一种改进的基于对数谱估计的语音增强算法

孔德廷
(中国西南电子技术研究所，四川成都 610036)

摘要：提出了一种基于对数谱估计的改进型语音增强算法。相对于传统语音增强算法，在语音信号存在不确定的条件下，利用软判决增益因子修正技术调正带噪语音信号的对数谱幅度，抑制背景噪声。引入的改进型先验信噪比估计和语音信号先验不存在概率估计方法，能够有效地估计得出语音信号的存在概率，进而求得语音信号存在时的谱增益因子函数，联合语音信号不存在时设定的增益因子函数加权求得谱增益函数。计算机仿真表明，即使在低信噪比条件下，输入背景噪声为高斯白噪声和粉红噪声等加性白噪声时，所提算法对噪声的抑制效果非常明显，且有效地克服了传统算法中引入的“音乐噪声”和语音信号畸变。

关键词：短时傅里叶变换；语音增强；幅度谱估计；先验信噪比；先验语音不存在概率
中图分类号：TN912.35 **文献标识码**：A **文章编号**：1000-3630(2020)-02-0208-06
DOI 编码：10.16300/j.cnki.1000-3630.2020.02.014

An improved speech enhancement algorithm based on log-spectral amplitude estimation

KONG Deting
(*Southwest China Institute of Electronic Technology*, *Chengdu* 610036, *Sichuan*, *China*)

Abstract: An improved speech enhancement algorithm is proposed for suppressing the background group noise in noisy speech. Compared with the traditional noise suppressor, the proposed method uses the soft-decision modification method to modify the log-spectral amplitude of noisy speech under the condition of uncertainty in speech signal presence. In this paper, an improved estimator for the priori signal-to-noise ratio and an effective estimator for the priori probability of speech absence are proposed, by which the conditional probability of speech presence can be obtained. Simulation results confirm the superiority of this method in noise suppression and speech enhancement under the background of Gaussian white noise and pink noise.

Key words: short-time Fourier transform; speech enhancement; acoustic magnitude estimation; priori signal-to-noise ratio; priori speech absence probability

0 引 言

语音增强算法作为语音信号处理中的一个重要分支领域，多年来一直吸引着广大学者对其进行不断的研究。语音增强算法旨在降低带噪语音信号中的背景噪声，提取纯净的语音信号，同时尽量减少信号畸变。语音增强算法不但涉及信号检测、波形估计等传统信号处理理论，而且与语音产生的生理特性、人耳的感知特性等生理学密切相关。

目前，国内外的研究成果一般分为三大类[1-2]：(1) 基于背景噪声估计的谱减法，提高输入带噪语音信号的信噪比；(2) 寻找稳健的、耐噪声的语音特征参数；(3) 基于模型参数适应化的噪声补偿算法。目前的补偿算法通常只考虑到噪声环境是平稳的，在低信噪比以及非平稳噪声环境中的效果并不理想。

语音增强算法一般基于短时傅里叶分析/修正/综 合 (Short-Time Fourier Analysis- Modification-Synthesis, STFT AMS)，例如：1979 年 Boll 提出的谱减法[3]，1949 年 Wiener 提出的 Wiener 滤波法，以及 1984 年 Ephraim 和 Malah 提出的 MMSE 短时幅度谱估计(Acoustic Magnitude Estimator, AME)算法[4]。Boll 提出的谱减法是直接从带噪语音信号频谱中减去噪声估计谱，尽管该方法能够有效地降低噪声，但在原始语音中会引入“音乐噪声”。为了克服上述问题，Ephraim 和 Malah 提出的 MMSE 短时幅度谱估计算法可有效避免“音乐噪声”的引入。Cappe[5]、Scalart 等[6]先后在 1994 年和 1996 年证明了利用先验信噪比的软判决方法可以得到较好的降噪性能。

收稿日期：2019-03-02；**修回日期**：2019-04-04
作者简介：孔德廷(1986－)，男，湖北荆门人，硕士，工程师，研究方向为基带信号处理，语音信号处理，智能优化算法等。
通讯作者：孔德廷, E-mail: dtkong@126.com

