基于 MGSC 与改进维纳滤波的麦克风阵列语音增强

肖 强，曾庆宁，王 瑶，谢先明，毛 维
(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，广西桂林 541004)

摘要：针对传统广义旁瓣抵消中阻塞矩阵语音泄露、非相干噪声消噪能力较弱及后置维纳滤波中相位不变等问题，提出一种基于改进广义旁瓣抵消与相位补偿维纳滤波的方法。该方法将阻塞矩阵变为阻塞滤波器从而减少了阻塞矩阵语音泄露，然后将相位补偿的维纳滤波用于估计纯净语音的幅度谱和相位谱，从而抑制广义旁瓣抵消残留的噪声。仿真及实测结果表明，该方法能够更加有效地抑制噪声的影响，提高语音的可懂度。

关键词：广义旁瓣抵消；相位补偿；阻塞滤波器；麦克风阵列
中图分类号：TN912.35 **文献标识码：**A **文章编号：**1000-3630(2017)-06-0567-07
DOI 编码：10.16300/j.cnki.1000-3630.2017.06.012

Speech enhancement of microphone array based on MGSC and improved Wiener filter

XIAO Qiang, ZENG Qing-ning, WANG Yao, XIE Xian-ming, MAO Wei
(*Guilin University of Electronic Technology*, *Key Laboratory of Cognitive Radio and Information Processing of Ministry of Education*, *Guilin* 541004, *Guangxi*, *China*)

Abstract: Traditional generalized sidelobe canceling algorithm is bad for non-coherent noise with the issue of speech leakage in block matrix, while Wiener post-filter also has the problem on unchanged phase. In order to modify this, a method is proposed based on modified generalized sidelobe canceling and phase compensation Wiener filtering for speech enhancement of microphone array. This method applies the idea of using a block filter instead of the block matrix to reducing speech leakage in block matrix, and meanwhile a way of Wiener filtering based on phase compensation is used to improve the performance of the post-filtering after generalized sidelobe canceling via estimating the magnitude spectrum and the phase spectrum of clean speech. The simulation results and perceptual evaluation of speech quality show that this method is able to eliminate the noise and to improve the speech intelligibility more effectively.

Key words: generalized sidelobe canceling; phase compensation; block filter; microphone array

0 引 言

各种各样的噪声严重影响了语音质量及语音识别系统的识别率，过大的干扰噪声甚至会影响到语音的可懂度。麦克风阵列语音增强技术的出现，打破了人们使用单一麦克风处理语音信号的固有模式。与传统的单麦克风语音增强技术相比，麦克风阵列技术既能利用阵元间的时域与空域信息抑制其他方向的干扰，又能确定声源在空间的位置。

麦克风阵列语音增强方法主要有波束形成(Beam Forming，BF)、自适应噪声抵消(Adaptive Interference Cancellation，AIC)、多路维纳滤波(Multichannel Wiener Filter，MWF)、子带分解法(Subband Decomposition，SD)、广义旁瓣抵消(Generalized Sidelobe Canceller，GSC)等[1-7]。比较这几种方法，GSC 不仅能消除相关性比较弱的噪声，而且也可消除相关性比较强的噪声，因此被广泛地研究和改进。GSC 最早由 Griffiths L[8]等提出，包括三部分：固定波束形成器(Fixed Beamforming，FBF)、阻塞矩阵(Blocking Matrix，BM)以及自适应干扰抵消器，该结构中 BM 存在严重的语音泄露，导致 AIC 中存在语音抵消的现象。针对这个问题，对 GSC 结构中 BM 部分进行改进成为了一个重要的研究方向，例如多路抗串扰自适应信号抵消、基于传输函数比的阻塞矩阵改进以及基于子空

收稿日期：2017-01-13; **修回日期：**2017-04-17
基金项目：国家自然科学基金(61461011)、“认知无线电与信息处理”教育部重点实验室 2016 年主任基金(CRKL160107)、广西自然科学基金重点项目(2016GXNSFDA380014)
作者简介：肖强(1992－)，男，湖南益阳人，硕士研究生，研究方向为语音信号处理。
通讯作者：肖强，E-mail: qiangxiao0116@163.com

间投影的 GSC(Subspace Generalized Sidelobe Canceller，SGSC)等[9-11]。由于 GSC 对弱相干噪声及非相干噪声的抑制能力差，1977 年 Allen 提出一种将维纳滤波(Wiener Filter，WF)作为后置滤波并与自适应波束形成结合的语音增强方法[12]，Cohen 提出将广义旁瓣抵消器和后置滤波器结合的方法对抑制弱相干及非相干噪声有较好的效果[13]。但该结构中不仅存在语音泄露问题，而且在后置维纳滤波中没有修正增强后的语音的相位。

本文提出一种基于改进广义旁瓣抵消(Modify Generalized Sidelobe Canceller，MGSC)与相位补偿维纳滤波(Phase Compensation Wiener Filter，PCWF)的麦克风阵列语音增强方法。该方法将 GSC 结构中的阻塞矩阵变为阻塞滤波器，在减少阻塞矩阵语音泄露问题上有较好的效果；后置相位补偿维纳滤波不仅估计了纯净语音的幅度谱，而且修正了相位谱。实验证明本文算法能够获得较好的性能。

