基于增益字典查询的语音增强算法

庞 亮[1] 陈 亮[1] 张翼鹏[2] 黄清泉[1]
(解放军理工大学通信工程学院 南京 210007)[1] (解放军南京炮兵学院 南京 211132)[2]

摘 要 对于基于统计模型的语音增强算法,不同分布模型对应于不同的增益函数,由于语音信号的不确定性,没有一种分布函数能准确对语音和噪声谱的分布建模,因此任何一种固定的统计模型均会存在一定的误差。所以提出一种增益字典查询的语音增强算法,该算法通过采用对数谱失真准则对一个语音噪声库进行增益的训练,得到一个增益的字典,其中输入为先验信噪比和后验信噪比的估计值。最后采用 ITU-T P. 826 PESQ、分段信噪比、总信噪比和对数谱失真对该算法进行了测试,并与基于高斯分布模型、拉普拉斯分布模型的算法进行了对比。实验结果表明,该算法无论在非平稳噪声还是平稳噪声环境下都比其他几种算法增强效果好,且音乐噪声和残留背景噪声也可以得到很好的抑制。

关键词 语音增强,字典查询,判决引导,改进递归平均算法
中图法分类号 TN912.35 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.10.004

Speech Enhancement Based on Gain Dictionary Queries

PANG Liang[1] CHEN Liang[1] ZHANG Yi-peng[2] HUANG Qing-quan[1]
(Institute of Communication Engineering, PLA University of Science and Technology, Nanjing 210007, China)[1]
(Nanjing Artillery Academy, Nanjing 211132, China)[2]

Abstract For speech enhancement algorithm based on statistical model, different distribution models are corresponding to different gain function, due to the uncertainty of the speech signal, no distribution function can accurately model the speech and noise spectra distribution, so any kind of fixed reference models will have some errors. We presented a gain dictionary queries based speech enhancement algorithm, getting a dictionary gain through training the voice of a noise library using log-spectral distortion criterion, for which the input is the estimate value of a priori and a posteriori SNR. Finally, we used ITU-T P. 826 PESQ, segmented SNR, total SNR and log-spectral distortion criterion to test the proposed algorithm, and compared this algorithm with Gaussian distribution model and Laplace distribution model. The experimental results show that the algorithm is better than the other algorithms, whether in stationary or non-stationary noisy environments, and musical noise and residual background noise can be well suppressed.

Keywords Speech enhancement, Dictionary queries, Decision-directed, IMCRA

1 引言

基于离散傅里叶变换(DFT)的单通道语音增强技术以其较低复杂度和较好的效果在目前得到广泛的应用和研究,尤其是频域的基于统计模型的语音增强算法。目前,应用最多的最小均方误差(MMSE)估计器[1]和对数谱最小均方误差(log-MMSE)估计器[2]都是基于傅里叶系数的实部和虚部是高斯分布的假设,在分析帧较长时,纯净信号的 DFT 系数的实部和虚部的分布是逼近高斯模型的,因为这种情况下信号相关区间的跨度比 DFT 长度小得多。这个假设可能对噪声 DFT 系数成立,而对于语音,一般我们在估计时会使用相对短(20ms～30ms)的窗,这个假设并不成立,因此有一些学者提出对语音部分的 DFT 系数使用非高斯模型,如利用高斯混合模型[3,4]、伽马(Gamma)分布[5,6]以及拉普拉斯(Laplacian)分布[7]对 DFT 系数的实部和虚部进行建模。从图 1 中可以看出,由于语音信号的不确定性,没有一种固定的分布,因此使用任何一种固定的分布函数对语音信号建模均会存在一定的偏差。

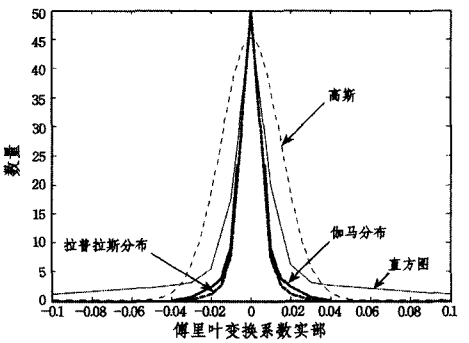


图 1 利用 10s 语音所得傅里叶变换系数实部的直方图

到稿日期:2014-10-20 返修日期:2015-01-25

庞 亮(1990－),男,硕士生,主要研究方向为语音信号处理,E-mail:pangliang411@163.com;陈 亮(1974－),男,博士,教授,主要研究方向为多媒体信息处理;张翼鹏(1988－),男,硕士,助教,主要研究方向为信息隐藏技术;黄清泉(1993－),男,硕士生,主要研究方向为数字图像处理。

