文章编号:1003-0530(2017)07-0918-09

采用复高斯分布模型的两步噪声幅度谱估计算法

欧世峰[1] 刘 伟[1] 宋 鹏[2] 赵晓晖[3]
(1. 烟台大学光电信息科学技术学院, 山东烟台 264005; 2. 烟台大学计算机与控制工程学院, 山东烟台 264005;
3. 吉林大学通信工程学院, 长春 130012)

摘 要: 噪声幅度谱估计是有效抑制外界噪声干扰、提高语音增强算法整体输出性能的重要环节。但目前针对该问题的研究相对较少, 常用的语音激活检测算法只能在语音不存在阶段对噪声信号的幅度谱进行更新或估计, 无法适用于更为复杂的非平稳噪声环境。为克服这一问题, 本文基于噪声频谱的复高斯分布模型假设, 提出了新型的两步噪声幅度谱估计算法。算法首先采用软判决技术计算噪声信号的功率谱, 然后再结合复高斯分布条件下信号幅度谱和功率谱之间的数学关系间接地获取噪声幅度谱的估计。文中基于这一结论给出了两种估计算法, 并在多种噪声环境下对它们的性能进行了仿真评估, 其测试结果有效表明了提出算法优良的估计性能。

关键词: 语音增强; 非平稳噪声; 幅度谱估计; 软判决算法; 复高斯分布模型
中图分类号: TN912.3 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2017.07.003

Two-Step Noise Amplitude Estimators Using Complex Gaussian Distribution Model

OU Shi-feng[1] LIU Wei[1] SONG Peng[2] ZHAO Xiao-hui[3]
(1. School of Science and Technology for Opto-electronic Information, Yantai University, Yantai, Shandong 264005, China;
2. School of Computer and Control Engineering, Yantai University, Yantai, Shandong 264005, China;
3. College of Communication Engineering, Jilin University, Changchun 130012, China)

Abstract: The estimate for the amplitude of noise signal plays an important role in many noise reduction or speech enhancement methods. However, compared with the noise power estimation, less attention has been paid to the amplitude estimation in the past years. In addition, the frequently-used voice activity detection (VAD) algorithm estimates or updates the noise amplitude during only speech absence area, which leads to an inferior performance in non-stationary noise conditions. To overcome such drawback, tow novel noise amplitude estimators working with two steps are proposed in this paper based on the assumption of complex Gaussian model. The estimation of noise power is achieved by soft decision (SD) method in the first step, and then the indirect estimators are subsequently obtained by using the relationship between the power and amplitude under the complex Gaussian distribution. The results of simulations indicated that the presented estimators can lead to significantly better speech quality than the frequently-used VAD method under various noise conditions.

Key words: speech enhancement; non-stationary noise; amplitude spectral estimation; soft decision approach; complex Gaussian distribution model

1 引言

现实工作和生活中,语音信号不可避免地会受到各类噪声信号的干扰。这些背景噪声的存在不仅破坏语音信号的声学模型,而且会对语音处理系统的整体性能造成极大影响。语音增强或噪声抑制技术的主要目的是滤除损坏信号可懂度和语音质量的背景噪声,尽可能地恢复出原有的纯净语

收稿日期: 2016-12-05; 修回日期: 2017-03-16
基金项目: 国家自然科学基金(61201457, 61005021); 山东省自然科学基金(ZR2014FQ016); 烟台大学研究生科技创新基金重点项目(YDZD1711)

音,从而提高语音信号处理系统的整体性能。过去三十年来,针对多种噪声环境下的语音增强问题,人们相继提出了许多经典、有效的算法,如:谱相减法、子空间算法、维纳滤波算法、最小均方误差(Minimum Mean Square Error, MMSE)算法及深度神经网络算法等[1-6]。这些算法多是基于声音信号的统计模型或某些特征信息来设计具有针对性的噪声抑制技术,在不同的应用环境和背景噪声下,其语音增强性能也有较大不同。但几乎所有的语音增强算法均具有同一特征,即它们的噪声消除效果都依赖于系统对于背景噪声谱信息(功率谱、幅度谱)估计的准确性。理论分析及实际的听力测试表明:过估计的噪声谱将会致使输出的增强语音产生较大畸变,降低系统对于语音有效成分的保护能力;而噪声谱的欠估计则会导致较多的噪声残留,严重影响语音增强算法的整体去噪效果[7]。

由于功率谱与自相关函数间的傅里叶变换关系以及计算先验信噪比参数的需要,现有的谱估计算法多是致力于获取噪声信号的功率谱估计,如最小统计(Minimum Statistics, MS)算法、SD 算法、最小控制递归平均(Minima Controlled Recursive Averaging, MCRA)算法等[8-10]。而近来的研究发现,相对于功率谱,噪声信号的幅度谱对于语音增强系统具有同样重要的应用价值。如在谱相减语音增强系统中,幅度谱减法相比于功率谱减法能够带来更为优良的人耳听觉体验,且残留“音乐噪声”更少,输出语音的整体质量更为出色[11]。目前,人们对于噪声幅度谱估计的研究相对较少,现多是采用 VAD 方法对带噪语音进行有声/无声判断,再在无语音段对噪声信号的幅度谱进行估计或更新[12]。该 VAD 算法在平稳噪声环境下具有较为稳健的估计性能,但由于其不能对噪声信号的幅度谱进行实时估计或更新,因此无法适应于现实生活中更为普遍的非平稳噪声环境。

