第 18 卷　第 2 期　2018 年 1 月
1671—1815(2018)002-0093-06

科 学 技 术 与 工 程
Science Technology and Engineering

Vol. 18　No. 2　Jan. 2018
ⓒ 2018　Sci. Tech. Engrg.

引用格式:肖　强, 曾庆宁, 王　瑶, 等. 多噪声环境的麦克风小阵语音增强[J]. 科学技术与工程, 2018, 18(2): 93—98
Xiao Qiang, Zeng Qingning, Wang Yao, et al. Small microphone array speech enhancement in multiple noise environments[J]. Science Technology and Engineering, 2018, 18(2): 93—98

电子技术、通信技术

多噪声环境的麦克风小阵语音增强

肖　强　曾庆宁　王　瑶　郑展恒
(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　针对在非平稳和多种噪声并存的语音增强算法抑制噪声能力有限的问题,提出基于最小跟踪噪声功率谱估计的相干滤波与广义旁瓣抵消的麦克风小阵语音增强算法。该方法先利用最小跟踪噪声功率谱估计的相干滤波抑制弱相关噪声,再结合广义旁瓣抵消与端点检测抑制强相关噪声。实验结果表明,方法更加有效地抑制噪声的影响;并提高了语音的可懂度。

关键词　相干滤波　　广义旁瓣抵消　　麦克风小阵
中图法分类号　TN912.35;　　**文献标志码**　A

在非平稳和多种噪声并存的实际环境中,噪声的干扰往往会严重影响目标语音的获取,致使获取的语音信号不是纯净的语音信号。在嘈杂环境下,单通道语音增强效果会急剧下降。与之相比,多通道语音增强在利用时域与频域信息的同时还利用了空域信息[1],从而达到抑制不同方位噪声的效果,弥补了单通道语音增强的不足。

Allen 等针对波束形成混响干扰抑制能力不足的问题,提出相干滤波器(coherence filter, CF)算法[2],其基本思想是:从信号的语谱和相关性出发,假设阵元间目标语音相关且噪声不相关,利用信号的互功率谱密度,抑制噪声信号包括混响干扰。但是相干滤波对强相关性噪声抑制较弱。Griffiths 和 Jim 提出广义旁瓣抵消(generalized sidelobe canceller, GSC)不仅能消除弱相关噪声;而且能消除强相关噪声[3]。但 GSC 对抑制弱相关噪声的能力较弱。在实际麦克风小阵中在存在强相关性噪声的同时还存在弱相关性噪声。

针对二元麦克风小阵模型,杨立春等提出基于相干性滤波器的广义旁瓣抵消器的麦克风小阵语音增强方法[4],该方法利用二元结构先通过噪声谱估计获得 CF 以抑制非相干或弱相干噪声,再通过 GSC 抑制强相关性噪声。之后提出基于零陷谱减的 GSC 二元麦克风小阵列语音增强算法[5]。以上几种算法对抑制强相关性与弱相关性噪声都有较好的效果。

在多噪声环境语音增强方面,Yousefian 等[6]提出基于相干函数的麦克风阵列语音增强。该算法实现在多种噪声且存在竞争声源的环境下的语音增强。Niwa 等[7]提出基于估计功率密度谱的后置滤波的语音增强方法。马金龙等[8]提出基于麦克风小阵的多噪声环境语音增强算法。该算法是基于相干滤波广义旁瓣抵消(CF-GSC) 的多噪声环境下语音增强算法。以上几种算法在多噪声环境下亦能取得较好的效果。

针对多噪声环境,本文提出基于最小跟踪噪声功率谱估计的相干滤波(minimum tracking noise power spectrum estimation coherence filter, MTCF)与广义旁瓣抵消的麦克风小阵语音增强算法(MTCF-GSC)。该方法先利用最小跟踪噪声功率谱估计的相干滤波抑制弱相关噪声,再利用广义旁瓣抵消与端点检测的结合以抑制强相关噪声。

1　信号模型

考虑一般麦克风阵列的信号模型,每个麦克风接收到的含噪语音信号表示为

$$y_i(n) = a_i * s_i(n) + v_i(n);\quad i = 1,2,\cdots,M \quad (1)$$

式(1)中, y_i 为含噪语音信号; a_i 为第 i 个麦克风的传输函数; s_i 为纯净语音信号; v_i 为噪声信号; $*$ 表示线性卷积; M 为麦克风数量且在本文中 $M = 6$。

