

文章编号:1004-9037(2009)03-0362-07

基于小波变换的语音增强算法综述

李如玮 鲍长春 窦慧晶

(北京工业大学电子信息与控制工程学院,北京,100124)

摘要:基于小波变换的语音增强方法是一种新颖的去噪方法,值得大家去研究利用。本文根据大量的文献资料,从小波分解、模极大值、小波系数相关性、阈值去噪及混合去噪法等几个方面,详细地介绍了目前国内外基于小波变换的语音增强方法的研究现状,分析了各种改进方法的优缺点,并论述了小波去噪的发展趋势,对提高语音信号处理的性能具有较大的应用价值。

关键词:语音增强;小波变换;噪声估计;阈值去噪

中图分类号:TN912

文献标识码:A

Speech Enhancement Algorithm Based on Wavelet Transform

Li Ruwei, Bao Changchun, Dou Huijing

(School of Electronic Information and Control Engineering, Beijing University
of Technology, Beijing, 100124, China)

Abstract: Speech enhancement method based on the wavelet transform is a novel de-noising method. Several speech enhancement methods, such as wavelet decomposition, wavelet transform maximum, wavelet coefficient correlation, and threshold de-noising methods, are introduced. And their advantages and disadvantages are analyzed. In addition, the potential trend of the speech enhancement based on wavelet transform is discussed as well.

Key words: speech enhancement; wavelet transform; noise estimation; threshold de-noising

引 言

语音通信是人类传播信息,进行交流时使用最多、最自然、最基本的一种手段。而这种通信中的信息载体——语音信号却是一种时变的、非平稳的信号,只有在很短的一段时间内(通常为5~50 ms)^[1]才被认为是平稳的。在语音的产生、处理和传输过程中,不可避免地会受到环境噪声的干扰,使得语音信号处理系统,如语音编码和语音识别系统的性能大大降低。为了改善语音质量,提高语音的可懂度,人们根据语音和噪声的特点,采取各种语音增强方法抑制背景噪声。但是语音信号去噪是一个很复杂的问题,必须考虑语音本身的特点、千变万化的噪声的特点、人耳对语音的感知特性以及大脑如何处理信号等问题,所以,语音增强技术的研究是语音信号处理中永恒的课题。

尽管语音信号的去噪理论和方法目前还远远没有解决,但是40多年来,研究者们针对不同的噪声、不同的应用对加性噪声提出了很多方法。流行的语音增强方法有维纳滤波^[2]、卡尔曼滤波^[3]、谱减法^[4]和自适应滤波^[5]等。其中维纳滤波是在平稳条件下的基于最小均方误差的最优估计,但对语音这种非平稳信号不是很适合;卡尔曼滤波克服了维纳滤波的平稳条件,在非平稳条件下也可保证最小均方误差最优,但是仅适用于清音;谱减法是常用的一种方法,但是在信噪比低的情况下,对语音的可懂度和自然度损害较大,并且重建语音中产生了音乐噪声;自适应滤波是效果最好的一种语音增强方法,但是由于需要一个在实际环境中很难获得的参考噪声源,实际工作中并不好用,而且和谐减法一样伴有音乐噪声。同时,以上各种方法在进行语音增强时,都需要知道噪声的一些特征或统计特性,而在没有噪声先验知识的情况下,从带噪语音

基金项目:北京市自然科学基金(4042009)资助项目;北京市教委科技发展计划(KM200710005001)资助项目。

收稿日期:2007-10-23;修订日期:2008-04-29

信号中提取语音信号是比较困难的。

小波变换是近10年来迅速发展起来的一种时频局部分分析方法,它克服了短时傅里叶变换固定分辨率的缺点,能够将信号在多尺度多分辨率上进行小波分解,各尺度上分解得到的小波系数代表信号在不同分辨率上的信息。同时小波变换与人耳的听觉特性非常相似^[6],便于研究者利用人耳的听觉特性,是分析语音这种非平稳信号的有力工具,所以近年来很多研究者都利用小波变换来处理语音信号。小波变换法去噪的原理是:语音信号的能量集中在低频段,而噪声能量则主要集中在高频段,这样就可将噪声小波系数占主要成分的那些尺度上的噪声小波分量置零或给予很小的权重,然后用处理后的小波系数重构恢复信号。同时,随着小波变换理论的发展,小波变换去噪不断丰富,并且取得了良好的效果,如1992年Mallat提出了利用小波变换模极大值去噪^[7],Donobo在1995年提出了非线性小波变换阈值去噪^[8],这种方法使得小波去噪得到广泛应用,吸引了众多的研究者。

