基于时频阈值的小波包语音增强算法

徐耀华[①②] 王 刚[①②] 郭 英[①]
[①](空军工程大学电讯工程学院 西安 710077)
[②](大连大学信息科学与工程重点实验室 大连 116000)

摘 要：该文考虑小波域应用语音降噪中听觉掩蔽效应，提出了一种基于时频阈值的小波包语音增强算法。新算法首先通过频域增强方法得到语音粗估计，通过跟踪估计语音时频特性的细节变化，及时调节降噪阈值，然后利用时频阈值对小波包系数进行处理，以达到语音降噪的目的。实验表明，较传统小波域语音降噪方法，新算法在抑制平稳白噪声的同时减小了语音信息的损失，其增强语音的 MOS (Mean Opinion Score)评分、输出信噪比、MBSD (Modified Bark Spectral Distortion)测度性能均有明显提高。

关键词：语音增强；小波；阈值；时频
中图分类号：TN912.3 **文献标识码**：A **文章编号**：1009-5896(2008)06-1363-04

Wavelet Package Based Speech Enhancement Algorithm Using Time-Frequency Threshold

Xu Yao−hua[①②] Wang Gang[①②] Guo Ying[①]
[①](*Telecommunication Engineering Institute, Air Force Engineering University, Xi'an* 710077, *China*)
[②](*University Key Lab of Information Sciences and Engineering, Dalian University, Dalian* 116000, *China*)

Abstract: Incorporating masking properties of the human auditory system in wavelet domain, this paper proposed a new algorithm of wavelet package speech enhancement based on the time-frequency threshold. New algorithm first obtains speech pre-estimation by frequency-based de-noising method, then, via tracing the variation of time-frequency information of the speech pre-estimation, the threshold is modulated adaptively. Finally, the noisy speech is de-noised by means of time-frequency thresholding the coefficients of the wavelet package. With comparing to the traditional wavelet algorithms, the proposed algorithm offers more pleasant enhanced speech with less distortion and residual noise in the additive Gaussian noise environments, and the experimental results demonstrate its better performance in Subjective test, input and output SNR test, and Modified Bark Spectral Distortion (MBSD) measurement tests.

Key words: Speech enhancement; Wavelet; Threshold; Time-Frequency

1 引言

噪声不仅影响语音可懂度和清晰度，而且造成人耳听觉疲劳。人们一直在努力研究有效的语音增强方法，以使得语音通信系统和自动语音处理系统能更好的应用于实际环境。

近年来，小波和小波包分析方法在语音增强领域中，受到学术界的广泛重视，它对非平稳信号的分析具有时频局部化分析的突出优点，能对时域和频域“聚焦”到信号的任意细节进行观察。因此，可用小波变换来处理带噪语音信号在不同尺度上的带噪小波系数，从而实现重构增强语音信号的目的。基于小波包的语音增强算法[1−4]的基本思路是：通过小波包变换，将带噪语音信号分解成小波包系数，选取合适的阈值，通过阈值函数，在小波域消除背景噪声，再经过小波包反变换，合成增强语音。其中，如何选取阈值是小波包增强算法的关键。早期的文献通常局限于不变阈值方法，如 Donoho 和 Johnstone 在基于小波阈值估计的降噪算法[1,2]中给出的适用于高斯白噪声的硬阈值函数和软阈值函数，该算法对小波各子带系数采用不随时间频率变化的阈值，在提高信噪比的同时很容易带来语音失真。针对这一问题，近来 Bahoura 和 Rouat 在阈值选择中引入了 Teager 能量算子，提出了基于 Teager 能量算子(Teager Energy Operator, TEO)的阈值算法[5,6]；Lei 和 Tung 通过估计带噪语音信噪比，提出了自适应噪声估计(Adaptive Noise Estimation, ANE)阈值算法[7]。这两种算法采用带噪语音的时频粗包络调整降噪阈值，能够在一定程度上减小语音失真，但是却无法把握语音的细节内容，算法只是利用语音的粗略时变信息，而没有充分利用时频的细节信息，因此称之为时变阈值方法。在考虑语音信号中掩蔽效应[8,9]的基础上，文章提出了时频(Time-Frequency, TF)阈值算法，新算法首先采用频域方法

2006-11-23 收到，2007-05-31 改回
国家自然科学基金(60601016)，陕西省自然科学基金(2006F40)和辽宁省高校重点实验室开放基金(2006-04)资助课题

得到估计语音，然后通过紧密跟踪语音的时频细节变化，及时调节小波包各子带的阈值，相比较而言，新算法在抑制背景白噪声的同时能够有效减小语音失真。研究和实验表明，新算法保持了小波域算法的高输出信噪比和低残留音乐噪声性能，并且具有失真低、抗干扰性能强的特点，尤其是在低信噪比情况下，也能获得良好的增强效果。