近年来，在语音信号增强算法中，利用软判决增益因子修改技术成为了一个热门的研究方向[7-10]。相对于使用传统的语音增强算法，在语音信号存在不确定(Speech Signal Presence Uncertainty)的假设条件下，联合改进的背景噪声估计(Optimized Noise Estimator)算法和先验语音不存在概率(Priori Speech Absence Probability, PSAP)，给出了比较好的降噪性能。本文提出了一种改进的对数谱幅度估计(Optimized Log-Spectral Amplitude Estimator, Optimized-LSA)方法，引入了一种新的先验信噪比估计(Priori Signal-to-Noise Ratio, Priori-SNR)方法以及一种有效的先验语音不存在概率估计方法。本文首先给出了语音信号存在或者缺失时增益因子的估计方法，然后结合先验语音不存在概率的估计值加权求得最终的谱增益因子。先验语音不存在概率基于软判决的方式获得，其在连续语音信号的相邻频点上呈现出很强的相关性。与传统的降噪算法相比，本文算法具有更好的降噪性能，即使在低信噪比条件下，仍然能够有效地避免“音乐噪声”的引入和语音信号的畸变。

1　增益估计模型

假设离散带噪语音输入信号 $y(n)$ 由语音信号 $x(n)$ 和不相关的加性高斯白噪声 $d(n)$ 叠加而成，对输入信号进行重叠、加窗后进行短时傅里叶变化(Short Time Fourier Transform，STFT)，则在时频域的表达式为

$$Y(k,l)=X(k,l)+D(k,l) \tag{1}$$

式中：k 表示频率索引，l 表示帧序号，$X(k,l)$ 为语音信号的短时傅里叶变换，$D(k,l)$ 为加性噪声的短时傅里叶变换。

假定 $H_0(k,l)$ 表示在第 l 帧的第 k 个频点上语音信号不存在，即仅有噪声存在；$H_1(k,l)$ 表示在第 l 帧的第 k 个频点上语音信号存在，即为带噪语音。则输入带噪语音信号的条件概率密度分布函数可表示为

$$p(Y(k,l)|H_0(k,l))=\frac{1}{\pi\lambda_d(k,l)}\exp\left\{-\frac{|Y(k,l)|^2}{\lambda_d(k,l)}\right\} \tag{2}$$

$$p(Y(k,l)|H_1(k,l))=\frac{1}{\pi(\lambda_x(k,l)+\lambda_d(k,l))}\cdot\exp\left\{-\frac{|Y(k,l)|^2}{\lambda_x(k,l)+\lambda_d(k,l)}\right\} \tag{3}$$

式中：$\lambda_x(k,l)\triangleq \mathrm{E}\{|X(k,l)|^2|H_1(k,l)\}$，为语音信号的功率谱估计；$\lambda_d(k,l)\triangleq \mathrm{E}\{|D(k,l)|^2\}$，为噪声信号的功率谱估计。应用贝叶斯原理，语音信号存在的条件概率 $p(k,l)\triangleq P(H_1(k,l)|Y(k,l))$ 可以表示为

$$p(k,l)=\left\{1+\frac{q(k,l)}{1-q(k,l)}[1+\xi(k,l)]\exp[-\upsilon(k,l)]\right\}^{-1} \tag{4}$$

式中：$q(k,l)\triangleq P[H_0(k,l)]$，定义为语音信号不存在的先验概率(PSAP)；$\xi(k,l)\triangleq\lambda_x(k,l)/\lambda_d(k,l)$，定义为输入信号的先验信噪比(Priori Signal-to-Noise Ratio)；$\gamma(k,l)\triangleq|Y(k,l)|^2/\lambda_d(k,l)$，为输入信号的后验信噪比(Posteriori Signal-to-Noise Ratio)，$\upsilon(k,l)\triangleq\gamma(k,l)*\xi(k,l)/[1+\xi(k,l)]$。

定义 $\hat{A}(k,l)$ 为输入信号的语音谱估计，则基于对数谱幅度估计算法为

$$\hat{A}(k,l)=\exp\{\mathrm{E}[\ln A(k,l)|Y(k,l)]\}\triangleq G(k,l)\cdot|Y(k,l)| \tag{5}$$

