

## 基于混合混响模型的多通道语音增强算法

解元<sup>1</sup>, 邹涛<sup>1</sup>, 孙为军<sup>2</sup>, 谢胜利<sup>2</sup>

(1. 广州大学机械与电气工程学院, 广东 广州 510006;

2. 广东工业大学物联网智能信息处理与系统集成教育部重点实验室, 广东 广州 510006)

**摘要:** 为了解决带混响和噪声场景下的语音增强问题, 构建了一个集成多通道线性预测模型和空间相干模型的语音增强模型, 设计了一种基于混合混响模型的多通道语音增强算法。该算法将后期混响分为 2 个分量, 分别用多通道线性预测模型和空间相干模型来建模, 为优化模型参数, 利用卡尔曼滤波器实施更新模型参数, 并用多项式矩阵特征值分解进行空间、时间和频率解相关, 实现去混响去噪声。实验结果表明, 所提算法可以实现高低混响带噪声环境下的语音增强, 相比于流行的语音增强算法, 其增强效果更优越, 其中语音质量客观评价 (PESQ) 值和短时客观可懂度 (STOI) 值最高分别提高了 30% 和 20%。

**关键词:** 多通道语音增强; 卡尔曼滤波器; 多项式矩阵特征值分解

**中图分类号:** TP18

**文献标志码:** A

**DOI:** 10.11959/j.issn.1000-436x.2024197

## Multichannel speech enhancement algorithm based on hybrid reverberation model

XIE Yuan<sup>1</sup>, ZOU Tao<sup>1</sup>, SUN Weijun<sup>2</sup>, XIE Shengli<sup>2</sup>

1. School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China

2. Key Laboratory of Intelligent Information Processing and System Integration of Internet of Things, Ministry of Education, Guangdong University of Technology, Guangzhou 510006, China

**Abstract:** To solve the speech enhancement problem in reverberation and noise scenarios, a new speech enhancement model was constructed integrating multichannel linear prediction model and spatial coherence model, and then a multichannel speech enhancement algorithm based on a hybrid reverberation model was designed. The post-reverberation was divided into two components, which were modeled using a multichannel linear prediction model and a spatial coherence model, respectively. To optimize the model parameters, a Kalman filter was used to update the model parameters and polynomial matrix eigenvalue decomposition was used for spatial, temporal, and frequency decorrelation to achieve reverberation and noise reduction. Experimental results show that the proposed algorithm can enhance speech in high and low-reverberation noise environments, and its enhancement effect is superior to popular speech enhancement algorithms, the performance indicators of speech enhancement, perceptual evaluation of speech quality score (PESQ) value and short-time objective intelligibility (STOI) value, have increased by 30% and 20%, respectively.

**Keywords:** multichannel speech enhancement, Kalman filter, polynomial matrix eigenvalue decomposition

收稿日期: 2024-07-04; 修回日期: 2024-11-01

通信作者: 邹涛, tzou@gzhu.edu.cn

基金项目: 广州市基础与应用基础研究基金资助项目 (No.SL2022A04J00289); 国家自然科学基金资助项目 (No.62003095, No.52171331); 广东省基础与应用基础研究基金资助项目 (No.2023A1515011311); 广州市市校联合实验室基金资助项目 (No.2023A03J0120)

**Foundation Items:** Guangzhou Basic and Applied Basic Research Project (No.SL2022A04J00289), The National Natural Science Foundation of China (No.62003095, No.52171331), Guangdong Provincial Basic and Applied Basic Research Fund (No.2023A1515011311), Guangzhou City School Joint Laboratory Project (No.2023A03J0120)

## 0 引言

在传声器语音增强中,记录的传声器信号是语音信号、混响以及噪声的混合,为了提高目标语音信号的清晰度和可懂度,需要增强目标信号,抑制混响和噪声。语音增强算法旨在减少干扰噪声和混响,同时最大限度地减少目标语音信号的失真,其已经广泛应用于电信、助听器、自动语音识别和语音控制系统等领域<sup>[1-4]</sup>。现有的语音增强算法可以分为单通道和多通道2类,其中单通道语音增强算法包括单通道维纳滤波法、谱减法、自适应滤波法、基于子空间的方法以及深度学习法等<sup>[5-10]</sup>。多通道语音增强,相对于单通道的来说,多通道语音增强需要考虑声音在空间中的传播特性,包括直达声和反射声。而且不同麦克风接收到的信号会有相位差异,从而影响信号的相干性。同时,多通道环境中的混响和噪声更加复杂,因为声音会在多个界面上反射,导致声音信号在空间中的分布更加不均匀。因此,本文考虑多通道语音增强问题,设计了一种基于混合混响模型的多通道语音增强算法。

在多通道混响环境下的语音增强问题建模方面,2种流行的混响模型分别是多通道线性预测模型和空间相干模型,其中多通道线性预测使用延迟的多通道线性预测器独立地对每个频点中的混响信号的短时傅里叶变换系数进行建模,并且将预测残差作为期望的早期反射分量,已用于估计阵列配置的多通道传声器处的早期反射分量<sup>[11-12]</sup>。Chetupalli等<sup>[13]</sup>提出一种基于多通道线性预测的延迟混响消除的贝叶斯学习方法;Yoshioka等<sup>[14]</sup>提出基于加权预测误差(WPE, weighted prediction error)算法,其从过去的观测信号中预测后期混响,然后从观测信号中减去后期混响,以恢复被分离的语音信号或期望信号(包括直接路径和早期反射);Talmon等<sup>[15]</sup>提出一种短时间傅里叶变换域中的传递函数广义旁瓣消除器波束形成器,依赖于不同传感器之间相对传递函数的卷积传递函数近似,抑制在混响环境中传感器捕获的语音信号反射;Schwartz等<sup>[16]</sup>提出基于最小方差无失真响应(MVDR, minimum variance distortionless response)算法,使用多传声器最小均方误差估计器来获得早期语音分量的空间滤波,同时抑制早期反射、后期混响和环境噪声;Dietzen等<sup>[17]</sup>提出

基于瞬时功率谱密度(IPSD, instantaneous power spectral density)算法,从新定义的瞬时广义特征值估计功率谱密度(PSD),产生时间平滑PSD估计;Dietzen等<sup>[18]</sup>提出基于集成旁瓣对消和线性预测(ISCLP, integrated sidelobe cancellation and linear prediction)算法,建立了多通道线性预测模型。然而,多通道线性预测模型更适用于短混响时间和中混响时间的情况,在长混响时间情况下,多通道线性预测模型的计算代价是巨大的,语音增强效果受限。

空间相干模型是在后期混响声场空间均匀的假设下,其功率谱密度矩阵可以建模为时变的功率谱密度乘以时不变的空间相干矩阵<sup>[19]</sup>。Kodrasi等<sup>[20]</sup>提出了一种基于特征值分解的延迟混响功率谱密度估计器;Neo等<sup>[21]</sup>提出基于多项式矩阵特征值分解(PEVD, polynomial matrix eigenvalue decomposition)算法,利用多项式矩阵作为语音的宽带多通道信号,建立空间相干模型,实现高混响下的语音增强。然而,在混响时间不够长时,空间相干模型的空间均匀声场假设的有效性值得怀疑。因此,该模型适用于混响时间较长的情况。

盲源分离是一种已被广泛研究的源信号分离技术,其是在混叠通道信息未知的情况下仅从观测到的混合信号中分离出源信号,该技术已经应用于语音增强领域<sup>[22-23]</sup>。Sekiguchi等<sup>[24]</sup>提出一种基于迭代投影或迭代源导向的迭代优化算法,用于多通道噪声混响混合信号的盲源分离和去混响。Liu等<sup>[25]</sup>将多通道线性预测模型和空间相干模型这2种传统模型相结合,提出了一种混合混响模型,并将混合混响模型与多通道非负矩阵分解技术相结合,实现去混响和分离。Ueda等<sup>[26]</sup>提出联合WPE和独立向量提取的盲源分离技术进行语音去混响、去噪声和分离,前期提出一种基于盲信号提取的逆滤波器(BSEIF, blind signal extraction-based inverse filtering)算法,该算法利用盲信号提取技术和自适应时频类维纳掩蔽逆滤波器设计一种无监督的多通道语音增强方法,在高混响环境下实现较好的去混响去噪声效果<sup>[27]</sup>。然而,以上算法并不能同时兼顾低混响和高混响场景。