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

$$y_i=s_i+v_i,\quad i=1,2,\cdots,M \tag{1}$$

式(1)中：y 为含噪语音信号；s 为纯净语音信号；v 为噪声；M 为麦克风数量。假设纯净语音与噪声互不相关，即 $\mathrm{E}(sv^{\mathrm{H}})=0$。对式(1)两边进行短时傅里叶变换得出其频域形式：

$$Y_i(u,w)=S_i(u,w)+V_i(u,w),\quad i=1,2,\cdots,M \tag{2}$$

式(2)中，u 表示帧数，w 表示离散频率。

其含噪语音频谱用极坐标表示为

$$Y_i(u,k)=|Y_i(u,k)|\mathrm{e}^{\mathrm{j}\angle Y_i(u,k)},\quad i=1,2,\cdots,M \tag{3}$$

式(3)中，$|Y_i(u,k)|$ 表示第 i 个麦克风的频域幅度谱，$\angle Y_i(u,k)$ 表示第 i 个麦克风的频域相位谱。

2 改进的广义旁瓣抵消与相位补偿维纳滤波

本文的语音增强算法结构框图如图 1 所示，从图 1 中可见，算法分为两个部分：改进的广义旁瓣抵消及相位补偿维纳滤波。其主要思想是 MGSC 抑制相干噪声，后置 PCWF 用以抑制 MGSC 输出信号残留的非相干及弱相干噪声。

2.1 改进的广义旁瓣抵消

GSC 抑制噪声的原理是利用参考噪声信号抵消 FBF 输出中的噪声信号，其中 FBF 一般用延迟

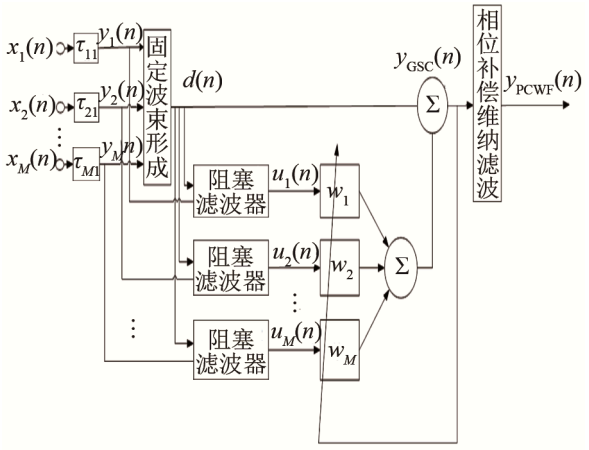


图 1 基于 MGSC-PCWF 的麦克风阵列语音增强原理框图

Fig.1 Block diagram of MGSC-PCWF based microphone array speech enhancement

求和波束形成(Delay Sum Beamforming，DSB)实现，麦克风接收到的语音信号经过 DSB 后表示为

$$d(n)=\frac{1}{M}\sum_{i=1}^{M}y_i \tag{4}$$

式(4)中，$d(n)$ 为语音期望信号；y_i 是麦克风接收到的信号经过时延补偿后的信号。图 1 中时延补偿通过式(5)获得：

$$\tau_{i1}=\frac{d_{i1}\cos(\alpha)}{c} \tag{5}$$

式(5)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。

传统的 GSC 结构中，阻塞矩阵存在严重的语音泄露问题，导致参考的噪声中存在大量的语音信息，从而使自适应干扰抵消结构中存在语音抵消，致使语音信号产生畸变。针对 GSC 结构中语音泄露的问题，本文引入阻塞滤波器产生噪声参考信号[14]，阻塞滤波器是在频域利用前一帧估计的噪声与前一帧的期望信号产生增益函数，然后再利用谱减法对语音信号进行阻塞以估计噪声。其中增益函数定义为[14]：

$$H_i(u+1,w)=H_i(u,w)+\beta\frac{U_i(u,w)}{P[D(u,w)]}D^*(u,w) \tag{6}$$

式(6)中，*表示共轭，β 为步长；$D(u,w)$ 表示 $d(n)$ 的短时傅里叶变换；$U_i(u,w)$ 表示第 i 个麦克风接收到参考噪声的短时傅里叶变换，由式(7)获得；$P[D(u,w)]$ 表示期望信号的功率谱，由式(8)获得[14]。

$$U_i(u,w)=Y_i(u,w)-D(u,w)H_i(u,w) \tag{7}$$

$$P[D(u+1,w)]=\lambda P[D(u,w)]+(1-\lambda)|D(u,w)|^2 \tag{8}$$

式(7)中，λ 表示平滑因子。

MGSC 的输出可以表示为

$$y_{\mathrm{GSC}}(n)=d(n)-\boldsymbol{w}^{\mathrm{T}}(n)\boldsymbol{u}(n) \tag{9}$$

