基于拉普拉斯模型和掩蔽效应的语音增强

徐翠香,马建芬
XU Cui-xiang,MA Jian-fen

太原理工大学 计算机与软件学院,太原 030024
College of Computer and Software,Taiyuan University of Technology,Taiyuan 030024,China
E-mail:xucuixiang2006@163.com

XU Cui-xiang,MA Jian-fen.Speech enhancement based on Laplacian model and masking.Computer Engineering and Applications,2010,46(7):153-154.

Abstract: An effective approach for attenuating acoustic noise and mitigating speech distortion is proposed.First,MMSE method is analysed when the clean speech is modeled by a Laplacian distribution and the noise is modeled by a Gaussian distribution. Then,human perceptual auditory masking threshold is incorporated into this approach when the threshold of spectral amplitude of enhanced speech is computed.The experiment result evaluated by objective measure shows the proposed method can achieve a more significant noise reduction and reduce the chances of speech distortion.

Key words: speech enhancement;masking properties;Minimum Mean-Square Error(MMSE)

摘 要: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。首先实现了语音信号服从 Laplacian 分布、噪声服从 Gaussian 分布假设下的 MMSE 增强算法。为了进一步提高语音增强效果,在增强语音谱幅度阈值的计算上将该方法与人的掩蔽特性相结合。通过语音增强方法性能客观评测表明,该语音增强方法更好地抑制了噪声,有效地减小语音失真。

关键词:语音增强;听觉掩蔽阈值;最小均方误差(MMSE)
DOI:10.3778/j.issn.1002-8331.2010.07.046 **文章编号**:1002-8331(2010)07-0153-02 **文献标识码**:A **中图分类号**:TN912

1 引言

在基于统计模型的语音增强算法中,对语音的估计通常都是基于高斯模型的,即假设语音和噪声信号的幅度都服从高斯分布。然而这种假设只有在语音信号帧较长时才成立。高斯模型主要是基于中心极限理论,当信号帧较长时语音信号谱系数的概率密度函数近似服从高斯分布,这种近似高斯分布主要集中在中心部分,而在两端这种近似高斯分布并不能精确反映语音谱系数的概率分布特性[1]。Porter 和 Boll[2]指出在变换域语音信号更符合 Gamma 分布。Martin[3]提出了基于 Gamma 模型的 DFT 域的短时幅度谱 MMSE 估计算法。Martin 和 Breithaupt[4]提出了基于 Laplacian 的 MMSE 算法,指出 Laplacian 模型下的 MMSE 算法与 Gamma 模型下的 MMSE 算法有相似的特性,而且较容易实现。

为了更好地提高语音增强效果,许多学者将人的听觉特性应用于语音处理中,例如:Chang Huai You[5]将人耳掩蔽效应与 MMSE 算法相结合;Virag[6]考虑人耳掩蔽效应,用来获得谱相减的参数;Johnston[7]将掩蔽效应应用于语音编码中。该文在基于 Laplacian-Gaussian 模型的 MMSE 算法中进一步考虑了人耳的掩蔽特性。实验结果证明,增强语音的信噪比和感知质量得到了提高。

2 基于 Laplacian-Gaussian 模型的 MMSE 幅度谱估计

用 $s(t)$和 $n(t)$分别表示纯净语音和不相关的加性噪声信号,则观察到的带噪语音信号可表示为:

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

用 $Y_k=R_k\exp(j\theta_k)$,N_k,$S_k=A_k\exp(j\alpha_k)$分别表示带噪语音信号 $y(t)$、噪声 $n(t)$和纯净语音信号 $s(t)$经傅里叶变换后的第 k 个频谱分量。假设语音信号服从 Laplacian 分布、噪声服从 Gaussian 分布,则可得到纯净语音信号的 MMSE 估计式为[4]:

$$\hat{A}_k=G_{MMSE}(\xi_k,\gamma_k)R_k=\frac{2}{L_{k+}-L_{k-}}\times\frac{L_{k+}erfcx(L_{k+})-L_{k-}erfcx(L_{k-})}{erfcx(L_{k+})+erfcx(L_{k-})}R_k \tag{2}$$

