

文章编号: 2096-1618(2024)03-0263-05

一种改进的广义旁瓣对消阵列语音增强算法

童环, 夏秀渝

(四川大学电子信息学院, 四川 成都 610065)

摘要:传统广义旁瓣对消(generalized sidelobe cancellation, GSC)算法在复杂声学环境下性能较差, 语音增强效果不理想, 需要加强其去噪能力以提高输出语音质量。对此, 提出一种改进型广义旁瓣对消算法。构建基于频域 GSC 和时域 GSC 两级滤波的系统结构, 利用一个选择滤波器输出两级滤波中质量较好的语音; 并将 GSC 上支路权值修改为可自适应调节的形式, 提高算法的适应性; GSC 权值迭代时采用一种变步长自适应算法, 步长因子根据信号信噪比实时调整, 防止滤波器权值发散。实验结果表明, 相比于传统 GSC 算法, 新算法在 SINR、PESQ、STOI 和 SDR 等指标上都有提升。

关键词:麦克风阵列; 语音增强; GSC; 变步长自适应算法

中图分类号: TN912.35

文献标志码: A

doi: 10.16836/j.cnki.jcuit.2024.03.001

0 引言

由于受到环境噪声干扰及系统噪声的影响, 由麦克风采集到的语音信号质量不高, 不能直接用于语音通话、语音识别等后续工作中, 需要设计算法对其进行增强。基于单麦克风的语音增强系统大多计算简单, 但去噪效果较差。而麦克风阵列语音增强系统可利用空间信息增强目标语音, 能获得比单麦克风更好的增强效果^[1]。

麦克风阵列波束形成方法可分为固定波束形成和自适应波束形成两类。固定波束形成算法的滤波器系数固定不变, 系统结构和计算较简单, 对特定场景和方向有较好去噪效果, 但对于不同声学环境的适应性较差。自适应波束形成算法的滤波器系数能根据环境自适应变化, 可以更好地保证波束性能。自适应波束形成算法中, 由 Griffiths 等^[2]提出的广义旁瓣对消(GSC)算法由于结构简单, 被广泛应用, 但仍有许多不足, 如 GSC 算法消除相干噪声的效果较好, 处理非相干噪声时表现一般, 对此提出增加后置滤波进行改进。Zelinski^[3]提出基于维纳滤波的后置滤波 GSC 方法, 刘力玮等^[4]提出基于谱减法的改进 GSC 算法等, 但后置滤波法在有效去除非相干噪声的同时也带来较大的语音失真。另外实际应用场景中, 语音、干扰均是随机和非平稳的, 即系统输入信噪比实时变化, 而目前的 GSC 算法对于声学环境变化的适应性不够好, 不能有效地跟踪信号和噪声统计特性的变化, 去噪效果不够稳定。

针对上述问题, 对 GSC 算法进行改进。将传统

GSC 固定支路权值改为自适应权值, 并利用时域和频域二次自适应滤波提高算法对于相干噪声和非相干噪声的去除能力; 根据信噪比实时调整自适应算法的迭代步长, 避免高输入信噪比时自适应对消器学习效果下降的问题; 最后结合选择滤波器, 进一步提升最终输出语音的质量。本文方法在 SINR、PESQ、STOI、SDR 等指标上全面优于传统 GSC 方法, 相比于其他带后置滤波的 GSC 改进方法, 本文算法在高输入信噪比时去噪效果表现优秀, 在减少语音失真方面也有较大优势。

1 改进 GSC 算法

传统 GSC 滤波器的原理框图如图 1 所示。

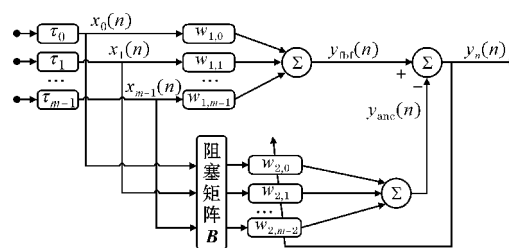


图1 传统 GSC 原理图

图 1 所示的传统时域 GSC 系统^[2]由固定支路和自适应支路组成。固定支路将时延补偿后各通道的信号以一个固定的权值求和, 可以简单消除一部分非相干噪声。自适应支路中信号先通过阻塞矩阵 B 以消除期望方向的目标信号, 只留下噪声信号, 再利用 LMS 算法调节自适应滤波器的权系数, 估计出固定支路中的噪声, 在系统输出端相减实现噪声抵消, 从而增强目标信号。

