

引用格式: 刘杨, 杨飞然, 梁兆杰, 等. 基于卡尔曼滤波的 STFT 域回声抵消算法[J]. 声学技术, 2022, 41(5): 757-762. [LIU Yang, YANG Feiran, LIANG Zhaojie, et al. Kalman filter based acoustic echo cancellation in the STFT domain[J]. Technical Acoustics, 2022, 41(5): 757-762.] DOI: 10.16300/j.cnki.1000-3630.2022.05.019

# 基于卡尔曼滤波的 STFT 域回声抵消算法

刘 杨<sup>1,2</sup>, 杨飞然<sup>1,2,3</sup>, 梁兆杰<sup>4</sup>, 杨 军<sup>1,2,3</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;

3. 中科新声(苏州)科技有限公司, 江苏苏州 215163; 4. 深圳供电局有限公司, 广东深圳 518000)

**摘要:** 提出了一种低复杂度的短时傅里叶变换域卡尔曼滤波算法来解决声学回声抵消问题。首先在短时傅里叶变换域建立了基于频域卷积传递函数的观测方程, 并利用一阶马尔科夫模型对频域回声路径进行建模, 给出了精确的卡尔曼滤波方程, 并讨论了过程噪声和观测噪声的估计问题。为降低算法计算复杂度, 提出了低复杂度卡尔曼滤波算法。另外, 在更新滤波器时加入远端信号邻近频点的信息来进一步提高回声抵消性能。实验结果表明, 所提算法对近端干扰不敏感, 不需要额外的双端对讲检测算法, 且比传统的频域自适应滤波算法具有更快的收敛速度。

**关键词:** 回声抵消; 短时傅里叶变换; 卡尔曼滤波; 低复杂度

中图分类号: TN912

文献标志码: A

文章编号: 1000-3630(2022)-05-0757-06

## Kalman filter based acoustic echo cancellation in the STFT domain

LIU Yang<sup>1,2</sup>, YANG Feiran<sup>1,2,3</sup>, LIANG Zhaojie<sup>4</sup>, YANG Jun<sup>1,2,3</sup>

(1. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. IOASONIC (Su zhou) Technologies Co. Ltd., Suzhou 215163, Jiangsu, China;

4. Shenzhen Power Supply Bureau Co. Ltd, Shenzhen 518000, Guangdong, China)

**Abstract:** In this paper, a low-complexity Kalman filter algorithm in the short-time Fourier transform (STFT) domain is proposed for acoustic echo cancellation. The observation equation is obtained based on the convolutive transfer function in the STFT domain, and the echo path in the frequency domain is modeled by using the first-order Markov process. The exact Kalman filter is derived, and the estimation of the process noise and observation noise is discussed. A fast implementation method is presented to reduce the complexity of the exact Kalman filter. In addition, more adjacent frequency points of the far-end signal are introduced to further improve the echo cancellation performance. Experimental results show that the proposed algorithm is insensitive to the near-end interference and unnecessary to have an explicit double-talk detection algorithm, and it converges faster than the frequency domain adaptive filtering algorithm.

**Key words:** echo cancellation; short-time Fourier transform (STFT); Kalman filter; low complexity

## 0 引言

受到新冠疫情全球大流行的影响, 在线会议系统得到了快速的发展, 语音通信是该系统的核心功能。声音也是人工智能与物联网融合时代最关键的入口之一。近年来, 人机交互技术在自动导航系统、智能音箱、语音遥控电视机等一系列产品中获得成功应用。回声抵消是语音通信和人机交互的关键问题之一<sup>[1-2]</sup>。上述这些技术的快速发展拓展了回

声抵消的应用场景并对回声抵消提出了更高的要求。

自适应滤波是回声抵消问题最成熟的解决方案。该方法通过辨识扬声器到近端传声器之间的回声路径来估计近端传声器采集到的回声信号, 再用传声器信号减去估计的回声信号得到误差信号, 即需要的近端语音信号。声学回声路径通常需要利用几千阶的 FIR 滤波器来进行建模, 因而算法的计算复杂度是首先需要考虑的问题之一。目前, 频域自适应滤波算法是声学回声抵消的标准解决方案<sup>[3-4]</sup>。约束的频域自适应滤波算法利用快速傅里叶变换 (Fast Fourier Transformation, FFT) 来实现精确的线性卷积和线性相关操作<sup>[2]</sup>, 因而相比直接在时域实现具有更低的复杂度。