本文针对上述问题，提出一种基于纯净语音谱增益函数字典查询的方法，该方法首先对一个带噪语音库运用对数谱失真准则进行训练，训练的带噪语音信噪比范围为[-19dB, 40dB]，得到一个增益字典，步进为 1dB，输入为先验信噪比和后验信噪比的估计值。先验信噪比的估计采用改进的低时延判决引导(Decision-Directed，DD)法，噪声功率谱的估计采用 IMCRA 算法[8]。从实验结果可以看出，该算法相比于其他几种算法能够获得较好的语音增强效果，并且采用查表的方法比增益计算更快速，占用的内存也较小。

2 先验信噪比估计

增益函数的两个输入为先验信噪比和后验信噪比，在实际情况下，这两个变量都是未知的，只能得到估计值，且准确的噪声估计和增益函数的估计都需要特殊的先验信噪比估计器。文献[9]提出的采用非因果先验信噪比估计算法，虽然估计先验信噪比延时较小，但该算法需要未来帧信息，实时性不够。大多数的算法都是采用了传统的判决引导法[10]，它是基于先验信噪比的定义及其与后验信噪比的关系，通过递归得到。先验信噪比和后验信噪比的表达式为：

$$\hat{\xi}(k,m)=\max\{a\frac{|\hat{X}(k,m-1)|^2}{\lambda_r(k,m-1)}+(1-a)\max[\gamma(k,m)-1,0],\xi_{\min}\} \tag{1}$$

$$\gamma(k,m)=\frac{|Y(k,m)|^2}{E\{|R(k,m)|^2\}}=\frac{|Y(k,m)|^2}{\lambda_r(k,m)} \tag{2}$$

其中，$a=0.98$ 为平滑因子，$\hat{X}(k,m-1)$为上一帧估计的纯净语音，$\lambda_r(k,m-1)$为上一帧估计的噪声功率谱，$\xi_{\min}$是 $\xi(k,m)$所允许的最小值，通过限定$\hat{\xi}(k,m)$的下限来控制可能产生的音乐噪声。但从图 2 可以明显看出，采用 DD 算法相比于 $\gamma-1$存在一帧的延时，且依赖于上一帧所估计的纯净语音，因此在话音的起端和终点处，DD 算法并不能很好地反映出当前帧状况，这些会对语音的质量产生较大影响。因此，本节使用了一种改进的低时延 DD 算法(MDD)，使用当前帧的语音信号代替上一帧的纯净语音，增益函数仍然使用上一帧计算的增益函数，同时噪声采用当前帧所估计的噪声。具体表达式如下：

$$\hat{\xi}(k,m)=\max\{a\frac{|G(k,m-1)Y(k,m)|^2}{\lambda_r(k,m)}+(1-a)\max[\gamma(k,m)-1,0],\xi_{\min}\} \tag{3}$$

其中，平滑因子和先验信噪比允许的最小值均与传统算法相同，$\xi_{\min}=-19$dB。

在式(3)中，由于第一项没有使用上一帧的先验信噪比，因此不再是一个递归平滑的算法。这样可能会导致算法对语音的突变较敏感，从而产生一定的音乐噪声。为此本文对后验信噪比计算方法进行了改进，不再直接使用当前帧的带噪语音，而是对当前帧的带噪语音先进行平滑，再计算后验信噪比，具体表达式如下：

$$\bar{\gamma}(k,m)=\frac{\lambda_y(k,m)}{\lambda_r(k,m)} \tag{4}$$

其中，$\lambda_y(k,m)=b\lambda_y(k,m-1)+(1-b)|Y(k,m)|^2$ 为平滑的带噪语音功率谱，$b=0.72$ 为平滑因子。

