

基于小波包与自适应维纳滤波的语音增强算法

董 胡^{1,2}, 徐雨明¹, 马振中¹, 李列文¹, 任 可¹

(1. 长沙师范学院 信息科学与工程学院, 湖南 长沙 410100;

2. 湖南师范大学 物理与电子科学学院, 湖南 长沙 410181)

摘 要: 语音增强主要用来提高受噪声污染的语音可懂度和语音质量, 它的主要应用与在嘈杂环境中提高移动通信质量有关。传统的语音增强方法有谱减法、维纳滤波、小波系数法等。针对复杂噪声环境下传统语音增强算法增强后的语音质量不佳且存在音乐噪声的问题, 提出了一种结合小波包变换和自适应维纳滤波的语音增强算法。分析小波包多分辨率在信号频谱划分中的作用, 通过小波包对含噪信号作多尺度分解, 对不同尺度的小波包系数进行自适应维纳滤波, 使用滤波后的小波包系数重构进而获取增强的语音信号。仿真实验结果表明, 与传统增强算法相比, 该算法在低信噪比的非平稳噪声环境下不仅可以更有效地提高含噪语音的信噪比, 而且能较好地保存语音的谱特征, 提高了含噪语音的质量。

关键词: 语音增强; 小波包; 自适应维纳滤波; 多分辨率分析; 多尺度分解

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2020)01-0050-04

doi: 10.3969/j.issn.1673-629X.2020.01.009

Speech Enhancement Algorithm Based on Wavelet Packet and Adaptive Wiener Filter

DONG Hu^{1,2}, XU Yu-ming¹, MA Zhen-zhong¹, LI Lie-wen¹, REN Ke¹

(1. School of Information Science and Engineering, Changsha Normal University, Changsha 410100, China;

2. School of Physics and Electronics, Hunan Normal University, Changsha 410181, China)

Abstract: Speech enhancement is mainly used to improve the speech intelligibility and speech quality of noise pollution. Its main application is related to improving the quality of mobile communication in noisy environments. Traditional speech enhancement methods include spectral subtraction, wiener filtering and wavelet coefficients. According to the problem of poor speech quality and music noise after speech enhancement for traditional speech enhancement algorithm in complex noise environment, we propose a speech enhancement algorithm combining wavelet packet transform and adaptive wiener filtering. The function of wavelet packet multi-resolution in signal spectrum division is analyzed, and the multi-scale decomposition of noise signal is carried out through wavelet packet. The wavelet packet coefficients of different scales are filtered out by self-adaptive wiener, and then the enhanced speech signal is reconstructed by using wavelet packet coefficients. The experimental simulation shows that this algorithm can not only improve the SNR of speech with noise more effectively under low SNR and non-stationarity noise environment in contrast with the traditional enhancement algorithm, but also can better preserve the spectral features of speech and improve the quality of speech with noise.

Key words: speech enhancement; wavelet packet; adaptive wiener filtering; multi-resolution analysis; multi-scale decomposition

0 引言

语音增强是语音和信号处理领域的一个重要问题, 它对许多基于计算机的语音识别、编码和通信应用都有一定的影响。语音增强的根本目标是提高语音质量和清晰度以被人类监听者感知到。语音增强算法包含传统的谱减法^[1-3]、维纳滤波法^[4-6]、小波系数阈值

法^[7-9]、子空间法^[10-11]及近年来提出的深度神经网络法^[12]等。这些语音增强算法或是基于统计模型或基于语音与噪声的先验信息, 在一定程度上改善了含噪语音的质量。然而, 在复杂噪声环境下, 尤其是在非平稳噪声环境下它们的语音增强性能出现下降。针对上述问题, 文中提出了一种基于小波包和自适应维纳滤

收稿日期: 2019-02-11

修回日期: 2019-06-17

网络出版时间: 2019-09-25

基金项目: 国家自然科学基金(11474090, 11174077, 61502164); 湖南省教育优秀青年项目(17B025); 湖南省自然科学基金青年项目(2018JJ3557); 长沙师范学院院级教学改革项目(P2016014)

作者简介: 董 胡(1982-), 男, 副教授, 研究方向为信号处理及嵌入式开发。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190925.1521.032.html>