式(9)中，$\boldsymbol{u}(n)=([u_1,u_2\cdots u_M])$；$\boldsymbol{w}=([w_1,w_2\cdots w_M])^{\mathrm{T}}$，表示 LMS 自适应滤波算法系数，其滤波系数向量 $\boldsymbol{w}$ 可用递归公式计算，见式(10)：

$$\boldsymbol{w}(n+1)=\boldsymbol{w}(n)+2\delta y_{\mathrm{GSC}}(n)\boldsymbol{u}(n) \tag{10}$$

式(10)中，δ 为步长因子。

2.2　相位补偿维纳滤波

传统的维纳滤波算法只对幅度谱进行改变，保持相位谱不变，其原因是之前普遍认为含噪语音的相位谱是纯净语音的某种最优估计。而最近的研究证明，改变相位谱能够获得较好的语音可懂度[15]，Wojcicki K 等人只改变含噪语音的相位谱、不改变幅度谱获得了较好的语音可懂度[16]。针对传统维纳滤波相位谱不变，本文提出一种基于相位补偿的维纳滤波方法。

相位补偿维纳滤波函数定义为[15]：

$$Y_{\mathrm{PCWF}}(u,w)=\left|\hat{S}(u,w)\right|\mathrm{e}^{\mathrm{j}\angle\hat{S}(u,w)} \tag{11}$$

式(11)中，$\hat{S}(u,w)$ 表示纯净语音的估计值，通过式(12)得到；$\angle$ 表示取其相位。

$$\hat{S}(u,w)=G(u,w)Y_{\mathrm{GSC}}(u,w) \tag{12}$$

式(12)中，$G(u,w)$ 是维纳滤波器的传递函数，定义为：

$$G(u,w)=\frac{SNR_r(u,w)}{SNR_r(u,w)+\varepsilon} \tag{13}$$

式(13)中，$SNR_r(u,w)$ 表示先验信噪比；ε 是噪声抑制因子，当 ε 值较大时，在低输入信噪比的条件下通过维纳滤波增强信号的衰减很大；当 ε 值较小时，在高输入信噪比条件下通过维纳滤波增强信号的衰减很小。针对这个问题，通过后验信噪比确定 ε 值，定义为[17]：

$$\varepsilon=\begin{cases}1 & SNR_s(u,w)>SNR_{th}(u,w)\\ 1.8 & \text{else}\end{cases} \tag{14}$$

式(14)中，$SNR_s(u,w)$ 表示后验信噪比，$SNR_{th}(u,w)$ 定义为：

$$SNR_{th}(u,w)=\frac{1}{K}\sum_{w=1}^{K}SNR_s^{(w)}(u,w) \tag{15}$$

式(15)中，K 表示帧长。

由于含噪信号是实信号，故经过短时傅里叶变换后是共轭对称的，即 $Y_{\mathrm{GSC}}(u,w)=Y_{\mathrm{GSC}}^{*}(u,N-w)$，利用反对称函数来改变共轭之间的角度关系，从而可以补偿相位谱。相位补偿的度数根据估计的噪声幅度谱来确定。

相位函数定义为：

$$\angle\hat{S}(u,w)=\angle Y_{\mathrm{GSC}}(u,w)+\theta(u,w) \tag{16}$$

式(16)中，$\angle\hat{S}(u,w)$ 是纯净语音的频域相位谱估计，由 $Y_{\mathrm{GSC}}(u,w)$ 相位谱叠加相位补偿函数得到，可以通过式(17)计算得到。

$$\theta(u,w)=\xi\psi(w)\left|\hat{v}(u,w)\right|^2 \tag{17}$$

式(17)中，ξ 是一个常数；$\left|\hat{v}(u,w)\right|$ 是估计的噪声幅度谱；$\psi(w)$ 是反对称函数，即：

$$\psi(w)=\begin{cases}1 & 0<\dfrac{w}{N}<0.5\\ -1 & 0.5<\dfrac{w}{N}<1\\ 0 & \text{其他}\end{cases} \tag{18}$$

3　实验结果与分析

实验使用 M-Audio 多路音频设备采集数据。实验分为仿真实验和实际含噪语音实验。麦克风阵列为四元平面矩形阵。如图 2 所示，M_1 与 M_3 之间的距离为 16 cm，M_1 与 M_2 之间的距离为 2 cm。录制环境为空旷的天台。仿真实验声源位于麦克风阵中心的正前方。噪声采用 Noise-92 数据库中 white 噪声。仿真实验中语音的采样率为 44.1 kHz，噪声采样率为 8 kHz。实际含噪语音实验如图 2 所示，说话人位于 M_1 与 M_3 的垂直平分线上且距离阵列中心 20 cm 处，噪声为 m109，位于 M_2 与 M_4 的垂直平分线上且距离阵列中心 30 cm 处。实际含噪语音的语音和噪声的采样率均为 44.1 kHz。