1 小波分解

在语音增强中,分解信号的目的是把信号的能量集中到某些频带的少数系数上,便于有效抑制噪声。在采用小波变换的方法中,研究者们一般采用正交小波,因为正交小波变换能最大程度地去除原信号中的相关性,将其能量集中在少数稀疏的、幅度相对较大的小波系数上。小波分解只是将每一级的低频成分逐级往下分解,对于高频成分不再作分解。这种分解方式不能满足既想得到好的时间分辨率又想拥有好的频率分辨率的场合,于是研究者们开始采用正交小波包来分解语音信号,这样便于利用人耳听觉掩蔽效应来进行语音增强。如文献[9]利用小波包算法有灵活的时频分析能力以及能较好地利用人耳基底膜的频率分析特性,按照Bark刻度与频率刻度之间的转换关系,采用固定小波包分解方式把0~4 000 Hz频带分成了52个频段,对应18个Bark刻度,从而在单声道条件下,其语音增强效果比传统的谱减法有更高的清晰度和可懂度。而文献[10]在利用小波包分解时,采用5级分解,得到17个频带与Bark刻度对应,文献[11]采用6级分解,得到24个临界带。它们的目的是为了充分地利用人耳的听觉特性,在进行语音增强时不需要把噪声完全抑制,只要残留的噪声不被感知即可,以便在去噪的同时减少不必要的语音失真。

通常研究者在采用小波包分解时一般采用固

定的分解级数,且一般在5级以上。大量的实验表明,小波分解级数对算法的降噪效果影响很大,分解级数过多,会造成信号的某些重要的局部特性丢失,信噪比反而下降,且运算量大,延时大;分解级数太小,则噪声对应的模极大值不能足够衰减,信号和噪声不能很好地区分,使得降噪效果不理想,信噪比提高有限^[12],因此采用固定小波分解级数在很大程度上限制了算法的降噪性能。为此,文献[13]提出了一种新颖的分解级数自适应选择方法,该方法有效提高了小波阈值降噪算法的性能,但进一步引入时延和计算量。

由于第一代小波存在时延大、算法相对复杂、对内存的需求大等缺点,文献[14]采用了自适应提升小波进行语音增强。实验表明该方法在算法复杂度降低的同时,能大大消除噪声且保持了语音良好的可懂度。

由小波理论可知:正交小波分解在分解过程中,不能保证中间过程的线性相位,这不利于语音信号的处理,而双正交小波分解则能够保证中间过程的相位不发生失真。为此,本文利用双正交小波包进行语音去噪,取得了不错的效果。同时,由于小波分解最终依赖滤波器组实现,不可避免地带来时延,限制了小波理论的应用范围。为此,有必要设计低时延的滤波器来实现小波分解,为小波理论的进一步应用打下坚实的基础^[15]。

总之,在采用小波分解的研究过程中,从最初的小波分解发展到小波包分解、利用人耳听觉特性进行小波包分解、自适应地选择分解级数、提升小波分解、为保证线性相位采用双正交小波包分解以及设计低时延的滤波器,都是为后续的处理工作做出充分的准备,以便更好地提高去噪性能。

2 模极大值去噪法

模极大值去噪的原理^[7]是:语音信号的模极大值随着尺度的增大而增大或不变,而噪声的模极大值却随着尺度的增大而减小。人们根据此特性,去除噪声的模极大值,保留语音的模极大值,然后利用保留的模极大值重构语音,达到去除噪声的目的。模极大值去噪的具体步骤是^[16]:对含噪语音进行离散二进小波变换,分解尺度一般为4或5;求每个尺度上小波系数对应的模极大值点;在最大尺度上,选取阈值,则模极大值小于该阈值的点被置零,反之不变;搜索传播点,保留语音产生的模极大值点,去除噪声产生的模极大值点;利用各尺度保留的模极大值点,重构去噪语音。

基于小波变换模极大值的去噪方法虽然具有很好的理论基础,但是在实际应用中存在很多影响计算精度的因素,去噪的效果并不满意。如文献[17]利用小波变换频响特性的插值法重建低尺度上的小波变换模极大值,最后根据压缩映射原理构造解析形式的迭代投影算子方法重构信号,但是这种方法提高的性能有限。同时,该方法在具体操作上有许多待解决的技术问题,如分解尺度取多大适合;重构只用有限个模极大值点,这样重构的信号肯定与原始信号有误差,那么如何构造与原始信号相似的小波系数等,这些都限制了该方法的进一步应用。目前研究这种方法的文献很少。

