第 30 卷第 4 期　　声 学 学 报　　Vol. 30, No. 4
2005 年 7 月　　ACTA ACUSTICA　　Jul., 2005

基于听觉掩蔽效应和 Bark 子波变换的语音增强*

陶 智[1]　赵鹤鸣[2]　龚呈卉[1]

(1 苏州大学物理学院　江苏苏州　215006)
(2 苏州大学电子信息学院　江苏苏州　215006)

2003 年 11 月 21 日收到
2004 年 5 月 17 日定稿

摘要　提出了一种适用于低信噪比下的提高语音的听觉效果的语音增强方法。该方法在谱减法的基础上有两个特点：首先减参数是根据人耳听觉掩蔽效应提出的且是自适应的；其次采用了与人耳听觉系统特性更为适应的 Bark 子波变换方法对增强前后的语音进行分析。对该算法进行了客观和主观测试，结果表明：与谱减法相比对低信噪比的语音信号，(1) 能更好地抑制残留噪声和背景噪声，(2) 增强后的语音具有更好的清晰度和可懂度。

PACS 数：43.60, 43.70

Speech enhancement based on masking properties of human auditory system and bark wavelet transform

TAO Zhi[1]　ZHAO Heming[2]　GONG Chenghui[1]

(1 *Dept.of phys., School of Sci., Suzhou Univ.* Jiangsu Suzhou 215006)
(2 *Dept.of Electron, School of Information, Suzhou Univ.* Jiangsu Suzhou 215006)

Received Nov. 21, 2003
Revised May 17, 2004

Abstract　A speech enhancement method under very low signal-to-noise ratios is proposed. There are two characteristics in the method based on and compared with the spectral subtraction. First, the subtraction parameter is proposed based on the masking properties of human auditory and is self-adaptive. Second, Bark wavelet transform is used, which is more suitable to human ears' auditory system, to give the before-and-after voice speech sound analysis. By performing an objective and subjective test, it shows that two improvements are made in the very low signal-to-noise ratio speech compared to the spectral subtraction algorithms. (1) reduce the residual noise and background noise more effectively. (2) the enhanced speech is more clearness and intelligibility.

引言

处理宽带噪声最通用的技术是谱减法[1]，但会产生不舒服的“音乐噪声”。谱减法的改进形式[2]是以改变减参数对噪声减弱、语音失真和音乐残留噪声作出权衡，但受到混合优化参数的限制。Berouti[3−5]等提出的方法部分解决了这种残留噪声，但门限的选取比较困难，当信噪比较低时，剩余噪声还是很大。

近年来，人们针对听觉外周提出了一些计算模型，并在语音编码、音频压缩和音质的客观度量等方面获得了应用，同时，基于人类听觉特性的语音增强研究也取得了一定的进展[6−7]。目前，在语音增强中用得比较成功的是掩蔽效应，它指出语音信号能够掩蔽与其同时进入听觉系统的一部分能量较小的噪声信号，而使得这部分噪声不为人所感知到。本文根据这种特性来自适应地设定和调整系统的相关参数，以有效地掩蔽残留噪声和最大限度地保留语音。

谱减法及其改进形式采用傅里叶变换的方法对语音信号进行分析，这种分析时频分辨率固定不变，有其局限性。许多研究者利用小波变换及小波包变换对语音信号进行分析，使本来不易察觉的信号特征在不同的分辨率的若干子空间中显露出来，但这一点对语音分析来说并不必要，因为听觉感知本身就存在较大的冗余；并且，无论是二进制、小波包还是 M 带小波变换[7−9]，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

* 国家自然科学基金 (60172016) 及苏州大学青年基金 (Q3108404) 资助项目

不完全吻合。

为此，本文提出一种针对语音信号的 Bark 子波变换的语音增强。其基函数满足时间－感知频率上的最佳不确定性，而分析尺度的伸缩则按照“临界带”的概念来变化，并使得每一尺度下的带宽为一个“频率群”。通过这样的构造方法所得的子波变换无疑具有与听觉系统十分吻合的分析特性，使增强后的语音有更高的清晰度和可懂度。