为进一步提高噪声幅度谱估计结果的准确性,扩展算法的适用范围,本文基于噪声信号幅度谱和功率谱的关系提出了新型的噪声幅度谱估计算法。算法首先采用软判决技术对噪声信号的功率谱进行实时估计,然后再利用复高斯分布条件下幅度谱和功率谱之间的数学关系间接地获取噪声幅度谱的估计。由于软判决算法在有声及无声段均能对噪声功率谱进行实时估计和更新,且复高斯模型已被证明能够较为完美地拟合多种噪声信号的实际分布特性,因此本文提出算法相对于传统的 VAD 算法在多种应用环境下都具有更为优良的估计性能。文中利用仿真实验对算法的估计效果进行了验证,其结果证明了本文提出算法对于多种噪声环境的适用性和有效性。

2　基于 VAD 的噪声幅度谱估计

加性噪声信号模型下,带噪语音信号 $y(t)$ 可表示为:

$$y(t) = x(t) + d(t) \tag{1}$$

其中,$x(t)$ 与 $d(t)$ 分别为纯净语音与噪声信号。将带噪语音信号 $y(t)$ 分帧加窗后进行短时傅里叶变换,可得:

$$Y(m,k) = X(m,k) + D(m,k) \tag{2}$$

这里,$Y(m,k)$、$X(m,k)$ 与 $D(m,k)$ 分别表示带噪语音、纯净语音以及噪声的频谱,m 与 k 为帧数及频点索引。式(2)的极坐标形式可表示如下:

$$|Y(m,k)|\mathrm{e}^{\mathrm{j}\theta_Y(m,k)} = |X(m,k)|\mathrm{e}^{\mathrm{j}\theta_X(m,k)} + |D(m,k)|\mathrm{e}^{\mathrm{j}\theta_D(m,k)} \tag{3}$$

其中,$|Y(m,k)|$、$|X(m,k)|$ 与 $|D(m,k)|$ 分别为带噪语音、纯净语音以及噪声信号的频谱幅度;$\theta_Y(m,k)$、$\theta_X(m,k)$ 与 $\theta_D(m,k)$ 则分别表示其相应的相位谱。

语音增强或噪声抑制技术即是通过对带噪语音频谱 $Y(m,k)$ 进行处理,以恢复或估计出纯净语音时域信号 $x(t)$ 的过程。基于不同估计理论和信号模型,人们已设计出许多成熟有效的语音增强算法,但多数算法均需要噪声信号谱信息的先验知识,如图 1 中所描述的幅度谱减法即需要对噪声幅度谱进行预估计,才能有效恢复出原始的纯净语音信号。幅度谱减法的原理简单描述如下:假设纯净语音与噪声信号的相位相同,则纯净语音信号的频谱幅度 $|X(m,k)|$ 可通过下式进行估计

$$|\hat{X}(m,k)| = |Y(m,k)| - \kappa \cdot \gamma_D(m,k) \tag{4}$$

这里,$\gamma_D(m,k) = E(|D(m,k)|)$ 表示噪声信号幅度谱,κ 为过减率[12]。结合带噪语音相位谱后,将(4)式进行傅里叶反变换可得纯净语音信号的时域估计

$$\hat{x}(t) = \mathrm{IFFT}(|\hat{X}(m,k)|\mathrm{e}^{\mathrm{j}\theta_Y(m,k)}) \tag{5}$$

由于纯净语音信号的相位谱 $\theta_X(m,k)$ 是未知的,谱

减法中采用带噪语音的相位谱 $\theta_Y(m,k)$ 对其进行近似。

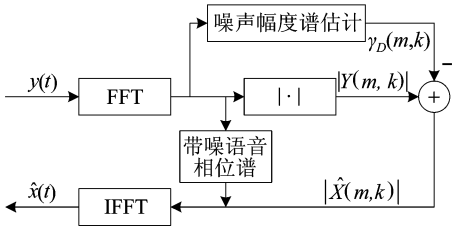


图 1 幅度谱减原理框图

Fig. 1 Diagram of amplitude spectral subtraction

影响幅度谱减法去噪性能的关键因素是系统对于噪声幅度谱 $\gamma_D(m,k)$ 估计的准确性。目前人们对于噪声幅度谱估计的研究相对较少,多是基于 VAD 技术在无声阶段对噪声信号的幅度谱来进行估计和更新。VAD 技术的实质是基于纯净语音与背景噪声信号的特征差别,检测出人们说话过程中的有声或无声片段。通常用于 VAD 的特征参数包括短时能量、短时过零率、峭度、高阶统计量及基音信息等[13-15]。根据实际应用环境或场所的不同,可以选择不同的特征参数或它们的组合来提高有声/无声判决的准确性。基于 VAD 技术来进行噪声幅度谱估计的过程可概括为以下两个部分:首先利用特征参数并结合判决阈值来对当前帧带噪语音信号进行有声或无声判决,即

$$I=\begin{cases}1 & \text{当 } \Pi_m \geqslant T \text{ 时}\\ 0 & \text{当 } \Pi_m < T \text{ 时}\end{cases} \tag{6}$$

式中 Π_m 表示算法对某一特征参数的计算结果,T 为判决阈值,I 表示最终判决结果。当 $I=1$ 时认为当前帧处于语音存在阶段,而当 $I=0$ 时则判决当前帧为无语音段。然后基于式(6)的判决结果对噪声信号幅度谱进行更新

$$\gamma_D(m,k)=\begin{cases}\gamma_D(m,k-1), & I=1 \text{ 时}\\ \alpha\gamma_D(m-1,k)+(1-\alpha)|Y(m,k)|, & I=0 \text{ 时}\end{cases} \tag{7}$$

