

子空间与维纳滤波相结合的语音增强方法

张雪英, 贾海蓉, 靳晨升

ZHANG Xueying, JIA Hairong, JIN Chensheng

太原理工大学 信息工程学院, 太原 030024

College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China

ZHANG Xueying, JIA Hairong, JIN Chensheng. Speech enhancement method based on combination of subspace and Winner filter. Computer Engineering and Applications, 2011, 47(14): 146-148.

Abstract: In view of the musical noise after the enhancement of speech corrupted by complicated additive noise, a speech enhancement method based on the combination of subspace and Winner filter is proposed. This method has following steps. By KL transformation the noisy speech is transformed into subspace domain, and the noisy speech eigenvalue is estimated. A Winner filter is formed by using the Signal-Noise-Ratio (SNR) formula in subspace domain. The estimated eigenvalue is filtered by the Winner filter. Thereby the new clean speech eigenvalue is gained. The clean speech is gained by KL reverse transformation. Simulation results show that under the background of white and train noise, the SNR in this method is more excellent than that in traditional subspace method. Meanwhile the musical noise after the enhancement is depressed effectively.

Key words: speech enhancement; subspace; Winner filter; musical noise

摘 要: 针对复杂背景噪声下语音增强后带有音乐噪声的问题, 提出一种子空间与维纳滤波相结合的语音增强方法。对带噪声语音进行KL变换, 估计出纯净语音的特征值, 再利用子空间域中的信噪比计算公式构成一个维纳滤波器, 使该特征值通过这个滤波器, 从而得到新的纯净语音特征值, 由KL逆变换还原出纯净语音。仿真结果表明, 在白噪声和火车噪声的背景下, 信噪比都比传统子空间方法有明显提高, 并有效抑制了增强后产生的音乐噪声。

关键词: 语音增强; 子空间; 维纳滤波; 音乐噪声

DOI: 10.3778/j.issn.1002-8331.2011.14.042 文章编号: 1002-8331(2011)14-0146-03 文献标识码: A 中图分类号: TN912

1 引言

在通信传输过程中信号不可避免地会受到各种背景噪声的干扰, 语音增强的目的就是将从被噪声污染的语音中恢复出原始的纯净语音, 改善语音质量, 提高语音可懂度。目前, 常用的语音增强方法有基于谱减^[1]、最小均方误差^[2]、掩蔽效应^[3]以及基于它们的改进算法, 但增强以后的语音都不同程度地存在一种具有一定节奏感的残余噪声, 一般称为“音乐噪声”。因此如何很好地去除音乐噪声就成为语音增强算法的关键。采用子空间与维纳滤波相结合的方法来消除音乐噪声, 改善语音质量。仿真结果表明, 在复杂背景噪声下, 该算法不仅提高了信噪比, 而且有效抑制了增强后产生的音乐噪声。

2 方法描述

2.1 子空间语音增强原理

子空间方法是通过空间分解, 将整个空间划分为两个独立子空间, 即噪声子空间和叠加噪声的信号子空间。然后对噪声子空间和信号子空间进行处理以实现语音增强。在加性噪声干扰的信号模型中, 假设语音信号和噪声信号互不相关,

设带噪声语音为:

$$Y(k) = S(k) + N(k) \quad (1)$$

Y, S 和 N 分别为 k 维带噪声语音矢量、纯净语音矢量和噪声信号矢量, 令 R_Y, R_S 和 R_N 分别表示 Y, S 和 N 的协方差矩阵, 令 $\hat{S} = HY$ 为纯净的语音信号的估计, H 为 $k \times k$ 的线性预测器。则预测值和真实值的误差为:

$$\epsilon = \hat{S} - S = (H - I) \cdot S + H \cdot N = \epsilon_S + \epsilon_N \quad (2)$$

