一种基于小波变换的频域语音增强方法

A Speech Enhancement Algorithm Based on Wavelet in Frequency Domain

朱 岩 李雪耀 张如波

(哈尔滨工程大学计算机科学与技术学院 哈尔滨150001)

Abstract A new speech enhancement algorithm in high noise environment is proposed in this paper. The whole algorithm is based on wavelet transform and the Stein's Unbiased Estimate of Risk (SURE) estimation. This work improves the basic wavelet shrinkage method from threshold rule and enhancement algorithm. Considering the phase has the potential to preserve a lot of the underlying speech formants' structure and composition, Enhancement method is implemented in frequency domain. This modification enables the system to handle colored noises. To evaluate the system performance, we have employed plentiful real-life speech data that are obtained under practical environment. Subjective and objective evaluations show that the proposed system improves speech intelligibility in noisy environment.

Keywords Wavelet, Detection, Enhancement, SURE estimation

1. 引言

语音增强就是减小噪音对语音的影响,提高语音通信系统的性能。在过去的几十年中,人们提出了各种方法来改善语音,例如:自适应子空间滤波、基于 HMM 的增强方法、谱减估计法等等。然而,在不利条件下的语音增强依然具有挑战意义。在1995年,Donoho 和 Johnstone[1]提出了小波消减法(wavelet shrinkage method)去除语音中的高斯白噪声。它包括 RiskShrink,VisuShrink,SureShrink 等阈值估计算法。虽然一些应用已经使用了这些方法,但是要得到能处理各种噪声的成功方法,还存在很多的问题。

在本文中,我们提出了一种新的语音增强方法来提高语音质量和可懂度,此方法是基于小波变换和 Stein 的无偏似然估计(SURE),我们从两方面对原算法进行了改进:阈值函数和增强算法。改进后的阈值函数具有软阈值函数的连续性,又比硬阈值函数平稳。由于相位信息中保留了大量的基本共振峰结构和成分,我们提出一种基于语音相位和在频域上的小波消减法。这些改进使算法能够处理有色非平稳的噪声。为了评估系统的性能,我们使用了大量的真实数据来测试系统的性能,主观测试和客观测试都表明改进的方法提高了语音的可懂度。

本文的结构如下,在第2节中,首先简要介绍了小波消噪方法,给出了系统的框架和存在的各种问题。接着,我们分别介绍了对阈值函数的改进及在频域上的小波消减算法。在第3节中,我们给出了计算机实验结果和系统性能的分析。最后对提出的增强方法进行了总结。

2. 语音增强

在过去的十几年里,小波变换已经广泛地应用到信号处理的各个领域。由于小波变换的多分辨率分析,对于处理像语音这样的随机的、非平稳的信号是一种强有力的分析工具,它能在低频看到缓慢的瞬时变化、在高频看到剧烈的变化。而且,小波变换是一种非参数的方法。不像 HMM 方法或者神经网络需要事先假定一定的模型并且进行大量的训练,只要选取的小波函数能满足提取信号的特征需要,就可以实时地进行处理。总之,小波变换能为语音增强提供一种合适的工具。

在语音增强过程中,我们依靠 Stein 的无偏似然估计方法(SURE)。由于小波分解后的语音信号中,语音能量主要集中在某些小波尺度上,语音信号在这些小波尺度上分解系数比其他信号(特别是噪声信号)要大得多,而且,噪声信号要覆盖大多数尺度。因而,通过把较小的小波分解系数置为0,就可以最大限度地去除噪声,保留原有信号中的语音成份。

2.1 系统结构

首先,假定信号在时域中表示为

$$y(n)=x(n)+e(n) \tag{1}$$

其中,$y(n)$是含噪语音,$x(n)$是不含语音的噪声信号,$e(n)$是附加的背景噪声。小波变换为

$$Wy(n)=Wx(n)+We(n) \tag{2}$$

其中,W 是离散小波变换(DTW)矩阵。

小波变换语音增强是通过消减小波分解系数趋于零实现的。我们所推荐的增强方法的系统结构图如图1所示。

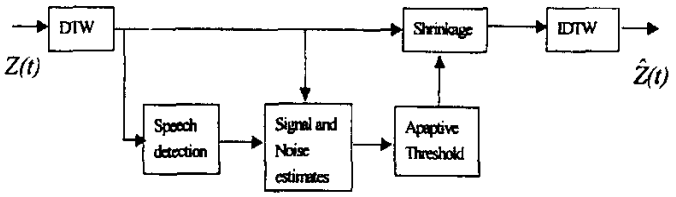

图1 系统结构图

由图可知，整个系统分为四步：

1. 语音信号小波变换进入变换域；

2. 使用语音流检测方法区分语音与噪声；

3. 由阈值估计器估计出修改小波分解系数所需的阈值，并且使用阈值消减函数进行小波系数变换；

4. 重构处理后的小波系数到达时域。

根据小波变换的统计特征，假设我们想要在高斯白噪声中估计出函数 f，当大多数的小波分解系数接近于零，对这种小波系数的消减处理能得到较好的效果。也就是当 f 系数的多数较小时，所保留较大系数就表征了 f 形态。由此我们可知阈值消减函数应具有下列两个特征：(1)能够去除较小的系数；(2)完好地保留较大的系数。