3 相关性去噪法

相关性去噪法的原理是:语音信号的小波系数在各尺度间具有较强的相关性,而噪声的小波系数在各尺度间没有明显的相关性。相关性去噪法的主要步骤是:计算相邻尺度同一空间位置小波系数的相关性 $CW_{j,k}$, $CW_{j,k}=\omega_{j,k}\omega_{j+1,k}$, j 代表尺度, k 代表位置, $\omega_{j,k}$ 表示第 j 尺度的第 k 个小波系数, $\omega_{j+1,k}$ 表示第 $j+1$ 尺度的第 k 个小波系数。比较相关性与小波系数的大小,如果相关性大,则说明是信号,小波系数保留;反之,认为是噪声,该小波系数置零。利用处理后的小波系数重构去噪信号。

该方法中,相关性的计算增强了信号的边缘特性,更便于提取信号的特征。文献[18]利用该方法进行去噪取得了较好的效果。

不过在该方法中,一旦小波分解的过程有偏差,则计算出的相关性不能准确代表 k 点的真实相关性,依赖相关性去噪法的性能则降低。文献[19]介绍了一种区域相关的去噪方法,较好解决了以上问题。该方法主要考虑了 k 点处的小波系数,也考虑了 k 点附近的小波系数,从而减弱了小波系数偏差带来的影响。但是该方法每一个点都要计算相关系数,计算相对复杂,并没有引起广泛的研究。

4 阈值去噪法

1995年,阈值去噪法首次由Donobo提出,他提出的非线性小波变换阈值去噪^[8]使得小波去噪得到深入研究和广泛应用。

阈值去噪法的理论依据是:噪声的小波系数和有用信号的小波系数在幅值上存在不同的表现形式,在低频段,语音信号的小波系数值大于噪声的小波系数值,在高频段,反之这样对各层小波系数

设定一个适当的阈值将信号与噪声分开。该算法的具体步骤为:对含噪信号进行小波变换;对小波系数进行非线性阈值处理;利用处理后的小波系数重构去噪信号。

在小波阈值去噪方法中,有2个关键的问题:(1)阈值施加方法;(2)对阈值的具体估计。这两个问题直接影响着去噪的性能。

4.1 阈值的施加

传统的阈值施加主要有两种常用方法^[20-21]:①硬阈值法,即当小波系数的绝对值小于给定的阈值时,令其为零;而大于阈值时,则保持不变,即

$$\omega_{\lambda} = \begin{cases} \omega & |\omega| \geq \lambda \\ 0 & |\omega| < \lambda \end{cases} \quad (1)$$

②软阈值法,即当小波系数的绝对值小于给定的阈值时,令其为零;而大于阈值时,令其都减去阈值,即

$$\omega_{\lambda} = \begin{cases} [\text{sign}(\omega)](|\omega| - \lambda) & |\omega| \geq \lambda \\ 0 & |\omega| < \lambda \end{cases} \quad (2)$$

式(1,2)中: ω 表示小波系数的大小; ω_{λ} 表示施加阈值后小波系数的大小, λ 为阈值; $\text{sign}(\cdot)$ 为符号函数; $|\omega|$ 表示取绝对值。

尽管这两种方法在实际中经常采用,但是方法本身具有缺点。例如:硬阈值法在 $\omega=\lambda$ 处, ω_{λ} 不连续,用 ω_{λ} 重构信号会产生振荡;软阈值法得到的 ω_{λ} 虽然连续性好,但在 $|\omega|>\lambda$ 时, ω_{λ} 与 ω 存在偏差,直接影响重构信号的精确性。同时,大量实验表明,由于上述方法的固有缺陷,使得硬阈值法对噪声消除不干净,且处理后的语音混有音乐噪声,而软阈值法虽然噪声去除较干净,但是对原始语音的损害很大,使语音的清晰度和可懂度大大降低。为此,研究者们对阈值的施加展开了大量的研究。