本文的设计思路是集成多通道线性预测模型和空间相干模型的优势,将后期混响分为2个分量,分别用多通道线性预测模型和空间相干模型对这2个

分量进行建模,提出一种基于混合混响模型的多通道语音增强算法,实现较好的去混响去噪声效果。

本文的创新点及贡献如下。

1) 构建了一个集成多通道线性预测模型和空间相干模型的语音增强模型,可以更加精确地对不同的混响环境建模。

2) 利用卡尔曼滤波器和多项式矩阵特征值分解设计了一种基于混合混响模型的多通道语音增强算法。

3) 对所提算法的去混响去噪声性能进行实验验证,并分别对比了几种流行的语音增强算法,实验结果证实了所提算法的有效性及其优越性。

## 1 模型

在噪声混响环境下,传声器接收的语音信号包括直接路径上的声音信号和混响,以及额外噪声,其中混响又分为早期反射声音和后期反射声音,由于早期反射声音对增加直接路径声音的强度有影响,从而对语音的可懂度产生积极影响。因此,将早期反射的声音和直接路径上的声音作为理想的语音信号。根据卷积传递函数的特点,后期混响可分为2个分量。第一个分量被认为是后期混响的空间异构分量,用多通道线性预测模型对其进行建模。第二个分量假定遵循空间均匀声场的假设,该声场是用空间相干模型建模的。因此,利用 $M$ 个传声器采集噪声混响语音信号,在短时傅里叶变换域中构建第 $m$ 个传声器采集的带噪声混响语音信号 $x_m(t)$ 的数学模型。

$$x_m(t) = h_m(t) * s(t) + v_m(t) = \sum_{\tau=0}^{\Delta-1} h_m(\tau) s(t-\tau) + \sum_{\tau=\Delta}^J h_m(\tau) s(t-\tau) + \sum_{\tau=J+1}^{L_h-1} h_m(\tau) s(t-\tau) + v_m(t) \quad (1)$$

其中,\*代表卷积, $s(t)$ 是纯净的目标语音信号; $h_m(t)$ 是声音到达第 $m$ 个传声器之间的房间脉冲响应,其由长度为 $L_h$ 的有限脉冲响应滤波器构建; $t$ 是短时傅里叶变换域中的时间框; $\sum_{\tau=0}^{\Delta-1} h_m(\tau) s(t-\tau)$ 为早期混响模型, $\Delta$ 定义为早期混响和后期混响的边界; $\sum_{\tau=\Delta}^J h_m(\tau) s(t-\tau)$ 为多通道线性预测模型;

$\sum_{\tau=J+1}^{L_h-1} h_m(\tau) s(t-\tau)$ 为空间相干模型; $J$ 定义为后期混响中多通道线性预测模型和空间相干模型之间的边界; $v_m(t)$ 是额外噪声。由于多通道线性预测模型适合于中短混响时间,空间相干模型适合于较长混响时间,本文构建的混合模型可以充分发挥多通道线性预测模型和空间相干模型的优势,针对不同的混响时间长度,实现较好的语音增强效果。

利用向量乘积的表示形式,式(1)可写为

$$x_m(t) = \mathbf{h}_{m,d}^T \mathbf{s}_d(t) + \mathbf{h}_{m,l_1}^T \mathbf{s}_{l_1}(t) + \mathbf{h}_{m,l_2}^T \mathbf{s}_{l_2}(t) + v_m(t) \quad (2)$$

其中, $(\cdot)^T$ 代表转置, $\mathbf{h}_{m,d} = [h_m(0), h_m(1), \dots, h_m(\Delta-1)]^T$ 是直接路径和早期反射的房间脉冲响应系数向量, $\mathbf{s}_d(t) = [s(t), s(t-1), \dots, s(t-\Delta+1)]^T$ 是期望的语音信号向量, $\mathbf{h}_{m,l_1} = [h_m(\Delta), h_m(\Delta+1), \dots, h_m(J)]^T$ 是长度为 $J-\Delta+1$ 后期混响的房间脉冲响应系数向量, $\mathbf{s}_{l_1}(t) = [s(t-\Delta), s(t-\Delta-1), \dots, s(t-J)]^T$ 是长度为 $J-\Delta+1$ 的后期混响语音信号向量, $\mathbf{h}_{m,l_2} = [h_m(J+1), h_m(J+2), \dots, h_m(L_h-1)]^T$ 是长度为 $L_h-1-J$ 的后期混响的房间脉冲响应系数向量, $\mathbf{s}_{l_2}(t) = [s(t-J-1), s(t-J-2), \dots, s(t-L_h+1)]^T$ 是长度为 $L_h-1-J$ 的后期混响语音信号向量,通常假设后期混响和附加噪声都与直接路径上的声音和早期混响不相关。因此,利用 $M$ 个传声器采集的语音信号可以建模为

$$\begin{cases} x_1(t) = \mathbf{h}_{1,d}^T \mathbf{s}_d(t) + \mathbf{h}_{1,l_1}^T \mathbf{s}_{l_1}(t) + \mathbf{h}_{1,l_2}^T \mathbf{s}_{l_2}(t) + v_1(t) \\ x_2(t) = \mathbf{h}_{2,d}^T \mathbf{s}_d(t) + \mathbf{h}_{2,l_1}^T \mathbf{s}_{l_1}(t) + \mathbf{h}_{2,l_2}^T \mathbf{s}_{l_2}(t) + v_2(t) \\ \vdots \\ x_M(t) = \mathbf{h}_{M,d}^T \mathbf{s}_d(t) + \mathbf{h}_{M,l_1}^T \mathbf{s}_{l_1}(t) + \mathbf{h}_{M,l_2}^T \mathbf{s}_{l_2}(t) + v_M(t) \end{cases} \quad (3)$$

将式(3)改写为矩阵向量的形式为

$$\mathbf{x}(t) = \mathbf{H} \tilde{\mathbf{s}}(t) + \mathbf{v}(t) \quad (4)$$

其中, $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$ ,  $\mathbf{H} =$

$$\begin{bmatrix} \mathbf{h}_{1,d}^T & \mathbf{h}_{1,l_1}^T & \mathbf{h}_{1,l_2}^T \\ \mathbf{h}_{2,d}^T & \mathbf{h}_{2,l_1}^T & \mathbf{h}_{2,l_2}^T \\ \vdots & \vdots & \vdots \\ \mathbf{h}_{M,d}^T & \mathbf{h}_{M,l_1}^T & \mathbf{h}_{M,l_2}^T \end{bmatrix}, \quad \tilde{\mathbf{s}}(t) = [\mathbf{s}_d(t) \quad \mathbf{s}_{l_1}(t) \quad \mathbf{s}_{l_2}(t)]^T,$$

$\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_M(t)]^T$ 。本文的目的是通过接收的传声器语音信号  $\mathbf{x}(t)$ ，获取理想的语音信号  $s_d(t)$ ，实现语音增强的效果。

## 2 算法

本文设计的算法将后期混响分为 2 个分量，分别用多通道线性预测模型和空间相干模型来建模，为优化模型参数，利用卡尔曼滤波器对模型参数进行更新，并用多项式矩阵特征值分解进行空间、时间和频率解相关，实现去混响去噪声。算法流程如图 1 所示。

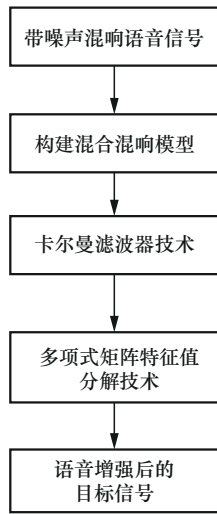


图 1 算法流程

首先对后期混响  $\mathbf{h}_{m,l_1}^T \mathbf{s}_{l_1}(t)$ ，基于多通道线性预测模型，在不同的传声器信号帧上利用旁瓣消除滤波器和线性预测滤波器并行操作，构造状态空间模型。

$$\hat{\mathbf{x}}(t) = \mathbf{F}\hat{\mathbf{x}}(t-1) + \mathbf{B}d(t) \quad (5)$$

其中， $\hat{\mathbf{x}}(t)$  是  $t$  时刻的预测值， $\mathbf{F}$  是状态转移矩阵， $\hat{\mathbf{x}}(t-1)$  是时刻  $t-1$  的最优估计值， $d(t)$  是输入向量； $\mathbf{B}$  是输入状态矩阵。计算状态协方差矩阵

$$\mathbf{P}(t) = \mathbf{F}\mathbf{P}(t-1)\mathbf{F}^H + \mathbf{B}\mathbf{Q}\mathbf{B}^H \quad (6)$$

其中， $\mathbf{P}(t)$  是  $t$  时刻的预测状态协方差矩阵， $\mathbf{P}(t-1)$  是  $t-1$  时刻的最优估计状态协方差矩阵， $(\cdot)^H$  代表 Hermitian 转置， $\mathbf{Q}$  是输入状态协方差矩阵。

