

一种适用于双讲状态的声学回声消除系统

周文俊, 夏秀渝

(四川大学电子信息学院, 成都 610065)

摘要: 声学回声消除中存在的两端讲话情况会导致回声消除器的性能发生恶化, 而传统的双端语音检测及其控制算法无法有效应对双讲。针对这一问题, 提出一种根据频谱信扰比(近端语音功率谱值和回声功率谱值之比)调整自适应步长的方法。为减少计算量以及降低处理时延, 采用分段频域块最小均方(Partitioned frequency block least mean square, PFBLMS)算法作为自适应滤波算法, 因而系统的自适应步长在频域上进行调整。首先建立频谱信扰比和相干函数之间的关系, 通过计算出的相干函数值得到步长因子, 最后根据步长因子实时调整每个频点的自适应步长。此外, 还结合双滤波器和稀疏控制算法进一步提高系统的鲁棒性和收敛性能。仿真实验表明, 该系统在双讲状态下依然能够保证良好的回声抑制能力, 并能够及时跟踪回声信道的变化。相比于基于归一化互相关函数的双端语音检测算法以及开源项目 Speex 中的回声消除算法, 系统在双讲情况下的回波抑制比(Echo return loss enhancement, ERLE)和语音质量得分(Perceptual evaluation of speech quality, PESQ)均有所提高。

关键词: 回声消除; 分段频域块自适应算法; 步长控制; 频谱信扰比; 相干函数

中图分类号: TN912.35 **文献标志码:** A

An Acoustic Echo Cancellation System for Double-Talk Scenario

ZHOU Wenjun, XIA Xiuyu

(School of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: The double-talk scenario will deteriorate the performance of echo canceller in acoustic echo cancellation, while traditional double-talk detection and other methods of controlling the adaptive step-size cannot effectively deal with it. To solve this problem, a method of adjusting the adaptive step-size according to the spectral signal-to-interference ratio (the ratio of the near-end speech's power spectrum to the echo's power spectrum) is proposed. In order to reduce computational complexity and processing delay, the partitioned frequency block least mean square (PFBLMS) algorithm is used as the adaptive filtering algorithm. So the adaptive step-size is adjusted in the frequency domain. First, the relationship between the spectral signal-to-interference ratio and the coherence function is established. Second, the step-size is obtained through the coherence function. Third, the adaptive step-size of each frequency point is adjusted in real time according to the calculated value. In addition, the dual filter and the sparse control algorithms are combined to further improve robustness and convergence performance of the system. The computer simulation shows that the system can not only guarantee good echo suppression ability in the double-talk scenario, but also track the changes of the echo channel in time. Compared with the double-talk

detection method based on the normalized cross-correlation function and the echo cancellation algorithm in the open source project Speex, the proposed system achieves better echo return loss enhancement (ERLE) and perceptual evaluation of speech quality (PESQ).

Key words: echo cancellation; partitioned frequency block least mean square(PFBLMS) algorithm; step-size control; spectral signal-to-interference ratio; coherence function

引言

随着线上语音通话的快速普及以及网络语音传输 (Voice over internet protocol, VoIP) 技术的不断发展, 如何实现高质量的网络语音通信成为一个急需解决的问题^[1]。在免提式通话的场景中, 声学回声的存在降低了语音通话质量, 严重时还会引发啸叫现象^[2], 极大地影响了用户的听觉体验。

声学回声消除系统中一般采用自适应滤波器来消除回声。回声消除的基本原理如图 1 所示。一般情况下, 由扬声器-环境-麦克风构成的声回授系统可以被建模成一个有限阶数的线性系统 $h(n)$, 来自远端的话音 $x(n)$ 经扬声器播放并通过声回授系统 $h(n)$ 后成为声学回声 $y(n)$, 即

$$y(n) = x(n) * h(n) \quad (1)$$

式中“*”为线性卷积。回声 $y(n)$ 与近端语音 $v(n)$ 混合后一同被麦克风采集, 得到自适应滤波器的期望信号, 也即近端麦克风信号

$$d(n) = v(n) + y(n) \quad (2)$$

回声消除的目的即是去除麦克风信号 $d(n)$ 中的回声 $y(n)$, 以此避免给远端用户造成语音干扰。 $\hat{y}(n)$ 为自适应滤波器的输出信号

$$\hat{y}(n) = x(n) * w(n) \quad (3)$$

误差信号 $e(n)$ 为 $d(n)$ 与 $\hat{y}(n)$ 之间的差值, 同时也作为经过回声消除处理后的信号传至远端

$$e(n) = d(n) - \hat{y}(n) = v(n) + y(n) - \hat{y}(n) \quad (4)$$

若不考虑近端语音 $v(n)$ 的影响, 自适应滤波器将根据误差信号 $e(n)$ 的大小不断调整自身权系数 $w(n)$ 。当 $w(n)$ 与 $h(n)$ 十分接近时, 自适应滤波器的输出 $\hat{y}(n) \approx y(n)$, 此时去远端信号 $e(n) \approx v(n)$, 这样回声就被很好地消除了。然而在实际应用中, 近端干扰是不可避免的, 因此需要采用步长控制等方法调节滤波器的迭代更新, 从而避免权值的发散, 并提高收敛性能。

