

基于频域逐级回归的声学回声控制

姜开宇* 吴超 国雁萌 付强 颜永红

(中国科学院声学研究所 北京 100190)

(中科院语言声学与内容理解重点实验室 北京 100190)

摘要: 传统声学回声控制算法一般采用基于随机梯度法更新的频域分块自适应滤波(PBFDAF)方法,但在以语音为主要回声信号的室内混响环境中,由于回声路径不稳定,往往收敛速度较慢,难以实现足够的回声抑制。该文提出一种基于频域逐级回归的声学回声控制算法。通过逐级回归分析远端信号和麦克风信号之间的线性关系,可以在保持较小的偏差的同时实现收敛较快的系统估计。同时,由于逐级分析了两通道间的短时相干性,因而该算法无需像常见方法一样,额外进行基于通道间相干函数的残余回声抑制或双讲检测,从而保持系统的紧凑性。若进一步假定近端背景噪声准平稳,则可利用基于近端信号非平稳程度的自适应平滑因子,在实现系统估计快速收敛的同时确保其稳定性。实验表明,该方法在常见的近端环境噪声水平下,在收敛速度和稳态误差上相对传统方法有显著优势,非常适合应用在室内远讲模式下的声学回声控制中。

关键词: 语音信号处理; 声学回声控制; 逐级回归; 声学回声抵消; 声学回声抑制

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2014)12-2896-06

DOI: 10.3724/SP.J.1146.2014.00131

Acoustic Echo Control Based on Frequency-domain Stage-wise Regression

Jiang Kai-yu Wu Chao Guo Yan-meng Fu Qiang Yan Yong-hong

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China)

(Key Laboratory of Speech Acoustics and Content Understanding, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Traditional echo control techniques as Partitioned Block Frequency Domain Adaptive Filter (PBFDAF) with stochastic gradient adaptive method usually endure slow convergence and insufficient echo suppression in reverberant room when the echo is speech and the echo path is unstable. An algorithm based on frequency domain stage-wise regression is proposed for acoustic echo control to achieve faster convergence of the system estimation with insignificant bias. Commonly used additional double-talk detector and inter-channel coherence based residual echo suppressor are not needed since short-time coherence analysis is performed in each stage. By further making mild assumptions on the quasi-stationarity of the near-end background noise, both fast convergence and stability of the estimation can be achieved simultaneously with a non-stationarity controlled smoothing factor. Experiments are carried out to show the superiority of the proposed approach in terms of convergence speed and steady state error in distant talking mode in ordinary room environment with various common levels of background noise.

Key words: Speech signal processing; Acoustic echo control; Stage-wise regression; Acoustic echo cancellation; Acoustic echo suppression

1 引言

在远讲或免提模式下的语音通信和人机交互系统中,远端语音信号经扬声器播放和房间传播后,会被本地麦克风接收,从而产生回声,严重影响语

音通信系统的质量和人机交互系统的可靠性。为此,一般通过回声抵消(echo cancellation)^[1-10]或回声抑制(echo suppression)^[11-16]等技术来实现声学回声控制。回声抵消技术假定远端信号和麦克风信号之间的关系可表示为一个定长的线性系统,通过在时域^[1]或频域^[2]中进行系统辨识,来估计麦克风信号中的回声成分,进而消除回声。自适应控制技术通过在强近端干扰出现时,停止或减慢系统估计的更新速率,从而防止自适应滤波器发散^[3-10]。回声抑制技术通常基于较松散的回声路径模型假定,它既可

2014-01-20 收到, 2014-04-18 改回

国家自然科学基金(10925419, 90920302, 61072124, 11074275, 11161140319)和中国科学院战略性先导科技专项(XDA06030100)资助课题

