

# 复广义高斯分布多通道最大似然联合去噪 去混响波束形成器

孟维鑫<sup>1,2</sup> 厉 剑<sup>1,2</sup> 郑成诗<sup>1,2</sup> 李晓东<sup>1,2</sup>

(1. 中国科学院声学研究所, 北京 100190; 2. 中国科学院大学, 北京 100049)

**摘 要:** 提出了一种基于复超高斯分布的多通道联合去噪去混响波束形成器。本文采用复超高斯模型对语音信号建模,在最大似然准则下首次推导出联合去噪去混响波束形成器的解析表达式,并证明了该式是现有多种联合去噪去混响波束形成器的一般化形式。同时通过理论推导证明本文所提算法性能优于传统多通道预测误差算法级联最小功率无失真波束形成器。仿真实验与实际实验结果均表明,本文提出的算法在多个客观指标上明显优于现有联合去噪去混响算法。

**关键词:** 语音增强; 去噪去混响; 波束形成

**中图分类号:** TN911.7      **文献标识码:** A      **DOI:** 10.16798/j.issn.1003-0530.2022.04.002

**引用格式:** 孟维鑫,厉剑,郑成诗,等. 复广义高斯分布多通道最大似然联合去噪去混响波束形成器[J]. 信号处理, 2022,38(4): 677-689. DOI: 10.16798/j.issn.1003-0530.2022.04.002.

**Reference format:** MENG Weixin, LI Jian, ZHENG Chengshi, et al. Jointly denoising and dereverberation with maximum likelihood beamformer under complex generalized Gaussian distribution [J]. Journal of Signal Processing, 2022, 38(4): 677-689. DOI: 10.16798/j.issn.1003-0530.2022.04.002.

## Jointly Denoising and Dereverberation with Maximum Likelihood Beamformer Under Complex Generalized Gaussian Distribution

MENG Weixin<sup>1,2</sup> LI Jian<sup>1,2</sup> ZHENG Chengshi<sup>1,2</sup> LI Xiaodong<sup>1,2</sup>

(1. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** This paper proposes a jointly denoising and dereverberation beamformer based on a complex super-Gaussian distribution. By modelling speech using a complex super-Gaussian distribution, we first derive the optimal denoising and dereverberation beamformer with a maximum likelihood criterion. The paper further proves that the proposed beamformer can be regarded as a generalized framework of many existing jointly denoising and dereverberation methods and also demonstrates that the proposed beamformer outperforms the weighted prediction error algorithm cascaded minimum power distortionless beamformer theoretically. Simulation results and experimental results show that the proposed beamformer does outperform many state-of-the-art joint denoising and dereverberation algorithms in terms of several objective measurements.

**Key words:** speech enhancement; denoising and dereverberation; beamformer

## 1 引言

在室内应用场景,传声器拾取的语音信号将不可避免地受到室内环境噪声和混响的共同影响,这不仅会造成语音质量和可懂度的下降,同时也会导致自动语音识别(Automatic Speech Recognition, ASR)系统识别率下降。因此,抑制传声器拾取信号中的噪声和混响成分,并提取高质量的目标语音一直是研究热点。

相比于单通道语音增强技术,基于传声器阵列的多通道语音增强技术可以充分利用目标信号与干扰信号的空间统计特性差异及通道间的相关性差异,更好地实现噪声抑制和混响去除<sup>[1-2]</sup>。因此,相比于单通道语音增强技术,多通道语音增强技术一般具有更好的去噪去混响性能。典型的传声器阵列自适应波束形成器包括最小方差无失真响应(Minimum Variance Distortionless Response, MVDR)波束形成器<sup>[3-4]</sup>、最小功率无失真响应波束形成器(Minimum Power Distortionless Response, MPDR)、广义旁瓣消除器(Generalized Sidelobe Cancellation, GSC)<sup>[5]</sup>、多通道维纳滤波(Multichannel Wiener filtering, MWF)<sup>[6]</sup>和最大似然无失真响应(Maximum Likelihood Distortionless Response, MLDR)<sup>[7]</sup>波束形成器等。在弱混响环境下,自适应波束形成器对方向性干扰噪声具有较好的抑制作用;但在强混响条件下,其抑制性能会有不同程度的下降。这是由于混响是由方向性干扰噪声多次反射形成,且随着混响增加,方向性干扰噪声将逐渐退化为各向同性均匀分布的扩散噪声,因而通过自适应零陷设计的方法的性能会有不同程度的下降。针对去混响任务,许多学者提出了多种经典算法,包括基于波束形成的算法<sup>[8]</sup>、基于谱增强的算法<sup>[9]</sup>,以及直接估计晚期混响逆滤波器的算法等<sup>[10-11]</sup>。其中,多通道加权预测误差(Weighted Prediction Error, WPE)自适应混响抵消算法及其改进算法<sup>[10-11]</sup>具有良好的去混响性能和鲁棒性,被广泛研究和应用。WPE算法使用复高斯模型建模期望语音的复频谱,在最大似然准则下推导得到了抑制晚期混响的最优滤波器。然而,WPE算法在推导过程中忽略了噪声的影响,因此在噪声环境中,WPE算法的去混响效果会有一定程度的下降。

去噪算法和去混响算法既可通过独立优化而后级联实现,也可通过联合优化实现。前者将去噪去混响任务分解成两个子任务,各自独立优化,实现较为简单,但理论上并非全局最优;后者通过联合优化去噪去混响两个子任务来设计算法,算法设计较为复杂,但理论上可实现全局最优。有学者将WPE算法和MPDR波束形成器进行级联<sup>[12-13]</sup>,先对阵列拾取信号去除混响,再在此基础上进一步去除噪声。这类方法中的WPE算法由于噪声的存在,难以达到最优的去混响性能;同理,由于WPE算法未实现理想的去混响性能,后续MPDR波束形成器的去噪性能也会受到负面影响。此外,推导WPE去混响滤波器和MPDR波束形成器使用的优化准则不同,因而难以保证级联算法的去噪去混响性能在同一个准则下达到最优。近年来,有学者在复高斯分布假设条件下推导得到了基于最大似然准则的联合去噪去混响算法,即加功率无失真响应(Weighted Power minimization Distortionless response, WPD)波束形成器<sup>[14]</sup>。同WPE算法相同,WPD算法也使用复高斯模型建模语音时频域信号,并将去噪去混响任务的多个滤波器合并为一个滤波器,在最大似然准则下推导得到了最优联合去噪去混响滤波器。Nakatani等人<sup>[14]</sup>研究表明,WPD波束形成器比WPE算法级联MPDR波束形成器具有更优的去噪去混响性能,同时解决了传统级联方法两个模块优化准则不一致的问题。

已有研究表明<sup>[11,15]</sup>,语音信号在统计上并非理想的高斯信号,其分布更偏向于超高斯模型,相应地其复频谱也服从复超高斯分布。2007年,Jensen等人<sup>[15]</sup>提出采用复广义高斯模型中的复超高斯模型来建模语音,并推导了单通道最小均方误差(Minimum Mean Squared Error, MMSE)准则下的增益因子表达式,取得了比基于复高斯模型的最优滤波器更好的语音增强效果。2015年,Jukić等人<sup>[11]</sup>将复超高斯模型引入到WPE算法中,提出了复广义高斯模型加权预测误差算法(Complex Generalized Gaussian distribution based Weighted Prediction Error, CGG-WPE)。不同于WPE算法,该算法使用复超高斯模型建模期望语音的复频谱并在最大似然准则下推导抑制晚期混响的最优滤波器,获得了更好的去混响效果。

综上所述,基于复超高斯模型的语音增强算法在单通道语音增强和传声器阵列去混响应用中已得到一定的研究,但基于复超高斯模型的传声器阵列联合去噪去混响算法尚未有完整的理论推导和实验研究。基于复超高斯模型,本文首次通过理论推导得到了一种传声器阵列联合去噪去混响方法,记为复广义高斯分布加权功率无失真响应(Complex Generalized Gaussian distribution based Weighted Power minimization Distortionless response, CGG-WPD)波束形成器。本文采用复广义高斯模型对目标语音进行建模,并使用联合优化的方法同时完成去噪与去混响两个任务。理论推导表明,本文所提算法是WPD波束形成器以及WPE算法和MPDR波束形成器级联算法的一般化形式,且可以通过选取适当的参数获得较现有联合去噪去混响算法更优的性能。仿真实验和实际实验结果表明,本文提出的联合去噪去混响算法比WPD波束形成器具有更好的去噪去混响性能,在语音质量等客观指标上的测试结果也优于WPD波束形成器以及WPE算法和MPDR级联算法。

