第 19 卷　第 23 期　2019 年 8 月
1671—1815(2019)0023-0126-05

科　学　技　术　与　工　程
Science Technology and Engineering

Vol. 19　No. 23　Aug. 2019
ⓒ 2019　Sci. Tech. Engrg.

引用格式:罗　瀛, 曾庆宁, 龙　超. 基于软语音存在概率的噪声互功率谱估计[J]. 科学技术与工程, 2019, 19(23): 126-130
Luo Ying, Zeng Qingning, Long Chao. Noise cross power spectral density estimation based on soft speech presence probability[J]. Science Technology and Engineering, 2019, 19(23): 126-130

电子技术、通信技术

基于软语音存在概率的噪声互功率谱估计

罗　瀛　曾庆宁*　龙　超
(桂林电子科技大学认知无线电与信息处理教育部重点实验室,桂林 541004)

摘　要　针对目前已有的改进相干滤波语音增强系统中噪声互功率谱估计方法运算效率低、准确性不足的问题,提出一种基于软语音存在概率的噪声互功率谱估计方法。通过计算语音信号的固定先验软语音存在概率代替语音活动检测器,得到噪声互功率谱的无偏估计值,以改善估计的准确性,同时还可避免复杂的偏差补偿值计算,使算法计算量得以减小。仿真实验结果表明,所提出的噪声互功率谱估计方法在应用于改进相干滤波语音增强系统时有更好的感知语音质量评价得分,且运算用时更短。

关键词　相干滤波　　无偏估计　　软语音存在概率　　噪声互功率谱密度
中图法分类号　TN912.35;　　**文献标志码**　A

与传统语音增强技术相比较,麦克风阵列语音增强技术[1]能获得更高质量的语音输出。但麦克风阵列增强往往依赖阵元的数量,这并不利于当今设备小型化的发展趋势。二元麦克风阵列[2]具有较小的体积,同时具备比传统增强技术更好的语音增强效果,现已广泛应用于助听器[3]、人工耳蜗[4]等方面。

相干滤波算法[5](coherence filter,CF)是目前应用较为广泛的二元麦克风增强技术,但其存在的问题是当麦克风之间距离较近时,两个通道产生的噪声具有相关性,此时相干滤波算法不能很好地滤除噪声。为解决这一问题,Jeannes 等[6]提出一种利用带噪语音互功率谱密度(cross power spectral density, CPSD)减去噪声 CPSD 估计得到纯净语音 CPSD 的方法,改进了相干滤波器。噪声的 CPSD 估计变为研究的主要问题。

为了估计噪声的 CPSD,Rahmani 等[7]提出了基于最小统计(minimum statics, MS)的噪声 CPSD 估计方法。之后 Fathi 等[8]提出改进最小追踪(improved minimum tracking, IMT)噪声 CPSD 估计方法。张正文等[9]则在 IMT 算法的基础上提出基于递归最小追踪(recursive minimum tracking,RMT)的噪声 CPSD 估计方法。RMT 算法与 MS 算法相比,对噪声 CPSD 的估计更为准确,但同时由于 RMT 算法的复杂性,其运算时间相对 MS 及 IMT 算法更长。

现提出一种基于软语音存在概率(soft speech presence probability,SPP)的最小均方误差(minimum mean-square error,MMSE)噪声 CPSD 估计方法。与传统的语音活动检测(voice activity detector,VAD)方法进行噪声更新估计相比,使用 SPP-MMSE 算法,无需进行偏差补偿计算就能获得无偏噪声功率谱估计,能更为快速准确对噪声进行跟踪。将该算法应用于改进相干滤波系统的噪声 CPSD 估计时,不仅能提高系统噪声抑制效果,还能大幅度的减少系统的计算复杂度。

1　软语音存在概率

将带噪语音信号经过分帧和离散傅里叶变换后得到信号的频谱信息。以 S 和 N 分别为纯净语音和噪声的频域表示。假设语音信号与噪声信号不相关,带噪语音信号频域 Y 由表达式 $Y_k(l)=S_k(l)+N_k(l)$ 给出,其中 k 表示频率点,l 表示帧数。同时给出带噪语音信号和噪声信号的频谱极坐标表达 $Y=R_e^{j\theta}$ 和 $N=De^{j\Delta}$,R 和 D 是带噪语音信号和噪声信号在极坐标下的极径,θ 及 Δ 是带噪语音信号和噪声信号在极

