第16卷第3期
2004年6月
重庆邮电学院学报
Journal of Chongqing University of Posts and Telecommunications
Vol. 16 No. 3
Jun. 2004

文章编号:1004-5694(2004)03-0065-04

基于短时对数谱的MMSE语音增强算法研究*

陈红梅,陈健
(西安电子科技大学 通信工程学院,陕西 西安 710071)

摘　要:研究了单话筒采集条件下基于语音短时对数谱的最小均方误差(MMSE-LSA)估计的语音增强算法,给出了其算法分析的基本流程图。由于语音是时变的,因此,假设语音频谱分布为高斯分布,在此基础上讨论了MMSE-LSA算法的先验信噪比ξ_k的2种估计方法——最大似然估计方法和直接判决估计方法。试验证明此方法的语音增强效果较好,尤其在较低信噪比时效果更明显。

关键词:语音增强;短时对数谱;最小均方误差
中图分类号:TN912.35　　**文献标识码**:A

0 引 言

语音增强技术近年来一直备受关注。由于噪声特性各异,语音增强的方法也各不相同。在语音处理的实际应用中,各类宽带噪声的污染较其它噪声更为普遍,也更难处理。因此,设法增强带有宽带干扰的语音信号具有重要意义,尤其是单话筒采集条件下如何消除背景噪声的影响更是许多人研究的课题。

语音增强的一个主要目标是从带噪语音信号中提取尽可能纯净的原始语音。单输入时消除宽带噪声的典型方法有谱减法、最小均方误差(MMSE: minimun mean square error)法、Wiener滤波器法、多带通滤波器法、自适应滤波器法等。这些方法都能使噪声得到减弱。我们对单输入情况下,短时对数谱的MMSE语音增强算法进行了详细的研究,在语言信号受平稳加性白色高斯噪声污染时,获得带噪语音时的语音增强是有效的。

1 短时对数谱的MMSE语音增强算法

对于人耳来说,语音信号的短时谱幅度(STSA: short time spectral amplitude)比相位更重要。由于人耳对频谱强度的感受与幅度的对数成正比,因此我们考察基于短时对数谱的最小均方误差(MMSE-LSA)估计。

由于语音信号的准平稳性,使得要对语音信号做数字处理必须先按短时段对语音信号分帧,这样每一帧信号都具有短时平稳性。带噪语音的短时谱可用快速傅立叶变换一帧一帧的计算得到,在相位提取后存储起来,然后对纯净语音的短时对数谱作最小均方误差估计。处理后的语音由估计得到的幅度谱和存储的相位重建。我们假设语音频谱分布为高斯分布,并在此假设下推导MMSE-LSA的估计公式,分析其特性。

设观察到的一帧带噪信号为$y(t)=x(t)+d(t)$,$0\leqslant t\leqslant T$,其中$x(t)$为纯净语音信号,$d(t)$为平稳加性高斯噪声。实际上,为避免分帧时的截断效应,应对$y(t)$加窗。为了书写方便,用$y(t)$代表加窗后的带噪信号。令$Y_k=R_k\exp[\mathrm{j}\theta_k]$,$X_k=A_k\exp[\mathrm{j}\alpha_k]$和$D_k$分别表示带噪语音$y(t)$、信号$x(t)$和噪声$d(t)$进行FFT变换后的第$k$个频谱分量。语音增强的任务就是利用已知的噪声功率谱信息$y(t)$,从中估计出$x(t)$。即由$\{Y_0,Y_1,\cdots\}$估计出X_k。由于我们仅对频谱幅度的对数感兴趣,而认为相位对语音质量影响不大,因而估值问题可以简化为估计A_k(设$\hat{A}_k$为A_k的估计值)。即求式(1)为最小

$$A=E\{(\log A_k-\log\hat{A}_k)^2\} \tag{1}$$

$\hat{A}_k$的估计式可写成

$$\hat{A}_k=\exp\{E[\ln A_k|y(t)]\},0\leqslant t\leqslant T \tag{2}$$

式(2)可写成

$$\hat{A}_k=\exp\{E[\ln A_k|Y_k]\} \tag{3}$$

基于高斯模型的假设,式(3)中$E[\ln A_k|Y_k]$即是由Y_k推导$\ln A_k$。设$Z_k=\ln A_k$,则有

$$\Phi_{Z_k|Y_k}(\mu)=E\{\exp(\mu Z_k)|Y_k\}=E\{A_k^{\mu}|Y_k\} \tag{4}$$

---

* 收稿日期:2003-05-19
作者简介:陈红梅,女,西安电子科技大学通信工程学院硕士研究生,主要研究方向为语音信号处理。

得到 $E\{\ln A_k|Y_k\}=\frac{\mathrm{d}}{\mathrm{d}\mu}\Phi_{Z_k|Y_k}(\mu)|_{\mu=0}$ (5)

