

文章编号: 1001-0920(2009)06-0849-05

一种时间自适应阈值的小波包改进语音增强算法

田 岚^{1,2}, 侯正信¹, 孙晋松²

(1. 天津大学 电子信息工程学院, 天津 300072; 2. 山东大学 信息科学与工程学院, 济南 250100)

摘 要: 针对传统小波语音增强算法存在过度阈值处理的问题, 提出一种改进的时间自适应阈值小波包去噪算法. 该方法采用听觉感知小波包对噪声语音进行分解, 得到小波包听觉感知节点上的系数, 并基于语音存在概率估计按帧自动调节去噪阈值. 因改进的阈值能更好地避免语音小波包系数被过度阈值处理的情况, 从而在抑制噪声的同时保留了更多的原始语音成分, 进一步提高了降噪效果. 实验结果表明, 该算法比常规小波自适应阈值算法能得到更清晰的语音增强信号.

关键词: 语音增强; 小波包; 自适应阈值

中图分类号: TN912.35

文献标识码: A

Improved speech enhancement algorithm based on time-adaptive threshold wavelet packet transform

TIAN Lan^{1,2}, HOU Zheng-xin¹, SUN Jin-song²

(1. School of Electronic and Information Engineering, Tianjin University, Tianjin 300072, China; 2. School of Information Science and Engineering, Shandong University, Ji 'nan 250100, China. Correspondent: TIAN Lan, E-mail: tianlan65@sdu.edu.cn)

Abstract: For the problem of over threshold for conventional wavelet speech enhancement method, this paper presents an improved time adaptive threshold (TAT) wavelet packet speech enhancement algorithm. By using the perceptual wavelet packet, the noisy speech signal is decomposed into wavelet packet node coefficients. For each frame, the speech-present probability estimation is calculated. The cutting noise TAT is adjusted with frame so that the recovered signal can contain more precise speech signal. Experimental results show that the enhanced speech produced by the improved algorithm has clearer and better speech quality than that by the usual TAT wavelet enhancement algorithms.

Key words: Speech enhancement; Wavelet packet; Time adaptive threshold

1 引 言

在语音识别系统中, 语音信号特征的准确提取对于识别结果起到重要的作用. 但实际语音信号中会不同程度地受到周围噪声的干扰, 从而使得识别系统很难准确地提取信号的特征参数. 语音增强是解决噪声污染的有效办法, 其主要任务是从带噪语音中恢复出原始纯净语音信号, 改善语音质量. 因此需要对语音识别系统采集到的语音信号进行语音增强预处理, 抑制噪声, 以便进一步处理.

小波是一种灵活的时频分析技术, 能够对信号进行有效的时频分解, 适合对语音这类非平稳信号进行处理, 其中基于小波变换的语音增强技术有了

很大的发展. 根据语音信号和噪声在小波分析尺度上分布特性的不同, Donoho 等^[1]提出了一种基于小波的阈值去噪方法, 其仅使用大于预定阈值的小波系数来重构原始信号, 从而达到抑制噪声的目的. 这种去噪方法中使用的阈值为全局阈值, 没有考虑噪声在不同尺度上的差异. 为了进一步提高去噪效果, Johnstone 等^[2]又提出一种与小波分解尺度相关的阈值设置方法. 这种尺度相关的小波阈值去噪方法可以改进恢复的原始语音信号. 但无论是全局阈值还是尺度相关的阈值, 都是随时间不变的, 因此在传统的基于小波变换的语音增强中, 常会出现过度阈值处理的现象, 即在抑制噪声的同时, 将语音信号的

收稿日期: 2008-07-09; 修回日期: 2008-11-03.

基金项目: 高等学校博士学科点专项科研基金项目(20060056051).

作者简介: 田岚(1965—), 女, 济南人, 教授, 博士生, 从事语音信号处理及应用等研究; 侯正信(1935—), 男, 天津人, 教授, 博士生导师, 从事信号及图像处理等研究.