常用的提高双讲状态下回声消除器性能的方法是采用双端语音检测。双端语音检测有能量比较法^[3]、基于信号自相关^[4]或互相关^[5-7]的检测方法、基于过零率的检测方法^[8]等。当双讲检测器判断处于双讲时刻时, 则冻结滤波器的权系数 (即令步长为零), 其他时刻进行正常的自适应迭代。但双端判决算法的准确性和鲁棒性并不理想^[9]: 首先, 双讲时刻往往通过设定一个固定阈值来进行判定, 而环境和噪声干扰往往是不断变化的; 其次, 完全冻结权系数导致了自适应滤波器的缓慢收敛。单独采用双滤波器^[10]的方式虽然可以在一定程度上防止自适应滤波器在双讲时刻的发散, 但稳态误差大, 收敛速度慢。还有一类步长控制方法以滤波器失调随迭代次数的增加而减小为准则, 通过估计泄漏系数进而确定当前时刻的最佳步长^[11-13]。这类方法鲁棒性较强, 能比较好地应对双讲情况, 但算法复杂度较高, 且

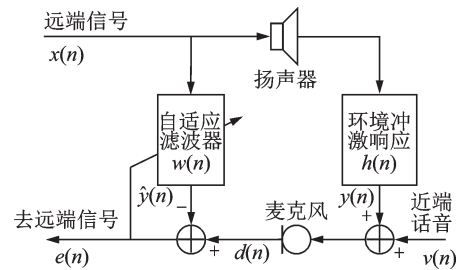


图 1 回声消除的基本原理

Fig.1 Fundamental principles of echo cancellation

收敛后的稳态误差仍然较大。本文提出一种根据频谱信扰比(近端话音功率谱值与回声功率谱值之比)确定自适应步长的方法,并结合双滤波器和稀疏控制算法进一步提高回声消除器的鲁棒性和收敛性能,实验表明该算法在回波抑制比、收敛速度及输出音频质量等方面均有所提升。

1 回声消除系统的具体实现

本文所提回声消除系统的整体框图如图2所示。自适应滤波器采用分段频域块最小均方(Partitioned frequency block least mean square, PFB LMS)算法,步长控制部分主要是为了防止双讲时刻滤波器权值的发散,稀疏控制算法则增加了滤波器的收敛速度,同时减小了稳态误差。双滤波器结构进一步增强了系统在双讲状态下的鲁棒性,辅助滤波器和自适应滤波器之间的系数传递由控制逻辑进行控制。

1.1 PFB LMS 算法

PFB LMS算法通过对信号分块划分然后再进行频域处理的方式,克服了归一化最小均方(Normalized least mean square, NLMS)算法计算复杂度高和频域块最小均方(Frequency block least mean square, FBLMS)算法处理延迟大的缺点^[14-15]。该算法首先在时域上对滤波器输入信号 $x(n)$ 和滤波器权系数 $w(n)$ 做分块处理。设每个分块的权值个数为 N ,总的分块数为 M ,则第 l 次迭代时第 p 个分块的输入信号向量和滤波器权向量分别为

$$x_p(l) = (x((l-p-1)N), \dots, x((l-p)N-1))^T \quad 0 \leq p \leq M-1 \quad (5)$$

$$w_p(l) = (w(l, pN), \dots, w(l, (p+1)N-1))^T \quad 0 \leq p \leq M-1 \quad (6)$$

为减小计算复杂度,自适应迭代在频域上进行,因此首先将滤波器输入信号和滤波器系数都变换到频域。根据重叠保留法,首先将 $x_p(l)$ 向前延伸 N 个样点,将 $w_p(l)$ 向后补 N 个零,然后再分别对延伸后的向量 $\tilde{x}_p(l)$ 和补零后的向量 $\tilde{w}_p(l)$ 做快速傅里叶变换(Fast Fourier transform, FFT),得到频域输入信号向量 $\tilde{X}_p(l)$ 及频域滤波器权向量 $\tilde{W}_p(l)$ 。

对于有约束的^[16]PFB LMS算法,其权系数更新公式为

$$\tilde{W}_p(l+1) = \tilde{W}_p(l) + \mu_0 F \begin{bmatrix} I_{N \times N} & 0_{N \times N} \\ 0_{N \times N} & 0_{N \times N} \end{bmatrix} F^{-1} \mathbf{A}(l) \tilde{E}(l) \odot \tilde{X}_p^H(l) \quad 0 \leq p \leq M-1 \quad (7)$$