2018 年 12 月 28 日收到　国家自然科学基金(61461011)、广西自然科学重点基金(2016GXNSFDA380018)和桂林电子科技大学研究生科研创新项目(2017YJCX16,2017YJCX20)资助
第一作者简介:罗　瀛(1992—),男,汉族,硕士研究生。E-mail:362012643@qq.com。
*通信作者简介:曾庆宁(1963—),男,汉族,博士,教授。E-mail:1586477188@qq.com。

坐标下的极角。先验信噪比(signal to noise ratio, SNR)与后验信噪比分别定义为 $\xi=\sigma_S^2/\sigma_N^2$ 和 $\gamma=|Y|^2/\sigma_N^2$,其中 σ_S 与 σ_N 分别表示 S 和 N 的方差。

根据文献[10]的理论,假设语音信号 s、噪声 n 和带噪语音 y 的频谱均服从复高斯分布,则噪声谱系数分布 $P_N(n)$、纯净语音谱系数分布 $P_S(s)$ 及带噪语音谱系数分布 $P_Y(y)$ 可表示为

$$P_N(n)=\frac{1}{\pi\sigma_N^2}\exp\frac{|n|^2}{\sigma_N^2} \tag{1}$$

$$P_S(s)=\frac{1}{\pi\sigma_S^2}\exp\frac{|s|^2}{\sigma_S^2} \tag{2}$$

$$P_Y(y)=\frac{1}{\pi\sigma_N^2(1+\xi)}\exp\frac{|y|^2}{\sigma_N^2(1+\xi)} \tag{3}$$

噪声谱系数的分布在极坐标的形式下可以写为

$$P_{D,\Delta}(d,\delta)=\frac{d}{\pi\sigma_N^2}\exp\frac{d^2}{\sigma_N^2} \tag{4}$$

式(4)中,d 和 δ 分别是噪声谱系数在极坐标系下的极径变量和极角变量。

根据语音信号和噪声信号互不相关,可得:

$$P_{Y|D,\Delta}(y|d,\delta)=\frac{1}{\pi\sigma_S^2}\exp\frac{2dr\cos(\delta-\theta)-r^2-d^2}{\sigma_S^2} \tag{5}$$

式(5)中,r 和 θ 分别是带噪语音谱系数在极坐标系下的极径变量和极角变量。

为了估计噪声的功率谱,通过贝叶斯准则得到噪声功率谱平方系数 N^2 的条件期望 $\mathrm{E}\{|N|^2|Y\}$ 计算公式:

$$\mathrm{E}\{|N|^2|y\}=\frac{\int_0^{+\infty}\int_0^{2\pi}d^2P_{Y|D,\Delta}(y|d,\delta)P_{D,\Delta}(d,\delta)\mathrm{d}\delta\mathrm{d}d}{\int_0^{+\infty}\int_0^{2\pi}P_{Y|D,\Delta}(y|n,\delta)P_{D,\Delta}(d,\delta)\mathrm{d}\delta\mathrm{d}d} \tag{6}$$

将式(4)、式(5)代入式(6)并利用文献[12]中的公式,得到:

$$\mathrm{E}\{|N|^2|y\}=\frac{1}{(1+\hat{\xi})^2}|y|^2+\frac{\hat{\xi}}{(1+\hat{\xi})}\hat{\sigma}_N^2 \tag{7}$$

这样,$E\{|N|^2|Y\}$ 的计算就变为了先验 SNR$\hat{\xi}$ 和噪声功率谱 $\hat{\sigma}_N^2$ 的估计问题。在噪声功率的估计中,一种常见的假设是噪声信号比语音信号更为平稳。也就是假定相邻信号段的噪声功率之间存在相关性,那么在式(7)中就能使用前一帧的噪声功率估计来代替当前的噪声估计,即 $\hat{\sigma}_N^2=\hat{\sigma}_N^2(l-1)$。由于语音频谱系数在连续的段之间通常表现出较大程度的波动,因此很难估计先验信噪比。在文献[10]中使用了最大似然估计(maximum-likelihood, ML)对 $\hat{\xi}$ 进行估值。最后通过参数为 $\alpha_{\mathrm{pow}}=0.8$ 的递归平滑得到噪声的功率谱估计:

$$\hat{\sigma}_N^2(l)=\alpha_{\mathrm{pow}}\hat{\sigma}_N^2(l-1)+(1-\alpha_{\mathrm{pow}})E[|N|^2|y(l)] \tag{8}$$