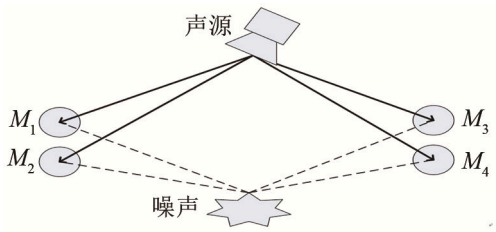


图 2　四元平面麦克风矩形阵

Fig.2　Four element rectangular microphone array

在仿真实验与实际含噪语音实验中，以 M_1 为参考麦克风。由于 M_1 与 M_2 之间的距离很小，因此 M_1 与 M_2 之间的时延可以忽略不计，即 $\tau_{21}=0$。由于声源位于 M_1 与 M_3 的垂直平分线上，因此 M_1 与 M_3 之间的时延为 0，M_1 与 M_4 之间的时延约等于 M_1 与 M_3 之间的时延。延迟采样点 $N_{i1}=ceil(\tau_{i1}*f_s)$，其中 $ceil$ 表示大于或者等于指定表达式的最小整数函数。f_s 是含噪语音的采样率。

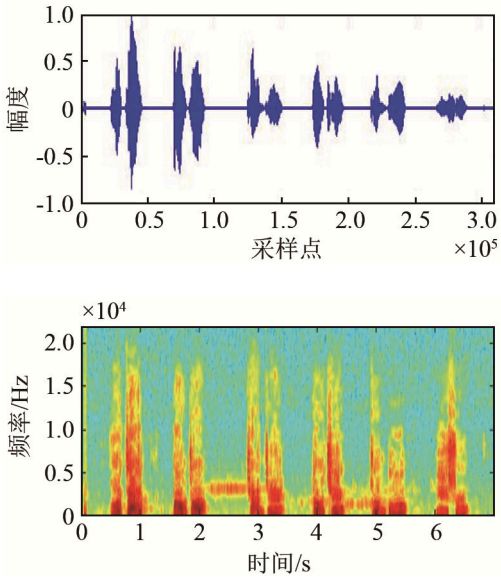
在仿真实验中帧长取 20 ms，帧移为 10 ms。在实际含噪语音实验中帧长取 40 ms，帧移取 20 ms。在阻塞滤波器中的步长 β 取值为 0.1，平滑

因子λ取值为 0.68。在 PCWF 中的相位补偿函数中ξ取值为 3.47。

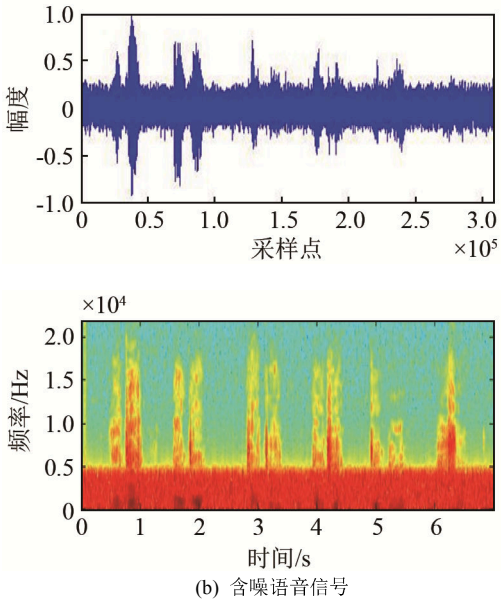
图 3 是仿真实验结果。在 4 个麦克风接收到的每条语音中分别加入信噪比为 0 dB 的白噪声。图 3 为以 white 噪声为背景噪声，在 0 dB 信噪比(Signal to Noise Ratio，SNR)的环境下不同算法对噪声处理后的时域和频域效果仿真图。

图 4 是实际含噪语音实验结果。图 4 为以 m109 为背景噪声，在 0 dB 信噪比(SNR)的环境下不同算法对噪声处理后的时域和频域效果仿真图。

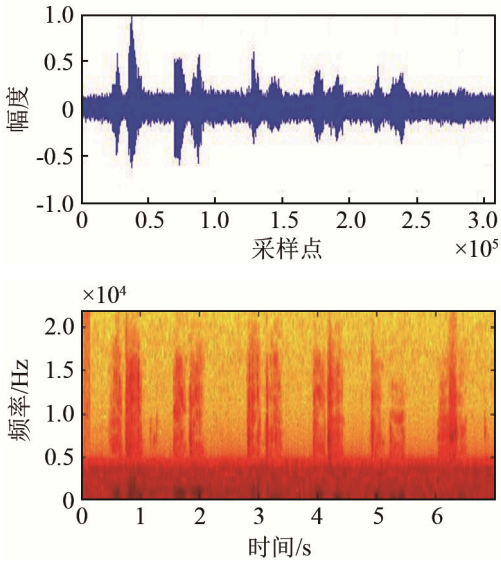
从图 3(c)与图 3(e)及图 3(d)与图 3(f)中可以明显地看出，后置维纳滤波能够有效地抑制残留的噪声，使处理后的语音波形更接近纯净语音波形。从图 3(e)与图 3(f)的时频仿真图可以看出，MGSC-



