

一种基于 HHT 的语音增强算法研究与仿真

汝 振¹, 李 昕¹, 陈 飞², 杨李箭¹, 李 翔¹

(1. 上海大学 机电工程与自动化学院, 上海 200072;

2. 上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 语音增强是语音识别、语音编码等语音分析系统中一个重要的环节, 直接影响到语音信号的处理过程。回顾了语音增强算法的发展历程, 提出一种基于希尔伯特黄(HHT)算法的语音增强算法。使用 HHT 对含噪语音进行 EMD 变换, 得到各阶 IMF 分量和残差, 对高频的 IMF 分量采用小波去噪的软阈值方法进行处理, 把经过阈值处理的高频 IMF 和低频 IMF 相加, 得到增强的语音信号。针对三种典型噪声(白噪声、babble 噪声和 volvo 噪声)的降噪问题, 应用 HHT 算法取得了良好的结果。

关键词: 语音增强; 希尔伯特-黄; 谱减法; 小波

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2010)08-0116-04

Research and Simulation of Speech Enhancement Algorithm Based on HHT

RU Zhen¹, LI Xin¹, CHEN Fei², YANG Li-jian¹, LI Xiang¹

(1. School of Electromechanical Engineering and Automation, Shanghai University,
Shanghai 200072, China;

2. School of Optical-Electrical and Computer Engineering, University of Shanghai for
Science and Technology, Shanghai 200093, China)

Abstract: Speech enhancement is a critical step in the fields of speech analysis, recognition, coding and synthesis, which directly affects the performance of the speech digital signal processing system. In this study, retrieved the history of speech enhancement and promoted a new pattern of speech enhancement based on HHT, using wavelet packet transform (WPT) to improve the pretreatment of HHT. After that got IMF derivation and residual. The high frequency IMF has been denoised by using wavelet. The high frequency and low frequency signal that have been dealt with threshold were added, then got the enhanced speech signal. As to denoise of three typical noise (white, babble, volvo), the experiment results indicate that the proposed algorithm get good results.

Key words: speech enhancement; HHT; spectral subtraction; wavelet

0 引言

语音增强主要有两个目的:第一是从主观度量说,改进语音质量,消除背景噪声,使听者易于接受,不会感觉疲劳。第二从客观度量说是提高语音可懂度。

用谱减法来抑制噪声^[1]是 Weiss, Aschkenasy 和 Parsons 在 1974 年最早提出的,使用噪声的平均谱来估计含噪语音段的噪声,使用谱减法抑制加性噪声。Lin 和 Oppeheim 在 1979 年提出了维纳滤波语音增强方法^[2],以迭代的方式进行维纳滤波。带噪语音的

AR 模型参数需要估计出来。软判决噪声抑制方法^[3]是在 1980 年由 Maulay 和 Malpass 提出的。基于 MMSE 短时谱估计的语音增强方法^[4]是在 1984 年由 Ephraim 和 Malah 提出的。新的语音增强方法在 90 年代后相继出现。例如:对带噪声的语音信号进行小波变换^[5],或者对带噪语音进行 K-L 变换^[6](KLT—Karhunen-Loeve Transform),又或者进行离散余弦变换^[7](Discrete Cosine Transform)来实现语音增强,以及使用基于人耳听觉掩蔽效应的语音增强方法^[8]。Hilbert-Huang 变换^[9]是 1998 年由 N. E. Huang 及其合作者提出的一种自适应的、能适用非线性、非平稳信号的新的信号分析方法。这种新方法的关键之处在于“经验模式分解”。近年来被许多学者用于语音增强^[9-11]。

收稿日期:2009-10-29;修回日期:2010-03-16

作者简介:汝 振(1983-),男,安徽亳州人,硕士生,研究方向为模式识别与智能系统、语音识别;李 昕,副教授,博士,研究方向为人工智能与模式识别、语音和话者识别、智能机器人。

1 HHT 变换

HHT^[5-7]信号处理方法分为两个步骤:

第一个步骤是经验模态分解(Empirical Mode Decomposition, 简称 EMD), 通过这一步能够获得有限数目的固有模态函数(Intrinsic Mode Function, 简称 IMF);

第二步则是利用 Hilbert 变换和瞬时频率方法获得信号的时频谱也即是 Hilbert 谱。有意义的瞬时频率 IMF 必须满足 2 个条件:

(1)在整个数据区间内, 一个 IMF 的极值点和过零点个数必须相等或最多只能相差一个点。

(2)在任意一点处包络线的平均值为零, 平均值是由极大值点定义的上包络线和由极小值点定义的下包络线的平均值。

满足此条件的 IMF 就被称为一个单分量信号。

EMD 算法的思路是: 利用了由极点定义包络, 在给定的信号中找到信号的极大值和极小值, 将所有极大值点和极小值点通过一条曲线连接起来, 得到的两条曲线包含信号所有数据, 能够获得信号曲线上包络曲线以及下包络曲线, 然后再计算上下包络曲线在每一个点上的平均值, 能够获得平均值曲线 m_1 。如果把分析信号定义为 $x(t)$, 则: $x(t) - m_1 = c_1$, c_1 为第一阶 IMF 分量。如果 c_1 不满足 IMF 的两个条件, 则求 c_1 与其包络中值的差值。直到 IMF 的两个条件得到满足, c_1 才是真正的第一阶 IMF 分量, 最后信号的逼近分量 R_1 可以通过从原始信号中减去 c_1 获得: $x(t) - c_1 = R_1$, 这就是筛分过程。通过对 R_1 重复上面的过程, 从而可以获得第二阶 IMF 分量。对信号的多次 EMD 筛分, 最终可以得到信号的多个 IMF 分量和一个逼近分量 R_n , 信号可以表示为: $x(t) = \sum_{i=1}^n c_i + R_n$ 。上面的筛分过程实际上是将原始信号分解为各种不同的特征波。这是由于每一个 IMF 分量都能反映信号的特征尺度。

与基于 Fourier 分析的方法相比, HHT 方法具有很多优越性。例如: 第一, 对信号的普遍自适应性, 不但对非线性、非平稳信号有效, 对线性、平稳信号同样有效。第二, 对信号表达的精确性, 它表达的是信号的最本质的特征, 而且不需要对数据施加任何限制。第三, 具有精确的时频局部性。第四, 简易性, 无论是硬件还是软件, HHT 方法的实现都非常简便。Hilbert 变换的物理含义反映到解析信号的极坐标形式: 信号局部的最佳逼近就是通过一正弦曲线的频率和幅值调制

获得。按照瞬时频率的定义, IMF 分量的瞬时频率可以表示为: $\omega(t) = \frac{d\theta(t)}{dt}$ 。相应的解析函数的幅值谱和瞬时频率可以通过对每一阶 IMF 作 Hilbert 变换求出, 原始信号表示为:

$$x(t) = \text{Re} \sum_{j=1}^n a_j(t) e^{i \int \omega_j(t) dt}$$

从而信号幅值、时间和瞬时频率之间的关系可以通过此式表达出来。其中表示为时间、瞬时频率的信号的幅值称作 Hilbert 谱。

2 基于 HHT 变换的语音增强算法

在实验室环境下录制中文“上海大学”, 并加入 SNR=5 的白噪声。噪声来自于 NOISEX-92 数据库。

图 1 为纯净语音“上海大学”与加噪语音。

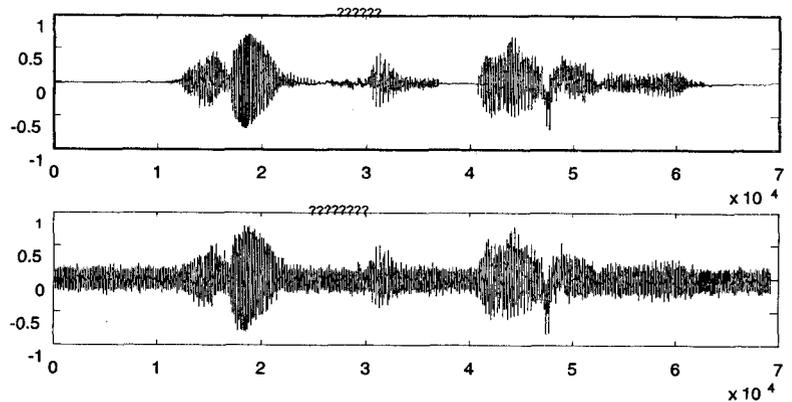


图 1 纯净语音“上海大学”与加噪语音

首先使用 HHT 对含噪语音进行 EMD 变换。得到各阶 IMF 分量和残差。如图 2 所示。

从图中可以看出噪声基本包含在 IMF1~IMF7 这些包含了信号较高频的成分中。由于宽带噪声的频谱分布与话音频谱重叠, 所以不能简单地采用低通尺度滤波, 否则会丢掉有用的语音信号。