文献[10]提出的 VAD-MMSE 算法中对噪声更新的判断是硬性的,即:

$$\mathrm{E}[|N(l)|^2|y(l)]=\begin{cases}\hat{\sigma}_N^2(l-1), & |y(l)|^2\geqslant\hat{\sigma}_N^2(l-1)\\ |y(l)|^2, & |y(l)|^2\leqslant\hat{\sigma}_N^2(l-1)\end{cases} \tag{9}$$

这样的估计是有偏的,因此文献[10]利用补偿因子 B 对结果进行无偏补偿,$\hat{\sigma}_N^2(l)=E(|N|^2|y)B$,补偿因子在一定程度上提高了算法的准确性,但其计算较为复杂,降低了算法的计算速度。文献[11]提出一种基于 SPP 的无偏 MMSE 噪声功率谱估计,与传统的 VAD-MMSE 算法相比,SSP-MMSE 算法不需要进行硬性的语音存在判决,因此也不需要设置并计算无偏估计补偿因子。这样便提高了算法的计算速率,同时也不影响噪声估计的准确性。MMSE 估计器的表达式为

$$\mathrm{E}(|N|^2|y)=P(H_0|y)E(|N|^2|y,H_0)+P(H_1|y)E(|N|^2|y,H_1) \tag{10}$$

式(10)中,H_0 指语音存在,H_1 指语音不存在。与式(4)类似的,假设带噪语音与噪声信号的离散傅里叶变换系数都服从复高斯分布,利用贝叶斯准则,能得到后验 SPP 表达式:

$$P(H_1|y)=\frac{P(H_1)P_{Y|H_1(y)}}{P(H_0)P_{Y|H_0(y)}+P(H_1)P_{Y|H_1(y)}} \tag{11}$$

为了计算后验 SPP,需要先计算先验概率 $P(H_1)=1-P(H_0)$及语音存在似然函数 $P_{Y|H_1(y)}$ 和语言音不存在似然函数 $P_{Y|H_0(y)}$。在未经观测的情况下,假设某一时频点是否包含语音的可能性相等,可选择一种最坏假设,即先验概率 $P(H_1)=P(H_0)=0.5$。先验概率值是与观测无关的固定值。

在式(11)中的两个似然函数 $P_{Y|H_1(y)}$ 和 $P_{Y|H_0(y)}$ 反映了所观测到的语音信号 y 在语音存在和语音不存在这两种情况中的可能性大小。由于假设参数服从复高斯分布,语音不存在时的似然函数表达式为

$$P_{Y|H_0(y)}=\frac{1}{\hat{\sigma}_N^2\pi}\exp-\frac{|y|^2}{\hat{\sigma}_N^2} \tag{12}$$

语音存在时的似然函数表达式为

$$P_{Y|H_1(y)}=\frac{1}{\hat{\sigma}_N^2(1+\xi_{H_1})\pi}\exp-\frac{|y|^2}{\hat{\sigma}_N^2(1+\xi_{H_1})} \tag{13}$$

需注意要将 ξ 和 ξ_{H_1} 的概念区分。ξ 代表真实的局部 SNR,而 ξ_{H_1} 则是语音存在似然函数的一个参数,ξ_{H_1} 可以认为是一个长时间的固定先验 SNR 而非短时的局部 SNR。式(12)、式(13)在计算上的区别就在于参数 ξ_{H_1}。因此需要选择一个最优的固定先

验 SNRξ_{H_1},保证语音存在和语音不存在两个似然函数的不同,使得在语音不存在的情况下使后验 SPP 的估计值接近于 0。

将式(12)、式(13)代入式(11)就能得到后验 SPP 的表达式:

$$P(H_1|y)=\left[1+\frac{P(H_0)}{P(H_1)}(1+\xi_{H_1})\mathrm{e}^{-\frac{|y|^2}{\hat{\sigma}_N^2}\frac{\xi_{H_1}}{1+\xi_{H_1}}}\right]^{-1} \tag{14}$$