收稿日期: 2021-05-25; 修回日期: 2021-07-05

基金项目: 南方电网科技项目(090000KK52190062)。

作者简介: 刘杨(1997—), 女, 江苏无锡人, 博士研究生, 研究方向为回声抵消。

通信作者: 杨军, E-mail: jyang@mail.ioa.ac.cn

还有一类回声抵消方法是在短时傅里叶变换域(Short Time Fourier Transformation, STFT)直接进行的。与上述基于线性卷积的方法不同,短时傅里叶变换域的算法实现的一般不是一个精确的线性卷积<sup>[5]</sup>,而只是一个逼近的方法。早期的算法<sup>[6-10]</sup>首先估计出回声信号的功率谱再利用语音增强中常用的维纳滤波来完成回声抑制,但这些算法都忽略了相位信息,因而得到的近端语音质量较差。最近发展的一类算法<sup>[11-14]</sup>同时估计回声信号的幅度和相位信息,因而近端语音质量得到了保证。文献[11]和[13]的算法没有讨论在双端对讲状态下滤波器的更新问题。文献[12]和[14]的算法分别直接采用卡尔曼滤波和递归最小二乘(Recursive Least Squares, RLS)在每一个频点进行滤波器系数更新,因而复杂度较高。

本文研究了短时傅里叶变换域的单通道线性回声抵消问题。首先给出了标准的卡尔曼滤波更新方程,讨论了过程噪声和观测噪声的估计方法,进一步给出了简化的卡尔曼滤波算法来降低计算复杂度。本文另一个贡献是引入了远端信号更多的临近频点来进一步提高回声抵消的性能。最后通过实验验证了所提算法的有效性。

## 1 模型

考虑如图1所示的单通道回声抵消模型,传声器接收到的信号可以表示为

$$y(t) = d(t) + v(t) \quad (1)$$

其中:回声信号  $d(t) = h(t) * x(t)$  是远端参考信号  $x(t)$  通过扬声器到传声器的回声路径  $h(t)$  后得到的,  $v(t)$  包括近端语音  $s(t)$  和噪声信号  $n(t)$ 。传声器接收信号在STFT域可表示为

$$Y(k, n) = D(k, n) + V(k, n) \quad (2)$$

其中:  $Y(k, n)$ 、 $D(k, n)$  和  $V(k, n)$  分别表示第  $n$  个时间帧和第  $k$  个频点的传声器信号、回声信号和近端信号。在STFT域的回声信号可以表示为<sup>[14-15]</sup>

$$D(k, n) = \sum_{k'=0}^{N-1} \sum_{l=-\infty}^{\infty} H(k, k', l) X(k', n-l) \quad (3)$$

其中:  $X(k, n)$  是第  $n$  个时间帧和第  $k$  个频率块的远端参考信号,  $H(k, k', l)$  为频点  $k$  和  $k'$  之间的交叉子带滤波器系数,  $N$  为傅里叶变换窗长。

直接使用式(3)进行计算复杂度较高,因而不是不现实的。下面我们给出几种近似的实现方法。假设STFT分析窗的频率分辨率足够高,则各频率之间信号相互独立,即当  $k \neq k'$  时,  $H(k, k', l) = 0$ 。当  $k = k'$  时,可以将  $H(k, k', l)$  重写为第  $n$  个时间帧和第  $k$  个频点的第  $l$  个时延块的回声路径  $H_l(k, n)$ 。随着  $l$

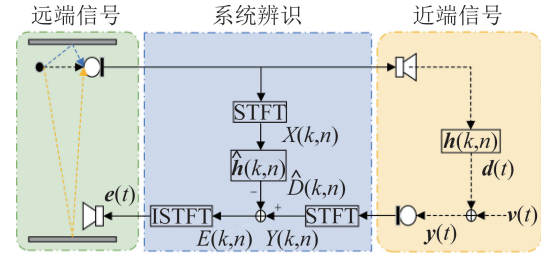


图1 单通道回声抵消模型

Fig.1 Single-channel echo cancellation model

的增加,  $H_l(k, n)$  的值随时延块的衰减越来越快,因而可以只保留  $L$  个系数对STFT域回声信号进行近似。从而,STFT域的回声信号可以近似表示为

$$\begin{aligned} D(k, n) &\approx \sum_{l=0}^{L-1} H_l(k, n) X(k, n-l) \\ &= \mathbf{h}^H(k, n) \mathbf{x}(k, n) \end{aligned} \quad (4)$$