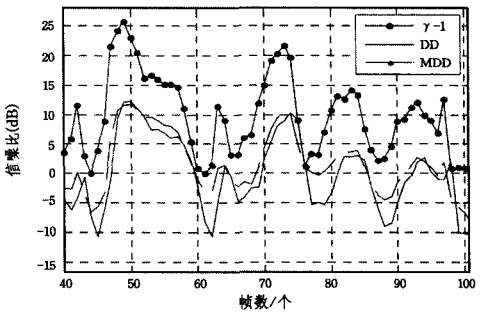


图 2 信噪比时延对比图

3 噪声估计

时间递归平均算法利用了噪声一般情况下对语音频谱具有不均匀影响这一现象，即有些频谱区域受到噪声的影响比其他一些区域所受影响更大。不同频谱分量具有不同的实际信噪比(SNR)，因此在某一特定的频带的实际 SNR 很低时，可以对噪声按频带来进行估计和更新。最小值控制递归平均算法(IMCRA)的噪声谱估计就是利用过去的噪声估计与当前带噪语音谱的加权平均，通过跟踪平滑后的功率谱的最小值求得当前频点的语音存在概率来自适应地改变权重。

算法共需要 2 次功率谱平滑，第 1 次平滑功率谱通过下式可得：

$$S(k,m)=\alpha_s S(k,m-1)+(1-\alpha_s)S_f(k,m) \tag{5}$$

其中，α_s 为平滑因子，且

$$S_f(k,m)=\sum_{i=-L_w}^{L_w}w(i)|Y(k-i,m)|^2 \tag{6}$$

其中，$w(i)$为 Hanning 窗函数，窗长为 $2L_w+1$。不断更新该平滑功率谱的最小值 $S_{\min}(k,m)$。然后通过以下规则得到一个粗略的语音存在性的判决结果：

$$I(k,m)=\begin{cases}1, & \text{如果 } \gamma_{min}(k,m)<\gamma_0 \text{ 且 } \zeta(k,m)<\zeta_0\\ 0, & \text{其他}\end{cases} \tag{7}$$

其中，γ_0 和 ζ_0 为阈值参数，且

$$\gamma_{\min}(k,m)=\frac{|Y(k,m)|^2}{B_{\min}S_{\min}(k,m)}$$
$$\zeta(k,m)=\frac{S(k,m)}{B_{\min}S_{\min}(k,m)} \tag{8}$$

其中，因子 $B_{\min}=1.66$ 代表了最小噪声估计的偏差。

第 2 次平滑仅针对已经被式(7)基本上判断为噪声的那些频率分量。具体形式如下：

$$\tilde{S}_f(k,m)=\begin{cases}\dfrac{\sum_{i=-L_w}^{L_w}w(i)I(k-i,m)|Y(k-i,m)|^2}{\sum_{i=-L_w}^{L_w}w(i)|I(k-i,m)|^2}, & \text{如果 } \sum_{i=-L_w}^{L_w}I(k-i,m)\neq 0\\ \tilde{S}(k,m-1), & \text{其他}\end{cases} \tag{9}$$

其中，$w(i)$为窗函数(频域)，L_w 为窗长度(频点数量)，$\tilde{S}(k,m)$定义如式(10)。在计算式(9)的 $\tilde{S}_f(k,m)$之后，进行如下的一阶递归平均：

$$\tilde{S}(k,m)=\alpha_s\tilde{S}(k,m-1)+(1-\alpha_s)\tilde{S}_f(k,m) \tag{10}$$

式(9)是在频域进行的平滑操作，而式(10)是在时域进行

平滑，并实时更新最小值 $\tilde{S}_{\min}(k,m)$，该帧不存在语音的先验概率 $\hat{q}(k,m)$ 计算如下：

$$\hat{q}(k,m)=\begin{cases}1, & \text{if } \tilde{\gamma}_{\min}(k,m)\leqslant 1 \& \tilde{\zeta}(k,m)<\zeta_0 \\ \dfrac{\gamma_1-\tilde{\gamma}_{\min}(k,m)}{\gamma_1-1}, & \text{if } 1<\tilde{\gamma}_{\min}(k,m)<\gamma_1 \& \tilde{\zeta}(k,m)<\zeta_0 \\ 0, & \text{others}\end{cases} \tag{11}$$

其中，γ_1 为一个阈值参数，且

$$\tilde{\gamma}_{\min}(k,m)=\frac{|Y(k,m)|^2}{B_{\min}\tilde{S}_{\min}(k,m)}$$
$$\tilde{\zeta}(k,m)=\frac{\tilde{S}(k,m)}{B_{\min}\tilde{S}_{\min}(k,m)} \tag{12}$$

