

DOI:10.19651/j.cnki.emt.2212365

# 改进 TF-GSC 和改进后置滤波语音增强算法

杨诗童 杨飞

(武汉大学电气与自动化学院 武汉 430072)

**摘要:** 由于声学环境中噪声的复杂性和不确定性,传统的多通道语音增强算法对于噪声的抑制效果不足,从而导致了较差的听觉体验。针对这一问题,提出了一种改进 TF-GSC 和改进后置滤波语音增强算法。算法使用最大似然法得到目标语音信号和噪声信号的功率谱密度,然后使用信号功率谱密度比值得到的变步长归一化最小均方算法来改进 TF-GSC。还提出了联合信号功率谱密度比值和先验信噪比估计语音存在概率的改进最优修正对数幅度谱估计器。不同信噪比环境下的仿真实验表明,本文提出的算法可以有效地滤除相干噪声和非相干噪声,与其他算法相比,增强后的语音信号具有更高的信噪比和语音质量。

**关键词:** 语音增强;传递函数广义旁瓣相消器;最优修正对数幅度谱估计器;最大似然;参数估计

**中图分类号:** TN912.35 **文献标识码:** A **国家标准学科分类代码:** 510.4040

## Improved TF-GSC and improved post filter speech enhancement algorithm

Yang Shitong Yang Fei

(School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China)

**Abstract:** Due to the complexity and uncertainty of noise in acoustic environment, the traditional multichannel speech enhancement algorithm has insufficient noise suppression effect, resulting in a poor auditory experience. To solve this problem, an improved TF-GSC and improved post filter speech enhancement algorithm was proposed in this paper. The algorithm used the maximum likelihood method to obtain the power spectral density of the target speech signal and noise signal, and then proposed an improved TF-GSC which used the variable step normalized least mean square algorithm obtained by the signal power spectral density ratio. An improved optimally modified log spectral spectrum estimator was also proposed using the estimated speech presence probability by combining the signal power spectral density ratio and a priori signal to noise ratio. The simulation experiments in different SNR environments show that the algorithm proposed in this paper can effectively filter coherent noise and incoherent noise. Compared with other algorithms, the enhanced speech signal has higher SNR and speech quality.

**Keywords:** speech enhancement; transfer function generalized sidelobe canceller; optimally modified log spectral amplitude estimator; maximum likelihood; parameter estimation

## 0 引言

相对于单通道语音增强算法,基于麦克风阵列的多通道语音增强算法可以利用麦克风阵列的空间信息来增强特定的语音信号,去除大部分的相干噪声和非相干噪声,从而有效地提高语音信号的信噪比和语音质量<sup>[1-3]</sup>。自适应波束形成器会根据麦克风阵列采集的实时语音数据自适应地调整滤波器加权系数,因此能够更好地抑制各种类型的噪声<sup>[4-6]</sup>。典型的自适应波束形成器主要包括最小方差无失真响应(minimum variance distortionless response, MVDR)波束形成器<sup>[6]</sup>、线性约束最小方差(linearly constrained minimum variance, LCMV)波束形成器<sup>[7]</sup>和广

义旁瓣相消器(generalized sidelobe canceller, GSC)<sup>[8]</sup>等。GSC 因为结构简单且容易实现,被广泛地应用在各种多通道语音增强算法中。GSC 主要由三部分组成,分别为固定波束形成器(fixed beamformer, FBF)、阻塞矩阵(blocking matrix, BM)和自适应噪声消除器(adaptive noise canceller, ANC)。李斌等<sup>[9]</sup>提出一种通过自适应调整阻塞矩阵以减少目标语音泄露的改进 GSC。Priyanka 等<sup>[10]</sup>提出了采用快速归一化最小均方算法的广义旁瓣相消器,在各种噪声下可以快速收敛并且复杂度较低。Barnov 等<sup>[11]</sup>提出了改进的递归最小二乘算法来实现 GSC 的 ANC 模块。

收稿日期:2022-12-11

自适应波束形成器可以有效地抑制相干噪声,增强特定的语音信号,但是对于非相干噪声的抑制能力有限。因此,通常在自适应波束形成器后使用后置滤波技术来消除残余噪声。Zhang 等<sup>[12]</sup>将传递函数广义旁瓣相消器 (transfer function generalized sidelobe canceller, TF-GSC) 与维纳滤波器相结合,使用后置维纳滤波器来消除 TF-GSC 增强后语音信号的残留背景噪声。Thüne 等<sup>[13]</sup>提出了一种基于最大似然法和贝叶斯改进的噪声功率谱密度估计算法,并应用于多通道的维纳滤波,取得了较好地语音增强效果。倪峰等<sup>[14]</sup>对经过 GSC 增强后的语音使用深度神经网络和最优修正对数幅度谱 (optimally modified log spectral amplitude, OMLSA) 估计器相结合的后置滤波算法来消除残留的背景噪声。