*通信作者: 姜开宇 jiangkaiyu@hcccl.ioa.ac.cn

独立工作^[11,12], 也可作为自适应滤波的后续处理来抑制残余回声^[13-16]。

自适应滤波器的参数估计通常被建模成多元线性回归问题, 但最小二乘估计的闭式求解, 需要对远端信号的协方差矩阵求逆, 不仅计算量很大, 而且在协方差矩阵病态的条件下存在数值不稳定的问题。随机梯度下降法可避免直接矩阵求逆, 计算量小, 因而得到广泛应用^[3-6]。但在远端信号为语音时, 由于其协方差矩阵特征值分散, 收敛速度较慢。因此, 研究一种计算量不大, 同时又能结合最小二乘估计, 不依赖远端信号统计特性的快速收敛算法, 对实际应用很有意义。

在语音通信和交互系统中, 由于近端存在语音和背景噪声, 远端扬声器信号和近端麦克风信号之间的线性关系会受到干扰, 这通常被称作“双讲”问题。而且, 由于室内混响的影响, 经常需要较多参数才能较精确地建模回声路径。更重要的是, 回声路径可能时变, 甚至因环境扰动而出现突变, 因此只能基于有限时长内的观测数据进行无偏估计, 这将导致估计方差较大。另外由于远端信号是语音, 在时频分布上比较稀疏, 能量较弱的频带因为淹没在近端噪声中而无法被有效辨识, 因而回声路径的估计相对于真实值经常是有差异的。因此, 在实际环境下, 要保持以较小的方差获得回声路径的无偏估计, 往往不太可能。相对而言, 一个偏差不大, 但收敛快速的系统估计, 对于确保算法在各种情况下都有足够的回声消除量, 从而保证系统的稳定性尤为重要。

注意到自适应滤波, 自适应控制和残余回声抑制模块虽然往往是独立实现的, 却用到了长短不同但同属线性模型的假设。特别地, 短阶线性系统的假设被广泛应用在基于相干函数估计的自适应控制^[7-9]和回声抑制^[15,16]算法中。值得一提的是, 它对瞬态干扰的鲁棒性和快速收敛特性主要应归因于单参数的加权最小二乘估计(Weighted Least Squares, WLS)。但是此种方法由于基于短时谱估计, 在长回声路径下存在较大的建模误差。文献[16]尝试在幅度谱减的框架内, 将其扩展到更长的系统中用作残余回声抑制。但是幅度谱减可以看作由三角不等式解释的过减, 在理论上具有一定偏差。可见, 研究一种将前述的 3 个模块整合到同一框架下, 同时避免不必要的偏差的方法, 将减少算法中的参数, 提高算法的性能和稳健性, 并使得算法在不同应用环境下参数的重新调试变得更加简单。

基于以上考虑, 本文提出在复频域的逐级回归方法, 将两通道之间的线性关系建模为逐级回归模

型, 并利用逐级的短时谱估计实现长系统的辨识, 在保持偏差较小的同时, 不仅能够实现系统估计的快速收敛, 而且具有抗瞬态干扰(近端语音)的能力。在本文的逐级回归中, 每级只进行一个单参数的简单回归, 所以不必进行矩阵求逆就能求得逐级的最小二乘估计, 能够达到较快的收敛速度。此外, 如果估计时使用了足够多的独立样本, 则该估计对于非持续的干扰(近端语音)将保持鲁棒。这是因为, 如果用恒定方差的白噪声序列对近端信号(包括近端语音和背景噪声)建模, 即使其分布不是高斯的, 根据 Gauss-Markov 定理, 最优的线性无偏估计子为最小二乘估计。尽管逐级回归相对于多元线性回归通常有一定偏差^[17], 但它具有更快的收敛速度和对双讲的鲁棒性。同时, 通过对回声路径衰减形状的合理假设, 可以预先合理确定逐级回归中引入回归变量的次序, 并对估计得到的回归系数做出合理约束, 从而将系统估计的偏差和均方误差控制在一个较低的范围, 因而它非常适合于以语音为主要回声信号的非平稳混响环境。为防止因快速收敛特性而可能导致的快速发散, 本文还提出利用基于近端信号非平稳程度的自适应平滑因子, 从而保证了算法的稳定性。

