第 37 卷 第 4 期
2018 年 7 月

应用声学
Journal of Applied Acoustics

Vol. 37, No.4
July, 2018

◇ 研究报告 ◇

基于卡尔曼滤波的低复杂度去混响算法*

齐园蕾[1,2] 杨飞然[1] 杨 军[1,2†]

(1 中国科学院噪声与振动重点实验室 (声学研究所) 北京 100190)

(2 中国科学院大学 北京 100049)

摘要 在电话会议、智能音箱等应用场景下，传声器往往处在声源的远场。混响信号的存在会掩蔽后续到达的直达声信号，降低传声器接收信号的语声质量以及语声识别系统的准确识别率。多通道线性预测算法是一种经典的盲去混响算法，但该算法往往具有较高的计算复杂度。该文提出了一种简化的卡尔曼滤波更新算法，通过对角化卡尔曼滤波器状态向量误差协方差矩阵，降低了自适应多通道线性预测去混响算法的复杂度。通过与现有分块对角简化算法对比发现，该文提出的简化算法在保证语声质量的同时，进一步降低了原卡尔曼滤波算法的复杂度。

关键词 卡尔曼滤波，低复杂度，自适应多通道线性预测，盲去混响

中图法分类号: TN912.3 **文献标识码:** A **文章编号:** 1000-310X(2018)04-0559-08

DOI: 10.11684/j.issn.1000-310X.2018.04.015

Kalman filter based low-complexity dereverberation algorithm

QI Yuanlei[1,2] YANG Feiran[1] YANG Jun[1,2]

(1 *Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Beijing* 100190, *China*)

(2 *University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract Microphones are always far away from the speech source in the video-conference systems and intelligent loudspeakers applications. Reverberation signal will smear successive direct signal, which severely degrades the audible speech quality of the captured signals and the performance of automatic speech recognition (ASR) system. The multi-channel linear prediction (MCLP) algorithm is one of the classical blind dereverberation methods, but it suffers from high computational cost. We propose a simplified Kalman filter algorithm, which reduces the complexity of adaptive MCLP dereverberation method by diagonalizing the state error correlation matrix. Compared with the original Kalman filter, the complexity of the proposed algorithm is reduced considerably without significant performance degration.

Key words Kalman filter, Low complexity, Multi-channel linear prediction, Blind dereverberation

2017-12-11 收稿; 2018-03-24 定稿

*国家自然科学基金项目 (61501449), 中国科学院声学研究所青年英才计划项目 (QNYC201722), 2016 年湖北省省院合作专项

作者简介: 齐园蕾 (1991-), 女, 辽宁阜新人, 博士研究生, 研究方向: 信号与信息处理。

†通讯作者 E-mail: jyang@mail.ioa.ac.cn

1 引言

在室内进行的电话会议、智能音箱等应用场景下，传声器往往处在声源的远场。由于房间边界及房间内物体对声波的反射作用，传声器除接收到声源发出的直达声外，还有来自各个方向的反射声。一般将到达时间在直达声之后30~50 ms的声信号称为早期反射声，在此之后到达的声信号称为晚期反射声[1]，即混响拖尾。心理声学研究发现，早期反射声可增强直达声的强度，提高语声可懂度。而混响信号会掩蔽后续到达的直达声信号，导致语声模糊。另外，混响信号还会降低传声器接收信号的语声质量以及语声识别系统的准确识别率。随着声源与传声器之间距离的增加，混响对传声器接收信号的破坏作用更加严重。因此，对传声器接收信号去混响是一项十分必要的工作。

盲去混响算法是指在去混响的过程中，对声源和传声器之间的房间冲激响应(Room impulse response, RIR)的先验知识是未知的。基于传声器阵列的多通道线性预测算法是一种经典的盲去混响算法。根据多输入输出求逆理论(Multiple input/output inverse theorem, MINT)，在各通道传递函数不含公共零点的条件下，多通道方法可以完美均衡时不变的房间冲激响应[2]。然而，MINT算法对系统辨识误差十分敏感，而且实际房间的冲激响应往往含有相近的零点[3]，因此MINT算法在实际中难以应用。由于时域线性预测算法往往要求很长的滤波器长度，并且存在白化目标信号的问题。

最近有学者提出在短时傅里叶变换(Short-time Fourier transform, STFT)域应用多通道线性预测算法在各子带独立处理信号。根据滤波器组理论，短时傅里叶变换具有完美重建性质，即通过逆STFT可以准确地重建输入信号[4]。通过选取合适的窗函数，短时傅里叶变换将时域的一帧信号变换为每个频率子带(频率柜)的一个点，由此可以减少每个子带的滤波器长度[5−6]。另外，可借助快速傅里叶变换(Fast Fourier transformation, FFT)提高计算效率。由于房间冲激响应实际上是随时间变化的，所以需要时变的预测模型系数建模。比较典型的算法有加权预测误差[6] (Weighted prediction error, WPE) 算法和加权递归最小二乘[7] (Weighted recursive least squares, WRLS)算法，二者都在STFT域实施。WPE算法基于多通道线性预测模型，在每个频带利用自回归模型描述混响信号，利用混响预测权估计晚期混响。通过在每个频带应用最大似然准则，迭代估计混响预测权系数和语声谱方差。随着迭代次数的增加，计算复杂度也不断增加，因而限制了该算法的实际应用。文献[8]提出了STFT域的时变多通道自回归(Multichannel autoregressive, MAR)信号模型，利用卡尔曼滤波器估计MAR系数，该算法可视为一种广义的RLS算法。