由于传统 TF-GSC 和后置滤波不能无失真地去除各种噪声,本文提出了改进 TF-GSC 和改进后置滤波语音增强算法。该算法使用最大似然法估计出目标语音信号和噪声信号的功率谱密度,然后使用参考麦克风接收到语音信号的功率谱密度和目标语音信号的功率谱密度得到信号功率谱密度比值。本文提出了使用信号功率谱密度比值得到的变步长归一化最小均方算法,用来改进 TF-GSC 的自适应噪声消除器。并且本文还提出了联合信号功率谱密度比值和先验信噪比估计语音存在概率的改进最优修正对数幅度谱估计器。经仿真实验表明,该算法可以有效地提高增强后语音信号的信噪比和语音质量。

## 1 信号模型及传统算法

### 1.1 语音信号模型

在含有噪声的声学环境中,假设有  $M$  个麦克风组成一个麦克风阵列,麦克风阵列接收到的信号为:

$$x_m(t) = a_m(t) * s(t) + n_m(t); m = 1, \dots, M \quad (1)$$

其中,  $x_m(t)$  表示第  $m$  个麦克风接收到的信号,  $a_m(t)$  表示纯净语音信号到第  $m$  个麦克风的传递函数,  $*$  表示卷积,  $s(t)$  表示纯净语音信号,  $n_m(t)$  表示第  $m$  个麦克风接收到的噪声信号,包括相干噪声和非相干噪声。对上式进行短时傅里叶变换并表示为向量形式,可以得到:

$$\mathbf{X}(k, l) = \mathbf{A}(k) S(k, l) + \mathbf{N}(k, l) \quad (2)$$

其中,  $k$  和  $l$  分别为频率索引和时间索引,  $\mathbf{A}(k)$  为时不变的传递函数。以麦克风阵列的第一个麦克风为参考麦克风,将参考麦克风接收到的纯净语音信号作为目标语音信号,则上述声学模型可以用相对传递函数模型表示为:

$$\mathbf{X}(k, l) = \mathbf{H}(k) A_1(k) S(k, l) + \mathbf{N}(k, l) \quad (3)$$

### 1.2 TF-GSC 算法

TF-GSC 是在传统 GSC 基础上,使用相对传递函数重新构造固定波束形成器和阻塞矩阵,从而能够更好地消除噪声。TF-GSC 是由上下两个通道组成,上方通道使用固定波束形成器得到一个目标语音无失真的信号,下方通道先使用阻塞矩阵将目标语音滤除,得到只包含噪声的参考

信号,然后使用自适应噪声消除器对固定波束形成器的输出信号和阻塞矩阵的输出信号进行自适应滤波,消除固定波束形成器输出信号中的残余噪声。

固定波束形成器通常使用延时求和波束形成器,其输出信号可以表示为:

$$Y_{\text{FBF}}(k, l) = \mathbf{W}^H(k, l) \mathbf{X}(k, l) = \frac{\mathbf{H}^H(k)}{\|\mathbf{H}(k)\|^2} \mathbf{X}(k, l) \quad (4)$$

其中,上标  $H$  表示向量的共轭转置,  $\|\cdot\|$  表示向量的二范数。噪声参考信号可以通过使用相对传递函数对齐的两个麦克风信号相减得到。因此,维度为  $M \times (M-1)$  的阻塞矩阵可以表示为:

$$\mathbf{B}(k) = \begin{bmatrix} -H_2^*(k) & -H_3^*(k) & \dots & -H_M^*(k) \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (5)$$

其中,上标  $*$  表示复数的共轭。因此,阻塞矩阵的输出为:

$$\mathbf{U}(k, l) = \mathbf{B}^H(k, l) \mathbf{X}(k, l) \quad (6)$$

自适应噪声消除器使用归一化最小均方算法得到,其滤波器系数的更新公式为:

$$G_m(k, l+1) = G_m(k, l) + \mu \frac{U_m(k, l) Y^*(k, l)}{P(k, l) + \epsilon}; m = 2, \dots, M \quad (7)$$

$$P(k, l) = \rho P(k, l-1) + (1-\rho) \|\mathbf{U}(k, l)\|^2 \quad (8)$$

其中,  $\mu$  为迭代步长,  $U_m(k, l)$  为阻塞矩阵的输出信号,  $\epsilon$  为防止分母为零的极小值,  $\rho$  为遗忘因子,  $Y(k, l)$  为 TF-GSC 的输出信号,即:

$$Y(k, l) = Y_{\text{FBF}}(k, l) - Y_{\text{ANC}}(k, l) = \mathbf{W}^H(k, l) \mathbf{X}(k, l) - \mathbf{G}^H(k, l) \mathbf{B}^H(k, l) \mathbf{X}(k, l) \quad (9)$$