式中：$G(k,l)$ 表示在第 l 帧、第 k 个频点上的修正因子。依据统计分析模型：

$$\begin{aligned}\mathrm{E}[\ln A(k,l)|Y(k,l)]=&\\ \mathrm{E}[\ln A(k,l)|Y(k,l),H_1(k,l)]\cdot p(k,l)+&\\ \mathrm{E}[\ln A(k,l)|Y(k,l),H_0(k,l)]\cdot[1-p(k,l)]&\end{aligned} \tag{6}$$

当语音不存在时，修正因子约束为大于一个指定阈值，该阈值可以依据对背景噪声属性的主观经验确定，基本准则是背景噪声越大，该值越小，反之越大。则在语音缺失的条件下，输入信号的语音谱估计可表示为

$$\exp\{\mathrm{E}[\ln A(k,l)|Y(k,l),H_0(k,l)]\}=G_{\min}\cdot Y(k,l) \tag{7}$$

当语音存在时，定义谱增益修正因子为 $G_{H1}(k,l)$，则输入信号的语音谱估计可以表示为

$$\exp\{\mathrm{E}[\ln A(k,l)|Y(k,l),H_1(k,l)]\}=G_{H1}\cdot Y(k,l) \tag{8}$$

代入条件概率密度分布函数，可以得：

$$G_{H1}(k,l)=\frac{\xi(k,l)}{1+\xi(k,l)}\exp\left(\frac{1}{2}\int_{\upsilon(k,l)}^{\infty}\frac{\mathrm{e}^{-t}}{t}\mathrm{d}t\right) \tag{9}$$

代入式(5)，可以得到改进型对数谱估计的谱增益修正因子可以表示为

$$G(k,l)=\{G_{H1}(k,l)\}^{p(k,l)}\times G_{\min}^{1-p(k,l)} \tag{10}$$

基于改进型对数谱估计的语音增强算法原理框图如图 1 所示。

2　语音缺失先验概率估计方法

Ephraim 等[11]提出的判决导引算法，提供一种有效的估计先验信噪比的方法，其公式表示为

$$\begin{aligned}\hat{\xi}(k,l)=&\alpha\times G_{H1}^2(k,l-1)\times\gamma(k,l-1)+\\ &(1-\alpha)\max\{\gamma(k,l)-1,0\}\end{aligned} \tag{11}$$