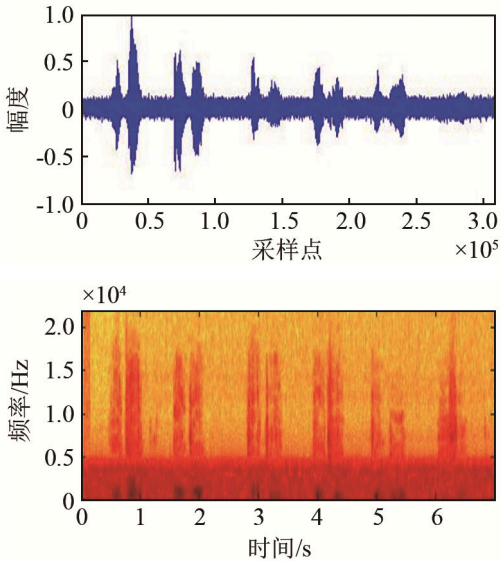
(a) 纯净语音信号



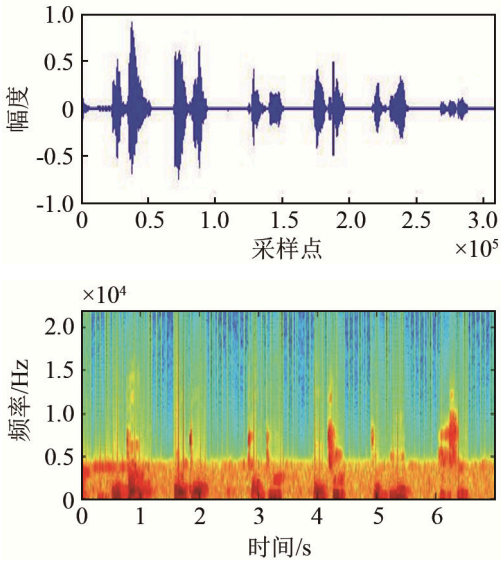
(b) 含噪语音信号



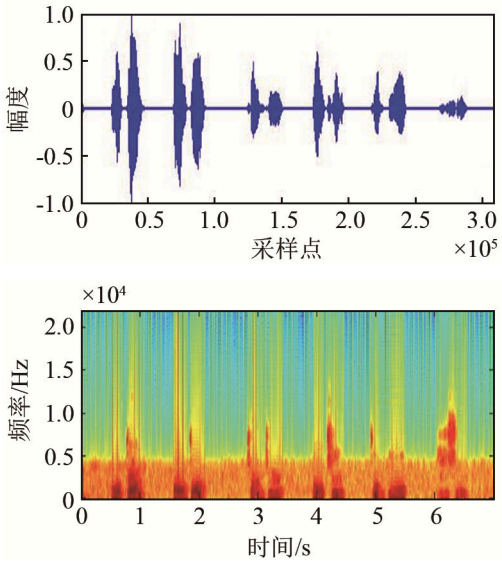
(c) GSC



(d) MGSC



(e) GSC-WF



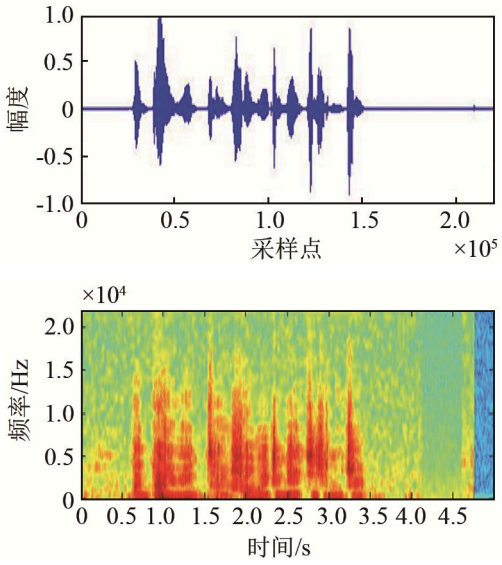
(f) MGSC-PCWF

图 3　背景为白噪声，信噪比为 0 dB 环境下不同算法处理前后的时频图

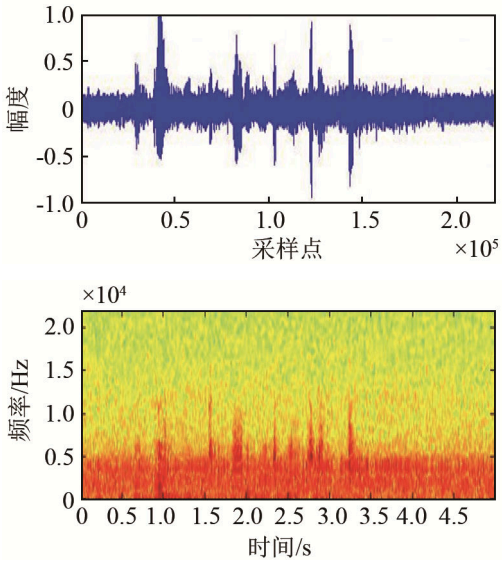
Fig.3　Time-frequency diagrams before and after different algorithm processing in the environment where the background noise is white and the signal to noise ratio is 0 dB