### 1.3 后置 OMLSA 算法

后置 OMLSA 算法是在 TF-GSC 算法之后进行的后置滤波,通过对 TF-GSC 的输出信号进行单通道的后置滤波,可以有效地消除残余噪声。OMLSA 是对考虑了语音存在不确定性的对数幅度谱估计器的最优修正,其增益函数为:

$$G_{\text{OMLSA}}(k, l) = G_H(k, l)^{p(k, l)} \cdot G_{\min}^{1-p(k, l)} \quad (10)$$

其中,  $p(k, l)$  为语音存在概率,  $G_{\min}$  为语音不存在时增益的下限约束,  $G_H(k, l)$  为语音存在时对数幅度谱估计器的增益函数,  $G_H(k, l)$  是通过最小化目标语音信号对数幅度谱和估计信号对数幅度谱误差的平方得到的,其表达式为:

$$G_H(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \exp\left(\frac{1}{2} \int_{v(k, l)}^{\infty} \frac{e^{-x}}{x} dx\right) \quad (11)$$

$$v(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \gamma(k, l) \quad (12)$$

其中,  $\xi(k, l)$  为先验信噪比,  $\gamma(k, l)$  为后验信噪比。

因此, 经过 TF-GSC 算法和后置 OMLSA 算法之后得到的目标语音信号的估计信号可以表示为:

$$\hat{S}(k, l) = G_{\text{OMLSA}}(k, l)Y(k, l) \quad (13)$$

## 2 本文提出的算法

本文在上述算法的基础上, 提出了改进 TF-GSC 和改

进后置滤波语音增强算法。算法先对麦克风阵列接收到的语音信号使用最大似然法得到目标语音信号和噪声信号的功率谱密度的估计值, 然后使用信号功率谱密度比值得到变步长归一化最小均方算法来改进 TF-GSC。并且本文还提出了联合信号功率谱密度比值和先验信噪比估计语音存在概率的改进后置 OMLSA。本文提出的改进 TF-GSC 和改进后置滤波语音增强算法的结构框图如图 1 所示。

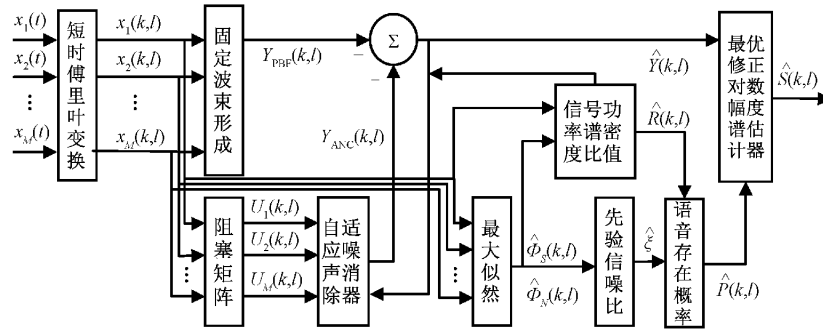


图 1 本文提出的算法结构框图

### 2.1 基于最大似然的参数估计

由于目标语音信号与噪声信号相互独立, 麦克风阵列接收到的语音信号的协方差矩阵可以表示为:

$$\Phi_X(k, l) = E\{X(k, l)X(k, l)^H\} = \Phi_S(k, l) \cdot H(k)H(k)^H + \Phi_N(k, l) \cdot \Gamma(k) \quad (14)$$

其中,  $E\{\cdot\}$  表示期望,  $\Phi_S(k, l)$  是目标语音信号的功率谱密度, 噪声信号的协方差矩阵是由两部分组成, 分别是时变的噪声功率谱密度  $\Phi_N(k, l)$  和时不变的噪声场空间相干矩阵  $\Gamma(k)$ ,  $\Gamma(k)$  可以提前在没有目标语音的声学环境中得到或者在目标语音暂停期间得到。

为了简洁起见, 后面的部分公式将省略频率索引和时间索引。假设麦克风阵列接收到的语音信号  $X$  为零均值的循环对称复高斯模型, 其协方差矩阵为  $\Phi_X$ 。则麦克风阵列接收到的语音信号  $X$  的对数似然函数可以表示为:

$$\mathcal{L}(X | \Phi_S, \Phi_N) = \ln(p(X | \Phi_S, \Phi_N)) = -M \ln(\pi) - \ln \det \Phi_X - X^H \Phi_X^{-1} X \quad (15)$$

使用最大似然法来联合估计目标语音信号和噪声信号的功率谱密度, 即最大化式(15)中的对数似然函数。因此, 可以得到:

$$\frac{\partial \mathcal{L}(X | \Phi_S, \Phi_N)}{\partial \Phi_S} = \text{tr} \left( \frac{\partial \mathcal{L}(X | \Phi_S, \Phi_N)}{\partial \Phi_X} \cdot \frac{\partial \Phi_X}{\partial \Phi_S} \right) = 0 \quad (16)$$

$$\frac{\partial \mathcal{L}(X | \Phi_S, \Phi_N)}{\partial \Phi_N} = \text{tr} \left( \frac{\partial \mathcal{L}(X | \Phi_S, \Phi_N)}{\partial \Phi_X} \cdot \frac{\partial \Phi_X}{\partial \Phi_N} \right) = 0 \quad (17)$$