例如:文献[9]采用小波包分析来模拟人耳的频率分析机制,利用人耳听觉系统可以抑制某些频率的信号以及对有用的信号通过降低听觉阈值而达到听清的目的特性,提出了如式(3)的类似听觉系统机制的动态阈值法除噪。该动态阈值能够跟踪语音信号的高频变化,提高辅音的增强效果。

$$\omega_{\lambda} = \begin{cases} \omega & |\omega| > \alpha \times \lambda \\ [\text{sign}(\omega)](|\omega| - \lambda) \times \frac{\alpha}{\alpha - 1} & \lambda < |\omega| \leq \alpha \times \lambda \\ 0 & |\omega| \leq \lambda \end{cases} \quad (3)$$

式中 α 为实验系数。该方法较好地模拟了听觉系统的声音感知阈值效应,在白噪声情况下,处理后的语音无论是清晰度还是可懂度都明显优于传统的

谱减法。但是这种方法对于听觉中枢神经的语音增强机理的模拟还有待研究,同时 α 的取值很关键,直接影响处理效果。对不同的噪声类型,该参数的取值需要调整。

文献[22]提出的阈值施加方法如下

$$\omega_{\lambda} = \begin{cases} (\omega^2 - \lambda^2)^{1/2} & |\omega| > \lambda \\ 0 & \text{其他} \end{cases} \quad (4)$$

该阈值施加法比硬阈值稳定,类似连续的软阈值。该方法,可以增强信号能量大的频率段,削弱信号能量小的频率段,从而提高语音的可懂度。

文献[23]提出了Step-garrote 阈值法,这种方法的关键是各级小波系数绝对值在 $0 \sim 2\lambda$ 范围内单步增长,但是这种方法只适合白噪声的情况。

文献[24]源于小波系数为噪声和语音对应的小波系数之和,提出了小波软硬阈值的折中算法,即

$$\omega_{\lambda} = \begin{cases} [\text{sign}(\omega)](|\omega| - \alpha\lambda) & |\omega| \geq \lambda \\ 0 & |\omega| < \lambda \end{cases} \quad (5)$$

式中 α 为 $0 \sim 1$ 之间的一个参数,此时, ω_{λ} 的值介于软阈值与硬阈值之间,且 ω_{λ} 的值更接近原语音对应的小波系数,通过调节该参数可以获得更好的增强效果。实验表明该方法优于单纯的硬阈值和软阈值。不过该方法中如果噪声不同,却使用相同的 α ,去噪效果会受到影响。

文献[25]也提出了改进的阈值法,即

$$\omega_{\lambda} = \begin{cases} \omega \left(\frac{|\omega|^{\beta} - \lambda^{\beta}}{|\omega|^{\beta}} \right) & |\omega| \geq \lambda \\ 0 & |\omega| < \lambda \end{cases} \quad (6)$$

式中 β 是一个关键参数,通过这个参数使阈值介于硬阈值和软阈值之间,该方法在抑制噪声的同时减少了语音信息的损失,信噪比和均方误差等性能都有明显的提高。不过这种方法对于清音和浊音采用的参数不同,故对实际语音处理时必须先判断语音是清音还是浊音。

以上各种方法中的相同变量的含义一致,并且各种方法在小波系数小于阈值时都简单地置零,这是不合理的。为了既保证施加阈值后小波系数的平滑过度,又保证大的小波系数能够去掉噪声,文献[26]提出一种改进的阈值,即当小波系数小于阈值时,让小波系数平滑减小到零,大于阈值时,小波系数减去阈值,从而减少音乐噪声。

总之,研究者们一直努力地寻求阈值施加方法,尽量去除噪声,并尽可能保持语音的可懂度和自然度。

4.2 阈值估计

小波去噪的关键是阈值的选取,因为阈值直接

影响着去噪的效果和重构信号的失真程度。如果阈值太小,噪声去除不干净,反之会去除一部分语音信号的分量。因此,首先要估计阈值的大小。目前,阈值估计的方法很多,如基于均方误差的最小风险阈值 $\lambda_{\text{MSE}} \sim \sigma * \sqrt{2 \ln N}$,其中, N 为数据长度, σ 为噪声强度^[20]。但是最小均方误差很难估计,所以在有关文献中,研究者们偏爱使用通用阈值 $\lambda_{\text{UNIV}} = \sigma * \sqrt{2 \ln N}$,这个阈值估计法对稳定的高斯白噪声效果最好,但对有色噪声或非稳定噪声不是很适合。另外还有利用Stein的无偏估计求出的SURE阈值^[27]等。这些方法都是采用Donobo提出的方法来估计噪声强度的,即取小波系数在各个尺度下绝对值的中值除以常数0.674 5的值作为该尺度下的小波系数中噪声强度的估计,此时 $\sigma = \text{median}(|\omega|)/0.674 5$, $|\omega|$ 表示小波系数的绝对值。