其中:  $\mathbf{h}(k, n) = [H_0(k, n) \dots H_{L-1}(k, n)]^T$  为滤波器系数向量,  $\mathbf{x}(k, n) = [X(k, n) \dots X(k, n-L+1)]^T$  为输入信号向量。

在实际应用中,STFT分析窗的频率分辨率可能不足以使  $k \neq k'$  时的  $H(k, k', l) = 0$ 。此时引进相邻时间帧和频率帧的信息,可以进一步提高算法的回声抵消量<sup>[10]</sup>。如图2所示是扩充邻近频点信息的两种方式。第一种扩充的远端参考信号向量  $\mathbf{x}^{(1)}(k, n)$  表示为

$$\mathbf{x}^{(1)}(k, n) = [\mathbf{x}^T(k-L, n) \dots \mathbf{x}^T(k, n) \dots \mathbf{x}^T(k+1, n)]^T \quad (5)$$

式(5)使用  $L$  个时间帧以及每个时间帧当前频点上下各  $K$  个频率帧的信息,向量长度为  $(2K+1)L$ 。第二种扩充的远端参考信号向量  $\mathbf{x}^{(2)}(k, n)$  表示为

$$\mathbf{x}^{(2)}(k, n) = [X(k, n-L) \dots X(k, n-1) \dots X(k-K, n) \dots X(k+K, n)]^T \quad (6)$$

式(6)使用  $L$  个时间帧以及当前时频点上下各  $K$  个频率帧的信息,向量长度变为  $L+2K$ ,图2中的箭头是向量元素的填充顺序。远端信号向量进行扩充之后,每个时频点对应的回声路径向量  $\mathbf{h}(k, n)$  的长度也扩充成相应的长度。

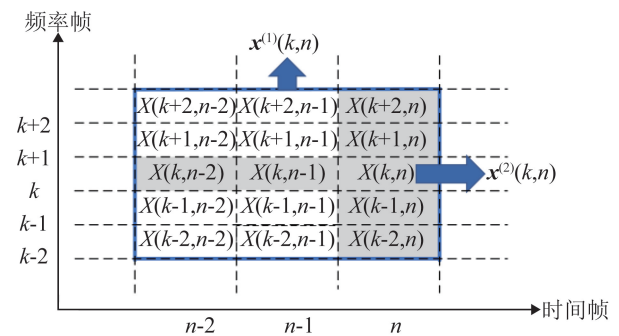


图2 邻近频率帧信息扩充类型

Fig.2 Expansion types of adjacent frequency frame information

假设估计的第 $n$ 个时间帧，第 $k$ 个频率块的房间回声路径系数向量为 $\hat{\mathbf{h}}(k, n)$ ，则估计的频域回声信号为

$$\hat{D}(k, n) = \hat{\mathbf{h}}^H(k, n) \mathbf{x}(k, n) \quad (7)$$

从而可以得到估计的频域近端信号：

$$E(k, n) = Y(k, n) - \hat{D}(k, n) \quad (8)$$

## 2 STFT域卡尔曼滤波

### 2.1 状态方程和观测方程

在回声抵消问题中，实际要估计的是房间回声路径的频域系数向量 $\mathbf{h}(k, n)$ 。本文采用一阶马尔可夫模型对回声路径的变化进行建模：

$$\mathbf{h}(k, n+1) = \mathbf{h}(k, n) + \mathbf{w}(n) \quad (9)$$

其中： $\mathbf{w}(n)$ 是体现回声路径变化的状态噪声。假设 $\mathbf{w}(n)$ 服从0均值高斯分布，状态噪声的协方差矩阵为 $\Phi_w(k, n) = E[\mathbf{w}(n)\mathbf{w}^H(n)]$ 。