## 2 信号模型及传统算法

### 2.1 信号模型

假设室内期望语音信号与不相关噪声被任意形状的传声器阵列拾取,该阵列包含 $M$ 个全指向性传声器。此时拾取信号不仅包含期望信号的直达声,也包含早期反射声以及晚期混响。对拾取的时域信号做短时傅里叶变换(Short-Time Fourier Transform, STFT),拾取信号的时频域系数可以表示为:

$$\mathbf{x}(k, l) = \sum_{\tau=0}^L \mathbf{a}(k, \tau) s(k, l - \tau) + \mathbf{n}(k, l) \quad (1)$$

其中, $k$ 代表频域索引, $l$ 代表帧数索引, $\mathbf{x}(k, l) = [x^{(1)}(k, l), x^{(2)}(k, l), \dots, x^{(M)}(k, l)]^T$ 为所有传声器拾取信号的时频域系数; $s(k, l)$ 代表纯净语音的时频域系数; $\mathbf{a}(k, l) = [a^{(1)}(k, l), a^{(2)}(k, l), \dots, a^{(M)}(k, l)]^T$ 代表声学传递函数, $L$ 代表卷积长度; $\mathbf{n}(k, l) = [n^{(1)}(k, l), n^{(2)}(k, l), \dots, n^{(M)}(k, l)]^T$ 代表干扰和加性噪声。进一步式(1)可以重写为:

$$\mathbf{x}(k, l) = \mathbf{d}(k, l) + \mathbf{r}(k, l) + \mathbf{n}(k, l) \quad (2)$$

其中:

$$\mathbf{d}(k, l) = \sum_{\tau=0}^{b-1} \mathbf{a}(k, \tau) s(k, l - \tau) \quad (3)$$

$$\mathbf{r}(k, l) = \sum_{\tau=b}^L \mathbf{a}(k, \tau) s(k, l - \tau) \quad (4)$$

其中, $b$ 为预测延迟帧数。式(2)第一部分 $\mathbf{d}(k, l)$ 包含直达声和早期反射声,统称为期望信号;第二部分为晚期混响 $\mathbf{r}(k, l)$ ,第三部分为噪声信号 $\mathbf{n}(k, l)$ 。在实际应用中,混响情况下语音增强的目标为保留含有直达声和早期反射声的期望信号部分。为了方便对比,本节将简要介绍WPE算法与MPDR波束形成器,以及WPE算法级联MPDR波束形成器的两种级联方式。为了表述方便,下文中将WPE和MPDR级联算法记为WPE+MPDR算法。

### 2.2 加权预测误差去混响算法

忽略加性噪声 $\mathbf{n}(k, l)$ ,采用多通道自回归模型将公式(1)重写为:

$$\mathbf{x}(k, l) = \sum_{\tau=b}^{L_w} \mathbf{W}^H(k, \tau) \mathbf{x}(k, l - \tau) + \mathbf{d}(k, l) \quad (5)$$

其中, $L_w$ 为自回归预测模型的阶数, $\mathbf{W}(k, \tau)$ 为 $M \times M$ 多通道线性预测滤波器系数。式(5)右侧第二项为期望信号,即预测误差。加权预测误差去混响算法使用时变方差复高斯模型建模期望信号,即通道 $m$ 的期望语音信号在时频域上服从零均值能量不断变化的复高斯模型:

$$N_c(d^{(m)}(k, l); 0, \lambda_d(k, l)) = \frac{1}{\pi \lambda_d(k, l)} e^{-\frac{|d^{(m)}(k, l)|^2}{\lambda_d(k, l)}} \quad (6)$$

其中,期望信号功率即时方差 $\lambda_d(k, l)$ 是需要估计的变量。此时在最大似然准则下构建的优化问题可表示为<sup>[10]</sup>:

$$\hat{\Xi}(k) = \arg \min \sum_{l=1}^L \frac{\left\| \mathbf{x}(k, l) - \sum_{\tau=b}^{L_w} \hat{\mathbf{W}}^H(k, \tau) \mathbf{x}(k, l - \tau) \right\|_2^2}{\lambda_d(k, l)} \quad (7)$$

其中, $\hat{\Xi}(k) = [\mathbf{W}(k, b), \mathbf{W}(k, b+1), \dots, \mathbf{W}(k, L_w)]^T$ 为多帧预测系数矩阵的集合; $\|\cdot\|_2$ 为 $\ell_2$ 范数。由于 $\lambda_d(k, l)$ 是未知变量,优化问题(7)不存在闭式解,需使用迭代方式进行求解。迭代方法分两步,第一步估计期望信号的能量,第二步根据估计得到的期望信号能量来设计最优滤波器,重复迭代这两步直到算法收敛。具体迭代过程如下:

第一步:

$$\hat{\lambda}_d(k, l) = \frac{1}{(2\delta + 1)M} \sum_{\tau=l-\delta}^{l+\delta} \sum_{m=1}^M |d^{(m)}(k, l)|^2 \quad (8)$$

第二步:

$$\hat{\Xi}(k) = \mathbf{R}_{\text{WPE}}^{-1}(k) \mathbf{P}(k) \quad (9)$$

其中:

$$\mathbf{R}_{\text{WPE}}(k) = \sum_{l=L_w+1}^L \frac{\tilde{\mathbf{x}}(k, l) \tilde{\mathbf{x}}^H(k, l)}{\hat{\lambda}_d(k, l)} \quad (10)$$

$$\mathbf{P}(k) = \sum_{\tau=L_w+1}^L \frac{\tilde{\mathbf{x}}(k, l) \mathbf{x}^H(k, l)}{\hat{\lambda}_d(k, l)} \quad (11)$$

且  $\tilde{\mathbf{x}}(k, l) = [\mathbf{x}(k, l-b)^T, \mathbf{x}(k, l-b-1)^T, \dots, \mathbf{x}(k, l-L_w)^T]^T$  为传声器阵列拾取信号的堆积向量。上述迭代算法最终收敛到最优滤波器  $\hat{\Xi}(k)$ , 此时拾取信号滤波后得到的估计期望信号为:

$$\hat{\mathbf{d}}(k, l) = \mathbf{x}(k, l) - \sum_{\tau=b}^{L_w} \hat{\mathbf{W}}^H(k, \tau) \mathbf{x}(k, l-\tau) \quad (12)$$

### 2.3 最小功率无失真响应波束形成器

不失一般性, 选取第一通道作为参考通道, 并假设期望信号在频域上可以近似为相对传递函数 (Relative Transfer Function, RTF) [3] 与纯净语音的乘积。进一步将晚期混响作为噪声的一部分, 此时公式(1)可改写为:

$$\mathbf{x}(k, l) = \mathbf{d}(k, l) + \mathbf{n}(k, l) \quad (13)$$

其中,

$$\mathbf{d}(k, l) = \mathbf{h}(k) s(k, l) \quad (14)$$

$\mathbf{h}(k) = \left[ 1, \frac{a^{(2)}(k, l)}{a^{(1)}(k, l)}, \dots, \frac{a^{(M)}(k, l)}{a^{(1)}(k, l)} \right]^T$  为期望信号的相对传递函数[5]。

MPDR 波束形成器旨在设计一个空间滤波器, 在无失真约束情况下最小化波束形成输出信号能量, 从而达到提取目标语音并抑制噪声的目的。设计 MPDR 波束形成器最优滤波器时, 优化问题可以写为[3]:

$$\begin{aligned} \mathbf{w}(k) &= \arg \min \sum_{l=1}^L |\mathbf{w}^H(k) \mathbf{x}(k, l)|^2 \\ \text{s.t. } & \mathbf{w}^H(k) \mathbf{h}(k) = 1 \end{aligned} \quad (15)$$

该优化问题的闭式解可表示为:

$$\mathbf{w}(k) = \frac{\hat{\mathbf{R}}_x^{-1}(k) \mathbf{h}(k)}{\mathbf{h}^H(k) \hat{\mathbf{R}}_x^{-1}(k) \mathbf{h}(k)} \quad (16)$$

其中:

$$\hat{\mathbf{R}}_x(k) = \sum_{l=1}^L \mathbf{x}(k, l) \mathbf{x}^H(k, l) \quad (17)$$

为样本协方差矩阵。此时拾取信号经过波束形成器最优滤波器滤波得到的期望信号可表示为:

$$\hat{\mathbf{s}}(k, l) = \mathbf{w}^H(k) \mathbf{x}(k, l) \quad (18)$$

从上述的推导可知, MPDR 波束形成器具有同时去除噪声和混响的能力。但在传声器数量较少时, MPDR 波束形成器没有足够多的自由度同时消除噪声和混响, 因此在实现去噪去混响任务时, 通常需要在 MPDR 波束形成器模块前级联多通道去混响模块。

### 2.4 去噪去混响级联算法

WPE 去混响算法级联 MPDR 波束形成器的去噪去混响算法通常分为两个阶段, 第一阶段通过多通道去混响算法去除混响, 第二阶段通过波束形成器抑制部分残留混响以及噪声。根据迭代方式的不同可以分为具有反馈模块的级联方式, 以及不含反馈模块的级联方式。其中不存在反馈模块的级联方式, 去混响模块和波束形成模块相互独立, 如图 1(a) 所示; 存在反馈模块的级联方式, 波束形成模块的结果将影响去混响模块的迭代过程, 如图 1(b) 所示。

