

6-4-2020

Eigen-value Combination Approach for Speech Enhancement

Huijuan Lu

1. *College of Information Engineering, China Jiliang University, Hangzhou 310018, China;;*

Yaqing Liu

1. *College of Information Engineering, China Jiliang University, Hangzhou 310018, China;;*

Yanqiu Liu

1. *College of Information Engineering, China Jiliang University, Hangzhou 310018, China;;*

Guan Wei

2. *College of Moder Science and Technology, China Jiliang University, Hangzhou 310018, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Eigen-value Combination Approach for Speech Enhancement

Abstract

Abstract: *In order to further suppress noise, a kind of speech enhancement algorithm based on Eigen-value merge was proposed.* Eigen-values combination was used to improve the speech quality on the basis of the classic embedded pre-whitening subspace methods. The study shows that, after decomposing the covariance matrix of speech signals with noise, the larger Eigen-value component mainly includes speech information, and the smaller Eigen-value component mainly contains noise. Sorted by Eigen-values from small to big, the adjacent large Eigen-value component replaces with small Eigen-value component, which can effectively suppress noise and improve the quality of speech. Compared with other speech enhancement algorithm, this algorithm based on Eigen value merger can work effectively in a variety of noisy environment, significantly improves the SNR, and has better speech intelligibility.

Keywords

speech enhancement, subspace method, eigen-value decomposing, speech quality

Recommended Citation

Lu Huijuan, Liu Yaqing, Liu Yanqiu, Guan Wei. Eigen-value Combination Approach for Speech Enhancement[J]. Journal of System Simulation, 2016, 28(7): 1622-1627.

基于特征值合并的语音增强算法

陆慧娟¹, 刘亚卿¹, 刘砚秋¹, 关伟²

(1. 中国计量学院信息工程学院, 浙江 杭州 310018; 2. 中国计量学院现代科技学院, 浙江 杭州 310018)

摘要: 为进一步抑制噪声, 提出一种基于特征值合并的语音增强算法。在经典的内嵌式预白化子空间方法的基础上, 用特征值合并来提高语音质量。研究发现, 对含噪语音的协方差矩阵进行特征值分解后, 大特征值分量主要包含语音信息, 而小特征值分量主要包含噪声, 特征值分量按特征值从小到大排序后, 剔除相邻的小特征值分量, 可有效抑制噪声, 提高语音质量。相比于其它方法, 基于特征值合并的语音增强算法能有效工作于各种噪声环境中, 显著提高信噪比, 并有更好的语音可懂度。

关键词: 语音增强; 子空间方法; 特征值分解; 语音质量

中图分类号: TN912.3

文献标识码: A

文章编号: 1004-731X (2016) 07-1622-06

Eigen-value Combination Approach for Speech Enhancement

Lu Huijuan¹, Liu Yaqing¹, Liu Yanqiu¹, Guan Wei²

(1. College of Information Engineering, China Jiliang University, Hangzhou 310018, China;

2. College of Moder Science and Technology, China Jiliang University, Hangzhou 310018, China)

Abstract: In order to further suppress noise, a kind of speech enhancement algorithm based on Eigen-value merge was proposed. Eigen-values combination was used to improve the speech quality on the basis of the classic embedded pre-whitening subspace methods. The study shows that, after decomposing the covariance matrix of speech signals with noise, the larger Eigen-value component mainly includes speech information, and the smaller Eigen-value component mainly contains noise. Sorted by Eigen-values from small to big, the adjacent large Eigen-value component replaces with small Eigen-value component, which can effectively suppress noise and improve the quality of speech. Compared with other speech enhancement algorithm, this algorithm based on Eigen value merger can work effectively in a variety of noisy environment, significantly improves the SNR, and has better speech intelligibility.