式中：α 为权重因子，在噪声衰减程度和语音信号畸变之间进行权衡。

在本文提出的改进型对数谱估计的增益修正

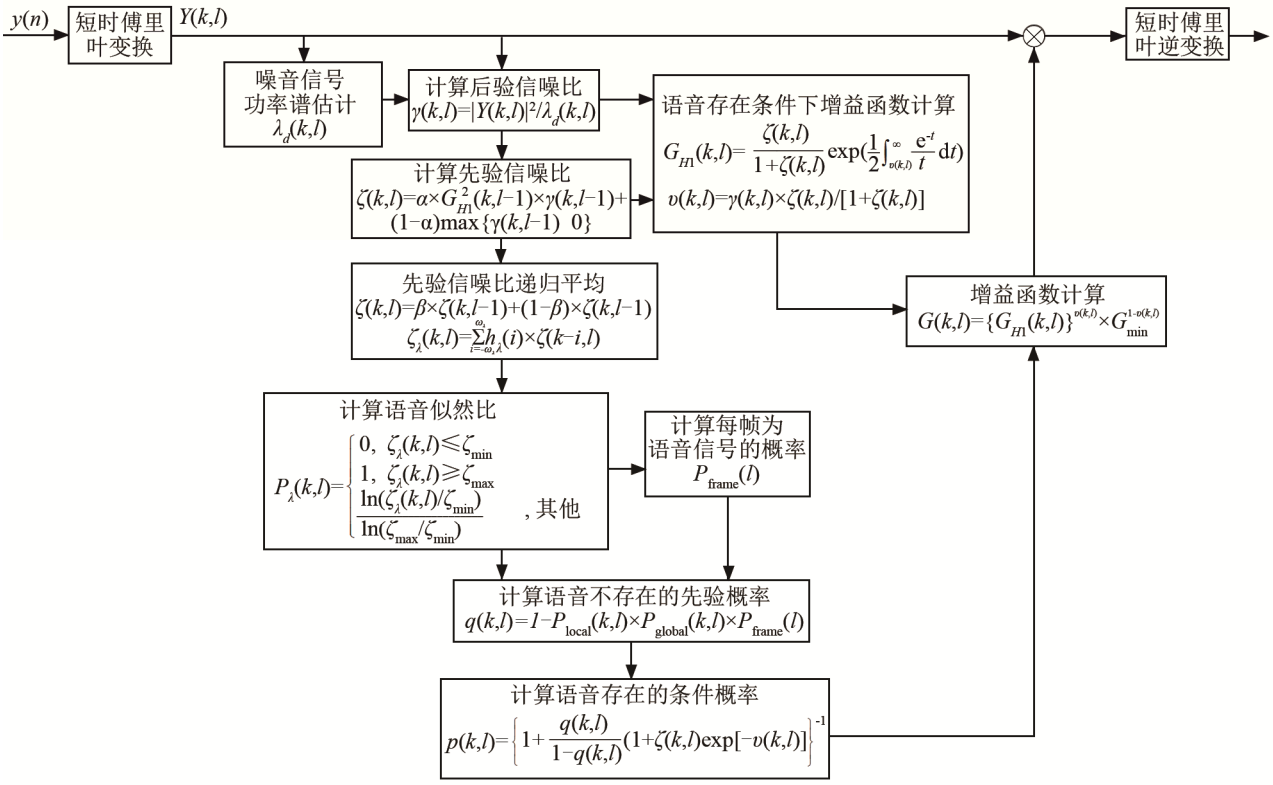


图 1 语音增强算法原理框图

Fig.1 Schematic diagram of the proposed speech enhancement algorithm

因子计算公式中，需要用到语音信号存在的条件概率 $p(k,l)$，而根据式(4)，计算 $p(k,l)$ 又需要知道语音不存在的先验概率 $q(k,l)\triangleq P(H_0(k,l))$。本文提出了一种有效的估计方法用于估计语音缺失先验概率，该方法基于时频域的三个分布参数，通过计算进而求得先验信噪比。

$$q(k,l)=1-P_{\text{local}}(k,l)*P_{\text{global}}(k,l)*P_{\text{frame}}(l) \tag{12}$$

式中：$P_{\text{local}}(k,l)$，$P_{\text{global}}(k,l)$ 分别表示局部或者全局先验信噪比均值与对应频点上语音似然比的比值，可表示为

$$P_{\lambda}(k,l)=\begin{cases}0, & \varsigma_{\lambda}(k,l)\leqslant\varsigma_{\min}\\ 1, & \varsigma_{\lambda}(k,l)\geqslant\varsigma_{\max}\\ \dfrac{\ln[\varsigma_{\lambda}(k,l)/\varsigma_{\min}]}{\ln(\varsigma_{\max}/\varsigma_{\min})}, & \text{其他}\end{cases} \tag{13}$$

式中：λ 可表示为 local 或者 global，ς_{λ} 为对应的先验信噪比均值。即如果先验信噪比均值小于一个指定阈值，表明该频点上语音存在概率较低，否则较高。

通过在频域加窗求得先验信噪比的均值，根据窗长的不同得到局部均值和全局均值，公式表示为

$$\varsigma_{\lambda}(k,l)=\sum_{i=-\omega_{\lambda}}^{\omega_{\lambda}}h_{\lambda}(i)\times\varsigma(k-i,l) \tag{14}$$

为了获得更好的性能，在使用估计先验信噪比之前，首先对其进行一阶递归平均，如式(15)所示：

$$\varsigma(k,l)=\beta\times\varsigma(k,l-1)+(1-\beta)\times\xi(k,l) \tag{15}$$

在式(12)中，$P_{\text{frame}}(l)$ 用于在语音信号不存在的条件下进一步衰减背景噪声，其取值取决于相邻连续输入信号能量的变化情况。对每帧的输入信号，定义信号比均值表示不同频点上的先验信噪比的均值，可表示为

$$\varsigma_{\text{frame}}(l)=\underset{1\leqslant k\leqslant N/2+1}{\text{mean}}\{\varsigma(k,l)\} \tag{16}$$

