

基于人耳掩蔽效应的电子耳蜗语音增强方法

黄雅婷¹, 陶智¹, 顾济华¹, 赵鹤鸣², 严冬明¹

(1. 苏州大学物理科学与技术学院, 苏州 215006; 2. 苏州大学电子信息学院, 苏州 215006)

摘要: 提出一种符合人耳听觉感知的语音增强方法, 使电子耳蜗能在噪声环境下获得准确的语音信息。利用 Bark 子波变换实现电子耳蜗中的语音处理, 结合人耳听觉系统特性实现语音增强。使用根据人耳听觉掩蔽效应提出的自适应减参数。实验结果表明该算法在低信噪比情况下, 信噪比可提高 30 dB 左右, 更好地抑制了残留噪声和背景噪声, 合成的语音具有较好清晰度和可懂度。

关键词: 电子耳蜗; 语音增强; 掩蔽效应; Bark 子波

Speech Enhancement Method in Electronic Cochlear Implant Based on Human Auditory Masking Properties

HUANG Ya-ting¹, TAO Zhi¹, GU Ji-hua¹, ZHAO He-ming², YAN Dong-ming¹

(1. School of Phys. Sci. and Tech., Suzhou Univ., Suzhou 215006; 2. School of Electronic and Information, Suzhou University, Suzhou 215006)

【Abstract】 To acquire accurate speech information in cochlear implant under noise environments, a speech enhancement in cochlear implant based on human auditory masking properties is proposed. Bark wavelet transform is used to realize speech in electronic cochlear implant and to realize speech enhancement combined with human ear's auditory system. Subtract parameter is got based on the masking properties of human auditory and self-adaptive. The result shows that the SNR output can be improved by 30 dB even if the SNR input is low and can reduce the residual noise and background noise more effectively. The enhanced speech is more clear and intelligible.

【Key words】 electronic cochlear implant; speech enhancement; auditory masking properties; Bark wavelet

电子耳蜗是目前唯一能使全聋患者恢复听力的装置, 它将声能转换成电能, 直接刺激耳蜗内残余的听神经纤维, 使聋人产生听觉^[1]。现有电子耳蜗设备在安静环境下能提供很好的语音信息给使用者, 但在噪声环境下使用者获得的语音信息量会下降。本文提出在电子耳蜗中利用人耳的听觉掩蔽效应进行语音增强, 采用Bark子波变换实现电子耳蜗的CIS语音信号处理。Bark子波基函数满足时间-感知频率上的最佳不确定性, 该子波变换具有与听觉系统很吻合的分析特性。人耳能掩蔽语音信号中能量较小的噪声信号, 使这部分噪声不被人感知, 本文根据此特性自适应地设定并调整算法的相关参数。

1 电子耳蜗语音处理器基本原理

1.1 电子耳蜗的机理

声音信号在耳蜗内以行波方式传播, 声音高频部分在蜗底振幅达到最大, 而低频部分在蜗顶振幅达到最小, 因此, 耳蜗是一个空间的机械频率分析器: 蜗底感受高频信号, 蜗顶感受低频信号, 沿基底膜, 从顶部到底部大致对应的音频范围为 16 Hz ~ 20 kHz。通常将此范围的频率划分为 24 个频率群, 即共有 24 个 Bark 域。

电子耳蜗是通过手术将一组电极植入耳蜗鼓阶内, 电极沿基底膜分布。体外语音处理器产生对应电极的电刺激信号, 体内的刺激电流生成电路产生电刺激脉冲, 兴奋聋人耳蜗内完好的听神经, 各电极附近听神经的兴奋传至大脑皮层语言中枢, 诱发听觉, 进而恢复初步的语言交流能力^[2]。

1.2 电子耳蜗的 CIS 语音信号处理方案

1991 年美国学者 Wilson 提出了电子耳蜗的连续交替取样

(CIS)方案^[3], 其工作原理如图 1 所示。语音信号在预加重后经过 6~8 个带通滤波器, 每个通道经全波整流、低通滤波可得到各通道的包络信号, 再用对数或平方律压缩到适当的动态范围。各电极的刺激脉冲由一组时间上交替的双相脉冲序列调制对应的包络信号获得。