Keywords: speech enhancement; subspace method; eigen-value decomposing; speech quality

引言

语音增强的主要目的是从带噪的语音信号中尽可能提取纯净的原始信号, 消除背景噪声和提高

语音的实际质量, 使听者更容易接受。谱减法是最早的语音增强算法, 它是由 Boll^[1-2]在 1979 年提出的。谱减法假设噪声与纯净的语音信号相互独立, 而语音信号具有短时平稳性, 所以从含噪语音的功率谱中减去噪声功率谱可以得到较为纯净的语音频谱, 再通过反短时傅里叶变换即可得到增强后的语音。谱减法的优点是运算量小, 易于实现, 但也存在自身的缺点。由它引进的“音乐噪声”是很多改进算法所不能弥补的缺陷。基于 MMSE 的语音



收稿日期: 2014-08-12 修回日期: 2015-02-01;
基金项目: 国家自然科学基金(61272315, 60842009),
浙江省自然科学基金(Y1110342);
作者简介: 陆慧娟(1962-), 女, 浙江, 博士, 教授,
研究方向为物联网、模式识别与云计算。

<http://www.china-simulation.com>

• 1622 •

增强算法^[3-4]在降噪和提高语音可懂性方面进行折中, 适用范围较广, 但计算量大, 模型参数不易调整, 而且语音频谱的先验分布的获得在很大程度上取决于统计结果的代表性和重要性。子空间语音增强算法^[5-8]是一种新颖高效的算法, 它具有语音失真小、残余噪声少、音乐噪声不明显等特点, 并且具有很好的控制残余噪声和语音失真对立关系的机制。它将带噪语音信号的向量空间分解为 2 个子空间, 一个是语音信号的子空间, 一个是噪声信号的子空间, 通过将噪声子空间置零, 得到纯净语音信号的近似估计。

子空间算法最早由 Dendrion^[9]在 1991 年提出, 该算法根据最小二乘估计准则, 对含噪语音进行 SVD(Singular Value Decomposition)分解, 忽略较小的奇异值, 从中恢复出语音信号, 这种方法只适合白噪声的情况。随后, Jensen^[10]将基于奇异值分解算法推广到有色噪声环境下, 将对噪声信号的白化作为子空间算法的一部分, 并采用最小方差准则估计语音信号子空间。1995 年, Ephraim 和 Van Trees^[5]提出采用 KLT(Karhunen-Loeve transform)进行子空间分解的方法, 通过对有色噪声进行预白化处理, 使得子空间算法可以处理有色噪声。Rezayee 和 Gazor^[11]假设语音信号的特征向量可以近似使有色噪声的协方差矩阵对角化, 提出了一种自适应的 KLT 方法, 但这种方法为次优估计方法。Yi Hu^[6]等人则利用广义特征值分解(GEVD)算法给出了一种最优的估计。

语音信号的短时平稳性具体表现为语音帧的相关性。现有的子空间算法中只考虑去噪, 并没有考虑提高语音帧的相关性。实验证明, 噪声的引入严重降低了语音帧之间的相关性。基于这点考虑, 本文在内嵌式预白化子空间方法的基础上, 提出特征值合并的方法来改善语音增强效果。即对含噪语音的协方差矩阵进行特征值分解后, 剔除相邻的小特征值分量, 提高语音帧的相关性, 从而进一步抑制噪声。实验证明, 该算法能够显著提高增强语音的可懂性, 同时有效抑制噪声。

1 子空间语音增强方法简介

子空间方法是将带噪语音信号的向量空间分解为相互正交的 2 个子空间, 其中一个只包含噪声, 而另一个同时包含语音和噪声。通过将噪声子空间置零, 得到的即为纯净语音信号的近似估计。

假设纯净语音信号 s 与噪声 n 互不相关且为零均值的, 那么含噪语音信号 y 可以表示为:

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

纯净语音信号 s 可由含噪语音信号 y 和线性评估器 H 估计得到

$$\hat{s} = Hy \quad (2)$$

估计误差为:

$$r = \hat{s} - s = (H - I)s + Hn = r_s + r_n \quad (3)$$

其中 r_s 代表信号失真, r_n 表示残余噪声能量。

定义信号失真能量为 $\overline{r_s^2}$, 残余噪声能量为 $\overline{r_n^2}$, 即

$$\overline{r_s^2} = E[r_s^T r_s] = \text{tr}(E[r_s^T r_s]) = \sigma^2 \text{tr}(HH^T) \quad (4)$$

$$\overline{r_n^2} = E[r_n^T r_n] = \text{tr}(E[r_n^T r_n]) =$$

$$\text{tr}((H - I)R_s(H - I)^T) \quad (5)$$

其中: σ^2 为一常数; R_s 为纯净语音信号的协方差矩阵; I 为单位矩阵。

把残余噪声能量控制在某一阈值之下, 同时保证信号失真最小, 即

$$\min \overline{r_s^2} \quad \text{Subject to: } \overline{r_n^2} \leq k\sigma^2 \quad (6)$$

在以上条件约束下, 最优线性评估器 H 可以通过求下式的极值得

$$\begin{aligned} \ell(H, \mu) = \overline{r_s^2} + \mu(\overline{r_n^2} - K\sigma^2) = \\ \text{tr}((H - I)R_s(H - I)^T) + \\ \mu(\sigma^2 \text{tr}(HH^T) - K\sigma^2) \end{aligned} \quad (7)$$

其中 $\mu \geq 0$ 为拉格朗日系数。通过对式(7)关于 H 求偏导, 很容易得到最优线性评估器:

$$H_{opt} = R_s(R_s + \mu R_n)^{-1} \quad (8)$$

其中 $R_n = \sigma^2 I$ 为噪声的协方差矩阵。对于高斯白噪声来说, 最优线性评估器可以表示为:

$$H_{opt} = U \Delta_s (\Delta_s + \mu \Delta_n)^{-1} U^T \quad (9)$$

U 为特征向量矩阵, 而 Δ_s 为含噪语音信号协方差矩阵特征值分解得到的特征值矩阵, Δ_n 为噪声协方差矩阵特征值分解得到的特征值矩阵。(9) 式是白噪声环境下的最优线性评估器, 但对于更具一般性的有色噪声却不是最优解。Yi Hu^[6]等人利用广义特征值分解(GEVD)对有色噪声的情况给出了一种最优估计。

在矩阵论中, 存在一个矩阵 V 同时可以对角化 2 个矩阵, 即 :

$$\begin{aligned} V^T R_s V &= A \\ V^T R_n V &= I \end{aligned} \tag{10}$$

这里 V 和 A 分别为 R_s 和 R_n 的特征向量矩阵及特征值矩阵。此算法广义特征值分解的过程就是一个噪声白化的过程, 事实证明内嵌式预白化子空间方法是一种处理有色噪声的最优估计。最优线性估计器为:

$$\begin{aligned} H_{opt} &= R_n V \Lambda_s (\Lambda_s + \mu I)^{-1} V^T = \\ &V^{-T} \Lambda_s (\Lambda_s + \mu I)^{-1} V^T \end{aligned} \tag{11}$$

其中 μ 满足式(12)

$$\kappa \sigma^2 = tr\{(V^T V)^{-1} \Lambda_s^2 (\Lambda_s + \mu I)^{-2}\} \tag{12}$$

2 特征值合并算法

研究表明, 一般情况下大的特征值分量主要包含着语音信息, 小的特征值分量则主要包含噪声。低信噪比时, 噪声信息比较多, 对含噪语音信号的协方差矩阵进行特征值分解后并进行特征值排序后, 如果相邻特征值比较接近(在阈值 L 之内), 剔除小的特征值分量, 即是剔除更多的噪声, 所以可以得到增强效果更好的语音信号。在高信噪比情况下, 噪声信息比较少, 同样剔除小的特征值分量, 会造成语音信号失真。但由于帧移的存在, 失真的语音信息大部分是重复的语音信息, 所以也能取得不错的增强效果。

另一方面, 对含噪语音信号的协方差矩阵进行特征值分解并进行特征值排序后, 剔除相邻的小的特征值分量, 此时相邻语音帧变为当前帧和第 3 帧(第 2 帧被剔除)。那么原来第 2 帧和第 3 帧的相

