

基于感知小波变换的语音增强方法研究

李 野¹, 吴亚锋¹, 刘雪飞²

(1. 西北工业大学 数据处理中心, 西安 710072; 2. 陕西师范大学 生命科学学院, 西安 710062)

摘要: 在 ERB 尺度下构造的感知小波符合人耳对固有语音的频率感知特性, 通过一种纯数学算法计算其参数, 在听觉感知上可以近乎完美地使信号进行重构。首先采用感知小波对带噪语音进行分解, 其次在语音信号的子带层次上用一种类似于软阈值的无穷阶可导的函数进行阈值处理, 最后应用谱减法进行二次增强。实验表明, 该算法使信噪比和 PESQ 得分都有较大提高, 特别是在信噪比较高时, 语音具有很好的清晰度和可懂度。

关键词: 感知小波; 语音增强; 阈值

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2009)04-1313-03

Research on speech enhancement based on perceptual wavelet transform

LI Ye¹, WU Ya-feng¹, LIU Xue-fei²

(1. Data Processing Centre, Northwestern Polytechnical University, Xi'an 710072, China; 2. College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China)

Abstract: A perceptual wavelet constructed under ERB scale accorded with the perceptual of human hearing system, used pure mathematics method to calculate its parameter and the perceptual wavelet could reconstruct faultlessly in auditory perception. Used perceptual wavelet to decompose noisy speech, then used infinite-order derivative function which was similar to the soft threshold on sub-band level of speech signals, then used spectral subtraction in the second time speech enhancement. Experiment result shows the algorithm is highly improved both SNR and PESQ scores, especially when the SNR is high, the speech has well definition and intelligibility.

Key words: perceptual wavelet; speech enhancement; threshold

0 引言

语音增强在解决语音噪声污染、改进语音质量、提高语音可懂度等方面有着重要的作用。近几年来, 基于小波的语音增强是该领域的一个研究热点并取得了许多研究成果^[14], 但在这些研究中几乎都是直接采用了现有的小波基函数进行分析和应用。对于语音分析来说, 由于听觉感知本身存在较大的冗余; 同时, 无论是二进小波、小波包还是 M 带小波变换, 其频域划分都是一种倍频程关系, 这与人耳所固有的对语音的频域感知特性并不完全吻合^[5]。人类听觉系统对于声音频率的感知与实际频率的对应关系是一种非线性映射关系。映射频率尺度按照其应用的测量方法可分为 Mel 尺度、Bark 尺度、ERB 尺度。本文在 ERB 尺度下采用高斯函数作为感知小波的基函数, 基函数能够满足时间—感知频率(弯折频率、主观频率)上的最佳不确定性(时间—带宽积最小^[6]), 因此通过这样的构造方法所得到的感知小波变换将具有与听觉系统十分吻合的分析特性。

本文研究感知小波变换在阈值语音增强方法中的应用。仿真实验表明基于感知小波变换的去噪结果优于小波阈值法的去噪结果, 具有明显的优越性。

1 感知小波实现

人类的听觉系统对于声音频率的感知与实际频率间呈现

一种非线性关系。Greenwood^[7]提出了基于耳蜗的声音感知频率与实际频率(Hz)间的映射关系如式(1)所示:

$$f_{CB} = 16.7 \log(1 + 0.006046f_{Hz}) \quad (1)$$

利用此映射关系式, 将高斯函数在临界频带轴上移动, 可得到基于临界频带尺度的滤波器组。由于各子带的时域波形满足小波分析的小波动、紧支撑等必要条件, 其被命名为感知小波滤波器组^[4]。用 b 代表临界频带尺度变量, 当单位带宽定义为 3 dB 的情况下高斯基本函数定义如式(2)。

$$W_0(b) = \exp(-cb^2) = 2^{-4b^2} \quad (2)$$

其中 $c = 41n(2)$ 。

$$W_k^{CB}(f) = W_0(b - b_1 - k\Delta b) = 2^{-4(b - b_1 - k\Delta b)^2} \quad (3)$$