利用式(4)计算 $\Phi_{Z_k|Y_k}(\mu)$，并由此得到 $E\{\ln A_k|Y_k\}$。由于

$$\Phi_{Z_k|Y_k}(\mu)=E\{A_k^{\mu}|Y_k\}$$

$$\frac{\int_0^{\infty}\int_0^{2\pi}a_k^{\mu}p(Y_k|a_k,\alpha_k)p(a_k,\alpha_k)\mathrm{d}a_k\mathrm{d}\alpha_k}{\int_0^{\infty}\int_0^{2\pi}p(Y_k|a_k,\alpha_k)p(a_k,\alpha_k)\mathrm{d}a_k\mathrm{d}\alpha_k} \tag{6}$$

此处基于高斯模型假设，$p(\cdot)$ 为概率密度函数，

$$(Y_k|a_k,\alpha_k)=\frac{1}{\pi\lambda_d(k)}$$

$$\exp\left\{-\frac{1}{\lambda_d(k)}|Y_k-a_k\exp(\mathrm{j}\alpha_k)|^2\right\} \tag{7}$$

$$p(a_k,\alpha_k)=\frac{a_k}{\pi\lambda_d(k)}\exp\left\{-\frac{a_k^2}{\lambda_x(k)}\right\} \tag{8}$$

$\lambda_d(k)\triangleq E\{|D_k|^2\}$；$\lambda_x(k)\triangleq E\{|X_k|^2\}$ 分别为语音和噪声的第 k 个频谱分量的方差，把式(7)和式(8)代入式(6)，应用零阶修正贝赛尔函数 $\mathrm{I}_0(\cdot)$，得到

$$\Phi_{Z_k|Y_k}(\mu)=\frac{\int_0^{\infty}a_k^{\mu+1}\exp(-a_k^2/\lambda_k)\mathrm{I}_0\left(2a_k\sqrt{v_k/\lambda_k}\right)\mathrm{d}a_k}{\int_0^{\infty}a_k\exp(-a_k^2/\lambda_k)\mathrm{I}_0\left(2a_k\sqrt{v_k/\lambda_k}\right)\mathrm{d}a_k} \tag{9}$$

λ_k 满足下式的关系：

$$\frac{1}{\lambda_k}=\frac{1}{\lambda_x(k)}+\frac{1}{\lambda_d(k)} \tag{10}$$

v_k 定义如下，

$$v_k\triangleq\frac{\xi_k}{1+\xi_k}\gamma_k;\xi_k\triangleq\frac{\lambda_x(k)}{\lambda_d(k)};\gamma_k\triangleq\frac{R_k^2}{\lambda_d(k)} \tag{11}$$

这里 ξ_k 和 γ_k 分别称之为先验和后验信噪比。计算式(9)中的积分可得

$$\Phi_{Z_k|Y_k}(\mu)=\lambda_k^{\mu/2}\Gamma(\mu/2+1)\mathrm{F}(-\mu/2;1;-v_k) \tag{12}$$

$\Gamma(\cdot)$ 是伽码函数，$\mathrm{F}(a;c;x)$ 为合流超几何函数。

$$\mathrm{F}(a;c;x)=\sum_{r=0}^{\infty}\frac{(a)_r}{(c)_r}\frac{x^r}{r!} \tag{13}$$

这里 $(a)_r\triangleq 1\cdot a\cdot(a+1)\cdots(a+r-1)$，$(a)_0\triangleq 1$。式(12)中出现的 $\mathrm{F}(-\mu/2;1;-v_k)$ 在 $|\mu|<2$ 时逐项微分，在 $\mu=0$ 可导出

$$\frac{\partial}{\partial\mu}\mathrm{F}(-\mu/2;1;-v_k)|_{\mu=0}=-\frac{1}{2}\sum_{r=1}^{\infty}\frac{(-v)^r}{r!}\frac{1}{r} \tag{14}$$