关系系数变为第 1 帧和第 3 帧的运算。由于第 1 帧为大的特征值分量, 包含更多的语音信息, 在改进后相邻帧的相关性会变大。这是内嵌式预白化子空间语音增强算法所没有考虑到的。

内嵌预白化子空间方法的实质是对含噪语音信号进行 V^T 变换($V^T y$), $V^T y$ 即为将含噪信号投影后得到的主要包含语音信号的部分。合并相近的特征值后, 得到的特征向量矩阵 V 不再是方阵。通常只有方阵才能求得其逆矩阵, 因此对特征向量矩阵进行广义逆运算^[12]求其逆矩阵。

实验表明, 该算法在所有情况下都强于经典的内嵌式预白化子空间算法。图 1 为特征值置换理论的算法流程图。首先对一段 Noisex-92^[13]语音库里的语音进行分帧处理。基于子空间的语音增强的处理过程也就是把语音信号转换为矩阵的处理过程。分帧能够把一段语音信号转换为矩阵, 考虑到语音信号前后的连续性, 设置了帧移。一般帧移的长度是帧长的一半, 即语音是以一半帧移的速度向前推动。在对语音信号进行加窗之后, 每一帧都可以当作平稳信号处理。之后求得噪声和混合语音的协方差矩阵, 进而求得 $R_n^{-1} R_s$ 协方差矩阵。进行广义特征值分解, 求得其特征值及对应的特征向量并且筛选排序之后, 可以得到线性评估器。再经过一个反叠加的过程, 即得到增强后的语音。

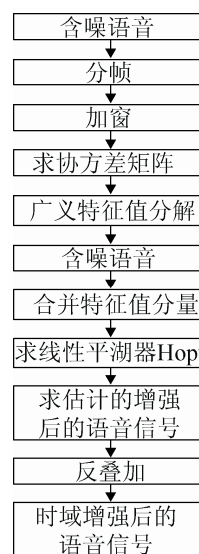


图 1 特征值合并语音增强算法流程图

对含噪语音信号的协方差矩阵进行特征值分解并进行特征值排序后,用 L 的大小决定相邻特征值分量的取舍, L 为相邻特征值分量的特征值之差。实验表明,在低信噪比下, $0.001 < L < 0.05$ 时有很好的语音增强效果,即相邻两个特征值分量的特征值之差在该范围内,剔除掉小的特征值分量;在高信噪比情况下,当 $0.001 < L < 0.005$ 时有很好的语音增强效果,即相邻两个特征值分量的特征值之差在该范围内时,剔除掉相邻的小特征值分量可以有效改善语音质量。

3 仿真实验与结果分析

在 Noisex-92 库中选择分别参杂了四种不同噪声的语音信号来验证特征值置换算法的语音增强性能和效果。分别是:白噪声,粉红噪声, F16 战机的噪声和 8000 Hz 的高频信道噪声。每种噪声从“sp01”到“sp30”共 30 组语音信号,保证结果的客观准确性。噪声信号和语音信号的采样频率均为 8 kHz,采样长度为 21 000。矩形窗的长度为 300,帧移为 50%。PESQ(Perceptual Evaluation of Speech

Quality)^[14] 代表语音质量, segSNR(segmental SNR)^[15]代表噪声残余即去噪程度。

由图 2 可得出,内嵌式预白化子空间方法去噪声是不彻底的。它和别的子空间方法一样,只是去掉了对应着噪声的子空间部分,而对应着语音信号子空间部分的噪声没有做任何处理。这将影响着语音信号质量及可懂性。作为参考的其它经典语音增强算法有:内嵌式预白化子空间语音增强算法,SSBoll79(spectral subtraction)^[11], SSPARA98(A Parametric Formulation of the Generalized Spectral Subtraction method)^[16]。

表 1 为各种语音增强算法在白噪声环境下语音增强效果的对比。可以看出,相比于其它语音增强算法,在白噪声环境下,语音信号的质量及分段信噪比都有小幅度的提高。从表 2 可以看出:对于粉红噪声环境,低信噪比下,特征值分量合并算法去噪效果特别明显。在高信噪比下,去噪效果相比于其它语音增强算法也有所提高。同时特征值合并算法也能保证较高的语音质量。