其中,α 为滑动因子。

通过式(7)可以看出,基于 VAD 技术的噪声幅度谱估计实际上只有在无语音段对噪声信号幅度谱进行估计和更新,其在语音有声段则是直接采用无声阶段的估计结果。对于平稳或统计特性变化较为缓慢的背景噪声,VAD 算法可以获得较为理想的噪声幅度谱估计结果。但对于现实生活中更为常见的非平稳噪声,其估计性能将大打折扣,进而会对语音增强算法的整体去噪效果产生严重影响。

3 两步噪声幅度谱估计算法

上文给出的 VAD 噪声幅度谱估计算法由于只能在语音无声阶段更新噪声幅度谱,无法适用于非平稳的噪声环境。针对这一问题,本文通过推导复高斯分布模型条件下噪声信号幅度谱与功率谱的关系,在利用软判决算法获取噪声信号功率谱估计的基础上,提出了两步噪声幅度谱估计算法。由于提出算法在语音存在与不存在阶段都能够对噪声特性进行实时估计和更新,从而可以有效跟踪非平稳噪声的幅度谱变化,获取了更为优良的估计效果。本节将首先给出软判决算法的基本原理和设计过程,然后通过推导噪声信号幅度谱及其功率谱之间的数学关系,进而获得本文算法的计算步骤。

3.1 基于软判决的噪声功率谱估计

仍采用 $X(m,k)$、$D(m,k)$ 与 $Y(m,k)$ 表示纯净语音、噪声及带噪语音信号在第 k 个频点、第 m 帧上的频谱,用 R_0 和 R_1 分别表示当前帧中语音信号不存在及存在性假设[16],即

$$\begin{aligned}R_0&: Y(m,k)=D(m,k)\\ R_1&: Y(m,k)=X(m,k)+D(m,k)\end{aligned} \tag{8}$$

假设语音和噪声信号频谱互不相关,且服从均值为零、方差为 $\lambda_X(m,k)$ 和 $\lambda_D(m,k)$ 的复高斯分布,则在 R_0 和 R_1 条件下带噪语音频谱的条件分布分别为:(为简单起见,以下公式推导中省略索引 m 与 k。)

$$P(Y\mid R_0)=\frac{1}{\pi\lambda_D}\exp\left(-\frac{|Y|^2}{\lambda_D}\right) \tag{9}$$

$$P(Y\mid R_1)=\frac{1}{\pi(\lambda_X+\lambda_D)}\exp\left(-\frac{|Y|^2}{\lambda_X+\lambda_D}\right) \tag{10}$$

根据贝叶斯估计理论,并结合式(9)与式(10)可得带噪语音信号频谱条件下语音的不存在概率如下式所示

$$P(R_0\mid Y)=\frac{P(Y\mid R_0)P(R_0)}{P(Y\mid R_0)P(R_0)+P(Y\mid R_1)P(R_1)}=\frac{1}{1+(P(R_1)/P(R_0))\cdot\Psi(Y)} \tag{11}$$

这里,$P(R_0)$与$P(R_1)$分别表示当前帧中语音不存在及存在的先验概率。综合式(9)、式(10)与式(11)可得似然率$\Psi(Y)$的表达式如下

$$\Psi(Y)=\frac{1}{1+\xi}\exp\left(\frac{\eta\xi}{1+\xi}\right) \tag{12}$$

这里,$\xi=E(|X|^2)/\lambda_D$为先验信噪比参数,其可通过经典的直接判决算法估计,$\eta=|Y|^2/\lambda_D$则为后验信噪比。

设$E(|D|^2|Y)$为R_0和R_1两种假设下噪声信号的瞬时功率估计,则其可通过下式的软判决技术计算获得

$$\begin{aligned}E(|D|^2|Y)=&E(|D|^2|Y,R_0)P(R_0|Y)+\\&E(|D|^2|Y,R_1)P(R_1|Y)\end{aligned} \tag{13}$$

其中,带噪语音信号频谱条件下的语音存在概率$P(R_1|Y)=1-P(R_0|Y)$。明显地,当语音不存在时,带噪语音中只包含噪声信号,有$E(|D|^2|Y,R_0)=|Y|^2$;而当语音存在时,则通过 MMSE 估计可得[9]

$$E(|D|^2|Y,R_1)=\frac{\xi}{1+\xi}\gamma_D+\left(\frac{1}{1+\xi}\right)^2|Y|^2 \tag{14}$$

结合式(13)与式(14),可得软判决算法对于噪声方差,即功率谱的估计如下[9]

$$\begin{aligned}\hat{\lambda}_D(m+1,k)=&\beta\hat{\lambda}_D(m,k)+\\&(1-\beta)E\{|D(m,k)|^2|Y(m,k)\}\end{aligned} \tag{15}$$

其中,β表示平滑因子。结合以上推导过程可以看出,软判决算法同时考虑了有声与无声情况下噪声信号功率谱的估计结果,因此能够较为准确地跟踪背景噪声统计特性的变化情况,已成为目前较为常用的噪声功率谱估计算法之一。

3.2　本文提出算法

噪声频谱D的实部与虚部分别用D_{R}与D_{I}表示,即$D=D_{\mathrm{R}}+jD_{\mathrm{I}}$。由于时域中噪声信号的均值为0,根据中心极限定理理论,经过短时傅里叶变换后,其频谱的实部与虚部相互独立且服从均值为0的高斯模型分布,故可以采用以下两式分别表示D_{R}与D_{I}的概率密度函数为

$$P(D_{\mathrm{R}})=\frac{1}{\sqrt{2\pi\lambda_{D_{\mathrm{R}}}}}\exp\left(-\frac{D_{\mathrm{R}}^2}{2\lambda_{D_{\mathrm{R}}}}\right)=\frac{1}{\sqrt{\pi\lambda_D}}\exp\left(-\frac{D_{\mathrm{R}}^2}{\lambda_D}\right) \tag{16}$$

$$P(D_{\mathrm{I}})=\frac{1}{\sqrt{2\pi\lambda_{D_{\mathrm{I}}}}}\exp\left(-\frac{D_{\mathrm{I}}^2}{2\lambda_{D_{\mathrm{I}}}}\right)=\frac{1}{\sqrt{\pi\lambda_D}}\exp\left(-\frac{D_{\mathrm{I}}^2}{\lambda_D}\right) \tag{17}$$