在实际应用中,由于声学环境复杂,阻塞矩阵并不能完全滤除目标信号,导致目标信号泄露到自适应支路中,这样系统也会对消掉部分目标信号,造成目标信号失真,噪声信号消除不完全。本文在此基础上进行了改进,原理图如图2所示。

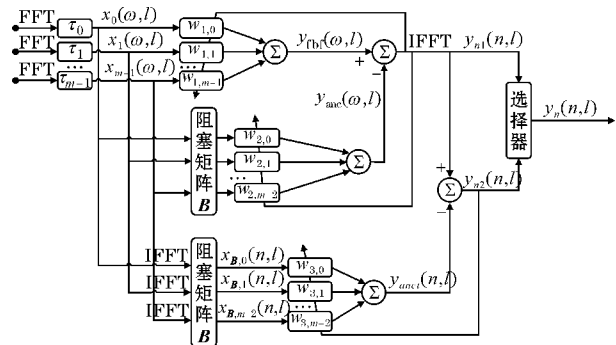


图2 改进 GSC 原理示意图

根据信号流向,本文 GSC 算法有 3 条支路。最上面一条信号支路对应于传统 GSC 算法中的固定波束形成支路。对阵列输入信号进行时延补偿以后,本文将上支路的固定权值修改为自适应的形式,可以更灵活地配合下支路进行噪声消除。下面两条支路均为自适应支路,用于模拟产生上支路中的噪声或干扰。其中中间支路采用频域 GSC 算法进行第一级消噪,下支路采用时域 GSC 算法进行第二级消噪,最后利用一个选择器选择两级消噪结果中质量较好的输出信号作为最终输出。自适应算法采用本文提出的变步长自适应算法。

1.1 上支路权值自适应

传统广义旁瓣抵消器中,固定支路多采用延迟求和的办法简单消除部分非相干噪声,这样虽然方法简单,但噪声消除效果有限,本文将固定权值更改为自适应权值,以更好地配合自适应支路进行噪声的估计和消除。

传统 GSC 固定支路输出为

$$y_{fb}(\omega, l) = w_1^T x(\omega, l)$$

其中 $w_1 = [w_{1,0}, w_{1,1}, \dots, w_{1,m-1}]^T$ 为 m 路信号的权值; $x(\omega, l) = [x_0(\omega, l), x_1(\omega, l), \dots, x_{m-1}(\omega, l)]^T$ 为经过延时补偿后的 m 路信号, w_1 的权值固定不变。本文将 w_1 设置成自适应调整的权值,增加系统灵活性。下文称此支路为上支路, w_1 的更新公式为

$$w_1 = w_1 - \mu_1 y_{n1}(\omega, l) x(\omega, l)$$

其中

$$y_{n1}(\omega, l) = y_{fb}(\omega, l) - y_{anc}(\omega, l)$$

由于本文第一级 GSC 是在频域进行的,为避免上下支路滤波器权值变为 0,限制 w_1 为实数,并对 w_1 权

值做归一化约束。这样上支路对各通道输入信号进行一定程度的自动选择,可以更灵活地配合下支路自适应地进行噪声消除。增加约束后上支路权值调整公式变为

$$y_{fb}(\omega, l) = \frac{w_1}{\sum w_1} x(\omega, l)$$

$$w_1 = w_1 - \frac{\mu_1 y_{n1}(\omega, l) x(\omega, l)}{\sum w_1}$$

每次运算时,上支路的权值之和为 1,不会出现权值及输出全部逼近于 0 的情况。上支路对各通道输入信号进行自适应选择,可以更灵活配合下支路消除噪声,提升系统的去噪能力和适应能力。