文中提出一种基于 HHT 的语音增强算法, 对高频的 IMF 分量采用小波去噪的软阈值方法进行处理, 即对每一个 IMF 分量采用一个浮动的阈值来甄别携带能量较少的数据, 即小于或等于该阈值的数据视为携带能量较少者, 实际处理中将这此值作为零处理, 而仅仅保留阈值以上的数据, 软阈值处理公式如下:

$$\text{IMF}_{j,n} = \begin{cases} \text{sign}(\text{IMF}_{j,n}) (|\text{IMF}_{j,n}| - \delta) & |\text{IMF}_{j,n}| > \delta \\ 0 & |\text{IMF}_{j,n}| < \delta \end{cases}$$

其中 sign 是符号函数, δ 为阈值。计算函数为: $\delta = \sqrt{2 \log N \sigma / \ln(j+1)}$, σ^2 为噪声的估计方差。显然随着尺度 j 的增加, 阈值随着减小。把经过阈值处理的

频 IMF 和低频 IMF 相加,得到增强的语音信号。

3 实验结果

在实验分析中,录制了一段语音信号 $s(t)$,信号的采样频率为 8kHz,量化精度为 16 比特,内容为中文“上海大学”。将这段纯语音信号分别叠加 SNR=5 的高斯白噪声、babble 和 volvo 噪声。噪声来自于 NOISEX-92 数据库。

