditory system

由于低频纯音可有效地掩蔽高频纯音, 所以根据听觉的掩蔽效应进行语音增强时, 不需要把噪声完全抑制掉, 只要使得残留的噪声信号不被人耳所感知即可. 这样做的好处是在消除噪声的同时可以减少不必要的语音失真. 在调整步长系数 μ 的同时, 利用听觉掩蔽阈值来调整 LMS 自适应滤波算法的参数, 可以提高语音质量.

2.1 听觉掩蔽阈值的计算

掩蔽阈值的计算主要包括: 纯净语音的初步估计, 临界带的分析, 误差函数的扩展, 掩蔽阈值的偏移量, 掩蔽阈值的归一化处理, 能量公式与绝对听觉掩蔽阈值的结果比较. 为了判断信号特性是纯音还是噪声, 引入纯音系数 $\alpha(\lambda)$, 它是根据临界频带宽度的平坦度来判断的.

$$\alpha(\lambda) = \max \left\{ \min \left[-\frac{10 \times \lg G_m(\lambda)}{\lg A_m(\lambda)}, 1 \right], \rho \right\}. \quad (4)$$

式(4)中, $G_m(\lambda)$, $A_m(\lambda)$ 分别是信号功率谱密度的几何平均值和算术平均值, λ 为音调系数, 用来衡量纯音和噪声特性.

阈值偏移量计算式^[7]为

$$O_a(\lambda) = \alpha(\lambda) \times (14.5 + \lambda) + [1 - \alpha(\lambda)] \times 5.5, \quad (5)$$

掩蔽阈值的计算式为

$$P(L, n) = 10^{\log_{10}(c_i - (O_a/10))}. \quad (6)$$

其中 c_i 为第 i 个临界带宽的扩展功率谱.

将得到的掩蔽阈值扩展到各个频谱段上, 记为 $P(L, n)$. 绝对听觉阈值的计算公式为 $P_a(L, n) = 3.64f^{-0.8} - 6.5\exp(f - 3.3)^2 + 10^{-3}f^4$, (7) 则最终掩蔽阈值

$$P_m(L, n) = \max(P(L, n), P_a(L, n)). \quad (8)$$

2.2 利用掩蔽阈值改进 LMS 算法

传统固定步长 LMS 算法在收敛的前提下, μ 取值较大时, 收敛速度加快, 但稳态误差会增大; 反之, 稳态误差减小时, 收敛速度就会变慢. 利用改变步长的方法可以克服这一矛盾. 自适应过程开始时, 取较大的 μ 值以保证收敛速度, 然后让 μ 值逐渐变小, 以保证收敛后得到较小的稳态误差.

听觉掩蔽阈值 $P(L, n)$ 大, 说明在此频率段中人耳对其他相近频率段的语音信号和噪声信号的抗干扰能力较强, 此时采用较小的 μ 系数, 不会减少影响语音质量的频率部分; 如果掩蔽阈值较小, 则采用较大的 μ 系数. 现定义二重收敛步长因子 $\mu' = 2\mu / \delta_x^2$. 其中 δ_x^2 是输入信号的方差 (用时间平均代替),

$\delta_x^2 = \sum_{i=0}^L x^2(n-i) = X^T X$, $\hat{\delta}_x^2$ 是输入信号的数学期望 $\hat{\delta}_x^2 = E \left[\sum_{i=0}^L x^2(n-i) \right] = \sum_{i=0}^L E[x^2(n-i)] = E[X^T X]$. 则原收敛步长因子 μ , 二重收敛步长因子 μ' 与掩蔽阈值 $P(L, n)$ 的关系可表示为

$$\frac{P_{\max} - P(L, n)}{P(L, n) - P_{\min}} = \frac{\mu - \mu_{\min}}{\mu_{\max} - \mu} = \frac{\mu' - \mu'_{\min}}{\mu'_{\max} - \mu'}. \quad (9)$$

其中, P_{\max} 和 P_{\min} 分别为听觉掩蔽阈值 P 的最大值和最小值. 由于 LMS 算法收敛和稳定性与权系数矢量的系数数目和信号功率直接相关, 为了进一步改进 LMS 算法的收敛特性, 减小稳态均方误差, 引入常数 $\theta = \sqrt{\delta^2 - \hat{\delta}^2} / \delta$, $0 < \theta \leq 1$. θ 的取值是由输入信号决定的, 通常 θ 近似为 1. 若 $\theta = 1$, 则是常规 LMS 算法; 当 $\theta \neq 1$ 时, 滤波器的权矢量逐渐变化, 并逐渐变为 0 矢量. 将它们代入式(2), 得到

$$w(n+1) = \theta w(n) + \frac{\mu'}{P(L, n)} e(n) \operatorname{sgn}[x(n)]. \quad (10)$$

综上所述, 改进后算法的实现步骤如下:

- 1) 对带噪语音进行初步估计, 然后计算出掩蔽阈值 $P(L, n)$;
- 2) 根据掩蔽阈值和收敛因子的关系式(9)计算出二重收敛步长因子 μ' ;
- 3) 由求得的系数估计自适应滤波器的权系数

矢量 $w(n)$,最后得到纯净语音的估计值.