基于STFT域的多通道线性预测算法的计算复杂度与每个子带滤波器阶数成平方关系[8−9]。该复杂度限制了算法在很多资源有限的系统平台上的应用。文献[9]针对STFT域的自适应多通道线性预测去混响算法，提出了一种简化的卡尔曼滤波求解方法，将计算复杂度降到与滤波器阶数成线性关系。然而，该简化方法会导致一定程度的语声质量下降。另外，该算法只估计一个通道的信号，实际中需要计算多个通道。本文将提出另一种卡尔曼滤波器的简化方法，在保证不损失语声质量的同时，进一步降低STFT域自适应多通道线性预测去混响算法的复杂度。

2 基于卡尔曼滤波的多通道线性预测算法

在多通道线性预测理论中，认为混响信号是传声器延迟信号经线性滤波的输出。因此，去混响问题的关键是通过某一自适应滤波器盲估计滤波器系数。

2.1 信号模型

令向量$\boldsymbol{y}(k,n)=\left[Y_1(k,n),\cdots,Y_m(k,n),\cdots,Y_M(k,n)\right]^{\mathrm{T}}$表示STFT域的传声器接收信号，其中$Y_m(k,n)$表示第$m$个传声器的接收信号，$k$为频率下标，$n$为时间下标，$M$为传声器数目，上标T表示转置操作。假设在每个频率子带中混响信号由MAR过程[5,10]产生，即

$$\boldsymbol{y}(k,n)=\underbrace{\sum_{p=D}^{L}\boldsymbol{C}_p(k,n)\boldsymbol{y}(k,n-p)}_{\boldsymbol{r}(k,n)}+\boldsymbol{x}(k,n),\quad(1)$$

其中，$M\times M$的矩阵$\boldsymbol{C}_p(k,n)$为时变的MAR系数，$p=[D,D+1,\cdots,L]$。L为线性预测长度，

延迟 $D > 1$ 的选择与STFT的帧重叠参数有关，取值要保证 $\boldsymbol{x}(k,n)$ 与 $\boldsymbol{r}(k,n)$ 的相关可以忽略。$\boldsymbol{x}(k,n) = [X_1(k,n),\cdots,X_M(k,n)]^{\mathrm{T}}$ 为目标信号，代表直达声和早期反射声，$\boldsymbol{r}(k,n)$ 代表晚期混响。如果能够估计出MAR系数，则可以通过FIR滤波器 $\boldsymbol{C}_p(k,n)$ 恢复目标信号 $\boldsymbol{x}(k,n)$。由于所有变量都是频率相关的，为了简化表示，下文中将忽略频率下标 k。定义如下变量：

$$\boldsymbol{Y}(n) = \boldsymbol{I}_M \otimes \left[\boldsymbol{y}^{\mathrm{T}}(n),\cdots,\boldsymbol{y}^{\mathrm{T}}(n-L+D)\right], \quad (2)$$

$$\boldsymbol{c}(n) = \mathrm{Vec}\left\{[\boldsymbol{C}_D(n),\cdots,\boldsymbol{C}_L(n)]^{\mathrm{T}}\right\}, \quad (3)$$

其中，$\boldsymbol{I}_M$ 是 $M\times M$ 的单位阵，$\otimes$ 代表Kronecker乘积[11]，$\mathrm{Vec}\{\cdot\}$ 为矩阵列堆叠操作因子，$\boldsymbol{c}(n)$ 的长度为 $L_c = M^2(L-D+1)$，$\boldsymbol{Y}(n)$ 是由传声器观测信号构成的尺寸为 $M\times L_c$ 的稀疏矩阵。利用式(2)和式(3)重写式(1)，可得到混响信号的向量化表示：

$$\boldsymbol{y}(n) = \boldsymbol{Y}(n-D)\,\boldsymbol{c}(n) + \boldsymbol{x}(n), \quad (4)$$

式(4)称为卡尔曼滤波器模型中的观测方程。假设向量 $\boldsymbol{x}(n)$ 服从零均值的复正态分布：

$$\boldsymbol{x}(n) \sim \mathcal{N}(\boldsymbol{0}_{M\times 1}, \boldsymbol{\Phi}_x(n)), \quad (5)$$

其中，$\boldsymbol{\Phi}_x(n) = E\left\{\boldsymbol{x}(n)\,\boldsymbol{x}^{\mathrm{H}}(n)\right\}$ 是 $\boldsymbol{x}(n)$ 的协方差矩阵。根据文献[10]，采用最大对数似然函数准则迭代估计 $\boldsymbol{\Phi}_x(n)$：$\hat{\boldsymbol{\Phi}}_x^{(\mathrm{RML})}(n) = \alpha\hat{\boldsymbol{\Phi}}_x(n-1) + (1-\alpha)\,\boldsymbol{e}(n)\boldsymbol{e}^{\mathrm{H}}(n)$，其中 α 为指数迭代平滑因子，$\hat{\boldsymbol{\Phi}}_x(n) = \alpha\hat{\boldsymbol{\Phi}}_x(n-1) + (1-\alpha)\,\hat{\boldsymbol{x}}(n)\,\hat{\boldsymbol{x}}^{\mathrm{H}}(n)$。另外，假设目标信号在不同的时间帧之间是不相关的，这一假设已被普遍接受并被广泛应用[12−14]于非混响语声信号的STFT系数的建模。