其中,$\lambda_{D\mathrm{R}}=\lambda_D/2$与$\lambda_{D\mathrm{I}}=\lambda_D/2$分别表示$D_{\mathrm{R}}$与$D_{\mathrm{I}}$的方差。结合以上两式得$D_{\mathrm{R}}$与$D_{\mathrm{I}}$的联合概率密度函数为

$$P(D_{\mathrm{R}},D_{\mathrm{I}})=\frac{1}{\pi\lambda_D}\exp\left(-\frac{D_{\mathrm{R}}^2+D_{\mathrm{I}}^2}{\lambda_D}\right) \tag{18}$$

由$D=D_{\mathrm{R}}+jD_{\mathrm{I}}$,通过上式易知噪声信号频谱$D$服从均值为0的复高斯分布,其概率密度函数表示如下

$$P(D)=\frac{1}{\pi\lambda_D}\exp\left(-\frac{|D|^2}{\lambda_D}\right) \tag{19}$$

显然,当带噪语音中语音信号不存在时,式(19)与(9)等价。为推导方便,此处设

$$L_1=D_{\mathrm{R}}^2,L_2=D_{\mathrm{I}}^2,Z=L_1+L_2 \tag{20}$$

根据式(16)与式(17),可得[17]

$$P(L_1)=\frac{1}{\sqrt{L_1}}P(D_{\mathrm{R}})|_{D_{\mathrm{R}}=\sqrt{L_1}}=\frac{1}{\sqrt{L_1\pi\lambda_D}}\exp\left(-\frac{L_1}{\lambda_D}\right) \tag{21}$$

$$P(L_2)=\frac{1}{\sqrt{L_2}}P(D_{\mathrm{I}})|_{D_{\mathrm{I}}=\sqrt{L_2}}=\frac{1}{\sqrt{L_2\pi\lambda_D}}\exp\left(-\frac{L_2}{\lambda_D}\right) \tag{22}$$

由$Z=L_1+L_2$,得$P(Z)=P(L_1)*P(L_1)$,这里符号“*”表示卷积[17]。结合式(21)与式(22),可得

$$P(Z)=P(L_1)*P(L_2)=\frac{1}{\pi\lambda_D}\int_0^{\infty}\frac{1}{\sqrt{L_2(Z-L_2)}}\exp\left(-\frac{Z}{\lambda_D}\right)\mathrm{d}L_2=\frac{1}{\lambda_D}\exp\left(-\frac{Z}{\lambda_D}\right) \tag{23}$$

设$A=Z^{1/2}=|D|$,则根据式(23),可得噪声频谱幅度A的分布如下

$$P(A)=2A\cdot P(Z)|_{Z=A^2}=\frac{2A}{\lambda_D}\exp\left(-\frac{A^2}{\lambda_D}\right) \tag{24}$$

根据上式,经计算可得A的数学期望$E(A)$,即噪声信号的幅度谱为

$$\gamma_D=E(|D|)=E(A)=\int_0^{\infty}AP(A)\mathrm{d}A=$$

$$\frac{2}{\lambda_D}\int_0^{\infty} A^2 \exp\left(-\frac{A^2}{\lambda_D}\right) \mathrm{d}A = \frac{\sqrt{\pi\lambda_D}}{2} \tag{25}$$

加入帧数和频点索引 m 与 k 后，通过式(25)即可获得复高斯分布模型条件下，噪声幅度谱 $\gamma_D(m,k)$ 与其功率谱 $\lambda_D(m,k)$ 的数学关系如下

$$\gamma_D(m,k) = \frac{\sqrt{\pi}}{2}\sqrt{\lambda_D(m,k)} \tag{26}$$

通过式(26)可以看出，复高斯模型条件下噪声信号的功率谱与幅度谱存在着简单的对应关系，在无法直接获取噪声幅度谱的情况下，可先采用 3.1 节中的软判决算法计算噪声功率谱，然后再利用式(26)间接地获得噪声信号幅度谱的估计。综合式(13)、式(15)与式(26)，本文提出两种噪声幅度谱的估计算法，其步骤分别如下：

提出算法 1：

1)利用式(15)的软判决算法计算噪声功率谱 $\lambda_D(m,k)$；

2)通过式(26)间接获取噪声幅度谱的估计。

提出算法 2：

1)利用式(13)计算噪声信号的瞬时功率；

2)结合式(26)，通过平滑获得噪声幅度谱的估计如下

$$\hat{\gamma}_D(m+1,k) = \rho\hat{\gamma}_D(m,k) + (1-\rho)\frac{\sqrt{\pi}}{2}\sqrt{E\{|D(m,k)|^2|Y(m,k)\}} \tag{27}$$