2017 年 6 月 20 日收到　　国家自然科学基金(61461011)和教育部重点实验室 2016 年主任基金(CRKL160107)资助
第一作者简介:肖　强(1992—),硕士研究生。研究方向:语音信号处理。E-mail:qiangxiao0116@163.com

假设纯净语音与噪声互不相关,即 $E(sv^H)=0$,对式(1)两边进行短时傅里叶变换得出其频域形式:

$$Y_i(u,w)=A_iS_i(u,w)+V_i(u,w) \tag{2}$$

式(2)中,u 表示帧数;w 表示离散频率。

2 基于最小跟踪噪声功率估计的相干滤波

Jeannes 等提出从含噪语音的互功率谱减去噪声估计的互功率谱的方法从而改进相干滤波器的算法[9,10]。因此噪声谱估计直接影响改进相干滤波算法的噪声抑制效果。常用的噪声谱估计方法是语音端点检测技术(voice activity detection, VAD)。但 VAD 适用于平稳噪声环境下。因此,在实际环境中,VAD 技术估计的噪声谱不准确。

针对 VAD 算法对非平稳噪声的噪声谱估计不准确的问题,Martin 等学者提出最小跟踪噪声功率谱估计算法[11,12]。基于最小跟踪噪声功率谱估计的相干滤波原理框图如图 1 所示。

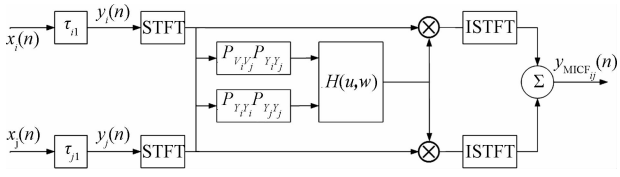

图 1 最小跟踪噪声功率谱估计的相干滤波原理框图

Fig. 1 Block diagram of coherent filtering for minimum tracking noise power spectrum estimation

图 1 中时延补偿通过式(3)获得。

$$\tau_{i1}=\frac{d_{i1}\cos\alpha}{c} \tag{3}$$

式(3)中,τ_{i1} 是以第 1 个麦克风为参考麦克风的第 i 个麦克风的时延;d_{i1} 是第 i 个麦克风到参考麦克风的距离;c 为声速;α 为声源的入射角。

根据 Jeannes 提出的改进相干滤波器算法可得传递函数表达式:

$$H(u,w)=\frac{|P_{Y_iY_j}(u,w)|-|P_{V_iV_j}(u,w)|}{\sqrt{P_{Y_iY_i}(u,w)P_{Y_jY_j}(u,w)}} \tag{4}$$

式(4)中,$P_{Y_iY_j}(u,w)$ 是含噪语音互功率谱;$P_{V_iV_j}(u,w)$ 噪声互功率谱;$P_{Y_iY_i}(u,w)$ 是含噪语音的功率谱;其中 $P_{Y_iY_j}(u,w)$ 用式(5)迭代实现。

$$P_{Y_iY_j}(u,w)=\beta P_{Y_iY_j}(u,w)+(1-\beta)Y_i(u,w)Y_j^*(u,w) \tag{5}$$

式(5)中,β 表示平滑因子,其值在[0,1]之间;$*$ 表示共轭。

针对存在噪声功率谱估计偏差与平滑参数而导致估计的噪声过大或过小的问题。通过偏差因子来补偿偏低的噪声功率谱估计值以及使用随时间和频率改变的平滑参数。

引入平滑因子 $\lambda(u,w)$ 对含噪语音进行噪声互功率谱估计,表达式为

$$\hat{P}_{V_iV_j}(u,w)=\lambda(u,w)\hat{P}_{V_iV_j}(u-1,w)+[1-\lambda(u,w)]Y_i(u,w)Y_j^*(u,w) \tag{6}$$

式(6)中,$\hat{P}_{V_iV_j}(u,w)$ 是含噪语音的噪声互功率谱估计值;平滑因子 $\lambda(u,w)$ 通过式(7)获得:

$$\lambda(u,w)=\begin{cases}\sqrt{1-\delta}, & \Gamma_{Y_iY_j}(u,w)\leqslant\delta\\ \sqrt{1-\Gamma_{Y_iY_j}(u,w)}, & \delta<\Gamma_{Y_iY_j}(u,w)<\gamma\\ \sqrt{1-\gamma}, & \Gamma_{Y_iY_j}(u,w)\geqslant\gamma\end{cases} \tag{7}$$