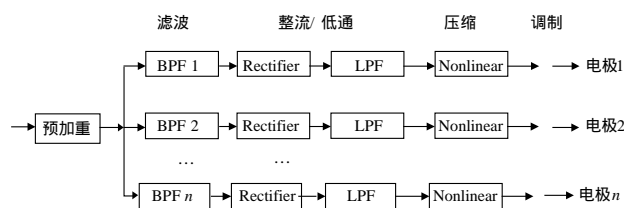


图 1 CIS 语音信号处理方案示意图

CIS 方案是电子耳蜗目前采用的信号处理方法之一, 本文采用 Bark 子波变换代替带通滤波实现电子耳蜗的 CIS 语音信号处理。Bark 子波基函数满足时间-感知频率(弯折频率, 主观频率)上的最佳不确定性(时间-带宽积最小), 而分析尺度的伸缩则按临界带来变化, 并使每一尺度下的带宽为一个频率群。通过这样的构造方法得到的子波变换将具有与听觉系

基金项目: 国家自然科学基金资助项目(60572076); 江苏省高校自然科学基金资助项目(05JKB510113); 苏州大学学生科研基金资助项目

作者简介: 黄雅婷(1983 -), 女, 硕士研究生, 主研方向: 语音信号处理; 陶智, 副教授; 顾济华、赵鹤鸣, 教授、博士生导师; 严冬明, 本科生

收稿日期: 2007-07-28 **E-mail:** jhgu@suda.edu.cn

统相吻合的分析特性。

2 电子耳蜗内语音增强

2.1 电子耳蜗中的语音增强

假设语音为平稳信号，而噪声和语音为加性信号且彼此不相关。此时带噪语音信号可表示为

$$y(t) = s(t) + n(t) \quad (1)$$

其中， $s(t)$ 为纯净语音信号； $n(t)$ 为噪声信号。对功率谱有

$$|Y(w)|^2 = |S(w)|^2 + |N(w)|^2 \quad (2)$$

在实际的增强过程中，更多使用谱减法的推广形式^[4]为

$$G(w) = G[SNR_{post}(w)] = \begin{cases} (1 - \alpha [\frac{\hat{D}(w)}{Y(w)}]^{\gamma_1})^{\gamma_2} & [\frac{\hat{D}(w)}{Y(w)}]^{\gamma_1} < \frac{1}{\alpha + \beta} \\ (\beta [\frac{\hat{D}(w)}{Y(w)}]^{\gamma_1})^{\gamma_2} & \text{else} \end{cases} \quad (3)$$

其中， α 为谱减阈值系数，其值越大，剩余噪声衰减越大，语音失真也越大； β 为谱减噪声系数， β 增大可降低剩余噪声的音乐噪声，但会增加增强后语音的背景噪声。 $\gamma = \gamma_1 = 1/\gamma_2$ ，该参数决定了增益参数从 $G(w) = 0$ 到 $G(w) = 1$ 的平滑程度。对于每一帧 m 每个通道 n ，掩蔽阈值 $T_{m,n}(w)$ 的最小值与参数 $\alpha_{m,n}(w)$ 和 $\beta_{m,n}(w)$ 的最大值有关：

$$\alpha_{m,n}(w) = F_\alpha[\alpha_{\min}, \alpha_{\max}, T(w)] \quad (4)$$

$$\beta_{m,n}(w) = F_\beta[\beta_{\min}, \beta_{\max}, T(w)] \quad (5)$$

当 $T(w) = T(w)_{\min}$ 时， $F_\alpha = \alpha_{\max}$ ；当 $T(w) = T(w)_{\max}$ 时， $F_\alpha = \alpha_{\min}$ 。 $T(w)_{\min}$ 和 $T(w)_{\max}$ 分别是逐通道得到的掩蔽阈值的最小值和最大值。

