

◎网络、通信与安全◎

一种基于快速噪声估计的 MMSE 语音增强算法

陈照平¹, 马建芬², 张雪英¹CHEN Zhao-ping¹, MA Jian-fen², ZHANG Xue-ying¹

1. 太原理工大学 信息工程学院, 太原 030024

2. 太原理工大学 计算机与软件学院, 太原 030024

1. College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China

2. College of Computer and Software, Taiyuan University of Technology, Taiyuan 030024, China

E-mail: zhaopingchen2001@yahoo.com.cn

CHEN Zhao-ping, MA Jian-fen, ZHANG Xue-ying. Minimum-Mean-Square-Error algorithm based on fast noise-estimation. Computer Engineering and Applications, 2007, 43(22): 113-114.

Abstract: A fast noise-estimation algorithm used in the speech enhancement algorithm of MMSE is studied. Experiments results prove that this algorithm can dramatically improve objective performance compared with spectrum subtraction algorithm and MMSE-LSA algorithm, and it can estimate the noise spectrum quickly in the non-stationary noise environment.

Key words: noise estimation; Minimum-Mean-Square-Error(MMSE); speech enhancement

摘要: 提出一种基于快速噪声估计的 MMSE 语音增强算法, 实验表明这种算法比起谱相减法和基于语音短时对数谱的最小均方差(MMSE-LSA)算法能更显著地提高算法的客观性能, 在非平稳噪声环境中能快速估计出变化的噪声功率谱。

关键词: 噪声估计; 最小均方差; 语音增强

文章编号: 1002-8331(2007)22-0113-02 文献标识码: A 中图分类号: TN912

语音通信领域中, 由于受到周围环境以及传输信道的影响, 纯净语音被添加了背景噪声, 影响了语音的质量与自然度。语音增强就是从带噪声语音信号中提取纯净的原始语音。目的是降低噪声分量, 提高语音清晰度和可懂度, 减轻听觉疲劳。

在大多数语音增强算法中, 都假定噪声功率谱的估计是已知的。这些噪声功率谱的估计对语音增强算法是十分重要的。如果噪声功率谱估计太低, 会听到残留噪声。同样, 如果噪声功率谱估计太高, 会损害语音信号。因此, 在非平稳环境下, 快速估计出变化的背景噪声有重要意义。

1 快速噪声估计算法^[1]

设观察到的带噪语音信号为

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

其中, $x(n)$ 是纯净语音, $d(n)$ 是噪声。先用一个递归的平滑过程来计算带噪语音的功率谱:

$$P(\lambda, k) = \eta P(\lambda - 1, k) + (1 - \eta) |Y(\lambda, k)|^2 \quad (2)$$

其中, $P(\lambda, k)$ 是平滑功率谱, λ 是帧的标号, k 是频率分量的标号, $|Y(\lambda, k)|^2$ 是带噪语音的短时功率谱, 然后用递归法则去跟踪最小值 $P_{\min}(\lambda, k)$, 而更新语音存在概率 $p(\lambda, k)$ 的方法

类似于 Cohen 在 2002 年噪声估计算法^[2]中所使用的方法, 利用语音存在概率计算和频率相关的平滑常数 $\alpha_s(\lambda, k)$ 。最后, 在计算了和频率相关的平滑常数 $\alpha_s(\lambda, k)$ 后, 利用式(3)来更新噪声功率谱:

$$D(\lambda, k) = \alpha_s(\lambda, k) D(\lambda - 1, k) + (1 - \alpha_s(\lambda, k)) |Y(\lambda, k)|^2 \quad (3)$$