利用式(4)，得到STFT域观测方程的表达式：

$$Y(k, n) = \mathbf{h}^H(k, n) \mathbf{x}(k, n) + V(k, n) \quad (10)$$

### 2.2 卡尔曼滤波器的更新过程

在卡尔曼滤波中，状态向量估计误差的协方差矩阵为

$$\Phi_e(k, n) = E\{[\mathbf{h}(k, n) - \hat{\mathbf{h}}(k, n)][\mathbf{h}(k, n) - \hat{\mathbf{h}}(k, n)]^H\} \quad (11)$$

最小化目标函数(11)，得到卡尔曼滤波器的更新过程为<sup>[15]</sup>

$$E(k, n) = Y(k, n) - \hat{\mathbf{h}}^H(k, n-1) \mathbf{x}(k, n) \quad (12)$$

$$\mathbf{K}(k, n) = \Phi_e(k, n-1) \mathbf{x}(k, n) =$$

$$[\mathbf{x}^H(k, n) \Phi_e(k, n-1) \mathbf{x}(k, n) + \Phi_v(k, n)]^{-1} (13)$$

$$\hat{\mathbf{h}}(k, n) = \hat{\mathbf{h}}(k, n-1) + \mathbf{K}(k, n) E(k, n) \quad (14)$$

$$\Phi_e(k, n) = [\mathbf{I}_L - \mathbf{K}(k, n) \mathbf{x}^H(k, n)] \Phi_e(k, n-1) + \Phi_w(k, n) \quad (15)$$

其中： $\mathbf{K}(k, n)$ 是卡尔曼增益。

### 2.3 状态噪声和观测噪声的估计

状态噪声协方差 $\Phi_w(k, n)$ 估计的准确性影响到系统对通道变化的跟踪能力和算法的稳态性能。我们假设 $\mathbf{w}(n)$ 的各个元素之间互不相关，则 $\Phi_w(k, n)$ 变为对角矩阵 $\Phi_w(k, n) = \phi_w(k, n) \mathbf{I}$ 。利用下式估计 $\phi_w(k, n)$ ：

$$\phi_w(k, n) = \frac{1}{L} (1 - c^2) \|\hat{\mathbf{h}}(k, n)\|_2^2 \quad (16)$$

其中：参数 $c$ 的值越大，算法的稳态收敛误差越小，跟踪能力越差，反之亦然。

观测噪声协方差 $\Phi_v(k, n)$ 的估计准确性影响到系统的双端对讲性能。当系统达到稳态时，误差信

号近似为近端信号，因而可以直接利用误差信号来估计观测噪声方差：

$$\Phi_v(k, n) = \alpha \Phi_v(k, n-1) + (1 - \alpha) \|E(k, n)\|^2 \quad (17)$$

其中： $\alpha$ 是遗忘因子。

### 2.4 低复杂度实现

为了简化卡尔曼滤波的计算复杂度，首先将误差协方差矩阵 $\Phi_e(k, n)$ 简化为对角矩阵 $\Phi_e(k, n) \approx \phi_e(k, n) \mathbf{I}_L$ <sup>[16]</sup>。则卡尔曼增益的更新公式变为<sup>[16]</sup>

$$\mathbf{K}(k, n) = \frac{\phi_e(k, n-1) \mathbf{x}(k, n)}{\phi_e(k, n-1) \|\mathbf{x}(k, n)\|^2 + \Phi_v(k, n)} \quad (18)$$

为了计算式(15)，将矩阵 $\mathbf{K}(k, n) \mathbf{x}^H(k, n)$ 对角化为

$$\mathbf{K}(k, n) \mathbf{x}^H(k, n) \approx \frac{1}{L} \mathbf{x}^H(k, n) \mathbf{K}(k, n) \mathbf{I}_L \quad (19)$$

从而 $\phi_e(k, n)$ 的更新变为

$$\phi_e(k, n) = \left(1 - \frac{1}{L} \mathbf{x}^H(k, n) \mathbf{K}(k, n)\right) \phi_e(k, n-1) + \phi_w(k, n) \quad (20)$$

低复杂度STFT域卡尔曼滤波器(Low Complexity STFT-Kalman, LC-STFT-Kalman)更新时，按顺序更新式(12)、(19)、(14)、(20)，其中状态噪声协方差的估计仍然采用式(16)和(17)。

## 3 计算复杂度

表1列出了STFT域的归一化最小均方误差<sup>[11]</sup>(STFT-NLMS)、STFT-RLS<sup>[14]</sup>、变步长分块频域自适应滤波<sup>[4]</sup>(VS-PBFDFAF)、STFT-Kalman和LC-STFT-Kalman 5种算法更新每个采样点所需的乘法数。