2.2 SURE 阈值

对于一个 N 维正态分布的矢量 $\vec{X}=\{X_0,X_1,\cdots,X_{N-1}\}$，每个分量是一个独立的正态分布的随机变量 $X_i\sim N(u_i,\sigma^2)$，其中 $i=0,1,\cdots,N-1$。由 Stein 定义的对均值 $\vec{u}=(u_0,u_1,\cdots,u_{N-1})$ 的估计 $\hat{\vec{u}}=(\hat{u}_0,\hat{u}_1,\cdots,\hat{u}_{N-1})$ 表示如下：

$$\hat{\vec{u}}=\vec{X}+g(\vec{X}) \tag{3}$$

其中 $g=(g_i)_{i=0}^{N-1}:R^N\to R^N$ 是无误差的函数。

由矢量 $\vec{X}$ 使用无偏似然估计来预测均值 $\vec{u}$ 的定义如下所示：

$$E_u\|\hat{\vec{u}}-\vec{u}\|^2=N\sigma^2+E_u\|g(\vec{X})\|^2+2\sigma^2E_u[\nabla\cdot g(\vec{X})] \tag{4}$$

其中 $\nabla\cdot g\equiv\Sigma_i\frac{\partial}{\partial X_i}g_i$。

Stein 估计表明对于人意的、非线性的、有偏的信号仍然能够估计它的最小无偏量。现在，考虑到在软阈值估计下，$\hat{u}_i^t=\eta_t(X_i)$ 是对真正均值 u_i 的估计，其中 $t(t\geqslant 0)$。使用 Stein 公式演算，$\hat{u}^t$ 应当是无偏差的估计，由公式(4)我们能推导出等式：

$$SURE_{soft}(t;\vec{X})=N\sigma^2+\sum_{i=0}^{N-1}\min(|X_i|,t)^2+2\sigma^2\sum_{i=0}^{N-1}\chi_{[-t,t]}(X_i) \tag{5}$$

其中 $\chi_{[-t,t]}(X_i)=\begin{cases}1 & |X_i|>t\\ 0 & |X_i|\leqslant t\end{cases}$ 是特征函数。根据此公式可知无偏似然估计式(4)可表示为：

$$E_u\|\hat{u}^t-u\|^2=E_u[SURE_{soft}(t,\vec{X})] \tag{6}$$

因此，在无偏似然估计下的软阈值大小定义如下：

$$t^{soft}=\arg\min_{t\geqslant 0}SURE(t,\vec{X}) \tag{7}$$

其中 $t\in\{0\}\cup\{|X_i|\}_{i=0}^{N-1}$。由式(7)生成的阈值被称为全局 SURE 阈值(global SURE threshold)。根据大数定理可以保证 SURE 估计收敛于真值。

2.3 阈值消减函数的改进

通常的阈值函数定义如下：

$$T_{t_k}^G(x)=\begin{cases}\left(1-\alpha\left|\frac{t_k}{x}\right|^{r_1}\right)^{r_2}x & ,\left|\frac{t_k}{x}\right|<\frac{1}{\alpha+\beta}\\ \left(\beta\left|\frac{t_k}{x}\right|^{r_1}\right)^{r_2}x & ,otherwise\end{cases} \tag{8}$$

其中 x 是含噪语音的小波分解系数。设 t_k 是第 K 层尺度上的估计阈值，α 和 β 是消减系数，r_1 和 r_2 是指数参数。

由 Donoho 和 Johnstone 提出的小波阈值消减函数分别是软阈值消减函数(Soft Shrinkage Function)和硬阈值消减函数(Hard Shrinkage Function)，其定义分别如下所示：

$$T_t^{hard}(x)=\begin{cases}x & |x|>t\\ 0 & |x|\leqslant t\end{cases} \tag{9}$$

$$T_t^{soft}(x)=\begin{cases}sgn(x)(|x|-t) & |x|>t\\ 0 & |x|\leqslant t\end{cases} \tag{10}$$

其中 $t\in[0,\infty]$ 是阈值。

软硬阈值在处理语音时各有利弊。由于软阈值方法对较大系数的消减作用，估计值趋于有更大的偏差。而硬阈值函数具有的不连续性，硬阈值消减估计趋于带来更大变化并且产生不平稳——即对数据中的较小变化敏感。当这些基本的小波阈值方法用来处理像语音这样的被真实噪音污染的复杂信号时，就存在较大的问题。