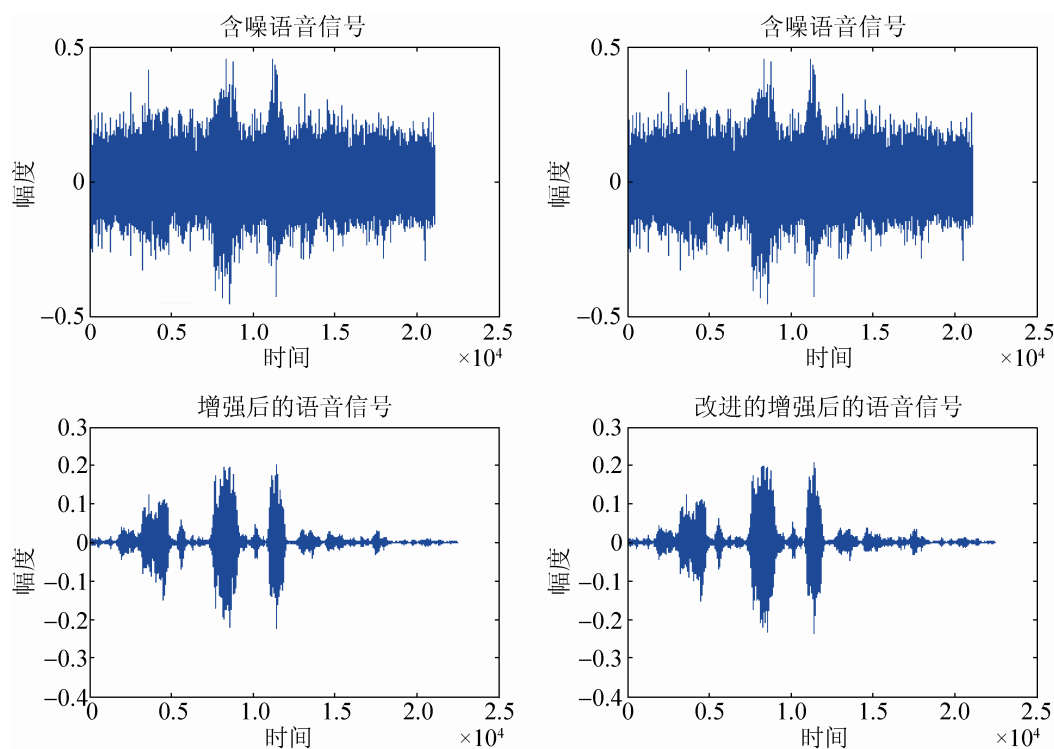


图 2 改进前后时域波形对比图

<http://www.china-simulation.com>

从表 3 很明显看出:特征值合并算法能够有效抑制噪声并且保证很好的语音质量,说明该算法适用于 F16 机舱噪声环境。由表 4 可知,高频信道噪声中,特征值合并算法在各信噪比下都优于内嵌式

预白化子空间方法,同时相对于其它在空间算法也是最优的。在不同信噪比下,特征值合并算法均提高了语音质量并且大幅度地去除了语音信号中的残余噪声。

表 1 白噪声环境下各种语音增强方法的 PESQ 值以及 segSNR 值

语音增强算法	-5 dB	0 dB	5 dB	10 dB	15 dB
特征值合并	1.633 0, -0.450 5	2.094 0, 1.923 4	2.450 1, 4.340 4	2.774 0, 6.448 7	3.118 5, 8.307 5
内嵌式预白化	1.592 9, -0.662 6	2.055 7, 1.746 7	2.439 4, 4.193 0	2.771 2, 6.429 3	3.113 3, 8.293 1
SSBoll79	1.817 2, -1.969 2	1.753 2, -1.619 2	1.153 5, -1.431 4	1.333 8, -1.271 5	1.519 7, -0.412 3
SSPARA98	0.969 3, -0.895 4	1.040 6, -0.794 5	1.241 1, -0.616 3	1.395 3, -0.511 9	1.595 1, -0.456 6