在这里,假设 $P(H_0)=P(H_1)$。同时,在式(12)~式(14)中的噪声功率谱谱估计是当前帧的前一帧的噪声功率谱估计,即 $\hat{\sigma}_N^2=\hat{\sigma}_N^2(l-1)$。

通过式(14)能找到后验 SNR$\hat{\gamma}=|y|^2/\hat{\sigma}_N^2$ 和参数 ξ_{H_1}与 $P(H_1|y)$之间的关系为

$$\hat{\gamma}=\lg\left[\frac{1+\xi_{H_1}}{P(H_1|y)-1}\right]\frac{1+\xi_{H_1}}{\xi_{H_1}} \tag{15}$$

这里需要计算噪声先验 SNR 有限长的最大似然(ML)估计 $\hat{\xi}^{\mathrm{ML}}=\hat{\gamma}-1$,再利用条件 $\hat{\xi}(l)=\max[0,\hat{\xi}^{\mathrm{ML}}(l)]$以及式(7),对噪声更新进行决断。

假设固定先验 SNR 为 $10\lg\xi_{H_1}=15$ dB,语音存在参数 $P(H_1|y)>0.075$,后验 SNR$\hat{\gamma}>1$,先验 SNR 有限长的最大似然(ML)估计 $\hat{\xi}^{\mathrm{ML}}=\hat{\gamma}-1>0$。语音存在下的参数最优估计量参数表示为

$$\mathrm{E}(|N|^2|y,\hat{\xi},H_1)=E(|N|^2|y,\hat{\xi}^{ml}=\hat{\gamma}-1)=\hat{\sigma}_N^2 \tag{16}$$

语音不存在下的参数最优估计量参数表示为

$$\mathrm{E}(|N|^2|y,H_0)=E(|N|^2|n)=|y|^2 \tag{17}$$

将式(16)、式(17)代入式(10),这样在语音存在不确定的情况下,MMSE 估计就变成了一种语音观测值$|y|^2$ 和噪声功率谱估计 $\hat{\sigma}_N^2$ 的软权重计算:

$$\mathrm{E}(|N|^2|y)=P(H_0|y)|y|^2+P(H_1|y)\hat{\sigma}_N^2 \tag{18}$$

式(18)中:$P(H_0|y)=1\text{-}P(H_1|y)$,噪声功率谱估计 $\hat{\sigma}_N^2$ 采用前一帧的功率谱估计,即 $\hat{\sigma}_N^2=\hat{\sigma}_N^2(l-1)$。然后通过式(8)递归平滑计算当前帧的噪声谱功率估计。

当噪声功率谱估计值 $\hat{\sigma}_N^2$ 低于真实噪声功率谱 σ_N^2 时,通过式(14)计算出的后验 SPP 就会出现过估计,此时对噪声频谱的跟踪效率将会变低。如果考虑到极端的情况,当 $\hat{\sigma}_N^2$ 的值严重低于 σ_N^2 时,后验 SPP 的参数值会趋近于 1,即 $P(H_1|y)=1$。此时噪声的更新将会停止。为防止这一情况出现,首先需要对后验 SPP 的值进行递归平滑处理:

$$\bar{P}(l)=0.9\bar{P}(l)+0.1P(H_1|y) \tag{19}$$

当这一平滑值大于 0.99 时,可认为此时后验 SPP 的更新已经停滞,为了让后验 SPP 的参数值 $P(H_1|y)$小于 0.99,此时有:

$$P(H_1|y)\leftarrow\begin{cases}\min\{0.99,P[H_1|y(l)]\}, & \bar{P}(l)>0.99\\ P(H_1|y), & \text{else}\end{cases} \tag{20}$$

最后是寻找最优的固定先验 SNRξ_{H_1},这需要通过最小化误差概率来获得。文献[11]通过最小化误差概率实验计算出的最优固定先验 SNR 是 $10\lg\xi_{H_1}=15$ dB。这个固定值将应用于每一帧噪声功率谱估计的计算中。通过固定先验信噪比计算的软语音存在概率在应用于噪声功率谱估计时就能获得无偏估计值。

