

低信噪比下基于信号子空间的 语音增强算法研究

王 红,王爱平,邹 海

(安徽大学 计算机科学与技术学院,安徽 合肥 230601)

摘 要: 内嵌式预白化子空间算法与其它经典语音增强算法相比有很好的去噪效果,但是在低信噪比环境下,仍存在大量残留噪声. 针对这一点,本文提出信号子空间优化与维纳滤波方法相结合的改进方法. 仿真结果表明,在几种常见背景噪声的低信噪比下,相比传统的谱减法、维纳滤波法、内嵌式预白化子空间算法,本文算法效果更佳,能有效的抑制背景噪声,改善语音质量,并且保证一定的语音可懂度.

关键词: 内嵌式预白化;子空间优化;维纳滤波;语音质量;可懂度

中图分类号: TN912.35

文献标识码: A

文章编号: 1000-7180(2017)02-0043-05

Speech Enhancement Algorithm based on Signal Subspace with Low Signal-to-Noise Ratio

WANG Hong, WANG Ai-ping, Zou Hai

(College of Computer Science and Technology, Anhui University, Hefei 230601, China)

Abstract: Compared with other classic speech enhancement algorithm, built-in pre-whitening subspace algorithm has a good denoising effect, but in a low signal-to-noise ratio environment, there is still a large amount of residual noise. For this, the method that combining signal subspace optimization and wiener filtering is put forward to improve the speech quality. Simulation results show that under several common background noises with low signal-to-noise ratio, the performance of the proposed approach is more excellent, it can effectively restrain background noise and improve the speech quality, meanwhile, in some extent, guarantee the speech intelligibility, compared with the traditional spectral-subtraction, wiener filtering method, built-in pre-whitening subspace algorithm.

Key words: built-in pre-whitening; subspace optimization; wiener filtering; speech quality; speech-intelligibility

1 引言

生活中充满了噪音,语音增强是为了从带噪语音中恢复出干净语音,改善语音质量,提高语音可懂度. 经典语音增强算法主要有谱减法、维纳滤波法、子空间算法等. 谱减法^[1]会使增强后的语音经常伴随一定节奏感的音乐噪声. 维纳滤波法^[2]可以稍提高语音可懂度,但也会产生音乐噪声. 信号子空间方法可以在整体上达到最佳的语音增强效果,因为它在语音信号失真和残留噪声之间提供了一种控制机

制,可以对二者进行权衡.

信号子空间方法最早是由 Denrinos 提出,他利用奇异值分解 (Singular Value Decomposition, SVD) 对子空间进行划分^[3]. Ephraim 和 Van Trees 提出使用特征值分解 (Eigen Value Decomposition, EVD) 将带噪语音向量空间分解为两个相互正交的子空间,信号子空间 (以纯净语音信号为主,包含少量噪声) 和噪声子空间 (仅包含噪声). 利用 KLT 变换,将噪声子空间置零,从信号子空间中获

优估计器,时域约束估计器(TDC)和频域约束估计器(SDC)^[4].接着 Rezayee 和 Gazor 推广了 Ephraim 和 Van Trees 的子空间算法,提出有色环境下的次优估计方法^[5].之后, Yi Hu 等人提出了一种有色环境下基于内嵌式预白化的子空间算法^[6].文献[7]则提出了一种新的子空间优化准则重构误差方差(Variance of Reconstruction Error, VRE),用于平衡噪声抑制和语音失真.

虽然信号子空间方法已经有着很好的去噪效果,但是噪声覆盖了整个带噪语音的向量空间,信号子空间不可避免残留噪声成分,尤其是在低信噪比环境下语音增强效果仍不尽人意.因此,如何在低信噪比环境下进行语音增强仍然是目前研究的热点和难点.本文提出信号子空间优化与维纳滤波方法相结合的改进方法.

2 内嵌预白化信号子空间语音增强算法

信号子空间语音增强算法将 K 维带噪语音向量空间分解成两个相互正交的子空间,信号子空间和噪声子空间.假设信号满足模型:

$$y(t) = x(t) + n(t) \quad (1)$$

其中, y, x 和 n 分别表示 $K \times 1$ 维带噪语音向量,干净语音向量和噪声向量.在被加性噪声干扰的信号模型中,假设干净语音信号 x 与噪声 n 互不相关且均值为零,则:

$$R_y = R_x + R_n = E[xx^T] + E[nn^T] \quad (2)$$

R_y, R_x 和 R_n 分别是带噪语音,干净语音和噪声的协方差矩阵. R_x 秩为 M ,则有 M 个非零特征值.定义 $\hat{x} = Hy$ 是对干净语音的线性估计, H 是 $K \times K$ 维线性估计器,则估计值和真实值之间误差表示为:

$$e = \hat{x} - x = (H - I) \cdot x + H \cdot n = e_x + e_n \quad (3)$$

这里, e_x 和 e_n 分别表示语音信号的失真和增强后的残留噪声.定义语音失真能量 $\bar{e}_x^2$ 和残留噪声能量 $\bar{e}_n^2$ 分别表示为:

$$\begin{aligned} \bar{e}_x^2 &= E[e_x^T e_x] = \text{tr}(E[e_x e_x^T]) \\ &= \text{tr}((H - I)R_x(H - I)^T) \\ \bar{e}_n^2 &= E[e_n^T e_n] = \text{tr}(E[e_n e_n^T]) = \text{tr}(HR_n H^T) \end{aligned} \quad (4)$$

这里, $E[\cdot]$ 表示均值运算符.

语音失真能量和残留噪声能量不可能同时达到最小,子空间方法一般采用两种线性估计器,时域约

束估计器(TDC)和频域约束估计器(SDC)来实现语音增强.由于频域约束估计器考虑到听觉意义的掩蔽效应,本文将改进的地方应用到频域约束估计器中.这里先介绍一下频域约束估计器(SDC)的估计准则,即:将每一个频谱分量约束在某一范围内的同时,最小化语音失真^[4].内嵌预白化信号子空间方法^[6]中用下式来表达:

$$H_{\text{opt}} = \arg \min_H \text{tr}(E[e_x e_x^T])$$

Subject to:

$$E\{|v_k^T e_n|^2\} \leq \beta_k, k = 1, 2, \dots, M$$

$$E\{|v_k^T e_n|^2\} = 0, k = M + 1, \dots, K \quad (5)$$

这里, $v_k^T e_n$ 表示残留噪声 e_n 在第 k 阶谱分量上的映射, v_k 是接下来介绍的联合矩阵 $\Sigma = R_n^{-1} R_x$ 特征向量矩阵第 k 个列向量, β_k 取值范围在 $(0, 1)$ 之间.

因为对于白噪声而言, $R_n = \delta_n^2 I$ 为对角阵,而有色噪声的协方差矩阵不再是对角阵,所以在文献[6], YiHu 等人提出基于内嵌预白化的子空间方法.由于 R_x 和 R_n 皆为非负定矩阵,根据矩阵理论相关定理知,存在这样一个非奇异矩阵 V ,可同时对角化 R_x 和 R_n ,即:

$$V^T R_x V = \Lambda_x$$

$$V^T R_n V = I \quad (6)$$

这里, V 和 Λ_x 分别是联合矩阵 $\Sigma = R_n^{-1} R_x$ 的特征向量矩阵和特征值矩阵.根据 Kuhn-Tucker 必要条件可得频域约束估计器 H_{opt} 需要满足:

$$H_{\text{opt}} R_x + L H_{\text{opt}} R_n - R_x = 0 \quad (7)$$

其中, $L = V \Lambda_\mu V^T$, Λ_μ 表示拉格朗日乘子对角阵,将式(6)带入式(7),得到如下:

$$Q \Lambda_x + V^T V \Lambda_\mu Q = \Lambda_x \quad (8)$$

这里, $Q = V^T H_{\text{opt}} V^{-T}$,这是控制系统中常见的 Lyapunov 方程,选取 SDC2 解法,假设 Q 和 $V^T V$ 都是对角阵^[4],则最后的最优线性估计器为:

$$H_{\text{opt}} = V^{-T} Q V^T \quad (9)$$

估计的纯净语音为 $\hat{x} = V^{-T} Q V^T \cdot y$

3 信号子空间优化与维纳滤波结合的子空间语音增强算法

信号子空间维度估计在整个语音增强的过程中非常关键.大的特征值分量主要包含语音信息,小的特征值分量主要包含噪声信息.内嵌预白化子空间方法中信号子空间的维度估计依据特征值大于零来确定,对小于零的特征值置零,与之对应的特征向量不变.而本文我们依旧采用特征值大于零来估计

信号子空间维度,但不同是本文将对特征值小于零的特征值进行去除,同时将对应特征值小于零的特征向量也进行去除.我们首先定义投影向量为:

$$VV^T = [V_1 V_2] \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = V_1 V_1^T + V_2 V_2^T \quad (10)$$