目前很多具体的阈值估计法都是基于以上几种方法的改进。为了能跟踪噪声水平的变化,文献[9]首先对无语音段的噪声进行分析,获得噪声的频谱特征作为阈值的基准,然后取基准值的 $2 \sim 4$ 倍作为初始阈值。初始阈值确定后,各频段的阈值随输入信号按式(7)而变化,即

$$\lambda(t) = \lambda(t-1) - c \times \text{sign}(\omega(t-1) - \lambda(t-1)) \times \log(|\omega(t-1) - \lambda(t-1)| + 1) \quad (7)$$

式中 c 为常数,一般为 $1.5 \sim 3$,决定阈值变化的快慢。 $\lambda(t)$ 为 t 时刻的阈值, $\omega(t-1)$ 为 $t-1$ 时刻信号的小波系数。由式(7)可以看出:各频段每时刻的阈值由上一时刻的阈值和信号的小波系数共同决定,即当上一时刻信号的小波系数比阈值大,则阈值下降,反之阈值增大。实验表明,这种方法能灵敏地跟踪语音的高频变化,提高清音的增强效果,并且在信噪比较高时语音增强效果接近谱减法,信噪比较低时效果则明显优于谱减法,不过这种方法只限于高斯白噪声。文献[28]引入Teager能量算子来自适应地估计阈值,但这种方法在低信噪比情况下存在过度阈值的问题。文献[29]对每个子带采用鲁棒性的权重因子来修正噪声掩蔽阈值,即如果剩余噪声的能量大于噪声掩蔽阈值,则含噪语音的小波系数被抑制从而移走更多的剩余噪声;如果剩余噪声的能量小于噪声掩蔽阈值,则利用权重因子来保证语音的质量。实验证明该方法能够提升增强语音的质量。不过由于引入了较为复杂的权重因子,该算法的复杂度增大。文献[30]在小波包的基础上,根据信号的奇异性理论,即白噪声具有负的奇异性,

其幅度和稠密度随尺度的增大而减小,而信号则相反^[31],提出了自适应阈值法。在该方法中,噪声强度的估计采用了谱熵的估计方法,从而对每一尺度上的噪声进行最大量去噪,进一步提高含有白噪声语音的信噪比。实验表明,该方法具有较好的去除白噪声的效果。文献[32]对细节采用Stein的无偏似然估计原理进行阈值估计^[27],这种阈值只将部分系数置零,当信号的高频信息有很少一部分在噪声范围内时,可将弱小信号提取出来;对概貌采用Donobo提出的非线性软阈值,在抑制噪声时有效地表达了信号特征。不过在这种方法中,由于对概貌采用存在缺陷的软阈值,使得对代表重要信息的概貌去噪时,导致重建语音存在偏差。文献[33]考虑到白噪声信号的幅值随着小波尺度 j 的增大而迅速衰减,对不同尺度使用不同的阈值 λ_j ,即 $\lambda_j = \sigma * \sqrt{2 \ln N / \ln(j+1)}$,以便减少语音信息的损失,这种方法应该适合所有场合,不过最关键的还是 σ 估计。

以上方法都是建立在稳定白噪声的基础上的。但是,语音所处的环境噪声是非常复杂的,既有白噪声,还有工厂噪声、嘈杂人声及马达声等,既有稳定噪声还有非稳定噪声,这些噪声与稳定白噪声具有不同的特性。为了能去除有色噪声,文献[22]采用最小均方误差估计阈值,该方法能自适应跟踪系统的变化以及根据某些统计认识能够找出最优阈值。但是这种方法在信噪比很低情况下,效果并不理想。文献[12]采用了一种自适应的时频阈值估计方法。在该方法中噪声强度 σ 的估计采用 $\sigma = \frac{\sum_{n=0}^{\text{int}(0.2 * N)} \omega(n)}{(\text{int}(0.2 * N) + 1)}$,其中, $\omega(n)$ 为第 n 点小波系数, $\text{int}(0.2 * N)$ 表示取离 $(0.2 * N)$ 最近的整数。然后采用 $\lambda_j = \sigma * \sqrt{2 \ln(N * \log_2(N))}$ (式中 λ_j 表示第 j 尺度的阈值)计算阈值。实验表明,该方法能够去除非稳定和有色噪声。但是当信噪比低时,对语音的损害很大。文献[11]提出了随信噪比而改变的阈值算法,而信噪比是根据自适应噪声来估计的。该方法能有效去除各种噪声且能减少语音失真。但是当信噪比很低时,效果不佳。文献[34]提出了根据噪声帧频谱的平整度 γ 判断出噪声的类型,然后对不同噪声采用不同的自适应阈值对带噪语音进行处理。该自适应阈值采用一个噪声类型函数 $g(\gamma)$ 修正通用阈值,即 $\lambda_j = \sigma * \sqrt{2 \lg N} * g(\gamma)$, $g(\gamma) = \sqrt{2 / \lg(0.05\gamma)}$ (γ 为平整度)。实验表明该方法具有较好的噪声抑制作用。不过这种方法