本文后续部分安排如下: 首先介绍用于声学回声控制的模型近似及频域逐级回归算法, 同时提出一个时频依赖的平滑因子及随机幅度谱最小值约束, 从而实现一个完整的声学回声控制算法。然后, 对该方法和一个分块频域自适应滤波器(Partitioned Block Frequency Domain Adaptive Filter, PBFDAF)结合自适应控制的实现^[2,3,18]进行了多种条件下的性能对比测试, 最后给出结论。

2 问题分析及逐级回归介绍

2.1 模型近似

设麦克风信号表示为

$$y(t) = \mathbf{h}^T(t)\mathbf{s}(t) + n(t) \quad (1)$$

其中 $\mathbf{h}(t) = [h_1(t) \cdots h_M(t)]^T$, 表示待估计的长度为 M 的回声路径, $\mathbf{s}(t)$ 是远端信号向量, $n(t)$ 为近端噪声, $(\cdot)^T$ 表示转置。将 $\mathbf{h}(t)$ 和 $\mathbf{s}(t)$ 分为长度为 B 的 L 个块, 则 $M = LB$, 且有

$$\mathbf{h}_l(t) = [h_{(l-1)B+1}(t) \cdots h_{lB}(t)]^T \quad (2)$$

$$\mathbf{s}_l(t) = [s(t-lB+1) \cdots s(t-lB+B)]^T \quad (3)$$

其中 $l = 1, 2, \dots, L$ 。用 \mathbf{F} 和 \mathbf{F}^{-1} 表示 $2B \times 2B$ 的 DFT 和 IDFT 矩阵。并定义

$$\mathbf{H}_l(k) = \mathbf{F} \left[\mathbf{h}_l^T(kB) 0 \cdots 0 \right]^T \quad (4)$$

$$\mathbf{S}_l(k) = \mathbf{F} \begin{bmatrix} \mathbf{s}_l((k-1)B) \\ \mathbf{s}_l(kB) \end{bmatrix} \quad (5)$$

其中 k 是块的编号。将其代入式(1), 则有

$$\mathbf{y}(k) = \begin{bmatrix} \mathbf{0}_B & \mathbf{I}_B \end{bmatrix} \sum_{l=1}^L \mathbf{F}^{-1} \mathbf{H}_l(k) \odot \mathbf{S}_l(k) + \mathbf{n}(k) \quad (6)$$

其中 \odot 表示 Hadamard 积, $\mathbf{0}_B$ 和 \mathbf{I}_B 表示 $B \times B$ 的 0 矩阵和单位阵, $\mathbf{y}(k) = [y(kB-B+1) \cdots y(kB)]^T$, $\mathbf{n}(k) = [n(kB-B+1) \cdots n(kB)]^T$ 。根据这个表达式可以得到常用的 PBFDAF 结构^[2]。如果 $\mathbf{s}(t)$ 以 B 为周期, 则

$$\mathbf{Y}(k) = \sum_{l=1}^L \mathbf{H}_l(k) \odot \mathbf{S}_l(k) + \mathbf{N}(k) \quad (7)$$

$$\text{其中 } \mathbf{Y}(k) = \mathbf{F} \begin{bmatrix} \mathbf{y}(k-1) \\ \mathbf{y}(k) \end{bmatrix}, \quad \mathbf{N}(k) = \mathbf{F} \begin{bmatrix} \mathbf{n}(k-1) \\ \mathbf{n}(k) \end{bmatrix}.$$

文献[11]表明, 该近似具有较小的误差。尽管该模型近似并不能减少模型参数, 但却可以使用在谱修正处理框架中常用的重叠相加方法来实现信号合成, 而不再需要圆周卷积到线性卷积的转换。同时,

还可以在对 $\begin{bmatrix} \mathbf{y}(k-1) \\ \mathbf{y}(k) \end{bmatrix}$ 和 $\mathbf{s}_l(kB)$ 进行谱分析前加上适