实际应用中, 两种级联方式在各阶段滤波器均分开独立设计, 因此其设计难度低, 易于实现。但该方式也存在以下问题: WPE 算法是在无噪声环境下使用最大似然准则推导出的最优去混响滤波器, 而 MPDR 波束形成器是在最小输出能量准则下推导出的最优空间滤波器, 两者的优化准则不同, 因此没有在同一准则下实现全局最优。

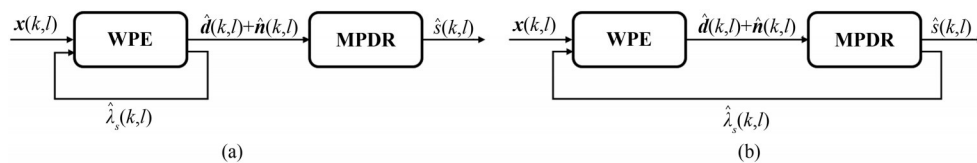


图1 级联去噪去混响算法流程图

Fig. 1 Cascade denoising and dereverberation algorithm processing flow



### 3 CGG-WPD算法

#### 3.1 理论推导

本节将在最大似然准则下首次推导得到基于复广义高斯模型先验的联合去噪去混响算法,即CGG-WPD算法。根据式(12)和式(18),可以得到:

$$\hat{s}(k, l) = \mathbf{w}^H(k) \left( \mathbf{x}(k, l) - \sum_{\tau=b}^{L_w} \hat{\mathbf{W}}^H(k, \tau) \mathbf{x}(k, l - \tau) \right) = \bar{\mathbf{w}}^H(k, l) \bar{\mathbf{x}}(k, l) \quad (19)$$

其中,  $\bar{\mathbf{x}}(k, l) = [\mathbf{x}^T(k, l), -\mathbf{x}^T(k, l-b), \dots, -\mathbf{x}^T(k, l-L_w)]^T$  代表传声器阵列拾取信号堆积矢量,  $\bar{\mathbf{w}}(k, l) = [\mathbf{w}^T(k), \dots, (\hat{\mathbf{W}}(k, \tau) \mathbf{w}(k))^T]^T$  代表滤波器系数堆积

矢量。在式(19)中,由于 $\bar{\mathbf{w}}(k, l)$ 同时包含了去噪滤波器与去混响滤波器,因此该波束形成器可同时具备去噪去混响功能。假设目标语音信号在时间上是彼此独立的,且都服从零均值的复广义高斯分布,其概率密度函数可以表示为<sup>[11]</sup>:

$$\rho(s(k, l)) = \frac{p}{2\pi\gamma\Gamma(2/p)} e^{-\frac{|s(k, l)|^p}{\gamma^{2/p}}} = e^{-\log\left(\frac{p}{2\pi\gamma\Gamma(2/p)}\right) - \frac{|s(k, l)|^p}{\gamma^{2/p}}} \quad (20)$$

其中, $\gamma$ 为尺度参数, $p$ 为复广义高斯分布的形状参数, $\Gamma(\cdot)$ 为伽马函数(Gamma Function)。根据 $p$ 的取值范围,复广义高斯分布可以分为以下三种:复超高斯分布( $0 < p < 2$ )、复高斯分布( $p = 2$ )以及复亚高斯分布( $p > 2$ )。假设期望信号服从复超高斯分布,即 $0 < p < 2$ ,根据凸分析理论<sup>[16]</sup>,式(20)所描述的概率密度函数还可以表示为缩放时变方差复高斯模型的最大值,即:

$$\rho(s(k, l)) = \max_{\lambda_s(k, l) > 0} N_c(s(k, l); 0, \lambda_s(k, l)) \psi(\lambda_s(k, l)) \quad (21)$$

其中, $\psi(\lambda_s(k, l))$ 为缩放函数。

根据上述模型,在最大似然准则下联合去噪去混响的优化问题可以表示为:

$$\max_{\bar{\mathbf{w}}(k)} \prod_{l=1}^L \max_{\lambda_s(k, l) > 0} N_c(s(k, l); 0, \lambda_s(k, l)) \psi(\lambda_s(k, l)) \quad (22)$$

s.t.  $\bar{\mathbf{w}}^H(k) \bar{\mathbf{h}}(k) = 1$

其中,  $\bar{\mathbf{h}}(k) = [\mathbf{h}^T(k), \mathbf{0}^T, \dots, \mathbf{0}^T]^T$ 。

值得注意的是,式(22)中的优化问题与期望信号 $s(k, l)$ 及其功率 $\lambda_s(k, l)$ 有关,但在实际应用中期望信号一般是未知的,只能获得估计值。因此,为了得到优化问题的解,本文使用期望信号的估计值代替其理论值。对式(22)取对数并使用估计值代替真实值,式(22)的等价优化问题可表示为:

$$\min_{\bar{\mathbf{w}}(k), \hat{\lambda}_s(k, l) > 0} \sum_{l=1}^L \left( \frac{|\hat{s}(k, l)|^2}{\hat{\lambda}_s(k, l)} + \log(\pi \hat{\lambda}_s(k, l)) - \log \psi(\hat{\lambda}_s(k, l)) \right) \quad (23)$$

s.t.  $\bar{\mathbf{w}}^H(k) \bar{\mathbf{h}}(k) = 1$

采用拉格朗日乘子法求解上述优化问题,构造代价函数如下:

$$\mathcal{J}_k = \sum_{l=1}^L g(\bar{\mathbf{w}}(k), \hat{\lambda}_s(k, l)) + \alpha_k (\bar{\mathbf{w}}^H(k) \bar{\mathbf{h}}(k) - 1) \quad (24)$$

其中:

$$g(\bar{\mathbf{w}}(k), \hat{\lambda}_s(k, l)) = \left( \frac{|\hat{s}(k, l)|^2}{\hat{\lambda}_s(k, l)} + \log(\pi \hat{\lambda}_s(k, l)) - \log \psi(\hat{\lambda}_s(k, l)) \right) + \left( \frac{|\bar{\mathbf{w}}^H(k, l) \bar{\mathbf{x}}(k, l)|^2}{\hat{\lambda}_s(k, l)} + \log(\pi \hat{\lambda}_s(k, l)) - \log \psi(\hat{\lambda}_s(k, l)) \right) \quad (25)$$

$\alpha_k$ 为拉格朗日乘子。优化问题式(23)需要同时优化两个变量 $\hat{\lambda}_s(k, l)$ 和 $\bar{\mathbf{w}}(k)$ ,导致该优化问题不存在闭式解。为了解决这一问题,本文提出一种交替迭代优化 $\hat{\lambda}_s(k, l)$ 和 $\bar{\mathbf{w}}(k)$ 的算法,最终收敛到全局最优解。假设 $\bar{\mathbf{w}}(k)$ 已知,此时式(24)对 $\hat{\lambda}_s(k, l)$ 求偏导并令其偏导为0可得:

$$\hat{\lambda}_s(k, l) = \frac{2\gamma^{p/2}}{p} |\hat{s}(k, l)|^{2-p} \quad (26)$$

假设 $\hat{\lambda}_s(k, l)$ 已知,代价函数式(24)变为:

$$\mathcal{J}_k = \sum_{l=1}^L \frac{|\bar{\mathbf{w}}^H(k, l) \bar{\mathbf{x}}(k, l)|^2}{\hat{\lambda}_s(k, l)} + \alpha_k (\bar{\mathbf{w}}^H(k) \bar{\mathbf{h}}(k) - 1) \quad (27)$$

式(27)对 $\bar{\mathbf{w}}(k)$ 求偏导并令其偏导为0可得:

$$\bar{\mathbf{w}}(k) = \frac{\hat{\mathbf{R}}_{\text{CGG}}^{-1}(k) \bar{\mathbf{h}}(k)}{\bar{\mathbf{h}}^H(k) \hat{\mathbf{R}}_{\text{CGG}}^{-1}(k) \bar{\mathbf{h}}(k)} \quad (28)$$

其中:

$$\mathbf{R}_{\text{CGG}}(k) = \sum_{l=1}^L \frac{\bar{\mathbf{x}}(k, l) \bar{\mathbf{x}}^H(k, l)}{\hat{\lambda}_s(k, l)} = \sum_{l=1}^L \frac{p \bar{\mathbf{x}}(k, l) \bar{\mathbf{x}}^H(k, l)}{2\gamma^{p/2} |\hat{s}(k, l)|^{2-p}} \quad (29)$$