一些有用成分也去除了. 为了减少这种情况, Chen 等^[3] 提出了一种基于 TEO (Teager energy operator) 的小波包去噪算法. 首先对带噪语音的小波包分解系数进行 TEO 运算; 然后根据 TEO 结果的包络来调节阈值, 使得阈值在语音出现时具有相对较小的值而在噪声出现时保持较高的值, 从而实现小波阈值的时间自适应调节. 相比传统的小波去噪算法, 基于 TEO 的语音增强方法在去除噪声的同时, 更多地保留了原始语音信号的成分. 但是由于这种阈值的时间自适应性依赖于小波系数 TEO 结果的包络, 而这种包络的变化比较缓慢, 仍存在部分过度阈值处理的情况.

为进一步减少语音小波系数的过度阈值处理, 本文对带噪语音信号采用听觉感知小波包分解 (PYPD), 并基于小波包节点的时间帧能量来估计语音出现的概率, 进而利用该概率值调节去噪阈值, 从而得到小波包去噪的时间自适应阈值 (TAT), 以期在抑制噪声的同时, 减少语音信号有效成分的丢失, 从而获得更好的语音增强效果.

2 小波包变换理论

小波分析可以对信号进行有效的时频分析, 但因其对信号频段进行二进倍频程的划分, 故在高频段其频率分辨率较差, 而低频段的时间分辨率较差. 小波包分析则能提供一种更加精细的分析方法, 它将频带进行多层次划分, 对小波分析没有进一步划分的高频成分进行分解, 并根据信号分析的特点, 选择合适的小波包分解树, 从而提高了时频分辨率, 因此小波包具有更广泛的应用^[4].

令正交小波基的滤波器系数为 h_n 和 g_n , 尺度函数和小波函数分别为 $\phi(t)$ 和 $\psi(t)$, 则称由公式

$$\begin{cases} w_n = \sqrt{2} \sum_k h_k u_n(2t - k), \\ w_{n+1} = \sqrt{2} \sum_k g_k u_n(2t - k) \end{cases}$$

定义的函数集合 $\{u_n(t)\}_{n \in \mathbb{Z}}$ 为由 $u_0(t) = \phi(t)$ 所

确定的小波包^[5]. 因此, 小波包 $\{u_n(t)\}_{n \in \mathbb{Z}}$ 是包括尺度函数和小波函数在内的具有一定联系的函数的集合.

在本文所提出的语音增强算法中, 听觉感知小波包分解被用来对语音信号进行分析. 人耳内耳的基底膜具有与频谱分析仪相似的作用, 在 20 ~ 16000 Hz 上可以分成 24 个临界带, 每一个临界带对应一个 Bark 尺度, 而且在相同临界带内的频率信号在大脑中是叠加在一起进行评价的^[6]. 利用小波包灵活的频率划分方法设计的与临界带相似的小波包分解结构, 称其为听觉感知小波包分解^[7]. 听觉感知小波包分解已成功应用于语音信号处理的许多领域中. 实验证明, 相对于传统的小波包分解方式, 采用听觉感知小波包对语音信号进行分析能够得到更好的处理结果^[8]. 因此, 为了使增强后的语音具有更好的音质, 本文采用听觉感知小波包对语音信号进行分解. 对于采样频率为 16 kHz 的信号, 其频率范围对应标号为 1 ~ 21 的临界带. 图 1 给出了对应的听觉感知小波包分解树. 其中 $w_{j,k}$ ($j = 1, 2, \dots, 8, k = 0, 1, \dots, 2^j - 1$) 代表第 j 层小波包分解的第 k 个子带的系数, 而相邻的末端节点的中心频率相差 $1/4$ Bark 频率间隔.

3 自适应阈值的设定

文献[9]提出了一种信号功率谱的估计方法, 能够很好地跟踪噪声功率谱的变化. 本文则基于这种功率谱的估计方法, 利用小波包分解节点系数的时间帧的能量来估计信号为语音帧的概率, 进而利用此概率值调节小波包的去噪阈值.