当从假设条件 $H_1(k,l)$ 向 $H_0(k,l)$ 过渡时，$P_{\text{frame}}(l)$ 的引入可以防止弱语音信号被过度衰减或者削幅。伪代码表示如下：

```
If    ς_frame(l)>ς_min    then
   If   ς_frame(l)>ς_frame(l−1)    then
      P_frame(l)=1
      ς_peak(l)=min{max[ς_frame(l),ς_pmin], ς_pmax}
   Else
      P_frame(l)=μ(l)
Else
   P_frame(l)=0
```

其中：

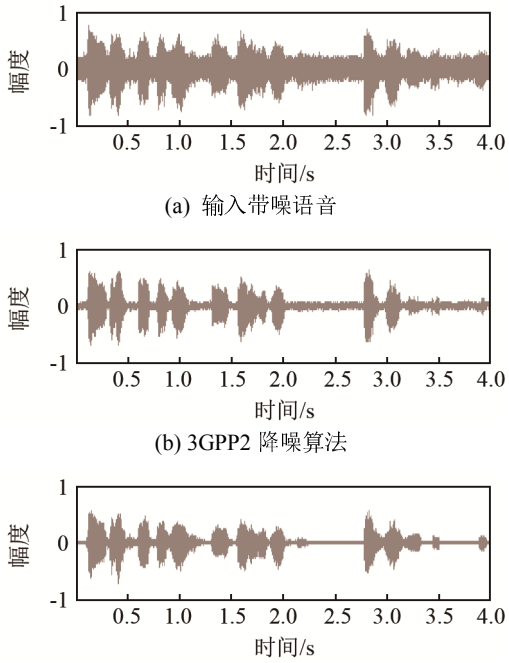
$$\mu(l)=\begin{cases}0, & \varsigma_{\text{frame}}(l)\leqslant\varsigma_{\min}\cdot\varsigma_{\text{peak}}(l)\\ 1, & \varsigma_{\text{frame}}(l)\geqslant\varsigma_{\max}\cdot\varsigma_{\text{peak}}(l)\\ \dfrac{\ln[\varsigma_{\text{frame}}(l)/\varsigma_{\text{peak}}(l)/\varsigma_{\min}]}{\ln(\varsigma_{\max}/\varsigma_{\min})}, & \text{其他}\end{cases} \tag{17}$$

3　仿真分析

文中提出的改进型对数幅度谱估计算法，相对于传统对数幅度谱估计算法，提出了一种新的先验信噪比估计方法和先验语音不存在概率估计方法。本节将在不同的背景噪声及不同的输入信噪比下，仿真分析文中所提改进算法的降噪性能。

仿真条件如下：带噪语音信号采样率为 8 kHz，噪声类型分别为：高斯白噪声、粉红噪声(Pink Noise)，输入信噪比为 0 dB。短时傅里叶变换(STFT)的长度为 256，每帧有效样本点数为 160，相邻两帧的重叠样本数为 48。最小增益因子阈值 $G_{\min}$ 设为 0.03，最小先验信噪比阈值 $\varsigma_{\min}$ 设为 0.1，最大先验信噪比阈值 $\varsigma_{\max}$ 设为 0.3，最小信噪比峰值阈值 $\varsigma_{p\min}$ 设为 1，最大信噪比峰值阈值 $\varsigma_{p\max}$ 设为 10。

图 2 和图 3 分别给出了本文算法在 0 dB 高斯白噪声条件下，与 3GPP2 文献中提出的降噪算法[12]在时域和频域的噪声抑制性能对比图。由图可知，本文算法相对于 3GPP2 文献中提出的降噪算法对高斯白噪声的抑制性能较好，并且未引入任何的音乐噪声及很小的信号畸变。



