

文章编号: 1007-9432(2011)02-0117-04

用小波包改进子空间的语音增强方法

贾海蓉, 张雪英, 牛晓薇
(太原理工大学 信息工程学院, 太原 030024)

摘 要: 针对传统子空间语音增强方法中, 由于不能去除整个噪声子空间而导致语音特征值估计的偏差, 致使增强语音中仍有残留噪声的问题, 提出一种新的用小波包改进的方法, 该方法利用小波包对噪声的抑制功能, 首先对带噪语音进行 KL(Karhunen-Loeve Transform)变换, 得到带噪语音的特征值, 并对该特征值进行 Daubechies8 小波尺度分解, 利用新的改进的软判决阈值函数去除一部分噪声子空间; 然后再在子空间内用统计信息的方法实时跟踪此时噪声特征值, 进一步消除所有噪声子空间, 从而得到最终估计的语音特征值; 最后由 KL 逆变换还原出纯净语音。仿真结果表明, 在输入信噪比相同的情况下, 经过该方法的增强语音的输出信噪比明显高于传统子空间方法, 听觉感受上增强语音也具有更好的清晰度和可懂度。

关键词: 语音增强; 子空间; 小波包; 阈值函数
中图分类号: TN912.3 **文献标识码:** A

语音增强是解决噪声污染的一种有效的预处理技术, 随着数字通信系统的快速发展和广泛使用, 人们越来越重视语音增强在语音处理方面的作用。近年来, 子空间方法在语音增强研究中已有了较大发展^[1,2], 它是通过空间分解将整个空间分为噪声子空间和信号加噪声子空间, 然后通过去除噪声子空间, 并对信号加噪声子空间进行估计来实现语音增强。但是传统的子空间方法, 在低信噪比和非平稳噪声情况下, 噪声的方差用无声段的方差来代替, 增强后的语音常伴有一种具有一定节奏感的残余噪声, 一般称为“音乐噪声”。为了去除音乐噪声, 提高语音质量, 笔者提出一种用小波包改进子空间的语音增强算法。小波包变换是近年来发展起来的去噪新方法, 是从小波分析延伸出来的一种对信号进行更加细致地分析和重构的方法, 不但对低频部分进行分解, 而且对高频部分也做了二次分解, 对信号的分析能力更强, 特别适于处理非平稳信号^[3-7]。该算法采用小波包变换的方法对子空间域中带噪语音特征值进行处理, 即采用改进的小波软判决阈值函数, 得到更新的带噪语音的特征值, 估计出纯净语音特征值, 并采用基于统计信息算法更新噪声特征值, 从而实现了噪声特征值更加准确的估计^[8]。

1 子空间语音增强算法

语音矢量的协方差矩阵有很多零特征值, 这说明纯净语音信号矢量的能量分布在它对应空间的某个子集中。而语音信号处理中, 噪声方差通常都假设已知, 且严格正定。噪声矢量存在于整个带噪信号张成的空间中, 因此带噪语音信号的矢量空间可以认为由一个信号加噪声的子空间和一个纯噪声的子空间构成。可以利用信号子空间处理技术, 消除纯噪声子空间, 并对语音信号进行估计, 实现语音增强。它的原理如图 1 所示。

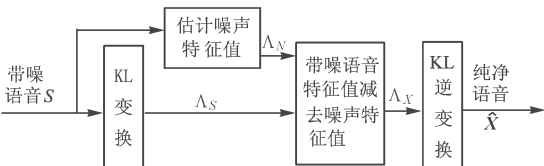


图1 子空间语音增强方法原理图

假设带噪语音信号可以表示为:

$$S = X + N. \tag{1}$$

式中, S 是带噪语音; X 是纯净语音; N 是方差为 σ^2 的高斯白噪声, S, X, N 为 K 维矢量。

首先, 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 U 和特征值对角阵 Λ_S ; 其

次,估计噪声特征值 Λ_N ,一般取带噪语音前 3 000 点求其方差的平均值作为噪声的特征值组成对角阵 Λ_N ;再次,带噪语音特征值 Λ_S 减去噪声特征值 Λ_N ,得到纯净语音特征值 Λ_X ,由 $G = \Lambda_X (\Lambda_X + \mu \sigma_N^2 I)^{-1}$ 可以得到时域约束估计器 (TDC) 增益 G ;最后,由 KL 逆变换可得最优估计器 H_{opt} ,即

$$H_{opt} = U \Lambda_X (\Lambda_X + \mu \sigma_N^2 I)^{-1} U^T,$$

由 $X = H_{opt} S$ 可得增强语音。

从上面的分析可以看出,子空间语音增强算法存在的两个问题使得它存在残留噪声:

- 1) 采用无声段的方差的平均值作为噪声估计的特征值,不能实时跟踪噪声的变化。
- 2) 直接用带噪语音的特征值减去有偏差的噪声特征值,进一步增加了残留噪声。

2 用小波包改进的子空间语音增强算法

2.1 小波包去噪原理

语音和噪声的小波系数在不同尺度下具有相反的性质。对带噪语音信号进行小波变换时,语音信号的小波变换模值随尺度的增加而增加,而噪声则随尺度的增加而减小,经过多层小波分解后,噪声对应的小波系数幅值变小,因此,可以认为较小的小波系数由噪声产生,而较大的小波系数由语音信号产生。通过在各层的不同分辨率下设定不同的阈值门限,调整小波系数,将小于阈值的小波系数归零,并由剩余的小波系数重构信号,达到去噪的目的。

带噪语音信号进行 N 层离散小波包分解,得到一组小波包分解系数 $W_{j,k}$, j 为小波变换的尺度, k 为离散时间,它由语音信号对应的小波包系数 $u_{j,k}$ 和噪声信号对应的小波包系数 $v_{j,k}$ 组成,对于每一个小波包分解系数,选择一个适当的阈值 λ ,并对系数进行阈值量化,得出估计的小波系数 $D_{j,k}$,使得 $\|D_{j,k} - u_{j,k}\|$ 最小。

所以,选择合适的小波阈值 λ 对噪声的消除非常关键。如果选取 λ 太小,那么,施加阈值以后的小波系数中将包含过多的噪声分量; λ 太大,那么将去除一部分语音信号的分量,从而使由小波系数重建后的信号产生过大的失真。因此,在实际工作中,估计阈值的大小尤为关键。常用的估计阈值 λ 的阈值函数为:

1) 硬阈值函数^[8]

$$D_{j,k} = \begin{cases} W_{j,k}, & |W_{j,k}| \geq \lambda, \\ 0, & |W_{j,k}| < \lambda; \end{cases} \quad (2)$$

2) 软阈值函数^[9]

$$D_{j,k} = \begin{cases} \text{sgn}(W_{j,k}) (|W_{j,k}| - \lambda), & |W_{j,k}| \geq \lambda; \\ 0, & |W_{j,k}| < \lambda. \end{cases} \quad (3)$$

但是,硬阈值和软阈值函数存在如下缺陷:

- 1) 硬阈值函数整体不连续,使去噪后的信号中会出现突变的震荡点,尤其是当噪声水平较高时。
- 2) 软阈值函数虽然整体连续性好,但是当小波系数较大时,处理过的系数与原系数之间总存在恒定偏差,直接影响重构信号与真实信号的逼近程度。

笔者提出了一种改进的阈值函数,能克服硬判决和软判决的缺陷,有效减少失真,其阈值函数:

$$D_{i,j} = \begin{cases} W_{j,k}, & |W_{j,k}| \geq \lambda; \\ \tanh(W_{j,k}) \frac{\lambda(e^{2\lambda} - e^{-2\lambda})}{e^{2|W_{j,k}|} - e^{-2\lambda}}, & |W_{j,k}| < \lambda. \end{cases} \quad (4)$$

2.2 新算法描述

小波包改进的语音增强算法主要针对子空间中由于噪声估计的偏差而导致的语音特征值估计的偏差,最终引起语音失真和音乐噪声残留问题。它利用了小波包对语音信号的去噪原理,并结合改进的新软判决阈值函数,在子空间域中对带噪语音的特征值进行小波包分解,得到初步去噪的语音信号的特征值;然后在此基础上用基于统计信息^[6]的非平稳噪声自适应算法估计出噪声特征值,并通过两个特征值相减得到最终语音信号的特征值,最后通过 KL 逆变换恢复出最终的纯净语音。新算法原理如图 2 所示。

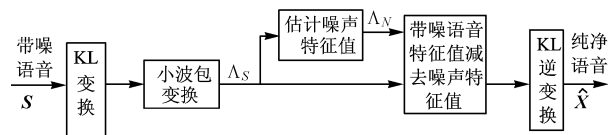


图 2 新方法原理图

新算法的实现步骤为:

- 1) 对带噪语音进行 KL 变换,得到子空间域的带噪语音的特征向量 U 和特征值 Λ_S 。
- 2) 采用小波包变换的方法对带噪语音特征值 Λ_S 进行处理,得到新特征值 Λ_X 。

对带噪语音的特征值采用 8 阶 Daubechies 小波进行小波包分解,得到每一级的小波包分解系数 $W_{j,k}$,采用小波系数在各尺度下绝对值的中值估算噪声方差 σ_j

$$\sigma_j = \frac{\text{median}(|W_{j,k}|)}{0.6745}. \quad (5)$$

利用通用阈值 λ_{long} 计算每一级的节点阈值,并对该节点的小波包分解系数进行改进的软判决函数的处理,更新小波包系数,为

$$\lambda_{\text{long}} = \sigma_j \sqrt{2\ln N} . \tag{6}$$

小波包重建, 估计出纯净语音。

3) 对 Λ_S 采用基于统计噪声的非平稳自适应算法得到噪声特征值 Λ_N , 从而纯净语音特征值为:

$$\Lambda_X = \Lambda_S - \Lambda_N . \tag{7}$$

4) 由 KL 逆变换可得最优估计器 H_{opt} , 由此可得增强语音 X 。

$$H_{\text{opt}} = U\Lambda_X(\Lambda_X + \mu\sigma_N^2I)^{-1}U^T ,$$
$$X = H_{\text{opt}}S . \tag{8}$$

3 实验仿真及分析

为了证明用小波包改进子空间语音增强算法的有效性, 本文对不同信噪比下带噪语音进行增强, 并与传统子空间增强算法在时域波形图和频谱图以及增加的信噪比进行对比。仿真软件采用 MATLAB, 语音取自 863 语音库, 噪声为 NOISEX.92 数据库的高斯白噪声, 语音信号的采样率为 8 kHz, 帧长为 80 个采样点, 帧叠为 50%。

3.1 时域波形图和频谱图的仿真比较

时域波形图和频谱图分别包括纯净语音、0dB 带噪语音、传统的子空间语音增强方法后的语音、本文新方法增强后的语音时域波形图和频谱图, 它们是相互对应的。从理论上来说, 增强后的语音时域波形应该与纯净语音越逼近越好, 证明算法的可行性更高; 从频谱图上来看, 深色表示语音, 浅色代表噪声, 同理, 增强后的频谱图应该与纯净语音越逼近越好。时域波形图和频谱图分别如下图 3, 图 4 所示。

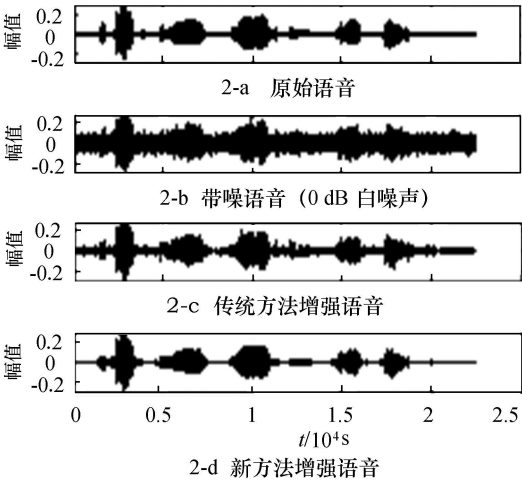


图 3 波形比较图

从时域波形图可以看出, 带噪语音采用本文方法增强后, 噪声得到了明显的抑制, 与原始语音很接近, 而传统方法还存在残留噪声。从频域语谱图可

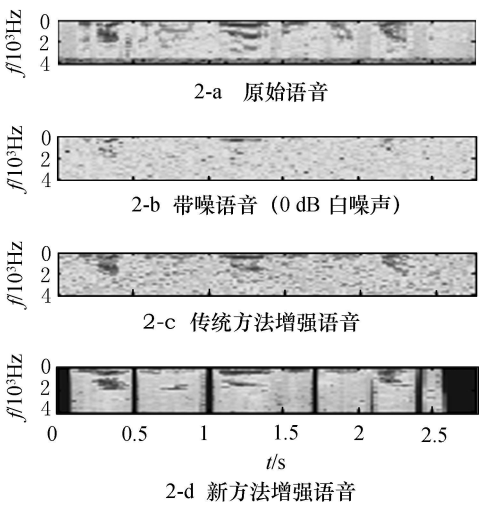


图 4 语谱图比较

以看出, 采用本文方法后得到的增强语音, 语谱图的能量分布已很接近于原始语音, 效果明显优于传统的方法。主观听测也表明, 采用本文方法后, 音乐噪声得到明显抑制, 语音质量和听觉舒适度有了很大提高。