定义观测状态方程为

$$\mathbf{Z}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{V} \quad (7)$$

其中， $\mathbf{A}$  是状态观测矩阵， $\mathbf{V}$  是观测噪声矩阵。设

计卡尔曼增益为

$$\mathbf{K}(t) = \frac{\mathbf{P}(t)\mathbf{A}^H}{\mathbf{A}\mathbf{P}(t)\mathbf{A}^H + \mathbf{P}_{VV}} \quad (8)$$

其中， $\mathbf{P}_{VV}$  是观测噪声协方差矩阵。

对状态  $\hat{\mathbf{x}}(t)$  进行更新学习，有

$$\hat{\mathbf{x}}^+(t) = \hat{\mathbf{x}}(t) + \mathbf{K}(t)(\mathbf{Z}(t) - \mathbf{A}\hat{\mathbf{x}}(t)) \quad (9)$$

且

$$\mathbf{P}^+(t) = (\mathbf{I} - \mathbf{K}(t)\mathbf{A})\mathbf{P}(t) \quad (10)$$

其中， $\hat{\mathbf{x}}^+(t)$  和  $\mathbf{P}^+(t)$  分别是更新后的状态和状态协方差矩阵。

通过以上的卡尔曼滤波器更新方程，每个递归方程通常包含两组不同的更新，即反映状态演化的先验时间更新和当前测量值的后验测量更新，利用这些递归方程估计真实状态，实现模型参数的实时更新<sup>[18]</sup>。

然后对后期混响  $\mathbf{h}_{m,l_2}^T \mathbf{s}_{l_2}(t)$ ，基于空间相干模型，同时假设后期混响跟随扩散声场，满足声场空间均匀，基于空间相干模型，利用多项式矩阵特征值分解进行空间、时间和频率解相关。

定义协方差矩阵为

$$\mathbf{R}_{xx}(\tau) = \mathbf{E}[\mathbf{x}(t)\mathbf{x}^H(t-\tau)] \quad (11)$$

其中， $\mathbf{E}[\cdot]$  代表期望， $\mathbf{R}_{xx}(\tau)$  的第  $(p,q)$  元素表示为  $r_{p,q}(\tau) = \mathbf{E}[x_p(t)x_q^*(t-\tau)]$ ，且  $(\cdot)^*$  代表复共轭，对式(11)利用  $z$  变换，有

$$\mathbf{R}_{xx}(z) = \sum_{\tau=-\infty}^{\tau=+\infty} \mathbf{R}_{xx}(\tau)z^{-\tau} \quad (12)$$

选择  $z$  变换的窗函数长度为  $W$ ，且  $W \gg \Delta$ ，每一帧的尺寸为  $T$ ，假设每一帧上的信号是平稳的，则式(11)可以近似估计为

$$\hat{\mathbf{R}}_{xx}(\tau) \approx \frac{1}{T+1} \sum_{t=0}^T \mathbf{x}(t)\mathbf{x}^H(t-\tau) \quad (13)$$

同时，式(11)近似为

$$\hat{\mathbf{R}}_{xx}(z) \approx \sum_{\tau=-W}^{\tau=W} \mathbf{R}_{xx}(\tau)z^{-\tau} \quad (14)$$

结合式(2)、式(4)和式(11)，得

$$\begin{aligned} \mathbf{R}_{xx}(z) &= \mathbf{H}_d \mathbf{s}_d(z) \mathbf{s}_d^T(z) \mathbf{H}_d^T + \mathbf{H}_{l_1} \mathbf{s}_{l_1}(z) \mathbf{s}_{l_1}^T(z) \mathbf{H}_{l_1}^T + \\ &\mathbf{H}_{l_2} \mathbf{s}_{l_2}(z) \mathbf{s}_{l_2}^T(z) \mathbf{H}_{l_2}^T + \mathbf{R}_{vv}(z) = \\ &\mathbf{R}_{s_d s_d}(z) + \mathbf{R}_{s_{l_1} s_{l_1}}(z) + \mathbf{R}_{s_{l_2} s_{l_2}}(z) + \mathbf{R}_{vv}(z) \end{aligned} \quad (15)$$

其中， $\mathbf{H}_d = [\mathbf{h}_{1,d}^T, \mathbf{h}_{2,d}^T, \dots, \mathbf{h}_{M,d}^T]^T$ ， $\mathbf{H}_{l_1} = [\mathbf{h}_{1,l_1}^T, \mathbf{h}_{2,l_1}^T, \dots, \mathbf{h}_{M,l_1}^T]^T$ ， $\mathbf{H}_{l_2} = [\mathbf{h}_{1,l_2}^T, \mathbf{h}_{2,l_2}^T, \dots, \mathbf{h}_{M,l_2}^T]^T$ ， $\mathbf{R}_{s_d s_d}(z) = \mathbf{H}_d \mathbf{s}_d(z)$

$\mathbf{s}_d^T(z)\mathbf{H}_d^T$ ,  $\mathbf{R}_{s_1s_1}(z) = \mathbf{H}_{l_1}\mathbf{s}_{l_1}(z)\mathbf{s}_{l_1}^T(z)\mathbf{H}_{l_1}^T$ ,  $\mathbf{R}_{s_2s_2}(z) = \mathbf{H}_{l_2}\mathbf{s}_{l_2}(z)\mathbf{s}_{l_2}^T(z)\mathbf{H}_{l_2}^T$ ,  $\mathbf{R}_v(z)$ 是噪声协方差矩阵。

同时后期混响  $\mathbf{h}_{m,l_2}^T\mathbf{s}_{l_2}(t)$  功率谱密度矩阵可以建模为时变的功率谱密度乘以时不变的空间相干矩阵, 即

$$\mathbf{R}_{s_2s_2}(z) = \varphi(z)\mathbf{\Gamma}(z) \quad (16)$$

其中,  $\varphi(z)$ 是时变后期混响的功率谱密度;  $\mathbf{\Gamma}(z)$ 是时不变的空间相干矩阵, 对于全向传声器, 其第  $(i,j)$  元素表示为

$$\Gamma_{ij}(z) = \sin c \frac{2\pi d_{ij}(t-1)f_s}{Wv} \quad (17)$$

其中,  $\sin c(\cdot) = \frac{\sin(\cdot)}{(\cdot)}$ ,  $d_{ij}$  是第  $(i,j)$  传声器对之间的距离,  $f_s$  是样本频率,  $v$  是声音的速度。

定义准厄米特矩阵的多项式特征值分解为

$$\mathbf{R}_{xx}(z) \approx \mathbf{U}^P(z)\mathbf{A}(z)\mathbf{U}(z) \quad (18)$$

其中,  $(\cdot)^P$  代表 para-Hermitian 矩阵运算, 且  $\mathbf{U}^P(z) = \mathbf{U}^H(z^{-1})$ ,  $\mathbf{U}(z)$  是特征向量多项式矩阵,  $\mathbf{A}(z)$  是对角多项式矩阵, 由于假设后期混响和附加噪声都与直接路径上的声音和早期混响不相关, 结合式(15)和式(18), 得

$$\mathbf{R}_{xx}(z) \approx \begin{bmatrix} \mathbf{U}_{s_d}^P(z) \\ \mathbf{U}_{s_1}^P(z) \\ \mathbf{U}_{s_2}^P(z) \\ \mathbf{U}_v^P(z) \end{bmatrix}^T \begin{bmatrix} \mathbf{A}_{s_d}(z) & & & \\ & \mathbf{A}_{s_1}(z) & & \\ & & \mathbf{A}_{s_2}(z) & \\ & & & \mathbf{A}_v(z) \end{bmatrix} \begin{bmatrix} \mathbf{U}_{s_d}(z) \\ \mathbf{U}_{s_1}(z) \\ \mathbf{U}_{s_2}(z) \\ \mathbf{U}_v(z) \end{bmatrix} \quad (19)$$

其中, 特征向量多项式矩阵  $\mathbf{U}(z) = \begin{bmatrix} \mathbf{U}_{s_d}(z) & \mathbf{U}_{s_1}(z) & \mathbf{U}_{s_2}(z) & \mathbf{U}_v(z) \end{bmatrix}^T$  可利用序列矩阵对角化迭代算法获得, 其第  $i$  次迭代满足

$$\mathbf{U}_i^P(z)\mathbf{U}_i(z) = \mathbf{U}_i(z)\mathbf{U}_i^P(z) = \mathbf{I} \quad (20)$$