$\Gamma(\mu/2+1)$ 可由 $\ln\Gamma(\mu/2+1)$ 从式(15)得到，

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\Gamma\left(\frac{\mu}{2}+1\right)=\Gamma\left(\frac{\mu}{2}+1\right)\frac{\mathrm{d}}{\mathrm{d}\mu}\ln\Gamma\left(\frac{\mu}{2}+1\right) \tag{15}$$

$$\ln\Gamma(\mu/2+1)=-c\frac{\mu}{2}+\sum_{r=2}^{\infty}\frac{(-\mu)^r}{2^r r}\alpha_r,|\mu|<2 \tag{16}$$

这里 $\alpha_r\triangleq\sum_{n=1}^{\infty}\frac{1}{n^r}$，$c=0.57721566490$ 是欧拉常数。把式(16)逐项微分并利用式(15)有

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\Gamma(\mu/2+1)|_{\mu=0}=-c/2 \tag{17}$$

利用式(14)和式(17)，从式(12)可得

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\Phi_{Z_k|Y_k}(\mu)|_{\mu=0}=$$

$$\frac{1}{2}\ln\lambda_k-\frac{1}{2}\left(c+\sum_{r=1}^{\infty}\frac{(-v_k)^r}{r!}\frac{1}{r}\right)=$$

$$\frac{1}{2}\ln\lambda_k+\frac{1}{2}\left(\ln v_k+\int_{v_k}^{\infty}\frac{\mathrm{e}^{-t}}{t}\mathrm{d}t\right) \tag{18}$$

把式(18)代入式(5)，应用式(12)和式(3)，我们得到谱估计：

$$\hat{A}_k=\frac{\xi_k}{1+\xi_k}\exp\left\{\frac{1}{2}\int_{v_k}^{\infty}\frac{\mathrm{e}^{-t}}{t}\mathrm{d}t\right\}R_k \tag{19}$$

可以看出，给 R_k 乘一个非线性增益函数可得到 $\hat{A}_k$，而增益函数仅依赖于先验和后验信噪比。增益函数可写成：

$$G(\xi_k,\gamma_k)\triangleq\frac{\hat{A}_k}{R_k}=\frac{\xi_k}{1+\xi_k}\exp\left\{\frac{1}{2}\int_{v_k}^{\infty}\frac{\mathrm{e}^{-t}}{t}\mathrm{d}t\right\} \tag{20}$$

MMSE-LSA 语音增强算法包括 3 部分：谱分析/合成(通过窗函数的 FFT/IFFT 和叠迭)；噪声功率谱估计；增益函数的计算，其算法流程图见图 1。

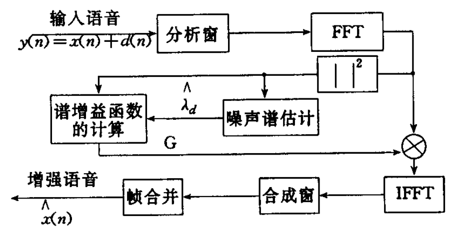


图1 MMSE-LSA语音增强算法流程图
Fig. 1 Flow diagram of MMSE-LSA speech enhancement numeration

2 先验信噪比的确定

MMSE-LSA 估计是在假定先验信噪比 ξ_k 和噪声方差 $\lambda_d(k)$ 已知的条件下得到的。然而，此处讨论的语音增强，所用的参数是提前未知的，仅有带噪语音可以利用。因而，在实际系统中，这些参数通常由估计得到。与噪声方差相比，先验信噪比是一个关键的参数。

由式(20)可以看出，先验信噪比 ξ_k 对最终的增益值的确定有很大影响。$\xi_k=\lambda_x(k)/\lambda_d(k)$，其中 $\lambda_d(k)$ 可以在无语音时通过对噪声的统计求平均获得，对于 $\lambda_x(k)$，由于语音是时变的，必须在每一帧重新进行估计。在此，给出 2 种估计方法：最大似然估