通过估计得到的MAR系数 $\hat{\boldsymbol{c}}(n)$ 对传声器信号线性滤波，得到目标信号的估计值：

$$\hat{\boldsymbol{x}}(n) = \boldsymbol{y}(n) - \boldsymbol{Y}(n-D)\,\hat{\boldsymbol{c}}(n). \quad (6)$$

为建模MAR系数的不确定性，假设向量 $\boldsymbol{c}(n)$ 由独立的复随机变量构成。通过一阶马尔科夫模型描述MAR系数状态向量：

$$\boldsymbol{c}(n) = \boldsymbol{A}(n)\boldsymbol{c}(n-1) + \boldsymbol{w}(n), \quad (7)$$

式(7)称为卡尔曼滤波器模型中的状态方程，矩阵 $\boldsymbol{A}(n)$ 描述状态向量随时间的传播过程。该模型最早由GeraldEnzner等[15]提出，并得到进一步推广应用[16]。由于MAR系数随时间的传播过程是未知的，一般选择状态转移矩阵 $\boldsymbol{A}(n) = \boldsymbol{I}_{L_c}$。也就是说，我们利用 $\boldsymbol{w}(n)$ 来描述实际中系数状态向量 $\boldsymbol{c}(n)$ 的时变特性。$\boldsymbol{w}(n)$ 是零均值复高斯扰动过程，即 $\boldsymbol{w}(n)\sim\mathcal{N}(\boldsymbol{0}_{L_c\times 1}, \boldsymbol{\Phi}_w(n))$。假设 $\boldsymbol{x}(n)$ 和 $\boldsymbol{w}(n)$ 是不相关的，且 $\boldsymbol{w}(n)$ 的协方差矩阵为 $\boldsymbol{\Phi}_w(n) = \varphi_w(n)\boldsymbol{I}_{L_c}$。由连续两帧MAR系数向量的变化量估计方差 $\varphi_w(n)$，即

$$\hat{\varphi}_w(n) = \frac{1}{L_c}E\left\{\|\hat{\boldsymbol{c}}(n) - \hat{\boldsymbol{c}}(n-1)\|_2^2\right\} + \eta, \quad (8)$$

其中，η 是一个小正常数。

2.2 卡尔曼滤波求解

定义如下状态向量误差协方差矩阵：

$$\boldsymbol{\Phi}_\varepsilon(n) = E\left\{[\boldsymbol{c}(n) - \hat{\boldsymbol{c}}(n)]\,[\boldsymbol{c}(n) - \hat{\boldsymbol{c}}(n)]^{\mathrm{H}}\right\}. \quad (9)$$

由式(7)的状态方程和式(4)的观测方程，以及当前时刻获得的传声器观测数据，利用卡尔曼滤波器最小化均方误差 $E\left\{\|\boldsymbol{c}(n) - \hat{\boldsymbol{c}}(n)\|_2^2\right\}$，可迭代估计出状态向量 $\boldsymbol{c}(n)$。卡尔曼滤波器更新方程[8]如下：

$$\boldsymbol{\Phi}_\varepsilon(n\,|n-1) = \hat{\boldsymbol{\Phi}}_\varepsilon(n-1) + \boldsymbol{\Phi}_w(n), \quad (10)$$

$$\boldsymbol{e}(n) = \boldsymbol{y}(n) - \boldsymbol{Y}(n-D)\,\hat{\boldsymbol{c}}(n-1), \quad (11)$$

$$\boldsymbol{R}_e(n) = \boldsymbol{Y}(n-D)\,\boldsymbol{\Phi}_\varepsilon(n\,|n-1)\,\boldsymbol{Y}^{\mathrm{H}}(n-D) + \boldsymbol{\Phi}_x(n), \quad (12)$$

$$\boldsymbol{K}(n) = \boldsymbol{\Phi}_\varepsilon(n\,|n-1)\,\boldsymbol{Y}^{\mathrm{H}}(n-D)\,\boldsymbol{R}_e^{-1}(n), \quad (13)$$

$$\hat{\boldsymbol{\Phi}}_\varepsilon(n) = [\boldsymbol{I}_{L_c} - \boldsymbol{K}(n)\,\boldsymbol{Y}(n-D)]\,\boldsymbol{\Phi}_\varepsilon(n\,|n-1), \quad (14)$$

$$\hat{\boldsymbol{c}}(n) = \hat{\boldsymbol{c}}(n-1) + \boldsymbol{K}(n)\,\boldsymbol{e}(n), \quad (15)$$

其中，$\boldsymbol{e}(n)$ 为预测误差向量，$\boldsymbol{K}(n)$ 为卡尔曼增益矩阵。通过观察式(6)和式(11)可知，预测误差 $\boldsymbol{e}(n)$ 是在给定MAR系数 $\hat{\boldsymbol{c}}(n-1)$ 的条件下对目标信号 $\boldsymbol{x}(n)$ 的估计。

3 低复杂度的卡尔曼滤波器

在卡尔曼滤波器的更新过程中，状态向量误差协方差矩阵 $\boldsymbol{\Phi}_\varepsilon(n)$ 和目标信号 $\boldsymbol{x}(n)$ 的协方差矩阵 $\boldsymbol{\Phi}_x(n)$ 对算法的性能表现和计算复杂度具有非常重要的作用。通过采取某种方法合理简化二者的结构，可在算法的性能表现和计算复杂度之间得到一个折中。