其中,  $\mathbf{I}$  是单位矩阵, 利用获得的  $\mathbf{U}(z)$ , 实现语音增强。

$$\mathbf{y}(z) = \mathbf{U}(z)\mathbf{x}(z) \quad (21)$$

其中,  $\mathbf{x}(z)$  是  $\mathbf{x}(t)$  的  $z$  变换,  $\mathbf{y}(z)$  是强去相关的, 满足

$$\mathbf{E}[\mathbf{y}(z)\mathbf{y}^P(z)] = \mathbf{E}[\mathbf{U}(z)\mathbf{x}(z)\mathbf{x}^P(z)\mathbf{U}^P(z)] \approx \mathbf{A}(z) \quad (22)$$

因此, 通过组合信号子空间中的分量和噪声子

空间中零化分量来实现输出信道中的去噪去混响, 从而获取理想的语音信号  $\mathbf{s}_d(z)$ 。

$$\mathbf{s}_d(z) = \mathbf{U}_d(z)\mathbf{x}(z) \quad (23)$$

且

$$\mathbf{E}[\mathbf{s}_d(z)\mathbf{s}_d^P(z)] = \mathbf{E}[\mathbf{U}_d(z)\mathbf{x}(z)\mathbf{x}^P(z)\mathbf{U}_d^P(z)] \approx \mathbf{A}_d(z) \quad (24)$$

由于频谱优化特性, 基于混合混响模型的多通道语音增强算法倾向于按信号能量的降序对强去相关输出进行排序。因此, 定义具有最大总频谱功率的信号为第一信道, 语音增强后的信号  $\mathbf{s}_d(t)$  在  $\mathbf{s}_d(z)$  信号的第一通道中。

综上, 基于混合混响模型的多通道语音增强算法利用以上的卡尔曼滤波器技术和多项式矩阵特征值分解技术, 实现高低不同混响且带噪声环境下的语音增强, 整体流程如算法 1 所示。

**算法 1** 基于混合混响模型的多通道语音增强算法

**输入** 多通道带噪声混响语音信号  $\mathbf{x}(t)$

**输出** 语音增强后的信号  $\mathbf{s}_d(t)$

步骤 1 构建时频域语音信号模型

$$\mathbf{x}(t) = \mathbf{H}\tilde{\mathbf{s}}(t) + \mathbf{v}(t)$$

步骤 2 卡尔曼滤波器技术

1) 构造状态空间模型

$$\hat{\mathbf{x}}(t) = \mathbf{F}\hat{\mathbf{x}}(t-1) + \mathbf{B}d(t)$$

2) 计算状态协方差矩阵

$$\mathbf{P}(t) = \mathbf{F}\mathbf{P}(t-1)\mathbf{F}^H + \mathbf{B}\mathbf{Q}\mathbf{B}^H$$

3) 定义观测状态方程:

$$\mathbf{Z}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{V}$$

4) 设计卡尔曼增益

$$\mathbf{K}(t) = \frac{\mathbf{P}(t)\mathbf{A}^H}{\mathbf{A}\mathbf{P}(t)\mathbf{A}^H + \mathbf{P}_{VV}}$$

5) 对状态  $\hat{\mathbf{x}}(t)$  进行更新学习

$$\hat{\mathbf{x}}^+(t) = \hat{\mathbf{x}}(t) + \mathbf{K}(t)(\mathbf{Z}(t) - \mathbf{A}\hat{\mathbf{x}}(t))$$

步骤 3 多项式矩阵特征值分解技术

1) 定义协方差矩阵

$$\mathbf{R}_{xx}(\tau) = \mathbf{E}[\mathbf{x}(t)\mathbf{x}^H(t-\tau)]$$

2) 利用  $z$  变换

$$\mathbf{R}_{xx}(z) = \sum_{\tau=-\infty}^{\tau=\infty} \mathbf{R}_{xx}(\tau)z^{-\tau}$$

3) 定义准厄米特矩阵的多项式特征值分解

$$\mathbf{R}_{xx}(z) \approx \mathbf{U}^P(z)\mathbf{A}(z)\mathbf{U}(z)$$

4) 得

$$U(z) = \begin{bmatrix} U_{s_d}(z) & U_{s_{t_1}}(z) & U_{s_{t_2}}(z) & U_v(z) \end{bmatrix}^T$$

$$A(z) \approx E[U(z)x(z)x^P(z)U^P(z)]$$

5) 获取理想的语音信号

$$s_d(z) = U_d(z)x(z)$$

步骤 4 语音增强后的信号  $s_d(t)$  位于  $s_d(z)$  的第一通道中。

### 3 实验

#### 3.1 实验环境和评估准则

在本文的仿真实验中, 假设早期混响和后期混响的边界为  $\Delta = 0.1$  s, 后期混响中多通道线性预测模型和空间相干模型之间的边界为  $J = 0.7$  s, 根据不同的数据集,  $J$  的选取也有一定偏差, 从理论上严格证明 2 个模型之间的边界值是一项极具挑战性的工作。语音信号的采样频率为 16 kHz。在应用卡尔曼滤波器技术阶段, 短时傅里叶变换分析和合成使用平方根 Hanning 窗, 采样频率为 512 个采样点 50% 重叠。在多项式矩阵特征值分解阶段, 选择  $z$  变换的窗函数长度为  $W = 1600$ , 每一帧的尺寸为  $T = 1600$ 。实验选用混响时间分别为 0.61 s 和 1.22 s, 其中混响时间为 0.61 s 的语音信号来自线性传声器阵列的多通道音频数据库, 线性传声器阵列包含 5 个传声器, 传声器间距为 8 cm, 声源位于声器阵列的 2 m 距离处<sup>[28]</sup>; 混响时间为 1.22 s 的语音信号选用来自国际上公开的语音库<sup>[29]</sup>; 脉冲响应测量和噪声记录取自完整的环境的声学特性 (ACE, acoustic characterization of environment) 数据库, 使用 3 通道的传声器阵列<sup>[30]</sup>。噪声来自公开的噪声数据库, 选用其中的啾呀嘈杂语音作为额外噪声<sup>[31]</sup>。定义信噪比 (SNR, signal-to-noise ratio) 为语音信号和噪声在时域中的功率比, 本文考虑 SNR 为 -10~20 dB, 分别验证所提算法的去噪去混响性能。

为评价所提算法的性能, 以直达声信号为参考信号, 选择 2 种国际上公用的语音增强性能评价准则: 语音质量客观评价 (PESQ, perceptual evaluation of speech quality score), 其值为 -0.5~4.5<sup>[32]</sup>; 短时客观可懂度 (STOI, short-time objective intelligibility), 其值为 0~1<sup>[33]</sup>。PESQ 值和 STOI 值越大, 说明语音增强的性能越好。

为了对比本文所提算法的优越性, 选择以下对比算法: 基于最小方差无失真响应 (MVDR) 算法<sup>[16]</sup>、基于瞬时功率谱密度 (IPSD) 算法<sup>[17]</sup>、基

于集成旁瓣对消和线性预测 (ISCLP) 算法<sup>[18]</sup>、基于多项式矩阵特征值分解 (PEVD) 算法<sup>[21]</sup>, 以及基于盲信号提取的逆滤波器 (BSEIF) 算法<sup>[27]</sup>, 其中 MVDR、IPSD 和 ISCLP 算法是经典的基于多通道线性预测模型的多通道语音增强算法, 在处理低混响下的语音增强显著; PEVD 算法是基于空间相干模型的无监督多通道语音增强算法, BSEIF 算法是基于盲信号提取技术和自适应时频类维纳掩蔽逆滤波器设计的一种多通道语音增强方法, 在高混响场景下具有较好的优势。本文所提算法是集成多通道线性预测模型和空间相干模型设计的一种多通道语音增强算法, 对比算法的实施参数与所提算法一致, 通过与这些流行算法对比, 可以验证所提算法的有效性和优越性。

#### 3.2 实验结果

本节进行去混响和去噪声的仿真实验验证所提算法的有效性, 同时对实验结果进行分析, 证实所提算法的优越性。

##### 3.2.1 去混响实验

###### 1) 实验 1

本实验考虑无噪声带混响语音信号的去混响效果, 混响时间为 0.61 s, 为了验证通道数对去混响的影响, 测试通道数设置为 2~5, 实验结果如图 2 所示。由图 2 可以看出, 基于多通道线性预测模型的算法在 0.61 s 混响时间下的性能比基于空间相干模型的算法具有明显的优势。基于盲信号提取的逆滤波器算法在两通道情形下效果最好。所提算法在不同通道数下都优于对比算法, 特别是当通道数为 3 时, 所提算法呈现了最好的去混响效果, 其中 PESQ 值和 STOI 值明显优于其他对比算法。因此, 在接下来的仿真实验中, 以通道数 3 为基础进行实验。