本文电子耳蜗中的语音增强方法如下：

电子耳蜗中的语音信号处理采用 Bark 子波代替以有限的频率点为中心频率的一组带通滤波器，将每帧声音信号分成 18 个通道，在每个通道里分别估计噪声，根据人耳的听觉掩蔽效应，来自适应地设定和调整算法的相关参数，并根据式(3)达到降噪的目的，最后将处理后的语音数据进行 Bark 子波合成。

本文提出的增强系统如图 2 所示。

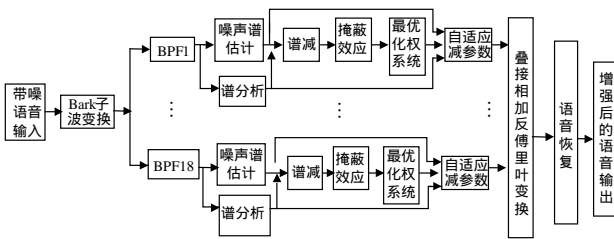


图 2 电子耳蜗中的增强系统

2.2 Bark 子波变换及其在电子耳蜗中的应用

人耳基底膜对外来声音信号有频率选择和调谐作用，在 20 Hz ~ 16 kHz 的听阈内组成 24 个临界频段：

$$b = 13 \arctan[0.76f] + 3.5 \arctan[f]^2 \quad (6)$$

其中， b 代表 Bark 频率； f 为线性频率，单位为 kHz。

构造临界子波所需子波母函数选择形式为 $w(b) = e^{-c_b b^2}$ 。假设被分析语音信号的线性频率带宽 $|f| \in [f_1, f_2]$ ，相应的 Bark 频率带为 $[b_1, b_2]$ ，则定义子波函数 Bark 域形式为

$$w_k(b) = w(b - b_1 - k\Delta b), k = 0, 1, \dots, k-1 \quad (7)$$

其中， Δb 为 $w_k(b)$ 转换到线性频率下的临界频带子波函数表

示形式 $w_k(f)$ ，可得到在频域定义的临界频带子波变换，即

$$s_k(t) = BW_k(t) = \int_{-\infty}^{\infty} S(f)W_k(f)e^{i2\pi ft} df \quad (8)$$

线性频率域下 Bark 子波族如图 3 所示。

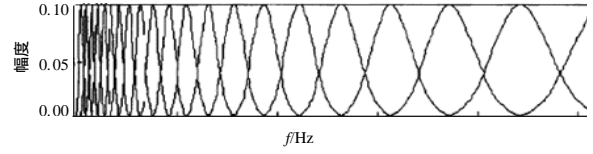


图 3 线性频率域下 Bark 子波族示意图

2.3 听觉掩蔽效应

一个较弱的声音(被掩蔽音)引起的听觉感受被另一个较强的声音(掩蔽音)影响的现象称为人耳的掩蔽效应。在安静环境中能被人耳听到的纯音的最小值称为绝对闻阈。掩蔽类型分为频域掩蔽和时域掩蔽。声音频率与掩蔽曲线不是线性关系，为从感知上统一度量声音频率，引入了临界频带的概念^[5]：当某个纯音被以它为中心频率、具有一定带宽的连续噪声所掩蔽时，如果在这一频带内噪声功率等于该纯音的功率，则此时该纯音处于刚好能被听到的临界状态，称这一带宽为临界带宽。

本文利用人耳的生理听觉特性，计算每个通道音频的子带掩蔽阈值，从而自适应地设定和调整式(6)中的参数 α 和参数 β ，有效掩蔽了残留噪声。

噪声掩蔽阈值^[6]的计算由以下几个部分组成：

(1) 计算临界带的功率谱。根据文献[7]临界带的划分，可以得到每个临界带(Bark 带)中的语音能量，即

$$B(i) = \sum_{k=l_i}^{h_i} E(k) \quad (9)$$

其中， l_i 为临界带 i 的下边界； h_i 为临界带 i 的上边界； $i=1, 2, \dots, i_{\max}$ ， i_{\max} 由采样频率决定。