这里，ρ 同式(15)中 β，表示平滑因子。

4 仿真实验与结果分析

为验证本文提出的两种噪声幅度谱估计算法的有效性能，本节中采用 40 段纯净语音信号(其中 20 段为男声，20 段为女声)作为测试数据，背景噪声为取自 Noisex-92 标准噪声库的 4 种不同类型噪声信号，分别为 White 噪声、Babble 噪声、Factory 噪声以及 Destroyerengine 噪声。所有声音信号的采样频率均为 8 kHz/s，基于语音信号的短时平稳特性，设计每一语音帧包含 256 个采样点，即 N 取值为 256，帧间重叠 50%。所有算法中的平滑因子均取值为 0.9，幅度谱减算法中过减率 κ 的设置采用[12]中所介绍算法。在输入信噪比分别为 0 dB、5 dB、10 dB 及 15 dB 四种条件下对 VAD 及本文提出算法的噪声估计性能进行了仿真测试，同时基于图 1 的幅度谱减系统对三种算法的语音增强效果进行了详细对比和验证。

图 2(a)给出的是在 Factory 噪声背景下、输入信噪比为 5 dB 时带噪语音信号的时域波形图。带噪语音信号持续时间为 5.6 s，其中 0.9 s 至 2.9 s(对应 56 帧至 161 帧)以及 3.6 s 至 4.8 s(对应 225 帧至 300 帧)为语音存在段。图 2(b)给出的则是三种算法在第 12 个频点(约为 188 Hz)处对于噪声幅度谱的估计及跟踪结果。为更好地评价算法的估计效果，图中同时给出了噪声信号的实际(真实)幅度谱。图 3 给出的是在 Babble 噪声背景下、输入信噪比为 10 dB 时带噪语音信号以及三种算法的噪声幅度谱估计结果。综合图 2 和图 3 可以看出：两种背景噪声均为典型的非平稳噪声信号，其幅度谱具有较为明显的时变特性；VAD 算法只能在语音不存在阶段对噪声幅度谱的估计进行更新，在有语音段则是直接采用无语音阶段的估计结果，其整体的估计结果与实际的噪声幅度谱存在较大差异；本文给出的两种算法可以有效地跟踪噪声信号的时变特性，其无论在有声阶段还是无声阶段都可以对噪声的幅度谱进行有效的估计和跟踪；相对于提出算法 2，本文提出算法 1 估计结果中的过估计成分较多，整体估计误差稍大。

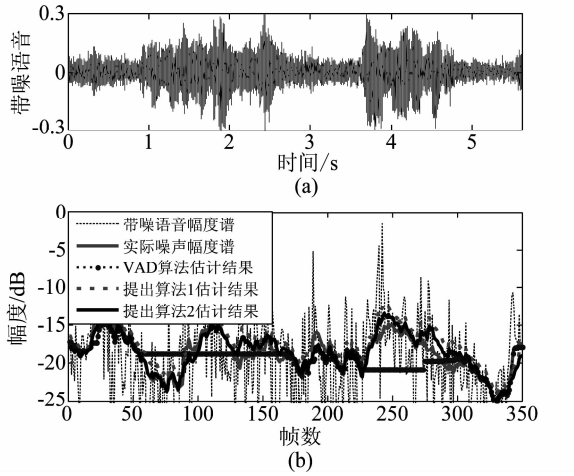


图 2 算法估计结果对比情况
(Factory 噪声背景，SNR=5 dB)

Fig. 2 Performance comparison between VAD and our proposed methods
(Factory noise condition, SNR=5 dB)

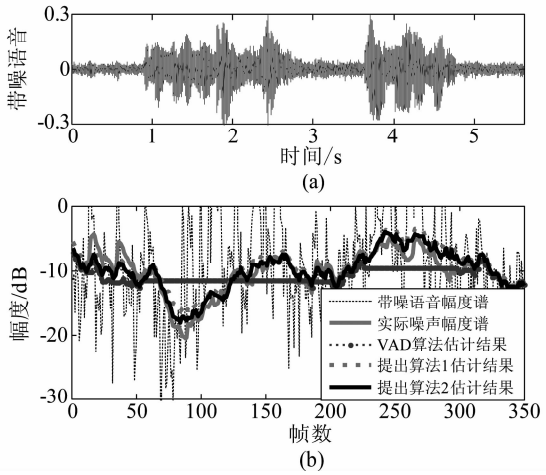


图 3　算法估计结果对比情况(Babble 噪声背景,SNR = 10 dB)

Fig. 3　Performance comparison between VAD and our proposed methods(Babble noise condition,SNR = 10 dB)

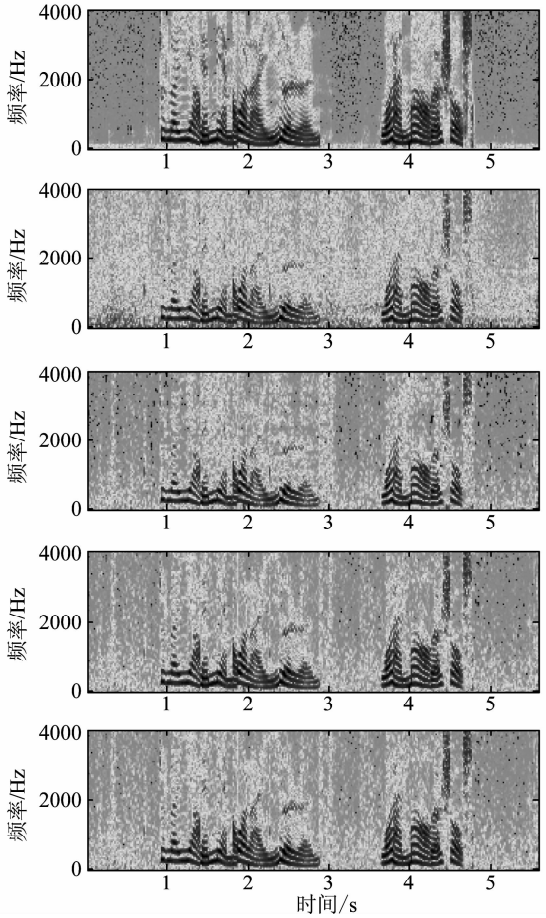


图 4　算法输出语音语谱图对比
(Factory 噪声背景,SNR = 10 dB)