则语音存在的条件概率为：

$$p(k,m)=\frac{1}{1+\dfrac{q(k,m)}{1-q(k,m)}(1+\xi(m))\exp(-v(k,m))} \tag{13}$$

其中，$v(k,m)=\gamma(m)\xi(m)/(1+\xi(m))$，$\gamma(m)$ 和 $\xi(m)$ 分别为当前语音帧的后验信噪比和先验信噪比。从而可以得到噪声功率谱的平滑参数 $\alpha_d(k,m)$：

$$\alpha_d(k,m)=\alpha+(1-\alpha)p(k,m) \tag{14}$$

其中，α 为一常数值 0.85，则噪声功率的估计可以表示为：

$$\lambda_d(k,m)=\alpha_d(k,m)\lambda_d(k,m-1)+[1-\alpha_d(k,m)]|Y(k,m)|^2 \tag{15}$$

此外，对噪声估计引入了一个偏差补偿因子 β：

$$\tilde{\lambda}_d(k,m)=\beta\lambda_d(k,m) \tag{16}$$

其中，β 设为 1.47，引入偏差项是为了最小化语音失真。

4 增益字典的训练

基于统计模型的语音增强均可以表示为如下形式：

$$\tilde{A}_k=\tilde{G}(\hat{\xi},\gamma)R \tag{17}$$

其中，G 是关于先验信噪比 $\hat{\xi}$ 和后验信噪比 γ 的一个非线性函数，基于字典的语音增强就是通过大量的数据训练，得到一个增益的表格，输入为先验信噪比和后验信噪比的估计值，每一对 $(\hat{\xi},\gamma)$ 都对应一个最佳的增益值。训练的数据使用整个 TIMIT-TRAIN 库的语音，这些语音信号的频率都是限制在 300Hz～3400Hz 的常用范围内，通过对纯净语音增加不同大小的高斯白噪声，使得输入信号的信噪比从－19dB～40dB 不等。对于训练的数据，每一帧、每一个频点都需要计算出相应的先验信噪比和后验信噪比，这些值被放入在一个横坐标为－19dB～40dB、纵坐标为－30dB～40dB、步进为 1dB 的表格中，且每一对 $(\hat{\xi},\gamma)$ 都有一个对应的 (A^2,R^2)，最后利用对数最小均方差的准则计算落在同一格内所有点的最佳增益值，对数最小均方误差的表达式为[11]：

$$D=\sum_{m=1}^{M_{ij}}\{\log[A_{ij}(m)]-\log[G_{ij}R_{ij}(m)]\}^2 \tag{18}$$

则最佳增益的表达式表示为：

$$\tilde{G}_{ij}=\sqrt[M_{ij}]{\prod_{m=1}^{M_{ij}}\frac{A_{ij}(m)}{R_{ij}(m)}} \tag{19}$$

其中，M_{ij} 表示落在同一格的点数。

在增益的训练中，使用对数最小均方误差估计器求增益的估计值，表达式为：

$$G_{LSA}=\frac{\hat{\xi}}{\hat{\xi}+1}\exp\{\frac{1}{2}\int_{v_k}^{\infty}\frac{e^{-t}}{t}dt\} \tag{20}$$

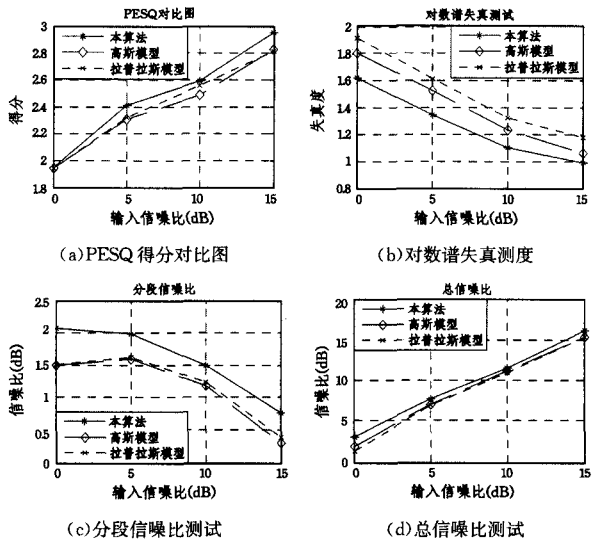
具体的训练步骤如下：

(1)根据式(2)和式(3)计算所有帧的先验信噪比 $\hat{\xi}(m)$ 和后验信噪比 $\gamma(m)$。噪声的估计采用 IMCRA 算法(式(5)－式(16))。所有的点 $(\hat{\xi},\gamma)$ 都有与之对应的 (A^2,R^2)。