波的语音增强算法。

1 小波包变换

小波包变换是一种直观且有效的语音增强方法。语音与噪声的小波包变换所表现的特性截然相反,语音信号小波包变换的模值随小波尺度的增加而递增,但噪声的模值却随小波包尺度的增加而递减。这样,接连多次小波包变换后,噪声对应的小波包系数基本被去除或者幅值非常小,剩余系数主要由语音信号控制。利用较显著的语音小波包系数来重构语音信号,进而较好地去除噪声。

小波包由两组正交小波基滤波器系数生成。若 $\{h_k\}_{k \in \mathbb{Z}}$ 和 $\{g_k\}_{k \in \mathbb{Z}}$ 是一组共轭镜像滤波器(QMF),满足:

$$\sum_{n \in \mathbb{Z}} h_{n-2k} h_{n-2l} = \delta_{kl}, \quad \sum_{n \in \mathbb{Z}} h_n = \sqrt{2} \quad (1)$$

$$g_k = (-1)^k h_{1-k}, \quad l, k \in \mathbb{Z} \quad (2)$$

小波包与小波分解不同,它不仅对低频信号部分作分解,还对高频信号部分作分解,因此小波包是一种比小波更加精细的分解算法^[13]。图1为小波包对一维时间序列分解特性图,其中A代表低频,D代表高频,末尾下标序号代表小波包分解的层数。三层小波包分解关系为:

$$S = AAA_3 + DAA_3 + ADA_3 + DDA_3 + AAD_3 + DAD_3 + ADD_3 + DDD_3 \quad (3)$$

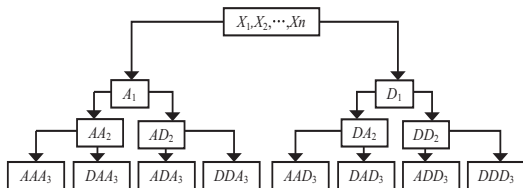


图1 小波包时间序列分解

2 自适应维纳滤波

自适应维纳滤波与变化的局部语音信号有关。自适应维纳滤波原理见图2。在滤波过程中,针对降噪语音 $x(n)$ 的局部平均值 m_x 和局部变量 σ_x^2 作如下处理。

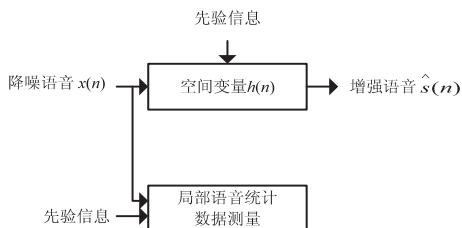


图2 自适应维纳滤波语音增强

假设加性噪声 $v(n)$ 是零均值且与变量 σ_v^2 一样具有白色特性,于是,功率谱 $p_v(w)$ 可近似表达如下:

$$p_v(w) = \sigma_v^2 \quad (4)$$

考虑到 $x(n)$ 中一小段语音信号是平稳的,则 $x(n)$ 可表达如下:

$$x(n) = m_x + \sigma_x w(n) \quad (5)$$

其中, m_x 和 σ_x 分别表示 $x(n)$ 的局部均值和标准偏差; $w(n)$ 表示零均值的单位噪声变量。

采用“直接判别”算法来预估当前帧的先验信息^[14]。在一小段语音中,维纳滤波转换方程可近似作如下表达:

$$H(w) = \frac{P_s(w)}{P_s(w) + P_v(w)} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2} \quad (6)$$

由式6可得维纳滤波脉冲响应:

$$h(n) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2} \delta(n) \quad (7)$$

由式7可知,局部增强的语音信号可表达如下:

$$\hat{s}(n) = m_x + (x(n) - m_x) \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2} \delta(n) = m_x + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2} (x(n) - m_x) \quad (8)$$

假设每一帧语音信号的 m_x 和 σ_s 都得到更新,增强的信号可表达如下:

$$\hat{s}(n) = m_x(n) + (x(n) - m_x(n)) \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2} \quad (9)$$