2 基于噪声 CPSD 估计的改进相干滤波算法

图 1 是基于噪声 CPSD 估计的改进相干滤波算法原理图,从图 1 中看出在二元阵列模型下,麦克风所接收的信号为

$$y_i=s+n_i \tag{21}$$

式(21)中:y_i、s 和 n_i 分别是带噪语音信号、纯净语音信号及噪声信号。其经过傅里叶变换后的频域表达式为

$$Y_i(k,l)=S(k,l)+N_i(k,l) \tag{22}$$

式(22)中:$Y_i(k,l)$、$S(k,l)$和 $N_i(k,l)$分别是带噪语音信号、纯净语音信号及噪声信号的频域变换表达。由 Jeannes 提出的改进相干滤波器算法,其传递函数表达式为

$$H(k,l)=\frac{|P_{Y_1Y_2}(k,l)|-|P_{N_1N_2}(k,l)|}{\sqrt{P_{Y_1Y_1}(k,l)P_{Y_2Y_2}(k,l)}} \tag{23}$$

式(23)中:$P_{Y_1Y_2}(k,l)$是带噪语音的 CPSD 密度;$P_{N_1N_2}(k,l)$是噪声信号的 CPSD 密度;$P_{Y_1Y_1}(k,l)$和 $P_{Y_2Y_2}(k,l)$和带噪语音的自功率谱密度。传递函数中只有噪声 CPSD 是非观测值,因此,对噪声的 CPSD 估计是研究的重点。

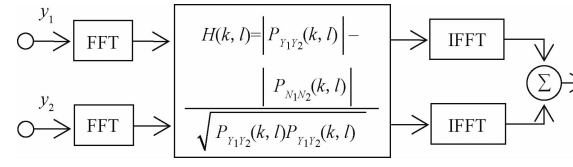

图 1 基于噪声互功率谱密度估计的改进相干滤波算法原理框图

Fig. 1 Block diagram of improved coherent filtering based on noise cross power spectral density estimation

3 基于软语音存在概率的噪声互功率谱估计

利用软语音存在概率对改进相干滤波算法中的噪声互功率谱进行估计,令带噪语音信号的互功率后验信噪比 λ 为

$$\lambda=\frac{|P_{Y_1Y_2}(k,l)|}{|P_{N_1N_2}(k,l)|} \tag{24}$$

已经证明基于软语音存在概率的噪声功率谱最小均方误差估计是一种无偏估计,因此当满足噪声更新条件 $\lambda-1>0$ 时,噪声的 CPSD 无偏估计可由式(10)推导得

$$\begin{aligned}E[\,|P_{N_1N_2}(k,l)|\,|y_1,y_2\,]=&\\ P(H_0|y_1,y_2)|P_{Y_1Y_2}(k,l)|+&\\ P(H_1|y_1,y_2)|P_{N_1N_2}(k,l)|&\end{aligned}\tag{25}$$

由于该噪声的 CPSD 估计值是无偏估计,因此对该估计结果不需要进行偏差补偿计算。在式(25)中$P(H_0|y_1,y_2)=1-P(H_1|y_1,y_2)$,所以只需要计算后验 SPP 参数 $P(H_1|y_1,y_2)$,就能得到噪声的 CPSD 无偏估计。后验 SPP 参数 $P(H_1|y_1,y_2)$ 的参数值计算方式可由式(8)推导得到:

$$P(H_1|y_1,y_2)=\left[1+\frac{P(H_0)}{P(H_1)}(1+\xi'_{H_1})\mathrm{e}^{-\frac{|P_{Y_1Y_2}(k,l)|\ \xi'_{H_1}}{|P_{N_1N_2}(k,l)|1+\xi'_{H_1}}-1}\right]\tag{26}$$

假定语音存在或不存在的概率相等,即 $P(H_1)=P(H_0)=0.5$。在此条件下最小化误差概率实验计算得到的软语音存在概率固定先验信噪比 ξ'_{H_1} 的值为 $10\lg\xi'_{H_1}=15$ dB。把这一参数值应用于每一帧的噪声 CPSD 计算,就可获得噪声 CPSD 的无偏估计。同时还利用式(19)、式(20)类似的方式对 $P(H_1|y_1,y_2)$ 进行平滑递归处理以确保参数的更新不会停滞:

$$\bar{P}(l)'=0.9\bar{P}(l)'+0.1P(H_1|y_1,y_2)\tag{27}$$

$$P(H_1|y_1,y_2)\leftarrow\begin{cases}\min\{0.99,P[H_1|y_1(l),y_2(l)]\}, & \bar{P}'(l)>0.99\\ P(H_1|y_1,y_2), & \text{else}\end{cases}\tag{28}$$