当的窗函数, 用来更好地折中谱估计中的分辨率和旁瓣。如果将 $\mathbf{S}_l(k)$ 看作回归变量, 则可以直接使用回归分析技术。

2.2 逐级回归

对于式(6)中的回归模型, 由于 $\mathbf{n}(k)$ 是由近端语音和背景噪声共同组成的, 可以近似看作独立同分布的随机向量, 只是概率密度函数尾部比高斯分布更重。依据 Gauss-Markov 定理, 此时最优的线性无偏估计子仍为最小二乘估计。无偏意味着对双讲鲁棒, 在同样长度数据上具有最小的方差则意味最快的收敛速度。虽然频域逐级回归的求解与直接求最小二乘解存在一定差异, 但是可以预期, 其在收敛速度上相对于随机梯度法仍具有优势。第3节中将通过实验来评估本文方法在实际环境下的表现。下面给出频域逐级回归的算法过程。

将 $\mathbf{Y}(k)$, $\mathbf{S}_l(k)$ 和 $\mathbf{N}(k)$, $i = 1, 2, \dots, 2B$ 的第 i 个元素分别表示为 $Y(i, k)$, $S_l(i, k)$ 和 $N(i, k)$ 。对第 k 个块, 其第 i 个频谱分量 $S_l(i, k)$ 的自谱可以通过指数加权的方法递归估计,

$$\widehat{\Phi}_{S_l S_l}(i, k) = \alpha \widehat{\Phi}_{S_l S_l}(i, k-1) + (1-\alpha) S_l(i, k) S_l^*(i, k) \quad (8)$$

其中 $(\cdot)^*$ 表示复共轭, α 是平滑系数, 与平均时用到的独立帧数有关。本文针对语音应用, 设置 $\alpha = 0.98^{8000B/(128f_s)}$, 其中 f_s 为采样率。因为 $S_l(i, k) = S_{l-1}(i, k-1)$, $l = 2, 3, \dots, L$, 则由

$$\widehat{\Phi}_{S_l S_l}(i, k) = \widehat{\Phi}_{S_{l-1} S_{l-1}}(i, k-1), \quad l = 2, 3, \dots, L \quad (9)$$

将 m 级回归后的残差表示为 $V_m(i, k)$, $m=0, 1, \dots, L$, 其中 $V_0(i, k) = Y(i, k)$ 。可以假定当 $L > Q$ 时, $S_l(i, k)$ 与 $V_m(i, k)$ 的相关性比 $S_{l-1}(i, k)$ 更强, 其中 Q 由两通道之间已知的时延决定。简化起见, 本文假定远端信号与麦克风信号之间的时延已对齐到不超过一个块长度, 则进入逐级回归的回归变量的顺序从前到后可以指定为 $S_1(i, k), \dots, S_L(i, k)$ 。 $S_m(i, k)$ 和 $V_m(i, k)$ 之间的互相关用下式估计

$$\begin{aligned} \widehat{\Phi}_{V_m S_{m+1}}(i, k) &= \alpha \widehat{\Phi}_{V_m S_{m+1}}(i, k-1) \\ &+ (1-\alpha) V_m(i, k) S_{m+1}^*(i, k), \\ m &= 0, 1, \dots, L-1 \end{aligned} \quad (10)$$

第 $m+1$ ($m = 0, 1, \dots, L-1$) 级一元回归的系数由式(11)给出

$$\widehat{H}_{m+1}(i, k) = \frac{\widehat{\Phi}_{V_m S_{m+1}}(i, k)}{\widehat{\Phi}_{S_{m+1} S_{m+1}}(i, k)} \quad (11)$$

相应的残差为

$$V_{m+1}(i, k) = V_m(i, k) - \widehat{H}_{m+1}(i, k) S_{m+1}(i, k) \quad (12)$$

在 L 级回归后, 对回声和近端信号的最终估计分别为 $\widehat{E}_L(i, k) = \sum_{m=0}^{L-1} \widehat{H}_{m+1}(i, k) S_{m+1}(i, k)$ 和 $V_L(i, k)$ 。