其中, ϵ_S 和 ϵ_N 分别表示语音信号的失真和增强后残留的噪声, 相应的能量分别为:

$$\begin{aligned} \overline{\epsilon_S^2} &= E[\epsilon_S^T \epsilon_S] = \text{tr}(E[\epsilon_S^T \epsilon_S]) = \text{tr}\{(H - I)R_S(H - I)^T\} = \\ &\quad \text{tr}(HR_S H^T - HR_S - RH^T + R_S) \end{aligned} \quad (3)$$

$$\begin{aligned} \overline{\epsilon_N^2} &= E[\epsilon_N^T \epsilon_N] = \text{tr}(E[\epsilon_N^T \epsilon_N]) = \\ &\quad \text{tr}\{(H \cdot N)R_N(H \cdot N)^T\} = \text{tr}(HR_N H^T) \end{aligned} \quad (4)$$

求解下面时域约束条件(TDC)下的方程^[4], 就可以得到优化的线性估计器:

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60472094); 山西省自然科学基金(No.20051039, No.2008011031)。

作者简介: 张雪英(1964—), 女, 教授, 博士生导师, 主要从事语音信号处理方面的研究。E-mail: tyzhangxy@163.com

收稿日期: 2009-09-04; **修回日期:** 2010-01-25

$$\min_H \bar{\epsilon}_S^2 \text{ (在 } (1/k)\bar{\epsilon}_N^2 \leq \sigma^2 \text{ 条件下)} \quad (5)$$

其中, σ^2 为正常数, 上述方程的解为:

$$\mathbf{H}_{\text{opt}} = \mathbf{R}_S (\mathbf{R}_S + \mu \mathbf{R}_N)^{-1} \quad (6)$$

对式(6)特征分解为:

$$\mathbf{H}_{\text{opt}} = \mathbf{U} \mathbf{A}_S (\mathbf{A}_S + \mu \mathbf{A}_N)^{-1} \mathbf{U}^T \quad (7)$$

$$\mathbf{A}_S = \text{diag}(\lambda_S^1, \lambda_S^2, \dots, \lambda_S^k) \quad (8)$$

$$\mathbf{A}_N = \text{diag}(\lambda_N^1, \lambda_N^2, \dots, \lambda_N^k) \quad (9)$$

其中, \mathbf{A}_S , \mathbf{A}_N 为 \mathbf{R}_S 和 \mathbf{R}_N 的特征值组成的对角阵, \mathbf{U} 为特征向量矩阵。因此, 最优线性估计器^[5]可写为:

$$\mathbf{H}_{\text{opt}} = \mathbf{U} \mathbf{G} \mathbf{U}^T \quad (10)$$

$$\mathbf{G} = \mathbf{A}_S (\mathbf{A}_S + \mu \mathbf{A}_N)^{-1} \quad (11)$$

Rezayee 等在文献[5]中通过对噪声向量的协方差矩阵 \mathbf{R}_N 近似对角化, 可以将其推广到有色噪声的情况, 当 $\mathbf{A}_N = \sigma_n^2 \mathbf{I}$ 即 \mathbf{R}_N 的特征值矩阵为单位阵时, 式(11)退化为文献[4]中所作白噪声假设的情况。

根据 $\hat{\mathbf{S}} = \mathbf{H} \mathbf{Y}$ 和式(11), 可通过对带噪信号进行KL变换即左乘矩阵 \mathbf{U}^T , 再将变换的结果乘以一个增益矩阵 \mathbf{G} , 然后对所得结果进行KL逆变换即左乘矩阵 \mathbf{U} , 最后可得到增强的语音信号:

$$\hat{\mathbf{S}} = \mathbf{H}_{\text{opt}} \cdot \mathbf{Y} \quad (12)$$

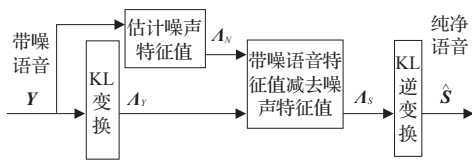


图1 传统子空间语音增强方法原理图

传统子空间语音增强方法步骤为:

步骤1 对带噪语音进行KL变换, 得到子空间域中的带噪语音的特征向量 \mathbf{U} 和特征值 \mathbf{A}_Y 。

步骤2 估计噪声特征值 \mathbf{A}_N , 一般取带噪语音前3000点方差的平均值作为噪声的特征值估计值。

步骤3 带噪语音特征值 \mathbf{A}_Y 减去噪声特征值 \mathbf{A}_N , 得到纯净语音特征值 \mathbf{A}_S , 由式(11)可以得到增益 \mathbf{G} 。

步骤4 由KL逆变换可得最优估计器 \mathbf{H}_{opt} , 由式(12)可得纯净语音的估计 $\hat{\mathbf{S}}$ 。

2.2 子空间与维纳滤波相结合的语音增强方法原理

传统子空间方法增强后的语音带有明显的音乐噪声, 而维纳滤波最大的好处就是可以有效地抑制音乐噪声。本文算法的改进之处就是利用子空间域中的信噪比公式构造一个维纳滤波器, 并对估计出的纯净语音特征值进行滤波, 从而达到抑制音乐噪声的目的。

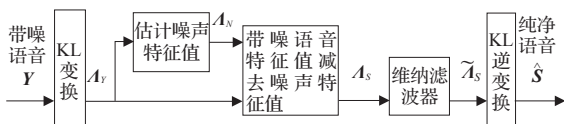


图2 子空间与维纳滤波相结合的语音增强方法原理图

子空间与维纳滤波相结合的语音增强方法具体步骤为:

步骤1 同传统子空间语音增强算法步骤1。

步骤2 估计噪声特征值 \mathbf{A}_N :

\mathbf{A}_N 的估计采用最小值跟踪算法, 跟踪带噪语音协方差矩阵每个特征向量上的特征值, 在每一帧内更新噪声的特征值。可分为以下3步:

第一步: 在每个特征向量上对带噪语音 \mathbf{R}_Y 的特征值 $A_Y^i(n)$ 进行一阶平滑。

$$A(i, n) = \alpha_s A(i, n-1) + (1 - \alpha_s) A_Y^i(n) \quad (13)$$

式中, $A(i, n-1)$ 为第 i 个特征向量上的平滑特征值, n 表示当前帧 $\alpha_s (0 < \alpha_s < 1)$ 为一常数。

第二步: 最小值跟踪。将带噪语音 \mathbf{R}_Y 的特征值 $\lambda_Y^i(n)$ 进行一阶平滑后, 找出其局部最小值, 在相邻帧的每个分量上连续搜索和更新。

$$A_{\min}(i, n) = \min\{A_{\min}(i, n-1), A(i, n)\} \quad (14)$$

第三步: 定义一个临时变量:

$$A_{\text{tmp}}(i, n) = \min\{A_{\text{tmp}}(i, n-1), A(i, n)\} \quad (15)$$

每当处理第 jL (j 为正整数) 帧时, 更新局部最小值 $A_{\min}(i, n) \lim_{x \rightarrow \infty}$ 和临时变量 $A_{\text{tmp}}(i, n)$, 求出特征值的最小值得到噪声的特征值 \mathbf{A}_N 。

$$A_{\min}(i, n) = \min\{A_{\text{tmp}}(i, n), A(i, n)\}$$

$$A_{\text{tmp}}(i, n) = A(i, n) \quad (16)$$

其中, L 为一个包含多帧的搜索窗口大小, L 的选取要足够长, 以保证跟踪的最小值与噪声相关, 实验表明, 窗长 L 可以在 0.5~1.5 s 之间选择, $\alpha_s = 0.8$ 。