表1 各算法更新每个采样点所需的乘法数  
Table 1 Numbers of multiplications required for each output of different algorithms

算法名	输出一个采样点所需乘法数
STFT-NLMS	$(5P + 14 + 3 \log_2 N)/(1 - R)$
STFT-RLS	$(5P^2 + 7P + 14 + 3 \log_2 N)/(1 - R)$
VS-PBFDFAF	$50B_b + 102B + (10 + 4B_b + 4B) \log_2 N$
STFT-Kalman	$(4P^2 + 8P + 15 + 3 \log_2 N)/(1 - R)$
LC-STFT-Kalman	$(9P + 15 + 3 \log_2 N)/(1 - R)$

设定用于估计回声路径的滤波器长度为 $M$ ，STFT域算法的FFT点数为 $N$ ，重叠率为 $R \in [0, 1]$ ，即每次帧移 $N(1 - R)$ 个采样点，每个时频点估计的滤波器长度为 $P = M/[N(1 - R)]$ 。VS-PBFDFAF算法的FFT点数为 $N$ ，重叠率为50%，即每次帧移 $N/2$ 点，则主滤波器块数为 $B = 2M/N$ ，阴影滤波器块数为 $B_b = B/2$ 。已知计算所需要的乘法次数 $N$ 点FFT

(IFFT)所需实数乘法次数为 $N\log_2 N$ 、复数乘法为实数乘法的4倍、开方运算为实数乘法的40倍、除法运算为实数乘法的20倍。

图3给出了5种算法更新每个采样点所需的乘法次数随滤波器长度 $M$ 的变化,其中短时傅里叶变换点数 $N=512$ ,重合率 $R=75\%$ 。从图中可以看出,STFT-Kalman和STFT-RLS更新一个采样点所需的乘法次数随着回声路径的长度增加的趋势比其他三种算法快很多,其原因是这两种算法的计算量与每个时频点的滤波器长度的平方成正比。LC-STFT-Kalman的计算量比STFT-NLMS算法大,但比VS-PBFDFA小。

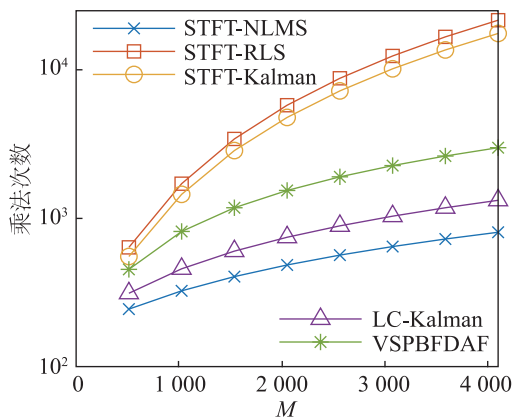


图3 5种算法更新每个采样点所需乘法次数随 $M$ 的变化  
Fig.3 The numbers of multiplications required for each output of different algorithms as a function of filter length  $M$

## 4 实验结果和分析

为了区分回声和近端语音,把TIMIT数据集中截取的英文作为远端信号,汉语普通话识别句表<sup>[17]</sup>中截取的中文作为近端信号。每段信号的总长度为40 s,采样率为16 kHz。如图4所示,远端信号为持续干净语音,近端信号在20~30 s出现语音,在近端传声器接收信号中加入白噪声,信噪比(Signal to Noise Ratio, SNR)为30 dB。在生成回声时,回声路径采用实际录制的房间脉冲响应并截断至2048点。在混合近端传声器接收信号时,近端信号与回声信号之比(Signal to Echo Ratio, SER)为0 dB。STFT的帧重合率为75%,分析窗和合成窗都采用汉宁窗。采用回声往返损耗增强(Echo Return Loss Enhancement, ERLE)对回声抵消算法的性能进行评估,其计算方法为

$$E_{\text{ERLE}} = 10 \lg \frac{E \left\{ [y(n) - v(n)]^2 \right\}}{E \left\{ [e(n) - v(n)]^2 \right\}} \quad (21)$$