式中: μ_0 为全局步长,符号“ \odot ”为两个向量点乘,上标“H”为取共轭。 F 和 F^{-1} 分别为大小为 $2N \times 2N$ 的离散傅里叶变换(Discrete Fourier transform, DFT)矩阵和离散傅里叶逆变换(Inverse discrete Fourier transform, IDFT)矩阵, $\mathbf{A}(l)$ 为大小为 $2N \times 2N$ 的归一化子带步长矩阵

$$\mathbf{A}(l) = \begin{bmatrix} \sigma^{-2}(l, 0) & 0 & \dots & 0 \\ 0 & \sigma^{-2}(l, 1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^{-2}(l, 2N-1) \end{bmatrix} \quad (8)$$

式中 $\sigma^2(l, k)$ 为信号 $x(n)$ 在第 k 个频点处的功率。 $\sigma^2(l, k)$ 的值采用求递归平均的方式进行估计

$$\sigma^2(l, k) = \gamma \sigma^2(l-1, k) + (1-\gamma) |\tilde{X}_0(l, k)|^2 \quad 0 \leq k \leq 2N-1 \quad (9)$$

式中: γ 为平滑系数,取值范围为 $(0, 1)$; $\tilde{X}_0(l, k)$ 为向量 $\tilde{X}_0(l)$ 中的第 k 个元素。

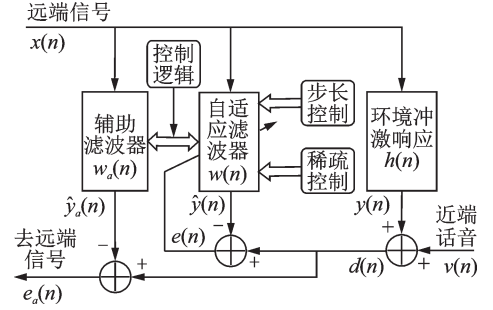


图2 回声消除系统的整体框图

Fig.2 Block diagram of echo cancellation system

$\tilde{E}(l)$ 为第 l 次迭代时的频域误差信号,它采用如下的方式进行计算。

$$\hat{y}(l) = \begin{bmatrix} 0_{N \times N} & 0_{N \times N} \\ 0_{N \times N} & I_{N \times N} \end{bmatrix} F^{-1} \sum_{p=0}^{M-1} \tilde{W}_p(l) \odot \tilde{X}_p(l) \quad (10)$$

$$\tilde{d}(l) = \begin{bmatrix} 0_{N \times 1} \\ d(l) \end{bmatrix} \quad d(l) = [d(lN - N), \dots, d(lN - 1)]^T \quad (11)$$

$$\tilde{e}(l) = \tilde{d}(l) - \hat{y}(l) \quad (12)$$

$$\tilde{E}(l) = F\tilde{e}(l) \quad (13)$$

1.2 步长控制

自适应滤波器对环境冲激响应的建模过程中,将远端语音和回声作为有用信号,而近端语音相当于噪声干扰,会导致滤波器的权值发散。本文实现步长控制的主要思路为:根据频谱信扰比的大小来实时调整自适应滤波器每个频点的迭代步长。本文推导了相干函数与频谱信扰比的关系式,给出依据实时估计的两种相干函数调整自适应滤波器步长的方法。

对于整个回声消除系统而言,近端语音 $v(n)$ 是有用信号而回声 $y(n)$ 为干扰信号,因此可定义频谱信扰比为

$$\xi(k) = \frac{P_{vv}(k)}{P_{yy}(k)} \quad (14)$$

式中 $P_{vv}(k)$ 和 $P_{yy}(k)$ 分别为 $v(n)$ 和 $y(n)$ 在第 k 个频点处的自功率谱值。

远端信号 $x(n)$ 和近端麦克风信号 $d(n)$ 的相干函数定义为^[17]

$$C_{xd}^2(k) = \frac{|P_{xd}(k)|^2}{P_{xx}(k)P_{dd}(k)} \quad (15)$$

式中 $P_{xd}(k)$ 为 $x(n)$ 和 $d(n)$ 的互功率谱, $P_{xx}(k)$ 和 $P_{dd}(k)$ 分别为 $v(n)$ 和 $y(n)$ 的自功率谱。假设近端信号 $v(n)$ 和远端信号 $x(n)$ 不相关, $h(n)$ 是一个线性系统,而 $y(n)$ 为 $x(n)$ 通过系统 $h(n)$ 后的输出值,则有^[18]

$$P_{dd}(k) = P_{vv}(k) + P_{yy}(k) \quad (16)$$

$$P_{xd}(k) = P_{xy}(k) = P_{xx}(k)H(k) \quad (17)$$

$$C_{xd}^2(k) = \frac{P_{yy}(k)}{P_{vv}(k) + P_{yy}(k)} = \frac{1}{1 + \xi(k)} \quad (18)$$