步骤3 同传统子空间语音增强算法步骤3。

步骤4 构造子空间域中的维纳滤波器, 这也是本文算法的改进之处。

文献[6]子空间域中信噪比计算公式可表示为:

$$SNR(k) = \frac{\text{tr}(\mathbf{U} \mathbf{R}_S \mathbf{U}^T)}{\text{tr}(\mathbf{U} \mathbf{R}_N \mathbf{U}^T)} = \frac{\sum_{k=1}^M \lambda_{\Sigma}^{(k)}}{k} = \frac{E_1}{E_2} \quad (17)$$

其中, λ 为纯净语音特征值, k 为纯净语音特征值大于零的个数, $E_1 = \sum_{k=1}^M \lambda_{\Sigma}^{(k)}$ 为信号能量, $E_2 = k$ 为噪声能量。

在子空间域中, 由维纳滤波在频域中的表达式:

$$W(k) = \frac{P_s(k)}{P_s(k) + P_n(k)} \quad (18)$$

其中 $P_s(k)$ 和 $P_n(k)$ 分别为语音和噪声功率谱密度。

可以构造一个子空间域中的维纳滤波器:

$$W(k) = \frac{E_1}{E_1 + E_2} = \frac{\sum_{k=1}^M \lambda_{\Sigma}^{(k)}}{\sum_{k=1}^M \lambda_{\Sigma}^{(k)} + k} \quad (19)$$

步骤5 利用维纳滤波器估计出新的纯净语音特征值。

由子空间域中构造的维纳滤波器 $W(k)$, 带噪语音特征值 \mathbf{A}_Y 和估计出的噪声特征值 \mathbf{A}_N , 可得纯净语音特征值通过维纳滤波器滤波的表达式为:

$$\tilde{\mathbf{A}}_S = \max\{(\mathbf{A}_Y - \mathbf{A}_N), 0\} \cdot W(k) \quad (20)$$

其中, $\tilde{\mathbf{A}}_S$ 为通过维纳滤波器估计出的新的纯净语音特征值。

步骤6 由KL逆变换可得新的最优估计器为:

$$\widehat{H}_{opt} = U \widehat{A}_s (\widehat{A}_s + \mu A_N)^{-1} U^{-T} \tag{21}$$

由式(12)可得新纯净语音的估计 \hat{S} 。

3 仿真

语音取自 863 语音库,噪声为 NOISEX.92 数据库的高斯白噪声和火车噪声,语音信号的采样率为 8 kHz,帧长 K 为 80 个采样点,帧叠为 50%。