3.2 输出信噪比较

采用在纯净语音中分别加入高斯白噪声, 全局信噪比为 -10, -5, 0, 5, 10, 15 dB 的带噪语音作为测试对象。在输入相同信噪比的情况下, 比较经过本文新方法与传统子空间方法增强后的输出信噪比, 仿真结果如表 1 所示。

表 1 传统方法与本文新方法性能比较

输入信噪比/ dB	传统方法/ dB	本文新方法/ dB
-10	-6.81	-4.827 0
-5	-1.167 6	3.322 8
0	6.510 4	9.417 8
5	11.563 4	12.331 8
10	14.476 8	16.026 8
15	17.802 1	17.935 8

从表 1 的实验结果可以看出, 本文新方法比传统子空间方法信噪比增加明显, 由此证明本文新方法的有效性。

4 结束语

本文将小波包去噪原理应用到子空间域中, 提出了一种改进子空间的语音增强方法。该算法通过改进的软判决函数对子空间域中带噪语音特征值进行处理, 更新带噪语音特征值, 从而实现了语音特征值更加准确的估计。仿真结果和主观听测表明, 本文方法比传统子空间方法语音增强效果明显, 很好地抑制了增强后产生的音乐噪声, 提高了语音质量, 降低了听觉疲劳度。

参考文献:

- [1] Ephraim Y, Van Trees H L. A signal subspace approach for speech enhancement[J]. IEEE Trans Speech and Audio Processing, 1995, 3(4): 251-266.
- [2] 吴北平, 李辉, 戴蓓倩, 等. 基于子空间噪声特征值估计的语音增强方法[J]. 信号处理, 2009, 25(3): 460-463.
- [3] 王文良. 基于小波包变换的语音增强算法研究[D]. 天津: 天津大学, 2007.
- [4] 王侠, 顾明亮. 一种基于小波变换的语音增强新方法[J]. 徐州师范大学学报, 2005, 23(3): 57-60.
- [5] 段其昌, 邓玉娟, 应泽贵. 基于改进阈值函数的小波包语音增强算法的研究[J]. 通信技术, 2009, 42(5): 86-88.
- [6] Donoho D L, Johnstone I M. Ideal spatial adaptation by wavelet shrinkage[J]. Biometrika, 1994, 81(3): 425-455.
- [7] Donoho D L. De-noising by soft thresholding[J]. IEEE Trans On Inform Theory, 1995, 41(3): 613-627.
- [8] Jia Hairong, Zhang Xueying, Jin Chensheng. A modified speech enhancement algorithm based on the subspace[C] //2nd International Symposium on Knowledge Acquisition and Modeling, 2009: 344-347.

A Modified Subspace Speech Enhancement Method Using Wavelet Packet

JIA Hai-rong, ZHANG Xue-ying, NIU Xiao-wei

(College of Information Engineering of TUT, Taiyuan 030024, China)

Abstract: In the traditional subspace speech enhancement method, the noise subspace can not be reduced that leads to the estimating bias of speech eigenvalues, and results in residual noise in enhanced speech. A new improved method using wavelet packet was advanced, which is based on the noise suppression of wavelet packet. Karhunen-Loeve Transform to the noisy speech was conducted firstly, its eigenvalues were attained and then decomposed by db8 wavelet packet, a part of noise subspace was reduced by new modified wavelet soft-decision threshold function. Secondly, in the subspace residual noise eigenvalues were estimated by statistical information method to further remove the noise subspace, and then attain final estimated speech eigenvalues. Lastly, the clean speech was restored by KL inverse transform. The simulation results show that the new method has higher SNR in the output side, the better speech quality than the traditional method in the same input SNR condition.

Key words: speech enhancement; subspace; wavelet packet; threshold function

(编辑: 贾丽红)