其中, $D(\lambda, k)$ 是噪声功率谱的估计值。图 1 描绘了真实的噪声功率谱和快速噪声估计算法估计的噪声功率谱。

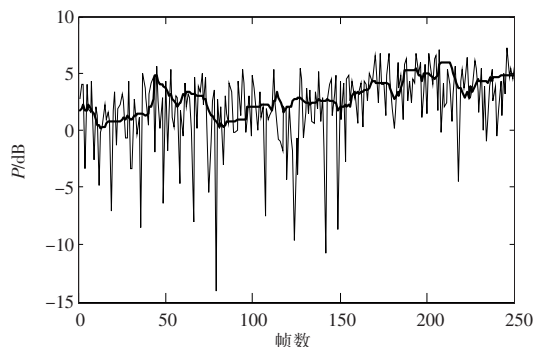


图 1 真实噪声功率谱和估计噪声功率谱

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60472094); 山西省高校科技开发项目(No.197-12909083)。

作者简介: 陈照平(1981-), 男, 硕士研究生, 主要研究方向: 语音信号处理; 马建芬(1967-), 女, 副教授, 硕士生导师, 主要研究方向: 语音信号处理、神经网络、小波处理。

图1中,细线表示真实噪声功率谱,粗线表示估计噪声功率谱,频率分量 $k=8$ 。可以看出在大约160帧处,噪声功率突升,估计噪声功率也即刻突升,说明算法能够很好的跟踪噪声的变化。

2 最小均方误差(MMSE)估计法^[3]

带噪声语音信号 $y(i)=s(i)+n(i)$, $s(i)$ 和 $n(i)$ 分别表示纯净语音和噪声,用 $Y_k=R_k \exp(j\theta_k)$, $N_k, S_k=A_k \exp(j\alpha_k)$ 分别表示带噪声信号、噪声和纯净语音的第 k 个频谱分量。 A_k 的估计式为: $\hat{A}_k=E(A_k|Y_0, Y_1, \dots, Y_N)$,进一步假设各个频谱分量之间彼此独立的。此时,由贝叶斯公式有

$$\hat{A}_k = E(A_k | Y_k) = \frac{\int_0^{2\pi} \int_0^{\infty} a_k P(Y_k | a_k, \alpha_k) p(a_k, \alpha_k) da_k d\alpha_k}{\int_0^{2\pi} \int_0^{\infty} P(Y_k | a_k, \alpha_k) p(a_k, \alpha_k) da_k d\alpha_k} \quad (4)$$

假设噪声谱是服从零均值高斯分布,则有

$$p(Y_k | a_k, \alpha_k) = \frac{1}{\pi \lambda_n(k)} \exp\left\{-\frac{1}{\lambda_n(k)} |Y_k - a_k e^{j\alpha_k}|^2\right\} \quad (5)$$

假设语音频谱服从零均值高斯分布,则其幅值和相位的联合分布为

$$p(a_k, \alpha_k) = \frac{a_k}{\pi \lambda_s(k)} \exp\left\{-\frac{a_k^2}{\lambda_s(k)}\right\} \quad (6)$$

这里 $\lambda_s(k)=E\{|S_k|^2\}$, $\lambda_n(k)=E\{|N_k|^2\}$ 为第 k 个频点下的语音和噪声的方差。将式(5)、(6)代入式(4)得

$$\hat{A}_k = \Gamma(1.5) \frac{\sqrt{v_k}}{\gamma_k} \exp\left(-\frac{v_k}{2}\right) \left[(1+v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right) \right] R_k \quad (7)$$

$\Gamma(\cdot)$ 表示伽玛函数, $\Gamma(1.5)=\sqrt{\pi}/2$, $I_0(\cdot)$ 和 $I_1(\cdot)$ 分别表示零阶和一阶贝叶斯函数, v_k 定义为:

$$v_k = \frac{\xi_k}{1+\xi_k} \gamma_k \quad (8)$$

其中 ξ_k 和 γ_k 为

$$\xi_k = \frac{\lambda_s(k)}{\lambda_n(k)} \quad (9)$$

$$\gamma_k = \frac{R_k^2}{\lambda_n(k)} \quad (10)$$

ξ_k 和 γ_k 分别代表先验与后验信噪比(SNR),若将 \hat{A}_k 看作 R_k 乘以一个增益,定义这个增益为:

$$G_{MMSE}(\xi_k, \gamma_k) = \frac{\hat{A}_k}{R_k} = \Gamma(1.5) \frac{\sqrt{v_k}}{\gamma_k} \exp\left(-\frac{v_k}{2}\right) \left[(1+v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right) \right] \quad (11)$$