2.3 由非平稳程度控制的自适应平滑

逐级回归方法收敛较快, 但同时也可能出现快速发散。近端短时扰动(双讲)带来的影响可以通过设置足够大的 α 来缓解, 但持续的扰动则可能导致 $\widehat{H}_l(i, k)$ 在仅有背景噪声而没有回声的情况下有较大方差, 这是由于当回声再次出现时, 用于估计 $\widehat{H}_l(i, k)$ 的样本过少, 会导致对近端信号的过拟合, 并造成失真。因而, 算法需要检测远端回声信号是否存在。假定远端信号非平稳而近端背景噪声平稳, 则可以引入由麦克风信号非平稳程度控制的平滑因子 $\alpha(i, k)$, 用来在麦克风信号较平稳时减慢统计量估计的更新速度, 保持 $\widehat{H}_l(i, k)$ 缓变, 而不至于快速发散。

在判定 $Y(i, k)$ 非平稳时, 应设置 $\alpha(i, k)$ 为第2.1节中的常规取值, 但当 $Y(i, k)$ 近似平稳时, 则应设置 $\alpha(i, k)$ 更接近于1。在单通道语音增强中, 有很多估计准平稳背景噪声幅度谱均值的方法, 本文采用文献[19]中较简洁的方法, 并类似地估计幅度谱方差。将估计的均值和方差表示为 $\widehat{\mu}_N(i, k)$ 和 $\widehat{\sigma}_N^2(i, k)$, 本文提出的时频依赖的平滑因子 $\alpha(i, k)$ 如下:

$$\alpha(i, k) = \alpha \min \left[\max \left[\frac{Y(i, k) - \widehat{\mu}_N(i, k) + b \sigma_N(i, k)}{(a+b) \sigma_N(i, k)}, 0 \right], 1 \right] \quad (13)$$

其中 a 和 b 与准平稳背景噪声的幅度谱概率分布上、下分位点有关。对于近似高斯的噪声, 本文中将其分别设置为6和3。

2.4 信号重建和过减

与语音增强中常见的谱增强框架类似，信号重建可采用 $V_L(i, k)$ 的重叠相加，考虑到通过精确的相位恢复有利于减小近端语音失真，这里用复谱减代替谱幅度减。复谱减也不存在传统幅度谱减法^[1]所常见的方差较大和音乐噪声问题，因而不需进行额外的平滑。但是，如果估计的 $\hat{H}_L(i, k)$ 与真实值相差太大，复谱减可能会增大麦克风信号。这个问题可以通过对 $V_L(i, k)$ 引入一个幅度约束来规避。

$$\tilde{V}_L(i, k) = \min \{V_L(i, k), cY(i, k)\} \quad (14)$$

其中 c 是一个接近 1 的常量。

由于本文的方法具有快速收敛的特性，可以不必采用基于相干函数估计的后处理。但是在混响环境中，实际回声路径的模型阶数较高，所以可能存在建模不足问题。而且，第 2.1 节中的模型近似会带来一定偏差，可能导致一定残余回声。为进一步抑制回声，可以考虑采用幅度谱过减，并通过在时间维度上加入平滑处理来消除过减带来的音乐噪声。然而，这种处理会在近端语音的起始时刻引入失真。为此，本文不采用时间维度上的平滑，而利用随机的幅度谱最小值约束，不仅避免了音乐噪声，而且保持了语音起始端的音质。随机幅度谱最小值约束不需引入额外的随机噪声添加模块，并能获得更自然的背景噪声。在此约束下，近端语音的过减估计为

$$\tilde{V}_L(i, k) = \max \left\{ \left| \tilde{V}_L(i, k) \right| - d \left| \hat{E}_L(i, k) \right|, \hat{\mu}_N(i, k) + \beta \sigma_N(i, k) \right\} e^{\sqrt{-1} \tilde{V}_L(i, k)} \quad (15)$$

其中 $\beta \sim N(0, 1)$ ，是标准正态分布， d 是幅度谱过减因子。此处之所以不采用卡方分布等非对称分布，是因为 β 只会影响算法输出的背景噪声特性，而对听觉的影响很小。最终，近端信号为式(16)的重叠相加：

$$\hat{n}(k) = \mathbf{F}^{-1} \left[\tilde{V}_L(1, k) \cdots \tilde{V}_L(2B, k) \right]^T \quad (16)$$