(2) 与扩展函数 $SF(i)$ 进行卷积，计算扩展临界带能量 $C(i)$ ，公式如下：

$$C(i) = B(i) * SF(i) \quad (10)$$

其中， $SF(i) = 15.81 + 7.5(i + 0.474) - 17.5\sqrt{1 + (i + 0.474)^2}$ 。

(3) 减去一个相对阈值偏移量 $O(i)$ ，以计算掩蔽阈值，以 dB 为单位的掩蔽阈值偏移量定义如下：

$$O(i) = \alpha \times (14.5 + i) + (1 - \alpha) \times 5.5 \quad (11)$$

其中， $\alpha = \min\left\{\frac{SFM_{dB}}{SFM_{dB\max}}, 1\right\}$ 为每个 Bark 中噪声特性因子；

$SFM_{dB} = 10 \lg \frac{Gm}{Am}$ 为谱平坦测度； Gm 为功率谱的几何平均值； Am 为功率谱的算术平均值。当 $SFM_{dB\max} = -60$ dB 时，信号为完全类似于单音，若 $SFM = 0$ dB 则信号完全类似于噪声。

(4) 确定了相对掩蔽阈值的偏移量后，噪声掩蔽阈值可以由下式得到：

$$T(i) = 10^{\lg[C(i)] - [O(i)/10]} \quad (12)$$

3 计算机仿真及结果分析

实验中通过麦克风采集一女声的带噪语音“为国争光”，噪声材料为平稳白噪声。语音和噪声信号经 8 kHz 采样，8 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(-5, -3, 0, 3, 5)的带噪语音。带噪语音的长度为 512 点的语音帧，相邻两帧之间重叠 256 点，通过 Bark 子波变换把每一帧分成 18 个通道，对每一帧每通道的带噪语音进行语音增强

