

基于子带谱熵的仿生小波语音增强

刘 艳*, 倪万顺

(大连大学 信息工程学院, 辽宁 大连 116622)

(* 通信作者电子邮箱 linda_whb@163.com)

摘 要: 前端噪声处理直接关系到语音识别的准确性和稳定性, 针对小波去噪算法所分离出的信号不是原始信号的最佳估计, 提出一种基于子带谱熵的仿生小波变换(BWT)去噪算法。充分利用子带谱熵端点检测的精确性, 区分含噪语音部分和噪声部分, 实时更新仿生小波变换中的阈值, 精确地区分出噪声信号小波系数, 达到语音增强目的。实验结果表明, 提出的基于子带谱熵的仿生小波语音增强方法与维纳滤波方法相比, 信噪比(SNR)平均提高约8%, 所提方法对噪声环境下语音信号有显著的增强效果。

关键词: 语音增强; 子带谱熵; 仿生小波变换; 去噪; 阈值

中图分类号: TN912.3 **文献标志码:** A

Speech enhancement based on bionic wavelet transform of subband spectrum entropy

LIU Yan*, NI Wanshun

(College of Information Engineering, Dalian University, Dalian Liaoning 116622, China)

Abstract: Front end noise processing has a direct impact upon the accuracy and stability of the speech recognition. According to the fact that the signal separated by wavelet denoising algorithm isn't its optimal estimation, a novel Bionic Wavelet Transform (BWT) de-noising algorithm based on subband spectrum entropy was proposed. To achieve the purpose of speech enhancement, the subband spectrum entropy, which has a good accuracy of the endpoint detection, was taken full advantage to distinguish the parts of speech and noise, to real-time update the threshold of BWT, and to precisely determine the noise signal wavelet coefficients. The experimental results indicate that the Signal-to-Noise Ratio (SNR) of the proposed algorithm is 8% higher than the Wiener filter algorithm. The proposed method has significant enhancement effect on speech signal in noisy environments.

Key words: speech enhancement; subband spectrum entropy; Bionic Wavelet Transform (BWT); denoising; threshold

0 引言

在自然环境中, 语音信号不可避免地会受到周围环境、传输媒介等带来的噪声影响。这些干扰信号使得语音信号不是纯净的原始信号。而语音训练模板是在高保真环境下录制的, 这就使得语音信号和训练模板很难匹配, 从而使语音识别率降低。语音增强就是要从含噪的语音中提取出尽可能纯净的原始信号, 提高语音识别率。

近年来, 语言增强的主要技术包括小波滤波技术^[1]、Wiener滤波^[2]、自适应滤波器^[3]和能量熵^[4], 它们在噪声去除方面都取得了一定效果。Donoho等^[5]在Mallat的小波理论指导下, 将小波分析应用到图像、语音等领域, 从此小波方法成为语音增强的一个重要研究方向。小波阈值去噪方法, 已被证明效果好于其他方法。小波阈值方法中最关键的是阈值的选取, 硬阈值函数对噪声的去除不够干净, 软阈值函数能彻底去除噪声, 但对语音信号的损害很大^[6]。由于小波去噪不能很好地反映人耳对语音信号幅度和频率的听觉特性, 2001年Yao等^[7]将小波变换和人耳的非线性传输模型相结合, 提出了仿生小波变换(Bionic Wavelet Transform, BWT)去噪方法。

针对现有小波去噪方法中存在的不足, 本文将子带谱熵

端点检测方法与BWT结合, 提出一种基于子带谱熵的仿生小波语音增强方法。采用基于子带谱熵端点检测算法, 从信号中分离出噪声信号, 计算出噪声标准方差。然后, 将经子带谱熵所分离的语音信号进行仿生小波动态硬阈值去噪, 以达到良好的语音增强效果。

1 子带谱熵端点检测

1.1 谱熵

Shen等^[8]将谱熵引入到语音信号端点检测中。由于信息熵的值只和语音信号能量的随机性有关, 与能量幅值无关^[9], 语音信号和噪声信号的熵存在着很大的差异。在一段频带较宽的语音信号中, 相对于噪声信号而言, 语音信号主要集中于某几个频段(主要集中在300~3400 Hz), 其动态范围广, 随机事件多, 平均信息量大, 对应的熵值大; 而噪声信号幅度小, 分布相对集中, 相应的熵值较小, 因此, 能比较好地区分语音段和非语音段。