设小波包分解的系数为 $w_{j,k}(n)$, 小波包系数的时间帧能量表示为 $E(\cdot)$, 其中 \cdot 为帧号, 每一时间帧的长度为 10 ~ 40 ms, 半帧重叠. 首先对每一帧的能量利用下式进行一阶平滑处理:

$$E_p(\cdot) = E_p(\cdot - 1) + (1 - \alpha) E(\cdot), \quad (1)$$

其中 $E_p(\cdot)$ 为平滑后的能量. 然后根据下式获得信号能量的局部最小值 E_{\min} :

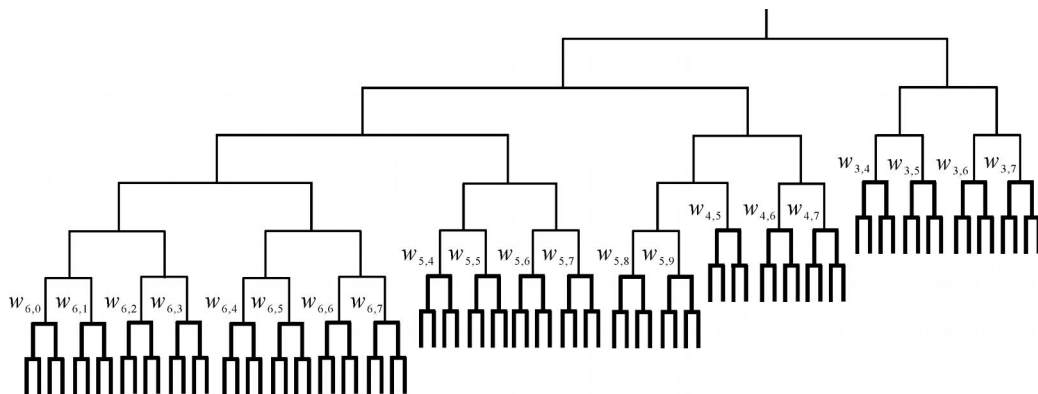


图 1 听觉感知小波包分解树

如果 $E_{min}(n-1) < E_P(n)$,

则 $E_{min}(n) = E_{min}(n-1) +$

$$\frac{1}{1-\alpha} (E_P(n) - E_P(n-1));$$

否则 $E_{min}(n) = E_P(n)$. (2)

如果一帧小波包系数主要为噪声系数, 则其能量 $E_P(n)$ 与对应的局部能量最小值 $E_{min}(n)$ 相差较小; 而当该帧小波包系数为语音信号的分解系数时, 则 $E_P(n)$ 与 $E_{min}(n)$ 具有较大的差别. 因此, 可通过计算 $E_P(n)$ 与局部最小能量 $E_{min}(n)$ 的比值来确定该时间帧的小波包系数主要是语音信号的分解系数还是噪声的分解系数.

设

$$S_r(n) = E_P(n) / E_{min}(n), \quad (3)$$

若 $S_r(n)$ 大于门限值, 则认为该帧为语音帧; 否则该帧为噪声帧, 并同时得到估计的语音帧出现概率 $p(n)$. 即

如果 $S_r(n) > \tau$,

则 $p(n) = \alpha p(n-1) + (1-\alpha)$;

否则 $p(n) = \alpha p(n-1)$. (4)

其中 α 为平滑算子. 于是该帧内的小波包系数的语音存在概率 $P(n)$ 为 $p(n)$, 即

$$P(n) = p(n), \quad n = (n-1) \frac{M}{2} \sim \frac{M}{2} - 1,$$

M 为每一帧的数据长度.

根据每个小波包节点的系数获得时间自适应阈值为

$$\text{Thr}_{j,k}(n) = |c_{j,k}| \sqrt{2 \log(N)} * (1 - P(n)). \quad (5)$$

其中: $|c_{j,k}| = \frac{\text{MAD}_{j,k}}{0.6745}$, $\text{MAD}_{j,k}$ 为第 j 层第 k 个子带的小波包系数的绝对值的中值; N 为信号的长度.

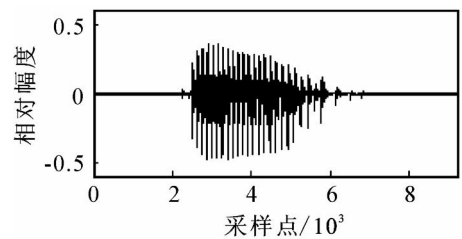
图 2 为采用 TEO 与本文方法根据带噪声语音的小波包分解系数所获得的自适应阈值. 其中: 噪声类型为白噪声, 信噪比为 5 dB. 因为语音存在概率值相比传统的 TEO 能更准确地跟踪语音信号的小波包系数的分布, 所以能够更好地实现阈值的时间自适应性.