1 语音增强的原理

设语音增强系统的增益函数为 $G(\omega)$，则增强语音的频谱 $|\widehat{S}(\omega)|$ 为带噪语音的短时频谱 $|Y(\omega)|$ 乘以系统增益函数 $G(\omega)$[3]，即：

$$\widehat{S}(\omega)=G(\omega)|Y(\omega)|,\qquad 0\le G(\omega)\le 1, \tag{1}$$

采用功率谱的形式，可得：

$$G(\omega)=\sqrt{1-\frac{|\widehat{D}(\omega)|^2}{|Y(\omega)|^2}}, \tag{2}$$

其中 $|\widehat{D}(\omega)|$ 为噪声功率谱。

增益函数 $G(\omega)$ 在噪声成分和语音成分之间变化：只有语音信号时，$G(\omega)=1$；只有噪声成分时，$G(\omega)=0$。在两种极限情况之间，增益值取决于 SNR:

$$\mathrm{SNR}_{\mathrm{post}}(\omega)=\frac{|Y(\omega)|^2}{|\widehat{D}(\omega)|^2}, \tag{3}$$

Berouti[3] 等人提出的算法采用的增益函数为：

$$G(\omega)=G[\mathrm{SNR}_{\mathrm{post}}(\omega)]=\begin{cases}\left\{1-\alpha\left[\dfrac{|\widehat{D}(\omega)|}{|Y(\omega)|}\right]^{r_1}\right\}^{r_2}, & \left[\dfrac{|\widehat{D}(\omega)|}{|Y(\omega)|}\right]^{r_1}<\dfrac{1}{\alpha+\beta},\\ \left\{\beta\left[\dfrac{|\widehat{D}(\omega)|}{|Y(\omega)|}\right]^{r_1}\right\}^{r_2}, & \text{else},\end{cases} \tag{4}$$

其中 $\alpha(\alpha>1)$ 为过减因子，增加 α 可以使残留噪声的峰值减少，但也增加了听觉失真；$\beta(0\le\beta\le 1)$ 为频谱阶，导致残留噪声减少，但增加了增强语音中的背景噪声；γ 为指数，它决定频谱从 $G(\omega)=1$(频谱成分未发生改变) 至 $G(\omega)=0$(频谱成分完全抑制) 的平滑转变。

参数 α 和 β 的选择是语音增强的关键所在，传统的方法可看作公式 (4) 的特例：当 $\alpha=1,\beta=0$ 且固定不变时，即为经典的功率谱减法 (PSS)[1]，其特点是残留噪声较大；对 α 和 β 进行优化，但为固定时，即为修正功率谱减法 (MSS)[4]，其效果虽有所改善但不理想；而在非线性谱减法 (NSS)[5] 中 β 取较小值 (如 0.01)，α 可以根据信噪比的变化而变化，其减噪效果有很大的改善，但在一帧语音信号中参数 α 也是固定的。另外，在低信噪比的情况下，这些方法是无法同时将语音失真和残留噪声降到最低的。

为此在我们实现的系统中，首先计算每一帧语音的不同 Bark 域的噪声掩蔽阈值，然后根据噪声掩蔽阈值得到自适应的减参数 α 和 β：若掩蔽阈值较高，残留噪声会很自然地被掩蔽而使人耳听不见，在这种情况下，减参数取它们的最小值；掩蔽阈值较低时，残留噪声对人耳的影响很大，有必要去减少它。对于每一帧 m，掩蔽阈值 $T_m(\omega)$ 的最小值与参数 $\alpha_m(\omega)$ 和 $\beta_m(\omega)$ 的最大值有关。减参数的应用有如下关系式：

$$\alpha_m(\omega)=F_\alpha[\alpha_{\min},\alpha_{\max},T(\omega)], \tag{5}$$

$$\beta_m(\omega)=F_\beta[\beta_{\min},\beta_{\max},T(\omega)], \tag{6}$$

其中 $\alpha_{\min}$，$\beta_{\min}$ 和 $\alpha_{\max}$，$\beta_{\max}$ 分别是参数 α 和 β 的最小值和最大值，F_α 和 F_β 是求减少最大残留噪声的函数：当 $T(\omega)=T(\omega)_{\min}$ 时，$F_\alpha=\alpha_{\max}$；当 $T(\omega)=T(\omega)_{\max}$ 时，$F_\alpha=\alpha_{\min}$。式中 $T(\omega)_{\min}$ 和 $T(\omega)_{\max}$ 分别是逐帧得到的掩蔽阈值的最小值和最大值。