该滤波器组的所有感知小波的加权变换函数 $H(f)$ 可表示为式(4)的形式。为满足滤波器组的重构特性, 即通过滤波器组的信号可近似实现完全重构, 则其加权变换函数 $H(f) = 1$ 。

$$H(f) = \alpha \sum_{k=0}^{M-1} W_0(b - b_1 - k\Delta b) = \alpha \sum_{k=0}^{M-1} W_k^{CB}(f) = 1 \quad (4)$$

其中: $b = 16.7 \log(1 + 0.006046f) \in [b_1, b_2]$; $b_1 = 16.7 \log(1 + 0.006046f_1)$; $k = 0, 1, \dots, M-1$, M 为滤波器组个数。

定义原始信号为 $s(t)$, 通过感知小波重构的信号为 $\hat{s}(t)$, 并假设已经找到合适的 α 和 Δb 使得 $H(f) = 1$ 成立, 那么重构信号还可以表示成频域形如式(5)。又定义 $y_k(t)$ 为原始信号经各感知小波滤波器的输出信号, 那么重构信号可由 $y_k(t)$ 加权实现, 如式(6)所示。

$$\hat{S}(f) = S(f)H(f) = S(f)\alpha \sum_{k=0}^{M-1} W_k^{CB}(f) = S(f) \quad (5)$$

收稿日期: 2008-07-12; 修回日期: 2008-09-22

作者简介: 李野(1983-), 男, 辽宁锦州人, 硕士研究生, 主要研究方向为语音信号处理、信号测试与状态检测等(ibly@mail.nwpu.edu.cn); 吴亚锋(1961-), 男, 教授, 博导, 博士, 主要研究方向为信号与信息处理; 刘雪飞(1982-), 女, 辽宁锦州人, 硕士研究生。

$$\hat{S}(f) = S(f)H(f) = \alpha \sum_{k=0}^{M-1} S(f) W_k^{CB}(f) \quad (6)$$

本文采用了一种纯数学算法^[8], 计算 α 和 Δb 以满足 $H(f) = 1$ 。

```

初始化  $\Delta\gamma$  为任意小实数
 $k$  为滤波器索引
 $M$  为滤波器子带个数
 $b_1 = 16.7 \log(1 + 0.006046f_1)$ 
 $b_2 = 16.7 \log(1 + 0.006046f_2)$ 
for  $\xi = 1.0 : \Delta\xi : M$ 
   $\gamma = \Delta\gamma$ ;
   $\max W_k(b) = \xi$ ;
  while ( $\gamma \leq (b_2 - b_1)/M - 1$ ) and ( $\max W_k(b) \geq \xi$ )
     $\max W_k(b) = \max | \sum_{k=0}^{M-1} W_k^{CB}(b - b_1 - k\gamma) |, b \in [b_1, b_2]$ 
     $\gamma = \gamma + \Delta\gamma$ 
  end
  if  $\gamma \leq b_2 - b_1/M - 1$ 
     $\Delta b = \gamma$ ;
    break;
  end
end
 $\alpha = 1 / \max | \sum_{k=0}^{M-1} W_k^{CB}(b - b_1 - k\Delta b) |$ 

```

由上文算法原理可知: $H(f)$ 的曲线形状由式(4)决定, 各感知小波的中心频率则通过 Δb 来改变。

本文针对 8 kHz 的采样率, 以 $f[\text{Hz}] \in [0, 4000]$, $\Delta b \approx 0.6114$ 及 $\alpha \approx 0.5743$ 及 $M = 39$ 为参数设计的感知小波滤波器组如图 1 所示。由计算可知: 该滤波器组所有子带的加权是一条波动最大为 0.003 的接近 1 的平线, 即 $H(f) = \alpha \sum_{k=0}^{M-1} W_k^{CB}(f)$ 为一条幅值接近 1 的平线。其中第 11 个及第 14 个加权感知小波的时域波形如图 2 所示。