4 实验步骤

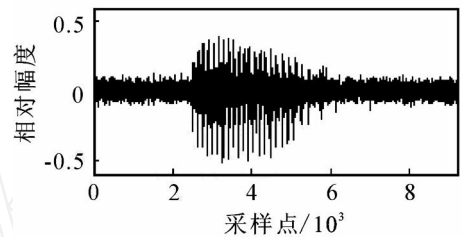
本文所用语音增强算法如图 3 所示. 算法的具体描述如下:

Step1: 对于带噪语音信号, 按照图 1 所示的听觉感知小波包分解方式进行分解, 其中黑色实线代表未进行二抽取的小波包分解.

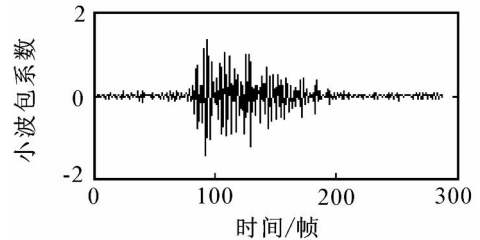
Step2: 对小波包分解的节点进行语音帧存在概率估计, 得到对应序列 $P(n)$, 其中使用的参数可根据实验具体测定. 本文所使用的参数设置如下:—



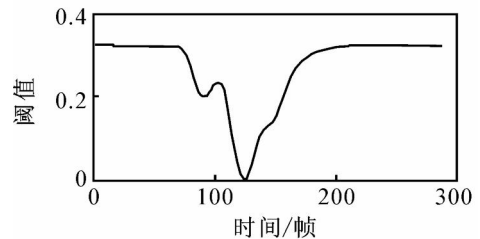
(a) 纯净语音



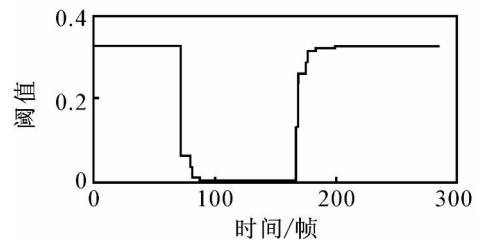
(b) 带白噪声的语音(信噪比=5dB)



(c) 小波包分解



(d) 基于TEO的阈值



(e) 本文方法得到的阈值

图 2 小波包节点的时间自适应阈值对比

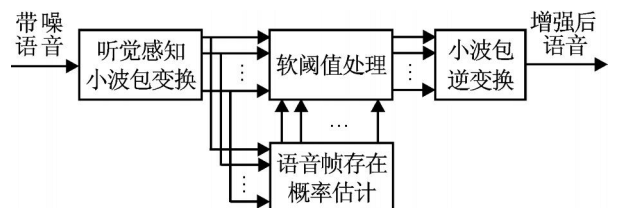


图 3 语音增强算法

帧信号的长度为 32 ms, $\alpha = 0.3$, $\tau = 0.998$, $\beta = 0.8$, $M = 8$, $\alpha = 0.2$.

Step3: 根据每个小波包节点的系数得到时间

自适应阈值,并利用时间自适应阈值对小波包分解系数 $w_{j,k}(n)$ 进行软阈值处理. 设阈值处理后的小波包分解系数为 $w_{j,k}(n)$, 则

$$w_{j,k}(n) = \begin{cases} \text{sgn}(w_{j,k}(n)) (|w_{j,k}(n) - \text{Thr}_{j,k}(n)|), & \\ w_{j,k}(n) & \text{Thr}_{j,k}; \\ 0, & w_{j,k}(n) < \text{Thr}_{j,k}. \end{cases}$$

Step4: 对 $w_{j,k}(n)$ 进行小波包逆变换, 得到增强后的语音信号.