图 3 为 SNR=5 的高斯白噪声信号,分别用基于 HHT 的去噪方法、小波软值法和谱减法对带噪信号进行处理的对比图。

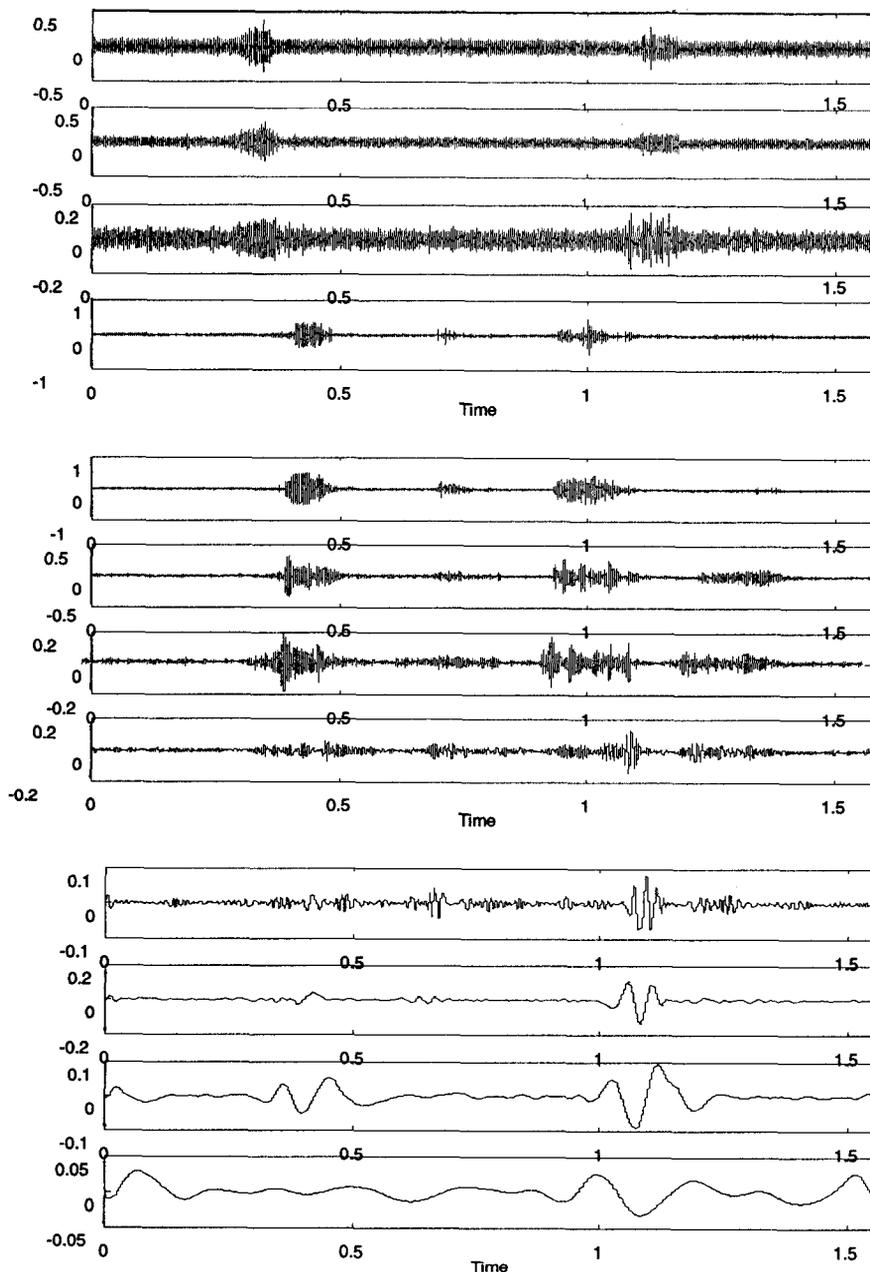


图 2 对含噪语音进行 EMD 变换得到的各阶 IMF 分量和残差

图 4 为 SNR=5 的 babble 噪声信号,分别用基于 HHT 的去噪方法、小波软值法和谱减法对带噪信号进行处理的对比图。

图 5 为 SNR=5 的 volvo 噪声信号,分别用基于 HHT 的去噪方法、小波软值法和谱减法对带噪信号进行处理的对比图。

加入三种噪声:白噪声、babble 噪声和 volvo 噪声,然后分别使用谱减法、小波法和文中提出的基于 HHT 的方法进行语音增强后,效果对比如图 3~图 5 所示。可以明显看出,谱减法的去噪效果最差,把大部分语音都去掉了。小波变换增强的语音效果较好,残存的背景噪声较少,但是由于局部波动特性被当作噪声而滤掉了,高频成分损失较多,所以听起来较低沉,语音的自然度有所破坏。HHT 在多种噪音下的去噪效果相对于另外两种方法来说都是最好的。不但去噪效果明显,而且波形相似度也高,实际倾听效果也令人满意。

景噪声较少,但是由于局部波动特性被当作噪声而滤掉了,高频成分损失较多,所以听起来较低沉,语音的自然度有所破坏。HHT 在多种噪音下的去噪效果相对于另外两种方法来说都是最好的。不但去噪效果明显,而且波形相似度也高,实际倾听效果也令人满意。

4 结束语

文中提出一种基于 HHT 的语音增强方法,该方法在提高语音信噪比的前提下保留了较多的高频语音成分,语音失真较小,语音的自然度较好,可懂性好。克服了小波在高频方面损失过多的缺点。实验证明该方法和谱减法、小波法相比较有着较大的优势。

但是,HHT 变换也有一些不足之处,例如:会在低频带产生一些虚假的频率分量,依赖于被分析的信号,首个 IMF 由于覆盖了太宽的频率范围导致特性不能保证,容易忽略低能量成分。所有这些缺点都有待于今后进一步的改进。

参考文献:

[1] Weiss M R, Aschkenasy E, Parsons T W. Study and development of the INTEL technique for improving speech intelligibi-