计方法和直接判决估计方法。

2.1 最大似然估计方法

假设有 L 个连续的观测量 $\{Y_k(m), Y_k(m-1), \cdots, Y_k(m-L+1)\}$，其中 $Y_k(j)$ 代表第 j 帧的第 k 个频率点的带噪频谱观测值。由于语音是缓变的，上述 L 帧语音的第 k 个频率点的方差也是缓变的，可以近似认为保持不变。另外，假设 $Y_k(i)$ 与 $Y_k(j)$ 是彼此独立的（这是为了简化计算考虑，事实上由于语音的缓变特性及处理时常常使用邻帧叠接窗，导致 $Y_k(i)$ 与 $Y_k(j)$ 存在一定的相关性）。

由于 $Y_k(m) = X_k(m) + D_k(m)$，$X_k(m)$ 与 $D_k(m)$ 均为高斯分布，所以 $Y_k(m)$ 也是高斯分布，方差为 $\lambda_x(k) + \lambda_d(k)$，有

$$p[Y_k(m), Y_k(m-1), \cdots | \lambda_x(k), \lambda_d(k)] = \prod_{l=0}^{L-1} \frac{\exp\left[-\frac{|Y_k(m-l)|^2}{\lambda_x(k)+\lambda_d(k)}\right]}{\pi[\lambda_x(k)+\lambda_d(k)]} \quad (21)$$

求最大似然估计，即求解：

$$\frac{\partial p[Y_k(m), Y_k(m-1), \cdots | \lambda_x(k), \lambda_d(k)]}{\partial \lambda_x(k)} = 0 \quad (22)$$

化简后可以得到 $\lambda_x(k)$ 的估值：

$$\hat{\lambda}_x(k) = \frac{1}{L}\sum_{l=0}^{L-1} R_k^2(m-l) - \lambda_d(k) \quad (23)$$

由于 $\lambda_x(k)$ 总是非负的，所以将上式修正后得到 ξ_k 的估计式为

$$\hat{\xi}_k = \max\left[\frac{1}{L}\sum_{l=0}^{L-1} \gamma_k(m-l) - 1, \varepsilon\right] \quad (24)$$

ε 为非负常数。

在实际使用时，式(24)的滑动平均用迭代平均来代替，即使用以下的估计式

$$\hat{\gamma}_k(m) = \max[\alpha\hat{\gamma}_k(m-1) + (1-\alpha)\gamma_k(m)/\beta, 1+\varepsilon], \varepsilon \geqslant 0$$

$$\xi_k(m) = \max[\hat{\gamma}_k(m) - 1, \varepsilon], \varepsilon \geqslant 0 \quad (25)$$

式(25)中增加了可调参数 α、β（$0 \leqslant \alpha \leqslant 1, \beta \geqslant 1$），它们的值由经验和主观试听决定。

2.2 直接判决估计方法

由最大似然估计式(24)，当 $L = 1$ 时，$\xi_k(m) = \max[\gamma_k(m) - 1, \varepsilon]$。另一方面，由 ξ_k 的定义及假设，有 $\xi_k(m) \approx \xi_k(m-1) = E\{A_k^2(m-1)\}/\lambda_d \approx \hat{A}_k^2(m-1)/\lambda_d$。其中，$(m-1)$ 为上一帧处理的结果。所以，构造 ξ_k 的估计式为

$$\xi_k(m) = \alpha\hat{A}_k^2(m-1)/\lambda_d + (1-\alpha)\max(\gamma_k(m) - 1, \varepsilon)/\beta \quad (26)$$

此方法利用了上一帧的处理结果。

3 试验结果分析

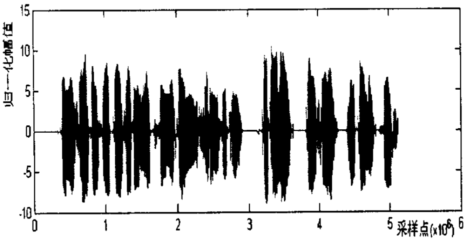
输入语音经 3.4 kHz 低通滤波/8 kHz 采样后，与平稳白色高斯噪声相加。性能评价是利用信噪比改善和主观试听方法。实验结果示于表 1 中。

表 1 3 种算法信噪比改善比较

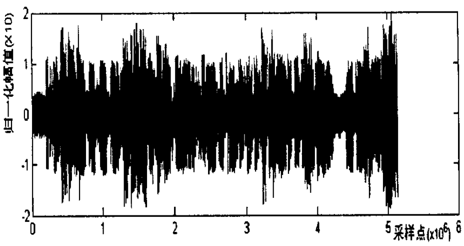
Tab. 1 Comparison of three numeration SNR