利用语谱图可视化了所提算法的去混响效果, 如图 3 所示, 图 3(a) 和图 3(b) 分别给出了语音源信号和带混响的语谱图, 图 3(c) 为语音增强后的语谱图。对比图 3(a) 和图 3(c) 可知, 增强后的语音信号与目标语音源信号的语谱图很接近, 说明所提算法对语音中的混响具有抑制作用, 具有去混响效果。

###### 2) 实验 2

本实验考虑混响时间为 1.22 s 的去混响效果, 如图 4 所示, 在 1.22 s 混响时间下, 基于空间相干模型的算法性能优于基于多通道线性预测模型的算

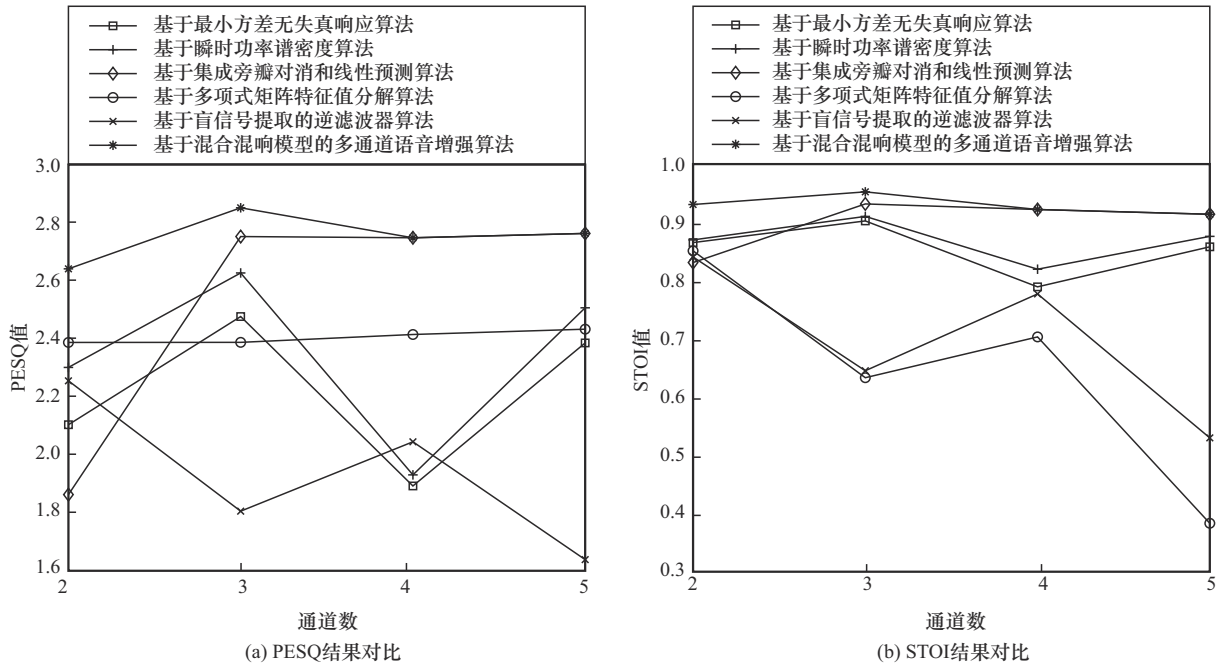


图2 通道数对去混响结果的影响

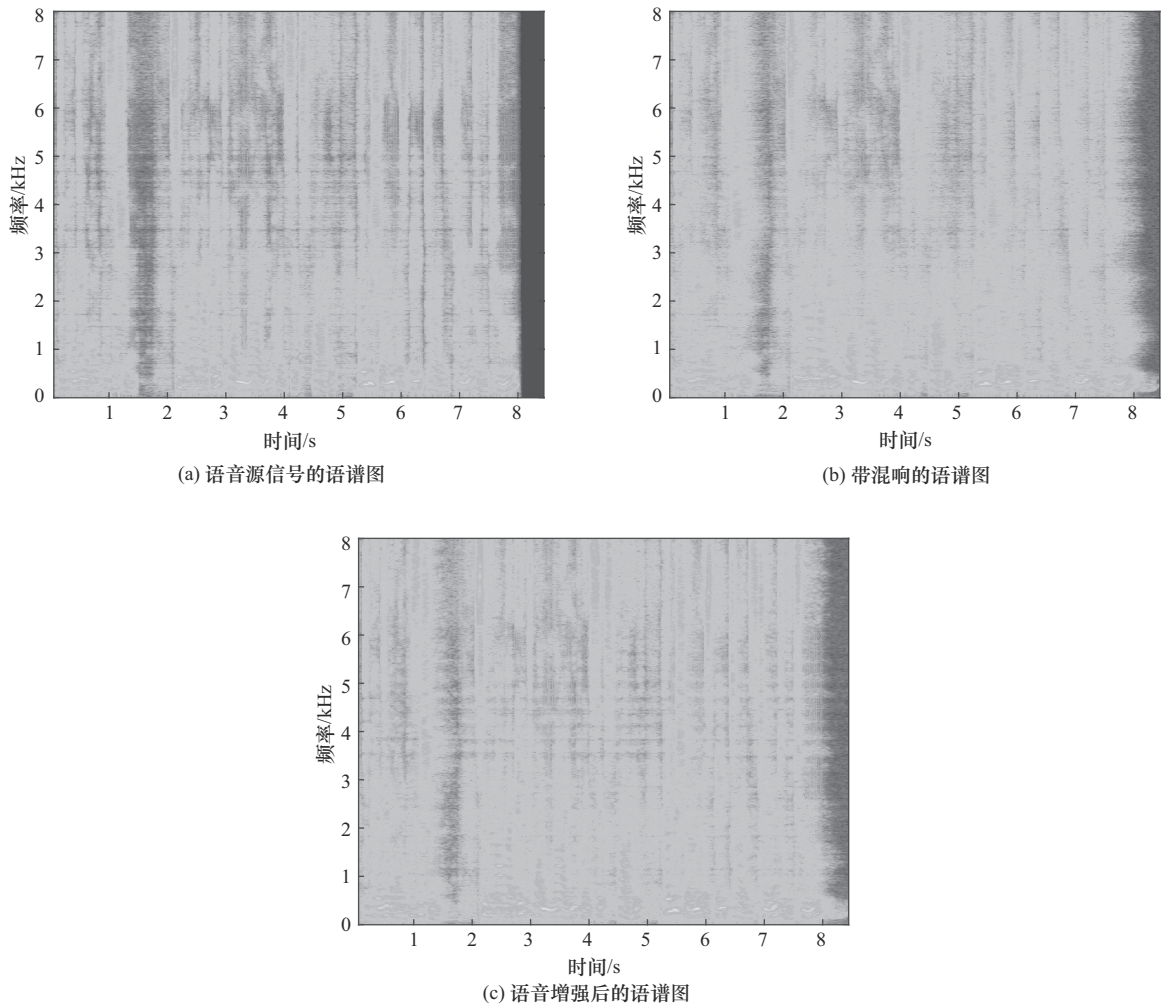


图3 0.61 s混响时间下的去混响效果

法。基于盲信号提取的逆滤波器算法去混响效果较好，其性能已经在文献[25]中得到验证。所提算法获得的 PESQ 值优于其他对比算法的，STOI 值基本持平，证实了其在高混响情形下的有效性和优越性。另外，可视化算法的去混响效果，如图 5 所示，对比图 5(a)和图 5(c)，可见增强后的语音信号与目标语音源信号的语谱图很接近，其去混响效果明显。