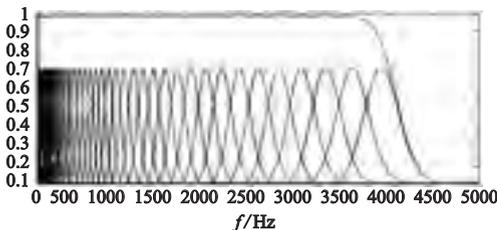


图1 感知小波滤波器组及 $H(f) = \alpha \sum_{k=0}^{M-1} W_k^{CB}(f)$

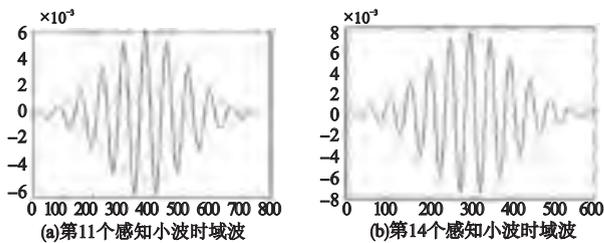


图2 第11个和第14个感知小波的时域波形

2 基于感知小波的语音增强方法

小波变换能使信号的能量在小波域集中于一些大的小波系数, 而噪声的能量却分布于整个小波域内, 因此经小波分解后, 信号的小波系数幅值要大于噪声的小波系数幅值。可以认为, 幅值较大的小波系数一般以信号为主, 而幅值较小的系数在很大程度上是噪声。于是, 采用阈值去噪的方法可以将信号系数保留, 而使大部分噪声系数减少为零。

该方法的基本做法是: 用感知小波滤波器组对信号进行滤波, 对于滤波后的小波系数可以设定一个阈值, 低于该阈值的小波系数置为 0, 高于该阈值的小波系数或完整保留, 或作相应的收缩处理。在实际应用中, 硬阈值和软阈值这两种方法得

到广泛的应用, 也取得了较好的效果, 但这两种方法也有一些明显的缺点。为了克服这些缺点, 本文采用了式(7)的改进阈值函数^[9]:

$$w_{j,k}(w_{j,k}, \lambda) = w_{j,k} - \lambda + 2\lambda / (1 + e^{2w_{j,k}/\lambda}) \quad (7)$$

该阈值函数可实现与软阈值函数相同的功能, 且具有无穷阶连续导数, 克服了传统阈值的不足, 满足了本文的要求, 具有很好的应用前景。在本文中阈值函数如式(8)取为

$$\lambda_j = \sigma \sqrt{2 \ln N / (a - k)} \quad (8)$$

其中: σ^2 为噪声的方差; N 为语音取样的长度; k 为滤波器组的序号; a 为实验参数, 本文取 $a = 100$ 。

本文采用感知小波阈值降噪时对清音段语音和浊音段语音采用不同的阈值处理方法^[1], 在信噪比提高的同时符合人耳的听觉感知, 但还有较小残留噪声存在。因此采用谱减法^[10]对其进行处理, 获得了较好的效果。本文提出的语音增强算法框图如图 3 所示。

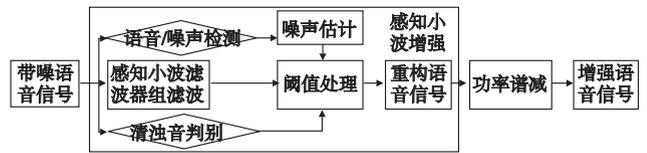


图3 语音增强算法框图

3 实验结果

实验中采用的语音材料选自 ITUT-T 推荐的 P. 862 建议的 PESQ^[11] 中的测试文件, 共 6 组(3 男 3 女)。该测试语音信号均为 8 kHz 采样、16 bit 量化的数字信号, 并在计算机中按一定比例混合组成不同信噪比(0, 3, 6, 10) dB 的带噪语音。带噪语音的长度为 256 点的语音帧, 相邻两帧重叠 128 点, 然后对带噪语音帧逐帧进行语音增强处理。