1.2 变步长自适应算法

传统 GSC 的自适应算法一般采用 LMS 算法^[5]或 NLMS 算法^[6], LMS 算法步长因子 μ 为固定值。若 μ 取值较大,算法收敛速度快,稳态误差大;若 μ 取值较小,算法收敛速度慢,稳态误差也较小。LMS 算法每次迭代梯度绝对值大小与输入信号幅值正相关,当自适应滤波器输入信号非平稳时权值收敛很不稳定。NLMS 算法选择了归一化步长因子,步长因子和滤波器输入信号短时能量的倒数成正比,于是参数 μ 变为

$$\mu = \frac{\alpha}{\delta + \|x\|_2}$$

其中 δ 为一个大于 0 的常数,防止除数为 0, x 为输入信号, α 为调节因子。NLMS 算法可以使权值迭代更为平稳。

NLMS 算法仅考虑了输入信号的非平稳性。在实际的噪声对消系统中,自适应对消器期待信号中不仅有需抵消的噪声,还包含目标语音,对于噪声对消器来说这时目标语音就是一种干扰,会导致自适应滤波器性能的恶化,甚至权系数的发散。针对此问题,提出一种变步长策略,当目标语音很强时,即输入信号中信噪比较高时,冻结对消器权值的调整,仅进行滤波和噪声对消;当目标语音较弱时,即信噪比较低时,对滤波器权值进行自适应调整,学习上下支路之间噪声或干扰的相关性。据此本文提出根据信噪比动态调整步长因子的思路,信噪比较高时,减小步长因子,信噪比较低时,增大步长因子。由于直接估计信噪比较困难,本文用相干函数代替信噪比进行步长控制。

对于一个输出端有噪声的单输入单输出系统,输入信号 $x(t)$ 与观测信号 $y(t)$ 的相干函数定义为

$$\gamma_{xy}^2(\omega) = \frac{|G_{xy}(\omega)|^2}{G_x(\omega) G_y(\omega)}$$

其中 $G_x(\omega)$ 为输入信号的自功率谱, $G_y(\omega)$ 为观测信号的自功率谱, $G_{xy}(\omega)$ 为 $x(t)$ 和 $y(t)$ 的互功率谱。设

观测信号 $Y(\omega)=H(\omega)X(\omega)+W(\omega)$, 当 $x(t)$ 和噪声 $w(t)$ 不相关时可以推导出^[7]:

$$\gamma_{XY}^2(\omega)=\frac{|H(\omega)|^2G_X(\omega)}{|H(\omega)|^2G_X(\omega)+G_W(\omega)}=\frac{\xi(\omega)}{\xi(\omega)+1}$$

其中 $\xi(\omega)=\frac{G_X(\omega)}{G_W(\omega)}$ 定义为谱信噪比。可见 $\gamma_{XY}^2(\omega)$ 和 $\xi(\omega)$ 有对应函数关系, $\gamma_{XY}^2(\omega)$ 大则此时谱信噪比也大, $\gamma_{XY}^2(\omega)$ 小此时谱信噪比也小, 所以可以用相干函数代替信噪比进行步长控制。本文变步长思路是让步长因子与 $\xi(\omega)$ 成反比, 最终采用的变步长因子为

$$\mu=\frac{\alpha}{\|X\|_1+\gamma_{XY}^2(\omega)+\delta}$$

其中 α 是算法总体步长因子, $\|X\|_1$ 是输入信号的 1 范数, 由于 1 范数计算更简单快速, 本文选取 1 范数代替 2 范数进行步长归一化。

本文计算 $\gamma_{XY}^2(\omega)$ 时, 取自适应支路中阻塞矩阵后的输出信号为 X , 由于下支路有 $m-1$ 路信号, 取其平均值作为最后的 X 。取上支路输出信号 y_{n1} 为观测信号 Y 。

1.3 二次消噪

采用频域 GSC 可以消除大部分噪声, 但还会残留部分噪声无法消除。于是进行二次消噪, 将阻塞后的时域噪声信号 $x_b(n)$ 作为输入信号, 频域 GSC 输出信号 $y_{n1}(n)$ 作为期望信号, 采用传统时域 GSC 方法再进行一次噪声抵消得到输出信号 $y_{n2}(n)$ 。同样的, 时域 GSC 时也进行变步长控制, 原理与 1.2 节相同。

当频域 GSC 对消噪声不完全时, 二次消噪可以进一步消除残留噪声, 得到质量更好的语音。但如果频域 GSC 已经比较完美地去除了噪声, 二次消噪可能会造成语音失真, 影响输出语音的质量。所以本文设置一个选择器, 通过自定义的一个语音质量评价指标对