但ERLE仅能评估回声抵消量,无法体现通过

算法后的近端语音失真程度,因此采用感知语音质量评价<sup>[18]</sup>(Perceptual Evaluation of Speech Quality, PESQ)对回声抵消算法收敛后在双讲阶段的性能进行评估。状态转移因子 $c \in (0, 1]$ ,越接近1,算法的稳态收敛误差越小,收敛后的回声抵消性能越好。本文中采用 $c=0.999\ 992$ 和 $\alpha=0.8$ ,状态噪声协方差和观测噪声方差的初值分别为 $\Phi_w(k, 0) = \phi_w(k, 0)\mathbf{I} = 0.05\mathbf{I}$ 和 $\Phi_v(k, 0) = 0.05$ 。

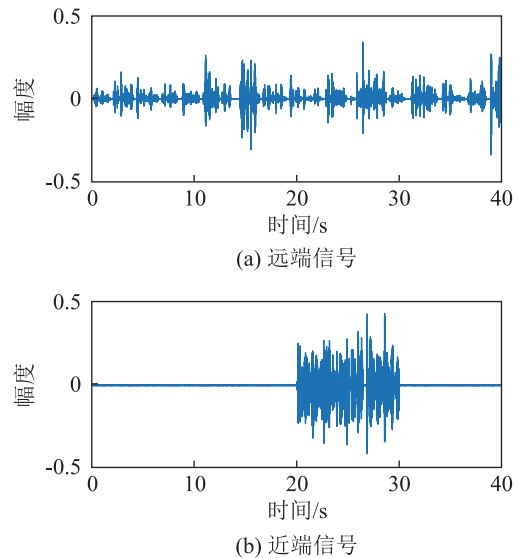


图4 实验采用的远端和近端信号  
Fig.4 Far-end and near-end signals used in the experiment

图5比较了不扩充邻近频点的卡尔曼滤波算法、STFT-RLS<sup>[14]</sup>和VS-PBFDFA<sup>[6]</sup>的性能。设SNR为30 dB, SER为0 dB,将STFT-Kalman和LC-STFT-Kalman的参数都设置成 $N=512$ 、 $L=16$ 。STFT-RLS的傅里叶变换点数为512,每个时频点的滤波器长度为16,重叠率为75%。VS-PBFDFA估计的时域滤波器长度为2048,分成4块,每个子滤波器的长度为512。从图5可以看出,基于STFT

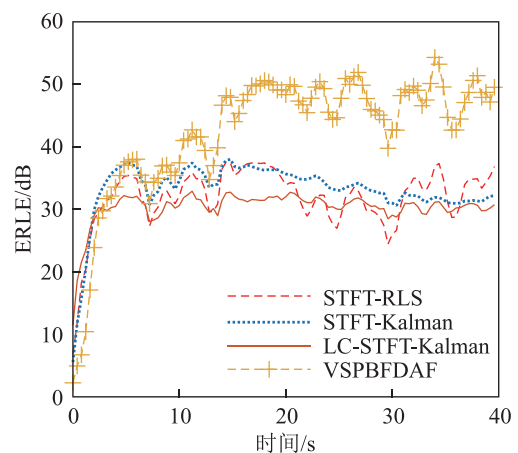


图5 4种不同算法的回声往返损耗增强(ERLE)随时间变化曲线  
Fig.5 Variations of ERLE with time for different algorithms

域的3种算法都比传统的VS-PBFDFAF算法具有更快的收敛速度，这主要是由于前者在每个频点需要更新的系数较少。算法达到稳态后，VS-PBFDFAF的ERLE最大，简化后的卡尔曼滤波算法LC-STFT-Kalman比原来的精确实现形式性能有所下降。STFT-RLS、STFT-Kalman、LC-STFT-Kalman和VS-PBFDFAF等4种算法更新每个采样点所需的乘法数目分别为5732、4776、744、1028。