为了弥补软/硬阈值方法带来的缺陷，考虑到能量谱消减原则，在语音增强算法中，我们提出了一种新的阈值函数。定义如下：

$$T_{t_k}^G(x)=\begin{cases}(x^2-t_k^2)^{1/2} & |x>t_k\\ 0 & otherwise\end{cases} \tag{11}$$

在式(8)中，系数分别为 $\alpha=1,\beta=0,r1=2,r2=1/2$。软/硬阈值和我们所提出的阈值函数在图2中绘出。

从图2中，可以看出我们提出的阈值函数具有软阈值函数的连续性，又比硬阈值函数稳定。这种阈值函数比软/硬阈值方法有较小的预测错误。而且，它

的计算的复杂度并不比基本方法大。

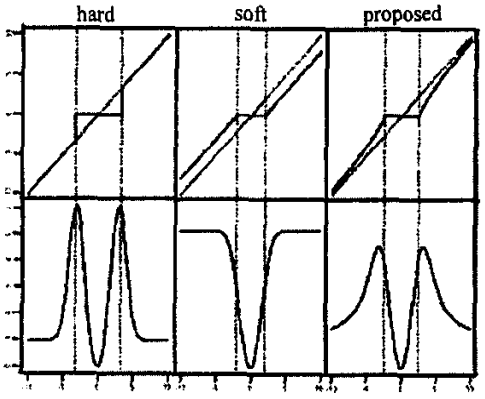

图2 顶部显示各种阈值函数，下部是相应的风险函数，其定义为 $R_t(\theta)=E\{T_t(X)-\theta\}^2$，其中 $T_t(.)$ 是具有阈值 t 和 $X\sim N(\theta,I)$ 的阈值函数。

2.4 增强算法

为了基本算法存在的问题，在原算法的基础上，我们提出了一种新的小波增强方法。这种方法是基于语音相位和在频率上的小波消减法。系统的结构框图如图3中所示。

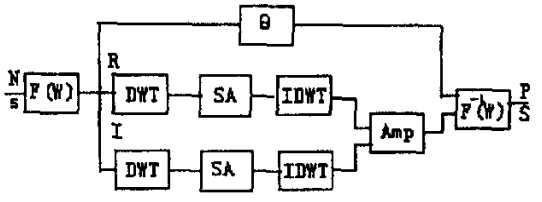

图3 增强系统的结构框图

基于 Stein 准则的基本方法在处理含噪语音时，初始信号是在时间域中，它并不能在所有的噪声层次上提高语音的可懂度。事实上，它产生了一种带有持续的、令人厌烦的、扭曲的语音信号，这主要是由于阈值技术非线性的结果。一些研究表明这样的实事：相位信息中保留了大量的基本共振峰结构和成分。因而，我们希望保留语音的相位信息。同时，对正态分布的多维随机信号进行频率变换。所产生的实部与虚部系数也是正态分布的，因为小波变换是一种线性和正交的变换，所以对于正态分布的多维随机矢量进行傅立叶变换再进行小波变换的结果，仍然是一个正态分布的矢量。

在我们的方法中，首先对时间域中的语声信号进行傅立叶变换，从中抽取出相位信息。然后，对实部与虚部的傅立叶变换系数分别进行小波变换，其结果再分别进行小波阈值消减处理，整个过程实域与虚域的处理是分别完成的。接着，为了产生增强处理的实部与虚部，再分别对阈值处理后的系数进行小波重构。最后，使用老的相位信息与重构后的幅值进行傅立叶反变换，得到时域的语音信息。总之，使用小波阈值消减处理与含噪的频域相位信息相结合的方法，是一种处理附加高斯白噪音(AWGN)语音增强的有力工具。

3 实验结果

为了细致地评估我们所提方法的性能，我们的语音流检测与增强的试验数据都来自于真实环境中。语音的采样频率是10kHz，8bit；相邻帧重叠50%；离散小波变换分解到6级。每帧长为1024个数据。为了准确地评价系统性能，我们进行了主观和客观两方面的测试。