在进行噪声类型判断时很困难。

总之,阈值的估计在小波去噪中是非常关键的,阈值的恰当与否直接关系到去噪的性能。

用阈值法去噪不仅能够几乎完全抑制噪声,而且能很好保留反映原始语音信号的特征尖峰点,具有很好的去噪效果,但是这种方法是强烈依赖所选的小波,同时,在有些情况下,阈值法去噪后的信号会在信号某些不连续点的附近和快变化点处,出现视觉上的干扰,如去噪信号在这些点处会在一个特定的目标水平上下跳变^[16],影响去噪的质量。所以有人提出了平移不变量小波阈值去噪法^[16],即通过平移含噪语音来改变不连续点的位置,再对平移后的信号进行阈值法去噪处理,然后把去噪后的信号再进行相反的平移,从而得到原信号的去噪信号。该方法与阈值法相比,不仅能得到更小的均方根误差,而且也进一步提高信噪比,只不过该改进阈值去噪法的计算复杂度较高。

5 混合去噪法

纯粹的小波去噪方法尽管能取得较好的效果,但是在低信噪比和有色噪声的情况下,语音的可懂度并不是很高。为了能利用小波变换的优点以及更好地去除噪声,现在的研究趋势是把各种小波方法相融合或小波方法与其他方法相结合。如文献[35]为了消除音乐噪声,提出了基于小波阈值的低方差谱估计方法,实验表明多带谱估计结合小波阈值能够抑制音乐噪声,增强语音的质量比谱减法好。不过这种方法对有色噪声的处理效果次于高斯白噪声。文献[36]在小波域对低尺度小波系数采用自适应滤波,对高尺度系数采用谱减法或维纳滤波等方法。实验表明这种方法结合了小波去噪、自适应滤波和谐减法的优点,对语音的损害小于阈值去噪,同时也减少了音乐噪声,不过引入了计算复杂度和时延。为了防止高频中的清音作为噪声被去除,文献[37]首先根据小波系数的能量进行清浊判断,如果是清音,则只对最小尺度的低频成分去噪,以便保留清音,否则对所有尺度去噪,从而在抑制噪声的同时尽可能地保留清音信息,提高增强语音自然度,降低增强语音的失真度,但是在有色噪声的情况下,噪声去得不干净。文献[11]采用的语音增强方法,添加的步骤是:含噪语音首先由小波变换分解成若干临界带,然后由前向反馈子系统提取一系列分量,利用平均归一化时频能量来引导前向反馈子系统阈值,抑制稳定噪声,同时,利用改进的小波阈值抑制非稳定噪声和有色噪声,最后利用固

定的软阈值对清音进行语音增强。该方法结合人工神经网络的去噪方法已经被成功用在语音识别领域^[38],只不过时延大而不能用在实时处理中。文献^[39]在小波域引入卡尔曼滤波,成功地把卡尔曼滤波的优点和小波分解能够模拟人耳感知特性的优点结合起来,在一定程度上能抑制非稳定噪声和有色噪声,并且语音有很少的失真。文献^[40]在小波域引入谱减法来计算掩蔽阈值和最优加权系数,引入基于噪声估计的参数法对清音进行增强,取得了很好的去除多种噪声的性能,只是算法更加复杂。考虑到虽然上述方法能够在一定程度上去除稳定、非稳定、白噪声和有色噪声,但是当信噪比很低的情况下,去噪效果欠佳,而且仍然含有少量音乐噪声,文献^[41]提出了利用仿生小波变换来对语音进行分解的方法。该仿生小波变换相对于小波变换的优点在于它在时频域的尺度不仅可以根据信号的频率进行调节,而且可以随信号的瞬时幅度以及一阶微分系数自适应调节。实验表明,该方法能更好地保留原始纯净语音。如果能够在仿生小波的基础上应用以上的各种方法,应该能得到更好的效果。

