一种低信噪比语音的增强算法

李 晔　王 童　崔慧娟　唐 昆
(清华大学电子工程系 微波与数字通信国家重点实验室　北京 100084)

摘　要：为改善低信噪比环境下语音的质量，论文提出了一种新的语音增强算法。算法首先根据噪声频谱的高斯统计模型得到用先验信噪比形式表示的噪声频谱估计值，然后利用帧内、帧间平滑算法估计每一个频点的先验信噪比，从而能够更好地跟踪先验信噪比的变化。算法接着引入一种简便的估计语音在每一个频点出现概率的方法，得出一种新的语音增强算法。客观测试和非正式听音测试表明：该算法在几乎不损伤语音清晰度的前提下，能够更好地抑制低信噪比语音增强所产生的音乐噪声，同时使语音信噪比得到了明显提高。

关键词：语音信号处理；语音增强；音乐噪声；信噪比
中图分类号：TN912.3　　**文献标识码**：A　　**文章编号**：1009-5896(2007)09-2054

Enhancement Algorithm for Low Signal to Noise Ratio Speech

Li Ye　Wang Tong　Cui Hui-juan　Tang Kun
(*State Key Laboratory of Microwave and Digital Communication*, *Department of Electronic Engineering*, *Tsinghua University*, *Beijing* 100084, *China*)

Abstract: This paper proposes a new speech enhancement algorithm to improve the quality of speech in low SNR environment. The algorithm estimates the noise spectrum in the form of prior SNR according to its Gaussian statistical model firstly. Then it estimates the prior SNR of each frequency bin utilizing the intra-frame as well as inter-frame smoothness to trace the quick change of the prior SNR. It also introduces a simple method to compute the speech presence probability to each frequency bin and obtain the new speech enhancement algorithm. Objective measurement combined with informal listening test shows that the new algorithm is more effective in eliminating musical noise and improving SNR obviously without significantly impairing the intelligibility of speech.

Key words: Speech signal processing; Speech enhancement; Musical noise; Signal to Noise Ratio (SNR)

1　引言

语音通信中，话音可能受到各种各样噪声的干扰，从而严重影响通信质量，采用语音增强技术是解决上述问题的有效途径。而基于单麦克输入的语音增强算法由于实现简单而成为研究的热点。目前基于单麦克输入的语音增强算法主要有谱相减法[1]、维纳滤波方法以及最小均方误差估计法和基于听觉掩蔽效应的方法[2]等。其中，谱相减法引入的约束条件最少，运算量小，物理意义最为直接，而且经过改进以后效果也较好，因此被广泛采用。

2　谱相减法简要介绍

设一帧加窗后的带噪语音的时域表示为

$$y_{n,i} = x_{n,i} + d_{n,i},\quad 0 \le n \le N-1 \tag{1}$$

其中 i 表示第 i 帧，n 表示第 n 个样点，N 为帧长，$x_{n,i}$ 为纯净语音，$d_{n,i}$ 为加性噪声，$y_{n,i}$ 为带噪语音。

相应的频域表示为

$$Y_{k,i} = X_{k,i} + D_{k,i} \tag{2}$$

由于通常 $x_{n,i}$ 和 $d_{n,i}$ 是不相关的，所以有

$$E[|Y_{k,i}|^2] = E[|X_{k,i}|^2] + E[|D_{k,i}|^2] \tag{3}$$

$E[|D_{k,i}|^2]$ 可以通过对端点检测算法检测到的语音间歇段统计平均得到，设为 $\lambda^2_{k,i}$，对于分析帧内的短时平稳过程，就可以得到原始语音的幅度谱的估计值：

$$|\widehat{X}_{k,i}| = \begin{cases} [|Y_{k,i}|^2 - \lambda^2_{k,i}]^{1/2}, & |Y_{k,i}|^2 \ge \lambda^2_{k,i} \\ a|Y_{k,i}|, & \text{其他} \end{cases} \tag{4}$$

式(4)中 a 是一个很小的常数。由于人耳对语音的感知是通过语音信号中各频谱分量的幅度获取的，对各分量的相位则不敏感[3]，因此直接用带噪语音的相位作为增强以后的语音的相位，就可得到原始语音的估计值。