式(7)中,δ、γ 为常量,取值范围为(0,1);$\Gamma_{Y_iY_j}(u,w)$ 表示相干系数,通过式(8)、式(9)获得。

$$\hat{\Gamma}_{Y_iY_j}(u,w)=\frac{P_{Y_iY_j}(u,w)}{\sqrt{P_{Y_iY_i}(u,w)P_{Y_jY_j}(u,w)}} \tag{8}$$

调整相干滤波阈值,获得相干滤波系数:

$$\Gamma_{Y_iY_j}(u,w)=\begin{cases}0.001, & |\hat{\Gamma}_{Y_iY_j}(u,w)|<\lambda_1\\ 0.99, & |\hat{\Gamma}_{Y_iY_j}(u,w)|>\lambda_2\\ |\hat{\Gamma}_{Y_iY_j}(u,w)|, & \text{其他}\end{cases} \tag{9}$$

在局部搜索窗长内搜索噪声功率谱密度来计算最小噪声互功率谱,即 $P_{\min}(u,w)$。$P_{\min}(u,w)$ 的计算是基于在 $\{u,u-1,\cdots,u-D+1\}$ 上计算得到的 $\hat{P}_{V_iV_j}(u,w)$,其中 D 为搜索窗长,即过去帧的帧数。因此,$P_{\min}(u,w)$ 表示为

$$P_{\min}(u,w)=\min[\hat{P}_{V_iV_j}(u,w),\hat{P}_{V_iV_j}(u-1,w),\cdots,\hat{P}_{V_iV_j}(u-D+1,w),] \tag{10}$$

由于 $P_{\min}(u,w)$ 总会小于噪声平均值,利用无偏修正因子 $B_{\min}(u,w)$ 进行修正,而无偏修正因子与 D 有关,$B_{\min}(u,w)$ 的值会随着 D 的变化而变化。Martin 通过渐进求取的方法计算无偏修正因子 $B_{\min}(u,w)$ 的近似值。$B_{\min}(u,w)$ 表示为

$$B_{\min}(u,w)\approx 1+(D-1)\frac{2}{\hat{Q}_{\mathrm{eq}}(u,w)} \tag{11}$$

因此修正后的噪声互功率谱估计值为

$$P_{V_iV_j}(u,w)=P_{\min}(u,w)B_{\min}(u,w) \tag{12}$$

3 MTCF-GSC

本文的语音增强算法结构框图如图 2 所示。

图 2表明,算法分为两部分:最小跟踪噪声功率谱估计的相干滤波与广义旁瓣抵消。其主要思想是首先 MTCF 抑制非相关噪声及弱相关噪声,然后 GSC 结合 VAD 抑制强相关噪声。

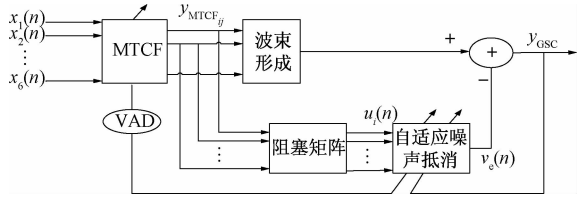

图 2　MTCF-GSC 原理框图

Fig. 2　MTCF-GSC block diagram

GSC 由三部分组成:固定波束形成(fixed beamforming,FBF)、阻塞矩阵(blocking matrix,BM)以及自适应干扰抵消器。GSC 抑制噪声的主要原理是通过固定波束形成产生期望信号、通过阻塞矩阵产生参考噪声信号;再利用参考噪声信号得到噪声估计值,从而抵消期望信号中的噪声。MTCF 已初步抑制麦克风接收到的信号中的部分噪声及混响,且它们所包含的语音信号基本相同,因此期望信号通过式(13)、式(14)获得。

$$\boldsymbol{y}_{\mathrm{MTCF}}(n) = [y_{\mathrm{MTCF}_{14}}(n), y_{\mathrm{MTCF}_{15}}(n), y_{\mathrm{MTCF}_{26}}(n), y_{\mathrm{MTCF}_{36}}(n), y_{\mathrm{MTCF}_{25}}(n), y_{\mathrm{MTCF}_{16}}(n)] \tag{13}$$

$$d(n) = \boldsymbol{y}_{\mathrm{MTCF}}(n)\boldsymbol{w}_{\mathrm{f}}^{\mathrm{T}} \tag{14}$$