2 相关知识

通常带噪语音信号 $x(n)$ 表示为 $x(n)=s(n)+w(n)$，$\{n=1,2,\cdots,N\}$。其中，$s(n)$ 和 $w(n)$ 分别表示原始语音和背景白噪声，N 为语音样本长度。

2.1 小波(包)阈值降噪

最初的小波(包)阈值降噪算法分为 4 步[1,2]：第 1 步，截取长度为 N_0 的噪声样本，计算小波(包)阈值 T_0；第 2 步，小波(包)分解带噪语音，得到 M 个子带小波(包)系数 $x^j(k)$ ($j=1,2,\cdots,M$)；第 3 步，通过阈值函数处理小波(包)系数；第 4 步通过小波(包)反变换得到增强语音。其中，第 2 步求阈值为

$$T_0=\sigma\sqrt{2\log(N_0\times\log_2 N_0)},\ \sigma=\mathrm{MAD}/0.6745 \tag{1}$$

式中 MAD 为噪声小波包高频子带系数绝对值的中值估计[1]。

第 3 步中的软阈值函数[1]和硬阈值函数[2]分别定义为

$$\tilde{x}^j(k)=\begin{cases}0, & \left|x^j(k)\right|<T_0\\ \mathrm{sign}(x^j(k))\cdot\left(\left|x^j(k)\right|-T_0\right), & \left|x^j(k)\right|\ge T_0\end{cases},\quad k=1,2,\cdots,N^j \tag{2}$$

和

$$\tilde{x}^j(k)=\begin{cases}0, & \left|x^j(k)\right|<T_0\\ x^j(k), & \left|x^j(k)\right|\ge T_0\end{cases},\quad k=1,2,\cdots,N^j \tag{3}$$

式中 j 表示第 j 个子带。各子带系数长度为 N^j。$\tilde{x}^j(k)$ 为阈值函数处理后的小波包子带系数，$\mathrm{sign}(\cdot)$ 为符号函数。

鉴于软阈值、硬阈值均不随信号时频信息变化，以下统一称为不变阈值。

2.2 ANE 阈值

ANE 算法使用信噪比调节阈值，归纳为：首先对语音信号分帧，并进行语音活动性检测，然后估计每一帧小波包子带信噪比(用 $\mathrm{SNR}^j(l)$ 表示第 l 帧 j 子带的信噪比)；通过式(4)计算阈值调节因子 $\beta^j(l)$，最后利用式(5)修正该帧的阈值[7]。

$$\beta^j(l)=\frac{1}{1+e^{-\eta(\mathrm{SNR}^j(l)-C)}} \tag{4}$$

η 为调解系数，C 为门限，这两个参数均由实验确定。

$$T^j(l)=T_0\beta^j(l) \tag{5}$$

3 基于时频阈值的新方法

3.1 时频阈值原理

为了更好的理解时频阈值，以一段频率随时间由低至高线性变化的 chirp 纯净信号为例，做时频分析，来论述各类阈值之间的关系。后面如无特殊说明，对 chirp 信号均采用 sym8 基，4 级小波包变换，取第 8 子带系数。图 1 中实线代表小波包子带系数，a、b、c 线段为不同阈值示意图，非实际计算所得，a 线段为不变阈值。不变阈值通过降低阈值以保留更多的有用信号，可以减小信号失真，但增加了无信号段的残留噪声；反之增大信号失真，减小残留噪声。本质上，不变阈值只能在改善信噪比和降低失真之间寻求一种折衷方案，不可能真正解决信噪比和失真之间的矛盾。