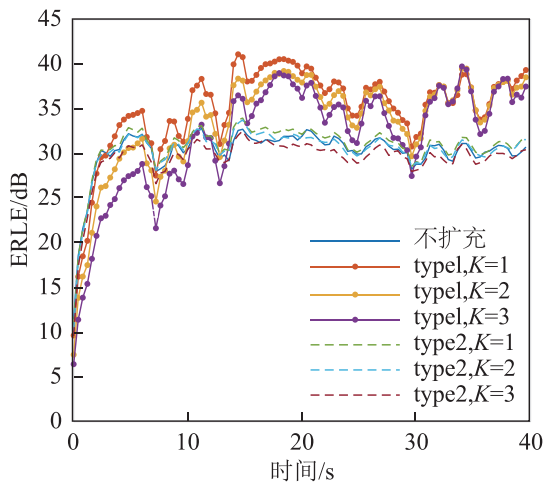
图6比较了不同的远端频段扩充方式对回声抵消性能ERLE的影响。图6(a)为 $N=512, L=16$ 的LC-STFT-Kalman算法的ERLE与扩充方式、扩充频点数的关系。从收敛速度来看，第一种扩充(type1)随着扩充频点的增加收敛速度明显变慢，第二种扩充(type2)对收敛速度的影响不大。这是由于采用第一种扩充方式后每个频点需要更新的滤波器系数明显增加。从算法收敛后的ERLE来看，第一种扩充收敛后的回声抵消量整体比未扩充时的高， $K=1$ 时效果最好。第二种扩充对回声抵消量的增

加并不明显。图6(b)为 $N=2048, L=4$ 的LC-STFT-Kalman算法的ERLE与扩充方式、扩充频点数的关系，其收敛规律与图6(a)相同。

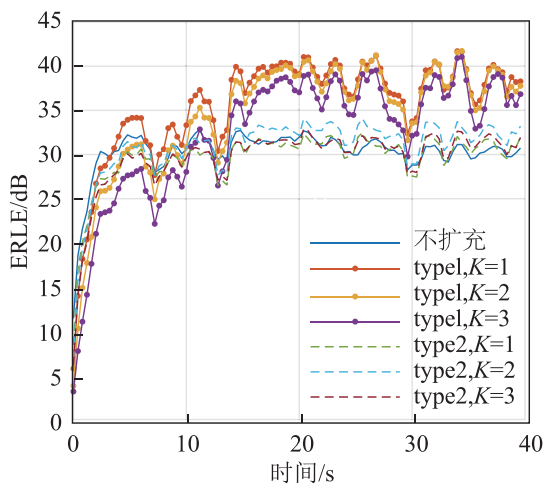
表2中给出了各算法收敛后双讲阶段的PESQ。对比表2和图5，VS-PBFDFAF因其回声抵消量最大，且收敛后对语音的损伤最小，在双讲阶段的PESQ是最大的。LC-STFT-Kalman使用第一种扩充的PESQ更高，且 $N=2048, L=4$ 时的双讲稳定性比 $N=512, L=16$ 时好，算法的PSEQ整体更高。

表2 各算法双讲阶段的感知语音质量评价(PSEQ)对比  
Table 2 Comparison of PESQ in the double-talk phase of the algorithms

算法类型	PESQ
VS-PBFDFAF	4.49
STFT-RLS	4.17
STFT-Kalman	4.11
LC-STFT-Kalman $N=512, L=16$	
不扩充	3.97
type1, $K=1$	4.27
type1, $K=2$	4.25
type1, $K=3$	4.26
type2, $K=1$	4.00
type2, $K=2$	3.97
type2, $K=3$	3.95
LC-STFT-Kalman $N=2048, L=4$	
不扩充	4.02
type1, $K=1$	4.37
type1, $K=2$	4.37
type1, $K=3$	4.35
type2, $K=1$	4.05
type2, $K=2$	4.13
type2, $K=3$	4.02



(a)  $N=512, L=16$ 的LC-STFT-Kalman



(b)  $N=2048, L=4$ 的LC-STFT-Kalman

图6 扩充频点对ERLE的影响

Fig.6 Influence of expanded frequency points on ERLE

表3给出了不同扩充方式和不同邻近频点数 $K$ 对算法计算复杂度的影响。从表3中可以看出， $N=512, L=16$ 的组合要比 $N=2048, L=4$ 的计算复杂度高很多。同一种组合中，在上下扩充频点数 $K$ 相同时，第二种扩充方式滤波器长度增加比第一种方式要少得多，因而其计算复杂度也小很多。

表3 更新单个采样点的计算复杂度随扩充方式和 $K$ 的变化  
Table 3 Variation of the computational complexity of updating a single sample point with expansion mode and  $K$