式(13)中,$y_{\mathrm{MTCF}_{ij}}$ 表示第 i 个麦克风与第 j 个麦克风接收到的含噪语音经过 MTCF 后的增强语音。式(14)中,$\boldsymbol{w}_{\mathrm{f}} = [w_{\mathrm{f1}}, w_{\mathrm{f2}}, w_{\mathrm{f3}}, w_{\mathrm{f4}}, w_{\mathrm{f5}}, w_{\mathrm{f6}}]$ 表示固定波束形成加权权值,其权值都为 $1/M$。

参考噪声由 $\boldsymbol{y}_{\mathrm{MTCF}}(n)$ 通过阻塞矩阵获得,即

$$\boldsymbol{u}(n) = \boldsymbol{y}_{\mathrm{MTCF}}(n)\boldsymbol{B}^{\mathrm{T}} \tag{15}$$

式(15)中,$\boldsymbol{B}$ 为 5 × 6 维矩阵,且满足约束条件 $\boldsymbol{B}_1 = 0$。

利用归一化最小均方算法得到噪声估计值 $v_e(n)$,结合 VAD 算法仅在非语音段更新滤波器系数,减少语音失真,即:

$$v_e(n) = \boldsymbol{u}(n)\boldsymbol{w}^{\mathrm{T}}(n) \tag{16}$$

式(16)中,滤波器权值系数 $\boldsymbol{w}(n) = [w_1(n)\ \ w_2(n)\ \ w_3(n)\ \ w_4(n)\ \ w_5(n)]$,其权值系数仅在非语音段更新滤波器权值系数;在语音段时则不更新滤波器的权值系数。即可用递归公式式(17)获得:

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + 2\eta y_{\mathrm{GSC}}(n)\boldsymbol{u}(n) \tag{17}$$

式(17)中,η 为自适应步长。因此,广义旁瓣抵消输出为

$$y_{\mathrm{GSC}}(n) = d(n) - v_e(n) \tag{18}$$

4　实验分析

4.1　数据采集

利用 M-Audio 多路音频设备采集实验数据。麦克风阵为六元线阵,在 10 m × 5 m × 5 m 的房间中录制,混响时间 $\mathrm{RT}_{60} \approx 0.35$ s。麦克风阵语音与噪声的位置如图 3 所示。说话人距离阵列中心垂直距离 25 cm。噪声源使用 Noisex-92 数据库中的部分噪声。噪声源 1 采用 babble 噪声,采用端射形式位于距离 M_1 为 20 cm 处;噪声源 2 采用 f16 噪声,位于距离阵列中心 40 cm,垂直距离 30 cm 处;噪声源 3 采用 m109 噪声,位于距离阵列中心 14 cm,垂直距离10 cm处;噪声源 4 采用 white 噪声,位于距离 M_6 为 20 cm 处。语音信号与噪声信号的采样率都为 44.1 kHz。

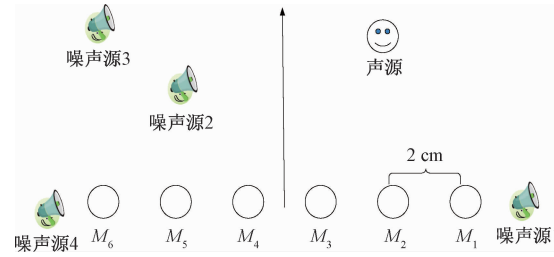

图 3　麦克风阵语音与噪声位置

Fig. 3　Microphone array speech and noise location

4.2　实验分析

实验采用 MTCF、GSC、CF-GSC 以及本文算法 MTCF-GSC 对抑制噪声效果进行对比。实验中帧长取 15 ms,帧移为 7 ms。本文算法中各参数取值:式(5)中,平滑因子 β 取值为 0.68;式(7)中,常量参数 δ 取值为 0.2,γ 取值为 0.8;式(9)中,λ_1 取值为 0.35,λ_2 取值为 0.8;式(10)中,搜索窗长 D 取值为 150;式(17)中自适应步长 η 取值为 0.002 5。