进一步假设自适应滤波器已经接近收敛,则 $e(n) \approx v(n)$,那么 $e(n)$ 和 $d(n)$ 的相干函数为

$$C_{ed}^2(k) = \frac{|P_{ed}(k)|^2}{P_{ee}(k)P_{dd}(k)} \approx \frac{P_{vv}(k)}{P_{vv}(k) + P_{yy}(k)} = \frac{\xi(k)}{1 + \xi(k)} \quad (19)$$

相干函数值的估计依赖于自功率谱 $P_{xx}(k)$ 、 $P_{dd}(k)$ 和 $P_{ee}(k)$ 以及互功率谱 $P_{xd}(k)$ 和 $P_{ed}(k)$,功率谱值采用求递归平均的方式进行计算。估算出各个功率谱后,即可根据定义式求出相干函数的估计值 $\hat{C}_{xd}^2(l, k)$ 与 $\hat{C}_{ed}^2(l, k)$,具体估算方法及误差分析见文献[19]。

由式(14)可知,频谱信扰比的比值越大,表明该频点上近端信号功率相对于回声信号功率的比例越大,这时要尽量减小该频点的步长,以防止滤波器的发散;而当频谱信扰比较小时,则可以增加该频点的步长,以加快滤波器的收敛。

由式(18、19)可以看出,相干函数 $C_{xd}^2(k)$ 及 $C_{ed}^2(k)$ 都能反映频谱信扰比的大小。 $C_{xd}^2(k)$ 的值越大,表明信扰比越小;而 $C_{ed}^2(k)$ 的值越大,表明信扰比越大。当频谱信扰比在 $[0, \infty)$ 变化时,相应的 $C_{xd}^2(k)$ 和 $C_{ed}^2(k)$ 的值正好在 $0 \sim 1$ 范围内变化,因此可以将这两个数值作为步长控制的参考依据。但 $C_{xd}^2(k)$

的估计值受信道时延估计准确性的影响很大,如果延时估计不准,则计算出的相干系数值将偏离真实值,从而导致步长的估计也受到影响;而 $C_{ed}^2(k)$ 的估计值虽然不需要考虑信道时延,但它假设了自适应滤波器接近收敛的条件,并且在滤波器系数迭代的初始阶段有 $e(n)=d(n)$,这时两个信号全相干,对应的步长为0,滤波器系数将始终冻结。

经过上面的分析,单独采用 $\hat{C}_{xd}^2(k)$ 或 $\hat{C}_{ed}^2(l, k)$ 作为步长控制参数都无法达到令人满意的效果。因此本文考虑结合两个相干函数估计值 $\hat{C}_{xd}^2(l, k)$ 和 $\hat{C}_{ed}^2(l, k)$,提出的可变步长为

$$\mu(l, k) = (1 - a)\hat{C}_{xd}^2(l, k) + a(1 - \hat{C}_{ed}^2(l, k)) \quad (20)$$

式中 a 为一个取值范围为 $(0, 1)$ 的参数,它决定了两个相干系数估计值在步长控制中所占的比例。 a 的最佳取值与信道时延估计的准确性以及滤波器分块长度等有关,通过大量实验测试,本文得出较为理想的取值为0.3。将可变步长 $\mu(l, k)$ 写成矩阵的形式,有

$$\mu(l) = \text{diag}(\mu(l, 0), \mu(l, 1), \dots, \mu(l, 2N - 1)) \quad (21)$$

引入步长控制后,PFBLMS算法的权系数更新公式为

$$\tilde{\mathbf{W}}_p(l+1) = \tilde{\mathbf{W}}_p(l) + \mu_0 F \begin{bmatrix} I_{N \times N} & 0_{N \times N} \\ 0_{N \times N} & 0_{N \times N} \end{bmatrix} F^{-1} \mu(l) \mathbf{\Lambda}(l) \tilde{\mathbf{E}}(l) \odot \tilde{\mathbf{X}}_p^*(l) \quad 0 \leq p \leq M - 1 \quad (22)$$

1.3 稀疏控制算法和双滤波器

在大多数回声消除的应用场景中,回声通道是稀疏的,即冲激响应系数存在很多零值或者系数值十分分散。当回声通道稀疏性很强时,传统的自适应滤波器收敛性能会大大降低。因此引入稀疏控制算法的主要目的是提高滤波器的收敛速度,并且减小滤波器收敛后的稳态误差^[20-21]。