组合方式	$K$	不扩充	第一种	第二种
$N=512, L=16$	0	744	—	—
	1	—	1896	816
	2	—	3048	888
	3	—	4200	960
$N=2048, L=4$	0	312	—	—
	1	—	600	384
	2	—	888	456
	3	—	1176	528

## 5 结 论

本文提出了用于单通道回声抵消的STFT域的卡尔曼滤波算法,并给出了其快速实现方法。本文还在简化模型的基础上应用两种扩充临近频点提高回声抵消性能的方式,使得扩充频点后的简化卡尔曼滤波算法的回声抵消量性能有了显著提升。使用 $K=1$ 的第一种扩充频点方式,可以使得LC-STFT-Kalman的回声抵消量达到40 dB,比精确实现的STFT-Kalman高5 dB。所提算法不需要双讲检测,具有比传统的频域自适应滤波算法更低的复杂度和更快的收敛速度,在很多场景下均具有较好应用前景。

### 参 考 文 献

- [1] BENESTY J, GÄNSLER T, MORGAN D R, et al. Advances in Network and Acoustic Echo Cancellation[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001.
- [2] HAYKIN S. Adaptive filter theory 4th edition[M]. USA: Prentice-Hall, 2002
- [3] YANG F R, ENZNER G, YANG J. Statistical convergence analysis for optimal control of DFT-domain adaptive echo canceler[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(5): 1095-1106.
- [4] YANG F R, YANG J. Optimal step-size control of the partitioned block frequency-domain adaptive filter[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2018, 65(6): 814-818.
- [5] AVARGEL Y, COHEN I. System identification in the short-time Fourier transform domain with crossband filtering[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(4): 1305-1319.
- [6] FALLER C, CHEN J D. Suppressing acoustic echo in a spectral envelope space[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 1048-1062.
- [7] FALLER C, TOURNERY C. Robust acoustic ECHO control using a simple ECHO path model[C]//2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. Toulouse, France. IEEE, 2006: V.
- [8] PARK Y S, CHANG J H. Frequency domain acoustic echo suppression based on soft decision[J]. IEEE Signal Processing Letters, 2009, 16(1): 53-56.
- [9] YANG F R, WU M, YANG J. Stereophonic acoustic echo suppression based on Wiener filter in the short-time Fourier transform domain[J]. IEEE Signal Processing Letters, 2012, 19(4): 227-230.
- [10] LEE C M, SHIN J W, KIM N S. Stereophonic acoustic echo suppression incorporating spectro-temporal correlations[J]. IEEE Signal Processing Letters, 2014, 21(3): 316-320.
- [11] LUIS VALERO M, HABETS E A P. Low-complexity multi-microphone acoustic echo control in the short-time Fourier transform domain[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(3): 595-609.
- [12] PARK J, CHANG J H. State-space microphone array nonlinear acoustic echo cancellation using multi-microphone near-end speech covariance[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(10): 1520-1534.
- [13] CARBAJAL G, SERIZEL R, VINCENT E, et al. Joint NN-supported multichannel reduction of acoustic echo, reverberation and noise[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2158-2173.
- [14] CHO B J, PARK H M. Stereo acoustic echo cancellation based on maximum likelihood estimation with inter-channel-correlated echo compensation[J]. IEEE Transactions on Signal Processing, 2020, 68: 5188-5203.
- [15] BRAUN S, HABETS E A P. Online dereverberation for dynamic scenarios using a Kalman filter with an autoregressive model[J]. IEEE Signal Processing Letters, 2016, 23(12): 1741-1745.
- [16] GUO J F, YANG F R, ENRNER G, et al. Tracking analysis and improvement of broadband Kalman filter using the two-echo-path model as a rapid tunnel[C]//2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC). Tokyo, Japan. IEEE, 2018: 341-345.
- [17] 榔听, 陈艾婷, 李佳楠, 等. 嘈杂语噪声下普通话儿童语句测试表的标准化[J]. 听力学及言语疾病杂志, 2009, 17(4): 318-322.  
XI Xin, Chen, Aiting, Li Jia', et al. Standardized mandarin sentence perception in babble noise test materials for children [J]. Journal of Audiology and Speech Pathology, 2009, 17(4): 318-322.
- [18] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs[C]//2001 IEEE International Conference on Acoustic Speech and Signal Processing Proceedings. Salt Lake City, Utah, USA. IEEE, 2001: 749-752.