本文利用信噪比(Signal-to-Noise Ratio,SNR)以及分段信噪比(Segmental Signal-to-Noise Ratio,SSNR)来测试 MTCF、GSC、CF-GSC 以及本文算法 MTCF-GSC 抑制噪声效果。信噪比衡量含噪语音信号所带的噪声大小。分段信噪比能够反映出量化器对不同电平输入端的量化质量,采用分段的方法计算每一个分段的信噪比,该方法为了避免检测语音段与非语音段将分段信噪比值限制在[− 10 dB, 35 dB]的范围。

实验中第一类输入信噪比 0 dB、第二类输入信噪比 0 dB 用 0/0 表示,为了保证噪声的不相关性第二类噪声添加录制噪声的不同段。图 4 为 MTCF、GSC、CF-GSC 以及本文算法 MTCF-GSC 在 f16 与 babble 为背景噪声且信噪比为 0/0 的噪声抑制效

果图。

图 5 为 MTCF、GSC、CF-GSC 以及本文算法 MTCF-GSC 在 f16 与 babble 为背景噪声且不同信噪比下的增强后的信噪比及分段信噪比对比图。

图 4 中显示，MTCF-GSC 抑制噪声的效果明显优于 MTCF 与 GSC，主要原因在于 MTCF 主要抑制非相关噪声，对相关噪声抑制能力较弱。GSC 主要抑制相关噪声，对非相关及弱相关噪声抑制能力较弱。因此，MTCF-GSC 既能抑制非相关噪声以及弱相关噪声，还能有效抑制相关噪声。MTCF-GSC 的语音失真度小于 CF-GSC。主要原因在于采用了最小跟踪噪声功率谱估计的相干滤波避免了因端点检测技术导致的噪声谱估计不准确的问题，从而减小了语音失真。几种算法增强后的语音都有一定的失真，即损失了部分语音信息。

图 5 中显示，MTCF-GSC 抑制噪声效果明显优于其他几种算法。由实验数据表明，MTCF-GSC 更具备稳健的增强效果，与 CF-GSC 相比基本都能提高 2 dB 左右的信噪比。

为进一步证明本文算法的有效性，采用不同的混合噪声、不同的信噪比模拟实际环境中的复杂噪声。表 1、表 2 为 MTCF、GSC、CF-GSC 以及本文算法 MTCF-GSC 在不同混合噪声和不同信噪比情况下的抑制效果。