PCWF 的去噪效果优于 GSC-WF。在此次仿真实验中发现，本文提出的 MGSC-PCWF 算法对噪声抑制有十分突出的效果。

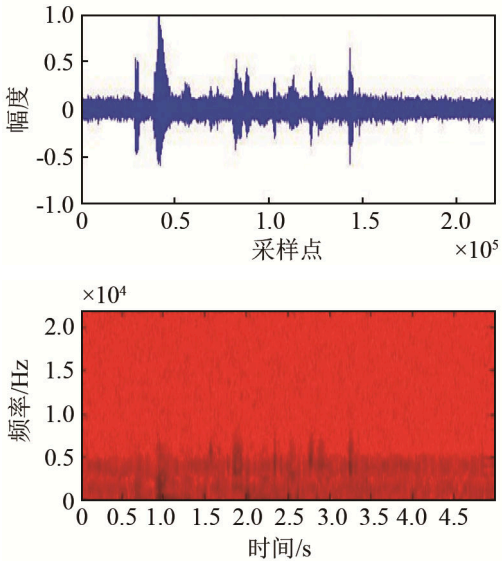
从图 4(d)中可以明显看出，GSC 结构中 BM 存在严重的语音泄露问题，而 MGSC 结构中的阻塞滤波器对语音泄露问题有较好的效果，较大程度上避免了因语音泄露而导致 AIC 结构中语音抵消的问题，从而降低了语音失真。从图 4(e)与图 4(f)中可以看出，PCWF 对抑制非相干噪声与弱相干噪声的能力优于传统的维纳滤波。从图 4(f)、4(g)及 4(h)中可以明显发现，MGSC 对抑制强相关噪声有较好的效果，且与 PCWF 结合还能抑制残留的非相干