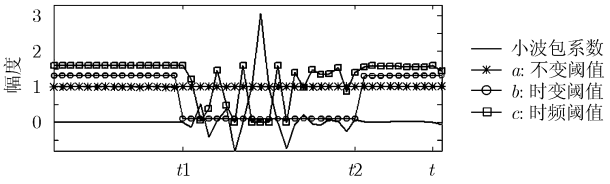

图 1 阈值函数示意图

如何解决信噪比和失真之间的矛盾，这里首先回顾一下频域语音增强算法中用到的方法。根据听觉掩蔽效应[8,9]，在频域上强语音信号对弱噪声信号有掩蔽作用，语音谱能量强的时候该频点噪声听起来不刺耳，当语音某频率点的谱能量强到一定程度(高于听觉掩蔽阈值)时，该频点的噪声不可闻。语音分为元音和辅音，语音能量主要集中元音上，而元音能量集中于 3 个主共振峰上。在语音谱能量强的频段，特别是元音的共振峰附近，由于掩蔽效应的存在，噪声听起来不明显，而语音谱能量弱的频段和语音的间隙期，噪声对人耳听觉的影响较大。因此，基于掩蔽效应的增强算法，其基本思路是，在语音谱能量强的频率段，尽可能地多保留信号，以减少失真；语音谱能量弱的频段，尽可能地抑制噪声，提高信噪比。听觉掩蔽的思想也可用于小波域降噪，可惜的是，小波域与频域是两个不同的信号域，小波包的各子带和频谱之间并没有严格的对应关系，小波包的每一层的子带都覆盖信号所占有的所有频率，只是各层的频率分辨率不同。在小波域应用掩蔽效应，可以简单地理解为小波域信号能量强的时刻，阈值设低一点，能量弱的时刻，阈值设高一点。

如图 1 b 线段所示阈值，若检测到有信号活动时，将阈值降低，无信号时增高阈值。这便是 ANE 阈值的基本思想。ANE 算法是在不变阈值的基础上，乘以调节因子 $\beta^j(l)$，调整阈值使之能够反映语音信号的时变包络，因此，文献[7]中称之为时变阈值。时频阈值与时变阈值的根本区别在于：时频阈值能够反映信号不同子带不同时刻的时频变化细节，如图 1c 线段所示；而时变阈值不能。首先分析时变阈值在同一子带的不同时刻情况。从图 1 可以看出，信号段 $t1$ 至 $t2$ 时段，小波系数由若干峰值和谷值组成，时变阈值使信号段的峰、谷值信号均得以较多的保留，信号段仍有较大的残留噪声。再看不同子带的情况。文章前面提到，小波包的每一层的子带都覆盖信号所占有的全部频率，同样，单一频率的信号也非均匀分布于小波包的各个子带上。在强噪声环境下，时变阈值仅能从信号能量强的子带中提取出能量包络，其他子带的信号分量都将被噪声淹没。

时频阈值采用频域语音粗估计信号经过小波域的变换，来提取信号时频信息，然后利用时频信息动态的调节阈值，如图 2 所示。时频阈值除了能够清晰准确地描述语音在小波域随时间、频率变化的细节，还具有很强的抗干扰性，能够较好地消除极低信噪比下带噪语音的背景噪声。这是因为：首先，语音有间隙，语音的能量仅存在于有语音的时段，有语音时段的瞬时信噪比要大于整段语音的平均信噪比；其次，在频域上，语音的能量主要集中在元音上的 3 个主共振峰上，在主共振峰频率点附近的信噪比要远远大于全频段语音信号的信噪比。在强背景噪声的环境下，语音短时频谱的局部频段上，仍保持较高的信噪比，采用频域算法增强后，频谱的失真并不是很大，但是残留音乐噪声却是频域增强算法难以克服的问题。TF 算法，采用频域算法估计语音，并经过小波包变换，提取语音的时频参数，只要纯净语音的频谱估计越接近于原始语音，时频阈值就越能反映语音信号时频变化的细节，同时，频域增强算法的强抗干扰性能，通过小波变换，传递给阈值，通过在小波域降噪，去除了频域增强残留的音乐噪声。

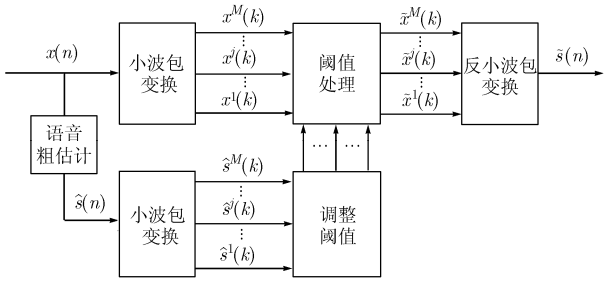

图 2　TF 算法原理框图

3.2 时频阈值的计算

时频阈值的计算方法详细过程描述如下：

(1)采用传统的频域增强算法如谱减法[10]，MMSE 法[11]，Log-MMSE 法[12]等，估计纯净语音 $\hat{s}(n)$；

(2)将估计纯净语音经过小波包变换分解，得到时频参数 $\hat{s}^j(k), k=1,2,\cdots,N^j$；

(3)利于用式(7)计算新的阈值 $T^j(k)$：

$$T^j(k)=\max(T_0-\left|\hat{s}^j(k)\right|,0), k=1,2,\cdots,N^j \qquad (6)$$

3.3 基于时频阈值的小波包增强算法

带噪语音信号被分为两路并行处理，其中一路用于小波包阈值降噪处理，另一路通过频域方法得到纯净语音的初估计，利用粗估计语音的时频信息调整阈值。考虑到频域初估计的方法是以帧为单位进行，而小波包变换是以样点为单位进行，为保障频域和小波域方法的同步，增强结果将产生一帧数据的延迟。这种延迟(依据频域初估计方法中的帧长确定)通常非常小，不会对听觉产生影响。如图 2，简述增强算法计算流程如下：