3.1 完全对角简化算法

文献[9]提出了一种分块对角结构简化状态向量误差协方差矩阵$\boldsymbol{\Phi}_{\varepsilon}(n)$的结构。该算法假设$\boldsymbol{\Phi}_{\varepsilon}(n)$为分块对角阵，认为$\boldsymbol{\Phi}_{\varepsilon}(n)$的非主对角线以外的子矩阵为零。这一简化使得矩阵$\boldsymbol{\Phi}_{\varepsilon}(n)$保持分块对角结构，同时使得卡尔曼滤波器向量$\boldsymbol{c}(n)$中的每一部分共享同一个误差信号$\boldsymbol{e}(n)$。这意味着分块对角简化算法中的目标信号和误差信号都简化为1×1的标量，也就是最终只恢复第一个通道的目标信号。由此将导致估计的协方差矩阵$\boldsymbol{\Phi}_{x}(n)$损失部分相关信息，对于对角化的协方差矩阵，损失了待估计目标信号的方差(能量)，进而影响最终输出信号的语声质量。文献[9]提出的算法虽然利用多个通道的空间信息，但最终只输出其中一个通道的去混响信号。为进一步去除残余混响，往往在某一去混响算法后级联其他多通道算法，如文献[17]提出的WPE算法后级联最小方差无失真响应(Minimum variance distortionless response, MVDR)波束形成器。由此可见，文献[9]中的分块对角简化方法限制了后续多通道算法的应用。另外，尽管文献[9]中的简化方法大大降低了文献[8]所提算法的计算复杂度，但输出的去混响信号的语声质量有所下降。

本文提出一种完全对角简化算法进一步降低卡尔曼更新过程的计算复杂度。根据式(14)可知，$\boldsymbol{I}_{L_c}-\boldsymbol{K}(n)\boldsymbol{Y}\left(n-D\right)$对$\hat{\boldsymbol{\Phi}}_{\varepsilon}(n)$的结构有直接影响[18]。我们采取如下近似：

$$\begin{aligned}&\boldsymbol{I}_{L_c}-\boldsymbol{K}(n)\boldsymbol{Y}\left(n-D\right)\\ \approx&\left[1-\frac{\mathrm{tr}\left[\boldsymbol{K}^{\mathrm{T}}(n)\boldsymbol{Y}^{\mathrm{T}}\left(n-D\right)\right]}{L_c}\right]\boldsymbol{I}_{L_c},\end{aligned}\tag{16}$$

由过程噪声的协方差矩阵$\boldsymbol{\Phi}_{w}\left(n\right)$为对角阵，可知上述近似是合理的。当滤波器开始收敛时，由于滤波器系数之间变得不相关，式(10)中的矩阵$\boldsymbol{\Phi}_{\varepsilon}\left(n\left|n-1\right.\right)$和$\hat{\boldsymbol{\Phi}}_{\varepsilon}\left(n-1\right)$将成为对角阵，即

$$\boldsymbol{\Phi}_{\varepsilon}\left(n\left|n-1\right.\right)\approx\sigma_{\varepsilon}^{2}(n)\boldsymbol{I}_{L_c},\tag{17}$$

$$\hat{\boldsymbol{\Phi}}_{\varepsilon}\left(n-1\right)\approx\sigma_{\mu}^{2}(n)\boldsymbol{I}_{L_c}.\tag{18}$$

为使协方差矩阵$\boldsymbol{\Phi}_{x}(n)$也对角化，对目标信号方差的估计作如下近似：

$$\hat{\varphi}_{x}^{(\mathrm{RML})}(n)\approx\alpha\hat{\varphi}_{x}\left(n-1\right)+\left(1-\alpha\right)\left\|\hat{\boldsymbol{e}}(n)\right\|_{2}^{2},\tag{19}$$

$$\hat{\varphi}_{x}(n)\approx\alpha\hat{\varphi}_{x}\left(n-1\right)+\left(1-\alpha\right)\left\|\hat{\boldsymbol{x}}(n)\right\|_{2}^{2},\tag{20}$$

定义

$$\boldsymbol{S}_{Y}(n-D)=\boldsymbol{Y}(n-D)\boldsymbol{Y}^{\mathrm{H}}(n-D),\tag{21}$$

和

$$\delta(n)=\frac{\hat{\varphi}_{x}^{(\mathrm{RML})}(n)}{\sigma_{\varepsilon}^{2}(n)},\tag{22}$$

则式(12)简化为

$$\begin{aligned}&\boldsymbol{R}_{e}(n)\\ =&\boldsymbol{Y}\left(n-D\right)\boldsymbol{\Phi}_{\varepsilon}\left(n\left|n-1\right.\right)\boldsymbol{Y}^{\mathrm{H}}\left(n-D\right)+\boldsymbol{\Phi}_{x}(n)\\ =&\boldsymbol{Y}\left(n-D\right)\boldsymbol{Y}^{\mathrm{H}}\left(n-D\right)\sigma_{\varepsilon}^{2}(n)+\hat{\varphi}_{x}^{(\mathrm{RML})}(n)\boldsymbol{I}_{M}\\ =&\boldsymbol{S}_{Y}\left(n-D\right)\sigma_{\varepsilon}^{2}(n)+\hat{\varphi}_{x}^{(\mathrm{RML})}(n)\boldsymbol{I}_{M}\\ =&\sigma_{\varepsilon}^{2}(n)\left[\boldsymbol{S}_{Y}\left(n-D\right)+\frac{\hat{\varphi}_{x}^{(\mathrm{RML})}(n)}{\sigma_{\varepsilon}^{2}(n)}\boldsymbol{I}_{M}\right]\\ =&\sigma_{\varepsilon}^{2}(n)\left[\boldsymbol{S}_{Y}\left(n-D\right)+\delta(n)\boldsymbol{I}_{M}\right].\end{aligned}\tag{23}$$