采用上述构架，可以在需要时方便的加入背景噪声抑制处理，本文在此不作进一步讨论。

3 实验和分析

根据引言分析，相对于传统方法，本文提出的估计子具有更快的收敛速度以及对双讲和回声路径突变的相对鲁棒性。并且，由于近端持续存在的背景噪声对系统辨识的方差下界的制约，可以预期模型近似和逐级回归的偏差在近端存在一定水平噪声时表现不明显。实验中，对本文提出的方法和 PBFDAF 结合自适应控制方法的一个公开实现

Speex^[2,3,18]，在不同的回声和本地平稳噪声比值的条件下进行了对比测试。

实验的回声路径由在一个普通会议室中实测获得，并截短为 2048 点，信号采样率为 16 kHz。在每个测试样例中，远端信号 $s(t)$ 是从 60 s 男女混合的 TIMIT 语音中随机抽取的 10 s 片段。近端语音为 10 s 的汉语语音，近端背景噪声是一个恒定方差的零均值高斯白噪声序列激励 AR(1) 系统 $1/(1 - 0.9z^{-1})$ 的输出。回声路径在 10 s 结束的时刻正负号反转，然后远端和近端语音重播一次，这样得到一个总长 20 s 的测试音频。

回声衰减量(Echo Return Loss Enhancement: ERLE)在短时段上定义为 $ERLE(k) = \frac{\mathbf{y}^T(k)\mathbf{y}(k)}{\hat{\mathbf{n}}^T(k)\hat{\mathbf{n}}(k)}$ ，对

数谱距离(Log Spectral Distance, LSD)如文献[20]中所定义，客观评价指标采用短时客观可懂度(Short-Time Objective Intelligibility, STOI)。在所有实验中，Speex 的块长和本文算法对应的 B 均设置为 256，Speex 中的滤波器长度和本文算法相应的参数 M 设为 2048。为对比公平起见，没有使用第 2.4 节中的过减，即 $d = 0$ 。另外，随机幅度最小值中 $\beta = 0.9$ 。本文算法采用汉宁(Hanning)窗，并对 50% 和 25% 帧移的情况分别进行考察。

算法的瞬态特性通过短时段上的 ERLE 和 LSD 来考察。回声和本地平稳噪声比(Echo-to-Noise-Ratio, ENR)为 10 dB 和 20 dB 时的情形分别如图 1，图 2 所示。为更清晰地进行对比，图中纵轴表示本文算法相对 Speex 的 ERLE 提高量。在两种情况下，本文方法的收敛速度都更快，并在 0~4 s 和 10~14 s 的时间段上取得了更高的 ERLE。当 ENR=20 dB，即近端背景噪声相对回声较弱时，Speex 在部分时段取得相对本文算法稍高的 ERLE。但是，当近端信号中包含一定程度的背景噪声时，如 ENR=10 dB 时，本文算法在几乎整个 20 s 的时间上都取得了更高的 ERLE。由此可见，本文算法非常适合于以语音为主要回声信号的室内混响环境。另外，本文算法 50% 和 25% 帧移的表现较为接近，但 25% 帧移的表现相对更好。

表 1~表 4 显示的结果由 10 次随机抽取音频后的测试指标平均得到，以综合考察非平稳的近端语音和回声信号在不同的时频重叠情况下的算法表现。ERLE 在整个时间段上计算。从表 1，表 2 中可以看出，本文的方法在各种情况下均取得了更大的回声衰减量，而表 3 和表 4 显示近端语音损伤没有增大，可懂度并没有受到太大影响。同时，25% 帧移的情况性能表现总体更好，但计算量更大。