3　噪声频谱估计

传统谱减算法中对噪声频谱的估计采用了长时平均，即通过对以往噪声段的频谱特征进行平均得到，而这样得到的噪声频谱值 $\lambda^2_{k,i}$ 有可能超过当前带噪语音的频谱值 $|Y_{k,i}|^2$，从而根据式(4)求 $|X_{k,i}|^2$ 时，往往采用半波整流的方法，即令 $|X_{k,i}|$ 为一个很小的正数值与带噪语音幅度谱的乘积，由此会引起严重的音调噪声。

假设噪声信号离散频谱服从复高斯分布，则其离散幅度谱序列服从瑞利分布[4]。设该离散幅度谱序列服从参数为 σ

2006-02-20 收到，2006-09-21 改回

的瑞利分布，即离散幅度谱满足如下分布：

$$f(x)=\begin{cases}\dfrac{x}{\sigma^2}\exp\left(-\dfrac{x^2}{2\sigma^2}\right), x\ge 0\\ 0, \qquad\qquad x<0\end{cases} \tag{5}$$

因为参数为 σ 的瑞利分布的均值为 $\sqrt{2\pi\sigma^2}/2$，且噪声幅度谱的均值又可以通过对以往噪声段的平均得到，所以可以得到 σ 的估计值。如果直接根据噪声幅度谱的瑞利统计模型求噪声幅度谱的估计值且不做任何限制，则由于一方面式(3)是一个统计概率意义上的等式，并不代表具体在每一帧的每一个频点处都能满足 $|D_{k,j}|<|Y_{k,j}|$；另一方面 σ 的计算是通过时域平均代替统计平均，所以对当前频点而言，根据以往噪声谱统计求得的 $\widehat{D}_{k,i}$ 仍有可能大于当前帧 $|Y_{k,j}|$。因此，为消除音调噪声，在求当前第 k 帧第 i 个频点噪声幅度谱估值 $|\widehat{D}_{k,i}|$ 的时候，运用瑞利统计模型并根据第 k 帧第 i 个频点处的带噪幅度谱限制 $|D_{k,j}|<|Y_{k,j}|$，具体求法如下：

$$\begin{aligned}\left|\widehat{D}_{k,i}\right| &= E\left(\left|D_{k,i}\right| \left\| \left|D_{k,i}\right|<\left|Y_{k,i}\right|\right)\right.\\ &=\int_0^{|Y_{k,i}|} xf(x)\mathrm{d}x \Big/ \int_0^{|Y_{k,i}|} f(x)\mathrm{d}x\\ &=\int_0^{|Y_{k,i}|}\frac{x^2}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)\mathrm{d}x \Big/ \int_0^{|Y_{k,i}|}\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)\mathrm{d}x\end{aligned} \tag{6}$$

根据瑞利分布性质，有

$$E[|D_{k,i}|]=\sqrt{\pi 2\sigma^2}/2 \tag{7}$$

将式(7)带入式(6)，中可得

$$\begin{aligned}&E\left(\left|D_{k,i}\right| \left\| \left|D_{k,i}\right|<\left|Y_{k,i}\right|\right)\right.\\ &=\int_0^{|Y_{k,i}|}\frac{\pi x^2}{2(E[|D_{k,i}|])^2}\exp\left(-\frac{\pi x^2}{4(E[|D_{k,i}|])^2}\right)\mathrm{d}x\\ &\quad\Big/\int_0^{|Y_{k,i}|}\frac{\pi x}{2(E[|D_{k,i}|])^2}\exp\left(-\frac{\pi x^2}{4(E[|D_{k,i}|])^2}\right)\mathrm{d}x\\ &=\frac{-\exp\left(-\dfrac{\pi|Y_{k,i}|^2}{4(E[|D_{k,i}|])^2}\right)+\dfrac{E[|D_{k,i}|]}{|Y_{k,i}|}\mathrm{erf}\left(\dfrac{\sqrt{\pi}|Y_{k,i}|}{2E[|D_{k,i}|]}\right)}{1-\exp\left(-\dfrac{\pi|Y_{k,i}|^2}{4(E[|D_{k,i}|])^2}\right)}\cdot|Y_{k,i}|\end{aligned} \tag{8}$$