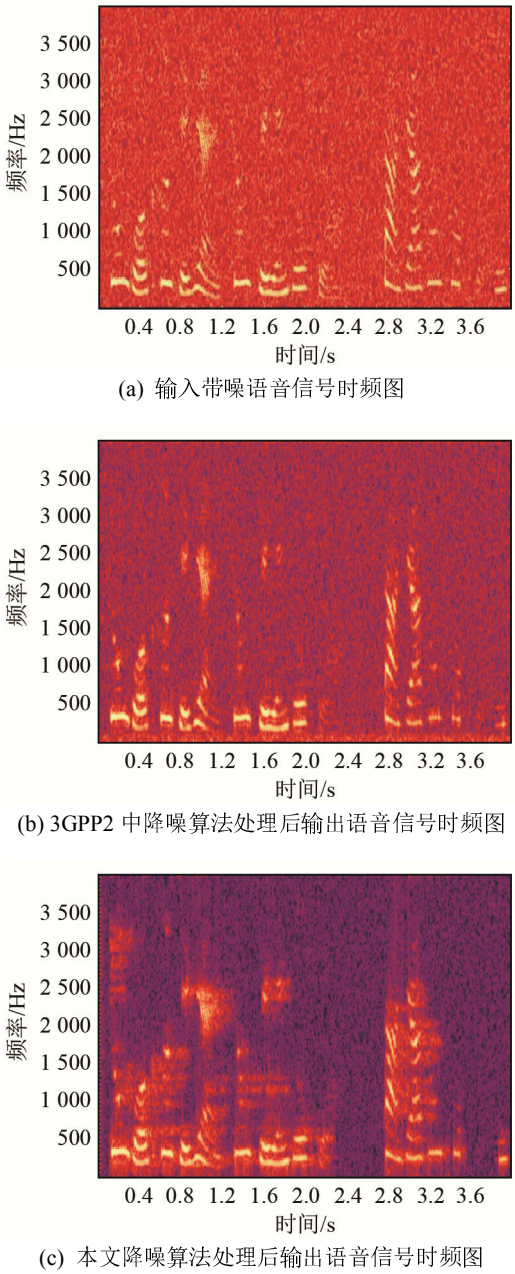
(a) 输入带噪语音

(b) 3GPP2 降噪算法

(c) 本文改进降噪算法

图 2　高斯白噪声条件下两种降噪算法的噪声抑制性能对比(时域信号)

Fig.2　Comparison of noise suppression performances between the improved and the 3GPP2 noise reduction algorithms under white Gaussian noise (see the signals in time domain)



(a) 输入带噪语音信号时频图

(b) 3GPP2 中降噪算法处理后输出语音信号时频图

(c) 本文降噪算法处理后输出语音信号时频图

图 3　高斯白噪声条件下两种降噪算法的噪声抑制性能对比(语谱图)

Fig.3　Comparison of noise suppression performances between the improved and the 3GPP2 noise reduction algorithms under white Gaussian noise (see the spectrograms)

图 4 给出了本文算法与 3GPP2 文献中提出的降噪算法以及基于对数谱估计的 MMSE 降噪算法在背景噪声为高斯白噪声时，不同输入信噪比下输出语音信噪比改进曲线对比图。从图 4 中可以看出，本文算法对低信噪比输入语音的噪声抑制能力明显优于其它两种降噪算法。

图 5 和图 6 分别给出了本文算法在 0 dB 粉红噪声条件下，在时域和频域的噪声抑制性能对比图。

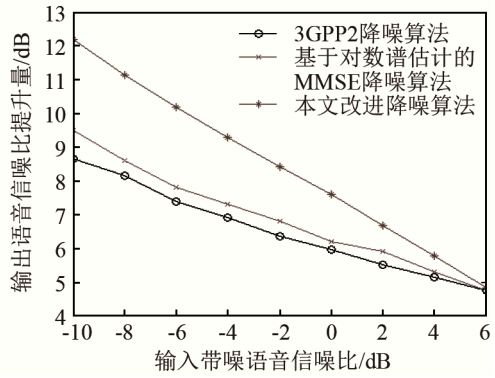
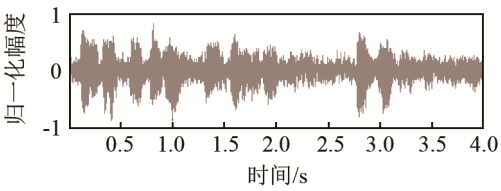
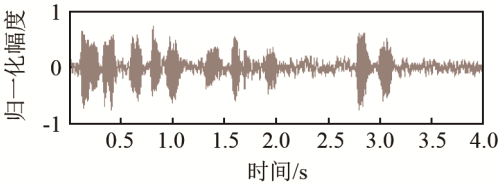