由于是基于掩蔽效应 (Masking Properties) 的 LMS 算法,故称此算法为 MPLMS 算法. 因为本算法的步长系数随着掩蔽阈值而不断变化,所以计算量较大. 由标准 LMS 算法的递推公式可知,自适应调整方向取决于 $e(n) \cdot x(n)$ 的符号. 对输入语音信号 $x(n)$ 取符号 $\text{sgn}[x(n)]$ 的意义就在于减少计算量. 对于高斯信号来说,对输入信号取符号也是一种变步长 LMS 算法,虽然在梯度估计性能上有一定程度的退化,但总体效果是增强的.

3 仿真和分析

Simulation and analysis

仿真分两个部分进行,先对加了噪声的语音信号进行分析,对比 LMS 算法和 MPLMS 算法的增强效果. 然后分析语音畸变的程度,调节噪声能量使输入信噪比变化,列表对比两种算法的输出信噪比.

3.1 对加入噪声的语音信号进行仿真实验

信号录入条件是采用 Windows 自带的录音机,选用抽样频率为 8 kHz、16 位、单声道的 PCM 格式进行录制. 在安静的环境下,朗读“我是中国人”,时间大约为 2 s,存储为 chinese.wav 文件. 文件采用 Matlab 函数 wavread() 调用,在 Matlab 中它以一维数组的形式存在. 噪声是利用 NOISE-92 噪声库的高斯白噪声(图 2). 最终算法的实现在 Matlab 的 GUI 工具上完成. 利用 GUI 工具可以看出传统 LMS 算法有效地削弱了噪声,但是语音信号有了一定程度的失真,且收敛速度较慢. 由图 3 和图 4 可知,不管是信号失真度还是收敛速度,MPLMS 算法都比传统最小均方算法好.

对实际试听的语音增强结果进行平均意见得分^[9] (Mean Opinion Score, MOS) 测试,也发现基于听觉掩蔽效应的 LMS 语音增强算法克服了传统 LMS 算法会引起语音畸变的缺点. 除了清晰度略有下降外,语音的懂度和自然度未受损失.

3.2 信噪比对算法的影响

选用不同能量的高斯白噪声,使带噪语音信噪比为 -5 dB、0 dB、5 dB,进行仿真实验,通过 Matlab 对 LMS 算法和 MPLMS 算法分别计算输出信噪比,结果见表 1. 由表 1 可知,在低信噪比的情况下两种算法的效果都不是很好,都不能很好地增强语音效果,但随着输入信噪比的提高,MPLMS 算法的输出

信噪比比传统 LMS 算法要高.

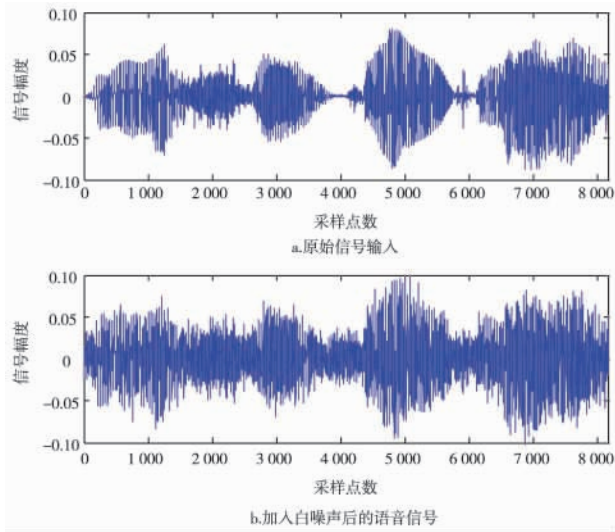


图 2 原始输入与加噪后的语音信号
Fig. 2 Original input and speech signal mixed with noise

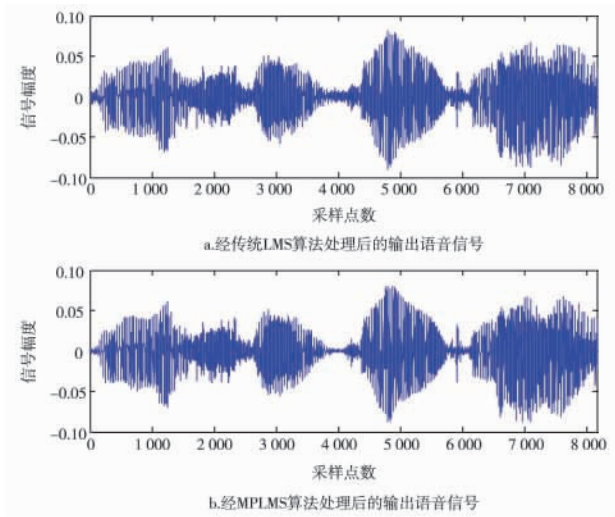


图 3 两种算法的增强效果对比
Fig. 3 Speech enhancement comparison between LMS & MPLMS

表 1 两种算法在不同输入信噪比情况下输出信噪比的比较
Table 1 The output-SNR comparison between LMS & MPLMS on the condition of different input-SNR