每一帧的输出信号 y_{n1} 和 y_{n2} 进行比较, 选择质量较好的语音作为最终输出。

本文把对消器输出 y_{n1} 或 y_{n2} 分别与 y_{fbf} 、 y_{anc} 的互相关比值作为选择因子, 它们近似反映了输出信号中目标语音和噪声的相对大小。

$$\gamma_1=\frac{R(y_{n1},y_{\text{fbf}})}{R(y_{n1},y_{\text{anc}})}$$
$$\gamma_2=\frac{R(y_{n2},y_{\text{fbf}})}{R(y_{n2},y_{\text{anc}})}$$

对于每一帧的输出信号, 比较 γ_1 和 γ_2 , 选择比值较大的输出信号作为最终系统的输出。

2 仿真分析

为分析验证本文算法的性能, 选取经典的时域广义旁瓣对消算法 (GSC)、基于后置滤波的广义旁瓣对消算法^[3]和本文算法进行仿真比较。

仿真环境设置: 麦克风数量为 8 的均匀线性阵列, 相邻麦克风之间的距离为 0.03 m; 信号采样率 16000 Hz, 对语音分帧处理, 帧长为 400, 帧移为 100。目标信号为长约 10 s 的一段纯净语音, 入射角度设为 10°。采用信干噪比 (SINR)、语音质量感知评价 (PESQ)、短时客观可懂度 (STOI) 和语音源失真比 (SDR) 作为语音增强算法的性能指标。

2.1 考察算法对相干噪声的去噪能力

2.1.1 单个干扰源的实验

设置干扰源为另一人的语音, 入射方向为 40°。

图 3 展示了输入信号信干比为 0 dB 时, 3 种 GSC 算法进行语音增强后得到的波形图。

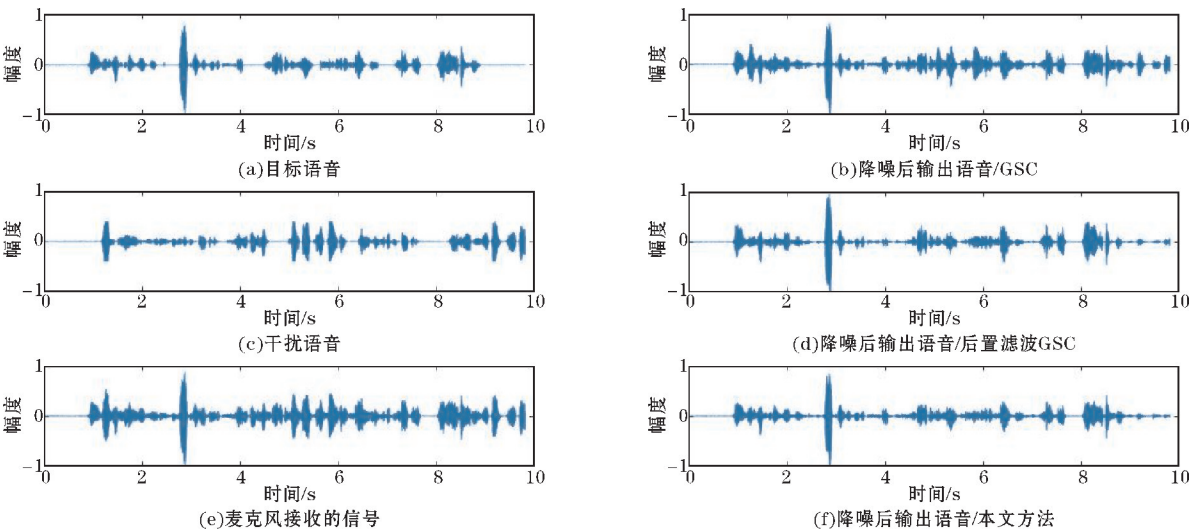


图 3 各种语音增强算法表现

从图 3 可以看出,带后置滤波的 GSC 算法去除干扰的效果比传统 GSC 好,但也带来一定程度的目标语音失真。而本文算法在去除噪声的同时,能够尽量避免目标信号失真,输出信号质量更好。