图 4、5 分别为输入信噪比为 3 dB 时采用本文方法进行语音增强的实验结果。

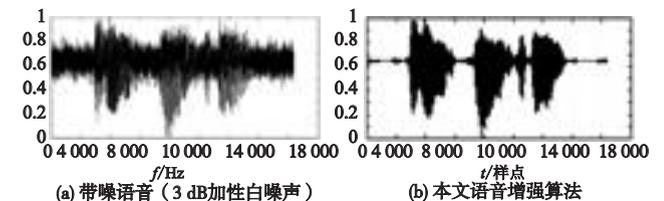


图4 语音时域图

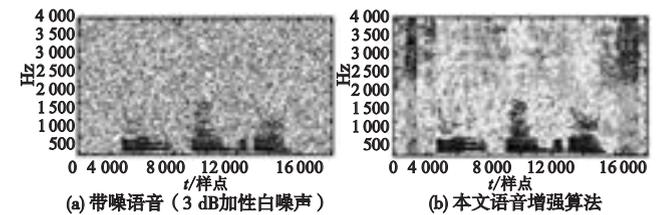


图5 语音谱图

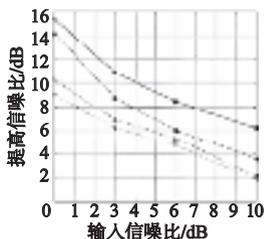
将噪声以变化的信噪比与语音信号混合, 用以下几种方法与本文提出的算法进行比较: a) 普通谱减法; b) 改进小波阈值法; c) 改进小波阈值法 + 谱减法。结果如图 6 所示。

语音信噪比(signal-to-noise ratio)计算公式如式下:

$$SNR = (1/L) \sum_{l=1}^{L-1} s^2(n + Nl) / (1/L) \sum_{n=0}^{N-1} [s(n + Nl) - \hat{s}(n + Nl)]^2 \quad (9)$$

其中: L 表示语音信号的总帧数; N 表示帧长; s 表示输入的语音信号; \hat{s} 表示增强语音信号。可以看出, SNR 表示时域信号失真, 式(9)表示对每一帧信号的信噪比进行计算, 最后取平

均值。



□ 本文方法 ○ 改进小波阈值+谱减 △改进小波阈值法▽ 谱减法

图6 不同语音增强算法下信噪比提高情况

为了验证语音增强算法的主观听觉质量,本文采用 PESQ 作为语音质量的评价标准对增强语音进行了客观评价。其中 PESQ 是目前主客观评价相关性最高的音质评估算法,能较好地反映语音的总体质量。表 1 比较了四种算法的输出语音的信噪比和 PESQ 得分。由表 1 可以看出,在 PESQ 得分方面,本文算法增强后的语音质量要好于其他三种。

表 1 PESQ 得分表

加噪语音信 号的信噪比	PESQ 得分				
	噪声信号	普通 谱减法	改进小波 阈值法	改进小波 阈值+谱减	本文 方法
10 dB	1.669	1.902	2.127	2.264	2.601
6 dB	1.534	1.192	1.982	2.112	2.381
3 dB	1.416	1.125	1.776	1.896	2.097
0 dB	1.391	1.048	1.557	1.620	1.900

4 结束语

本文在 ERB 尺度下构造的感知小波符合人耳对固有语音的频率感知特性,并给出了一种纯数学算法计算其参数。本文提出了应用感知小波对带噪语音进行阈值处理,并应用谱减法

进行二次增强的新算法。该算法对不同信噪比的带噪语音进行测试并与传统方法对比,通过实验结果表明,本文提出的算法不论是在信噪比还是 PESQ 得分都有较大提高,特别是在信噪比较高时,语音具有很好的清晰度和可懂度。

参考文献:

[1] 董胡,钱盛友.基于小波变换的语音增强方法研究[J].计算机工程与应用,2007,43(31):58-60.

[2] 王娜,郑德忠.节点阈值小波包变换语音增强新算法[J].仪器仪表学报,2007,28(5):952-955.

[3] DONOHO D L. De-noising by soft thresholding [J]. IEEE Trans on Inform Theory, 1995, 41(3):613-627.

[4] ISTVAN P. Perceptual wavelet-representation of speech signals and its application to speech enhancement [J]. Computer Speech and Language, 1996, 10(1):122.

[5] 付强,易克初.语音信号的 Bark 子波变换及其在语音识别中的应用[J].电子学报,2000,28(10):102-105.

[6] 科恩 L.时-频分析:理论与应用[M]. 白居宪,译. 西安:西安交通大学出版社,1998.

[7] GREENWOOD D D. Critical bandwidth and the frequency coordinates of the basilar membrane [J]. Journal of the Acoustical Society of America, 1961, 33(10):1344-1356.