本文提出的增强系统框图如图 1 所示。

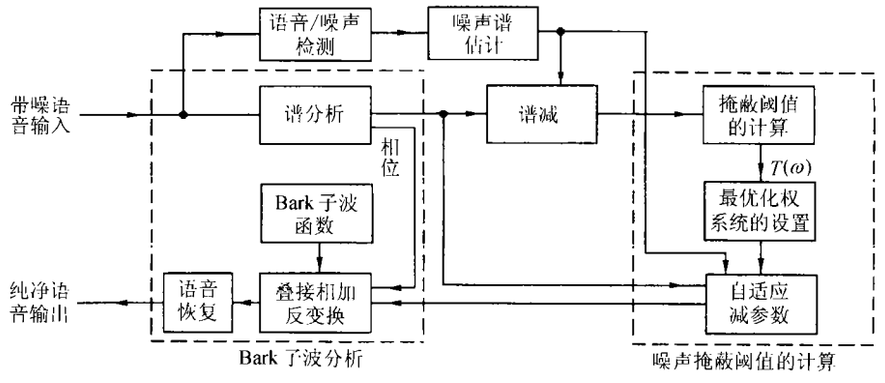


图 1 增强系统框图

2　噪声掩蔽阈值的计算

噪声掩蔽阈值[10]的计算由以下几个部分组成：

(1) 频率群的分析

时域语音信号 $x(t)$ 经过快速傅里叶变换 (FFT) 变成频域信号 $X(\omega)$, 信号的功率谱为：

$$P(\omega) = \mathrm{Re}^2X(\omega) + \mathrm{Im}^2X(\omega), \tag{7}$$

将语音信号的功率谱按频段 (Bark 域)[6] 逐一分成小段，计算每一段的能量，即：

$$B_i = \sum_{\omega=b_{li}}^{b_{hi}} P(\omega). \tag{8}$$

其中 B_i 表示第 i 段的能量，b_{li} 表示第 i 段最低的频率，b_{hi} 表示第 i 段最高的频率。

(2) 扩散 Bark 域功率谱

根据文献，引入扩散矩阵 $\boldsymbol{S}$, 满足条件：

$$\mathrm{abs}(j-i) \le 25, \tag{9}$$

其中 i 是已被掩蔽信号的 Bark 频率，j 是正被掩蔽信号的 Bark 频率，S_{ij} 为该矩阵 $\boldsymbol{S}$ 中的元素。

将矩阵 S_{ij} 与 B_i 相互卷积就可以得到扩散 Bark 频域谱 C_i, 即：

$$C_i = S_{ij} * B_i. \tag{10}$$

(3) 噪声掩蔽阈值的计算

有两种噪声掩蔽门限：一种是纯音掩蔽噪声 (TMN), 是在 C_i 下面 $14.5+i$ dB; 另一种是噪声掩蔽纯音 (NMT), 是在 C_i 下面 5.5 dB。其中 i 的值是相等的，但在 C_i 中的 i 是频段，而 $(14.5+i)$ 中 i 是指 dB。

为了能够辨别信号是纯音和噪声，给出系数 SFM:

$$SFM_{db} = 10\lg\frac{G_m}{A_m}, \tag{11}$$

其中为 G_m 该语音信号的几何平均，A_m 为该语音信号的算术平均。

设定系数 α:

$$\alpha = \min\left(\frac{SFM_{db}}{SFM_{db\max}}, 1\right). \tag{12}$$

当 $\alpha = 0$, 完全是噪声；$\alpha = 1$, 完全是纯音。实际语音信号既非噪声，又非纯音，α 介于两者之间。

通过 α 的数值可以判断这个信号是偏噪声的，还是偏语音的。

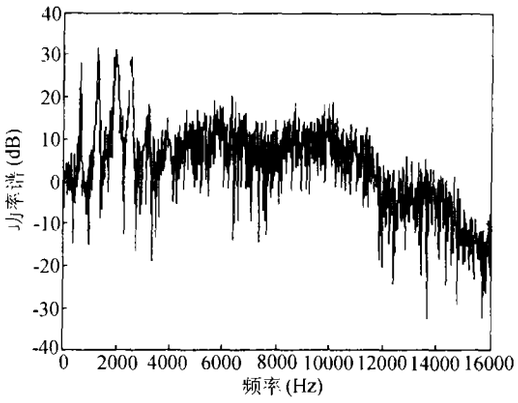
掩蔽能量的偏移函数为：

$$O_i = \alpha(14.5+i) + (1-\alpha)5.5. \tag{13}$$

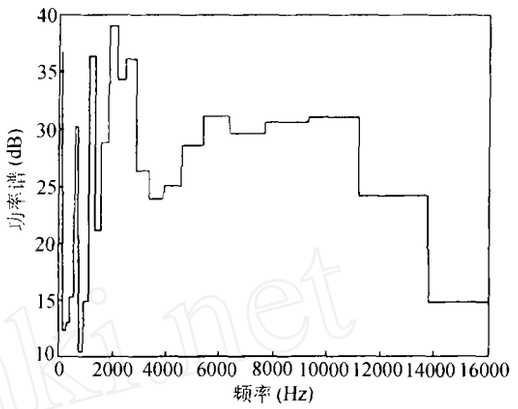
则噪声掩蔽阈值[9]为：

$$T_i = 10^{\lg(C_i)-(O_i/10)}. \tag{14}$$

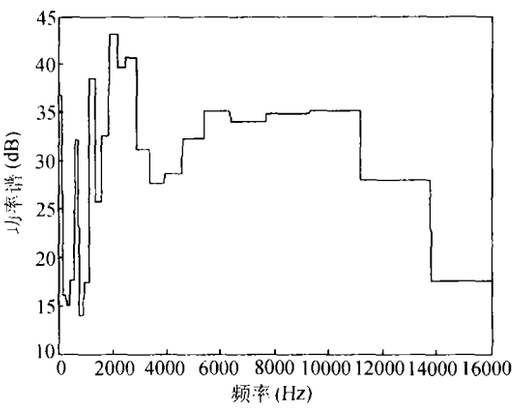
图2是一帧带噪语音掩蔽阈值计算的实验结果。



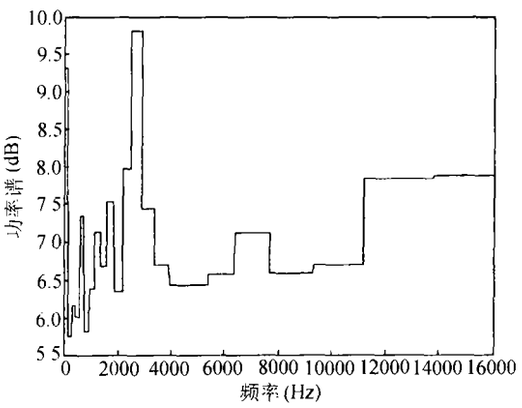
(a)　功率谱



(b)　Bark 域功率谱

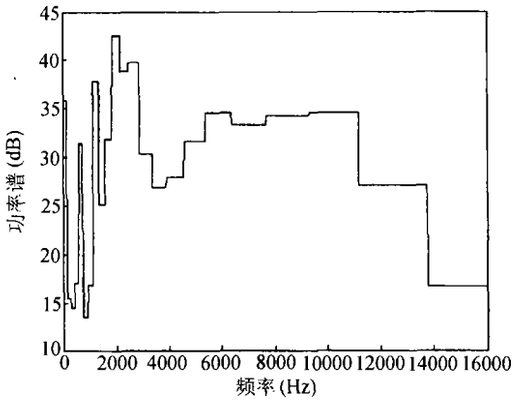


(c)　扩散 Bark 功率谱



(d)　扩散 Bark 功率谱补偿

图 2



(e) 掩蔽阈值

图 2 带噪语音掩蔽阈值计算举例

3 Bark 子波分析

3.1 Bark 子波简介

人耳基底膜具有与频谱分析器相似的作用，在 20 ~ 16000 Hz 范围内的频率可分成 24 个频率群 (临界带， Bark)[11]，频率群的划分相应于基底膜分成许多很小的部分，每一部分对应一个频率群，并且长度相等。对应于同一基底膜部分的那些频率的声音，在大脑中似乎是叠加在一起进行评价的[12]。这就是说，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。这就引出了所谓的 Bark 尺度的概念， Traunmullar 给出了线性频率与 Bark 频率之间的函数关系[13]，即：

$$b = 6.7a\sinh[(f-20)/600], \tag{15}$$

其中， b 代表 Bark 频率， f 为线性频率。

图 3 给出线性频率域 Bark 子波簇示意图。

构造 Bark 子波的基本思想是：首先由于在语音分析中时间与频率信息的同等重要性，所以选择的子波母函数 (以下简称母函数) 应在 Bark 域满足时间 - 带宽最小，即为 Bark 域的高斯函数；其次为与频率群的概念相一致，母子波在 Bark 域的带宽应相等且为单位宽度，即 1 个 Bark 。