其中, V 是矩阵 $\Sigma = R_n^{-1} R_x$ 的特征向量矩阵, $V = [V_1 V_2]$, 其中 $V_1 V_1^T$ 用来将带噪语音向量投影到信号子空间, $V_2 V_2^T$ 用来将带噪语音向量投影到噪声子空间,这种做法在某种程度上达到了分离语音和噪声向量的效果.对信号子空间进行处理,将特征值小于零的去除,与之对应的特征向量也去除.至此,信号子空间维度确定,噪声子空间也被置零.由于去除特征值小于零的特征向量所构成的特征向量矩阵可能不可逆,而最后语音增强处理中需要对特征向量矩阵求逆,我们引入广义逆运算^[8]思想,对于常规不可逆矩阵,通过广义逆运算求其逆矩阵.则最优线性估计器为:

$$H_{\text{opt}} = V_1^{-T} Q V_1^T \quad (11)$$

维纳滤波方法处理语音时能够有效抑制残留噪声,并且 Loizou 等人曾提出除维纳滤波算法对语音可懂度稍有提高外,其他算法都不能有效提高语音可懂度,甚至会降低^[9].我们将频域维纳滤波方法引入到子空间域,让估计的干净语音通过维纳滤波器.已知频域的维纳滤波器传递函数为:

$$W_k = \frac{P_x(k)}{P_x(k) + P_n(k)} \quad (12)$$

其中, $P_x(k)$ 和 $P_n(k)$ 分别是干净语音和噪声的功率谱密度.本文采用频域约束估计器,结合频域约束估计器估计准则以及式(12)维纳滤波传递函数,我们得到每一个谱分量对应的维纳滤波器为:

$$P_k = \frac{E_x^k}{E_x^k + E_n^k} = \frac{\lambda_x^k}{\lambda_x^k + \lambda_n^k} \quad k = 1, 2, \dots, K \quad (13)$$

根据以上思路,改进算法如下:

Step1 根据带噪语音前 120 ms, 得到噪声协方差矩阵、含噪语音协方差矩阵, 计算出联合矩阵 $\Sigma = R_n^{-1} R_y - I$.

Step2 对联合矩阵进行特征值分解, 特征值按从大到小排列, 对应特征向量也进行调整, 特征值和特征向量逐一对应.

Step3 将小于零的特征值去除, 同时去除与之对应的特征向量, 并对最终的特征向量矩阵求广义逆.

Step4 信号子空间维度估计:

$$M = \arg \max_{1 \leq k \leq K} \{\lambda_k > 0\} \quad (14)$$

Step5 让估计的干净语音通过维纳滤波器:

$$\hat{\mathbf{A}}_x = \mathbf{A}_x \cdot P_k = \text{diag}(\lambda_1^k, \lambda_2^k, \dots, \lambda_k^k) \quad (15)$$

Step6 对每一个 μ_k 值进行计算:

$$\mu_k = \begin{cases} \mu_0 - \frac{\lambda_k^k}{s}, & -5 \leq \lambda_k^k \leq 20 \\ 1, & \lambda_k^k \geq 20 \\ 20, & \lambda_k^k \leq -5 \end{cases} \quad (16)$$

Step7 计算增益矩阵和最优频域线性估计器:

$$q_{kk} = \begin{cases} \frac{\lambda_k^k}{\lambda_k^k + \mu_k}, & k = 1, 2, \dots, M \\ 0, & k = M+1, \dots, K \end{cases}$$

$$Q = \text{diag}(q_{11}, q_{22}, \dots, q_{kk})$$

$$H_{\text{opt}} = V_1^{-T} Q V_1^T \quad (17)$$

Step8 最后估计得到的干净语音为:

$$\hat{x} = H_{\text{opt}} \cdot y = V_1^{-T} Q V_1^T \cdot y \quad (18)$$

4 实验结果与分析

为了验证本文算法最终效果, 使用 Matlab 软件仿真实验. 实验采用 NOISEX-92 噪声库中 3 种不同噪声, white 高斯白噪声、Train 火车噪声、Street 街道噪声, 干净语音来自 NOISEUS 库 sp01 ~ sp30 文件和 IEEE 语音库 S_01_01 ~ S_02_10 文件, 每种噪声以 -10 db、-5 db、0 db、5 db 不同信噪比与干净语音合成测试带噪语音. 本实验干净语音信号和噪声信号采样频率均为 8 kHz, 语音分帧帧长取 32 ms, 采用重叠帧技术, 帧间重叠 50%, 加窗采用汉明窗.