式(8)中，$\mathrm{erf}(x)=\dfrac{2}{\sqrt{\pi}}\int_0^x\exp(-t^2)\mathrm{d}t$。

定义后验信噪比 $\gamma_{k,i}=\dfrac{|Y_{k,i}|^2}{(E[|D_{k,i}|])^2}$，先验信噪比 $\xi_{k,i}=P[\gamma_{k,i}-1]$，则式(8)可以近似表示为

$$\begin{aligned}&E\left(\left|D_{k,i}\right| \left\| \left|D_{k,i}\right|<\left|Y_{k,i}\right|\right)\right.\\ &=\frac{-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)+\dfrac{1}{\sqrt{1+\xi_{k,i}}}\mathrm{erf}\left(\dfrac{\sqrt{\pi(1+\xi_{k,i})}}{2}\right)}{1-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}\cdot|Y_{k,i}|\end{aligned} \tag{9}$$

从而，原始语音谱幅度的估计值可表示为

$$|\widehat{X}_{k,i}|=\left\{P\left[1-a\left(\frac{-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}{1-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}+\frac{\dfrac{1}{\sqrt{1+\xi_{k,i}}}\mathrm{erf}\left(\dfrac{\sqrt{\pi(1+\xi_{k,i})}}{2}\right)}{1-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}\right)^2\right]\right\}^{\frac{1}{2}}*|Y_{k,i}| \tag{10}$$

其中 a 为过减因子，目的是进一步去除噪声，$P[\,]$ 表示半波整流。

4　语音存在概率

上述推导过程中，一直假定带噪语音各频点包含纯净语音，实际上，在语音间隔处等许多地方，纯净语音并不存在。为此，文献[5]对传统的语音增强模型进行修正，假设第 k 帧第 i 个频率点存在语音的概率为 $p_{k,i}$，则可以得到最终的纯净语音频谱估计值为

$$|\widehat{X}_{k,i}|=(G_{k,i}|Y_{k,i}|)^{p_{k,i}}(\beta|Y_{k,i}|)^{(1-p_{k,i})} \tag{11}$$

其中 $G_{k,i}$ 为传统语音增强模型中所求得的各种增益因子，在本算法中，

$$G_{k,i}=\left\{P\left[1-a\left(\frac{-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}{1-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}+\frac{\dfrac{1}{\sqrt{1+\xi_{k,i}}}\mathrm{erf}\left(\dfrac{\sqrt{\pi(1+\xi_{k,i})}}{2}\right)}{1-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}\right)^2\right]\right\}^{\frac{1}{2}} \tag{12}$$

β 为一常数衰减因子，因此，将式(12)带入式(11)中即可以得到修正后的原始语音频谱估计值为

$$|\widehat{X}_{k,i}|=\left\{P\left[1-a\left(\frac{-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}{1-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}+\frac{\dfrac{1}{\sqrt{1+\xi_{k,i}}}\mathrm{erf}\left(\dfrac{\sqrt{\pi(1+\xi_{k,i})}}{2}\right)}{1-\exp\left(-\dfrac{\pi(1+\xi_{k,i})}{4}\right)}\right)^2\right]\right\}^{\frac{p_{k,i}}{2}}\cdot\beta^{1-p_{k,i}}*|Y_{k,i}| \tag{13}$$

语音存在概率的估计参考用文献[6]中的方法，计算如下：

$$p_{k,i}=\begin{cases}0, & \xi_{k,i}<\xi_{\min}\\ 1, & \xi_{k,i}>\xi_{\max}\\ \dfrac{\lg(\xi_{k,i}/\xi_{\min})}{\lg(\xi_{\max}/\xi_{\min})}, & \text{其他}\end{cases} \tag{14}$$

其中 $\xi_{\min}$ 和 $\xi_{\max}$ 为经验常数，第 k 帧的第 i 个频点的先验信噪比 $\xi_{k,i}$ 小于 $\xi_{\min}$ 时，则假定此频点不存在语音，反之当大于 $\xi_{\max}$ 时，存在语音的概率为 1，介于二者之间时，采用对数函数对语音存在概率进行估计，先验信噪比越大，语音存在的概率也越大，试验证明，这种计算概率方法是简单有效的[6]。