观察式(28)和式(29)可知,式(28)的波束形成器最优滤波器系数不受常数影响,因此式(29)可等价:

$$\mathbf{R}_{\text{CGG}}(k) = \sum_{l=1}^L \frac{\bar{\mathbf{x}}(k, l) \bar{\mathbf{x}}^H(k, l)}{|\hat{s}(k, l)|^{2-p}} \quad (30)$$

为避免式(30)中分母项为零,本文将在式(30)的分母项中添加小量  $\varepsilon$ , 即:

$$\mathbf{R}_{\text{CGG}}(k) = \sum_{l=1}^L \frac{\bar{\mathbf{x}}(k, l) \bar{\mathbf{x}}^H(k, l)}{\min(|\hat{s}(k, l)|^{2-p}, \varepsilon)} \quad (31)$$

在实际应用中,可以使用对角加载技术提高波束形成器的鲁棒性,此时波束形成器最优滤波器系数为:

$$\mathcal{J}_k = \mathbf{q}^H(k) \begin{bmatrix} \mathbf{I}_M \\ -\tilde{\mathbf{G}}(k) \end{bmatrix}^H \begin{bmatrix} \mathbf{R}_{\bar{x}\bar{x}}(k) & \mathbf{P}_{\bar{x}\bar{x}}^H(k) \\ \mathbf{P}_{\bar{x}\bar{x}}(k) & \mathbf{R}_{\bar{x}\bar{x}}(k) \end{bmatrix} \begin{bmatrix} \mathbf{I}_M \\ -\tilde{\mathbf{G}}(k) \end{bmatrix} \mathbf{q}(k) + \alpha_k \left( \mathbf{q}^H(k) \begin{bmatrix} \mathbf{I}_M \\ -\tilde{\mathbf{G}}(k) \end{bmatrix}^H \bar{\mathbf{h}}(k) - 1 \right) = \mathbf{q}^H(k) \mathbf{R}_{\bar{x}\bar{x}}(k) \mathbf{q}(k) - \mathbf{q}^H(k) \tilde{\mathbf{G}}^H(k) \mathbf{P}_{\bar{x}\bar{x}}(k) \mathbf{q}(k) - \mathbf{q}^H(k) \mathbf{P}_{\bar{x}\bar{x}}^H(k) \tilde{\mathbf{G}}(k) \mathbf{q}(k) + \mathbf{q}^H(k) \tilde{\mathbf{G}}^H(k) \mathbf{R}_{\bar{x}\bar{x}}(k) \tilde{\mathbf{G}}(k) \mathbf{q}(k) + \alpha_k (\mathbf{q}^H(k) \mathbf{v}(k) - 1) \quad (35)$$

其中:

$$\mathbf{R}_{\bar{x}\bar{x}}(k) = \sum_{l=1}^L \frac{\mathbf{x}(k, l) \mathbf{x}^H(k, l)}{|\hat{s}(k, l)|^{2-p}} \quad (36)$$

$$\mathbf{R}_{\tilde{x}\tilde{x}}(k) = \sum_{l=1}^L \frac{\tilde{\mathbf{x}}(k, l) \tilde{\mathbf{x}}^H(k, l)}{|\hat{s}(k, l)|^{2-p}} \quad (37)$$

$$\mathbf{P}_{\bar{x}\tilde{x}}(k) = \sum_{l=1}^L \frac{\tilde{\mathbf{x}}(k, l) \mathbf{x}^H(k, l)}{|\hat{s}(k, l)|^{2-p}} \quad (38)$$

代价函数式(35)对  $\tilde{\mathbf{G}}(k)$  求偏导并置零得:

$$\frac{\partial \mathcal{J}_k}{\partial \tilde{\mathbf{G}}(k)} = -2\mathbf{P}_{\bar{x}\tilde{x}}(k) \mathbf{q}(k) \mathbf{q}^H(k) + 2\mathbf{R}_{\tilde{x}\tilde{x}}(k) \tilde{\mathbf{G}}(k) \mathbf{q}(k) \mathbf{q}^H(k) = 0 \quad (39)$$

当  $\mathbf{R}_{\bar{x}\bar{x}}(k)$  为满秩矩阵时,式(39)的解为  $\tilde{\mathbf{G}}(k) = \mathbf{R}_{\tilde{x}\tilde{x}}^{-1}(k) \mathbf{P}_{\bar{x}\tilde{x}}(k)$ 。该解为文献[11]中提到的在复广义高斯模型下的多通道去混响算法对应的滤波器,即CGG-WPE算法对应的滤波器。将  $\tilde{\mathbf{G}}(k)$  的解带入式(35),经过化简得到:

$$\mathcal{J}_k = \mathbf{q}^H(k) \mathbf{R}_{\tilde{d}\tilde{d}}(k) \mathbf{q}(k) + \alpha_k (\mathbf{q}^H(k) \mathbf{v}(k) - 1) \quad (40)$$

其中,

$$\bar{\mathbf{w}}(k) = \frac{(\mathbf{R}_{\text{CGG}}(k) + \zeta \mathbf{I})^{-1} \bar{\mathbf{h}}(k)}{\bar{\mathbf{h}}^H(k) (\mathbf{R}_{\text{CGG}}(k) + \zeta \mathbf{I})^{-1} \bar{\mathbf{h}}(k)} \quad (32)$$

使用式(19)、(26)、(32)交替优化  $\hat{\lambda}_s(k, l)$  和  $\bar{\mathbf{w}}(k)$  直到收敛,或达到最大迭代次数后停止迭代。

### 3.2 CGG-WPD算法与传统算法的内在联系

本节将给出3.1节中CGG-WPD波束形成器的分解形式,同时将从理论上分析其与现有级联算法的内在联系,进而证明本文所提算法的优越性。假设CGG-WPD波束形成器可以拆分为两个独立滤波器:

$$\bar{\mathbf{w}}(k) = \bar{\mathbf{G}}(k) \mathbf{q}(k) \quad (33)$$

其中,  $\bar{\mathbf{G}}(k)$  具有去混响滤波器的功能,  $\mathbf{q}(k)$  具有空间滤波器的功能。在级联形式中,去混响滤波器应尽量减少期望信号失真,因此:

$$\bar{\mathbf{G}}(k) = \begin{bmatrix} \mathbf{I}_M \\ -\tilde{\mathbf{G}}(k) \end{bmatrix} \quad (34)$$

此时代价函数式(27)可展开写为:

$$\mathbf{R}_{\tilde{d}\tilde{d}}(k) = \sum_{l=1}^L \frac{\bar{\mathbf{G}}(k) \mathbf{x}(k, l) \mathbf{x}^H(k, l) \bar{\mathbf{G}}^H(k)}{|\hat{s}(k, l)|^{2-p}} = \sum_{l=1}^L \frac{\tilde{\mathbf{d}}(k, l) \tilde{\mathbf{d}}^H(k, l)}{|\hat{s}(k, l)|^{2-p}} \quad (41)$$

对  $\mathbf{q}(k)$  求偏导并令偏导为零得:

$$\mathbf{q}(k) = \frac{\mathbf{R}_{\tilde{d}\tilde{d}}^{-1}(k) \mathbf{h}(k)}{\mathbf{h}^H(k) \mathbf{R}_{\tilde{d}\tilde{d}}^{-1}(k) \mathbf{h}(k)} \quad (42)$$

式(42)得到的空间滤波器,即为根据去混响滤波器输出信号的空间特性构建的自适应波束形成器。该自适应波束形成器最优滤波器表达式类似于MPDR波束形成器以及文献[7]中提到的MLDR波束形成器。当  $p = 2$  时,此波束形成器为MPDR波束形成器最优滤波器;当  $p = 0$  时,此波束形成器为MLDR波束形成器最优滤波器;当  $0 < p < 2$  时,此波束形成器为在复超高斯分布先验条件下推导得到的最大似然波束形成器最优滤波器,本文将其称为CGG-MLDR波束形成器。下面本文将进一步证明,相比于MPDR波束形成器滤波器,CGG-MLDR波束形成器具有更优的去噪能力。

假设去混响滤波器的 $L$ 帧输出信号为 $\tilde{\mathbf{d}}(k, l) = \bar{\mathbf{G}}(k)\mathbf{x}(k, l)$ ,其中有 $L_1$ 帧只包含噪声和干扰,有 $L_2$ 帧不仅包含噪声并且同时包含期望信号,此时式(41)所代表的总协方差矩阵重写为:

$$\mathbf{R}_{\tilde{\mathbf{d}}\tilde{\mathbf{d}}}(k) = L_2(\lambda_s(k, l))^{p/2} \mathbf{r}_{ss}(k) + \left( L_1 \rho(k) + L_2(\lambda_s(k, l))^{p/2} \frac{1}{\kappa(k)} \right) \mathbf{r}_{nn}(k) \quad (43)$$