表 1 单源干扰情况下 3 种算法的性能

算法	SINR			PESQ			STOI			SDR		
	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
传统 GSC	4.0446	8.4969	12.1808	2.1763	2.4633	2.7730	0.7291	0.7996	0.8499	3.6708	8.5849	13.4081
后置滤波	6.7429	8.3418	9.3316	2.1176	2.3672	2.6100	0.7349	0.7963	0.8377	7.9186	10.6898	12.6531
本文算法	12.0256	14.9975	16.7350	2.7626	3.0719	3.2729	0.8210	0.8687	0.8955	12.0823	15.6493	17.9141

从表 1 可以看出,对于单个相干噪声的去除,本文方法在各个指标上都要全面优于其他两种方法,在各种信干比条件下都有较好的表现,而且与传统 GSC 和后置滤波 GSC 相比有较大的优势。

2.1.2 多个干扰源的实验

表 2 多个干扰源情况下 3 种算法的性能

算法	SINR			PESQ			STOI			SDR		
	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
传统 GSC	2.5338	7.1474	11.3484	1.9480	2.1992	2.5521	0.6130	0.7148	0.7928	2.0819	7.0341	11.9166
后置滤波	5.3434	7.6172	9.0789	1.9325	2.2104	2.4584	0.6484	0.7381	0.8068	5.4277	9.0468	11.7504
本文算法	4.7509	8.9472	12.8819	2.1094	2.3869	2.6777	0.6395	0.7314	0.8005	4.5711	8.8877	13.0408

对比表 1 和表 2 可以发现,相比单源干扰,存在多个干扰源时,3 种算法去噪能力均有所下降。就 PESQ 指标看,本文算法优于其他两种算法;STOI 指标上,本文算法与后置滤波 GSC 比较接近,优于传统 GSC 算法;就 SINR 和 SDR 指标来看,在输入信干比较低时,后置滤波算法表现最好,但随着输入信干比提高,本文

在上述仿真环境下,选取多组语音进行实验,输入信干比分别设为 0 dB、5 dB、10 dB,统计仿真实验中各评价指标的平均值,结果如表 1 所示。

设置 3 个干扰源,分别为 3 人的语音,入射角度分别为-20°、50°和 80°,假设 3 路干扰源具有相同的功率,总的输入信干比分别取 0 dB、5 dB、10 dB。选取多组语音完成实验,结果如表 2 所示。

算法逐渐优于其他两种算法。

2.2 考察算法对非相干噪声的去噪能力

为每个麦克风叠加白噪声,各麦克风之间的噪声不相干,信噪比分别设为 0 dB、5 dB、10 dB,3 种算法的实验结果如表 3 所示。

表 3 非相干噪声情况下 3 种算法的性能

算法	SINR			PESQ			STOI			SDR		
	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
传统 GSC	5.2259	9.5326	12.9539	2.1972	2.5719	2.9419	0.7044	0.7718	0.8257	4.9227	9.7934	14.6360
后置滤波	5.6500	7.5779	8.9940	2.1669	2.4540	2.7411	0.6496	0.7308	0.7882	7.4373	10.2140	12.4698
本文算法	6.0807	10.5771	14.3218	2.2449	2.6223	3.0206	0.7237	0.7833	0.8315	5.8284	10.4370	14.8178

从表 3 可以发现,在去除不相干的白噪声时,低信噪比情况下本文方法的 SDR 低于后置滤波 GSC 方法,但输入信噪比提高到 5 dB 以上后表现较好,超过后置滤波 GSC 算法,且在 SINR、PESQ 和 STOI 指标上优于其他两种方法。

2.3 考察算法在复杂场景下的去噪能力

表 4 不同算法在复杂场景下的表现

算法	SINR			PESQ			STOI			SDR		
	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
传统 GSC	3.7970	6.4267	12.3184	2.1509	2.5345	2.9052	0.6972	0.7614	0.8095	4.7171	9.8365	13.6598
后置滤波	5.8362	6.4207	9.2212	2.1397	2.4727	2.6937	0.6512	0.7380	0.8004	7.5215	10.5556	12.5513
本文算法	4.3275	6.8278	13.5849	2.2287	2.5929	2.9783	0.7064	0.7750	0.8175	5.5618	10.4320	14.0358