根据以上两点，选择母函数在 Bark 的形式为 $W(b) = e^{-c_1 b^2}$，并且在单位带宽定义为 3 dB 带宽的情况下 c_1 取为 $4\ln 2$ 。假定被分析语音信号 $s(t) \sim S(f)$ 的线性频率带宽 $|f| \in [f_1, f_2]$，相应的 Bark 频率带宽为 $[b_1, b_2]$ 。定义子波函数 Bark 域形式为：

$$W_k(b) = W(b - b_1 - k\Delta b), \quad k = 0, 1, \cdots, K-1 \tag{16}$$

式中， Δb 为 $W_k(b)$ 的平移步长，根据 Bark 域等带宽原则，有 $\Delta b = (b_2 - b_1)/K - 1$, K 为尺度参数。因为 $W(b) = e^{-c_1 b^2}$，所以

$$W_k(b) = e^{-4\ln 2(b-b_1-k\Delta b)^2} = 2^{-4(b-b_1-k\Delta b)^2}, \quad k = 0, 1, \cdots, K-1 \tag{17}$$

再将式 (15) 代入式 (16) 。就得到了线性频率下的 Bark 子波函数表达式

$$W_k(f) = c_2 2^{-4\{6.7a\sinh[(f-20)/600-(b_1+k\Delta b)]\}^2}, \tag{18}$$

这里 $b = 6.7a\sinh[(f_1 - 20)/600]$ 。式中 c_2 为规整因子，在频域定义 Bark 子波变换为：

$$s_k(t) = \int_{-\infty}^{\infty} S(f)W_k(f)e^{j2\pi ft}df = \int_{-\infty}^{\infty} c_2 S(f)2^{-4\{6.7a\sinh[(f-20)/600-(b_1+k\Delta b)]\}^2}e^{j2\pi ft}df = c_2\int_{-\infty}^{\infty} S(f)\exp\{-4\ln 2[6.7a\sinh((f-20)/600-(b_1+k\Delta b))]^2 + j2\pi ft\}df \quad k = 0, 1, \cdots, K-1 \tag{19}$$

3.2 Bark 子波变换在语音增强中的应用

设某一帧带噪语音信号 $s(t)$ 的均匀采样为 $s(n)$ 。采样率为 f_s，帧长为 N 。其 N 点 FFT 为 $S'(l)$, $l = 0, 1, \cdots, N-1$ 。对 $S'(l)$ 进行噪声掩蔽阈值的去噪，得到 $S(l)$ 。

实际应用采用 Bark 子波及其变换的离散形式。(18) 式的 Bark 子波变换可离散化为 $W_k(N-i) = W_k(i-1) = W_k[i(f_s/N)]$, $i = 1, 2, \cdots, N-1$ 。