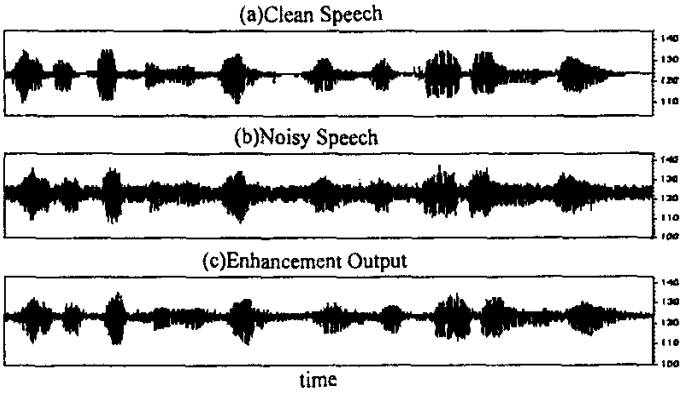

图4 原语音、噪声语音、增强后语音的对比

在客观测试中，我们建立了一个大型的语音库，在这个库中包含有120真实环境中采样得来的数据，同时，每条语音记录至少30分钟长。在测试中，主要是对输入信号的波形及 SNR 与输出信号的波形和 SNR 进行比较。在图4中，显示了信号增强的效果，图4(a)显示原始语音，图4(b)显示为加入电台的干扰后的波形，图4(c)显示为增强后的语音波形。可以看出，基本上去除了噪音干扰，同时，增强后没带来

其它的成分降低语音的可懂度。

在图5中(a),原始语音信号具有较低的信噪比。在图5(b)中是带有语音增强和检测算法的输出结果。因为语音检测算法的使用,它彻底地去除了噪声成分。而且,语音增强进一步增强了语音的可懂度。

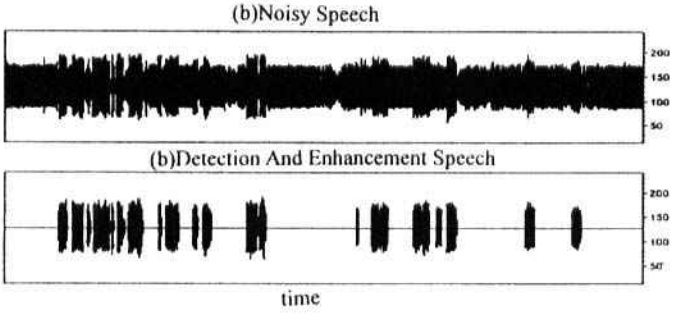

图5 含噪语音流及处理后语音

为了进行量化的评估,我们使用SNR作为客观的标准。因为它能够对比在含噪语音中的语音成分和噪声成分,所以可以定量地看出系统的性能,表1给出了3种不同SNR的客观测试结果。这些数据代表了我们大量试验的平均情况。可以看出对于有色噪音,算法也是有效的。

表1 语音增强前后SNR

No.	Noisy	Enhanced	SNR Gain
1	5 dB	11.29	6.29
2	10 dB	16.17	6.17
3	15 dB	18.65	3.65

主观试验是在专家的参与下完成的,我们发现,SNR值得的加不一定意味着语音可懂度的改善,因此我们进行了大量的主观试验。对于输入SNR在0到20dB的语音信号,系统比传统的小波增强方法有更好的性能。总之,评价结果显示我们提出的方案能在高噪音环境中获得更高的检测精确度和改善语音的可懂度。同时,算法的计算复杂度类似于谱减法,算法在实际应用中是可行的。

总结 在本文中提出了一种有效的基于小波变换的语音增强方法。我们的主要目标是发掘语音中的特征,从两个方面来改善基本的小波消减法:阈值消减函数和增强算法。由于考虑到了语音信号的相位信息,不同于传统的方法,我们的小波增强是在频域中完成的。由于使用了快速小波变换和模块化的程序设计方法。算法有效地实现了实时多通道处理。为了更好地评价系统的性能,使用了大量的真实环境中采集的数据来进行试验。试验结果表明我们提出的算法在实际环境下是可行的,同时算法也能精确地检测出高噪音环境下的微弱语音信号。

参考文献

1 Hosur S, Tewfik A H. Wavelet transform domain adaptive fir filtering. IEEE Trans. on Signal Processing, 1997, 45(3): 617～630

2 Erdol N, Basbug F. Wavelet transform based adaptive filters: analysis and new results. IEEE Trans. on Signal Processing, 1997, 44(9): 2163～2171

3 Donoho D L, Johnstone I M. Adapting to Unknown Smoothness via Wavelet Shrinkage. Journal of the American Statistical Association, 1995, 90: 1200～1224

4 Sheikhzadeh1 H, Abutalebi1 H R. An Improved Wavelet-based Speech Enhancement System. http://www.dspfactory.com/technology/technologypapers.html

5 Nason G P. Choice of the Threshold Parameter in Wavelet Function Estimation. in Wavelets and Statistics, 1995. 261～280

6 Sameti H, et al. HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise. IEEE Trans. Speech Audio Proc, 1998, 6(5)

7 Gao Hong-ye. Wavelet Shrinkage Denoising Using the Non-Negative Garrote. Journal of Computational and Graphical Statistics, 1998, 7: 469～488

8 Zhu yan, Li xueyao, Zhang rubo. Speech Stream Detection and Enhancement Wavelet-Based in High Noise Environment. International Conference on Control and Automation, 2002

9 Boll S F. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoustic, Speech, Signal Processing, 1979, ASSP-27: 113～120