本文采用成比例的稀疏控制算法,即简单地将每一块滤波器系数的总能量与全部分块滤波器系数的平均总能量的比值作为该块滤波器的稀疏控制系数。第 p 个分块的稀疏控制系数为

$$Q_p(l) = M \frac{\|\tilde{\mathbf{W}}_p(l)\|_2^2}{\sum_{p=0}^{M-1} \|\tilde{\mathbf{W}}_p(l)\|_2^2} \quad 0 \leq p \leq M - 1 \quad (23)$$

式中“ $\|\cdot\|_2$ ”表示求向量的二范数。可见,若某个滤波器分块的总能量越大,则相应的稀疏控制系数也越大,该分块系数的收敛速度将更快。引入稀疏控制算法后,PFBLMS算法的权系数更新公式为

$$\tilde{\mathbf{W}}_p(l+1) = \tilde{\mathbf{W}}_p(l) + \mu_0 Q_p(l) F \begin{bmatrix} I_{N \times N} & 0_{N \times N} \\ 0_{N \times N} & 0_{N \times N} \end{bmatrix} F^{-1} \mu(l) \mathbf{\Lambda}(l) \tilde{\mathbf{E}}(l) \odot \tilde{\mathbf{X}}_p^*(l) \quad 0 \leq p \leq M - 1 \quad (24)$$

对于双滤波器结构,回声消除系统中含有两个滤波器:主滤波器和辅助滤波器。其中主滤波器按照 PFBLMS算法作自适应调整,但其误差信号不作为回声消除系统的输出;而辅助滤波器则不作自适应调整,但误差信号会作为系统的输出。主滤波器和辅助滤波器系数的相互复制由一定控制逻辑进行控制,即算法在每次迭代后都会比较两个滤波器对回声抑制能力的大小。如果主滤波器的抑制能力更强,那么将把主滤波器的权系数复制给辅助滤波器,如果辅助滤波器的抑制能力更强,则会将辅助滤波器的权系数复制回主滤波器。两个滤波器对回声抑制能力的高低则主要根据两个滤波器的输出误差信号的能量大小进行判断^[22]。

2 仿真与分析

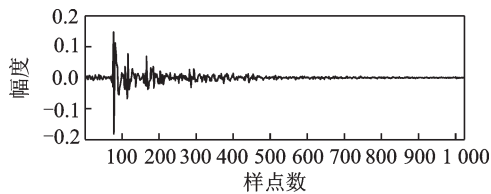
将本文算法的实验结果与文献[5]中基于归一化互相关函数的双端语音检测算法以及开源项目 Speex^[23]中的回声消除算法做对比。并采用了归一化失调(Normalized misalignment, NMIS)和回波抑制比(Echo return loss enhancement, ERLE)两个参数作为衡量回声消除效果的指标,两个指标分别定义为

$$\text{NMIS} = 10 \lg \frac{\|\mathbf{w} - \mathbf{h}\|_2^2}{\|\mathbf{h}\|_2^2} \quad (25)$$

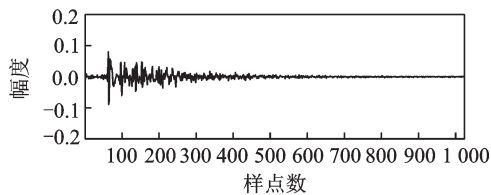
$$\text{ERLE} = 10 \lg \frac{E\{y^2(n)\}}{E\{[e(n) - v(n)]^2\}} \quad (26)$$

归一化失调值越小,表明自适应滤波器的系数越接近实际系统的冲激响应,即建模精度越高。回波抑制比的值越大,表明对回声的抑制程度越大。

实验 1 远端输入是均值为 0、方差为 1 的高斯白噪声,并在近端添加平均功率相对回声大小为 -20 dB 的高斯白噪声作为背景噪声。信号采样率设为 8 kHz。声回授系统的冲激响应由软件生成,冲激响应的长度为 1 024 点,随机选取的两个冲激响应如图 3 所示。高斯白噪声输入下的归一化失调和回波抑制比曲线如图 4 所示。假定冲激响应在 10 s 处发生突变,即由系统 1 转变为系统 2。

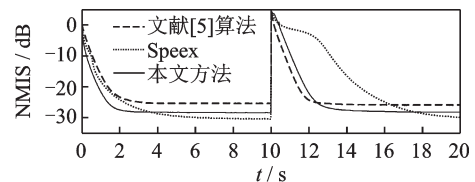


(a) Impulse response of system 1

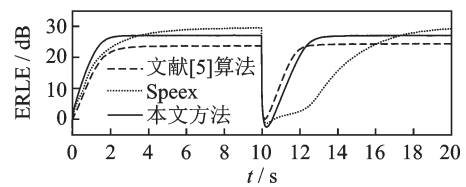


(b) Impulse response of system 2

图 3 声回授系统的冲激响应



(a) Normalized misalignment



(b) Echo return loss enhancement

图 4 白噪声输入下的归一化失调和回波抑制比

Fig.3 Impulse response of the acoustic feedback system Fig.4 NMIS and ERLE in the case of white noise input