表 2 粉红噪声环境下各种语音增强方法的 PESQ 值以及 segSNR 值

语音增强算法	-5 dB	0 dB	5 dB	10 dB	15 dB
特征值合并	1.731 6, -1.168 4	2.140 8, 1.047 5	2.445 5, 3.378 1	2.783 3, 5.609 5	3.196 1, 7.578 9
内嵌式预白化	1.706 7, -1.224 9	2.114 3, 0.996 5	2.435 2, 3.409 3	2.780 2, 5.597 9	3.188 3, 7.572 6
SSBoll79	0.963 5, -1.949 3	1.532 6, -1.856 9	1.512 2, -1.543 6	1.603 5, -0.725 0	1.618 1, -0.432 2
SSPARA98	0.976 3, -1.140 5	0.984 5, -0.939 5	1.139 4, -0.731 2	1.416 1, -0.562 7	1.595 2, -0.462 7

表 3 F16 噪声环境下各种语音增强方法的 PESQ 值以及 segSNR 值

语音增强算法	-5 dB	0 dB	5 dB	10 dB	15 dB
特征值合并	1.604 5, -1.422 7	2.037 3, 0.781 3	3.326 1, 2.445 5	2.783 3, 5.609 5	3.196 1, 7.578 9
内嵌式预白化	1.586 8, -1.430 0	2.030 5, 0.777 5	3.310 7, 2.435 2	2.780 2, 5.597 9	3.188 3, 7.572 6
SSBoll79	1.248 6, -2.596 7	1.682 5, -1.952 7	-1.782 3, 1.512 2	1.603 5, -0.725 0	1.618 1, -0.432 2
SSPARA98	0.827 9, -1.186 1	1.001 9, -0.929 1	-0.695 5, 1.139 4	1.416 1, -0.562 7	1.595 2, -0.462 7

表 4 高频信道噪声环境下各种语音增强方法的 PESQ 值以及 segSNR 值

语音增强算法	-5 dB	0 dB	5 dB	10 dB	15 dB
特征值合并	1.565 4, -0.994 2	2.099 8, 1.341 5	2.421 0, 3.767 3	2.730 2, 6.060 0	3.064 8, 8.097 0
内嵌式预白化	1.527 5, -1.216 6	2.056 2, 1.128 9	2.408 6, 3.650 9	2.725 8, 6.052 0	3.069 3, 8.092 3
SSBoll79	1.757 0, -2.340 8	1.070 6, -2.008 2	1.820 2, -1.619 9	1.561 9, -1.315 7	1.648 4, -0.429 8
SSPARA98	0.845 9, -1.055 5	0.992 9, -0.918 7	1.113 0, -0.681 9	1.510 7, -0.534 9	1.536 7, -0.437 1

4 结论

在内嵌式预白化子空间语音增强算法的基础上,提出了基于特征值合并的语音增强算法。对含噪语音信号协方差矩阵进行特征值分解并按特征值大小排序后,剔除相邻的小特征值分量,可以进一步有效抑制噪声,并保证较好的语音质量,弥补了内嵌式预白化去噪不彻底的缺陷,将语音质量及可懂度提高到一个新的水平。今后的工作将继续探究特征值分量对语音信号的其它方面的影响。

参考文献:

- [1] Boll S. Suppression of Acoustic Noise in Speech Using Spectral Subtraction [J]. IEEE Transactions on Acoustics, Speech and Signal Processing (S0096-3518), 1979, 27(2): 113-120.
- [2] Ephraim Y, Malah D. Speech Enhancement Using a Minimum Mean Square Error Short-time Spectral Amplitude Estimator [J]. IEEE Transactions on Acoustics, Speech and Signal Processing (S0096-3518), 1987, 32(6): 1109-1121.
- [3] Wiener N. Extrapolation, Interpolation, and Smoothing of Stationary Time Series [M]. USA: The MIT Press, 1964.