输入信噪比/dB	输出信噪比/dB	
	LMS 算法	MPLMS 算法
-5	-1.1	-1.7
0	3.9	4.7
5	7.2	10.8

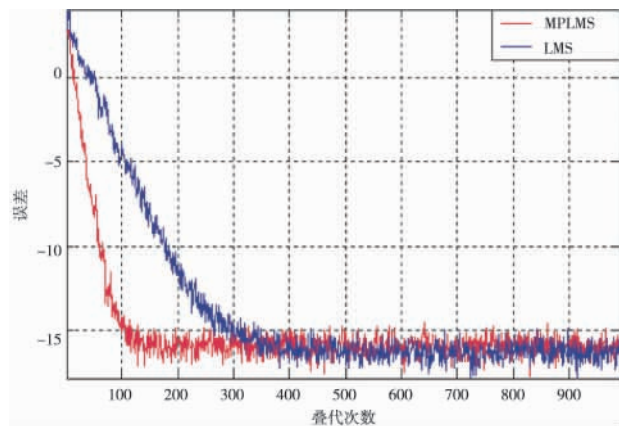


图4 两种语音增强算法的收敛曲线对比效果

Fig. 4 Convergence curve comparison of two speech enhancement algorithms

4 结束语

Concluding remarks

在分析与总结前人在自适应语音增强工作的基础上,提出了基于高斯模型和听觉掩蔽效应的 LMS 算法,并且和传统的 LMS 算法进行了仿真比较.结果表明:利用人耳的听觉掩蔽效应,有针对性地进行自适应滤波,能比较有效地兼顾去噪和保持语音信号的主观听觉质量,保存语音的自然度,提高了语音增强系统性能.

参考文献

References

[1] Poruba J. Speech enhancement based on nonlinear spectral sub-

traction[C]//Proceedings of the Fourth IEEE International Caracas Conference on Circuits and Systems 2002: 86-126

- [2] 衡霞,刘志镜. 基于自适应滤波的语音增强和噪声消除[J]. 微机发展, 2004, 14(1): 96-98
HENG Xia, LIU Zhijing. Speech enhancement and noise reduction based on self-adaptive filtering[J]. Microcomputer Development, 2004, 14(1): 96-98
- [3] 严雪艳,郭建中. 基于 LMS 自适应滤波器对噪声干扰的语音恢复研究[J]. 陕西师范大学学报: 自然科学版, 2009, 37(3): 42-46
YAN Xueyan, GUO Jianzhong. A study of renewing voice signal disturbed by noise based on Least Mean Square (LMS) adaptive filter[J]. Journal of Shaanxi Normal University: Natural Science Edition 2009, 37(3): 42-46
- [4] Virag N. Single channel speech enhancement based on masking properties of the human auditory system[J]. IEEE Transactions on Speech and Audio Processing, 1999, 7(2): 36-40
- [5] Lin P L, Huang P Y. Fusion methods based on dynamic-segmented morphological wavelet or cut and paste for multi-focus images[J]. Signal Processing, 2008, 88(6): 1511-1527
- [6] 易克初,田斌,付强. 语音信号处理[M]. 北京: 国防工业出版社, 2000: 425-489
YI Kechu, TIAN Bin, FU Qiang. Speech signal processing[M]. Beijing: National Defense Industry Press, 2000: 425-489
- [7] Quatieri T F. 离散时间语音信号处理: 原理与应用[M]. 赵胜辉,刘家康,谢湘,等译. 北京: 电子工业出版社, 2004: 530-556
Thomas F. Quatieri. Discrete-time speech signal processing: Principles and practice[M]. Prentice Hall International, Inc, 2004: 530-556
- [8] 冯流保. 基于听觉掩蔽效应的小波包语音增强[J]. 通信技术, 2010, 43(3): 138-141
FENG Liubao. Wavelet package speech enhancement based on mask property shrink[J]. Communication Technology, 2010, 43(3): 138-141
- [9] 张雄伟,陈亮,杨吉斌. 现代语音处理技术及应用[M]. 北京: 机械工业出版社, 2003: 172-176
ZHANG Xiongwei, CHEN Liang, YANG Jibin. Modern speech processing technology and applications[M]. Beijing: China Machine Press, 2003: 172-176

Modified adaptive speech enhancement based on masking properties of human auditory system

GONG Liang¹ ZHANG Yanping¹

¹ School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract Based on the analysis of classical adaptive filtering algorithm, an effective speech enhancement algorithm is proposed, which could reduce noise and distortion effectively. According to the auditory masking properties, this algorithm uses masking thresholds of the key frequency segments as dynamic coefficients, to further improve the effect of speech enhancement and estimate error function dynamically. Signals with different SNR are input to simulate this proposed algorithm on Matlab. The high SNR and low residual noise of the output signals show that the algorithm is better than traditional filtering algorithms.

Key words speech enhancement; auditory masking properties; adaptive filtering