其中, $\lambda_n$ 为噪声功率谱密度; $\rho(k) = \lambda_n / \varepsilon^{(1-p/2)}$ ;  $\kappa(k) = \lambda_s(k) / \lambda_n(k)$ 为输入信噪比; $\mathbf{r}_{ss}(k)$ 为信号相关矩阵, $\mathbf{r}_{nn}(k)$ 为噪声相关矩阵。由式(43)可知,总协方差矩阵为信号和噪声相关矩阵的线性组合,为便于分析,定义两个线性组合系数的比值为:

$$r_p(k) = \frac{\left( L_1 \rho(k) + L_2(\lambda_s(k, l))^{p/2} \frac{1}{\kappa(k)} \right)}{L_2(\lambda_s(k, l))^{p/2}} \quad (44)$$

由式(44)可知,此比值反映了同样快拍数下自适应波束形成器的鲁棒性程度,比值越大证明鲁棒性越强。当 $p = 0$ 时, $r_0(k) = (L_1 \rho(k) + L_2 / \kappa(k)) / L_2$ ;当 $p = 2$ 时比值为 $r_2(k) = L / L_2 \kappa(k)$ ,即此时比值取决于包含期望信号的帧数和输入信噪比。当 $\lambda_s(k, l) > \varepsilon$ 时, $r_0(k) > r_2(k)$ 。由于 $\varepsilon$ 是小量,因而 $\lambda_s(k, l) > \varepsilon$ 这一条件十分容易满足,所以当 $p = 0$ 时得到的波束形成器优于MPDR波束形成器,同理可证当 $0 < p < 2$ 时得到的CGG-MLDR波束形成器也同样优于MPDR波束形成器。同时根据文献[11]的研究,当 $p = 0.5$ 时CGG-WPE算法明显优于WPE算法( $p = 0$ ),而CGG-WPD去噪去混响算法等价于CGG-WPE算法与CGG-MLDR算法的级联,因此可知,当 $p = 0.5$ 时,本文提出的联合去噪去混响算法应明显优于WPE+MPDR算法。

在实际应用中不仅要考虑算法的性能,还应考虑应用算法的计算复杂度。表1给出了CGG-WPD算法、WPD算法以及WPD+MPDR算法的理论计算复杂度,以及本文第4节仿真实验所用参数即, $M = 6, L_w = 10, b = 4, I = 5$ 时的计算复杂度。从表1中可以发现,在相同迭代次数时,CGG-WPD算法与WPD算法的计算复杂度相当,都略高于WPE+MPDR算法。但图2的仿真结果表明CGG-WPD算

法在迭代2~3次时已经达到WPD算法迭代10次时的性能,所以在实际应用中CGG-WPD算法可以使用更少的迭代次数来降低计算复杂度。

表1 三种算法计算复杂度分析

Tab. 1 Computational complexity analysis of three algorithms

方法	计算复杂度	$M=6, L_w=10, b=4, I=5$
WPE+MPDR	$O(M^3(L_w - b + 1)^3 I) + O(M^3)$	$O(370656)$
WPD	$O((M(L_w - b + 1) + M)^3 I)$	$O(552960)$
CGG-WPD	$O((M(L_w - b + 1) + M)^3 I)$	$O(552960)$

## 4 实验测试

### 4.1 仿真实验设置及评估指标

本文评测CGG-WPD算法在不同声学场景下的去噪去混响性能,并将其与WPE+MPDR算法,以及WPD算法比较。测试所用纯净语音由从TIMIT数据集<sup>[17]</sup>中选取的30条男声与30条女声组成;测试所用噪声为NOISEX-92数据库<sup>[18]</sup>中的babble噪声。仿真实验所用的房间脉冲响应由镜像法<sup>[19]</sup>生成,房间尺寸为 $6\text{ m} \times 10\text{ m} \times 4\text{ m}$ 。这里选用 $M = 6$ 的均匀线阵,其阵元间距为 $0.043\text{ m}$ ,阵列放置在房间中心,期望信号在距离阵列中心 $2\text{ m}$ 处从端射方向的 $0^\circ$ 方向入射,另有一个干扰从 $45^\circ$ 方向入射。所有算法使用的预延迟 $b = 4$ ,自回归模型阶数 $L_w = 10$ ,帧长为 $64\text{ ms}$ ,帧移为 $16\text{ ms}$ 。本文使用语音质量感知评估测度(Perceptual evaluation of speech quality score, PESQ)<sup>[20]</sup>、扩展短时客观可懂度(Extended Short-Time Objective Intelligibility, ESTOI)<sup>[21]</sup>、信号失真比(Signal to Distortion Ratio, SDR)<sup>[22]</sup>、语音混响调制能量比(Speech-to-Reverberation Modulation energy Ratio, SRMR)<sup>[23]</sup>作为各算法去噪去混响能力的客观评价标准。实验所用相对传递函数使用协方差矩阵白化法<sup>[24]</sup>求得。

### 4.2 仿真实验结果及分析

首先设计仿真实验测试不同 $p$ 值下所提算法随迭代次数的变化情况。这里设置混响时间为 $T_{60} = 320\text{ ms}$ ,输入信干噪比为 $\text{SINR} = 5\text{ dB}$ 。图2给出了仿真实验结果,从中可以发现 $p$ 值的选取对

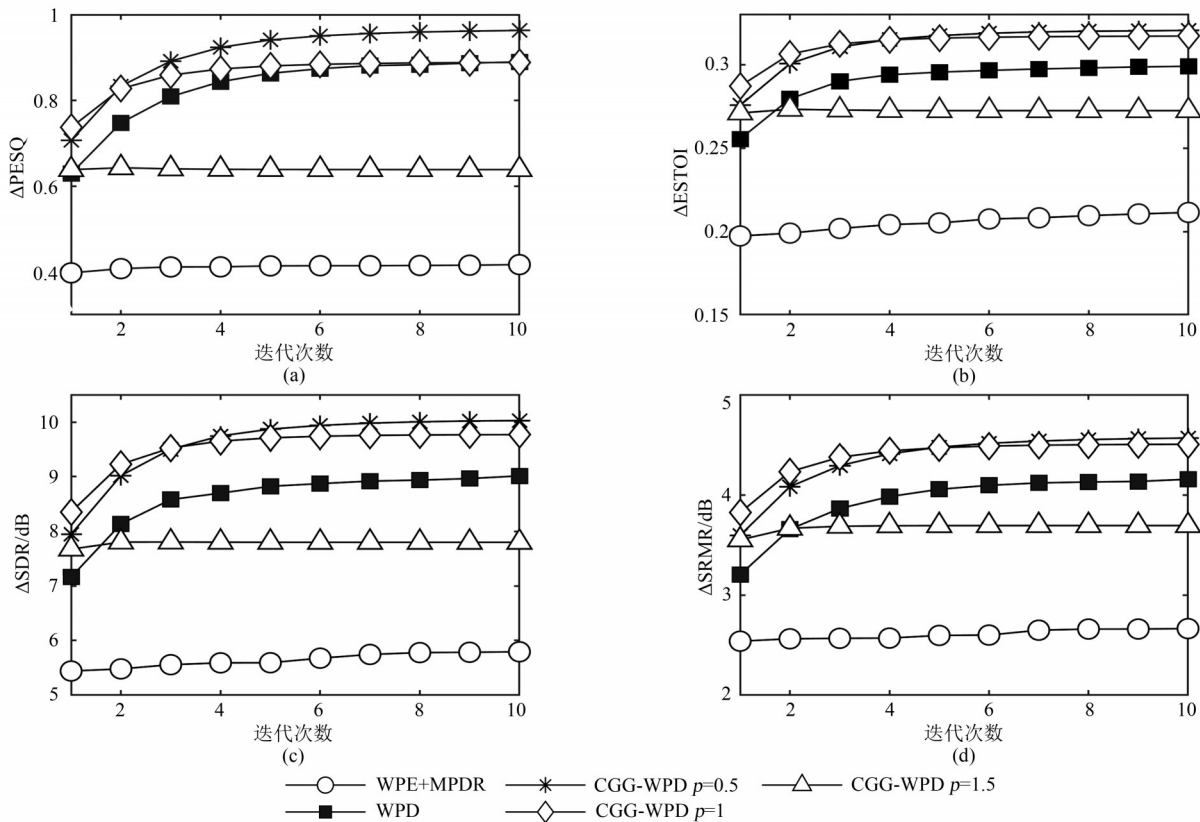


图2 不同迭代次数时的实验结果. (a) PESQ 提升量; (b) ESTOI 提升量; (c) SDR 提升量; (d) SRMR 提升量

Fig. 2 Results for different number of iterations. (a) PESQ improvements; (b) ESTOI improvements; (c) SDR improvements; (d) SRMR improvements