由式9可得, $x(n)$ 的局部均值变量 m_x 为:

$$\hat{m}_x(n) = \frac{1}{(2M+1)} \sum_{k=n-M}^{n+M} x(k) \quad (10)$$

$2M+1$ 是一小段估计语音信号的采样数。为测量统计的局部信号,定义 $\sigma_x^2 = \sigma_s^2 + \sigma_v^2$, 则可得下述公式:

$$\hat{\sigma}_s^2 = \begin{cases} \hat{\sigma}_x^2(n) - \hat{\sigma}_v^2, & \text{若 } \hat{\sigma}_x^2(n) > \hat{\sigma}_v^2 \\ 0, & \text{若 } \hat{\sigma}_x^2(n) \leq \hat{\sigma}_v^2 \end{cases} \quad (11)$$

$$\hat{\sigma}_x(n) = \frac{1}{(2M+1)} \sum_{k=n-M}^{n+M} (x(k) - \hat{m}_x(n))^2 \quad (12)$$

3 小波包与自适应维纳滤波语音增强算法

小波包变换具有能量集中特性,文中采用 db10 小波包进行多尺度分解。含噪语音信号经过小波包变换时,噪声的能量集中在高频部分且幅值较小的小波包系数上,而语音信号的能量则集中分布在低频部分且幅值较大的小波包系数上。语音信号的小波包系数值大于噪声的小波包系数值,因此通过小波包变换能够实现语音与噪声的分离。通过 db10 小波包的分解将含噪信号作频率谱划分得到不同尺度的小波包系数,然后对各个尺度的系数作自适应维纳滤波增强,最后将增强后的小波包系数进行重构即可得增强后的语音信号。整个语音增强算法的流程见图3。

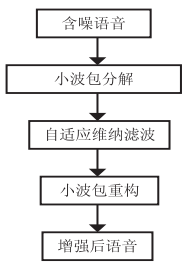
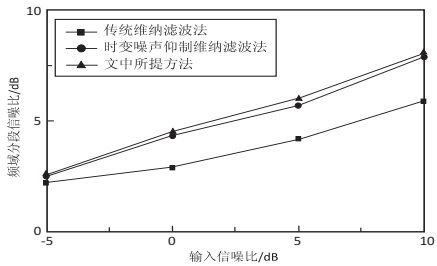


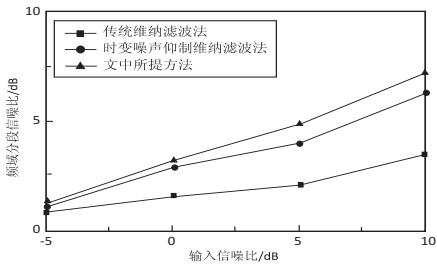
图 3 小波包与自适应维纳滤波语音增强算法流程

4 仿真实验与分析

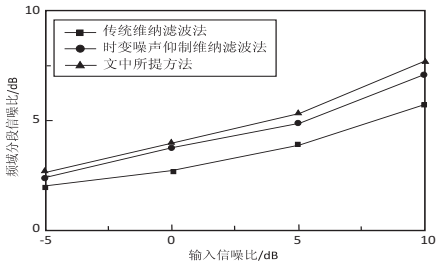
为验证所提算法的增强效果,利用 MATLAB 软件开展仿真实验。实验中使用录制的语音作为纯净语音样本。噪声信号取自 NOISEX-92 数据库中的白噪声、babble 噪声、pink 噪声和 factory 噪声。纯净语音和噪声的采样频率都为 16 kHz,16 bit 量化,将纯净语音与上述不同类型的噪声进行混合,合成信噪比不同的含噪信号。每帧信号帧长为 15 ms,帧移为 50%。为突出所提语音增强算法的有效性,分别将其与传统维纳滤波、文献[15]提出的时变噪声抑制维纳滤波进行比较,计算增强后语音的频域分段信噪比及客观语音质量评估(PESQ)。3 种语音增强算法的频域分段信噪比如图 4 所示,图 4(a)~图 4(d)分别为白噪声、babble 噪声、pink 噪声和 factory 噪声条件下,3 种语音增强算法的频域分段信噪比图。