6 结束语

语音增强的算法很多,传统的有依赖语音生成模型、需要提取语音参数的参数去噪方法,但是如果实际噪声或语音条件与该方法依赖的语音生成模型相差很大,或提取模型参数困难,该类方法失效;有如谱减法一类的非参数方法,虽然不需要从带噪信号中估计模型参数而广泛被研究者应用,但是这类方法利用了短时傅里叶变换,并不适合非平稳语音信号;有如听觉掩蔽效应一类的统计去噪方法,需要建立模型库,统计各种参数,算法运算量大,实时性不好,同时也是利用了短时傅里叶变换这种不适合非平稳语音信号的工具,去噪性能有限。但是小波变换不同,它既能在分解的过程中利用人耳的听觉掩蔽效应,同时能在没有噪声先验知识的情况下,从噪声和语音在各尺度的不同表现中把语音分离出来,它结合了语音增强传统方法的特点于一身。所以小波去噪应该是发展趋势。以上基于小波的去噪方法都是为了更大挖掘小波理论的特点,充分利用小波的优势,达到去噪的目的。

总之,为了能使小波去噪更有效,可以充分地利用小波系数的特征,再结合实际语音信号的特点,灵活选用各种小波去噪方法,或综合利用各种小波去噪方法,以及在小波域结合使用传统的自适应滤波、谱减法、维纳滤波等,以便利用各种方法的优

点,为提高语音信号处理的性能打下坚实的基础。

参考文献:

- [1] 鲍长春. 数字语音编码原理[M]. 北京:北京工业大学出版社,2001.
- [2] 杨行峻,迟惠生. 语音信号数字处理[M]. 北京:电子工业出版社,1995.
- [3] Gabrea M. Adaptive Kalman filtering-based speech enhancement algorithm [C]//IEEE Canadian Conference on Electrical and Computer Engineering. Toronto, IEEE, 2001(11):521-526.
- [4] Ogata S, Shimamura T. Reinforced spectral subtraction method to enhance speech signal[C]//Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology. Singapore: IEEE, 2001(1):242-245.
- [5] 胡啸,胡爱群,赵力. 一种新的自适应语音增强系统[J]. 电路与系统学报,2003,8(5):72-75.
- [6] Yang X, Wang K, Shamma S A. Auditory representations of acoustic signals[J]. IEEE Transaction on Information Theory, 1992,38(2):824-839.
- [7] Hwangwl M S. Singularity detection and processing with wavelets[J]. IEEE Transaction on Information Theory, 1992, 38(2):617-643.
- [8] Donobo D L. Denoising by soft-thresholding [J]. IEEE Transaction on Information Theory, 1995, 41(3): 613-627.
- [9] 王伟,杨道淳,方元,等. 基于听觉模型的小波包变换的语音增强[J]. 南京大学学报:自然科学版,2001,37(5):630-636.
- [10] Lei S F, Tung Y K. Speech enhancement for nonstationary noises by wavelet packet transform and adaptive noise estimation[C]//Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems. Hong Kong: IEEE, 2005: 41-44.
- [11] Shao Y, Chang C H. A versatile speech enhancement system based on perceptual wavelet denoising [C]//International Symposium on 2005 Circuits and Systems. Singapore: IEEE, 2005,2:864-867.
- [12] Mallat S G. 信息处理的小波导引[M]. 杨力华,译. 北京:机械工业出版社,2002.
- [13] 蔡铁,朱杰. 小波域降噪算法中最优分解层数的自适应选择[J]. 控制与决策,2006,21(2):217-220.
- [14] 陈立伟,赵春晖,姜海丽. 基于自适应提升小波变换的语音增强算法的研究[J]. 哈尔滨工程大学学报,2005,26(5):668-671.
- [15] Wang J, Kuang J, Zhao S, et al. A characteristic waveform decomposition algorithm using biorthogonal wavelet transform[C]//ICSP2006 Proceedings.