其中,  $\text{tr}(\cdot)$  表示矩阵的迹。对式(15)求导数可以得到:

$$\frac{\partial \mathcal{L}(X | \Phi_S, \Phi_N)}{\partial \Phi_X} = -\frac{\partial \ln \det \Phi_X}{\partial \Phi_X} - \frac{\partial X^H \Phi_X^{-1} X}{\partial \Phi_X} =$$

$$\Phi_X^{-1} X X^H \Phi_X^{-1} - \Phi_X^{-1} \quad (18)$$

对式(14)运用矩阵求逆引理可以得到:

$$\Phi_X^{-1} = \Phi_N^{-1} \left( \Gamma^{-1} - \frac{\Gamma^{-1} H H^H \Gamma^{-1} \Phi_S}{\Phi_N + H^H \Gamma^{-1} H \Phi_S} \right) \quad (19)$$

将式(18)和(19)带入到式(16)和(17)中解方程<sup>[15]</sup>可以得到目标语音信号和噪声信号功率谱密度的最大似然解为:

$$\hat{\Phi}_S(k, l) = \frac{M H^H \Gamma^{-1} X X^H \Gamma^{-1} H}{(M-1) (H^H \Gamma^{-1} H)^2} - \frac{\text{tr}(\Gamma^{-1} X X^H)}{(M-1) (H^H \Gamma^{-1} H)} \quad (20)$$

$$\hat{\Phi}_N(k, l) = \frac{\text{tr}(\Gamma^{-1} X X^H)}{M-1} - \frac{H^H \Gamma^{-1} X X^H \Gamma^{-1} H}{(M-1) (H^H \Gamma^{-1} H)} \quad (21)$$

### 2.2 改进的 TF-GSC 算法

传统的 TF-GSC 中的自适应噪声消除器通常使用固定步长的归一化最小均方算法, 迭代步长  $\mu$  通常为固定的经验值。对于归一化最小均方算法, 迭代步长  $\mu$  决定着算法的收敛性和稳定性。在仅含噪声时,  $\mu$  值应该较大, 以提高算法的收敛速度和跟踪速度; 在含有目标语音时,  $\mu$  值应该较小, 减少语音泄露, 提高算法的稳定性。

因此, 本文提出了一种利用目标语音信号的功率谱密度的改进 TF-GSC 算法。通过最大似然参数估计得到目标语音信号功率谱密度的估计值后, 计算参考麦克风接收语音信号功率谱密度与目标语音信号功率谱密度的比值, 利用信号功率谱密度比值使用软判决的方法得到一种变步长的自适应噪声消除器。首先, 求解参考麦克风接收语音信号的功率谱密度并进行时间递归平均, 可以得到:

$$P_r(k, l) = \alpha P_r(k, l-1) + (1-\alpha) |X_1(k, l)|^2 \quad (22)$$

其中,  $X_1(k, l)$  为参考麦克风接收到的语音信号,  $\alpha$  为时间平滑因子。对目标语音信号功率谱密度估计值进行同样的时间递归平滑可以得到:

$$P_s(k, l) = \alpha P_s(k, l-1) + (1-\alpha)\hat{\Phi}_s(k, l) \quad (23)$$

结合式(22)和(23)可以求得信号功率谱密度比值为:

$$R(k, l) = \frac{P_r(k, l)}{P_s(k, l) + \epsilon} \quad (24)$$

其中,  $\epsilon$  为防止分母为零的极小值。对信号功率谱密度比值进行软判决的方法,可以得到:

$$\mu(k, l) = \mu_f \max(\log_{10}(R(k, l)), 0) \quad (25)$$

其中,  $\mu_f$  为可变步长的调节系数。将式(25)带入到式(7)中,就可以得到变步长归一化最小均方算法。改进后的自适应噪声消除器可以在仅含噪声时快速收敛,可以更好地跟踪噪声的变化;在目标语音出现时,可以降低目标语音对系统的影响,防止目标语音失真,从而使 TF-GSC 更好地抑制噪声。

### 2.3 改进的后置 OMLSA 算法

本文提出利用目标语音信号和噪声信号功率谱密度的改进后置 OMLSA 算法。经过 TF-GSC 初步过滤噪声后,残余噪声信号的功率谱密度可以表示为:

$$\hat{\Phi}_N'(k, l) = (\mathbf{W}^H(k, l) - \mathbf{G}^H(k, l)\mathbf{B}^H(k)) (\mathbf{W}(k, l) - \mathbf{B}(k)\mathbf{G}(k, l)) \hat{\Phi}_N(k, l) \quad (26)$$