5 先验信噪比估计

从前面的推导可以看出，只要估计出先验信噪比，就可以通过式(13)、式(14)估计纯净语音，因此先验信噪比的估计至关重要。文献[7]中提出一种对先验信噪比采用反馈估计的方法，基本思想如下：

$$\hat{\xi}_{k,i}=a_{k,i}\frac{|\hat{X}_{k,i-1}|^2}{E[|D_{k,i-1}|^2]}+(1-a_{k,i})P[\gamma_{k,i}-1] \tag{15}$$

并且采用最小均方误差准则对 $a_{k,i}$ 进行估计，对给定的 $\xi_{k,i-1}$，均方估计误差为

$$J(a_{k,i})=E\{(\hat{\xi}_{k,i}-\xi_{k,i})^2\,|\,\hat{\xi}_{k,i-1}\} \tag{16}$$

假设语音谱系数和噪声谱系数是相互独立的、零均值复高斯随机变量，这与前面的假设模型一致，语音谱幅度服从瑞利分布，从而令 $\partial J(a_{k,i})/\partial a_{k,i}=0$，可以得到 $a_{k,i}$ 在最小均方误差准则下的估计值：

$$a_{k,i}=\frac{1}{1+\left(\dfrac{p[\gamma_{k,i}-1]-\hat{\xi}_{k,i-1}}{p[\gamma_{k,i}-1]+1}\right)^2} \tag{17}$$

在文献[7]算法的基础上，本文考虑到同一语音帧相邻频率点间先验信噪比的相关性，算法对先验信噪比进行帧内平滑如下：

$$\hat{\xi}_{k,i}=(\hat{\xi}_{k-1,i}+2\cdot\hat{\xi}_{k,i}+\hat{\xi}_{k+1,i})/4 \tag{18}$$

从而，将式(18)代入式(15)得到先验信噪比的估计，然后由先验信噪比的估计值根据式(14)求得语音存在概率，进而将先验信噪比估计以及语音存在概率估计代入式(13)估计纯净语音。

6 实验结果

本论文采用 8000Hz 采样的语音，以及 NOISEX-92 噪声库中的白噪声和坦克噪声,帧长设定为 32ms。图 1 和图 2 分别给出了在 0dB 白噪声和坦克噪声中的增强结果。其中(a)为带噪语音，(b)为增强后语音，横坐标为语音采样点序列，纵坐标为语音幅度。

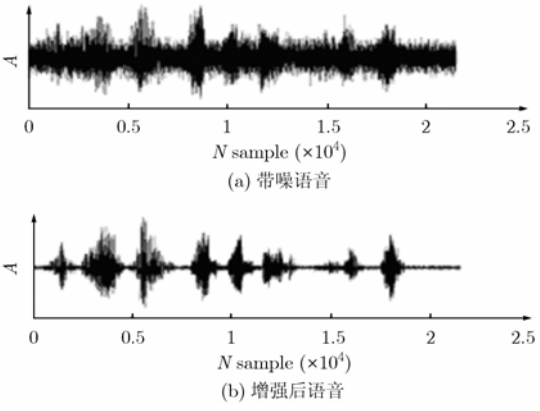

图 1 白噪声环境下的语音增强结果

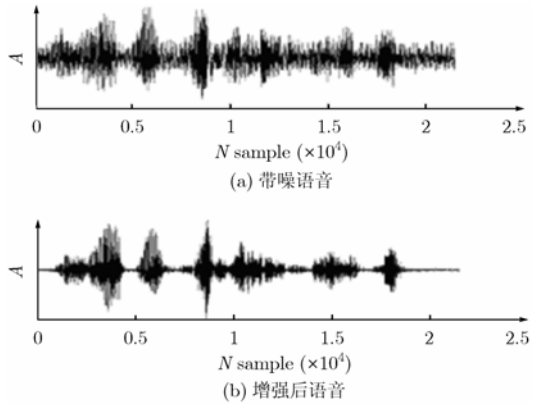

图 2 坦克噪声环境下的语音增强结果

语音增强算法在提高带噪语音信噪比以及听觉舒适度的同时，会对语音清晰度产生一定影响，为此，进行白噪声环境下的 0dB 和 5dB 带噪语音以及增强后语音诊断押韵测试(DRT)得分对比，DRT 测试使用若干对同韵母进行测试，让受听者每次听一对韵字中的某个音，判断是哪一个字，实验者判断正确的百分比就是 DRT 得分，结果如表 1 所示，可以看出，采用本文算法增强后的语音清晰度几乎没有损伤。