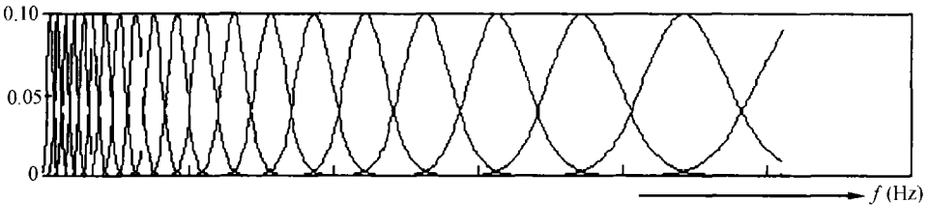


图 3 线性频率域下 Bark 子波簇示意图

则 k 尺度下的子波变换为：

$$S_k(n)=\sum_{l=0}^{N-1}S(l)w_k(l)e^{\mathrm{j}2\pi inN}.$$

Bark 子波的 N 个尺度下的分解信号叠加而成重构信号：

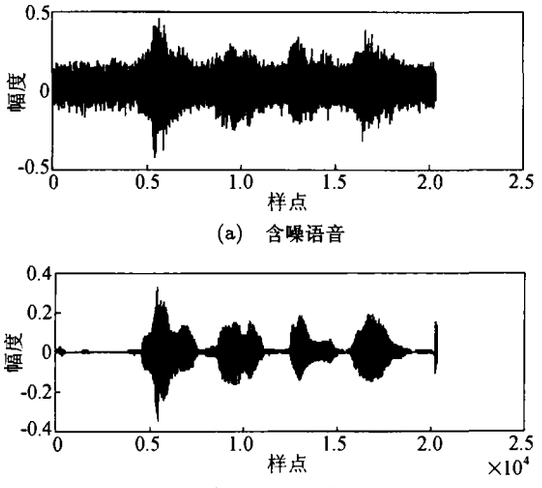
$$S(n)=\sum_{k=0}^{N-1}S_k(n).$$

本文做了以下两个实验：(1) 直接采用 FFT 和 IFFT 进行语音增强，(2) 采用 Bark 子波变换进行语音增强。与基于 FFT 的谱减法比较，本文方法具有以下特点：(1) 对于频率成分复杂的语音信号，在服从不确定性原理的前提下，使不同的时－频区都能获得比较合适的时－频分辨率。(2) 在感知特性上是与人耳的听觉系统特性十分吻合。实验结果也表明：采用 Bark 子波变换的增强语音比不用 Bark 子波变换的增强语音在信噪比上有明显的提高，而且增强后的语音更加符合人耳的听觉感知。

4　实验结果

实验中采用的语音材料选自中文语音平均意见得分 (MOS) 测试库[15]，共 6 位发音人 (3 男 3 女)，每人一句。噪声材料为平稳白噪声。语音和噪声信号经 44.1 kHz 采样，16 bit 量化为数字信号，并在计算机中按一定比例混合生成不同信噪比 $(-6,-3,0,3,6$ 和 9) 的带噪语音。带噪语音的长度为 256 点的语音帧，相邻两帧之间重叠 128 点，然后对每一帧带噪语音逐帧进行增强处理。

图 4 是一个语音增强实验结果的例子。(a) 为被平稳的白噪声污染后的带噪语音信号，(b) 为增强后的语音。



(a)　含噪语音

(b)　增强后的语音

图 4　含噪语音增强结果 (发音为“为国争光”)

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所提算法进行比较：(1) 谱减法 (2) 小波变换法 (3) 听觉模拟法。结果用表 1 来表示。

信噪比计算公式[19] 如下：

$$G_{\mathrm{SNR}}=\frac{1}{L}\sum_{m=0}^{L-1}10\lg\frac{\frac{1}{N}\sum_{n=0}^{N-1}d^2(n+Nm)}{\frac{1}{N}\sum_{n=0}^{N-1}[s(n+Nm)-\widehat{S}(n+Nm)]^2}\ [\mathrm{dB}],\tag{20}$$

其中 L 表示信号的帧长度，N 表示每帧的采样点。

表 1　4 种方法输出信噪比比较

输入信噪比 /dB	−6 dB	−3 dB	0 dB	3 dB	6 dB	9 dB
谱减法 /dB	2.47	4.72	7.20	9.55	12.05	14.37
小波变换法 /dB	4.71	6.75	8.73	10.09	12.33	14.49
听觉模拟法 /dB	4.95	7.07	8.96	10.48	12.50	14.56
本文方法 /dB	5.71	7.71	9.35	10.67	12.61	14.59

为了确证客观性能评估，我们进行了主观听觉测试。听觉测试是在 10 个听众中进行，内容是对语音的残留噪声、仍存在的背景噪声和语音失真进行全面认识。测试信号录在磁带上，用耳机进行实验。对于每个话音都有下列步骤：(1) 纯净语音和带噪语音均被重复播放两次，(2) 每个测试信号都被重复两次，且以随机顺序播放三次。