(1)截取语音起始段噪声信号，经过小波包变换分解，采用式(1)计算阈值初值；

(2)带噪语音经过小波包变换分解得到 M 个子带 $x^j(k)$；

(3)利用 3.2 节中的方法计算时频阈值；

(4)阈值函数降噪，采用式(2)软判决法；

(5)将阈值降噪后的系数经过小波包反变换，合成增强语音信号。

4　仿真实验

实验条件：从 863 语音库中截取 100 条语音样本，采样率为 8kHz，叠加高斯白噪声 $N(0,\sigma_w^2)$，通过改变噪声方差 σ_w^2 来控制污染噪声语音的输入信噪比，范围从−3dB 到 9dB；选择 sym8 小波基对所有语音样本进行 4 级分解，实验选用小波包不变阈值、ANE 阈值和 TF 算法结果进行比较。TF 算法采用 LogMMSE 方法[12]进行纯净语音估计。实验选用 MOS 评分主观方式，和信噪比、MBSD 测度[13]两种客观方式进行评测。为了对比时频阈值算法的抗干扰性能，实验同时列举了以原始纯净语音代替估计语音计算时频阈值算法(简称纯净语音 TF 算法)的评测结果，这是一种理想情况，这里仅仅用于对 TF 算法进行近一步的分析。

4.1 主观评测方法：MOS 评分

30 人参加测试。根据被测声音的残留噪声、语音清晰度和可懂度等情况给出综合选择评分(MOS)，MOS 评分范围从 1 到 5 代表主观感觉的优劣。计算评分平均值，见表 1。

表 1　MOS 评分

输入信噪比	不变阈值	ANE	TF	纯净语音 TF
9dB	3.2	4.0	**4.2**	4.3
6dB	2.8	3.5	**3.8**	3.9
3dB	2.3	3.0	**3.5**	3.7
0dB	2.1	2.6	**3.2**	3.5
−3dB	1.6	2.2	**2.8**	3.1

4.2　客观评价方法

(1)输入信噪比 $\mathrm{SNR}_{\mathrm{in}}$ 和输出信噪比 $\mathrm{SNR}_{\mathrm{out}}$　为比较去噪效果，定义输入信噪比(input SNR) $\mathrm{SNR}_{\mathrm{in}}$

$$\mathrm{SNR}_{\mathrm{in}}=10\lg\sum_{n=1}^{N}s^2(n)\Big/\sum_{n=1}^{N}w^2(n)$$

划分语音段和噪声段，令 N_x 和 N_w 分别为增强结果的语音段样本长度和噪声段样本长度，$\tilde{s}'(n)$ 和 $\tilde{w}'(n)$ 分别为增强结果的语音样本和噪声样本。定义输出信噪比(output SNR) $\mathrm{SNR}_{\mathrm{out}}$

$$\mathrm{SNR}_{\mathrm{out}}=10\lg\left[\sum_{n=1}^{N_x}\tilde{s}'^2(n)/N_x-\sum_{n=1}^{N_w}\tilde{w}'^2(n)/N_w\Big/\sum_{n=1}^{N_w}\tilde{w}'^2(n)/N_w\right]$$

无语音段采取手工划分，计算输入、输出信噪比，见图 3(a)。

(2)MBSD 测度　MBSD 测度是将噪声掩蔽阈值引进到

传统的 BSD(Bark Spectral Distortion)测度，得到不计噪声低于听觉掩蔽阈值的失真[13]。MBSD 能够更好地反映对人耳听觉影响的失真测度。实验结果见图 3(b)。