Fig. 4　Spectrogram comparison VAD and our proposed methods(Factory noise condition,SNR = 10 dB)

噪声估计的目的是提高语音增强系统的整体性能,因此本文基于图 1 所示的幅度谱减系统对三种算法的语音增强效果进行了重点验证和分析。图 4 给出的是 Factory 噪声背景下、输入信噪比为 10 dB时五种语音信号的语谱图,自上至下依次为纯净语音、带噪语音、VAD 算法输出语音、本文算法 1 及算法 2 输出语音。图 5 给出的则是在 Destroyerengine 噪声背景下、输入信噪比为 5 dB 时四种语音信号的语谱图,自上至下为带噪语音、VAD 算法输出语音、本文算法 1 及算法 2 输出语音。从图 4 中不难看出,在无语音段,三种算法对于背景噪声的抑制能力基本相当,但在语音存在阶段,本文提出的两种算法则具有更好的语音信号保护和恢复能力,同时消除噪声的效果也更为出色。通过图 5 中则能更加明显地观察到,由于在语音存在段对于噪声幅度谱的估计不够准确,VAD 算法输出语音在此阶段存在着大量的噪声残留,而本文提出的两种算法则对这些语音存在段的背景噪声进行了有效消除。

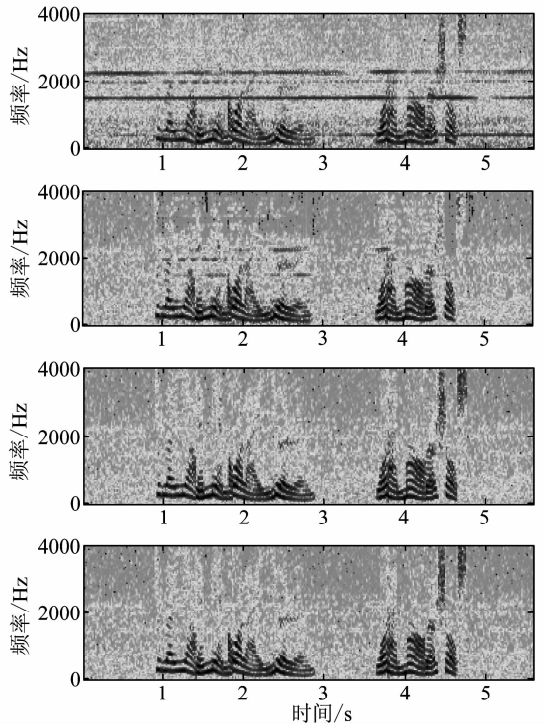


图 5　算法输出语音语谱图对比
(Destroyerengine 噪声背景,SNR = 5 dB)

Fig. 5　Spectrogram comparison VAD and our proposed methods(Destroyerengine noise condition,SNR = 5 dB)

为了定量地分析和对比三种噪声估计算法下的语音增强效果,采用分段信噪比(SegSNR),对数似然比(LLR)及语音质量感知评估(PESQ)等三种客观评价指标对于各个算法的性能进行了评估[18]。其中,SegSNR 可以有效地表征算法输出语音中背景

噪声的衰减情况,其数值越大表示算法抑制噪声的能力越强。该指标的计算公式如下

$$\mathrm{SegSNR} = \frac{10}{P}\sum_{p=1}^{P}\left(\log_{10}\left(\frac{\sum_{i=1}^{N}x^2(i+pN)}{\sum_{i=1}^{N}(x(i+pN)-\hat{x}(i+pN))^2}\right)\right) \quad (28)$$

这里,P 为总帧数,N 表示帧长。LLR 则是主要用于评测增强语音与原始语音间全极点模型的差异,其值越小表示算法对于语音的损伤程度越小。LLR 的定义式为

$$\mathrm{LLR} = \log\frac{\tilde{\alpha}_{\hat{x}}^{\mathrm{T}}\boldsymbol{R}_x\tilde{\alpha}_{\hat{x}}}{\alpha_x^{\mathrm{T}}\boldsymbol{R}_x\alpha_x} \quad (29)$$

其中,$\tilde{\alpha}$ 与 α 分别表示增强语音与原始纯净语音的 LPC 系数,$\boldsymbol{R}_x$ 为纯净语音信号的自相关矩阵。图 6 与图 7 分别给出了不同噪声背景及输入信噪比条件下三种算法输出语音的 SegSNR 与 LLR 对比情况,综合两图可以看出:在 White 噪声背景情况下,三种算法的输出 SegSNR 及 LLR 差别较小,整体的语音增强效果基本相当。其主要原因是 White 噪声为典型的平稳噪声信号,其统计特性随时间的变化情况较小,VAD 算法在此时则能够给出较为准确的估计结果;在其他三种非平稳噪声环境下,本文提出算法的输出 SegSNR 则明显高于 VAD 算法,而 LLR 又大幅低于 VAD 算法,从而说明本文提出算法的噪声抑制能力以及对于语音保护能力都要优于传统的 VAD 算法;由于提出算法 1 对于噪声幅度谱存在部分过估计的情况,因此提出算法 2 的整体语音增强性能要略优于提出算法 1。