(2) 在每一个格内采用对数谱最小均方差准则(式(18)、式(19))计算出落在该格内的最佳增益值 $\tilde{G}$，并存储在表格中。

5 实验结果

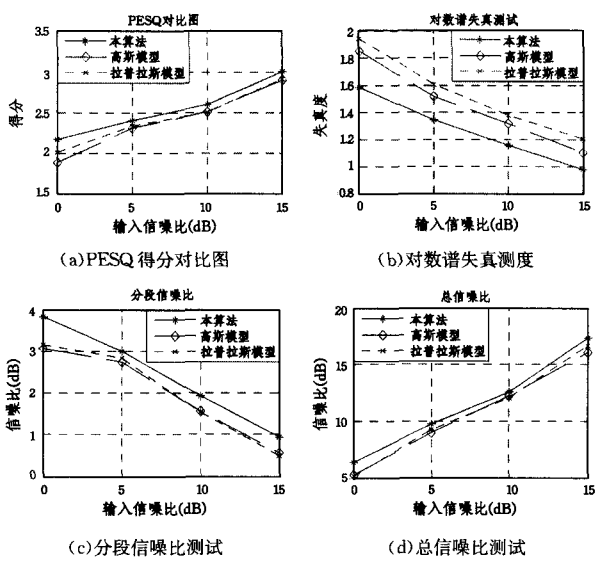
为验证基于增益字典查询的增强算法的性能，本文使用分段信噪比、对数谱失真、总信噪比和平均意见得分(PESQ)准则将该算法和基于拉普拉斯分布、基于高斯分布的增强算法做对比。测试语音采用标准语音库 NOIZEUS 中的语音，输入信噪比分别为 0dB、5dB、10dB、15dB，噪声分别为 babble 噪声、car 噪声和 street 噪声。仿真结果如图 3－图 5 所示。



(a)PESQ 得分对比图　(b)对数谱失真测度

(c)分段信噪比测试　(d)总信噪比测试

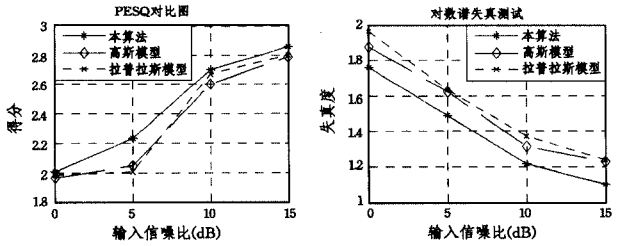
图 3　babble 噪声环境下的测试对比图



(a)PESQ 得分对比图　(b)对数谱失真测度

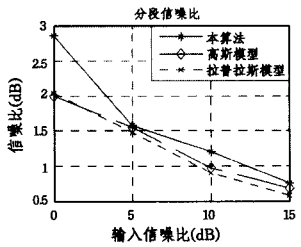
(c)分段信噪比测试　(d)总信噪比测试

图 4　car 噪声环境下的测试对比图

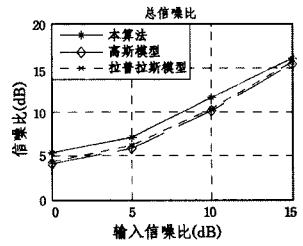


(a)PESQ得分对比图

(b)对数谱失真测度



(c)分段信噪比测试



(d)总信噪比测试

图 5　street 噪声环境下的测试对比图

从图 3—图 5 可以看出，在非稳态噪声 babble 噪声和 street 噪声条件下，基于增益字典查询的算法无论是主观评价的 PESQ 测试还是客观评价的分段信噪比、对数谱失真和总的信噪比测试都要优于基于高斯分布和基于拉普拉斯分布的算法。在平稳噪声 car 噪声环境下，它同样优于另外两种算法，且在平稳噪声环境下的效果相比于非平稳噪声环境下更好。因此本算法在平稳噪声条件和非平稳噪声条件下均可适用。图 6 为语音增强后的波形对比图，从图中明显可以看出，该算法能够很好地抑制音乐噪声和残留背景噪声。