增强算法	输入信噪比				
	10	5	0	−5	−10
SS	11.67	8.03	2.71	−0.31	−6.15
MMSE	12.65	9.12	6.10	3.24	0.85
MMSE-LSA	12.50	9.85	6.35	4.10	1.10

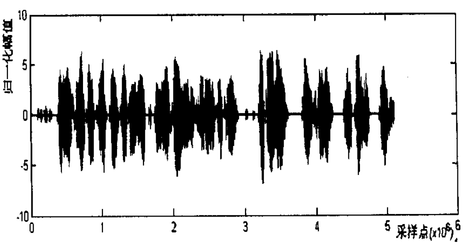
在输入信噪比较高时，MMSE-LSA 算法和最小均方误差（MMSE）算法、谱减法（SS：spectral subtract）的去噪效果相差不多，但当输入信噪比降低时，MMSE-LSA 算法去噪效果的优越性就表现出来了。与谱减法相比，MMSE-LSA 算法的可懂度和自然度较好。MMSE-LSA 估计效果的改善在很大程度上应归功于利用前一帧的信息。这种信息体现在先验信噪比的估计上。图2 给出了用MMSE-LSA 算法增强语音的时间波形，输入信噪比为 0 dB。



a 原始的纯净语音



b 信噪比为 0 dB 的带噪语音



c 采用 MMSE-LAS 算法的增强语音

图 2 语音增强波形

Fig. 2 Speech enhancement wave

在我们的具体解决方法中,由于纯净语音信号的谱实际上是未知的,在实际应用的过程中,我们并不是只利用前一帧的带噪信号(其特点是约束条件都是实际观测值,因而是准确的,不存在误差积累的问题),而是采用前一帧经过处理得到的增强语音作为代替,通过功率谱域的减谱法结合移动平均来实现。其算法稍微要复杂些。另外,还采用了有声/无声检测算法。

4 结论

本文主要研究了基于短时对数谱的语音增强(MMSE-LSA)算法,并将其降噪效果与谱减法进行了对比。在较低信噪比时,MMSE-LSA算法能有效地滤除带噪语音中的加性白色高斯噪声。此外还研究了MMSE-LSA算法中先验信噪比的两种估计。此方法可增强输入信噪比-10 dB至5 dB的带噪语音,且运算量不大,适应范围较广。

参考文献:

[1] WANG David L, LIM Jae S. The unimportance of phase in Speech enhancement[J]. IEEE Transactions on Acoustics,Speech and Signal Processing, 1982,Assp-30(4):679-681.

[2] AGAEAL Tarun, KABAL Peter. Pre-processing of noisy speech for voice codes [R]. Proc. IEEE Workshop on Speech Coding (Tsukaba, Japan)[C]. 169-171, Oct. 2002.

[3] COHEN Israel. Optimal speech enhancement under signal presence uncertainty using log-spectral Amplitude Estimator [J]. IEEE Signal Processing Letters. 2002,9(4):.

[4] COHEN Israel, Berdugo Baruch. Speech enhancement for non-stationary noise environments[J]. Signal Processing,2001,81 (11):2403-2418.

[5] EPRAM Y,MALAH D. Speech enhancement using a minumum mean-spuare error log-spectral amplitude estimate [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1985, ASSP-33(2): 443-445.

[6] EPRAM Y,MALAH D. Speech enhancement using a minimun mean-square error short-time spectral amplitude estimator[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing,1984,ASSP-32(6):.

(编辑:郭继笃)

Study of speech enhancement based on MMSE short time log spectral amplitude estimator

CHEN Hong-mei, CHEN Jian

(*School of Telecommunication Engineering, Xi'an University of Electronic science and Technology, Xi'an* 710071, *P. R. China*)

Abstract: In this paper, the authors study the speech enhancement algorithm based on minimun mean square error short time log spectral amplitude estimation(MMSE-LSA) under the single input condition,and present the basic flowchart of the algorithm analysis. Because the speech is non-stationary, and supposing the speech spectrum is the distribution of Gaussian,the authors discuss the estimation method of the parameter-priori SNR(signal-to-noise ratio)-maximum likelihood estimation approach and decision-directed estimation approach. Experiments show that the performance of the algorithm enhances the speech very well, especially in the conditions of low SNR.

Key words:speech enhancement; short-time log-spectral;minimun mean square error