从图 4 中的曲线可以看出,在滤波器迭代的初始阶段,本文算法的收敛速度最快,但当到达收敛以后,Speex 算法的稳态误差更小,本文算法的稳态误差介于两种算法之间。当环境发生突变以后,Speex 算法的收敛速度明显下降,无法及时恢复到收敛状态。综合来看,在白噪声输入的情况下,本文算法的稳态误差和收敛速度都能达到较好的水平。

实验 2 远端语音选用一段长 20 s 的语音,近端语音在第 5~9 s 为有话段,并在全部时刻叠加平均功率相对回声大小为 -15 dB 的高斯白噪声作为背景噪声,将近端语音(包含添加的背景噪声)和回声的混合信号作为麦克风采集信号。以上所有信号的采样率均为 8 kHz。远端语音、麦克风语音及近端语音的时域波形如图 5 所示。回声通道的冲激响应与实验 1 中相同,但此时设定为在第 12 s 处发生突变。

实验 2 中输出语音的归一化失调值和回波抑制比随时间的变化曲线如图 6 所示。从归一化失调曲线可见,3 种方法均能在不同程度上防止自适应滤波器系数在双讲时刻的发散,但无论是处于双讲时刻还是非双讲时刻,本文方法的滤波器失调都远低于另外两种方法,说明本文方法的建模精度更高。

从回波抑制比曲线可以看出,文献[5]中的方法虽然在双讲状态下性能没有发生恶化,但在自适应

迭代的初始阶段收敛速度很慢,滤波器接近收敛后的回声消除性能也无法达到满意的效果,稳态误差较大。Speex中的算法收敛性能有了很大的提高,但双讲时刻下的回波抑制比有着较为明显的下降。虽然本文方法的回波抑制比在双讲时刻也有所下降,但相比于文献[5]的方法高出10 dB左右,相比于Speex的回声消除算法也能高出5 dB左右。

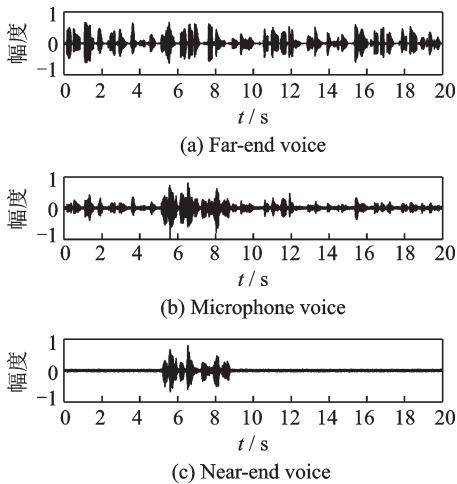


图5 远端语音、麦克风语音及近端语音的时域波形
Fig.5 Time-domain waveforms of far-end voice, microphone voice, and near-end voice

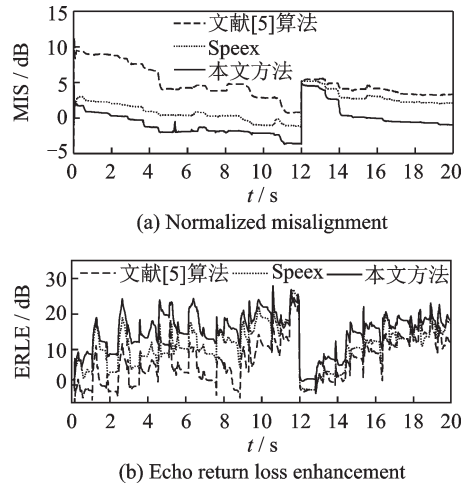


图6 语音输入下的归一化失调和回波抑制比
Fig.6 NMIS and ERLE in the case of voice input

在对回声通道的跟踪性能方面,当环境发生突变后,文献[5]方法的收敛速度明显慢于另外两种方法。且实验表明,该方法在某些情况下还会导致滤波器权系数发生冻结,以至于完全失去对回声通道的跟踪能力。而Speex的回声消除算法虽然能够稳定地跟踪环境变化,但收敛速度仍然较慢,本文提出的回声消除算法则在应对环境突变的情况下依然保证了更好的收敛速度。

图7展示了3种方法的输出信号时域波形图。可见,基于归一化互相关函数的双端检测方法残留回声最大,Speex回声消除算法次之,本文方法的残留回声最小。改变环境冲激响应以及远近端语音等条件,经过多次实验也可证实这一结论。