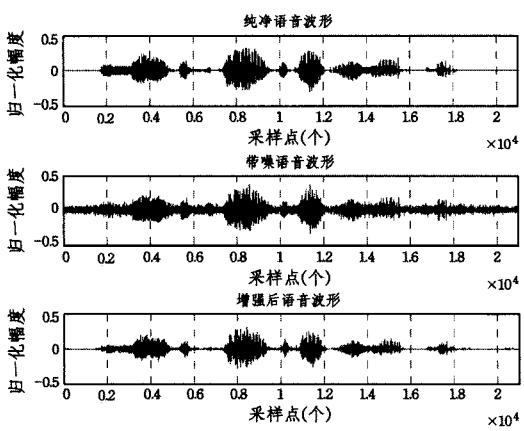


图 6　语音波形对比图

结束语　本文针对传统先验信噪比估计算法 DD 算法进行了改进，有效解决了 DD 算法存在一帧时延的问题。本文提出的基于增益的字典查询的模型，通过提前对噪声库进行训练得到一个增益的字典。采用增益查表的方法不仅有效提高了算法的效率，同时增强的效果无论是在平稳噪声还是非平稳噪声环境下都要比目前使用较多的基于高斯模型和拉普拉斯模型的算法效果更好，而且训练后的增益字典是一个 60×71 的矩阵，所占用的内存并不大，因此该算法同样具有较好的实用性。

参 考 文 献

[1] Loizuo P C. Speech Enhancement Theory and Practice[M]. CRC Press, 2007: 337-377

[2] 晏光华. 一种基于 MMSE-LSA 和 VAD 的语音增强算法[J]. 移动通信, 2014, 10(4): 59-62
Yan Guang-hua. A Speech Enhancement Algorithm Based on MMSE-LSA and VAD[J]. Mobile Communication, 2014, 10(4): 59-62

[3] 陈立伟, 王文姝, 袁頔. 自适应高斯混合模型语音增强方法[J]. 应用科技, 2009, 36(7): 11-15
Chen Li-wei, Wang Wen-shu, Yuan Di. A Speech Enhancement Method Based on Adaptive Gaussian Mixture Model[J]. Applied Science and Technology, 2009, 36(7): 11-15

[4] 梁岩, 鲍长春, 夏丙寅, 等. 基于高斯混合模型的压缩域语音增强方法[J]. 电子学报, 2012, 40(10): 2031-2038
Liang Yan, Bao Chang-chun, Xia Bing-yin, et al. Compressed Domain Speech Enhancement Based on Gaussian Mixture Model [J]. Acta Electronica Sinica, 2012, 40(10): 2031-2038

[5] 邹霞, 陈亮, 张雄伟. 基于 Gamma 语音模型的语音增强算法[J]. 通信学报, 2006, 27(10): 118-123
Zou Xia, Chen Liang, Zhang Xiong-wei. Speech enhancement with Gamma speech modeling[J]. Journal on Communications, 2006, 27(10): 118-123

[6] 赵改华, 周彬, 张雄伟. 修正的基于广义 Gamma 语音模型语音增强算法[J]. 计算机工程与应用, 2014, 50(18): 230-235
Zhao Gai-hua, Zhou Bin, Zhang Xiong-wei. Modified speech enhancement algorithm under signal presence probability with generalized Gamma speech model[J]. Computer Engineering and Applications, 2014, 50(18): 230-235

[7] 周彬, 邹霞, 张雄伟. 基于多元 Laplace 语音模型的语音增强算法[J]. 电子与信息学报, 2012, 84(7): 1562-1567
Zhou Bin, Zou Xia, Zhang Xiong-wei. Speech Enhancement with Multivariate Laplace Speech Model[J]. Journal of Electronics & Information Technology, 2012, 84(7): 1562-1567

[8] Wu D L, Zhu W P, Swamy M N S. Noise Spectrum Estimation with Improved Minimum Controlled Recursive Averaging based on Speech Enhancement Residue[C]// IEEE International Midwest Symposium on Circuits and Systems (MWSCAS). Boise, USA, 2012: 945-951

[9] 杨波, 王新房. 基于非因果先验信噪比估计的语音增强改进算法[J]. 计算机系统应用, 2012, 21(7): 200-202
Yang Bo, Wang Xin-fang. Improved Speech Enhancement Algorithm Based on Noncausal a Priori SNR Estimation[J]. Computer Systems & Applications, 2012, 21(7): 200-202

[10] Yong P C, Nordholm S, Dam H H. Trade-off Evaluation for Speech Enhancement Algorithms with Respect to The a Priori SNR Estimation[C]// IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Kyoto, Japan, 2012: 4657-4660

[11] Ekelens J, Jensen J, Heusdens R. A Data-Driven Approach to Optimizing Spectral Speech Enhancement Methods for Various Error Criteria[J]. Speech Communication, 2007, 49(5): 530-541