处理。

图4~图6给出了语音增强实验结果的例子。

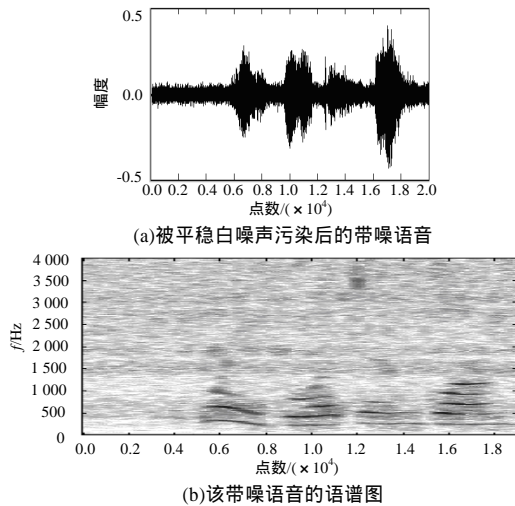


图4 被平稳白噪声污染后的带噪语音信号及其语谱图

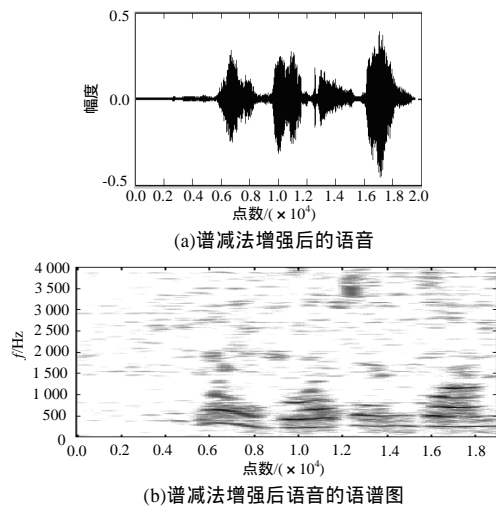


图5 谱减法增强后的语音信号及其语谱图

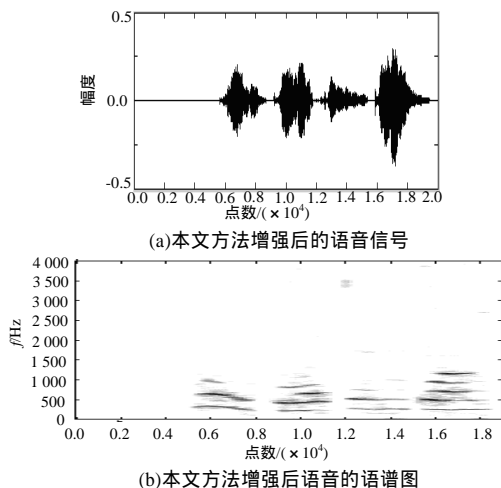


图6 本文方法增强后的语音信号及其语谱图

由图5和图6可看出,采用本文方法比采用谱减法所得的剩余噪声小。由语谱图可见,采用本文方法得到的增强信号能较好地保持原始信号的频谱特征。

对该带噪语音,用谱减法与本文所提算法进行比较,结果如表1所示。

输入信噪比	PSS方法输出信噪比	本文方法输出信噪比
-5	15.19	26.71
-3	17.18	27.43
0	20.15	29.37
3	23.12	31.35
5	25.09	34.49

信噪比计算公式如下:

$$SNR = 10 \lg \frac{\sum s(k)^2}{\sum (y(k) - s(k))^2} \quad (13)$$

其中, k 为采样点数; $s(k)$ 为增强后的语音; $y(k)$ 为带噪语音。

由表1中可以看出,本文方法的增强效果优于基本谱减法,噪声被更好地消除,信噪比得到了进一步提高。

主观听觉测试表明:用谱减法增强后虽然嘈杂噪声已明显减少,但夹杂有音乐噪声;用本文所提方法增强后其音乐噪声明显被抑制。在信噪比更低的情况下,本文方法的残留噪声对语音的干扰比谱减法要小得多,因此,更适用于电子耳蜗中的语音增强,使聋人在噪声环境中仍能获得具有一定的可懂度的语音信息。

4 结束语

环境的复杂多变导致传统电子耳蜗无法使聋人获得准确的语音信息。本文证明了在电子耳蜗中实现语音增强是可行的,并通过减小残留音乐噪声,提高了增强语音的感知质量。

许多环境下的噪声都是非平稳的,人类听觉系统能从这类噪声中提取有用信息,而本文只处理平稳随机噪声,对噪声特性变化剧烈的环境,此方法有一定局限性。

参考文献

- [1] 李春晓, 刘 樨, 聂开宝, 等. 频域和时域信息对连续交替取样策略汉语声调识别的贡献[J]. 生物医学工程杂志, 2006, 23(1): 41-44.
- [2] 聂开宝, 蓝 宁, 高上凯. 用小波变换实现电子耳蜗 CIS 语音信号的处理[J]. 清华大学学报, 1999, 39(9): 37-39.
- [3] Wilson B, Finley C C, Lawson D, et al. Design and Evaluation of Continuous Interleaved Sampling(CIS) Processing Strategy for Multichannel Cochlear Implants[J]. Journal of Rehab. and Research and Development, 1993, 30(1): 110-116.
- [4] 陶 智, 赵鹤鸣, 龚呈卉. 基于听觉掩蔽效应和Bark子波变换的语音增强[J]. 声学学报, 2005, 30(4): 367-372.
- [5] 徐 望, 王炳锡, 丁 琦. 一种基于信号子空间和听觉掩蔽效应的语音增强方法[J]. 声学学报, 2004, 20(2): 112-115.
- [6] 王晓娣. 多带谱相减结合感觉加权的语音增强方法研究[J]. 电力系统通信, 2005, 26(158): 51-52.
- [7] Nathalie Virag. Single Channel Speech Enhancement Based on Masking Properties of Human Auditory System[J]. IEEE Transactions on Speech and Audio Processing, 1999, 7(2): 126-137.