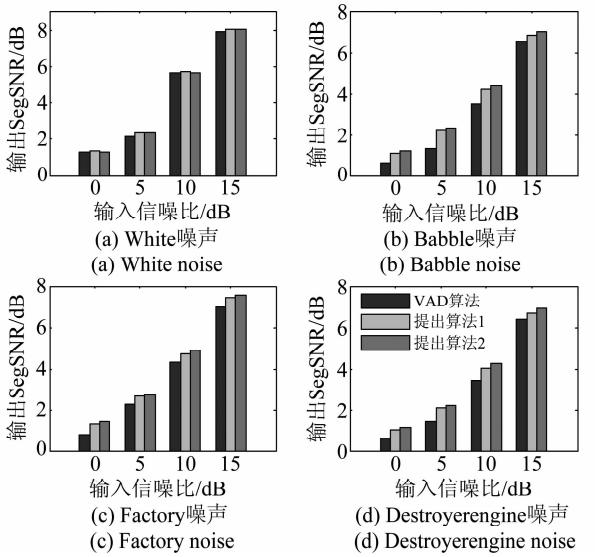


图 6 算法输出语音 SegSNR 对比情况

Fig. 6 SegSNR comparison between VAD and our proposed methods

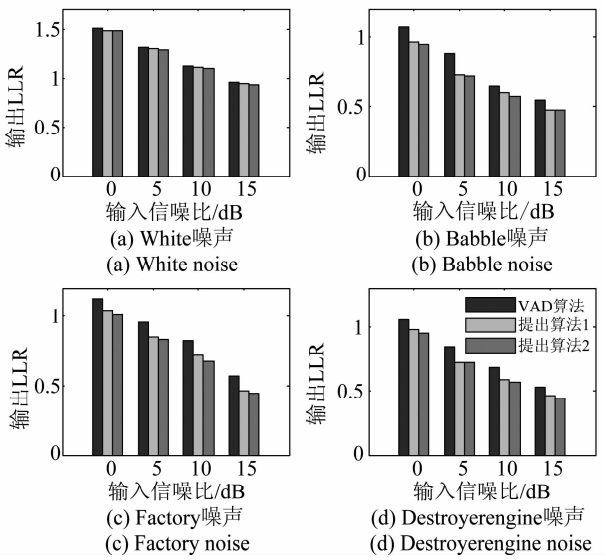


图 7 算法输出语音 LLR 对比情况

Fig. 7 LLR comparison between VAD and our proposed methods

图 8 给出的是不同噪声背景及输入信噪比下算法输出语音的 PESQ 得分情况。PESQ 计算结果与主观听觉测试具有高度的相关性,能够在多种环境下对测试语音的主观质量进行准确预测,已成为目前语音信号处理领域应用最为广泛的工业评测标准之一。PESQ 得分数值越高,表明被评测语音的质量越好,语音处理算法的性能也就越优越[19]。从图 8 可以明显看出,相对于 VAD 算法,本文提出算法的 PESQ 得分有了一定提升,特别是在三种非平稳噪声环境下,PESQ 得分提高较大,从而说明本文算法输出语音的整体质量更高,更加符合人们的主观听音感受。

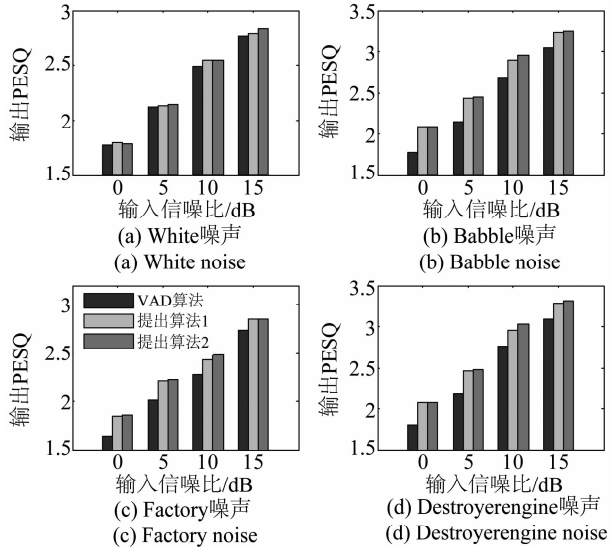


图 8 算法输出语音 PESQ 对比情况

Fig. 8 PESQ comparison between VAD and our proposed methods

为更加深入地分析本文提出算法的普适性与有效性,采用基于定量分析的语音增强技术来对三个噪声幅度谱估计算法的性能进行了对比验证和分析。由于定量分析语音增强算法多用于语音识别模块的前端以提高系统的识别率[20],故在此部分的仿真实验中,我们利用短时客观可懂度(Short-Time Objective Intelligibility, STOI)测度来对各个算法的综合性能进行评判[21]。表 1 给出的是不同输入信噪比条件下三个算法输出语音的平均 STOI 得分(对不同噪声背景下的得分结果进行了平均),其中加粗数值表示同等条件下结果最优。从表中可以看出,在不同信噪比环境下,本文提出算法的 STOI 得分相对于 VAD 算法均有不同程度的提高。对于可懂度测试,人们更为注重较低信噪比条件下的评测结果[6]。通过表 1 中数据不难看出,在输入信噪比为 0 dB 与 5 dB 时,本文算法输出语音的 STOI 得分提升更为显著,从而更加清晰地表明了提出算法相对于 VAD 算法的优越性。