CGG-WPD算法的性能有较大影响。在各项客观评价指标中,  $p = 0.5$ 时CGG-WPD算法都取得了最优的性能,且高于已有WPE+MPDR算法以及WPD算法。同时CGG-WPD算法( $p = 0.5$ )在低迭代次数时(2~3次)已经达到WPD算法接近收敛时的性能。从图中还可发现当混响时间较短时,WPE+MPDR算法随着迭代次数的增加,各项客观指标没有明显提升,但WPD和本文所提算法性能提升明显。由于WPD算法在第一阶段的去混响算法与WPE+MPDR算法一致,由此可知性能提升的主要原因是第二阶段的CGG-MLDR波束形成器较MPDR波束形成器具有更强的去噪能力。由于上述实验已经证明在 $p = 0.5$ 时所提算法取得最优性能,且在迭代5次时接近收敛,因此在后续所有实验中,CGG-WPD算法固定参数 $p = 0.5$ ,迭代次数为5次。

第二个仿真实验测试不同输入信干噪比条件下所提算法的去噪能力,混响时间设为 $T_{60} = 320$  ms。实验结果如图3所示。本文已在3.2节中证明了

CGG-WPD算法等价于CGG-WPE去混响算法和CGG-MLDR波束形成器的级联。为进一步通过仿真实验证明推导的正确性,本文在图3的仿真实验结果中同时给出联合算法以及与其等价的级联算法的实验结果。本文使用“(Jointly)”和“(Separate)”来区分联合滤波器以及其等价的级联滤波器,其中级联滤波器是通过图1(a)所示的级联结构实现滤波。从图3中可知,在不同输入信干噪比环境下,CGG-WPD算法在各项客观指标中都高于WPD算法及WPE+MPDR算法。但随着信干噪比的增加,CGG-WPD算法和WPD算法在PESQ打分和ESTOI打分这两项客观指标的差距逐渐缩小,在高信噪比时近乎相同。同时观察联合算法以及其相对应的级联算法的实验结果可发现,两者的各项客观指标近似一致,这也验证了3.2节理论推导的正确性。

最后,本文设计仿真实验比较不同混响条件下各算法的去混响性能。除了混响时间发生变化外,其余实验参数同仿真实验2,且固定输入信干噪比



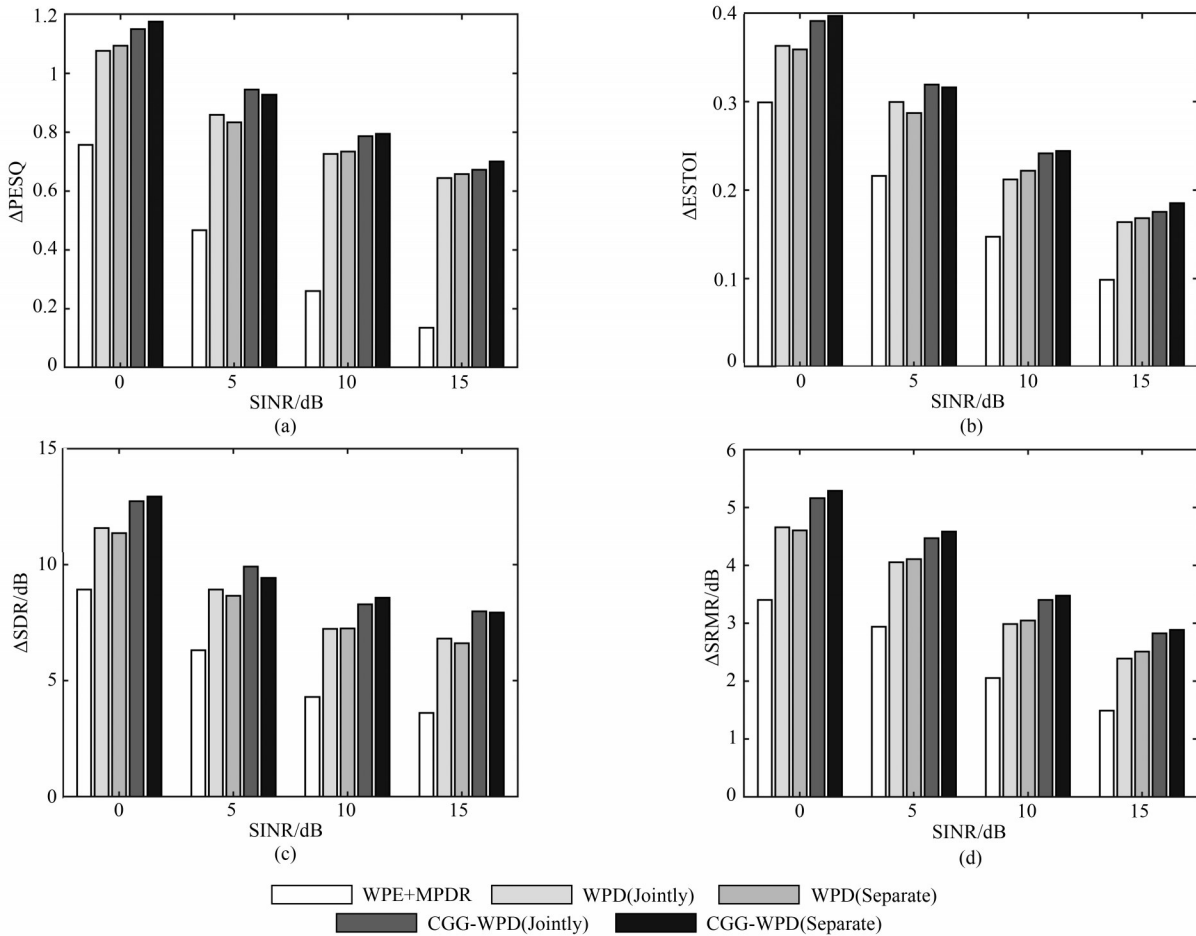


图 3 不同输入信干噪比的实验结果. (a) PESQ 提升量; (b) ESTOI 提升量; (c) SDR 提升量; (d) SRMR 提升量  
 Fig. 3 Results for different input SINR values. (a) PESQ improvements; (b) ESTOI improvements; (c) SDR improvements; (d) SRMR improvements

SINR = 5 dB。由图 4 中可知,在不同混响时间下,本文所提的 CGG-WPD 算法都优于 WPD 算法和 WPE+MPDR 算法。但随着混响时长的增加,各算法的各项客观指标的提升量都有不同程度的下降,且 WPE+MPDR 算法与其他算法在各项客观指标上的性能差距不断减小。

### 4.3 实际环境实验

本节测试实际噪声和混响环境下各算法的去噪去混响性能。实际房间脉冲响应选自 REVERB Challenge 数据集,选择其中一点播放语音信号作为期望信号,另一点播放噪声作为干扰。测试所用纯净语音由从 TIMIT 数据集中选取的 30 条男声与 30 条女声组成;测试所用噪声为 NOISEX-92 数据库中的 babble 噪声。同时本文还测试了 CHiME-3 数据集中不同场景的实录噪声下各算法的语音增强性

能。实验中的相对传递函数均使用协方差矩阵白化法估计得到。

表 2 至表 4 给出了各算法在不同房间 5 次迭代时的客观指标结果。从表 2~表 4 中可以发现 CGG-WPD 算法在不同信干噪比,不同房间的各项客观指标都是最优的。比较不同房间的实验结果可知,随着混响时间的增加,各算法的各项客观指标的提升量都有所下降。比较同一房间不同信干噪比的实验结果可知,高信干噪比时所提算法各项客观指标的提升量较低信干噪比时更高。这与仿真实验结果有差异,分析其主要原因是仿真实验中的混响时间较短,主要反映了所提算法的降噪能力;而真实房间混响时间较长,一般都在 0.5 s 以上,三种算法在高信干噪比下具有更好的去混响性能。图 5 为某测试样例的语谱图,从语谱图中可以发现 CGG-