首先在纯净语音中加入 5 dB 高斯白噪声作为带噪语音,采用本文的方法得到增强后仿真结果图 3 和图 4。

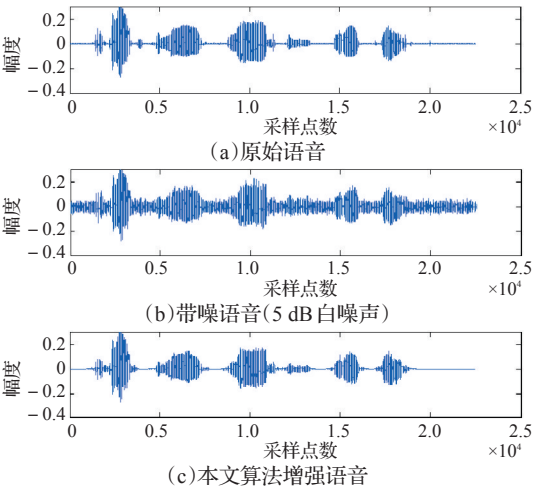


图3 波形图比较

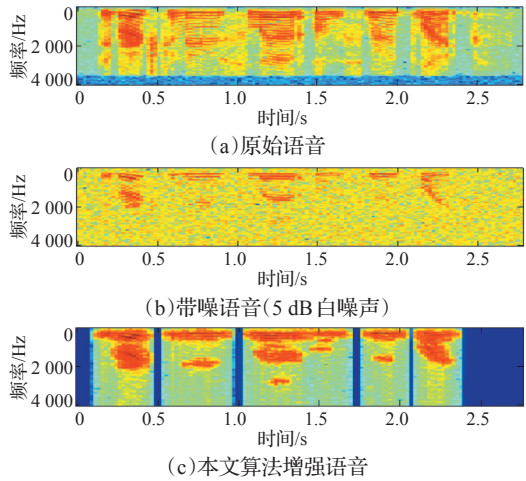


图4 语谱图比较

从图 3 和图 4 可以看出,增强语音和原始语音在波形图和语谱图上已非常接近,音乐噪声已得到很好的抑制,主观听觉感受基本一致。

其次,比较 Ephraim 和 Van Trees^[9]提出的白噪声条件下传统子空间语音增强算法和本文算法。测试对象为纯净语音分别加入信噪比为 0 dB, 5 dB, 10 dB, 15 dB 的高斯白噪声和信噪比为 -0.669 8 dB, 4.330 1 dB, 9.330 2 dB, 14.330 0 dB 的火车噪声,增强后语音质量性能的评估采用客观评估指标信噪

表1 传统子空间方法与本文方法性能比较表

噪声种类	输入信噪比/dB	传统子空间方法 SNR/dB	本文方法 SNR/dB
白噪声	0	9.170 9	9.756 6
	5	11.965 7	12.354 0
	10	15.476 0	16.184 1
	15	17.802 1	19.103 2
火车噪声	-0.669 8	2.962 2	3.622 1
	4.330 1	8.652 0	9.385 0
	9.330 2	12.619 2	13.096 5
	14.330 0	14.500 9	15.152 4

比 SNR。仿真结果如表 1 所示。

从表 1 的实验结果可以看出,在各种信噪比下本文方法比传统方法语音增强效果有明显提高。通过听测表明,采用传统方法增强后音乐噪声较多,而采用本文方法增强后明显地抑制了音乐噪声,取得了很好的增强效果。

4 结束语

将频域中的维纳滤波方法应用到子空间域中,提出了一种子空间与维纳滤波相结合的语音增强方法,仿真结果和主观听测都已表明,该方法比传统子空间方法语音增强效果有明显提高,很好地抑制了增强后产生的音乐噪声。另外,该方法计算过程简单,运算速度较快,利于实时实现。

参考文献:

[1] Li Ye, Cui Huijuan, Tang Kun. Speech enhancement algorithm based on spectral subtraction[J]. Journal of Tsinghua University, 2006, 46(10): 1685-1687.

[2] Rainer M. Speech enhancement based on minimum mean-square error estimation and supergaussian priors[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 845-856.

[3] 阔永红. 基于听觉掩蔽效应的 MMSE 语音增强算法[J]. 计算机工程与应用, 2007, 43(27): 140-141.

[4] Ephraim Y, Van Trees H L. A signal subspace approach for speech enhancement[J]. IEEE Trans on Speech and Audio Processing, 1995, 3(4): 251-266.

[5] Gazor S, Rezayee A. An adaptive KLT approach for speech enhancement[J]. IEEE Trans on Speech and Audio Processing, 2001, 9(2): 95-97.

[6] Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics[J]. IEEE Trans on Speech and Audio Processing, 2001, 9(5): 504-512.

[7] 陈照平, 张雪英, 马建芬. 基于短时谱估计的语音增强方法研究[D]. 太原: 太原理工大学, 2008.

[8] 赵胜跃, 戴蓓倩. 基于最小统计噪声估计的信号子空间语音增强[J]. 数据采集与处理, 2007(4).

[9] 吴周桥, 谈新权. 基于子空间方法的语音增强算法研究[J]. 声学电子工程, 2005(3).

[10] Lim J S, Oppenheim A V. Enhancement and bandwidth compression of noisy speech[J]. IEEE Proceedings, 1979, 67(1): 1586-1604.