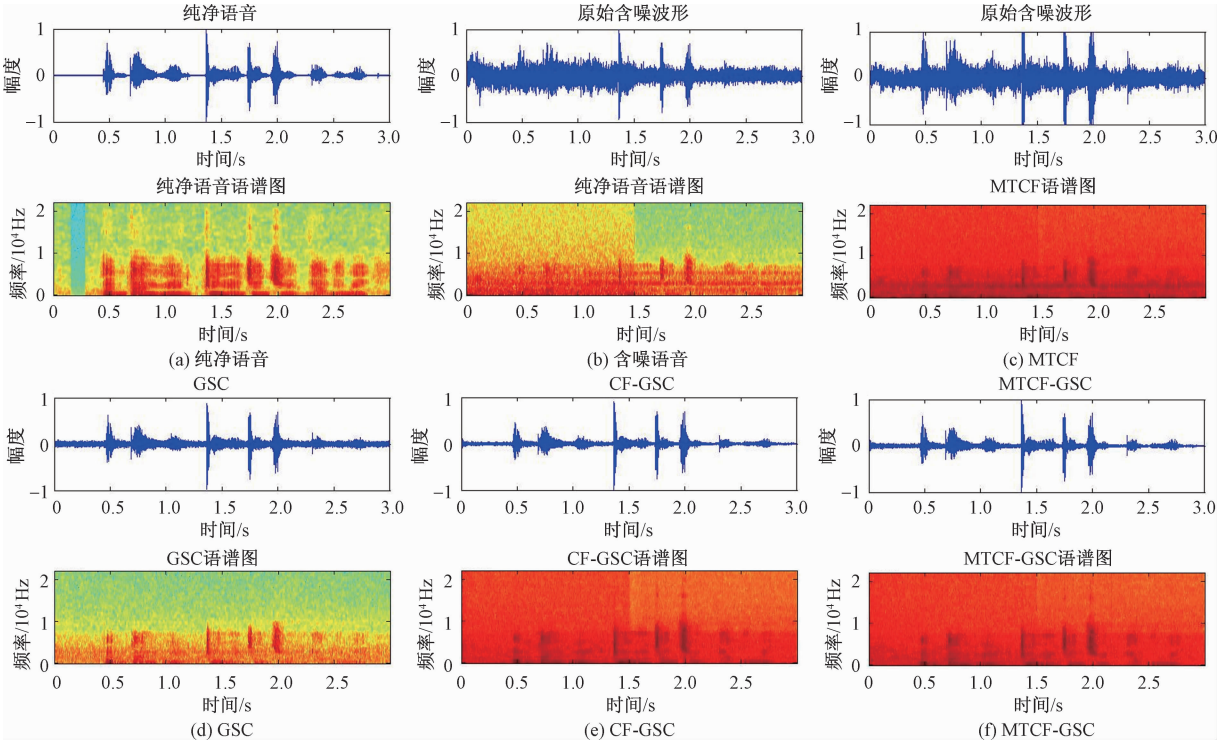

图 4 不同算法抑制背景噪声为 f16 及 babble 且信噪比为 0/0 的信噪比

Fig. 4 Different algorithm to suppress the background noise f16 and babble and the signal to noise ratio of 0/0 signal to noise ratic

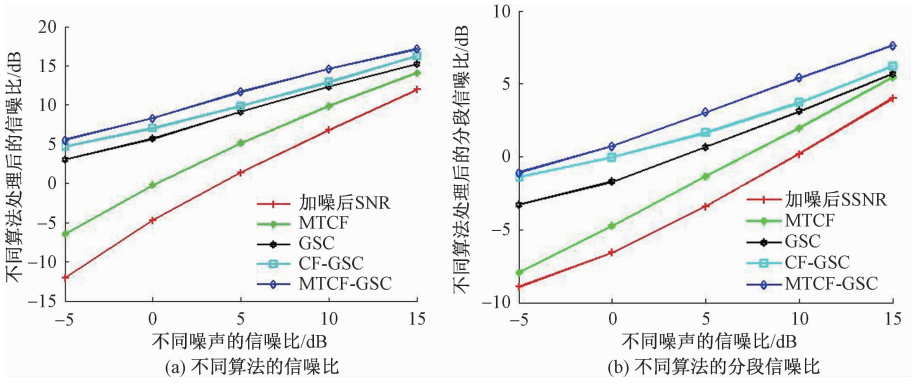

图 5 不同算法抑制背景噪声为 f16 及 babble 且不同信噪比的信噪比与分段信噪比

Fig. 5 Different algorithms to suppress the background noise f16 and babble and different signal to noise ratio of the signal to noise ratio and segmental signal to noise ratio

表 1　不同算法在不同混合噪声环境下的信噪比

Table 1　Signal-to-noise ratio of different algorithms in different mixed noise environments　dB

噪声	输入信噪比	加噪后语音信噪比	MTCF 信噪比	GSC 信噪比	CF-GSC 信噪比	MTCF-GSC 信噪比
m109 & f16	5/5	1.361 4	6.873 5	9.285 6	10.459 8	11.733 2
m109 & babble	0/5	-4.778 9	1.686 4	5.711 6	7.524 5	8.396 0
White & babble	5/5	1.440 9	7.500 9	9.771 9	10.558 3	11.792 0
f16 & m109 & babble	5/5/5	-1.038 8	3.639 6	7.810 1	8.349 7	10.016 9

表 2　不同算法在不同混合噪声环境下的分段信噪比

Table 2　segmental signal-to-noise ratio of different algorithms in different mixed noise environments　dB

噪声	输入信噪比	加噪后语音信噪比	MTCF 信噪比	GSC 信噪比	CF-GSC 信噪比	MTCF-GSC 信噪比
m109 & f16	5/5	-3.511 6	-0.490 6	0.783 0	2.015 4	3.041 6
m109 & babble	0/5	-6.651 0	-3.927 0	-1.675 7	0.102 7	0.673 7
White & babble	5/5	-3.350 7	-0.090 7	1.147 0	1.977 4	3.103 6
f16 & m109 & babble	5/5/5	-5.244 3	-2.686 9	-0.294 3	0.493 7	1.683 9