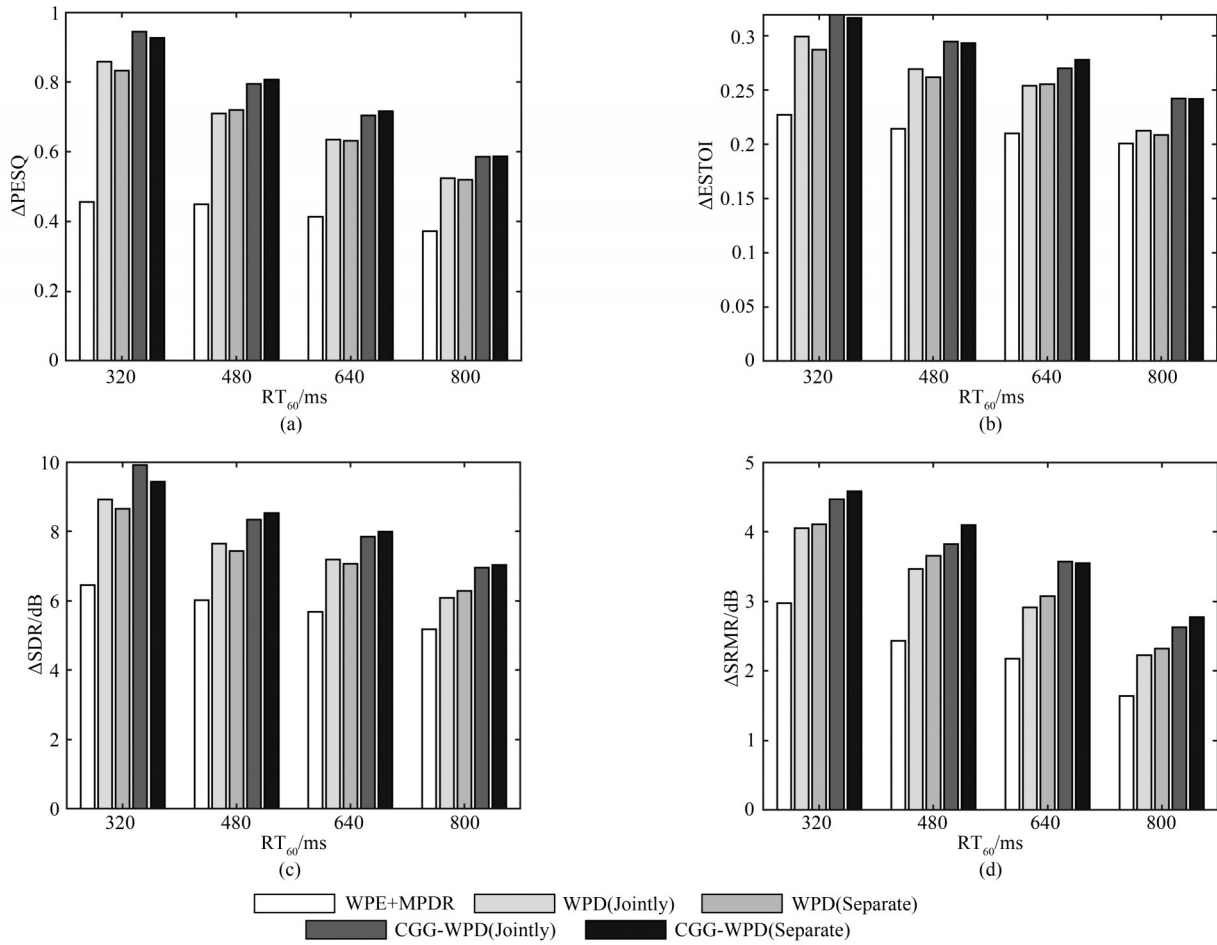


图4 不同混响时间下的实验结果. (a) PESQ 提升量; (b) ESTOI 提升量; (c) SDR 提升量; (d) SRMR 提升量  
 Fig. 4 Results for different reverberation time. (a) PESQ improvements; (b) ESTOI improvements; (c) SDR improvements; (d) SRMR improvements

表2 REVERB Challenge 小型房间测试结果

Tab. 2 Experiment results of small room in REVERB Challenge

方法	SINR=0 dB				SINR=10 dB			
	PESQ	ESTOI	SDR/dB	SRMR/dB	PESQ	ESTOI	SDR/dB	SRMR/dB
NOISY	1.35	0.38	-0.92	2.30	1.98	0.60	6.49	5.29
WPE+MPDR	1.37	0.39	-1.27	2.50	2.17	0.65	8.00	6.20
WPD	2.09	0.56	4.68	5.32	2.91	0.84	12.12	9.11
CGG-WPD	<b>2.26</b>	<b>0.62</b>	<b>6.17</b>	<b>6.61</b>	<b>3.04</b>	<b>0.85</b>	<b>12.30</b>	<b>9.35</b>

表3 REVERB Challenge 中等房间测试结果

Tab. 3 Experiment results of medium room in REVERB Challenge

方法	SINR=0 dB				SINR=10 dB			
	PESQ	ESTOI	SDR/dB	SRMR/dB	PESQ	ESTOI	SDR/dB	SRMR/dB
NOISY	1.10	0.31	-2.81	1.84	1.63	0.48	2.32	3.69
WPE+MPDR	1.30	0.40	-0.62	2.65	2.09	0.66	7.59	6.37
WPD	1.91	0.64	5.72	6.31	2.70	0.84	10.72	9.22
CGG-WPD	<b>2.01</b>	<b>0.66</b>	<b>6.09</b>	<b>6.85</b>	<b>2.80</b>	<b>0.85</b>	<b>10.96</b>	<b>9.47</b>

表 4 REVERB Challenge 大型房间测试结果

Tab. 4 Experiment results of large room in REVERB Challenge

方法	SINR=0 dB				SINR=10 dB			
	PESQ	ESTOI	SDR/dB	SRMR/dB	PESQ	ESTOI	SDR/dB	SRMR/dB
NOISY	1.12	0.28	-3.46	1.67	1.57	0.42	1.23	3.13
WPE+MPDR	1.39	0.40	-0.38	2.82	2.10	0.63	7.15	6.43
WPD	1.75	0.51	-1.27	5.60	2.55	0.77	9.87	8.77
CGG-WPD	2.02	0.62	5.92	6.64	2.61	0.78	9.97	8.98

WPD算法的噪声和混响残留明显少于WPE+MPDR算法。同时CGG-WPD算法相较WPD算法能更好

的保留期望信号的谐频结构,例如图5(e)和图5(f)的白色圆圈部分。

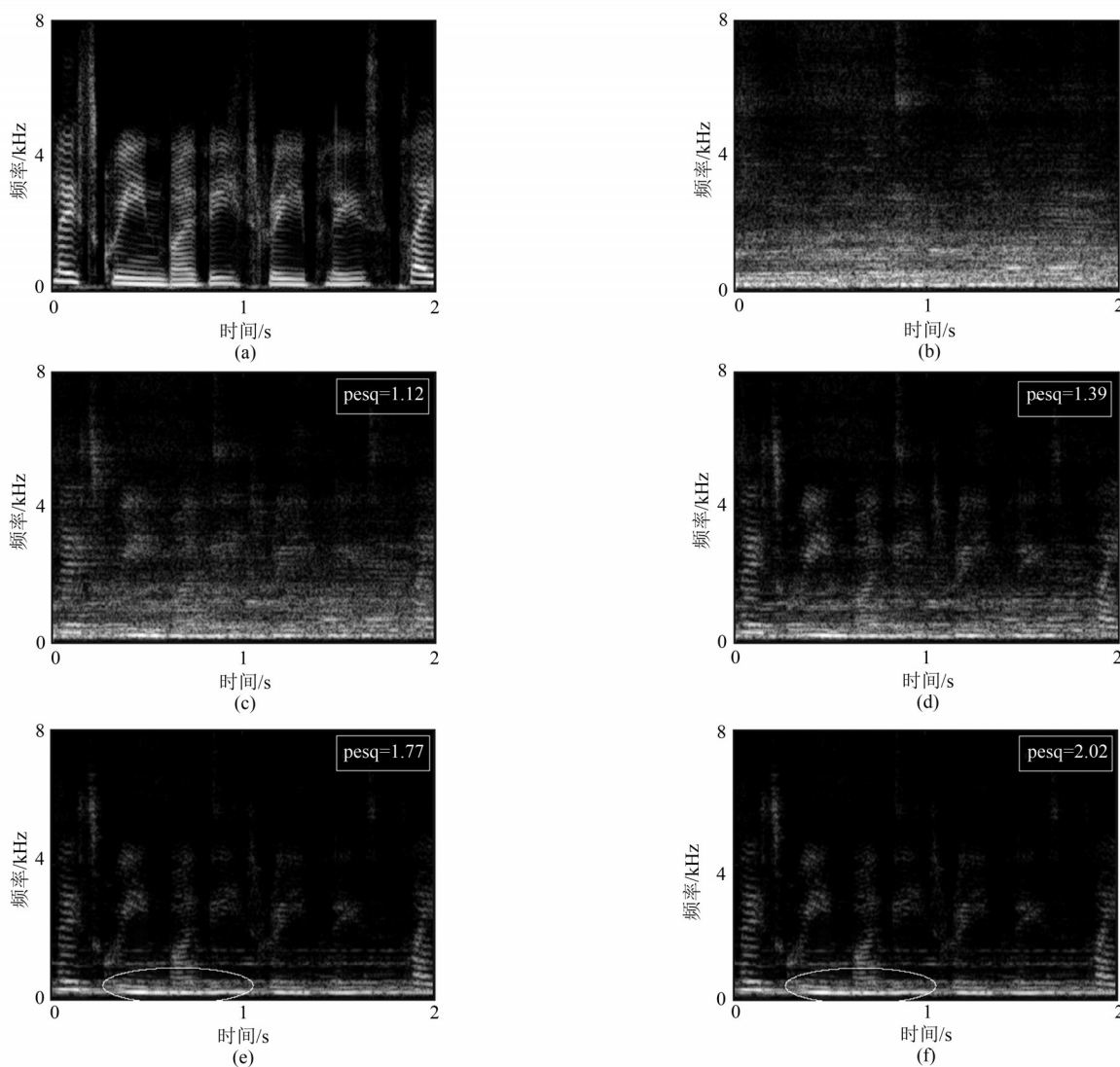


图 5 REVERB Challenge 测试样例语谱图. (a)纯净语音; (b)噪声信号; (c)带噪语音; (d)WPE+MPDR; (e)WPD; (f)CGG-WPD