谱熵原理:

对带噪语音信号分帧、加窗, 帧间重叠50%进行快速傅里叶变换(Fast Fourier Transformation, FFT), 得到频率分量

收稿日期: 2014-10-17; 修回日期: 2014-11-25。 基金项目: 辽宁省教育厅科学计划项目(L2013463)。

作者简介: 刘艳(1967-), 女, 辽宁大连人, 副教授, 博士, 主要研究方向: 电力传动系统建模、故障诊断与容错控制; 倪万顺(1989-), 男, 江苏泰州人, 硕士研究生, 主要研究方向: 模式识别、语音信号处理。

的能量谱 $s_{\omega}(i, j)$, 每个频率分量的归一化谱概率密度函数定义为:

$$P(i, j) = \frac{s_{\omega}(i, j)}{\sum_{j=0}^{N-1} s_{\omega}(i, j)} \quad (1)$$

其中: $s_{\omega}(i, j)$ 表示第 i 帧语音信号第 j 频率分量的功率谱幅度。 $s_{\omega}(i, j) = 0$, 如果频率 $f_j < 300$ Hz 或 $f_j > 3400$ Hz。

第 i 帧语音信号短时谱熵表达式如式(2)所示:

$$H_a(i) = - \sum_{j=1}^n p(i, j) \lg p(i, j) \quad (2)$$

1.2 子带谱熵

由于每一谱点的幅值容易受到噪声的干扰, 影响端点的检测。Wu 等^[10]提出子带谱熵的概念将每一帧语音信号分成若干个子带, 对每一个子带求谱熵, 这样就解决了幅值受噪声影响的问题。

子带谱熵^[11]表达式如式(3)所示:

$$H_b(i) = \sum_{m=1}^{N_b} P_b(m, i) * \lg[1/P_b(m, i)] \quad (3)$$

子带能量概率表达式如式(4)所示:

$$P_b(m, i) = \frac{E_b(m, i)}{\sum_{k=1}^N E_b(k, i)} \quad (4)$$

其中: $i \leq m \leq N$, $E_b(m, i)$ 表示第 m 个子带的能量, i 表示帧数, N 表示每一帧语音信号所划分的子带数。根据式(3)、(4)仿真得到的子带谱熵如图1所示。由图可以看出, 语音部分的谱熵值明显高于噪声部分的, 噪声部分的熵值平均在 0 ~ 0.5。

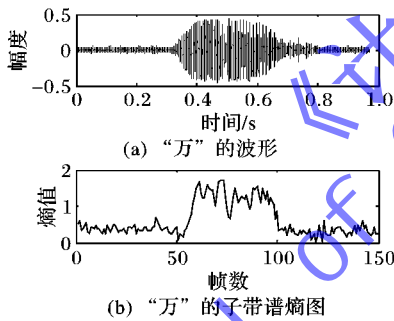


图1 “万”的波形和子带谱熵

2 仿生小波变换

人的耳朵由外耳、中耳和内耳组成, 中耳和内耳对接收到的声波信号进行增强和补偿, 最终由内耳耳蜗的基底膜接收^[12]。耳蜗基底膜对不同频率的声波信号具有选择性, 对低频信号的分辨能力强于高频信号。Yao 等^[7]将主动耳蜗理论与小波变换结合, 提出一种时频分析方法——仿生小波变换。仿生小波变换与基本小波变换相比, 它在时频域的尺度不仅可以按照语音信号的频率进行调节, 而且可以随着信号的幅度和一阶微分系数进行自适应调节。

2.1 仿生小波变换理论

小波变换是将语音信号在多个尺度上进行分解, 得到各尺度上的子波变换系数。在基本的小波变换中, 母小波函数 $h(t)$ 必须满足容许条件, 所以可以用其包络函数 $\tilde{h}(t)$ 来表示 $h(t)$ 以及信号 $h(t)$ 的小波变换。