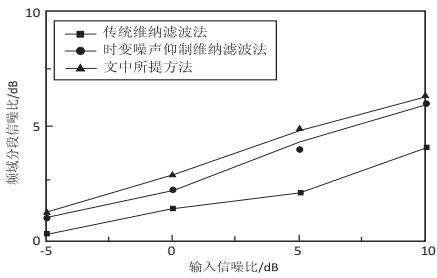
(a) 白噪声



(b) babble 噪声



(c) pink 噪声



(d) factory 噪声

图 4 四种不同噪声下的频域分段信噪比

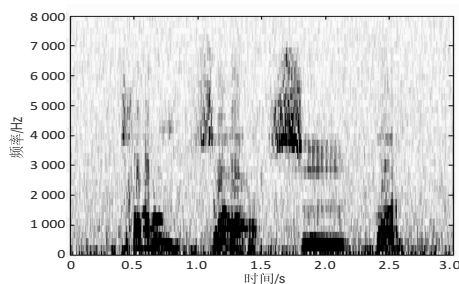
从图 4 可知,与传统维纳滤波算法比较,所提语音增强算法及文献[15]提出的时变噪声抑制维纳滤波算法频域分段信噪比较高,在低信噪比情况下尤为明显。文献[15]与所提语音增强算法相比,在白噪声和 pink 噪声条件下,其分段信噪比效果较接近,然而在 babble 和 factory 的非平稳噪声条件下,所提语音增强算法的分段信噪比较高,平均约高出 0.5 dB。

3 种不同算法语音增强后的 PESQ 如表 1 所示。由于传统维纳滤波在非平稳的噪声(factory、babble 噪声)环境下会对含噪语音产生过分抑制,容易造成语音失真和产生音乐噪声,在表 1 中也分别给出了其他两种语音增强算法的 PESQ 值。经主观听测表明,采用传统维纳滤波语音增强后有较明显的音乐噪声,而采用文中算法和文献[15]算法后都能较好地抑制音乐噪声,其 PESQ 均高于传统的维纳滤波。其中,文中的语音增强算法的 PESQ 亦高于文献[15]提出的语音增强算法。

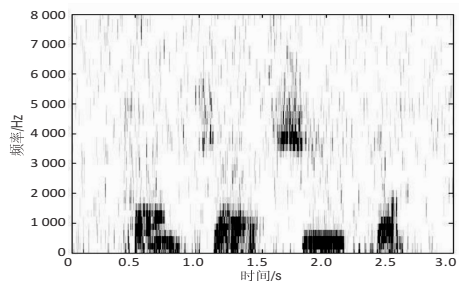
表 1 3 种语音增强算法 PSEQ 比较

噪声类型	语音增强算法	PESQ			
		10	5	0	-5
白噪声	传统维纳滤波算法	2.31	1.98	1.67	1.24
	文献[15]算法	2.57	2.10	1.85	1.46
	文中算法	2.69	2.34	2.12	1.71
Babble 噪声	传统维纳滤波算法	2.39	1.90	1.36	1.09
	文献[15]算法	2.64	2.03	1.46	1.18
	文中算法	2.58	2.25	1.63	1.37
Pink 噪声	传统维纳滤波算法	2.48	2.14	1.79	1.31
	文献[15]算法	2.66	2.26	1.95	1.68
	文中算法	2.82	2.47	2.31	2.05
factory 噪声	传统维纳滤波算法	2.73	2.38	1.92	1.49
	文献[15]算法	2.81	2.49	2.04	1.65
	文中算法	2.97	2.63	2.35	2.07

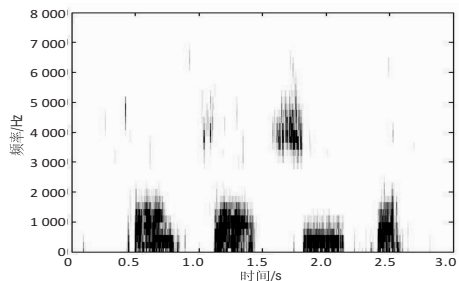
图 5 比较了 3 种语音增强算法的语谱图。其中,图 5(a)是含 pink 噪声(SNR = -5 dB)语谱图,图 5(b)、图 5(c)和图 5(d)分别为经过传统维纳滤波、文献[15]算法及文中算法的语音增强语谱图。