图 4 高斯白噪声不同输入信噪比条件下 3 种降噪算法性能曲线

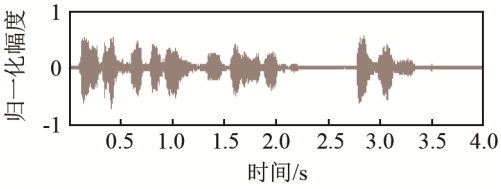
Fig.4 Noise reduction performances of MMSE, 3GPP2 and the improved algorithms under different SNRs in white Gaussian noise



(a) 输入带噪语音



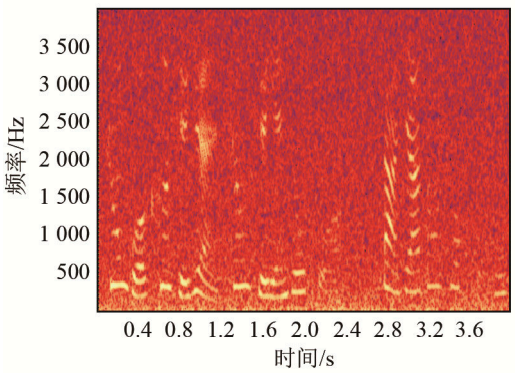
(b) 3GPP2 降噪算法



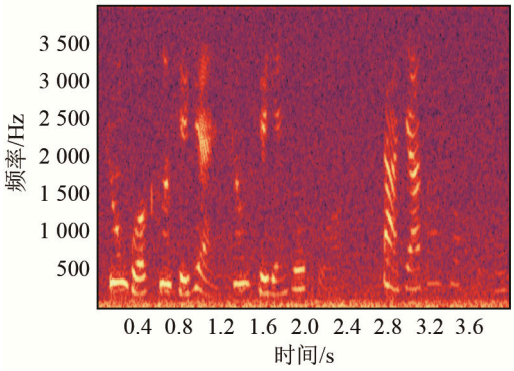
(c) 本文改进降噪算法

图 5 粉红噪声条件下两种降噪算法的噪声抑制性能对比(时域信号)

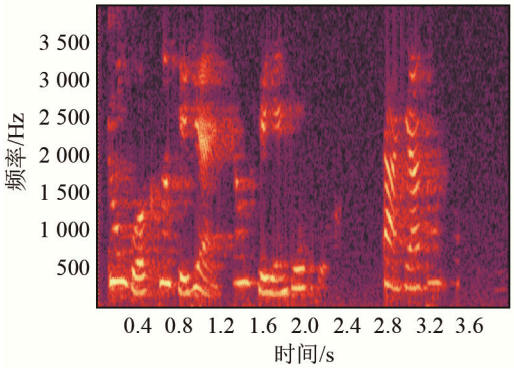
Fig.5 Comparison of noise suppression performances between the improved and the 3GPP2 noise reduction algorithms under pink noise (see the signals in time domain)



(a) 输入带噪语音信号时频图



(b) 3GPP2 中降噪算法处理后输出语音信号时频图



(c) 本文降噪算法处理后输出语音信号时频图

图 6 粉红噪声条件下两种降噪算法的噪声抑制性能对比(语谱图)

Fig.6 Comparison of noise suppression performances between the improved and the 3GPP2 noise reduction algorithms under pink noise (see the spectrograms)

从图中可以看出，本文算法能够很好地抑制粉红噪声，且有效地克服传统谱减算法中引入的音乐噪声。

图 7 给出了本文算法与 3GPP 文献提出的降噪算法以及基于对数谱估计的 MMSE 降噪算法在背景噪声为粉红噪声时，不同输入信噪比下输出语音信噪比改进曲线对比图。从图中可以看出，本文算法对粉红噪声的抑制能力明显优于其它两种降噪算法。