卡尔曼增益矩阵简化为

$$\begin{aligned}&\boldsymbol{K}(n)\\ =&\boldsymbol{\Phi}_{\varepsilon}(n|n-1)\boldsymbol{Y}^{\mathrm{H}}(n-D)\boldsymbol{R}_{e}^{-1}(n)\\ =&\frac{\sigma_{\varepsilon}^{2}(n)\boldsymbol{Y}^{\mathrm{H}}(n-D)}{\sigma_{\varepsilon}^{2}(n)[\boldsymbol{S}_{Y}(n-D)+\delta(n)\boldsymbol{I}_{M}]}\\ =&\boldsymbol{Y}^{\mathrm{H}}(n-D)\left[\boldsymbol{S}_{Y}(n-D)+\delta(n)\boldsymbol{I}_{M}\right]^{-1},\end{aligned}\tag{24}$$

则可进一步推导出简化的卡尔曼滤波器更新方程：

$$\sigma_{\varepsilon}^{2}(n)=\sigma_{\mu}^{2}\left(n-1\right)+\hat{\varphi}_{w}(n),\tag{25}$$

$$\boldsymbol{e}(n)=\boldsymbol{y}(n)-\boldsymbol{Y}\left(n-D\right)\hat{\boldsymbol{c}}\left(n-1\right),\tag{26}$$

$$\hat{\boldsymbol{c}}(n)=\hat{\boldsymbol{c}}\left(n-1\right)+\frac{\boldsymbol{Y}^{\mathrm{H}}(n-D)}{\boldsymbol{S}_{Y}(n-D)+\delta(n)\boldsymbol{I}_{M}}\boldsymbol{e}(n),\tag{27}$$

$$\begin{aligned}\sigma_{\mu}^{2}(n)=&\left[1-\frac{\mathrm{tr}\left[\left[\boldsymbol{S}_{Y}(n-D)+\delta(n)\boldsymbol{I}_{M}\right]^{-1}\boldsymbol{S}_{Y}(n-D)\right]}{L_c}\right]\\ &\times\sigma_{\varepsilon}^{2}(n).\end{aligned}\tag{28}$$

值得注意的是，本文提出的简化的卡尔曼滤波算法的误差信号$\boldsymbol{e}\left(n\right)$和目标信号$\boldsymbol{x}(n)$均为$M\times 1$的向量，这为后续级联其他多通道算法提供了方便。另外，也为计算方差$\hat{\varphi}_{x}\left(n\right)$提供了更多的可用信息。相比文献[9]提出的分块对角简化方法，本文提出的方法更具优势。

文献[15]提出了频域Kalman滤波算法，可以看成一种变步长频域自适应算法[19]。类似地，我

们发现上述简化的卡尔曼滤波算法可看作是一种变规整化因子的归一化最小均方 (Normalized least mean square, NLMS) 算法。根据式 (27) 可知，其中 $\delta(n)$ 可视为一个可变的规整化因子。根据式 (22) 和式 (25) 可知，方差 $\varphi_w(n)$ 对滤波器系数 $\boldsymbol{c}(n)$ 的估计具有重要作用。当算法还未收敛时，$\hat{\boldsymbol{c}}(n)$ 和 $\hat{\boldsymbol{c}}(n-1)$ 的差值较大，根据式 (8) $\varphi_w(n)$ 此时也取较大的值，因此提供了快速的收敛性能和跟踪性能。当算法开始收敛到稳态时，$\hat{\boldsymbol{c}}(n)$ 和 $\hat{\boldsymbol{c}}(n-1)$ 的差值减小，导致了较小的 $\varphi_w(n)$，也就是较低的失调。较小的 $\varphi_w(n)$ 值表征了良好的失调性能及差的跟踪性能，较大的 $\varphi_w(n)$ 值表征了良好的跟踪性能及差的失调性能。换句话说，$\varphi_w(n)$ 的取值高度决定了卡尔曼滤波器的跟踪性能和收敛性能，因此上述简化的卡尔曼滤波算法可看作是一种变规整化因子的 NLMS 算法。

3.2　计算复杂度比较

本文中只考虑乘除法的运算次数，忽略加减法运算产生的复杂度。表 1 中列出了原卡尔曼滤波算法、WPE 算法、分块对角简化方法以及本文提出的完全对角简化方法对应的计算复杂度。同时给出了当线性预测长度 $L=15$、延迟 $D=3$、$P=L-D+1=15-3+1=13$ 时，M 取某些特定值时的计算复杂度。

表 1　计算复杂度比较

Table 1　Compare of computational cost

通道数	算法			
	简化前	WPE	分块对角	完全对角
M	$P^3M^6+3P^2M^4+PM^4$ $+2PM^3+2PM^2+M^3$	$P^3M^3+2P^2M^2+2PM$	PM^3+3PM^2+6PM	$6PM^2+M^3$
2	149248	18980	416	320
4	9134176	146120	1768	1312
6	103183920	486876	4680	3024
8	578075776	1146704	9776	5504