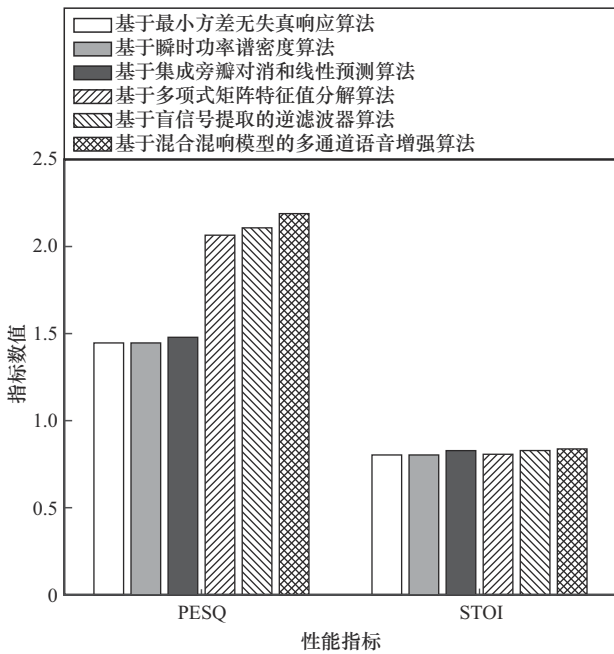
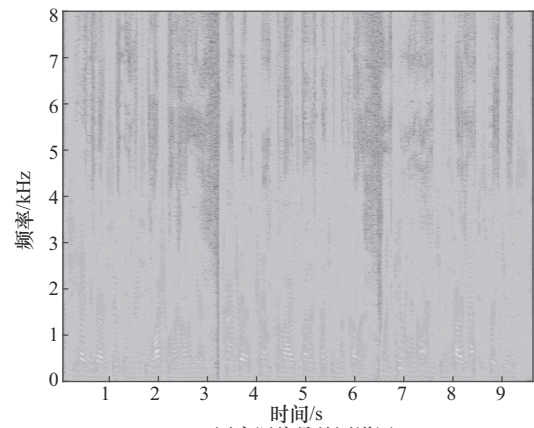


图 4 去混响性能对比

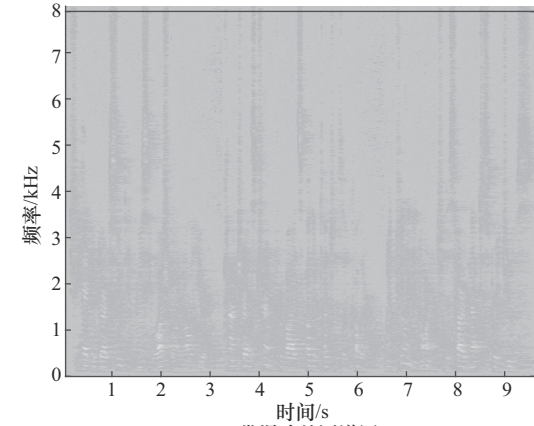
根据实验 1 和实验 2 的结果可以发现：通道数影响算法的性能，本文重点考虑通道数为 3 的仿真实验；在 0.61 s 混响时间下，基于多通道线性预测模型的算法性能优于基于空间相干模型的算法，而在 1.22 s 混响时间下，恰恰相反；所提算法兼顾 2 种模型的优势，在高低混响下都实现了较好的去混响性能。特别是在高混响环境下，与基于盲信号提取的逆滤波器算法比较，所提算法仍能实现较好的结果，验证了其优越性。

### 3.2.2 去噪声实验(实验 3)

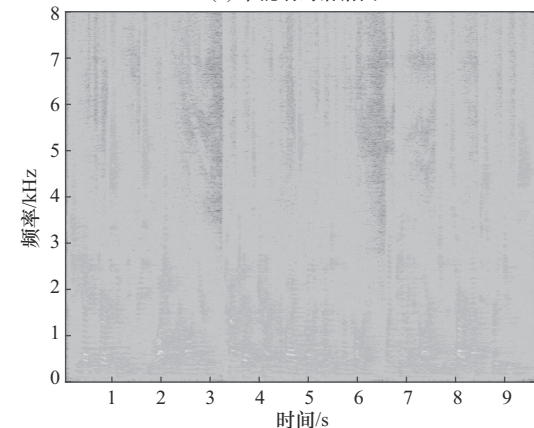
本实验选用纯净的无混响语音信号加上一组噪声信号，选用 SNR 为 -10~20 dB，合成无混响带噪声的语音混合信号进行去噪声测试，实验结果如图 6 所示。随着信噪比的增加，与预期一致，所有算法获得的 PESQ 值和 STOI 值逐渐增加，而所提算法获得结果一直优于对比算法，验证了所提算法去噪声的有效性及其优越性。



(a) 语音源信号的语谱图



(b) 带混响的语谱图



(c) 去混响后的语谱图

图 5 1.22 s 混响时间下的去混响效果

### 3.2.3 去混响去噪声实验

#### 1) 实验 4

考虑混响时间为 0.61 s 下的去混响去噪声效果，在实验 1 的带混响语音信号中添加一组噪声，信噪比仍然选用 SNR 为 -10~20 dB，实验结果如图 7 所示。对比于图 6，带有混响和噪声信号的语音增强效果下降，但是整体上仍然实现很好的去混响去噪声效果，PESQ 值和 STOI 值在较好的变化区间，特别是在 SNR=-10 dB 条件下 PESQ 值相比于当前最好的

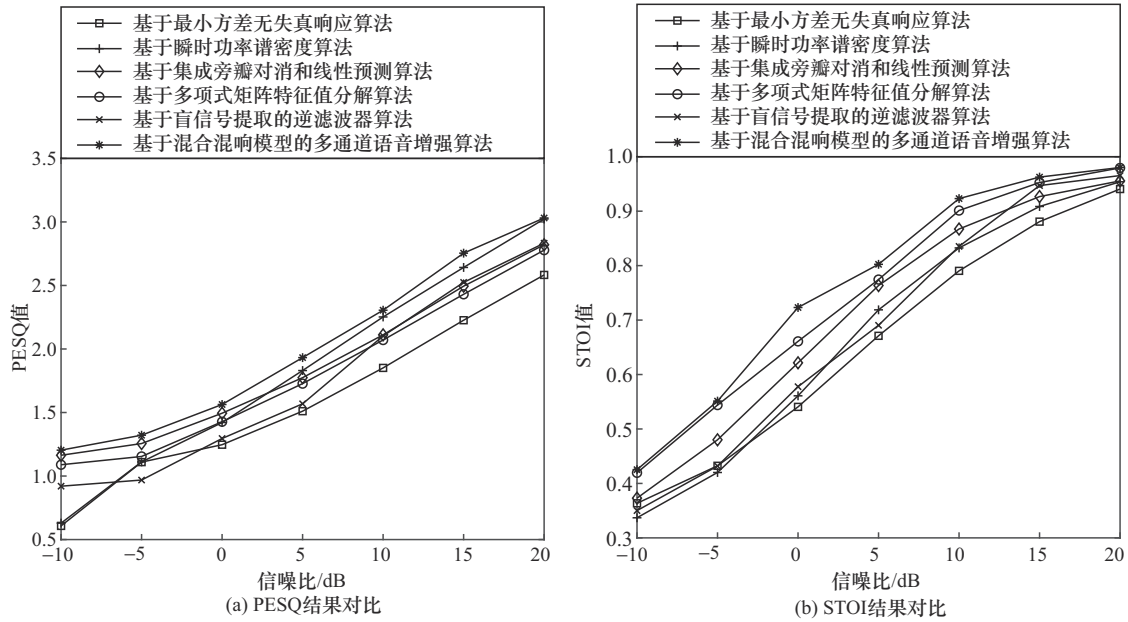


图6 去噪声性能对比

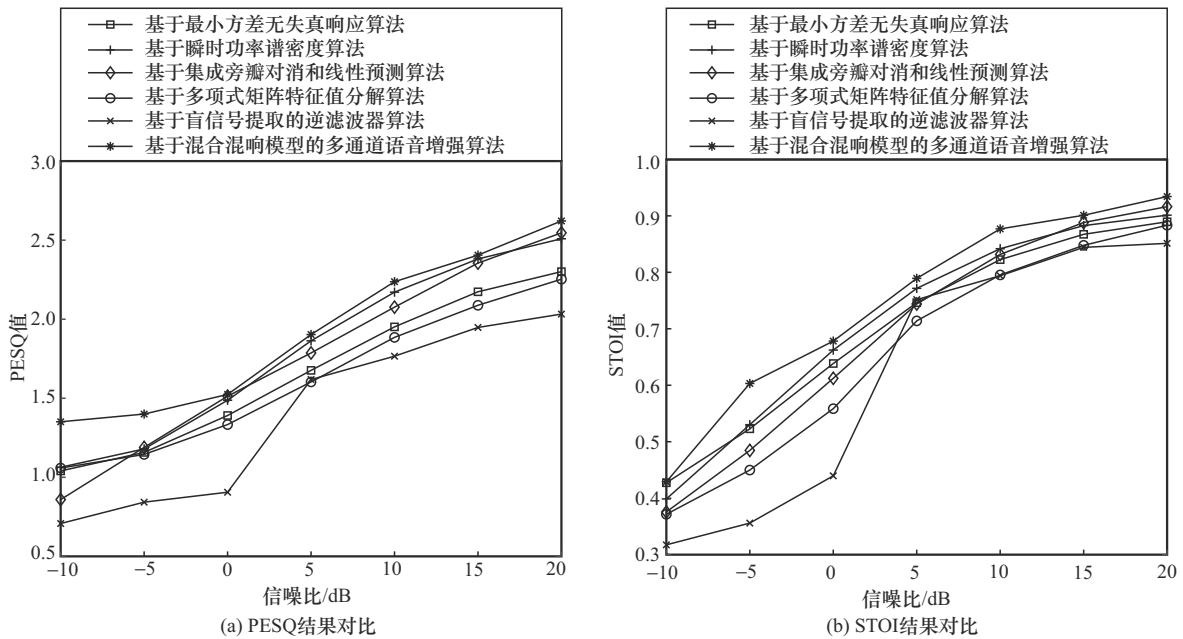


图7 混响时间为0.61 s的去混响去噪声性能对比

结果提高近30%，在SNR=-5 dB条件下STOI值相比于当前最好的结果提高近20%。另外，基于多通道线性预测模型的算法优于基于空间相干模型的算法，基于盲信号提取的逆滤波器算法最弱，而所提算法仍优于对比算法，说明所提算法在带混响和噪声环境下，具有语音增强性能的有效性和优越性。