最后，给出在高斯白噪声和粉红噪声背景下不

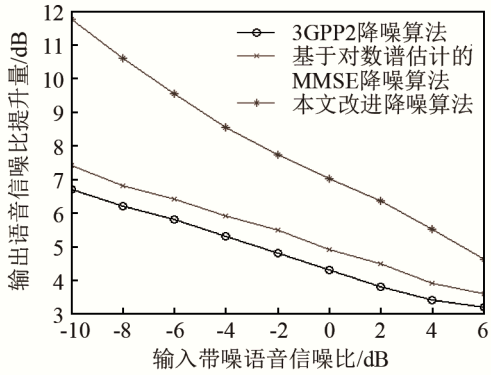


图 7 粉红噪声不同输入信噪比条件下 3 种降噪算法性能曲线

Fig.7 Noise reduction performances of MMSE, 3GPP2 and the improved algorithms under different SNRs in pink noise

同信噪比时的主观语音质量评估(Perceptual Evaluation of Speech Quality, PESQ)对比分析表格，如表 1 所示。从表 1 中可以看出，本文算法在输入低信噪比语音信号时，PESQ 得分明显高于 3GPP2 中给出的降噪算法以及基于对数谱估计的 MMSE 降噪算法。

表 1　高斯白噪声和粉红噪声条件下 3 种降噪算法的 PESQ 得分对比

Table 1　Comparison of PESQ betweem MMSE, 3GPP2 and the improved algorithms under white Gaussian noise and pink noise

SNR/dB	高斯白噪声			
	输入信号	3GPP2	MMSE	本文算法
−8	1.119	0.937	1.120	1.511
−4	1.262	1.258	1.265	1.790
0	1.495	1.793	1.803	2.097
4	1.763	2.260	2.265	2.302
SNR/dB	粉红噪声			
	输入信号	3GPP2	MMSE	本文算法
−8	1.021	1.052	1.102	1.519
−4	1.280	1.440	1.452	1.775
0	1.588	1.891	1.903	2.092
4	1.925	2.310	2.325	2.410

4　结束语

本文针对语音信号存在不确定条件下的语音增强问题，通过引入的改进型先验信噪比估计方法和先验语音不存在估计方法，提出了一种基于软判决技术的改进型对数幅度谱估计算法。仿真结果表明，即使在低信噪比条件下，本算法仍然能够有效地抑制带噪语音信号的加性白噪声，且有效地克服了传统降噪算法引入的“音乐噪声”及语音信号畸变。

参　考　文　献

[1] 赵力, 梁瑞宇, 等. 语音信号处理[M]. 北京: 机械工业出版社, 2017: 104-120.

[2] DEVYANI S, RATNADEEP R, PUKHRAJ P. A review of speech signal enhancement techniques[J]. International Journal of Computer Applications, 2016, **139**(14): 23-26.

[3] BOLL S. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Transactionson Acoustics Speech and Signal Process, 1979, ASSP-27(2): 113-120.

[4] EPHRAIM Y, MALAH D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator[J]. IEEE Transactions on Acoustics Speech and Signal Process, 1984, **32**(6): 1109-1121.

[5] CAPPE O. Elimination of the musical noise phenomenon with the Ephraim and malah noise suppressor[J]. IEEE Trans. Speech Audio Process, 1994, **2**(2): 345-349.

[6] SCALART P, FILLO J V. Speech enhancement based on a priori signal to noise estimation[C]//IEEE International Conference on Acoustics, Speech and Signal Process, 1996: 629-632.

[7] MALAH D, COX R V, ACCARDI A J. Tracking speech-presence uncertainty to improve speech enhancement non-stationary noise environments[J]. IEEE International Conference on Acoustics Speech and Signal Process, 1999, **10**(2): 1102-1105.

[8] COHEN I. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging[J]. IEEE Transactions on Acoustics Speech and Signal Process, 2003, **11**(5): 466-475.

[9] ABRAMSON A, COHEN I. Simultaneous detection and estimation approach speech enhancement[J]. IEEE Transactions On Audio, Speech, And Language Processing, 2007, **15**(8): 327-335.

[10] KIM N S, CHANG J H. Spectral enhancement based on global soft decision[J]. IEEE Signal Processing Letters, 2000, **7**(5): 108-110.

[11] EPHRAIM Y, MALAH D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator[J]. IEEE Trans. Acoust. Speech Signal Processing, 1985: 443-445.

[12] 3GPP2. Enhance variable rate codec, speech service options 3,68,70, and 73 for wideband spread spectrum digital systems[S]. 3GPP2 C.S0014-D, 2009: 83-92.