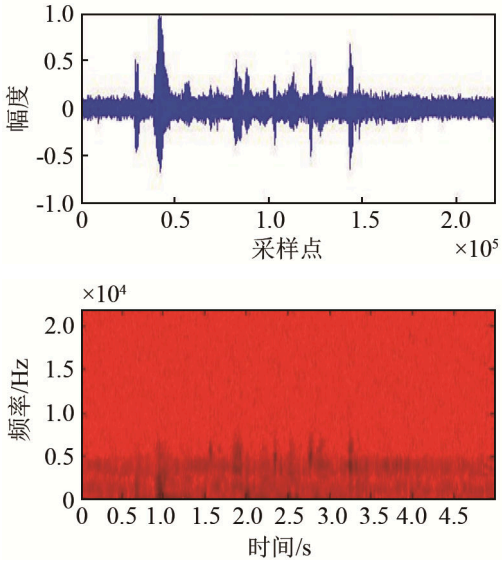
(a) 纯净语音信号



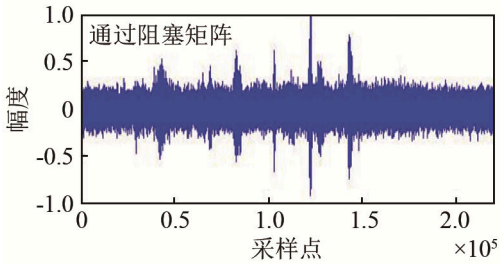
(b) 含噪语音信号

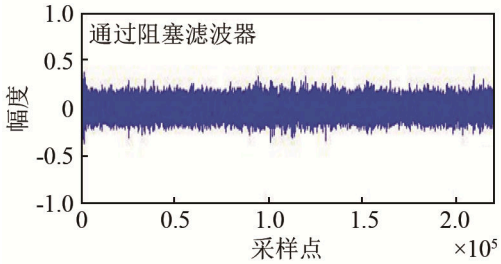


(c) GSC

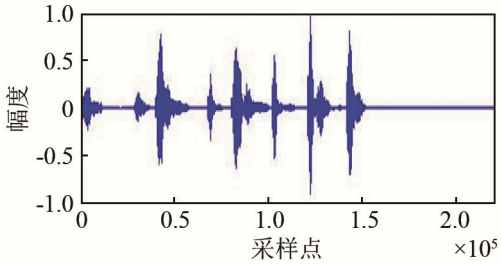


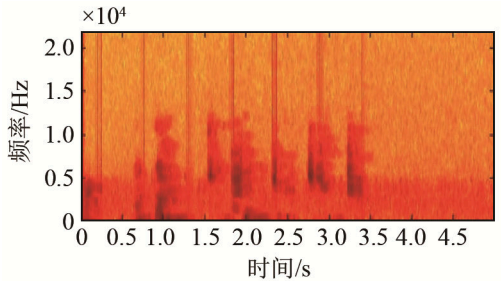
(d) MGSC



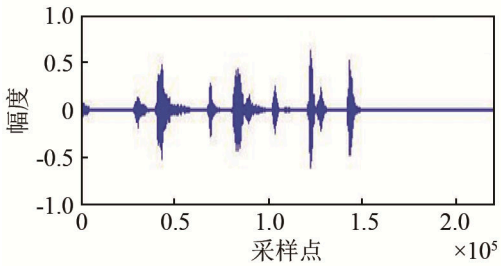


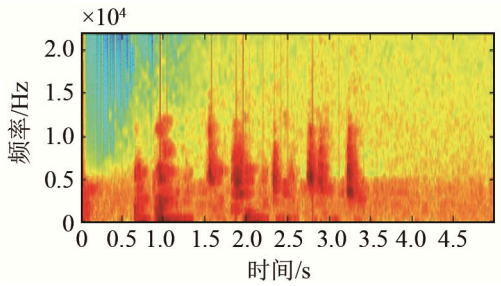
(e) 参考噪声



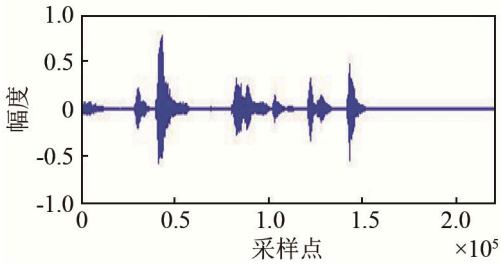


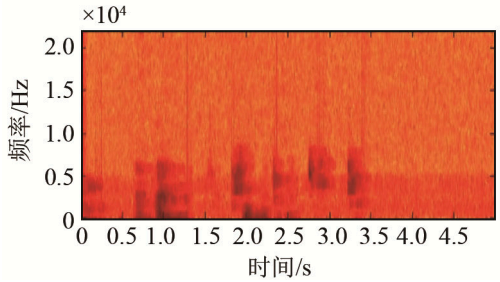
(f) WF



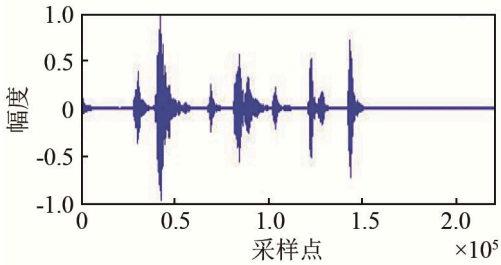


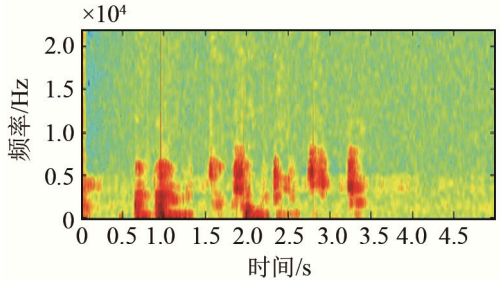
(g) PCWF





(h) GSC-WF





(i) MGSC-PCWF

图 4 背景噪声为 m109，信噪比为 0 dB 环境下不同算法处理前后的时频图

Fig.4 Time-frequency diagrams before and after different algorithm processing in the environment where the background noise is m109 and the signal to noise ratio is 0 dB

噪声和弱相干噪声。在此次实际含噪语音实验中发现本文提出的 MGSC-PCWF 算法对抑制噪声有十分突出的作用。

利用主观语音质量评估(Perceptual Evaluation of Speech Quality，PESQ)，可以进一步表明几种算法可以抑制含噪语音的噪声，提高语音可懂度。PESQ 是 ITU-T P.862 建议书提供的客观平均意见得分(Mean Opinion Score，MOS)值评价方法，与 MOS 打分相关度达到 0.97，且能避免进行 MOS 打分时引入的主观因素。原始语音信号和增强后的语音信号都重新采样至 8 kHz 后进行评估。

在此次试验中分别以 white、babble 与 f16 为背景噪声的环境对不同信噪比的 PESQ 进行评估。麦克风接收到的信号的信噪比分别为：−5、0、5、20 dB，评分结果如表 1、表 2 与表 3 所示。

表 1　white 噪声环境不同 SNR 的 PESQ 得分

Table 1　PESQ scores of different SNRs in the white noise environment

Input_SNR/dB	−5	0	5	20
含噪语音	0.574 6	0.811 0	1.164 6	2.276 5
SGSC	0.564 8	0.832 3	1.205 7	2.307 1
MGSC	0.684 0	0.938 8	1.330 5	2.316 0
GSC-WF	1.702 3	1.976 5	2.294 1	2.944 1
MGSC-WF	1.685 4	2.110 9	2.328 9	3.002 8
MGSC-PCWF	2.072 5	2.378 2	2.622 5	3.420 6

表 2　babble 噪声环境不同 SNR 的 PESQ 得分

Table 2　PESQ scores of different SNRs in the babble noise environment

Input_SNR/dB	−5	0	5	20
含噪语音	1.066 6	1.404 9	1.653 1	2.596 6
SGSC	0.975 3	1.423 7	1.723 8	2.684 7
MGSC	1.087 1	1.459 2	1.858 6	2.680 0
GSC-WF	1.575 0	1.722 0	1.840 5	2.820 1
MGSC-WF	1.719 7	1.790 7	1.953 5	2.905 6
MGSC-PCWF	1.688 9	1.900 3	2.099 1	3.139 5