小波变换表达式:

$$f_m(\tau, a) = \frac{1}{\sqrt{a}} \int f(t) \tilde{h}((t - \tau)/a) \cdot \exp(-j\omega_0((t - \tau)/a)) dt \quad (5)$$

其中: a, τ 分别为尺度和时间因子。Yao 等^[7]将主动耳蜗理论与小波变换结合, 在小波变换母函数中添加了一个调整因子 T , 得到仿生小波变换。

调整因子 T 表达式如式(6)所示:

$$T(\tau + \Delta\tau) = \left(1 - G_1 \frac{S_{BWT}}{S_{BWT} + |f_{BWT}(\tau, a)|} \right)^{-1} \cdot \left(1 + G_2 \left| \frac{\partial f_{BWT}(\tau, a)}{\partial t} \right| \right)^{-1} \quad (6)$$

其中: G_1, G_2 表示耳蜗的两个能动因子, S_{BWT} 是饱和因子, $\Delta\tau$ 是计算步长。

根据式(5)、(6)得到仿生小波变换表达式(7):

$$f_{BWT}(\tau, a) = \frac{1}{T\sqrt{a}} \cdot \int f(t) \tilde{h}\left(\frac{t - \tau}{aT}\right) \cdot \exp(j\omega_0((t - \tau)/a)) dt \quad (7)$$

2.2 阈值估计

小波变换中辨别语音信号和噪声的方法有模极大值消噪、阈值消噪等, 本文采用阈值消噪的方法。阈值的估计是去噪方法中的关键问题, 也是困扰很多学者的难点问题。阈值选取的合适与否将直接影响到降噪的效果。阈值估计得太大, 会将语音信号误当作噪声信号而去除掉, 产生语音失真; 反之, 阈值估计得太小, 去噪后的语音信号中仍存在大量噪声信号, 不能达到最佳去噪效果。

经过大量学者的艰苦探索, 已经提出了多种阈值估计的方法。本文采用一种基于尺度的阈值方法: 统一阈值估计方法。统一阈值估计是基于最小均方误差推导出来的, 表达式如下:

$$\lambda = \sigma_i \sqrt{2 \ln N} * (\sigma/\bar{\sigma}) \quad (8)$$

其中: σ_i 为噪声标准方差, N 为信号的长度, σ 为语音信号前一段噪声信号的标准方差, $\bar{\sigma}$ 所有噪声标准方差的平均值。在噪声段是进行实时更新。

$$\sigma_i = \sqrt{\text{median}(|w_{j,k}|)/0.6745} \quad (9)$$

其中: $w_{j,k}$ 为信号小波分解系数。该方法实现简单, 去噪效果不错。

2.3 阈值函数

常用的阈值函数有硬阈值函数和软阈值函数^[13], 但各自都有不足。

硬阈值函数公式如下:

$$\hat{w}_{j,k} = \begin{cases} w_{j,k}, & |w_{j,k}| \geq \lambda \\ 0, & |w_{j,k}| < \lambda \end{cases} \quad (10)$$

其中: $\hat{w}_{j,k}$ 为估计小波系数(经阈值函数处理后的小波系数), $w_{j,k}$ 为信号小波分解系数(带噪语音信号小波分解后得到的系数), λ 为阈值。

由图2可以看出, 硬阈值函数的波形图是不连续的, 语音信号重构是会引入振荡噪声, 影响去噪效果。

软阈值函数公式如下:

$$\hat{w}_{j,k} = \begin{cases} \text{sgn}(w_{j,k}) (|w_{j,k}| - \lambda), & |w_{j,k}| \geq \lambda \\ 0, & |w_{j,k}| < \lambda \end{cases} \quad (11)$$

式(11)中符号意义与硬阈值中相同。

由图 3 可以看出,软阈值函数的波形是连续的,避免了引入振荡噪声,但对大于 λ 的估计值 $\hat{w}_{j,k}$ 都减去了 λ ,信号重构后会出现偏差,影响去噪效果。