语音增强效果对比主要依据 pesq 和 segSNR 的测评值. pesq (Perceptual Evaluation of Speech Quality) 反映增强语音听觉质量, 得分范围一般在 -0.5 ~ 4.5 之间, 分数越高代表语音听觉质量越好. 时域分段信噪比 segSNR (segmental SNR) 反映噪声去除程度. 表 1 为各语音增强算法在白噪声环境下增强效果对比. 从表中可看出白噪声下, 本文方法相比其它三种方法, segSNR 得分有明显提高, 尤其在低信噪比下提升更为明显, 说明本文方法有很好降噪性能. pesq 得分也优于其它三种方法, 说明白噪声下本文方法同时改善了语音的听觉质量.

表 1 白噪声环境下各种语音增强算法效果对比

方法	−10 db		−5 db		0 db		5 db	
	pesq	segSNR	pesq	segSNR	pesq	segSNR	pesq	segSNR
谱减法	0.737 4	−5.794 0	0.830 3	−4.781 4	1.058 3	−4.136 5	1.368 9	−3.706 7
维纳滤波法	0.028 9	−1.494 5	0.735 5	−0.431 9	1.452 4	1.452 9	1.997 7	4.089 1
buit-in 子空间法	0.638 4	−1.814 1	1.313 8	−0.114 3	1.958 3	2.388 9	2.420 6	5.288 3
本文方法	0.758 3	−0.836 5	1.460 3	2.448 2	2.081 3	3.174 0	2.573 3	5.788 3

表 2、表 3 为各语音增强算法在街道噪声和火车噪声环境下语音增强效果对比,可看出,本文算法在 pesq、segSNR 得分上整体上优于其它三种算法,segSNR 得分在低信噪环境下提高明显.表明在街道噪声和火车噪声环境下本文方法有效去除残留噪声的同时,改善了增强语音的听觉质量.

表 2 Street 街道噪声环境下各种语音增强算法效果对比

方法	−10 db		−5 db		0 db		5 db	
	pesq	segSNR	pesq	segSNR	pesq	segSNR	pesq	segSNR
谱减法	0.961 4	−7.720 3	1.345 2	−7.166 6	1.729 3	−6.627 0	2.103 9	−5.611 4
维纳滤波法	0.806 0	−2.157 0	1.366 6	−0.916 4	1.759 3	0.520 6	2.095 1	2.265 7
buit-in 子空间法	1.107 6	−3.918 4	1.657 4	−2.373 7	1.966 0	−0.487 2	2.259 3	1.975 3
本文方法	1.234 4	−1.694 8	1.787 7	−0.657 7	2.052 3	0.588 7	2.314 6	2.407 3

表 3 Train 火车噪声环境下各种语音增强算法效果对比

方法	−10 db		−5 db		0 db		5 db	
	pesq	segSNR	pesq	segSNR	pesq	segSNR	pesq	segSNR
谱减法	1.708 8	−6.043 0	1.726 4	−5.112 6	1.815 0	−4.411 7	2.063 7	−3.724 9
维纳滤波法	0.370 5	−0.863 6	1.169 5	1.002 8	1.983 1	3.485 4	2.487 1	6.160 8
buit-in 子空间法	0.683 9	−1.246 1	1.380 2	1.423 8	1.956 5	4.250 5	2.461 6	6.760 5
本文方法	0.836 2	−0.224 8	1.547 2	2.043 5	2.126 1	4.361 1	2.515 0	6.559 1

另外,我们也要关注语音可懂度是否有提高.频域分段信噪比 fwSNRseg(Frequency weighted segmental SNR)与可懂度主观测试结果相关性较高^[10],因此它在一定程度上可反映增强后语音可懂度水平.fwSNRseg 值越大说明可懂度水平越高.使用测试 pesq 和 segSNR 得分的同一组语音数据来检测可懂度水平.图 1、图 2、图 3 分别表示 white 白噪声、Street 街道噪声以及 Train 火车噪声环境下各种语音增强算法及加噪未处理语音的可懂度对比.由图可知,本文方法的 fwSNRseg 得分整体上高于其它几种参考方法,虽然在有些信噪比处得分稍低于内嵌预白化算法.

综上,在低信噪比下,本文算法的 segSNR 得分提高是尤为突出的,同时 pesq 得分也有一定的提高,可懂度维持在一定范围.说明本文算法可以有效去除噪声,改善听觉质量,保证一定语音可懂度,适合工作在低信噪比环境下.