L_{k+}和 L_{k-}的定义如下:

$$L_{k\pm}=\frac{1}{\sqrt{\xi_k}}\pm\sqrt{\gamma_k} \tag{3}$$

$erfcx(x)$是缩放补足误差函数,定义如下:

$$erfcx(x)=e^{x^2}\frac{2}{\sqrt{\pi}}\int_x^{\infty}e^{-t^2}dt \tag{4}$$

其中 ξ_k 和 γ_k 是先验信噪比和后验信噪比,其表达式分别为:

$$\xi_k=\frac{\lambda_s(k)}{\lambda_n(k)} \tag{5}$$

$$\gamma_k=\frac{R_k^2}{\lambda_n(k)} \tag{6}$$

作者简介:徐翠香(1982-),女,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,中国电子学会高级会员,主要研究方向:语音信号处理、自然语言处理。

收稿日期:2008-08-27 **修回日期**:2008-11-11

$\lambda_s(k)$,$\lambda_n(k)$分别表示纯净语音和噪声的第 k 个谱分量的方差。先验信噪比的估计采用判决引导(Decision-Directed)法[8],其估计式如下:

$$\hat{\xi}_k(l)=(1-\alpha)\max[\gamma_k(l)-1,0]+\alpha\frac{\hat{A}_k^2(l-1)}{\lambda_n(k,l-1)} \tag{7}$$

l 表示当前帧号,α 取经验值 0.98。

3 听觉掩蔽阈值的计算

掩蔽阈值[5]是通过模拟人类听觉系统的掩蔽特性和频率选择特性得到。掩蔽阈值的计算主要包括:纯净语音的初步估计(采用谱减法),临界带的分析,扩展函数,掩蔽阈值偏移量,掩蔽阈值的归一化处理和与绝对听觉掩蔽阈值的比较。

掩蔽阈值偏移量的计算有两种方法:一种是固定掩蔽阈值偏移量;另一种方法是基于 Bark 临界带的音调特性,噪声和纯音情况下的掩蔽特性是不同的,纯音的掩蔽阈值偏移量为(14.5+λ),λ 是 Bark 频带,噪声的掩蔽阈值偏移量为 5.5 dB,采用的是第二种方法。为了判断信号特性是纯音还是噪声,引入纯音系数 $\alpha(\lambda)$,纯音系数是根据 Bark 谱的平坦度来判断的,平坦度的定义如下:

$$SFM_{dB}(\lambda)=10\lg\frac{G_m(\lambda)}{A_m(\lambda)} \tag{8}$$

$G_m(\lambda)$、$A_m(\lambda)$分别是功率谱密度的几何平均值和算术平均值。$\alpha(\lambda)$的定义如下:

$$\alpha(\lambda)=\max\left\{\min[\frac{SFM_{dB}(\lambda)}{-60},1],0\right\} \tag{9}$$

则相对掩蔽阈值偏移量为:

$$O_\alpha(\lambda)=\alpha(\lambda)(14.5+\lambda)+[1-\alpha(\lambda)]5.5 \tag{10}$$

用掩蔽阈值偏移量就可计算出掩蔽阈值 $T(1,\lambda)$。

4 结合掩蔽效应和 Laplacian-Gaussian 模型的 MMSE 算法

由式(2)得纯净的估计式为:

$$\widehat{A_k}=G_{MMSE}(\xi_k,\gamma_k)R_k \tag{11}$$

$G_{MMSE}(\xi_k,\gamma_k)$为增益函数。为了利用人耳的感知特性平滑语音段到无语音段听觉感知上的突然变化,引入频谱板(spectral flooring)[5],可将增益函数写为如下形式:

$$G_k=\begin{cases}\dfrac{2}{L_{k+}-L_{k-}}\times\dfrac{L_{k+}erfcx(L_{k+})-L_{k-}erfcx(L_{k-})}{erfcx(L_{k+})+erfcx(L_{k-})} & \gamma_k>\rho_1(k)+\rho_2(k)\\ \left\{\rho_2(k)\dfrac{1}{\gamma_k}\right\}^{\frac{1}{2}} & \text{其他}\end{cases} \tag{12}$$

上式中 $\rho_1(k)$、$\rho_2(k)$是门限系数,它们决定了最低的信噪比门限,当低于此门限时,用一个特殊的最低增益函数来代替 MMSE 增益。通过大量仿真这两个参数值定义如下:

$$\rho_1(k)=\frac{T(l,k)-T_{\min}(l)}{T_{\max}(l)-T_{\min}(l)}(\rho_{1\max}-\rho_{1\min})+\rho_{1\min} \tag{13}$$

$$\rho_2(k)=\frac{T(l,k)-T_{\min}(l)}{T_{\max}(l)-T_{\min}(l)}(\rho_{2\max}-\rho_{2\min})+\rho_{2\min} \tag{14}$$