针对它们存在的问题,本文用如式 (12) 所示的阈值函数。该阈值函数结合硬阈值函数与软阈值函数的优点,克服了硬阈值函数的不连续性,又克服了软阈值函数的恒定偏差问题。

$$\hat{w}_{j,k} = \begin{cases} \gamma w_{j,k} + (1 - \gamma) \operatorname{sgn}(w_{j,k}) (|w_{j,k}| - \lambda), & |w_{j,k}| \geq \lambda \\ \operatorname{sgn}(w_{j,k}) \frac{w_{j,k}^4}{\lambda^3}, & |w_{j,k}| < \lambda \end{cases} \quad (12)$$

其中: $\gamma = 1 - e^{-25 \sqrt{|w_{j,k}| - \lambda}}$, $\gamma (0 \leq \gamma \leq 1)$ 是调节因子。

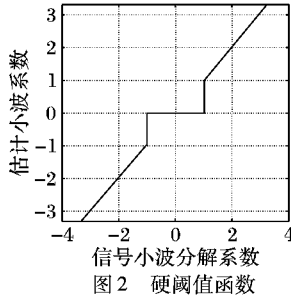


图 2 硬阈值函数

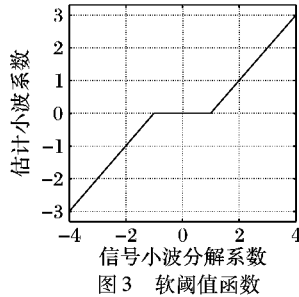


图 3 软阈值函数

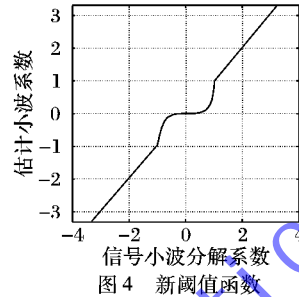


图 4 新阈值函数

3 基于子带谱熵的仿生小波语音增强方法

本文提出的基于子带谱熵的仿生小波语音增强方法流程如图 5 所示。

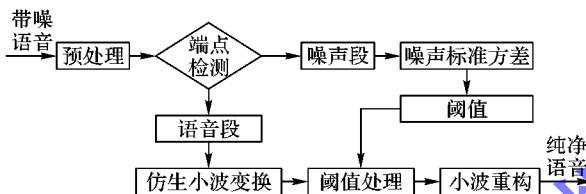


图 5 基于子带谱熵的仿生小波语音增强方法流程

1) 预处理。对输入的语音信号进行分帧与加窗处理。本文实验使用汉明窗,帧长为 20 ms,帧间重叠 50%。

2) 取前 5 帧信号(熵值在 $(0, 0.5]$) 为噪声信号,计算噪声标准方差 σ (对第一段语音信号进行仿生小波变换时, σ 与 $\bar{\sigma}$ 是相同的)。

3) 子带谱熵端点检测。如果检测表明当前输入语音帧为噪声段,计算并更新噪声标准方差 σ ,同时求得 $\bar{\sigma}$ 。如果检测表明输入语音帧为语音段,则进行仿生小波变换降噪。

① 仿生小波降噪。将语音段信号通过式 (7) 进行仿生小波变换,得到仿生小波系数 $f_{\text{BWT}}(\tau, \alpha)$ 。

② 阈值处理。根据上文,用式 (12) 作为阈值函数,阈值中的噪声标准方差是步骤 2) 所得。

③ 小波重构。噪声段直接置“0”,仿生小波变换部分进行小波逆变换,得到增强后的语音信号。

4 实验结果与分析

为验证本文提出算法的有效性,仿真采用两种噪声信号,分别为白噪声和汽车噪声,噪声信号和语音信号的采样率都是 8 kHz,选取 Morlet 函数作为仿生小波变化中的母小波函数。仿真实验比较了不同信噪比 (Signal-to-Noise Ratio, SNR) (0 dB, 5 dB, 10 dB) 下,混合噪声后的语音信号用基于子带谱熵的仿生小波语音增强方法、维纳滤波方法和常规小波去噪方法所得到的输出语音信号的信噪比,如图 6 所示。

从表 1 中可以看到,本文提出的基于子带谱熵的仿生小波语音增强方法较其他两种方法,去噪效果有明显提升。针对不同的噪声环境,本文提出的方法对于白噪声的增强效果较好。由图 7 可以看出,白噪声去噪效果明显优于其他两种方法。但在汽车噪声环境下,3 中方法的去噪效果没有太大