- Singapore: IEEE, 2006(1).
- [16] 孙延奎. 小波分析及其应用[M]. 北京:机械工业出版社, 2005:220-224.
- [17] 戴悟僧. 小波变换的频响特性及其在语音去噪中的应用[J]. 数据采集与处理, 2000, 15(1):12-17.
- [18] 潘泉, 戴冠中, 张洪才, 等. 基于阈值决策的子波域去噪方法[J]. 电子学报, 1998, 26(1):115-117.
- [19] 徐晨, 赵瑞珍, 甘小冰. 小波分析应用算法[M]. 北京:科学出版社, 2004.
- [20] Donobo D L, Johnstone I M. Ideal denoising in an orthogonal basis chosen from a library of bases[J]. C R Acad Sci I-Math, 1994, 319: 1317-1322.
- [21] Donobo D L, Johnstone I M. Ideal spatial adaptation by wavelet shrinkage[J]. Biometrika, 1994, 81(3): 425-455.
- [22] Zhu Y, Li X Y, Zhang R B. Adaptive speech enhancement based on wavelet in high noise environment [C]//Proceedings of International Conference on Machine Learning and Cybernetics. Beijing: IEEE, 2002(2):885-889.
- [23] Ayat S, Manzuri M T, Dianat R. Wavelet based speech enhancement using a new thresholding algorithm[C]//Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing. Hong Kong: IEEE, 2004:238-241.
- [24] 赵未莲. 基于小波变换的阈值语音信号去噪[J]. 重庆科技学院学报:自然科学版, 2005, 7(4):73-75.
- [25] 李倩, 王让定, 陈金儿. 基于改进阈值的小波域语音增强算法[J]. 宁波大学学报:理工版, 2005, 18(3): 344-348.
- [26] 胡广书. 现代信号处理教程[M]. 北京:清华大学出版社, 2004:397-405.
- [27] Jansen M. Noise reduction by wavelet thresholding [M]. New York: Springer-Verlag, 2001.
- [28] Mohammed B, Jean R. Wavelet speech enhancement based on the teager energy operator[J]. IEEE Signal Processing Letters, 2001, 8(1):10-12.
- [29] Lu C T, Wang S C. Speech enhancement using robust weighting factors for critical-band-wavelet-packet transform[C]//Proceedings of 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal: IEEE, 2004, 1: 721-724.
- [30] 付炜, 许山川. 基于阈值的小波域语音降噪新算法[J]. 计算机与数字工程, 2005, 33(11): 80-83.
- [31] Mallat S, Zhong S. Characterization of signals for multiscale edges [J]. IEEE Trans PAMI, 1992, 14 (7):710-732.
- [32] Wang Zhenli, Yang Jie, Zhang Xiongwei. Combined discrete wavelet transform wavelet packet decomposition for speech enhancement[C]//ICSP 2006 Proceedings. Singapore: IEEE, 2006(2).
- [33] 刘娟花, 李福德. 一种改进的小波域语音去噪方法研究[J]. 西安工程科技学院学报, 2006, 20(1):92-95.
- [34] 马晓红, 宋辉, 殷福亮. 自适应小波阈值语音增强新方法[J]. 大连理工大学学报, 2006, 46(4):20-26.
- [35] Hu Y, Philipos C L. Speech enhancement based on wavelet thresholding the multitaper spectrum [J]. IEEE Trans on Speech and Audio Processing, 2004, 12(1):59-66.
- [36] Mohammad A A, Ali A, Farokh A M. Speech enhancement by adaptive noise cancellation in the wavelet domain[C]//2005 Fifth International Conference on Information, Communications and Signal Processing. Tehran: IEEE, 2005:719-723.
- [37] 王慧琴, 何继爱, 张秋余. 小波变换在语音增强中的应用[J]. 甘肃科学学报, 2005, 17(4):79-82.
- [38] Shao Y, Chang C H. A novel hybrid neuro-wavelet system for robust speech recognition [C]//ISCAS 2006. Island of Kos, Greece: IEEE, 2006: 1852-1855.
- [39] Shao Y, Chang C H. A Kalman filter based on wavelet filter-bank and psychoacoustic modeling for speech enhancement[C]//ISCAS 2006 Proceedings. Singapore: IEEE, 2006:121-124.
- [40] Shao Y, Chang C H. A generalized time-frequency subtraction method for robust speech enhancement based on wavelet filter banks modeling of human auditory system[J]. IEEE Transactions on Systems, Man and Cybernetics, Part B, 2007, 37(4): 877-889.
- [41] 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.

作者简介:李如玮(1972-),女,博士研究生,副教授,研究方向:语音增强、小波变换及其在语音信号处理中的应用, E-mail: liruwei@bjut.edu.cn; 鲍长春(1965-),男,博士,教授,博士生导师,研究方向:语音与音频信号处理; 樊慧晶(1969-),女,副教授,研究方向:语音与音频信号处理。