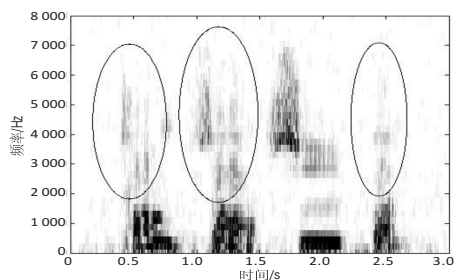
(a) 含-5 dB pink 噪声语谱图



(b) 传统维纳滤波算法



(c) 文献[15]算法



(d) 文中算法

图5 三种语音增强算法语谱图对比

从图5可知,与图5(b)相比较,图5(c)与图5(d)在较好地去除背景 pink 噪声的同时,更好地保留了语音的谐波成分,而图5(b)则有较明显的语音失真。图5(d)与图5(c)相比,图5(d)保留了较弱的语音部分的谱特征,提高了增强后的语音质量。

5 结束语

提出的小波包和自适应维纳滤波算法结合了小波包和自适应维纳滤波两种算法的优点,利用小波包变换对非平稳的含噪信号作分解处理从而实现噪声和语音的初级分离,然后利用自适应维纳滤波作进一步降噪处理,在较低信噪比的非平稳噪声环境下,对语音的损伤相对较小,较好地抑制了音乐噪声的产生,语音增

强效果较明显。文中算法比传统维纳滤波算法、文献[15]算法具有更高频域分段信噪比及 PESQ,且主观试听上也无明显的音乐噪声。因此,该算法在降低音乐噪声的基础上提升了语音增强效果。

参考文献:

- [1] 董 胡. 低信噪比环境下改进的语音端点检测算法[J]. 计算机技术与发展,2016,26(3):71-74.
- [2] 徐 娜,吴长奇. 结合差分阵列与幅度谱减的双麦语音增强算法[J]. 信号处理,2018,34(7):876-881.
- [3] 蔡 军,李 飞,张 毅. 基于听觉掩蔽效应的语音增强算法[J]. 计算机工程,2017,34(7):288-292.
- [4] 肖 强,曾庆宁,王 瑶,等. 基于 MGSC 与改进维纳滤波的麦克风阵列语音增强[J]. 声学技术,2017,36(6):567-573.
- [5] 孟 欣,马建芬,张雪英. 改进的参数自适应的维纳滤波语音增强算法[J]. 计算机工程与设计,2017,38(3):714-718.
- [6] 王 栋,贾海蓉. 改进相位谱补偿的语音增强算法[J]. 西安电子科技大学学报:自然科学版,2017,44(3):83-88.
- [7] 唐 鹏,郭宝平. 改进型阈值函数寻优法的小波去噪分析[J]. 信号处理,2017,33(1):102-110.
- [8] 刘冲冲,邹 翔,周正仙. 基于新阈值函数和自适应阈值的小波包语音增强研究[J]. 计算机应用研究,2017,34(11):3324-3328.
- [9] 钟建军,宋 健,由长喜,等. 基于信噪比评价的阈值优选小波去噪法[J]. 清华大学学报:自然科学版,2014,54(2):259-263.
- [10] KIM D K, CHANG J H. A subspace approach based on embedded prewhitening for voice activity detection[J]. Journal of the Acoustical Society of America, 2011, 130(5):304-314.
- [11] 孙成立,穆俊生. 基于特征值置换的子空间语音增强算法[J]. 计算机工程,2016,42(2):272-277.
- [12] KOLBAEK M, TAN Z H, JENSEN J. Speech intelligibility potential of general and specialized deep neural network based speech enhancement systems[J]. IEEE/ACM Transactions on Audio, Speech, & Language Processing, 2017, 25(1):153-167.
- [13] SANAM T F, SHAHNAZ C. Noisy speech enhancement based on an adaptive threshold and a modified hard thresholding function in wavelet packet domain[J]. Digital Signal Processing, 2013, 23(3):941-951.
- [14] SCALART P, FILHO J V. Speech enhancement based on a priori signal to noise estimation[C]//1996 IEEE international conference on acoustics, speech, and signal processing conference proceedings. Atlanta, GA, USA: IEEE, 1996: 629-632.
- [15] ARSLAN L M. Modified Wiener filtering[J]. Signal Processing, 2006, 86(2):267-272.