的差异。在低信噪比时,三种方法的语音增强效果差别不是很大。相对于低信噪比而言,高信噪比时提升效果更明显。常规的小波去噪,采用相同的阈值,忽略了噪声的特性。本文根据噪声的特性,动态地调整阈值,从而更好地抑制噪声。

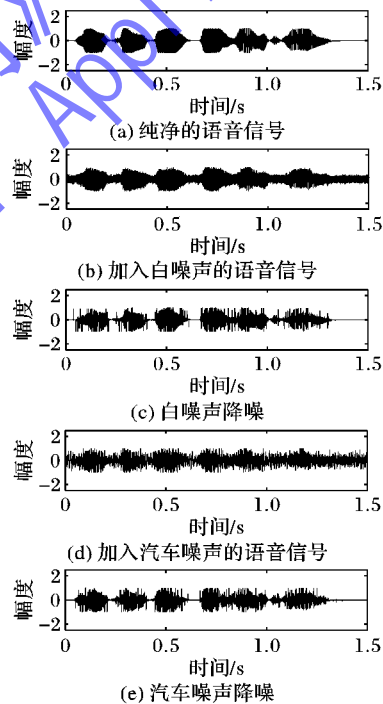


图 6 降噪效果

表 1 语音增强后的信噪比 (SNR) 值

噪声类型	输入信噪比/dB	输出信噪比/dB		
		本文方法	维纳滤波	常规小波去噪
白噪声	0	4.96	3.89	3.27
	5	10.38	7.72	5.56
	10	17.27	12.86	10.13
汽车噪声	0	3.81	2.87	2.72
	5	8.00	7.53	6.65
	10	14.39	12.88	11.31

5 结语

在分析和总结谱熵端点检测和仿生小波变换方法的基础上,提出采用子带谱熵端点检测与仿生小波变化相结合的去噪方法。该方法根据对端点的检测,精确地估计噪声标准方

差,实时地调整阈值,能实现对噪声信号和语音信号的最佳估计,达到去噪效果。实验结果也表明,该方法在多噪声情况下,输出信噪比比维纳滤波、常规小波去噪方法高。但本文方法的运算量比较大,在效率上有待提高。

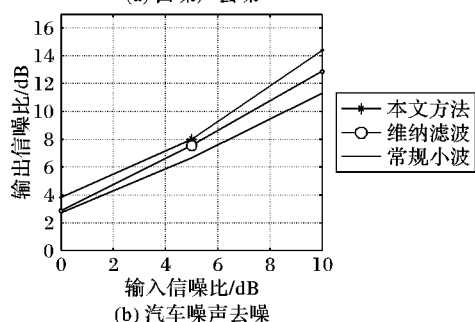
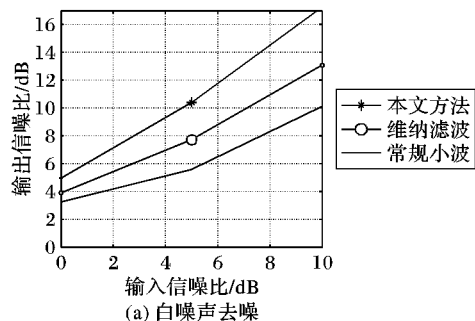


图7 去噪效果对比

参考文献:

- [1] LI F. Based on wavelet transform speech recognition of research and application [D]. Guilin: Guilin University of Technology, 2011. (李法强. 基于小波变换的语音识别的研究和应用[D]. 桂林: 桂林理工大学, 2011.)
- [2] DAI Y, ZHANG J. Reduction of $1/f$ noise in semiconductor devices based on wavelet transform and Wiener filter [J]. Acta Physica Sinica, 2011, 60(11): 185 – 190. (代煜, 张建勋. 基于小波变换和维纳滤波的半导体器件 $1/f$ 噪声滤波[J]. 物理学报, 2011, 60(11): 185 – 190.)
- [3] GENG Y, ZHANG D. Survey of adaptive filtering algorithms [J]. Information and Electronic Engineering, 2008, 6(4): 315 – 320. (耿妍, 张瑞金. 自适应滤波算法综述[J]. 信息与电子工程, 2008, 6(4): 315 – 320.)
- [4] ZHANG T, HE L, HUANG H, *et al.* Speech endpoint detection based on critical band and energy entropy [J]. Journal of Computer Applications, 2013, 33(1): 175 – 178. (张婷, 何凌, 黄华, 等. 基于临界频带及能量熵的语音端点检测[J]. 计算机应用, 2013, 33(1): 175 – 178.)
- [5] DONOHO D L, JOHNSTONE I M. Adapting to unknown smoothness via wavelet shrinkage [J]. Journal of the American Statistical Association, 1995, 90(432): 1200 – 1224.
- [6] REN Y, ZHANG X, JIA H. New threshold function of wavelet packet speech enhancement algorithm [J]. Application Research of Computers, 2013, 30(1): 114 – 116. (任永梅, 张雪英, 贾海荣. 一种新阈值函数的小波包语音增强算法[J]. 计算机应用研究, 2013, 30(1): 114 – 116.)
- [7] YAO J, ZHANG Y T. Bionic wavelet transform: a new time-frequency method based on all auditory model [J]. IEEE Transactions on Biomedical Engineering, 2001, 48(8): 856 – 863.
- [8] SHEN J-L, HUNG J-W, LEE L-S. Robust entropy-based endpoint detection for speech recognition in noisy environments [EB/OL]. [2014-08-12]. <http://wenku.baidu.com/view/e676ae0979563c1ec5da71d5.html>.
- [9] LI Y, WANG T, CUI H, *et al.* Voice activity detection in non-stationary noise [EB/OL]. [2014-09-01]. <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4281886&url=http%3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D4281886>.
- [10] WU B, WANG K. Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments [J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 762 – 775.
- [11] LI R, HU C, YU J. Research of speech endpoint detection based on spectral entropy algorithm [J]. Journal of Wuhan University of Technology, 2013, 35(7): 134 – 139. (李荣荣, 胡昌奎, 余娟. 基于谱熵的语音端点检测算法改进研究[J]. 武汉理工大学学报, 2013, 35(7): 134 – 139.)
- [12] YANG X, FAN X. Speech enhancement based on bionic wavelet transform and adaptive threshold functions [J]. Control and Decision, 2006, 21(9): 1033 – 1036. (杨玺, 樊晓平. 基于仿生小波变换和自适应阈值的语音增强方法[J]. 控制与决策, 2006, 21(9): 1033 – 1036.)
- [13] CHEN X, WANG Y, LIU L. Deep study on wavelet threshold method for image noise removing [J]. Laser and Infrared, 2012, 42(1): 3 – 5. (陈小曦, 王延杰, 刘恋. 小波阈值去噪法的深入研究[J]. 激光与红外, 2012, 42(1): 3 – 5.)
- [10] ACHANTA, R, SHAJI A, SMITH K, *et al.* SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274 – 2282.
- [11] YANG Y, XU L. Remote sensing image classification using layer-by-layer feature associative conditional random field [J]. Journal of Computer Applications, 2014, 34(6): 1741 – 1745. (杨耘, 徐丽. 基于分层特征关联条件随机场的遥感图像分类[J]. 计算机应用, 2014, 34(6): 1741 – 1745.)
- [12] XU B, BU J, CHEN C, *et al.* Efficient manifold ranking for image retrieval [C]// SIGIR '11: Proceedings of 34th ACM SIGIR International Conference on Research and Development in Information Retrieval. New York: ACM, 2011: 525 – 534.
- [13] XIAO Y, CAO Z, YUAN J. Entropic image thresholding based on GLGM histogram [J]. Pattern Recognition Letters, 2014, 40: 47 – 55.
- [14] ZHU L, KLEIN D A, FRINTROP S, *et al.* Multi-scale region-based saliency detection using W2 distance on N-dimensional normal distributions [C]// Proceedings of 20th IEEE International Conference on Image Processing. Piscataway: IEEE, 2013: 176 – 180.
- [15] XIAO Y, WU J, YUAN J. mCENTRIST: a multi-channel feature generation mechanism for scene categorization [J]. IEEE Transactions on Image Processing, 2014, 23(2): 823 – 836.

(上接第 857 页)