设置两个干扰源,第一个是入射方向为 50°的另一人的语音,信干比固定为 15 dB,第二个是入射方向为-40°的马路噪声,信噪比固定为 20 dB,另外每个麦克风叠加不相干的白噪声,使总的输入信干噪比为 0 dB、5 dB、10 dB,结果如表 4 所示。

从表4可知,面对复杂场景,本文方法的 PESQ 和 STOI 表现要优于其他两种方法,而在 SINR 和 SDR 指标下,在输入信干噪比较低时,本文方法表现不如后置滤波 GSC,但好于传统 GSC;且在输入信干噪比较高时,本文方法表现最好。

3 结束语

在传统 GSC 的基础上,结合时域 GSC 和频域 GSC,提出一种改进的 GSC 语音增强算法,通过步长控制与选择滤波器,进一步提高了系统的去噪能力。通过实验仿真可以看出,本文方法在消除相干干扰时表现优秀,特别对于单个相干干扰能够有效地去除;在面对多个干扰或者复杂环境时,在 PESQ 和 STOI 指标上表现优秀,对于 SINR 和 SDR 指标,如果输入信干噪比较低,表现不如后置滤波算法,但在高输入信干噪比时表现优于后置滤波算法。在一些实际应用场景中,如室内会议系统等,其存在的非相干噪声较小,干扰源个数有限,本文方法有较大的应用空间。本文方法消除非相干噪声的能力还有所欠缺,且采用时域频域二次滤波也加大了计算量。针对这些问题,后续将继续开展研究,期望提出更多有效的解决办法。

参考文献:

[1] 葛世超. 实时语音处理实践指南[M]. 北京:电子工业出版社,2020.

[2] L. Griffiths, C. Jim. An alternative approach to linearly constrained adaptive beamforming[J]. IEEE Transactions on Antennas and Propagation, 1982, 30(1):27-34.

[3] Zelinski R. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms[C]. ICASSP-88. International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1988.

[4] 刘力玮. 基于 GSC 结构的麦克风阵列语音增强技术研究[D]. 南京:南京信息工程大学,2021.

[5] 宋知用. MATLAB 语音信号分析与合成[M]. 第二版. 北京:北京航空航天大学出版社,2017.

[6] Turan C, Salman M S. A sparse function controlled variable step-size LMS algorithm for system identification[C]. 2014 22nd Signal Processing and Communications Applications Conference (SIU). IEEE, 2014:329-332.

[7] 周文俊,夏秀渝. 一种适用于双讲状态的声学回声消除系统[J]. 数据采集与处理, 2022, 37(2): 437-445.

[8] 张禄. GSC 自适应波束形成的后置滤波算法研究[D]. 哈尔滨:哈尔滨工业大学,2018.

[9] O L Frost. An algorithm for linearly constrained adaptive array processing[J]. Proceedings of the IEEE, 1972, 60(8).

[10] 续娇. 基于自适应波束成形的语音增强算法研究与实现[D]. 北京:北京交通大学,2020.

[11] Wu B, He S. Research on speech enhancement algorithm based on microphone array[C]. International Conference on Green Communication, Network, and Internet of Things (GCNIOT 2021). SPIE, 2021:53-60.

[12] 陈震昊. 基于麦克风阵列的语音增强算法的研究与实现[D]. 南京:南京邮电大学,2021.

An Improved Speech Enhancement Algorithm for GSC Microphone Array

TONG Huan, XIA Xiuyu

(School of Electronic Information, Sichuan University, Chengdu 610064, China)

Abstract: The traditional Generalized Sidelobe Cancellation (GSC) algorithm has poor performance in complex acoustic environments, and the effect of speech enhancement is not ideal. In this regard, an improved GSC algorithm is proposed. In this paper, a system structure based on frequency-domain GSC and time-domain GSC two-stage filtering is constructed, and a selected filter is used to output speech with better quality; Adaptability; and a variable step size adaptive algorithm is proposed, and the step size factor is adjusted in real time according to the Signal to Interference plus Noise Ratio (SINR) to prevent the divergence of the filter weight. The experimental results show that compared with the traditional GSC algorithm, the algorithm in this paper has improved in terms of SINR, PESQ, STOI, and SDR.

Keywords: microphone array; speech enhancement; GSC; variable step size adaptive algorithm