### 2) 实验5

考虑混响时间为1.22 s下的去混响去噪声效果，其中的噪声部分是啞呀嘈杂的语音噪声，为了

考虑其鲁棒性，选用信噪为SNR为-10~20 dB，实验结果如图8所示。对比于图7，由于混响时间较大，算法获得的PESQ值和STOI值有所降低，但整体还是在有效区间内变化，说明算法在高混响下仍然是有效的。其中，基于空间相干模型的算法优于基于多通道线性预测模型的算法，基于盲信号提取的逆滤波器算法也表现较好效果，而所提算法实现了更好的语音增强效果，证实了其更适用于高混响场景，其鲁棒性更好。

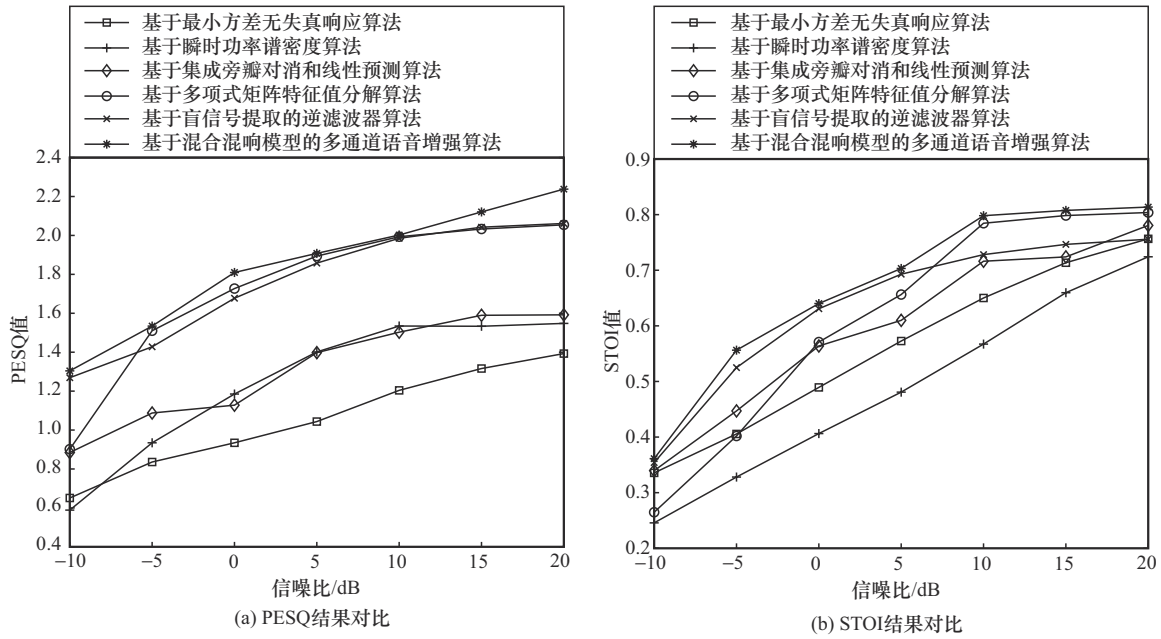


图8 混响时间为1.22 s的去混响去噪声性能对比

### 3) 实验6

在混响时间为0.61 s环境下添加一组干扰语音源信号，考虑带有2个混响语音源信号和噪声场景下的去混响去噪声效果，实验结果如图9所示。基于多通道线性预测模型的算法优于基于空间相干模型的算法和基于盲信号提取的逆滤波器算法，而所提算法一直处于较好的效果，证实了其在带有干扰性语音下的语音增强的鲁棒性。

### 3.3 实验结果分析

基于以上实验结果发现，本文设计的混合混响模型兼顾多通道线性预测模型和空间相干模型的优势，在低混响下基于多通道线性预测模型的算法优于基于空间相干模型的算法。相反，在高混响环境下基于空间相干模型的算法发挥更好的优势。而所提算法弥补了两者的缺陷，在高低混响环境下都能获得较好的增强效果，说明混合混响模型具有更好的有

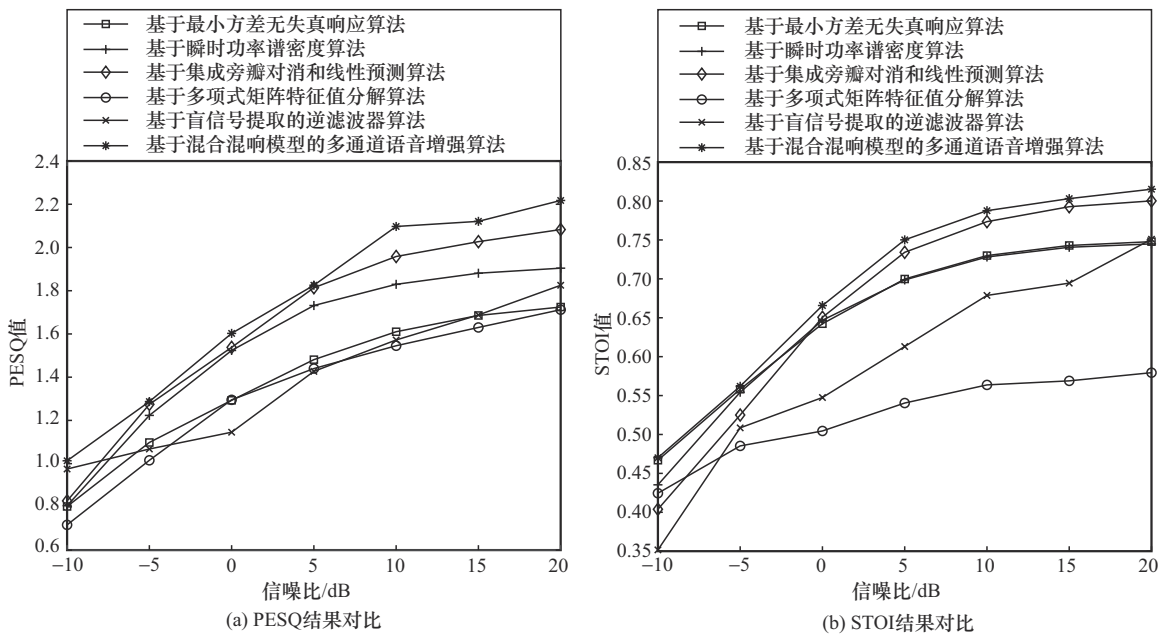


图9 带有干扰性语音信号下的语音增强性能对比

效性和通用性。特别是针对带有干扰性语音的场景,仍然可以抑制干扰性语音,增强目标语音,实现较好的去混响去噪声效果,验证了所提算法的鲁棒性。

### 3.4 算法复杂度分析

本文所提算法的计算复杂度主要集中于卡尔曼滤波器中的模型参数更新学习以及矩阵的多项式特征值分解。其中,卡尔曼滤波器的计算成本与状态向量维度成二次方关系;矩阵的多项式特征值分解的计算复杂度主要集中于式(18)中的矩阵乘积。随着迭代次数的增加,所提算法的运行时间也会增大。在实际应用中,针对不同场景需要设计状态向量维度,考虑算法的收敛性,调节算法迭代次数等,进一步减少算法的计算复杂度。