[8] KOH M S. Speech enhancement using a truncated and constrained minimum variance estimator in non-uniform wavelet filterbanks [C]// Proc of the 34th Conference on Signals, Systems and Computers. 2001:101-104.

[9] 谭东昱.基于小波变换的语音增强方法研究及实时实现[D].长沙:湖南大学,2007.

[10] BOLL S. Suppression of acoustic noise in speech using spectral subtraction [J]. Acoustics, Speech, and Signal Processing, 1979, 27(2):113-120.

[11] ITU T Rec. P 862, Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs [S]. 2001.

(上接第 1312 页)其中: S 是所有序列的集合;扩展质量函数及迭代期望质量定义同上。各序列的产生概率可根据图 1 和信息素模型参数很容易得到。显然,使用扩展更新规则,四条边在所有序列中出现的次数是一样的,其对应的信息素模型参数值增加的概率也是一样,也就是说,这是一个竞争平衡系统。那么,使用扩展更新之后,迭代期望质量 $W_F(\tau, t)$ 会不会随 t 增加呢?图 4 给出了使用扩展更新后 $W_F(\tau, t)$ 和 $augW_F(\tau, t)$ 随 t 的进化。可见,使用扩展更新能使 $M(KMST, HCF-AS, N_{ant})$ 收敛于 $2mst_simple_inst$ 的最优解。图 5 给出了产生两个最优解 s_4 和 s_{13} 的概率 $p(s_4, t)$ 和 $p(s_{13}, t)$ 随 t 的进化。可见,最优解之一 s_{13} 的产生概率 $p(s_{13}, t)$ 最终达到 1。

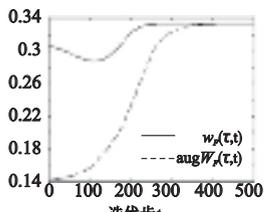


图4 $W_F(\tau, t)$ 和 $augW_F(\tau, t)$ 随 t 的进化

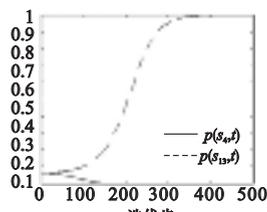


图5 $p(s_4, t)$ 和 $p(s_{13}, t)$ 随 t 的进化

5 结束语

本文研究了蚂蚁系统模型求解 k -最小生成树的性能。为此,设计了一种新的信息素模型和解的构造规则。在允许包括不可行解在内的全部解参与信息素模型更新的情况下,蚁群的迭代期望质量连续增加,算法收敛到最优解。

与此相关的另一问题是,算法为什么收敛到一个最优解而不是同时收敛到两个最优解(图 5),这是将来需要解释的问题。

参考文献:

[1] DORGO M, MANIEZZO V, COLORNI A. Ant system: optimization by a colony of cooperating agents [J]. IEEE Trans on Systems, Man, Cybernetics, Part B, 1996, 26(1): 29-41.

[2] 夏娜,蒋建国,魏星,等.改进型蚁群算法求解单任务 agent 联盟 [J]. 计算机研究与发展, 2005, 42(5):734-739.

[3] 李艳君,吴铁军.求解混杂生产调度问题的嵌套混合蚁群算法 [J]. 自动化学报,2003,29(1):95-101.

[4] DORIGO M, BLUM C. Ant colony optimization theory: a survey [J]. Theoretical Computer Science, 2005, 344(2): 243-278.

[5] BLUM C, DORIGO M. Search bias in ant colony optimization: on the role of competition-balanced system [J]. IEEE Trans on Evolutionary Computation, 2005, 9(2): 159-174.

[6] BLUM C, DORIGO M. The hyper-cube framework for ant colony optimization [J]. IEEE Trans on Systems, Man, Cybernetics, Part B, 2004, 34(2): 1161-1172.

[7] BLUM, SAMPLES, ZLOCHIN. On a particularity in model-based search [C]//Proc of Genetic and Evolutionary Computation Conference. San Francisco: Morgan Kaufmann Publishers, 2002.

[8] MONTGOMERY. Solution biases and pheromone representation selection in ant colony optimisation [D]. Queensland: Bond University, 2005.