上式中 $\rho_{1\min}$,$\rho_{2\min}$,$\rho_{1\max}$,$\rho_{2\max}$ 分别表示参数 ρ_1,ρ_2 的最大值和最小值,$T_{\max}(l)$,$T_{\min}(l)$分别表示掩蔽阈值 $T(l,k)$在第 l 帧的最大值和最小值。

5 算法的具体步骤

(1)实现基于 Laplacian-Gaussian 模型的 MMSE 算法,详细描述见第 2 章。

(2)对纯净语音的幅度谱进行初估计(采用谱相减法),然后计算出掩蔽阈值,掩蔽阈值计算的具体描述见第 3 章。

(3)根据掩蔽阈值计算出门限系数 $\rho_1(k)$、$\rho_1(k)$,根据门限系数实现等式(12),计算出增益函数,具体描述见第 4 章。

(4)最后由增益函数计算纯净语音的估计式。

6 实验结果及性能评价

所用实验语音数据的采样率为 8 kHz,帧长为 256,重叠 1/2,数据窗是汉明窗,计算掩蔽阈值时的临界带数为 18。选用的噪声来自 Noisex-92 数据库,语音语句选用 HNIT 数据库。$\rho_{1\min}$,$\rho_{2\min}$,$\rho_{1\max}$,$\rho_{2\max}$ 取经验值,分别为:$\rho_{1\min}=1$,$\rho_{2\min}=0$,$\rho_{1\max}=6.28$,$\rho_{2\max}=0.015$。语音增强算法的性能评测采用客观测试结合非正式听音测试来进行。客观测试指标主要包括段信噪比(segmental SNR),PESQ 值。选用 HNIT 语音数据库的 30 条句子,加入三种噪声(白噪声、F16 战斗机噪声和 babble 噪声),最后对增强后的 30 条语句的测试结果求平均做为实验测试结果,其结果如表 1。从表中可以看出,加入白噪声,信噪比为-5 dB 的情况下,30 句话的平均 Seg-SNR 值提高 0.12 dB,Pesq 值略有提高,在其他噪声信号和信噪比条件下均有提高。通过实验表明提出的算法提高了语音增强效果。

表 1 分段信噪比和 Pesq 值的比较表

噪声/dB		基于 Laplac-Gauss 模型的 MMSE 算法		该文算法	
		Seg-SNR	Pesq	Seg-SNR	Pesq
白噪声	-5	4.696	2.648	4.822	2.664
	0	7.302	2.946	7.435	2.986
	5	9.576	3.197	9.709	3.273
	10	11.459	3.427	11.581	3.524
F16 噪声	-5	4.817	2.808	4.963	2.815
	0	7.539	3.093	7.736	3.132
	5	9.868	3.338	10.060	3.400
	10	11.727	3.567	11.900	3.636
Babble 噪声	-5	4.341	2.936	4.409	2.954
	0	7.539	3.241	7.653	3.264
	5	10.277	3.508	10.391	3.539
	10	12.293	3.708	12.414	3.751

7 结论

提出的基于 Laplacian-Gaussian 模型和掩蔽效应的 MMSE 算法与基于 Laplacian-Gaussian 模型的 MMSE 算法相比,在消除噪声的同时减小了语音失真和残留的“音乐噪声”,提高了增强语音的段信噪比和感知质量,而且此算法计算较简单,实时性较好。

参考文献:

[1] Cohen I.Speech enhancement using super-Gaussian speech models and noncausal a priori SNR estimation[J].Speech Communication, 2005,47:336-350.

[2] Porter J,Boll S.Optimal estimators for spectral restoration of noisy speech[C]//Proc IEEE Internat Conf Acoust Speech,Signal Process (ICASSP),San Diego,CA,1984:18A.2.1-18A.2.4.

[3] Martin R.Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors[C]//Proc 27th IEEE Internat Conf Acoust Speech Signal Process,ICASSP-02,Orlando, FL,2002:I-253-I-256.

(下转 161 页)