### 3.5 算法的局限性

本文所提算法主要是基于多通道线性预测模型和空间相干模型框架进行建模的,其性能受到该框架的影响,主要是解决多通道语音增强问题,不适用于单通道情形,其计算复杂度相对于单一的线性预测模型和空间相干模型较高,实验成本较大。另外,在噪声和混响较大的场景下,例如目标语音信号淹没于噪声中,所提算法的性能急剧下降。针对实际复杂的应用场景,可以对信号进行预处理、数据清洗等,提高信号的清晰度和可懂性,接下来的工作还需要进一步拓展本文所提算法,探索复杂场景下的单通道的语音增强问题。

## 4 结束语

针对带混响和噪声环境下的语音增强问题,以多通道线性预测模型和空间相干模型为建模基础,提出一种基于混合混响模型的多通道语音增强算法。相比于传统模型下的其他算法,所提算法在高低混响环境下都能够获得更高的 PESQ 值和 STOI 值,相对于当前对比算法中最好的结果,在 SNR=-10 dB 条件下 PESQ 值提高近 30%,在 SNR=-5 dB 条件下 STOI 值提高近 20%。该算法可为电信、助听器、自动语音识别和语音控制系统等领域的实际应用提供一种技术支持。

另外,人工智能(AI, artificial intelligence)技术已经在语音增强领域取得较好的应用。在未来的工作中,需要结合本文提出的混合模型和 AI 技术,进一步提出新的语音增强算法。

## 参考文献:

- [1] HOANG P, HAAN J M D, TAN Z H, et al. Multichannel speech enhancement with own voice-based interfering speech suppression for hearing assistive devices[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2022, 30: 706-720.
- [2] OZTURK M Z, WU C S, WANG B B, et al. RadioSES: mmWave-based audioradio speech enhancement and separation system[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023, 31: 1333-1347.
- [3] 张琳, 王海涛, 杨爽, 等. 面向舱室声学环境的深度时域语音增强网络[J]. *声学学报*, 2023, 48(4): 890-900.  
ZHANG L, WANG H T, YANG S, et al. Single-channel deep time-domain speech enhancement networks for cabin environments[J]. *Acta Acustica*, 2023, 48(4): 890-900.
- [4] EVERS C, NAYLOR P A. Acoustic SLAM[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018, 26(9): 1484-1498.
- [5] CHEN J D, BENESTY J, HUANG Y T, et al. New insights into the noise reduction Wiener filter[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, 14(4): 1218-1234.
- [6] CHEN Z, WANG R, YIN F L, et al. Speech dereverberation method based on spectral subtraction and spectral line enhancement[J]. *Applied Acoustics*, 2016, 112: 201-210.
- [7] SAYOUD A, DJENDI M, MEDAHI S, et al. A dual fast NLMS adaptive filtering algorithm for blind speech quality enhancement[J]. *Applied Acoustics*, 2018, 135: 101-110.
- [8] SURENDRAN S, KUMAR T K. Oblique projection and cepstral subtraction in signal subspace speech enhancement for colored noise reduction[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018, 26(12): 2328-2340.
- [9] LUO Y, MESGARANI N. Conv-TasNet: surpassing ideal time-frequency magnitude masking for speech separation[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019, 27(8): 1256-1266.
- [10] 范君怡, 杨吉斌, 张雄伟, 等. U-net网络中融合多头注意力机制的单通道语音增强[J]. *声学学报*, 2022, 47(6): 703-716.  
FAN J Y, YANG J B, ZHANG X W, et al. Monaural speech enhancement using U-net fused with multi-head self-attention[J]. *Acta Acustica*, 2022, 47(6): 703-716.
- [11] YOSHIOKA T, NAKATANI T, MIYOSHI M. Integrated speech enhancement method using noise suppression and dereverberation[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2009, 17(2): 231-246.
- [12] DELCROIX M, YOSHIOKA T, OGAWA A, et al. Linear prediction-based dereverberation with advanced speech enhancement and recognition technologies for the reverb challenge[C]// *REVERB Challenge Workshop 2014*. [S.l.:s.n.], 2014: 1-8.
- [13] CHETUPALLI S R, SREENIVAS T V. Late reverberation cancellation using Bayesian estimation of multi-channel linear predictors and student's t-source prior[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019, 27(6): 1007-1018.
- [14] YOSHIOKA T, NAKATANI T. Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(10): 2707-2720.
- [15] TALMON R, COHEN I, GANNOT S. Multichannel speech enhancement using convolutive transfer function approximation in reverberant environments[C]// *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. Piscataway: IEEE Press, 2009: 3885-3888.
- [16] SCHWARTZ O, GANNOT S, HABETS E A P. Multi-microphone speech dereverberation and noise reduction using relative early transfer

- functions[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(2): 240-251.
- [17] DIETZEN T, MOONEN M, WATERSCHOOT T V. Instantaneous PSD estimation for speech enhancement based on generalized principal components[C]//Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO). Piscataway: IEEE Press, 2021: 191-195.
- [18] DIETZEN T, DOCLO S, MOONEN M, et al. Integrated sidelobe cancellation and linear prediction Kalman filter for joint multi-microphone speech dereverberation, interfering speech cancellation, and noise reduction[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 740-754.
- [19] BRAUN S, KUKLASINSKI A, SCHWARTZ O, et al. Evaluation and comparison of late reverberation power spectral density estimators[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 26(6): 1056-1071.
- [20] KODRASI I, DOCLO S. Analysis of eigenvalue decomposition-based late reverberation power spectral density estimation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 26(6): 1106-1118.
- [21] NEO V W, EVERS C, NAYLOR P A. Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 3255-3266.
- [22] 解元, 邹涛, 孙为军, 等. 面向高混响环境的欠定卷积盲源分离算法[J]. 通信学报, 2023, 44(2): 82-93.  
XIE Y, ZOU T, SUN W J, et al. Algorithm of underdetermined convolutive blind source separation for high reverberation environment[J]. Journal on Communications, 2023, 44(2): 82-93.
- [23] 解元, 邹涛, 孙为军, 等. 面向卷积混叠环境下的盲源分离新方法[J]. 自动化学报, 2023, 49(5): 1062-1072.  
XIE Y, ZOU T, SUN W J, et al. Novel blind source separation method for convolutive mixed environment[J]. Acta Automatica Sinica, 2023, 49(5): 1062-1072.
- [24] SEKIGUCHI K, BANDO Y, NUGRAHA A A, et al. Autoregressive moving average jointly-diagonalizable spatial covariance analysis for joint source separation and dereverberation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 2368-2382.
- [25] LIU T Z, LU Z H, COSTA J P J D, et al. A hybrid reverberation model and its application to joint speech dereverberation and separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31: 3000-3014.
- [26] UEDA T, NAKATANI T, IKESHITA R, et al. Blind and spatially-regularized online joint optimization of source separation, dereverberation, and noise reduction[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2024, 32: 1157-1172.
- [27] 解元, 邹涛, 余锦视, 等. 面向噪声和声学混响场景下的语音增强[J]. 信号处理, 2024, 40(12): 2238-2248.  
XIE Y, ZOU T, YU J S, et al. Speech enhancement for noise and acoustic reverberation scenarios[J]. Journal of Signal Processing, 2024, 40(12): 2238-2248.
- [28] HADAD E, HEESE F, VARY P, et al. Multichannel audio database in various acoustic environments[C]// Proceedings of the 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC). Piscataway: IEEE Press, 2014: 313-317.
- [29] VICTOR Z, SENEFF S, GLASS J. TIMIT acoustic-phonetic continuous speech corpus[R]. 1993.
- [30] EATON J, GAUBITCH N D, MOORE A H, et al. Estimation of room acoustic parameters: the ACE challenge[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(10): 1681-1693.
- [31] 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.
- [32] 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]//Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Piscataway: IEEE Press, 2001: 749-752.
- [33] TAAL C H, HENDRIKS R C, HEUSDENS R, et al. An algorithm for intelligibility prediction of time - frequency weighted noisy speech[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2011, 19(7): 2125-2136.

## [作者简介]



解元 (1989-), 男, 安徽利辛人, 博士, 广州大学讲师, 主要研究方向为盲信号分离、信号处理和机器学习等。



邹涛 (1975-), 男, 辽宁沈阳人, 博士, 广州大学教授, 主要研究方向为工业过程建模与仿真、模型预测控制、先进过程控制和实时优化技术研究与应用。



孙为军 (1975-), 男, 安徽马鞍山人, 博士, 广东工业大学副教授, 主要研究方向为模式识别、机器学习等。



谢胜利 (1956-), 男, 湖北荆州人, 博士, 广东工业大学教授, 主要研究方向为无线网络、自动控制、盲信号处理等。