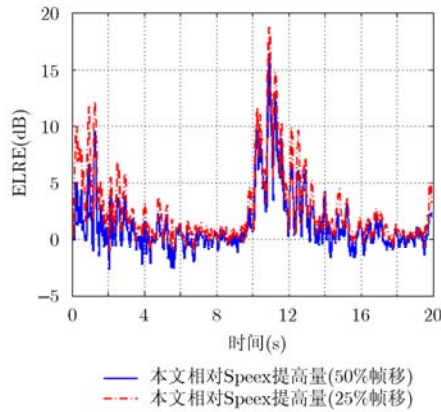


图 1 10 dB 回声噪声比时, 单讲情况下的短时 ERLE 对比测试

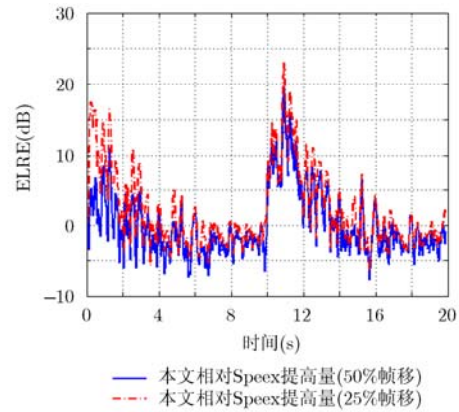


图 2 20 dB 回声噪声比时, 单讲情况下的短时 ERLE 对比测试

由于常见的室内远讲或免提语音通信和人机交互应用环境中, 通常存在一定程度的背景噪声以及不可避免的电路噪声, 且声学环境可能存在各种因素引起的扰动或者突变, 本文方法相对于传统方法会在总体回声抑制量上表现出显著优势。另外在非正式的主观测听中, 本文注意到由于本文方法较快的收敛速度和自适应的随机幅度谱最小值约束, 残余回声通常较白, 因而也相对不容易被听觉感知。

表 1 “单讲”时的 ERLE (dB)

回声噪声比(dB)	Speex	50%帧移	25%帧移
-5	0.50	1.09	1.36
0	1.44	2.56	3.01
10	4.03	7.33	8.22
20	4.82	9.89	11.28
30	4.94	10.48	12.04

表 2 “双讲”时的 ERLE(dB)

回声噪声比(dB)	Speex	50%帧移	25%帧移
-5	0.22	0.98	0.17
0	0.67	1.80	2.09
10	1.31	2.90	3.20
20	1.42	3.15	3.45
30	1.43	3.20	3.51

表 3 “双讲”时的 LSD(dB)

回声噪声比(dB)	Speex	50%帧移	25%帧移
-5	26.94	26.70	26.51
0	22.70	22.41	22.23
10	15.40	14.97	14.66
20	10.74	10.43	10.01
30	9.39	9.15	8.72

表 4 “双讲”时的 STOI

回声噪声比(dB)	Speex	50%帧移	25%帧移
-5	0.54	0.53	0.53
0	0.64	0.64	0.64
10	0.82	0.83	0.82
20	0.88	0.89	0.89
30	0.89	0.90	0.89

4 结束语

针对声学回声控制应用中, 作为回声的语音信号非白, 以及实际声学环境中常见的回声路径较长且往往不能确保持续稳定的特点, 本文提出了一种将逐级回归分析方法在频域处理框架下应用于声学回声控制问题的算法。分级的加权最小二乘估计确保了算法的快速收敛和抗近端非平稳干扰的稳健特性。同时在不同近端噪声水平下的实验表明, 本文的算法在常见噪声水平下的语音应用中, 能够获得很好的模型近似, 估计偏差较小。与一个公开的 PBFDAF 结合自适应控制的算法实现的对比实验显示, 在存在系统突变的环境下, 本文方法在总体回声抑制量上显示出明显的优势, 同时在双讲时很好地保持了近端语音质量。进一步的研究可以考虑针对具体应用场景的特点, 在建模误差和估计偏差以及方差之间取得更适当的平衡。

参考文献

[1] Sondhi M. An adaptive echo canceller[J]. *Bell System Technical Journal*, 1967, 46(3): 497-511.

[2] Soo J S and Pang K K. Multidelay block frequency domain adaptive filter[J]. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1990, 38(2): 373-376.