表 2 给出了在白噪声和坦克噪声环境下不同信噪比语音增强的结果，并与实际中广泛采用的谱减法做了比较。

表 1 带噪语音与增强后语音 DRT 测试结果 (%)

DRT 测试项		浊音性	鼻音性	送气性	低沉性	紧密性	持续性	DRT 总分
0dB	带噪语音	98.63	91.78	94.65	82.15	89.10	69.52	87.64
	增强语音	100.00	96.06	91.80	91.67	77.91	63.94	86.90
5dB	带噪语音	100.00	94.56	98.66	99.17	84.90	79.22	92.75
	增强语音	100.00	95.93	98.78	95.92	84.83	77.95	92.24

表 2　4 种输入语音文件增强后的输出信噪比 (dB)

输入信噪比(dB)		−10	−5	0	5
白噪声	谱减法	0.2	3.2	6.5	8.8
	本文算法	2.5	4.6	10.5	12.6
坦克噪色	谱减法	0.2	3.1	6.2	8.5
	本文算法	2.3	4.4	9.8	12.0

实验结果表明，本文算法能够比谱减法提高更多的信噪比，同时，能够更好地抑制音调噪声并且基本不损坏语音的清晰度。

7　结束语

论文利用噪声幅度谱的瑞利模型提出了新的以先验信噪比形式表达的噪声频谱估计算法，同时，对先验信噪比进行帧间反馈估计以及帧内频域平滑，最后利用语音存在概率模型对上述算法进一步进行改进。主观听觉测试表明在几乎不损坏语音清晰度的同时，算法能够很好地抑制低信噪比语音增强后所产生的音乐噪声，同时大幅度提高输入带噪语音的信噪比，已经成功运用于声码器前端作为预处理模块。

参 考 文 献

[1] Berouti M, Schwartz R, and Makhoul J. Enhancement of speech corrupted by acoustic noise [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Washington DC, 1979, Vol. 4: 208−211.

[2] 张金杰, 曹志刚, 马正新. 一种基于听觉掩蔽效应的语音增强方法 [J]. 清华大学学报(自然科学版)，2001，41(7): 1−4.
Zhang Jin-jie, Cao Zhi-gang, and Ma Zheng-xin. Speech enhancement method based on auditory masking [J]. *J Tsinghua Univ.* (*Sci & Tech*), 2001, 41(7): 1−4. (in Chinese)

[3] Wang D L and Lim J S. The unimportance of phase in speech enhancement [J]. *IEEE Trans. on Acoustic, Speech, Signal Processing*, 1982, 30(4): 679−681.

[4] Reibman A R and Nolte L W. Design and performance comparison of distributed detection networks [J]. *IEEE Trans.on Aerospace and Electronic Systems*, 1987, 23(6): 789−797.

[5] McAulay R J and Malpass M L. Speech enhancement using a soft-decision noise suppression filter [J]. *IEEE Trans.on Acoustic, Speech, Signal Processing*, 1980, 28(2): 137−145.

[6] Cohen I and Berdugo B.Speech enhancement for non-station ary noise environments [J]. *Signal Processing*, 2001, 81(11): 2403 −2418.

[7] Hasan M K, Salahuddin S and Khan M R. A modified *a priori* SNR for speech enhancement using spectral subtraction rules [J]. *IEEE Signal Processing Letters* , 2004, 11(4): 450−453.

李　晔：男，1981 年生，博士生，研究方向为语音信号处理等.

王　童：男，1983 年生，硕士生，研究方向为语音信号处理等.

崔慧娟：女，1945 年生，教授， 主要研究方向为语音信号处理、图像处理等.

唐　昆：男，1945 年生，教授，博士生导师，主要研究方向为语音信号处理、图像处理等.