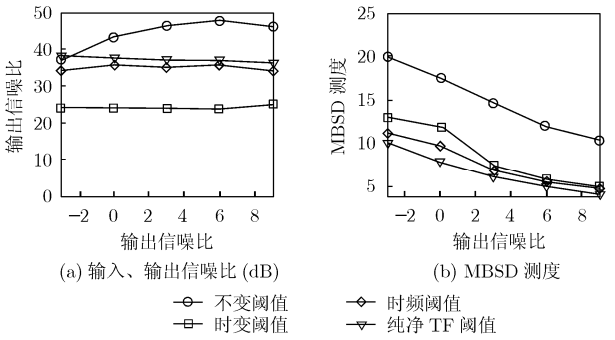

图 3 客观评价结果图

4.3 分析

通过主客观评测可知：主观听觉上，TF 算法清晰度和可懂度均优于不变阈值算法和 ANE 算法，且没有频域增强算法所特有的残留音乐噪声，听起来较悦耳。比较输出信噪比可以看出，TF 算法保持了小波域降噪的高信噪比的特点，增强后的语音几乎听不到残留噪声。由于利用了听觉掩蔽的原理，ANE 算法和 TF 算法输出信噪比较不变阈值算法有所降低，但换来低失真的增强语音。相比 ANE 算法，TF 算法输出信噪比提高，而失真降低。这一点可以从 MBSD 测度看出：TF 算法的失真远小于不变阈值算法，并且低于 ANE 算法。

比较 TF 算法和纯净语音代替估计语音的 TF 算法，后者主客观评价均优于前者一点，这说明：(1)选择更好的频域估计算法，可以进一步提高 TF 算法的性能。LogMMSE 算法虽然是频域较优异的算法，它通过平滑，衰减了部分音乐噪声，但却增大了谱失真。小波域增强可以抑制那种扰人的音乐噪声，因此，LogMMSE 对于 TF 算法而言，并不一定是好的估计算法。(2)频域估计的提高对 TF 算法增强效果的改进有限。(3)频域增强算法的强抗干扰特性，通过时频变换，传递到小波域，使得 TF 算法在低信噪比环境下也有良好表现。

5 结束语

在分析和总结前人小波包阈值去噪方法的基础上，提出时频阈值的概念，并应用于改进小波包降噪阈值，进行语音增强。与传统小波包阈值的区别在于，时频阈值是在强噪声干扰下，依然能够准确反映原始语音在小波域变化的时频细节。因此，新算法保持了小波域增强算法高信噪比，无残留音乐噪声的特点，同时又具备频域干扰性强，失真小的优点。改进粗估计算法和阈值调制方法，可以进一步地提升增强性能。

参 考 文 献

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

[2] Donoho D L and Johnstone I M. Ideal spatial adaptation by wavelet shrinkage [J]. *Biometrika*, 1994, 81(3): 425–455.

[3] Shao Yu and Chang C H. A versatile speech enhancement system based on perceptual wavelet de-noising [C]. Kobe, Japan, IEEE International Sym. on Circuits and Systems, 2005, 2: 864–867.

[4] Ayat S and Manzuri M T. Wavelet based speech enhancement using a new thresholding algorithm [C]. Hongkong, International Sym. on Intelligent Multimedia, Video and Speech Processing, 2004, 10: 238–241.

[5] Chen S H and Wang J F. Speech enhancement using perceptual wavelet packet decomposition and teager energy operator [J]. *Journal of VLSI Signal Processing*, 2004, 36(2): 125–139.

[6] Bahoura M and Rouat J. Wavelet speech enhancement based on the teager energy operator [J]. *IEEE Signal Processing Letters*, 2001, 8(1): 10–12.

[7] Lei S F and Tung Y K. Speech enhancement for nonstationary noises by wavelet packet transform and adaptive noise estimation [C]. Hongkong, International Sym. on Intelligent Signal Processing and Comm. Systems, 2005: 41–44.

[8] Virag N. Single channel speech enhancement based on masking properties of the human auditory system [J]. *IEEE Trans. on Speech and Audio Processing*, 1999, 7(2): 126–137.

[9] Chen Q, and Guo Y, *et al*.. A LSA-MMSE speech enhancement approach incorporating masking properties [M]. China, International Conference on Comm., Circuits and Systems Proc., 2006: 455–458.

[10] Boll S F. Suppression of acoustic noise in speech using spectral subtraction [J]. *IEEE Trans. on Acoust. Speech Signal Processing*, 1979, 28(2): 113–120.

[11] Ephraim Y and Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator [M]. *IEEE Trans. on Acoust. Speech Signal Processing*, 1984, 32(6): 1109–1121.

[12] 蔡斌. 语音增强方法研究及其应用. [硕士论文], 空军工程大学, 2004.

[13] Yang W, Dixon M, and Yantorno R. A modified bark spectral distortion measure which uses noise masking threshold [M]. Pocono Manor, USA, Pennsylvania, Speech Coding for Telecommunications Proc., 1997: 55–56.

徐耀华： 男，1975 年生，博士生，研究方向为信号处理.

王 刚： 男，1976 年生，讲师，博士，研究方向为统计信号处理、模式识别.

郭 英： 女，1963 年生，教授，博士生导师，研究方向为信号与信息处理、模式识别.