表 1　算法输出语音 STOI 得分对比情况

Tab. 1　STOI comparison between VAD and our proposed methods

输入 SNR	STOI 得分/%		
	VAD 算法	提出算法 1	提出算法 2
0 dB	70.8	73.6	**74.0**
5 dB	80.6	82.2	**82.4**
10 dB	91.2	91.7	**91.8**
15 dB	94.8	**95.1**	**95.1**
平均值	84.4	85.7	**85.9**

5　结论

针对非平稳噪声环境下的噪声幅度谱估计问题,本文通过分析复高斯分布模型条件下信号幅度谱与其功率谱的数学关系,给出了两种新型的噪声幅度谱估计算法。提出算法采用两步的形式,先经过软判决技术完成对噪声信号功率谱的估计,再通过推导其与幅度谱的数学关系间接地获取噪声信号幅度谱的计算结果。由于复高斯分布模型已被验证能够完美地拟合多种噪声信号的分布特性,且软判决算法在有声及无声阶段均能对噪声的统计特性进行有效跟踪,因此本文提出算法在多种噪声背景下都呈现出较为出色的估计性能。

参考文献

[1] Li J, Sakamoto S, Hongo S. Adaptive β-order generalized spectral subtraction for speech enhancement[J]. Signal Processing, 2008, 88(11): 2764-2776.

[2] Borowicz A, Petrovsky A. Signal subspace approach for psychoacoustically motivated speech enhancement[J]. Speech Communication, 2011, 53(2): 210-219.

[3] Chen Jingdong, Benesty J, Huang A, et al. New insights into the noise reduction Wiener filter[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(4): 1218-1234.

[4] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator[J]. IEEE Transactions on Acoust. Speech Signal Processing, 1984, 32(6): 1109-1121.

[5] Xu Yong, Du Jun, Dai Lirong, et al. An experimental study on speech enhancement based on deep neural networks[J]. IEEE Signal Processing Letters, 2014, 21(1): 65-68.

[6] Xu Yong, Du Jun, Dai Lirong, et al. A regression approach to speech enhancement based on deep neural networks[J]. IEEE Transactions on Audio, Speech and Language Processing, 2015, 23(1): 7-19.

[7] Djendi M, Scalart P. Reducing over-and under-estimation of the a priori SNR in speech enhancement techniques[J]. Digital Signal Processing, 2014,32: 124-136.

[8] Marti R. Noise power spectral density estimation based on optimal smoothing and minimum statistics[J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(5): 504-512.

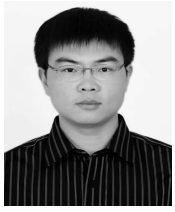
[9] Park Y S, Chang J H. A probabilistic combination method of minimum statistics and soft decision for robust noise power estimation in speech enhancement[J]. IEEE Signal Processing Letters, 2008, 15(1): 95-98.

[10] Cohen I. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging[J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(5): 466-475.

[11] Inoue T, Saruwatari H, Takahashi Y. Theoretical analysis of musical noise in generalized spectral subtraction based on higher order statistics[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(6): 1770-1779.

[12] Loizou P. Speech enhancement: theory and practice[M]. CRC Press, Boca Raton, FL, 2007.
[13] 廖逢钗, 李鹏, 徐波. 音乐噪声环境下的双声道语音活动检测[J]. 信号处理, 2009, 25(11): 1820-1824. Liao Fengchai, Li Peng, Xu Bo. Dual-channel voice activity detection in music noise envoronments[J]. Signal Processing, 2009, 25(11): 1820-1824. (in Chinese)
[14] Zhang Xiaolei, Wang Deliang. Boosting contextual information for deep neural network based voice activity detection[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2016, 24(2): 252-264.
[15] Zhang Xiaolei, Wu Ji. Deep belief networks based voice activity detection[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(4): 697-710.
[16] Abramson A, Cohen I. Simultaneous detection and estimation approach for speech enhancement [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(8): 2348-2359.
[17] Papoulis A, Pillai U. Probability, random variables and stochastic processes[M]. McGraw-Hill. 2011.
[18] Quackenbush S R, Barnwell T P, Clements M A. Objective measures of speech quality[M]. Prentice Hall, 1988.
[19] Hu Yi, Loizou P. Evaluation of objective quality measures for speech enhancement[J]. IEEE Transactions on Speech and Audio Processing, 2008, 16(1): 229-238.
[20] Zhu Qifeng, Alwan A. The effect of additive noise on speech amplitude spectra: A quantitative analysis[J]. IEEE Signal Processing Letters, 2002, 9(9): 275-277.
[21] Taal C H, Hendriks R C, Heusdens R, et al. An algorithm for intelligibility prediction of time-frequency weighted noisy speech[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(7): 2125-2136.

作者简介



欧世峰 男,1979 年生,山东巨野人。2008 年毕业于吉林大学,获工学博士学位。现为烟台大学光电信息科学技术学院副教授,主要研究方向为语音信号处理与盲信号处理。
E-mail: ousfeng@126.com

刘 伟 男,1992 年生,山东潍坊人。烟台大学光电信息科学技术学院硕士研究生,主要研究方向为语音信号处理。
E-mail: 1139983526@qq.com

宋 鹏(通讯作者) 男, 1983 年生,山东莱阳人。2014 年毕业于东南大学,获工学博士学位。现为烟台大学计算机与控制工程学院讲师,主要研究方向为语音信号处理、模式识别。
E-mail: pengsongseu@gmail.com

赵晓晖 男, 1957 年生, 北京人。1993 年毕业于法国贡比涅科技大学,获工学博士学位。现为吉林大学通信工程学院教授、博士生导师,国内外发表学术论文 100 余篇。主要研究方向为自适应信号处理理论及其在通信中的应用。
E-mail: xhzhao@jlu.edu.cn