因为 TF-GSC 可以几乎无失真的保留目标语音信号,所以目标语音信号的功率谱密度保持不变。因此,经过 TF-GSC 初步增强后语音信号的后验信噪比和先验信噪比分别可以表示为:

$$\hat{\gamma}(k, l) = \frac{|Y(k, l)|^2}{\hat{\Phi}_N'(k, l)} \quad (27)$$

$$\hat{\xi}(k, l) = \frac{\hat{\Phi}_s(k, l)}{\hat{\Phi}_N'(k, l)} \quad (28)$$

语音存在概率是 OMLSA 算法中的重要参数,决定着 OMLSA 算法对噪声的抑制程度和目标语音的失真程度。本文提出了联合信号功率谱密度比值和先验信噪比的语音存在概率估计算法。语音存在概率和先验信噪比成正相关的关系,因此使用先验信噪比的 S 型函数来估计语音存在概率,即:

$$p_s(k, l) = 1 / (1 + \exp(-\beta(\hat{\xi}(k, l) - 1))) \quad (29)$$

其中,  $\beta$  为 S 型函数的调节参数。图 2 为不同  $\beta$  值语音存在概率和先验信噪比的关系曲线。由图可知,  $\beta$  值越大,曲线越陡峭,语音存在概率随先验信噪比的变化越快。当先验信噪比趋近于正无穷时,语音存在概率趋近于 1,当先验信噪比趋近于负无穷时,语音存在概率趋近于 0。

语音存在概率和信号功率谱密度比值成负相关的关系,因此使用类似 S 型函数的形式来估计语音存在概率,即:

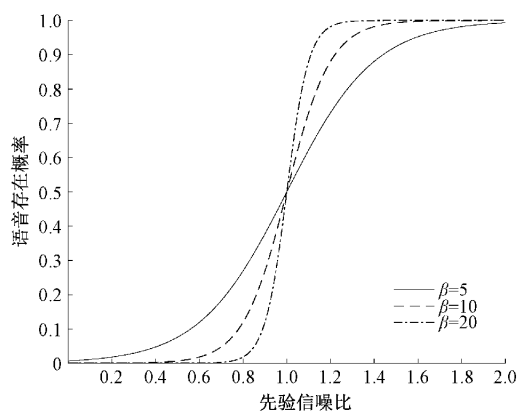


图 2 语音存在概率和先验信噪比的曲线图

$$p_r(k, l) = 1 - 1 / (1 + \exp(-\beta(R(k, l) - 5))) \quad (30)$$

然后,使用一种判决方法联合先验信噪比估计的语音存在概率和信号功率谱密度比值估计的语音存在概率来得到语音存在概率,即:

$$p(k, l) = \begin{cases} \min(p_s(k, l), p_r(k, l)), & p_s(k, l) \leq 0.5 \\ \max(p_s(k, l), p_r(k, l)), & p_s(k, l) > 0.5 \end{cases} \quad (31)$$

这样可以在目标语音出现时更好地保留目标语音信号,减少语音失真;在含有噪声时,能够更好地抑制噪声,提高信噪比。

最后,对得到的语音存在概率进行时域平滑,可以得到:

$$\hat{p}(k, l) = \alpha \hat{p}(k, l-1) + (1-\alpha)p(k, l) \quad (32)$$

将后验信噪比、先验信噪比和估计出的语音存在概率带入到式(10)~(13)中,就可以求得 OMLSA 算法的增益函数,进而可以得到目标语音信号的估计值。

### 3 仿真实验

本文在 MATLAB 上进行了仿真实验来验证本文提出的改进 TF-GSC 和改进后置 OMLSA 语音增强算法的性能,并与传统的 TF-GSC、文献[13]提出的基于最大似然的多通道维纳滤波算法和本文提出的改进 TF-GSC 加后置 OMLSA 算法进行对比。仿真实验中采用由 9 个麦克风组成的均匀线性麦克风阵列,相邻麦克风之间的距离设置为 5 cm。设置最左侧的麦克风为参考麦克风,目标语音信号与麦克风阵列垂直方向的夹角为  $45^\circ$ ,噪声信号与垂直方向的夹角为  $30^\circ$ ,如图 3 所示。目标语音信号选自 TIMIT 语音库,噪声信号使用 NOISEX-92 噪声数据库中的 babble 噪声来模拟相干噪声,此外还加入了相同信噪比的非相干噪声。实验中的采样频率为 8 kHz,短时傅里叶变换使用的是 Hamming 窗函数,帧长为 256,帧移为 64。下面将通过对比不同算法增强后语音信号的语音评价标准、波形图和语谱图来比较不同算法的性能。