表 3　f16 噪声环境不同 SNR 的 PESQ 得分

Table 3　PESQ scores of different SNRs in the f16 noise environment

Input_SNR/dB	−5	0	5	20
含噪语音	0.875 1	1.186 7	1.575 7	2.500 3
SGSC	0.735 9	1.238 5	1.658 7	2.586 9
MGSC	1.006 5	1.341 5	1.762 8	2.594 3
GSC-WF	1.835 5	2.071 9	2.331 1	2.903 4
MGSC-WF	1.915 6	2.162 7	2.475 2	3.054 8
MGSC-PCWF	2.057 3	2.351 6	2.669 0	3.361 2

从表 1、表 2 及表 3 中可以看出，即使 MGSC-PCWF 过程中存在一定的语音失真，但本文算法在不同背景噪声以及不同信噪比的环境下仍然能获得相对较好的性能。

4　结 论

本文以四元麦克风平面阵为基础，提出一种基于改进广义旁瓣抵消与相位补偿维纳滤波的麦克风阵列语音增强方法。在 GSC 结构中用阻塞滤波器替换阻塞矩阵，较大程度地避免了语音泄露。在后置相位补偿维纳中根据后验信噪比的变化，对噪声抑制因子进行修正且采用相位补偿函数进一步提高语音的可懂度。通过实验证明了本文算法的有效性。

参　考　文　献

[1] Benesty J, Makino S, Chen J. Speech enhancement[J]. Signals & Communication Technology, 2005.

[2] Ngo K, Spriet A, Moonen M, et al. A combined multi-channel Wiener filter-based noise reduction and dynamic range compression in hearing aids[J]. Signal Processing, 2012, **92**(2): 417-426.

[3] Markovich-Golan S, Gannot S, Cohen I. Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks[J]. IEEE Transactions on Audio Speech & Language Processing, 2013, **21**(2): 343-356.

[4] Li Kai, Fu Q, Yan Y. A subband feedback controlled generalized sidelobe canceller in frequency domain with multi-channel postfilter[C]//International Workshop on Intelligent Systems and Applications IEEE, 2010: 1-4.

[5] Cornelis B, Moonen M, Wouters J. Performance analysis of multichannel wiener filter-based noise reduction in hearing aids under second order statistics estimation errors[J]. IEEE Transactions on Audio Speech & Language Processing, 2011, **19**(5): 1368-1381.

[6] Jinsoo P, Wooil K, Han K D, et al. Two-microphone generalized sidelobe canceller with post-filter based speech enhancement in composite noise[J]. Etri Journal, 2015, **38**(2).

[7] Yu G J, Shin J W, Kim N S. Spectro-temporal filtering for multichannel speech enhancement in short-time fourier transform domain[J]. IEEE Signal Processing Letters, 2014, **21**(3): 352-355.

[8] Griffiths L J, Jim C W. An alternative approach to linearly constrained adaptive beamforming[J]. IEEE Transactions on Antennas & Propagation, 2010, **30**(1): 27-34.

[9] 曾庆宁, 欧阳缮. 子带 MCRASC—MGSC 微型麦克风阵语音增强算法[J]. 西安电子科技大学学报(自然科学版), 2010, **37**(6): 1011-1016.
CENG Qingning, OUYANG Shan. Subband MCRASC-MGSC algorithm for speech enhancement with the mini-typed microphone array[J].Journal of Xidian University(Natural Science Edition), 2010, **37**(6): 1011-1016

[10] 宋辉, 刘加. 基于广义奇异值分解的通用旁瓣消除算法[J]. 数据采集与处理, 2011, **26**(3): 241-246.
SONG Hui, LIU Jia. GSVD-based generalized sidelobe canceller[J]. Journal of Data Acquisition & Processing, 2011, **26**(3): 241-246.

[11] 郭庆华, 廖桂生. 一种稳健的自适应波束形成器[J]. 电子与信息学报, 2004, **26**(1): 146-150.
GUO Qinghua, LIAO Guisheng. A robust adaptive beamformer [J]. Journal of Electronics Information Technology, 2004, **26**(1): 146-150.

[12] Allen C. Adaptive multi-beam antannas for spacelab[C]//Antennas and Propagation Society International Symposium. IEEE, 1977: 420-423.

[13] Gannot S, Cohen I. Speech enhancement based on the general transfer function GSC and postfiltering[J]. IEEE Transactions on Speech & Audio Processing, 2004, **12**(6): 561-571.

[14] Wang D, Yin F. Subband adaptive beamforming for microphone array speech enhancement[C]//International Conference on Signal Processing. IEEE Xplore, 2006.

[15] Li Z, Wu W, Zhang Q, et al. Speech enhancement using magnitude and phase spectrum compensation[C]//International Conference on Computer and Information Science. IEEE Computer Society, 2016: 1-4.

[16] Wojcicki K, Milacic M, Stark A, et al. Exploiting conjugate symmetry of the short-time fourier spectrum for speech enhancement[J]. IEEE Signal Processing Letters, 2008, **15**(5): 461-464.

[17] Zhang M, Wu S, Guo W, et al. A microphone array dereverberation algorithm based on TF-GSC and postfiltering[C]//IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. IEEE, 2016: 1-4.