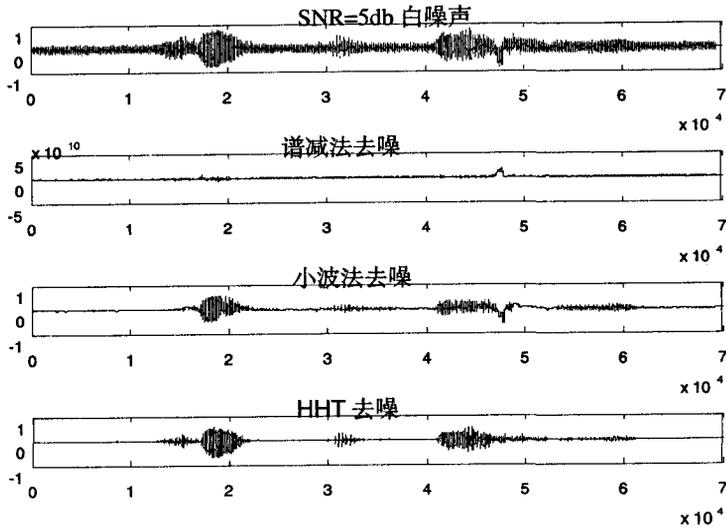


图 3 加入白噪声的去噪效果对比图

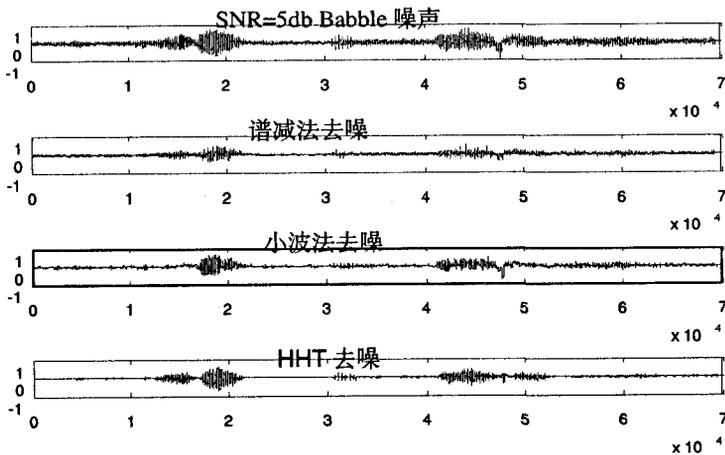


图 4 加入 babble 噪声的去噪效果对比图

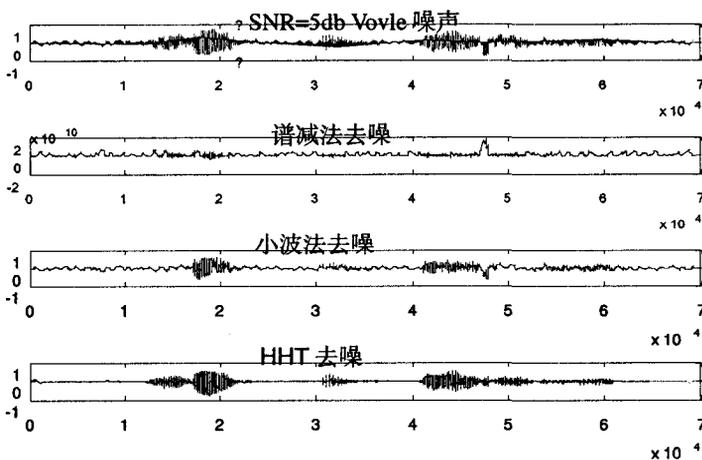


图 5 加入 volvo 噪声的去噪效果对比图

lity[M]. [s. l.]: Nicolet Scientific Corp., 1974.

[2] Lim J S, Oppenheim A V. Enhancement and bandwidth compression of noisy speech[J]. Proc. IEEE, 1997, 67: 1585 - 1604.

[3] McAulay R J, Malpass M L. Speech enhancement using a soft - decision noise suppression filter[J]. IEEE Trans. on Acoustics, Speech and Signal Processing, 1980, 28(2): 137 - 145.

[4] Ephraim V, Malah D. Speech enhancement using a minimum - mean square error short - time spectral amplitude estimator[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(6): 1109 - 1121.

[5] Lallouani A, Gabrea M, Gargour C S. Wavelet - based speech enhancement using two different threshold - based denoising algorithms [C]// Canada Conference on Electrical and Computer Engineering. Canada: [s. n.], 2004: 315 - 318.

[6] Rezayee A, Gazor S. An adaptive KLT approach for speech enhancement [J]. IEEE Transactions and Audio Processing, 2001, 9(2): 87 - 95.

[7] Hasan M K, Zilany M S A, Khan M R. DCT speech enhancement with hard and soft thresholding criteria [J]. Electronics Letters, 2002, 38 (13): 669 - 670.

[8] 王 霞, 卢建国, 张秀珍, 等. 一种基于人耳掩蔽效应和无语音概率的谱减语音增强算法 [J]. 现代电子技术, 2008(10): 23 - 25.

[9] Huang N E, Shen Zheng. The empirical mode decomposition and the Hilbert spectrum for non - linear and non - stationary time series analysis [J]. Proc. R. Soc. Lond, 1998, 454 (A): 903 - 995.

[10] Zou Xiaojie, Li Xueyao, Zhang Rubo. Speech Enhancement Based on Hilbert - Huang Transform Theory [C]// Proceedings of the First International Multi - Symposiums on Computer and Computational Sciences. [s. l.]: [s. n.], 2006.

[11] Li Xueyao, Zou Xiaojie, Zhang Rubo, et al. Method of Speech Enhancement Based on Hilbert - Huang Transform [C]// Proceedings of the 7th World Congress on Intelligent Control and Automation. Chongqing, China: [s. n.], 2008.

中国计算机学会会刊、中国科技核心期刊
《计算机技术与发展》欢迎订阅, 邮发代号: 52 - 127