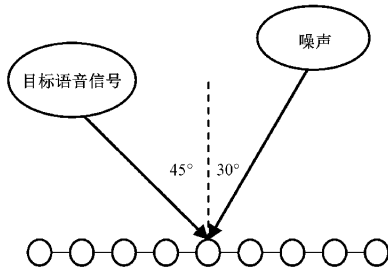


图 3 麦克风阵列示意图

3.1 客观评价标准

为了对比本文提出的算法和其他算法的性能,本节实验使用了 3 种客观评价标准在不同信噪比环境下做了对比实验,分别是分段信噪比(segmental signal noise ratio,

SegSNR)、语音质量感知评估(perceptual evaluation of speech quality, PESQ)和短时客观可懂度(short-time objective intelligibility, STOI)。

分段信噪比是计算所有含有语音信号的时间帧内信噪比的平均值,因此可以反应出语音信号的信噪比水平。本文提出的算法和其他算法增强后语音信号的分段信噪比如表 1 所示。从表中可以看出,本文提出的改进 TF-GSC 的分段信噪比高于传统的 TF-GSC。本文提出的改进后置 OMLSA 的分段信噪比也都高于传统的后置 OMLSA。在输入信噪比较高时,本文提出的算法的分段信噪比略低于文献[13]提出的 ML-MWF,但在低输入信噪比时,本文提出的算法较高。说明本文提出的算法在低信噪比时有更好的噪声抑制能力。

表 1 不同算法的分段信噪比

SNR/dB	TF-GSC	Proposed-TF-GSC	ML-MWF	TF-GSC-OMLSA	Proposed-TF-GSC-OMLSA
12.031 3	15.909 9	23.159 3	25.570 9	23.714 6	24.569 0
7.018 7	14.386 4	18.611 0	20.980 1	20.125 2	20.502 2
2.035 1	11.637 7	13.947 0	16.568 9	16.292 2	16.335 2
-2.944 7	8.111 4	9.268 6	12.399 2	12.601 1	12.637 0
-7.994 6	3.949 6	4.433 5	7.962 8	9.084 4	9.112 2

PESQ 是评价增强后语音信号质量的一项客观评价指标,其值通常在 1.0~4.5,值越大表明增强后语音信号的质量越高。表 2 为本文提出的算法和其他算法增强后语音信号的 PESQ。从表中可以看出,在输入信噪比较高时,本文提出的改进 TF-GSC 的 PESQ 远高于传统的 TF-GSC,本文提出的改

进后置 OMLSA 也远高于文献[13]提出的 ML-MWF 和传统的后置 OMLSA;在输入信噪比较低时,本文提出的改进 TF-GSC 和改进后置 OMLSA 也高于传统的 TF-GSC、文献[13]提出的 ML-MWF 和传统的后置 OMLSA。说明本文提出的算法增强后的语音信号有更高的质量。

表 2 不同算法的 PESQ

SNR/dB	TF-GSC	Proposed-TF-GSC	ML-MWF	TF-GSC-OMLSA	Proposed-TF-GSC-OMLSA
12.031 3	2.695 0	3.043 3	3.459 7	3.450 3	3.664 7
7.018 7	2.511 2	2.687 2	3.162 5	3.133 6	3.369 4
2.035 1	2.252 7	2.372 8	2.792 9	2.926 8	3.033 0
-2.944 7	1.933 4	2.002 6	2.387 4	2.525 3	2.577 4
-7.994 6	1.581 5	1.621 1	2.047 9	2.097 2	2.194 3

STOI 是评价增强后语音信号可懂度的一项客观评价指标,其值的范围在 0~1,值越大表明增强后语音信号的可懂度越好。表 3 为本文提出的算法和其他算法增强后语音信号的 STOI。从表中可以看出,本文提出的改进 TF-

GSC 的 STOI 明显高于传统的 TF-GSC,本文提出的改进后置 OMLSA 也略高于文献[13]提出的 ML-MWF 和传统的后置 OMLSA,说明本文提出的算法增强后的语音信号具有更高的可懂度。

表 3 不同算法的 STOI

SNR/dB	TF-GSC	Proposed-TF-GSC	ML-MWF	TF-GSC-OMLSA	Proposed-TF-GSC-OMLSA
12.031 3	0.942 3	0.964 2	0.971 7	0.971 6	0.972 9
7.018 7	0.898 2	0.924 0	0.937 7	0.937 1	0.941 5
2.035 1	0.838 1	0.855 6	0.877 7	0.879 8	0.888 9
-2.944 7	0.737 9	0.758 9	0.779 0	0.784 4	0.796 4
-8.031 7	0.642 6	0.655 0	0.677 6	0.682 7	0.686 2