[3] Valin J M. On adjusting the learning rate in frequency domain echo cancellation with double-talk[J]. *IEEE*

- Transactions on Audio, Speech, and Language Processing*, 2007, 15(3): 1030-1034.
- [4] Gupta V K, Chandra M, and Sharan S N. Acoustic echo and noise cancellation system for hand-free telecommunication using variable step size algorithms[J]. *Radio Engineering*, 2013, 22(1): 200-207.
- [5] Mayyas K. A variable step-size selective partial update LMS algorithm[J]. *Digital Signal Processing*, 2012, 23(1): 75-85.
- [6] 张琦, 王霞, 王磊, 等. 自适应回波抵消中变步长 NLMS 算法[J]. *数据采集与处理*, 2013, 28(1): 64-68.
Zhang Qi, Wang Xia, Wang Lei, et al. Variable step-size NLMS algorithm in echo cancellation[J]. *Journal of Data Acquisition & Processing*, 2013, 28(1): 64-68.
- [7] Gansler T, Hansson M, Ivarsson C J. A double-talk detector based on coherence[J]. *IEEE Transactions on Communications*, 1996, 44(11): 1421-1427.
- [8] Tashev I J. Coherence based double talk detector with soft decision[C]. *IEEE International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, 2012: 165-168.
- [9] Benesty J, Morgan D, and Cho J H. A new class of doubletalk detectors based on cross-correlation[J]. *IEEE Transactions on Speech and Audio Processing*, 2000, 8(2): 168-172.
- [10] Schuldt C, Lindstrom F, and Claesson I. A delay-based double-talk detector[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(6): 1725-1733.
- [11] Avendano C. Acoustic echo suppression in the STFT domain[C]. *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, New Platz, NY, USA, 2001: 175-178.
- [12] Faller C and Chen Jing-dong. Suppressing acoustic echo in a spectral envelope space[J]. *IEEE Transactions on Speech and Audio Processing*, 2005, 13(5): 1048-1062.
- [13] Wada T S and Juang B H. Enhancement of residual echo for robust acoustic echo cancellation[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(1): 175-189.
- [14] Shrawankar U and Thakare V M. Acoustic echo cancellation postfilter design issues for speech recognition system[J]. *International Journal of Science and Advanced Technology*, 2011, 1(5): 38-43.
- [15] Gustafsson S, Martin R, and Vary P. Combined acoustic echo control and noise reduction for hands-free telephony[J]. *Signal Processing*, 1998, 64(1): 21-32.
- [16] Enzner G, Martin R, and Vary P. Partitioned residual echo power estimation for frequency-domain acoustic echo cancellation and postfiltering[J]. *European Transactions on Telecommunications*, 2002, 13(2): 103-114.
- [17] Draper N R and Smith H. *Applied Regression Analysis*[M]. New York: Wiley Series in Probability and Mathematical Statistics, 1981: 337-341.
- [18] Jonathan Rouach: Ported Speex AEC mdf algorithm from C to Matlab[OL]. <https://github.com/wavesaudio/Speex-AEC-matlab>. 2014.01.
- [19] Doblinger G. Computationally efficient speech enhancement by spectral minima tracking in subbands[C]. *Proceedings of EUROSPEECH*, Madrid, Spain, 1995: 1513-1516.
- [20] Cohen I. Analysis of two-channel generalized sidelobe canceller (GSC) with post-filtering[J]. *IEEE Transactions on Speech and Audio Processing*, 2003, 11(6): 684-699.
- 姜开宇: 男, 1986 年生, 博士生, 研究方向为语音信号处理、阵列信号处理.
- 吴超: 男, 1988 年生, 博士生, 研究方向为语音信号处理.
- 国雁萌: 女, 1976 年生, 副研究员, 研究方向为语音信号处理、传声器阵列、语音识别.
- 付强: 男, 1972 年生, 研究员, 研究方向为语音信号处理、传声器阵列.
- 颜永红: 男, 1967 年生, 研究员, 研究方向为语音识别、语音搜索、机器学习、模式识别.