实验3 语音样本采用ICASSP 2021声学回声消除挑战^[24]中开源的回声消除测试集,测试集中的语音共分为4类:单讲无移动、单讲有移动、双讲无移动和双讲有移动。有无移动指的是在通话过程中麦克风或扬声器的位置是否发生变化。在测试集中随机选取上述每类语音样本各100组,分别采用不同算法进行测试。表1列举了不同算法在4种不同情况下的平均ERLE值,平均ERLE值的计算方式为:先对每组样本基于所有时域采样点计算ERLE均值,然后对100组测试样本的均值ERLE再进行算术平均。

此外,对于双讲情况,本文还采用了语音质量的感知评价

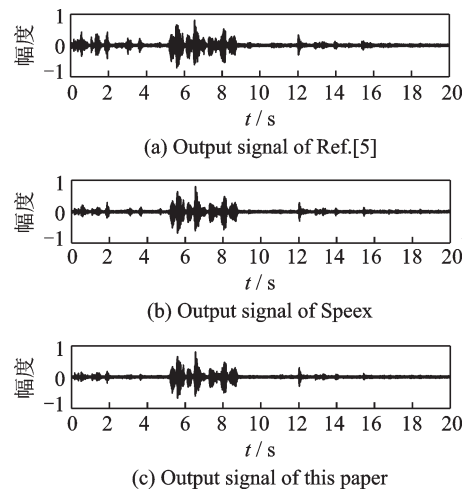


图7 不同算法的输出信号时域波形
Fig.7 Time domain waveforms of output signal for different algorithms

(Perceptual evaluation of speech quality, PESQ)作为评价指标。PESQ是一种用客观评价算法来模拟人类主观听觉感受的评估方法,PESQ分值在 $-0.5\sim 4.5$ 之间,分值越高,表明语音质量越好。不同算法分别在双讲无移动和双讲有移动情况下的平均PESQ值如表2所示。

表1 不同算法的平均ERLE值

算法	平均ERLE值			
	单讲无移动	单讲有移动	双讲无移动	双讲有移动
文献[5]	17.19	13.68	8.14	5.61
Speex	23.12	16.60	9.60	7.35
本文算法	22.35	18.12	11.76	9.45

表2 不同算法的平均PESQ值

算法	平均PESQ值	
	双讲无移动	双讲有移动
文献[5]	3.08	3.00
Speex	3.22	3.06
本文算法	3.35	3.22

从表1数据可以看出,除单讲无移动情况下,本文算法的平均ERLE值略低于Speex的回声消除算法外,其他3种情况均高于另外两种算法。特别是在双端通话的情况下,相比与另外两种算法有着较为明显的提升。从表2数据可以看出,本文算法在两种双讲情况的PESQ值均高于其他算法,说明经过本文算法处理后的语音有着更好的音频质量。

3 结束语

本文提出了一种基于频谱信扰比的步长控制方法,并结合稀疏控制算法以及双滤波器,组成了一种新的回声消除系统。仿真实验表明,该系统的性能高于文献[5]中基于归一化互相关函数的双端语音检测算法以及开源项目Speex中的回声消除算法。算法复杂度方面,所提算法与Speex的回声消除算法相当,但若考虑加入非线性处理,则本文算法更具优势,因为步长控制中估算的相干函数值可以直接用于非线性处理中。

同时本文提出的算法也存在一些有待改进的方面,由于引入了双滤波器,算法的复杂度仍然过高;步长控制中比例系数 a 的取值是固定的,这将导致系统性能在不同的条件下存在一定的差异;经过处理后的近端语音依然存在少量的残留回声。接下来的工作考虑针对这些缺点不断完善和优化所提出的回声消除系统,并结合非线性处理,在尽量减小语音失真的前提下进一步消除残留回声。

参考文献:

- [1] JYOTHIRMAI M, MOUNIKA A, PRATHIMA K, et al. Echo cancellation in voip using block-based adaptive filter[J]. *Research Journal of Engineering & Technology*, 2018, 9(2): 195-200.
- [2] 何培宇. 面向实时处理的自适应声学语音信号处理[D]. 成都:四川大学, 2005.
HE Peiyu. Adaptive acoustic and speech signal processing towards real-time operation[D]. Chengdu: Sichuan University, 2005.
- [3] CHO J H, MORGAN D R, BENESTY J. An objective technique for evaluating doubletalk detectors in acoustic echo cancelers [J]. *IEEE Transactions on Speech and Audio Processing*, 1999, 7(6): 718-724.
- [4] CHOWDHURY D, SARKAR M, HAIDER M Z, et al. Autocorrelation based double talk detection system with an NLMS acoustic echo canceler[C]//*Proceedings of 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. Piscataway: IEEE, 2019: 1-5.
- [5] GANSLER T, BENESTY J. A frequency-domain double-talk detector based on a normalized cross-correlation vector[J]. *Signal Processing*, 2001, 81(8): 1783-1787.
- [6] 张正文,饶鼎,巩朋成,等. 基于信号包络和互相关的双端通话检测算法研究[J]. *现代电子技术*, 2020, 43(3): 22-26.
ZHANG Zhengwen, RAO Ding, GONG Pengcheng, et al. Double-talk detection algorithm based on signal envelope and cross correlation[J]. *Modern Electronics Technique*, 2020, 43(3): 22-26.
- [7] BERIESTY J, MORGAN D R, 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.
- [8] IKRAM M Z. Double-talk detection in acoustic echo cancellers using zero-crossings rate[C]//Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2015: 1121-1125.
- [9] 谢鹏, 刘加. 采用变步长算法的无双端检测回声消除系统[J]. 数据采集与处理, 2010, 25(6): 736-740.
XIE Peng, LIU Jia. Adaptive acoustic echo cancellation system without double-talk detection using vss-nlms algorithm[J]. Journal of Data Acquisition & Processing, 2010, 25(6): 736-740.
- [10] HAMIDIA M, AMROUCHE A. A new structure for acoustic echo cancellation in double-talk scenario using auxiliary filter [C]//Proceedings of 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC). Piscataway: IEEE, 2014: 253-257.
- [11] YANG F, YANG J. Optimal step-size control of the partitioned block frequency-domain adaptive filter[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2017, 65(6): 814-818.
- [12] VALIN J M. On adjusting the learning rate in frequency domain echo cancellation with double-talk[J]. IEEE Transactions on Audio Speech & Language Processing, 2007, 15(3): 1030-1034.
- [13] NITSCH B H. A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain[J]. Signal Processing, 2000, 80(9): 1733-1745.
- [14] FARHANG-BOROJENY B. Adaptive filters: Theory and applications[M]. 2nd ed. New Jersey: John Wiley & Sons, 2013: 267-276.
- [15] SOO J S, PANG K K. Multidelay block frequency domain adaptive filter[J]. IEEE Transactions on Acoustics Speech & Signal Processing, 1990, 38(2): 373-376.
- [16] CHAN K S, FARHANG-BOROJENY B. Analysis of the partitioned frequency-domain block LMS (PFBLMS) algorithm [J]. IEEE Transactions on Signal Processing, 2001, 49(9): 1860-1874.
- [17] 王永德, 王军. 随机信号分析基础[M]. 北京:电子工业出版社, 2013: 84-109.
WANG Yongde, WANG Jun. Fundamentals of random signal analysis[M]. Beijing: Publishing House of Electronics Industry, 2013: 84-109.
- [18] 夏秀渝, 王永德, 何培宇. 相干函数在房间声回授通道特性测量中的应用[J]. 四川大学学报(自然科学版), 2004, 41(3): 542-546.
XIA Xiuyu, WANG Yongde, HE Peiyu. The application of coherence function in the measurement of room acoustic feedback channel characteristics[J]. Journal of Sichuan University (Natural Science Edition), 2004, 41(3): 542-546.
- [19] CARTER G, KNAPP C, NUTTALL A. Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing[J]. IEEE Transactions on Audio & Electroacoustics, 1973, 21(4): 337-344.
- [20] LOGANATHAN P, KHONG A W H, NAYLOR P A. A class of sparseness-controlled algorithms for echo cancellation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2009, 17(8): 1591-1601.
- [21] 刘立刚, MASAHIRO F, 张世永. 一种变步长 Proportionate NLMS 自适应滤波算法及其在网络回声消除中的应用[J]. 电子学报, 2010, 38(4): 973-978.
LIU Ligang, MASAHIRO F, ZHANG Shiyong. A variable step-size proportionate nlms adaptive filtering algorithm and its application in network echo cancellation[J]. Acta Electronica Sinica, 2010, 38(4): 973-978.
- [22] OCHIAI K, ARASEKI T, OGIHARA T. Echo canceler with two echo path models[J]. IEEE Transactions on Communications, 1977, 25(6): 589-595.
- [23] VALIN J M. Speex: A free codec for free speech[EB/OL]. (2016-02-28). <https://arxiv.org/pdf/1602.08668.pdf>.
- [24] SRIDHAR K, CUTLER R, SAABAS A, et al. ICASSP 2021 acoustic echo cancellation challenge: Datasets, testing framework and results[C]//Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway:IEEE, 2020: 151-155.

作者简介:



周文俊(1997-),男,硕士研究生,研究方向:声学回声消除, E-mail: zwjwz@fox-mail.com。



夏秀渝(1970-),通信作者,女,副教授,研究方向:语音增强、语音分离、计算听觉场景分析等, E-mail: xiaxy@163.com。