### 3.2 波形图和语谱图分析

实验还对比了在 7.018 7 dB 的输入信噪比下不同算法增强后的波形图和语谱图,如图 4 所示。其中,图 4(a)和(b)分别为参考麦克风接收到的目标语音信号和含噪语音信号,图 4(c)为传统 TF-GSC 增强后的语音信号,图 4(d)为本文提出的改进 TF-GSC 增强后的语音信号,图 4(e)为文献[13]提出的 ML-MWF,图 4(f)为改进 TF-GSC 加传统后置 OMLSA 增强后的语音信号,图 4(g)为本文提出的改进 TF-GSC 和改进后置 OMLSA 增强后的语音信号。从图中可以看出,传统的 TF-GSC 增强后语音信号在低频时有更多的残留噪声,而且在仅含噪声时由于系统跟踪较慢而增加噪声;相反改进的 TF-GSC 残留噪声更少。但是可以看到没有后置滤波的 TF-GSC 增强后的语音信号含有较多的非相干噪声,后置 OMLSA 可以滤除大部分的非相干噪声。文献[13]提出的 ML-MWF 和传统的后置 OMLSA 增强后语音信号仍然会有一些残留噪

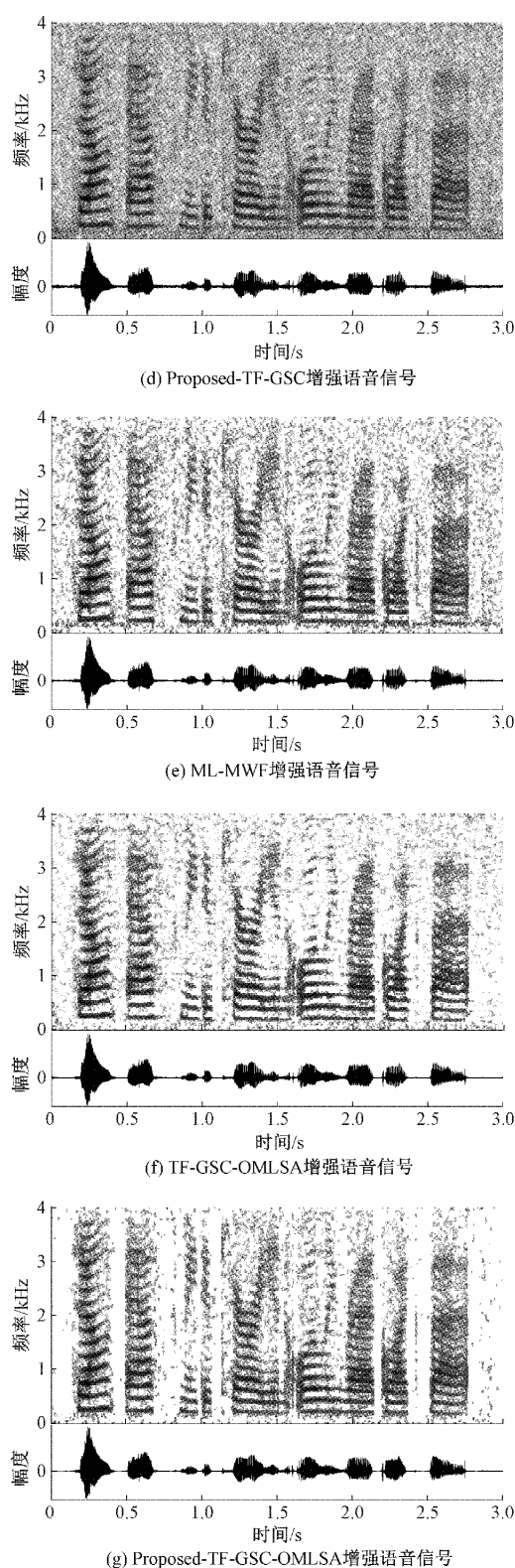
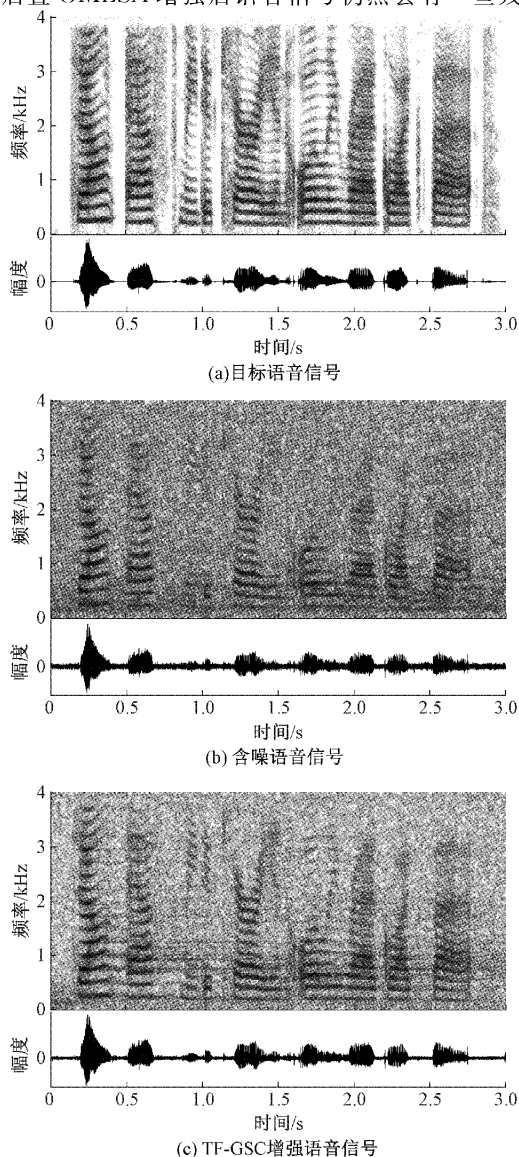