由表 1 可知，计算复杂度与滤波器阶数 P 和通道数目 M 有关。当 P 的取值固定时，随着通道数的增加，上述四种滤波算法的计算复杂度均有所增加。分块对角法和完全对角法均大大降低了原卡尔曼滤波的计算复杂度，且都低于 WPE 算法的计算复杂度，但分块对角简化方法的复杂度始终高于完全对角法。

4　仿真实验

由于汉语的时程特点与很多连读的拼音外语不同，混响对二者的干扰机理也不同。仿真实验中将测试三组输入信号，一段英文语声和两段中文语声。英文语声取自 TIMIT 数据库 [20] 中的前 20 个语声文件，长度约为 50.78 s。中文语声取自 GSBM 6001-89 国家标准样件中专门编写的“美谈不美”两段汉语普通话 (方明、雅坤朗诵)，长度分别为 44 s 和 38 s。选取的两段汉语普通话记录样件中不仅包含了汉语普通话的所有声母与韵母以及五个声调，而且它们的出现概率与相关国标对汉语因素统计的误差不超过 ±5%，已经被许多研究汉语的项目取用。

房间冲激响应由虚源法 [21] (Image method) 产生。线性预测长度 $L=15$，延迟 $D=3$，传声器阵列为等间隔线列阵 $M=4$，传声器间隔为 0.2 m。采样率为 16 kHz，FFT 点数为 1024，窗函数采用平方根汉宁窗，窗长为 512 点，50% 重叠。式 (8) 中的正常数 $\eta=10^{-5}$，平滑因子 $\alpha=0.2$。

为测试分块对角简化方法和完全对角简化方法在不同冲激响应条件下的性能表现，由虚源法产生 6 个不同参数的 RIR，如表 2 所示，d 表示声源与传声器之间的距离。

表 2　房间冲激响应

Table 2　Room impulse response

T60	d		
	0.5 m	1 m	2 m
0.63 s	$h1$	$h2$	$h3$
1 s	$h4$	$h5$	$h6$

算法的性能评价指标选择PESQ(Perceptual evaluation of speech quality)得分和SRMR(Speech to reverberation modulation energy ratio)得分。因为二者均与混响的主观感知高度相关，并被广泛应用于去混响算法的研究中。作为对照，本文还对WPE算法[22]做了相应的仿真，仿真实验中用到的WPE 算法的MATLAB代码见参考文献[23]。表3、表4分别给出了输入信号为英文语声条件下，原卡尔曼算法、WPE算法、分块对角简化方法、完全对角简化方法在6种RIR条件下两种指标的得分结果。表5、表6、表7、表8分别给出了输入信号为两段汉语普通话条件下，原卡尔曼算法、WPE算法、分块对角简化方法、完全对角简化方法在6种RIR条件下两种指标的得分结果。

从表3、表5、表7中可以看出，无论是英文语声还是中文语声，分块对角简化方法和完全对角简化方法相比于简化前的卡尔曼算法，PESQ得分均有所下降，但完全对角方法更接近简化前的性能表现。另外，分块对角简化方法和完全对角简化方法在大多数房间冲激响应条件下均好于目前通用的WPE算法。从表4、表6、表8中可以看出，WPE算法、分块对角和完全对角两种简化方法SRMR得分均高于简化前的卡尔曼算法，但并不能说明这三种算法的性能好于简化前，原因是WPE 算法和简化后的算法得到的去混响信号过于“干净”，导致语声信号存在一定失真。在混响较强的条件下，如表3、表5、表7中的RIR $h6$，分块对角简化方法略好于完全对角法，可见两种简化方法各自具有优势。为更具体地观察各算法的性能表现，图1给出了在$h5$条件下目标信号、混响信号以及两种简化方法对应的语谱图。

表3 TIMIT-PESQ得分

Table 3 TIMIT-PESQ score

RIR	混响信号	简化前	WPE	分块对角	完全对角
$h1$	3.01	3.95	3.72	3.67	3.83
$h2$	2.76	3.91	3.50	3.58	3.81
$h3$	2.62	3.71	3.09	3.49	3.52
$h4$	2.69	3.82	3.32	3.58	3.64
$h5$	2.46	3.72	3.10	3.44	3.54
$h6$	2.27	3.42	2.85	3.29	3.21

表4 TIMIT-SRMR得分

Table 4 TIMIT-SRMR score

RIR	混响信号	简化前	WPE	分块对角	完全对角
$h1$	3.00	4.15	4.38	4.64	4.35
$h2$	2.55	4.12	4.70	4.68	4.36
$h3$	1.94	3.10	3.89	3.51	3.28
$h4$	2.36	4.16	4.37	4.68	4.30
$h5$	1.88	4.01	4.45	4.63	4.23
$h6$	1.54	2.87	3.13	3.31	3.03

表5 方明-PESQ得分

Table 5 Fang Ming-PESQ score

RIR	混响信号	简化前	WPE	分块对角	完全对角
$h1$	3.06	3.72	3.55	3.51	3.65
$h2$	2.81	3.72	3.32	3.46	3.61
$h3$	2.71	3.61	3.05	3.39	3.38
$h4$	2.64	3.65	3.19	3.42	3.42
$h5$	2.43	3.60	2.99	3.33	3.34
$h6$	2.31	3.44	2.74	3.21	3.09