5 实验结果

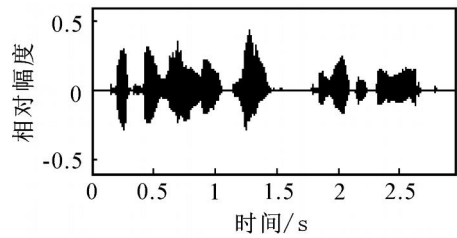
实验中所用到的纯净语音信号和噪声的采样频率均为 16kHz, 采样精度为 16bits, 其中噪声采用 Noisex-92 噪声库中的白噪声和 f16 战斗机噪声. 将一段纯净的语音信号叠加上信噪比 SNR = -3dB 的白噪声, 并分别使用传统小波包去噪算法、基于 TEO 算子的语音增强算法以及本文算法对带噪语音进行处理(其中小波函数选择为 sym6 小波), 其处理结果如图 4 所示.

从图 4 中可以看出, 传统的小波包去噪算法的过阈值处理现象比较严重, 许多有用的语音信号成分都被去除; 而基于 TEO 的语音增强方法则在去除噪声的同时, 保留了更多的语音信号的成分; 但相比较而言, 本文算法则具有更好的性能, 不但更好地抑制了噪声, 而且基本上保留了原始信号的成分, 主观测听增强后的语音, 其音质更清晰. 表 1 给出了噪声为白噪声条件下上述 3 种语音增强算法的结果对比.

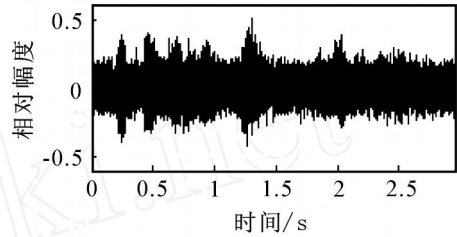
表 1 白噪声下 3 种语音增强的效果对比 dB

| 原始信噪比 | 传统小波包去噪算法 | 基于 TEO 的语音增强算法 | 本文算法 |
|-------|-----------|----------------|---------|
| 5 | 8.1194 | 8.84 | 11.0639 |
| 3 | 6.942 | 7.8544 | 9.9894 |
| 0 | 5.2365 | 6.2627 | 8.4736 |
| -3 | 3.7642 | 5.0652 | 6.8728 |
| -5 | 3.1548 | 4.2667 | 5.7757 |

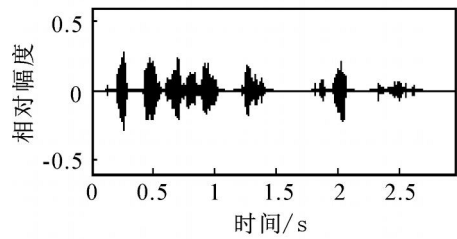
除白噪声外, 实验中还采用 f16 噪声对原始语音信号进行不同程度的干扰, 然后利用 3 种不同的语音增强方法进行增强处理. 表 2 给出了在 f16 噪声情况下得到增强后的语音信噪比. 从表 2 中可以看出, 相比传统的算法, 本文算法能更好地从非平稳噪声中恢复出纯净的语音信号, 提高了增强后的语音质量.



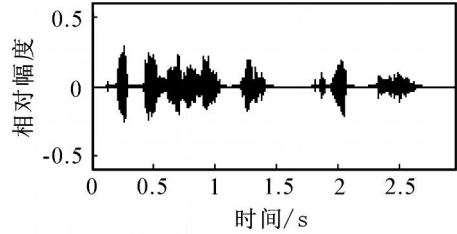
(a) 纯净语音信号



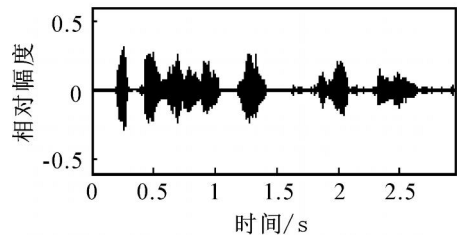
(b) 带噪语音信号(SNR=-3dB白噪声)



(c) 传统小波包的语音增强



(d) 基于TEO算子的语音增强



(e) 本文方法语音增强

图 4 3 种语音增强算法的图形对比

表 2 f16 噪声下 3 种语音增强的效果对比 dB

| 原始信噪比 | 传统小波包去噪算法 | 基于 TEO 的语音增强算法 | 本文算法 |
|-------|-----------|----------------|--------|
| 5 | 5.352 | 6.4352 | 8.5957 |
| 3 | 4.5454 | 5.469 | 7.3969 |
| 0 | 3.2267 | 3.9791 | 5.5539 |
| -3 | 2.2596 | 2.2287 | 3.5569 |
| -5 | 1.7329 | 0.75795 | 2.1663 |