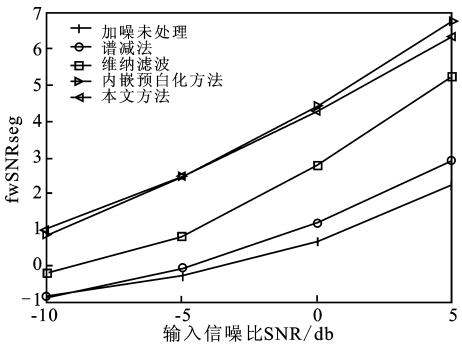


图 1 白噪声环境下可懂度对比

5 结束语

本文提出在低信噪比环境下,信号子空间优化与维纳滤波结合的子空间语音增强算法.特征值和特征向量一一对应,去除小于零的特征值的同时,相应的特征向量进行去除,否则反而会产生残留噪声.这有助于对信号子空间优化,尽可能去除信号子空

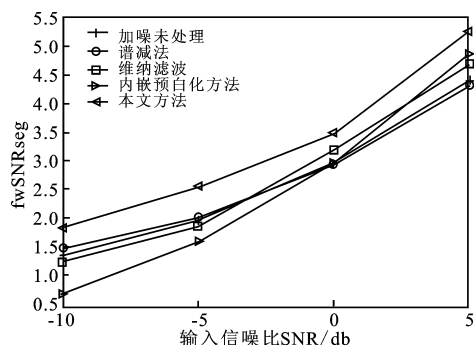


图2 街道噪声环境下可懂度对比

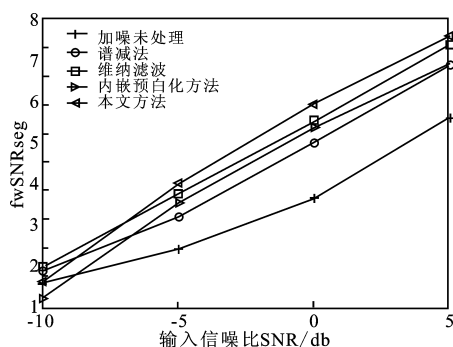


图3 火车噪声环境下可懂度对比

间中存在残留噪声。与此同时,将频域维纳滤波方法引入子空间域,让估计的干净语音通过维纳滤波器,可以进一步抑制残留噪声,保证语音可懂度。最终实验结果也表明,当输入信噪比在-10 dB和5 dB之间时,本文算法可以有效去除残留噪声,改善语音听觉质量,同时保证一定语音可懂度。

参考文献:

- [1] Boll S F. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans on Acoustics, Speech and Signal Processing, 1979, 27(2): 113-120.
- [2] 武夷峰,贾海荣,郭欣. 用子空间提高有色噪声背景下的语音质量[J]. 太原理工大学学报, 2015, 46(2): 192-194.
- [3] Denrinis M, Bakamidis S, Carayanni G. Speechen

hancement from noise: a regenerative approach[J]. Speech Communication, 1991, 10(2): 45-57.

- [4] Ephraim Y, Van Trees H L. A signal subspace approach for speech enhancement[J]. IEEE Trans on Speech and Audio Proc, 1995, 3(4): 251-266.
- [5] Rezayee A, Gsot S. An adaptive KLT approach for speech enhancement[J]. IEEE Trans on Speech and Audio Processing, 2001, 9(2): 87-95.
- [6] Yi Hu, Loizou P C. A generalized subspace approach for enhancing speech corrupted by colored noise[J]. IEEE Trans on Speech and Audio Processing, 2003, 11(4): 334-341.
- [7] Saadoun A, Amrouche A, Selouani S. A perceptual s-subspace speech enhancement using variance of the reconstruction error[J]. Digital Signal Proc, 2014, 24(1): 187-196.
- [8] Chang Jing, Li Zhibin. The generalized inverse eigenvalue problem for generalized periodic jacobi matrices[C]//Proceedings of International Conference on Computer Design and Applications. Washington DC, USA; IEEE Press, 2010: 354-356.
- [9] Loizou P C, Kim G. Reasons why current speech enhancement algorithms do not improve speech intelligibility and suggested solutions[J]. IEEE Trans on Audio, Speech and Language Processing, 2011, 19(1): 47-56.
- [10] J Ma, Yi Hu, Loizou P C. Objective measures for predicting speech intelligibility in noisy conditions based on new band importance functions[J]. The Journal of the Acoustical society of America, 2009, 125(5): 3387-3405.

作者简介:

王红女, (1990-), 硕士研究生. 研究方向为语音信号处理. E-mail: 229614452@qq.com.

王爱平女, (1956-), 教授, 硕士生导师. 研究方向为人工智能、编译技术、滤波算法收敛性.

邹海男, (1969-), 副教授, 硕士生导师. 研究方向为中间件理论与技术、智能计算.