表6 方明-SRMR得分

Table 6 Fang Ming-SRMR score

RIR	混响信号	简化前	WPE	分块对角	完全对角
$h1$	3.16	5.05	5.61	5.45	5.15
$h2$	2.27	4.32	5.46	4.63	4.28
$h3$	1.81	3.37	4.50	3.58	3.24
$h4$	2.37	4.84	5.34	5.23	4.76
$h5$	1.65	4.04	4.79	4.36	3.88
$h6$	1.44	3.12	3.36	3.30	2.87

表7 雅坤-PESQ得分

Table 7 Ya Kun-PESQ score

RIR	混响信号	简化前	WPE	分块对角	完全对角
$h1$	2.98	3.67	3.29	3.44	3.54
$h2$	2.71	3.63	3.13	3.35	3.47
$h3$	2.54	3.48	2.89	3.27	3.27
$h4$	2.50	3.53	2.96	3.31	3.34
$h5$	2.30	3.43	2.76	3.20	3.25
$h6$	2.13	3.24	2.54	3.07	3.00

表8 雅坤-SRMR得分

Table 8 Ya Kun-SRMR score

RIR	混响信号	简化前	WPE	分块对角	完全对角
$h1$	4.36	6.85	8.12	7.08	6.68
$h2$	3.56	5.81	7.66	5.95	5.72
$h3$	2.36	4.20	5.82	4.35	4.09
$h4$	3.27	6.48	7.66	6.67	6.18
$h5$	2.54	5.47	6.55	5.67	5.32
$h6$	1.86	3.97	4.28	4.09	3.73

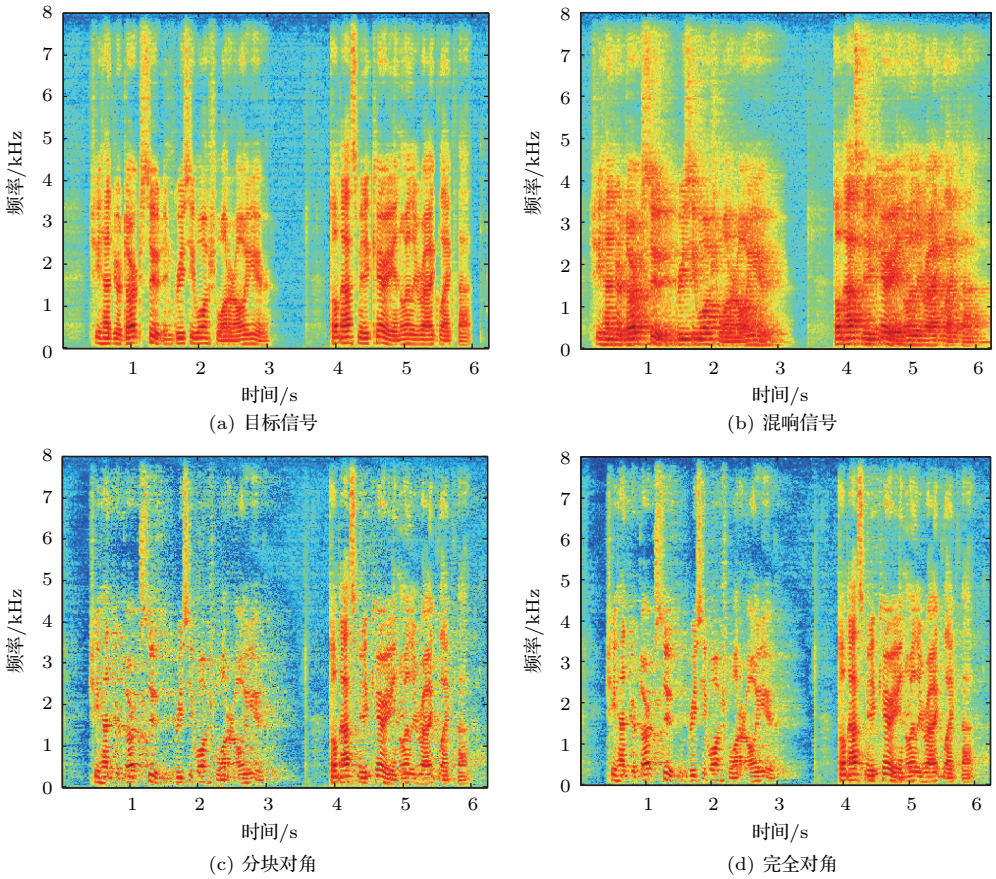

图 1　语谱图

Fig. 1　Spectrograms

从图 1 中可以看出，两种简化方法都在很大程度上去除了混响干扰，但相比于目标信号，两种方法输出的去混响信号都损失了部分语声信号。完全对角简化方法由于保留了更多的目标信号方差信息，性能表现优于分块对角方法。图 1 的仿真测试结果与表 3、表 4 给出的两种指标的得分是一致的。

另外，观察表 5 中的 $h4$ 和表 7 中的 $h3$，分块对角和完全对角两种简化方法的 PESQ 得分相同。为进一步比较两种算法的性能，将对第一段汉语普通话在 $h4$ 条件下的各算法输出语声和第二段汉语普通话在 $h3$ 条件下的各算法输出语声进行主观实验。语声质量的主观评价方法选取平均意见分 (Mean opinion score, MOS) 指标。MOS 的五级评分标准如表 9 所示。实验中由十位测试者对混响信号及四种算法的输出语声信号给出 MOS 得分，十位测试者的平均分作为该算法的最终得分，评分结果如表 10 所示。