图 4 不同算法的波形图和语谱图

声,部分低频频谱区域语音信号明显减弱,本文提出的改进后置 OMLSA 增强后的语音信号残留噪声更少,可以更好地保留语音信号。

## 4 结 论

针对存在相干噪声和非相干噪声的声学环境,本文提出了利用信号功率谱密度比值的改进 TF-GSC 和联合信号功率谱密度比值和先验信噪比估计语音存在概率的改进后置 OMLSA,可以更好地消除噪声。经过仿真实验表明,本文提出的改进 TF-GSC 和改进 OMLSA 可以有效地去除相干噪声和非相干噪声,而且相比于其他算法残留噪声更少,增强后的语音信号失真更小,具有更高的信噪比和语音质量。

## 参考文献

- [1] GANNOT S, VINCENT E, MARKOVICH-GOLAN S, et al. A consolidated perspective on multimicrophone speech enhancement and source separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(4): 692-730.
- [2] DOCLO S, KELLERMANN W, MAKINO S, et al. Multichannel signal enhancement algorithms for assisted listening devices: exploiting spatial diversity using multiple microphones [J]. IEEE Signal Processing Magazine, 2015, 32(2): 18-30.
- [3] 徐中明, 李怡, 张志飞, 等. 弹性网正则化广义逆波束形成算法改进[J]. 仪器仪表学报, 2021, 42(6): 243-252.
- [4] PRIVANKA S S. A review on adaptive beamforming techniques for speech enhancement[C]. Proceedings of Innovations in Power and Advanced Computing Technologies(i-PACT). Vellore, India: IEEE Press, 2017: 1-6.
- [5] 郝旺身, 刘雨曦, 冀科伟, 等. 基于空间平滑差分的鲁棒自适应波束形成[J]. 电子测量技术, 2022, 45(9): 50-55.
- [6] 郭翔宇, 鄢社锋, 王文侠. 基于迭代梯度方法的线性约束稳健 Capon 波束形成快速算法[J]. 信号处理, 2021, 37(5): 712-723.
- [7] 郭业才, 陈小燕, 王超. LCMV 分频的改进维纳滤波后置波束形成算法[J]. 电子测量与仪器学报, 2017, 31(10): 1646-1652.
- [8] DAI SH W, LI M H, ABBASI Q H, et al. A fast blocking matrix generating algorithm for generalized sidelobe canceller beamformer in high speed rail like scenario[J]. IEEE Sensors Journal, 2021, 21(14): 15775-15783.
- [9] 李斌, 张玲华. 一种广义旁瓣抵消器结构的语音增强改进算法[J]. 数据采集与处理, 2017, 32(2): 307-313.
- [10] PRIYANKA S S, KUMAR T K. GSC adaptive beamforming using fast NLMS algorithm for speech enhancement[C]. Proceedings of the 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC), Singapore: IEEE Press, 2019: 160-165.
- [11] BARNOV A, GENDELMAN A, SCHREIBMAN A, et al. A robust RLS implementation of the ANC block in GSC structures [C]. Proceedings of the 29th European Signal Processing Conference (EUSIPCO). Dublin, Ireland: IEEE Press, 2021: 261-265.
- [12] ZHANG M, WU SH H, GUO W X, et al. A microphone array dereverberation algorithm based on TF-GSC and postfiltering[C]. Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). Nara, Japan: IEEE Press, 2016: 1-4.
- [13] THÜNE P, ENZNER G. Maximum-Likelihood approach with bayesian refinement for multichannel-wiener postfiltering[J]. IEEE Transactions on Signal Processing, 2017, 65(13): 3399-3413.
- [14] 倪峰, 周翊, 刘宏清. 利用信号功率谱密度的麦克风阵列噪声消除方法[J]. 信号处理, 2020, 36(3): 373-381.
- [15] THUENE P, ENZNER G. Maximum-Likelihood approach to adaptive multichannel-wiener postfiltering for wind-noise reduction[C]. Proceedings of Speech Communication; 12. ITG Symposium. Paderborn, Germany: VDE Verlag, 2016: 1-5.

## 作者简介

杨诗童, 硕士研究生, 主要研究方向为语音信号处理和麦克风阵列语音增强。

E-mail: 2016301470005@whu.edu.cn

杨飞(通信作者), 博士, 副教授, 主要研究方向为机器视觉、智能传感器与检测技术和信号处理。

E-mail: f.yang@whu.edu.cn