表 1、表 2 明显表明,在复杂环境下,MTCF-GSC 算法相对于 MTCF、GSC 以及 CF-GSC 算法的噪声抑制能力更强。

本文算法比 MTCF、GSC 以及 CF-GSC 抑制噪声更明显的主要原因在于采用最小跟踪噪声功率谱估计的相干函数并采用与 VAD 结合的 GSC 算法。MTCF-GSC 能够更准确的估计噪声功率谱以及减少语音失真。

5　结论

本文以六元麦克风小阵为模型,针对多噪声环境,提出基于最小跟踪噪声功率谱估计的相干滤波与广义旁瓣抵消的麦克风小阵语音增强算法。采用最小跟踪噪声功率谱估计噪声功率谱产生相干滤波器的传递函数,用以抑制非相关及弱相关噪声;利用 GSC 结合 VAD 算法抑制强相关噪声。通过实验证明了本文算法的有效性。

参　考　文　献

1　林静然. 基于麦克风阵列的语音增强算法研究. 成都:电子科技大学,2007:1—8
Lin Jingran. On speech enhancement based on microphone arrays. Chengdu: University of Electronic Science and Technology of China, 2007: 1—8

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

3　Griffiths L J, Jim C W. An alternative approach to linearly constrained adaptive beamforming. IEEE Transactions on Antennas & Propagation, 2010; 30(1):27—34

4　杨立春, 钱沄涛. 基于相干性滤波器的广义旁瓣抵消器麦克风小阵列语音增强方法. 电子与信息学报, 2012; 34(12): 3027—3033
Yang Lichun and Qian Yuntao. Speech enhancement with generalized sidelobe canceller based on a coherence-based filter for small microphone arrays. Journal of Electronics & Information Technology, 2012; 34 (12): 3027—3033

5　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. 浙江大学学报(工学版), 2013; 47(8): 1493—1499
Yang Lichun, Qian Yuntao, Wang Wenhong. A GSC algorithm based on null spectral subtraction for dual small microphone array speech enhancement. Journal of Zhejiang University (Engineering Science), 2013; 47 (8): 1493—1499

6　Yousefian N, Loizou P C. A dual-microphone speech enhancement algorithm based on the coherence function. IEEE Transactions on Audio Speech & Language Processing, 2011; 20(2):599—609

7　Niwa K, Hioka Y, Kobayashi K. Post-filter design for speech enhancement in various noisy environments. International Workshop on Acoustic Signal Enhancement. IEEE, 2014:35—39

8　马金龙, 曾庆宁, 胡　丹,等. 基于麦克风小阵的多噪声环境语音增强算法. 计算机应用, 2015;35(8):2341—2344
Ma Jinlong, Zeng Qingning, Hu Dan, *et al.* Speech enhancement algorithm based on microphone array under multiple noise environments. Journal of Computer Applications, 2015; 35 (8): 2341—2344

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

10　Wang Y. A post-filtering technique for enhancing acoustic echo cancelation system. International Conference on Web Information Systems and Mining. 2012:73—80

11　Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Transactions on Speech & Audio Processing, 2002; 9(5):504—512

12　Martin R. Bias compensation methods for minimum statistics noise power spectral density estimation. Signal Processing, 2006; 86(6): 1215—1229

Small Microphone Array Speech Enhancement in Multiple Noise Environments

XIAO Qiang, ZENG Qing-ning, WANG Yao, ZHENG Zhan-heng
(Guilin University of Electronic Technology, Key Laboratory of Cognitive Radio and Information Processing of Ministry of Education, Guilin 541004, China)

[**Abstract**] Aiming at limited the suppressing the noise under the circumstance of non-stationary and noise, the speech enhancement algorithm based on coherent filtering of the minimum tracking noise power spectrum estimation and generalized sidelobe canceller is proposed. The method uses the coherent filtering of the minimum tracking noise power spectrum to suppress the weak correlation noise, and then combines the generalized sidelobe canceller and the voice activity detection to suppress the strong correlation noise. The experimental results show that this method can effectively suppress the influence of noise and improve the intelligibility of speech.

[**Key words**] coherent filtering generalized sidelobe canceller microphone small array