根据表 10 可知，分块对角简化算法和完全对角简化算法输出的去混响语声 MOS 得分十分接近。这与表 5、表 7 给出的客观评价指标的结果是一致的。相比混响语声，两种简化算法都明显提高了语声质量。相比 WPE 算法，两种简化算法也有明显的优势。在输入信号为男声的条件下，相比简化前的卡尔曼算法，两种简化算法输出的语声质量有所下降。但在输入信号为女声的条件下，两种简化算法

表 9　平均意见分五级标准

Table 9　MOS standard

MOS 得分	语声质量	失真程度
5	优	不易觉察
4	良	刚刚觉察但不讨厌
3	一般	可觉察但稍微讨厌
2	差	讨厌但能忍受
1	极差	非常讨厌且不能忍受

表 10　平均意见分

Table 10 MOS score

输入	RIR	混响信号	简化前	WPE	分块对角	完全对角
方明 (男)	$h4$	2.6	4.4	2.5	3.7	3.6
雅坤 (女)	$h3$	2.2	3.7	2.3	3.7	3.7

输出的语声质量与简化前的卡尔曼算法输出的语声质量十分接近。综上，本文提出的完全对角简化方法具有比分块对角简化算法更低的计算复杂度，同时具有相近的语声质量。

5 结论

本文通过对角化卡尔曼滤波器状态向量误差协方差矩阵，提出了一种简化的卡尔曼滤波更新算法，降低了STFT域自适应多通道线性预测去混响算法的计算复杂度。通过与现有算法对比分析，发现本文提出的算法在保证语声质量的同时，进一步降低了计算复杂度。但相比于简化前的卡尔曼滤波算法，本文提出的简化算法的语声质量比原算法稍有降低，需要进一步研究以解决该问题。

参 考 文 献

[1] Naylor P, Gaubitch N. Speech dereverberation[M]. London: Springer-Verlag, 2010: 6–8.

[2] Miyoshi M, Kaneda Y. Inverse filtering of room acoustics[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1988, 36(2): 145–152.

[3] Bekrani M, Khong A. A robust MINT equalization algorithm based on near-common zero concept[C]. Electrical Engineering (ICEE), IEEE, 2017: 1685–1690.

[4] Smith J. Spectral audio signal processing[M]. USA: W3K Publishing, 2011: 300.

[5] 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.

[6] Yoshioka T, Nakatani T. Generalization of multichannel linear prediction methods for blind MIMO impulse response shortening[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(10): 2707–2720.

[7] Yoshioka T, Tachibana H, Nakatani T, et al. Adaptive dereverberation of speech signals with speaker position change detection[C]. International Conference on Acoustics, Speech and Signal Processing, IEEE, 2009: 3733–3736.

[8] Braun S, Habets E. Online dereverberation for dynamic scenarios using a Kalman filter with an autoregressive model[J]. IEEE Signal Processing Letters, 2016, 23(12): 1741–1745.

[9] Dietzen T, Doclo S, Spriet A, et al. Low-complexity Kalman filter for multi-channel linear-prediction-based blind speech dereverberarion[C]. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE, 2017.

[10] Nakatani T, Juang B H, Yoshioka T, et al. Speech dereverberation based on maximum-likelihood estimation with time-varying Gaussian source model[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2008, 16(8): 1512–1527.

[11] Graham A. Kronecker products and matrix calculus: with applications[M]. New York: John Wiley & Sons, 1982: 130.

[12] Schwartz B, Gannot S, Habets E. Online speech dereverberation using Kalman filter and EM algorithm[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(2): 394–406.

[13] Esch T, Vary P. Speech enhancement using a modified Kalman filter based on complex linear prediction and supergaussian priors[C]. International Conference on Acoustics, Speech and Signal Processing, IEEE, 2008: 4877–4880.

[14] Erkelens J S, Heusdens R. Correlation-based and model-based blind single-channel late-reverberation suppression in noisy time-varying acoustical environments[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(7): 1746–1765.

[15] Enzner G, Vary P. Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones[J]. Signal Processing, 2006, 86(6): 1140–1156.

[16] Enzner G. Bayesian inference model for applications of time-varying acoustic system identification[C]. 18th European Signal Processing Conference, 2010: 2126–2130.

[17] Cohen A, Stemmer G, Ingalsuo S, et al. Combined weighted prediction error and minimum variance distortionless response for dereverberation[C]. International Conference on Acoustics, Speech and Signal Processing, IEEE, 2017: 446–450.

[18] Paleologu C, Benesty J, Ciochină S. Study of the general Kalman filter for echo cancellation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(8): 1539–1549.

[19] Yang F, Enzner G, Yang J. Frequency-domain adaptive Kalman filter with fast recovery of abrupt echo-path changes[J]. IEEE Signal Processing Letters, 2017, 24(12): 1778–1782.

[20] Garofolo J. Getting started with the DARPA TIMIT CD-ROM: Anacoustic-phonetic continuous speech database[S]. National Institute of Standards and Technology(NIST). Gaithersburg, MD, USA, 1993.

[21] Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics[J]. Journal of the Acoustical Society of America, 1979, 65(4): 943–950.

[22] 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.

[23] Nippon Telegraph and Telephone Corporation (NTT). WPE speech dereverberation[EB/OL]. [2017-11-26]. http://www. kecl.ntt.co.jp/icl/signal/wpe/.