则有:

$$\hat{A}_k = G_{MMSE} \cdot R_k \quad (12)$$

由此,可得到纯净语音频谱幅度的估值 \hat{A}_k ,对其添加带噪声信号的相位,并进行IFFT即可得到增强后的声音。

3 算法的仿真实现及性能评价

实验中采用了一个加入高斯白噪声和F16战斗机混合组成的非平稳噪声,信噪比是5dB,语音信号采样率是8kHz,采

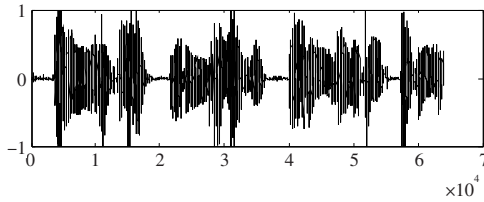


图2 纯净语音时域图

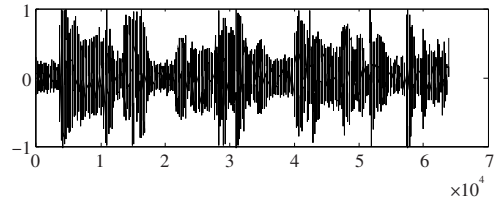


图3 谱减法增强时域图

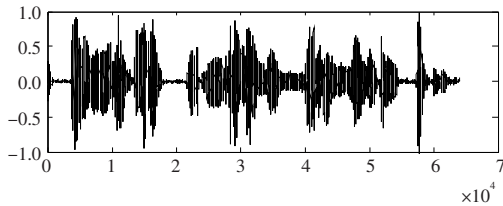


图4 MMSE-LSA 增强时域图

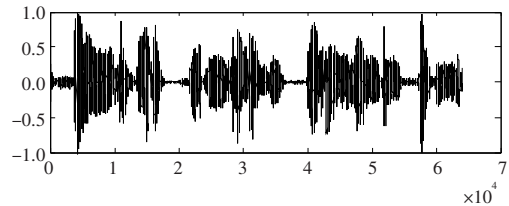


图5 基于快速噪声估计的MMSE 增强时域图

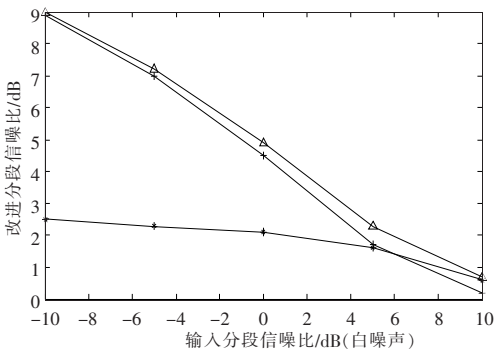
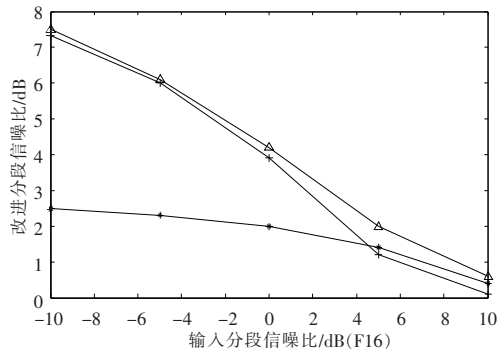


图6 几种语音增强算法改进的分段信噪比较



(下转 119 页)