6 结 论

本文利用与人耳听觉感知特点相符合的听觉感知小波包变换对含噪语音信号进行分解,并利用能够更好跟踪小波包系数分布的语音存在概率调节去噪阈值,再通过软阈值处理和小波包逆变换获得增强后的语音,实现了一种改进的时间自适应阈值的小波包语音增强方法.实验结果表明,与传统的小波包去噪方法和基于 TEO 的语音增强方法相比,本文所提出的算法能够更好地抑制噪声,以恢复更多的有效语音信息.因此,将该算法置于语音识别系统的前端,有利于语音识别系统更好地提取语音信号的特征参数,进而提高识别系统的抗噪性.

参考文献(References)

- [1] Donoho D, Johnstone I M. Ideal spatial adaptation by wavelet shrinkage[J]. *Biometrika*, 1994, 81(3): 425-455.
- [2] Johnstone I M, Silverman B W. Wavelet threshold estimators for data with correlated noise[J]. *J of the Royal Statistical Society*, 1997, 59(2): 319-351.
- [3] Chen S H, Wang J F. Speech enhancement using perceptual wavelet packet decomposition and teager energy operator[J]. *J of VLSI Signal Processing*, 2004, 36(2/3): 125-139.
- [4] 成礼智, 郭汉伟. 小波与离散变换理论及工程实践[M]. 北京: 清华大学出版社, 2005.
(Cheng L Z, Guo H W. Wavelet and discrete transform theory and application [M]. Beijing: Tsinghua University Press, 2005.)
- [5] 彭玉华. 小波变换与工程应用[M]. 北京: 科学出版社, 1999.
(Peng Y H. Wavelet transform and engineering application[M]. Beijing: Science Press, 1999.)
- [6] 易克初, 田斌, 付强. 语音信号处理[M]. 北京: 国防工业出版社, 2000.
(Yi K C, Tian B, Fu Q. Speech signal processing[M]. Beijing: National Defence Industry Press, 2000.)
- [7] Pinter I. Perceptual wavelet-representation of speech signals and its application to speech enhancement[J]. *Computer Speech and Language*, 1996, 10(2): 1-22.
- [8] Cohen I. Enhancement of speech using bark scaled wavelet packet decomposition [C]. Proc of the 7th European Conf on Speech, Communication and Technology. Aalborg, 2001: 1933-1936.
- [9] Sundarajan R, Philipos C L. A noise-estimation algorithm for highly non-stationary environment [J]. *Speech Communication*, 2006, 48(2): 220-231.
- [10] 陈文钢, 田岚, 姜晓庆, 等. 一种噪声谱快速跟踪的语音增强方法[J]. *山东大学学报*, 2006, 36(4): 26-28.
(Chen W G, Tian L, Jiang X Q, et al. A speech enhancement method based on quickly tracking the noise spectrum[J]. *J of Shandong University*, 2006, 36(4): 26-28.)
- [11] 王雪, 姜爱国, 王晟. 无线传感器网络的改进 GASA 优化设计[J]. *控制理论与应用*, 2006, 23(4): 593-596.
(Wang X, Jiang A G, Wang S. Optimal designs of wireless sensor network by adapted GASA[J]. *Control Theory and Applications*, 2006, 23(4): 593-596.)
- [12] Whitley D. The genetic algorithm and selection pressure: Why rank-based allocation reproduction trials is best [C]. Proc of the 3rd Int Conf on Genetic Algorithm. Los Altos: Morgan Kaufmann Publishers, 1989: 116-121.
- [13] Yee Leung, Gao Y, Xu Z B. Degree of population diversity—A perspective on premature convergence in genetic algorithms and its Markov chain analysis[J]. *IEEE Trans on Neural Network*, 1997, 8(5): 1165-1175.
- [14] 罗小平. 人工免疫遗传学习算法及其工程应用研究[D]. 杭州: 浙江大学, 2002.
(Luo X P. A study on immune genetic algorithm and its application [D]. Hangzhou: Zhejiang University, 2002.)

(上接第 848 页)