最后通过对 SPP-MMSE 条件下计算出的噪声 CPSD 估计进行平滑处理,得到最终的噪声 CPSD 估计:

$$\begin{aligned}|P_{N_1N_2}(k,l)|=&\alpha_{pow}|P_{N_1N_2}(k,l-1)|+(1-\alpha_{\mathrm{pow}})\\ &\mathrm{E}(|P_{N_1N_2}(k,l)|\,|y_1,y_2)\end{aligned}\tag{29}$$

式(29)中:$|P_{N_1N_2}(k,l-1)|$ 表示前一帧的噪声 CPSD 估计。

4　实验分析

实验使用 m-audio 多路音频采集器进行数据采集,二元麦克风阵列两个阵元之间的距离设置为 2 cm 来模拟实际应用中的小型语音设备,噪声源使用 Noisex—92 数据库中的部分噪声。语音录制与噪声录制在同环境中进行。共录制 8 段 3 s 左右的二元阵列纯净语音段作为参考,同时采用同环境下录制的 f16 噪声、babble 噪声、white 噪声,按照不同的信噪比(−5、−0、5、10 dB)混入纯净语音段,生成带噪的二元阵列语音文件。实验在 MATLAB 环境下对 MS 算法、RMT 算法、本文算法应用于改进相干滤波系统中的噪声抑制效果进行仿真对比实验。经过系统增强后的语音采用感知语音质量评价(perceptual evaluation of speech quality,PESQ)得分进行客观语音质量评价。PESQ 得分与主观评测方法平均意见得分(mean opinion score,MOS)的相关程度较高,模拟主观评价的同时又兼具良好的客观性,可以较为真实地反映消噪系统性能的优劣。

表 1 给出在相同噪声环境、相同信噪比条件下应用 MS 算法、RMT 算法、本文算法的改进相干滤波系统对带噪语音文件进行增强处理后的结果进行 PESQ 算法评测所得的平均得分。

表 1　不同算法在单一噪声环境下的 PESQ 平均得分

Table 1　PESQ scores of different algorithms in a single noise environment

噪声类型	SNR/dB	PESQ 得分		
		MS	RMT	本文算法
f16	−5	1.05	1.10	1.15
	0	1.24	1.27	1.35
	5	1.38	1.49	1.61
	10	1.61	1.73	1.87
babble	−5	0.83	0.88	0.95
	0	1.01	1.06	1.12
	5	1.16	1.18	1.29
	10	1.42	1.46	1.58
White	−5	0.96	1.03	1.12
	0	1.16	1.21	1.29
	5	1.35	1.43	1.51
	10	1.62	1.69	1.81

通过表 1 三种算法 PESQ 平均得分情况对比,能看出,应用本文算法进行噪声 CPSD 估计时,较 MS 算法及 RMT 算法改进相干滤波系统的消噪效果均有所提高。在同一噪声环境,同一信噪比的条件下应用本文算法的消噪系统的 PESQ 得分的平均值较 MS 算法提高了 0.2 分左右,较 RMT 算法提高了 0.1 分左右。这说明应用本文算法的改进相干滤波系统具有更强的消噪效果。

实验还对应用 MS 算法、RMT 算法、本文算法时改进相干滤波系统对语音段进行增强的时间进行了统计。表 2 给出了在 MATLAB 仿真环境下,应用不同算法进行噪声 CPSD 估计的改进相干滤波系统处理时长为 3 s 左右语音段的用时。

表 2　不同算法计算用时

Table 2　Computation times of different algorithms

算法名称	MS	RMT	本文算法
时间/s	0.83	2.12	0.19

从表 2 的统计结果能看出,应用了本文算法的改进相干滤波系统在对语段进行增强处理时有更快的处理速度,即验证了本文算法在对噪声 CPSD 进行估计时具有更小的计算量。