Fig. 5 REVERB Challenge test sample speech spectrograms. (a) clean speech; (b) noise; (c) noisy speech; (d) WPE+MPDR; (e) WPD; (f) CGG-WPD

表5给出了在CHiME-3比赛中提供的四种常见场景噪声下不同算法的实验结果,各场景下的实验结果为40条测试数据的平均值。从表5的实验结果中可知,CGG-WPD算法的PESQ及ESTOI指标在

BUS和STR两种噪声场景下明显优于WPD算法,在CAF和PED噪声场景下略优于WPD算法。此外,CGG-WPD与WPD算法在各噪声场景下都显著优于WPE+MPDR算法。

表5 CHiME-3测试结果  
Tab. 5 Experiment results of CHiME-3

方法	BUS		CAF		PED		STR	
	PESQ	ESTOI	PESQ	ESTOI	PESQ	ESTOI	PESQ	ESTOI
NOISY	2.40	0.73	1.90	0.56	1.81	0.54	2.03	0.67
WPE+MPDR	2.77	0.80	2.35	0.70	2.20	0.65	2.51	0.76
WPD	2.92	0.87	2.48	0.78	2.33	0.74	2.73	0.85
CGG-WPD	<b>3.02</b>	<b>0.88</b>	<b>2.54</b>	<b>0.80</b>	<b>2.39</b>	<b>0.76</b>	<b>2.83</b>	<b>0.86</b>

## 5 结论

采用复广义高斯模型建模期望语音信号,本文提出了一种新的联合去噪去混响算法。通过理论推导证明了该算法是WPD等联合去噪去混响算法的一般化形式,同时其也等价于去混响模块与去噪模块的级联形式。仿真实验证明通过选择合适的参数,本文提出的CGG-WPD算法的去噪去混响性能在多项客观指标上都优于现有的WPD以及WPE+MPDR算法。对比其他参数发现, $p = 0.5$ 是较优的参数选择。仿真实验及实际实验结果证明,CGG-WPD算法在不同的声学场景下的去噪去混响性能也都优于WPD及WPE+MPDR算法。但本文所有研究的波束形成器在最优滤波器迭代更新中,都需要对方差矩阵进行求逆运算,当矩阵维度较高,且迭代次数较多时,计算复杂度较高。因此,可将加快其收敛速度或降低每一次迭代的计算复杂度作为下一步研究方向。

### 参考文献

- [1] 潘超,黄公平,陈景东. 面向语音通信与交互的麦克风阵列波束形成方法[J]. 信号处理, 2020, 36(6): 804-815.  
PAN Chao, HUANG Gongping, CHEN Jingdong. Microphone array beamforming: An overview [J]. Journal of Signal Processing, 2020, 36(6): 804-815. (in Chinese)
- [2] 冷艳宏,郑成诗,李晓东. 功率比相关子带划分快速独立向量分析[J]. 信号处理, 2019, 35(8): 1314-1323.  
LENG Yanhong, ZHENG Chengshi, LI Xiaodong. Fast independent vector analysis using power ratio correlation-

- based bands partition [J]. Journal of Signal Processing, 2019, 35(8): 1314-1323. (in Chinese)
- [3] CAPON J. High-resolution frequency-wavenumber spectrum analysis [J]. Proceedings of the IEEE, 1969, 57(8): 1408-1418.
- [4] 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.  
GUO Xiangyu, YAN Shefeng, WANG Wenxia. A fast algorithm for linear constrained robust capon beamforming based on iterative gradient method [J]. Journal of Signal Processing, 2021, 37(5): 712-723. (in Chinese)
- [5] GANNOT S, COHEN I. Speech enhancement based on the general transfer function GSC and postfiltering [J]. IEEE Transactions on Speech and Audio Processing, 2004, 12(6): 561-571.
- [6] ZHENG Chengshi, DELEFORGE A, LI Xiaodong, et al. Statistical analysis of the multichannel Wiener filter using a bivariate normal distribution for sample covariance matrices [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 26(5): 951-966.
- [7] CHO B J, LEE J M, PARK H M. A beamforming algorithm based on maximum likelihood of a complex Gaussian distribution with time-varying variances for robust speech recognition [J]. IEEE Signal Processing Letters, 2019, 26(9): 1398-1402.
- [8] HABETS E A P, BENESTY J. A two-stage beamforming approach for noise reduction and dereverberation [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(5): 945-958.
- [9] SCHWARZ A, KELLERMANN W. Coherent-to-diffuse power ratio estimation for dereverberation [J]. IEEE/ACM Transactions on Audio, Speech, and Language Pro-



- cessing, 2015, 23(6): 1006-1018.
- [10] NAKATANI T, YOSHIOKA T, KINOSHITA K, et al. Speech dereverberation based on variance-normalized delayed linear prediction [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(7): 1717-1731.
- [11] JUKIĆ A, VAN WATERSCHOOT T, GERKMANN T, et al. Multi-channel linear prediction-based speech dereverberation with sparse priors [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(9): 1509-1520.
- [12] DRUDE L, BOEDDEKER C, HEYMANN J, et al. Integrating neural network based beamforming and weighted prediction error dereverberation [C] // Interspeech 2018, Hyderabad, India. IEEE, 2018: 3043-3047.
- [13] SONG Siyuan, CHENG Longbiao, LUAN Shuming, et al. An integrated multi-channel approach for joint noise reduction and dereverberation [J]. Applied Acoustics, 2021, 171(7): 107526-107534.
- [14] NAKATANI T, BOEDDEKER C, KINOSHITA K, et al. Jointly optimal denoising, dereverberation, and source separation [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28(14): 2267-2282.
- [15] ERKELENS J S, HENDRIKS R C, HEUSDENS R, et al. Minimum mean-square error estimation of discrete Fourier coefficients with generalized gamma priors [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(6): 1741-1752.
- [16] PALMER J, WIPF D, KREUTZ-DELGADO K, et al. Variational EM algorithms for non-Gaussian latent variable models [J]. Advances in Neural Information Processing Systems, 2006, 18(5): 1059-1066.
- [17] ZUE V, SENEFF S, GLASS J. Speech database development at MIT: Timit and beyond [J]. Speech Communication, 1990, 9(4): 351-356.
- [18] VARGA A, STEENEKEN H J M. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems [J]. Speech Communication, 1993, 12(3): 247-251.
- [19] ALLEN J B, BERKLEY D A. Image method for efficiently simulating small-room acoustics [J]. The Journal of the Acoustical Society of America, 1979, 65(4): 943-950.
- [20] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs [C] // 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Salt Lake City, UT. IEEE, 2001: 749-752.
- [21] JENSEN J, TAAL C H. An algorithm for predicting the intelligibility of speech masked by modulated noise maskers [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(11): 2009-2022.
- [22] VINCENT E, GRIBONVAL R, FEVOTTE C. Performance measurement in blind audio source separation [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(4): 1462-1469.
- [23] FALK T H, ZHENG Chenxi, CHAN W Y. A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(7): 1766-1774.
- [24] MARKOVICH S, GANNOT S, COHEN I. Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2009, 17(6): 1071-1086.

#### 作者简介



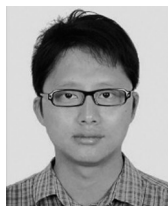
**孟维鑫** 男,1997年生,山东东营人。中国科学院大学、中国科学院声学研究所博士生,主要研究方向为语音信号处理、阵列信号处理。

E-mail: mengweixin@mail.ioa.ac.cn



**厉剑** 男,1989年生,江苏扬州人。中国科学院声学研究所助理研究员,主要研究方向为语音信号处理、阵列信号处理。

E-mail: lijian@mail.ioa.ac.cn



**郑成诗** 男,1980年生,福建三明人。中国科学院声学研究所研究员,博士生导师,主要研究方向为语音信号处理、阵列信号处理以及机器学习。

E-mail: cszheng@mail.ioa.ac.cn



**李晓东** 男,1966年生,江苏扬州人。中国科学院声学研究所研究员,博士生导师,主要研究方向为音频/语音信号处理、主动噪声与振动控制、声与振动信号监测与分析、声学测量和计量等。

E-mail: lxd@mail.ioa.ac.cn