测试的结果表明：利用本方法增强的语音在初始信噪比为 −5 dB 以上时，没有残留音乐噪声。在信噪比更低的情况下，残留噪声对语音的干扰比频谱减法要小得多。

5　结论

单通道谱减系统在减少背景噪声上很有效，然而它带来了可感知的令人烦躁的“残留噪声”。在本文中，提出了基于人耳听觉掩蔽效应的语音增强过程，并对增强前后的语音进行 Bark 子波变换。提出的算法提高了对低输入信噪比的改进。本文用该算法对不同信噪比的带噪语音进行测试并和传统的方法做对比。通过实验结果和信噪比得到的客观评估，结合主观听觉结果显示：与传统方法相比，背景噪声和残留噪声都减少了，而语音失真也在可接受的范围内。因此，可以得出这样的结论：基于听觉特性的 Bark 子波变换的语音增强过程，因为更多地考虑了人耳听

觉特性，所以比传统算法有较大改进，特别是在低信噪比的情况下，语音具有更好的清晰度和可懂度。

本文所处理的还只限于平稳随机噪声，对噪声特性变化剧烈的含噪语音，上述方法有其局限性。而许多环境下的干扰噪声是非平稳的，人类的听觉系统还是能从这类噪声中提取有用信息，因此研究非平稳随机噪声下的语音增强具有重要的意义，这方面的工作还有待进一步研究。

参 考 文 献

1 Boll S F. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech. Signal Processing. Apr.*, 1979; **ASSP-27**: 113—120

2 Lim J S, Oppenheim A V. Enhancement and bandwidth compression of noisy speech. *Proc. of the IEEE. Dec.*, 1979; **67**(12): 1586—1604

3 Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise. In: proc. IEEE ICASSP, Washington. DC, 1979: 208—211

4 Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans Acoust Speech Signal Processing.* 1985; **33**(2): 443—445

5 Lockwood P, Boudy J. Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection,for robust recognition in cars. *Speech Commun*, 1992; **11**(6): 215—228

6 Tsoukalas D E, Mourjopoulos J N, Kokkinakis G. Speech enhancement based on audible noise suppression. *IEEE Transactions on SPEECH and Processing.* 1997; **5**(6): 497—514

7 朱学文等. 帧同步混合小波包变换模拟听觉模型的语音增强的研究. 声学学报, 2003; **28**(1): 12—16

8 Seok J W, Bae K S. Speech enhancement with reduction of noise components in the wavelet domain. Copyright 1997 IEEE: 1323—1326

9 沈亚强, 金洪震. 一种基于子波变换的语音增强. 科技通报, 2000; **16**(3): 206—211

10 Johnston J D. Transform coding of audio signal using perceptual noise criteria. *IEEE J. Select. Areas Commun*, 1983; **6**(2): 314—323

11 赵 力. 语音信号处理. 机械工业出版社, 2003

12 Evangelista G, Cavaliere S. Discrete frequency warped wavelets:theory and applications. *IEEE Trans. on Signal Processing*, 1998; **46**(4): 874—885

13 Traunmullar H. Analytical expression for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, 1990; **88**(6): 97—100

14 Daubechies I. Ten lectures on wavelets. Philadelpha: SIAM, 1992

15 Deller J, Proakis J, Hansen J. Discrete-time processing of speech signals. Englewood Cliffs. NJ: Prentice-Hall, 1993

16 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. 电子学报, 2000; **28**(10): 102—105

17 Gulzow T, Engelsberg A, Heute U. Comparison of a discrete wavlet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement. suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech. Signal Processing.* 1998; **64**(7): 5—19

18 Soon I Y, Koh S N, Yeo C K. Noisy speech enhancement using discrete cosine transform. *Speech Communication*, 1998; **24**(3): 249—257

19 Nathalie Virag. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on Speech and Audio Processing*, 1999; **17**(2): 126—137

20 Unoki M, Akagi M. A method of signal extraction from noisy signal based on auditory scene analysis. *Speech Communication*, 1999; **27**(10): 261—279

21 Bronkhorst. The cocktail party phenomenon:a review of research on speech intelligibility in multiple-talker condition. *Acustica*, 2000; **86**(3): 117—128

22 WEI Wei, CHEN Yanpu. Speech enhancement by spectral component selection. Proceedings of ICSP 2000: 674—678

23 Mark Klein, Peter Kabal. Sinal subspace speech enhancement with perceptual post-filtering. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 2002: I-537-I-540

24 蒋文建, 韦 岗. 基于掩蔽特性的噪声环境下语音识别特征. 声学学报, 2001; **26**(6): 516—520

25 姚峰英, 张 敏. 一种增强带噪语音可懂度的新算法. 声学学报, 2002; **27**(6): 529—530