5 结论

通过模拟小型语音设备的仿真实验,证明了本研究所提出的基于软语音存在概率的噪声互功率谱估计算法在应用于改进相干滤波系统时较目前已有的算法有更好的消噪效果。通过与 MS 及 RMT 噪声 CPSD 算法的运算处理用时进行比较,证明了本文算法具有更小的计算量。因此本文算法在应用于改进相干滤波系统时,具有更好的实时性。本文算法兼具准确性高、实时性强的特点,能更好地满足一些实际场景(助听器、人工耳蜗、智能手机等)的应用需求。

参 考 文 献

1 林静然. 基于麦克风阵列的语音增强算法研究[D]. 成都: 电子科技大学, 2007
Lin Jingran. On speech enhancement based on microphone arrays [D]. Chengdu: University of Electronic Science and Technology of China, 2007

2 杨立春, 钱沄涛, 王文宏. 基于零陷谱减的 GSC 二元麦克风小阵列语音增强算法[J]. 浙江大学学报(工学版), 2013, 47(8): 1493-1499
Yang Lichun, Qian Quntao, Wang Wenhong. A GSC algorithm based on null spectral subtraction for dual small microphone array speech enhancement[J]. Journal of Zhejiang University (Engineering Science), 2013, 47(8): 1493-1499

3 Yousefian N, Loizou P C, Hansen J H. A coherence-based noise reduction algorithm for binaural hearing aids[J]. Speech Communication, 2014, 58(1): 101-110

4 Kallel F, Frikha M, Ghorbel M, et al. Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant[J]. Applied Acoustics, 2012, 73(1): 12-20

5 Allen J B, Berkley D A, Blauert J. Multimicrophone signal processing technique to remove room reverberation from speech signals[J]. Journal of the Acoustical Society of America, 1977, 62(4): 912, 915

6 Jeannes W L B, Scalart P, Faucon G, et al. Combined noise and echo reduction in hands-free systems: a survey[J]. IEEE Transactions on Speech & Audio Processing, 2001, 9(8): 808-820

7 Mohsen R, Ahmad A, Beghdad A, et al. A noise cross PSD estimator for dual-microphone speech enhancement based on minimum statistics[J]. Journal of Zhejiang University Science A, 2009, 10(6): 805-809

8 Kallel F, Ghorbel M, Frikha M, et al. A noise cross PSD estimator based on improved minimum statistics method for two-microphone speech enhancement dedicated to a bilateral cochlear implant[J]. Applied Acoustics, 2012, 73(3): 256-264

9 张正文, 赵晓晴, 尹 波. 基于递归最小追踪的噪声互功率谱估计算法[J]. 科学技术与工程, 2016, 16(10):164-166, 173
Zhang Zhengwen, Zhao Xiaoqing, Yin Bo. A noise cross power spectral density estimation method based on recursive minimum tracking [J]. Science Technology and Engineering, 2016, 16(10): 164-166, 173

10 Richard C H, Richard H, Jesper J. MMSE based noise PSD tracking with low complexity[J]. IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, 23(3): 4266-4269

11 Gerkmann T, Hendriks R C. Unbiased MMSE-Based noise power estimation with low complexity and low tracking delay[J]. IEEE Transactions on Audio, Speech and Language Processing, 2012, 20(4): 1383-1393

12 Gradshteyn I S, Ryzhik I M. Table of integrals series and products[M]. 6th ed. San Diego: Academic Press, 2000

Noise Cross Power Spectral Density Estimation Based on Soft Speech Presence Probability

LUO Ying, ZENG Qing-ning*, LONG Chao

(Key Laboratory of Cognitive Radio and Information Processing of Ministry of Education, Guilin University of Electronic Technology, Guilin 541004, China)

[**Abstract**] According to solve the problem of slow operation efficiency and inaccuracy of existing noise cross-power spectrum estimation methods in the improved coherent filtering speech enhancement system, a noise cross-power spectrum estimation method based on the soft speech presence probability is proposed. By calculating the soft speech presence probability with fixed priors instead of voice activity detector to obtain the unbiased estimation of the noise cross power spectrum to improve the accuracy of the estimation, and also avoid the calculation of the complicated bias compensation, so that the algorithm calculates can be reduced. The experimental results show that the proposed noise cross-power spectrum estimation method has better perceptual evaluation of speech quality score and shorter computation time when applied to the improved coherent filtering speech enhancement system.

[**Key words**] coherent filtering unbiased estimation soft speech presence probability cross power spectral density