

数字语音信号

信号包络提取

算法

语音信号处理

23

88-90

数字语音信号包络提取算法研究

On the Envelops Extracting Algorithms for Speech Signals

袁晓* 刘光远* 邱玉辉** 虞厥邦*

(电子科技大学光电子技术系 成都610054)*

(西南师范大学计算机科学系 重庆400715)**

TN912.3

摘要 Based on the properties of envelop filters and the characteristics of speech digital signals, two fast envelop extracting algorithms for speech signals are proposed in this paper. The complexity of computation and the error performances of these novel algorithms proposed are discussed. The simulation results show that the algorithms are efficient for extracting the envelop of speech signals.

关键词 Envelop filters, Fast algorithm, Analytic wavelet transform, Speech signals

1 信号包络提取原理

在许多领域,比如通讯、自动测试与控制、机械故障检测与诊断,特别是在语音信号处理与识别等领域中,从原始信号中提取包络信息往往是十分重要的^[1-5]。对于包络信号的提取,经典方法是采用Hilbert变换(HT)将原始信号 $s(t)$ 转换成复解析信号 $\tilde{s}(t)$,取其模,即 $|\tilde{s}(t)|$ 作为 $s(t)$ 的幅度包络信号^{[1][2]}。这种包络提取法—HT方法对于提取窄带载波信号的包络是十分有效的,但对于宽带时变信号,比如语音信号的包络提取却存在许多固有的缺陷^[5]。

实信号 $s(t)$ 的复解析表示 $\tilde{s}(t)$ 形式为^{[2][3]}

$$\tilde{s}(t) = s(t) + j\hat{s}(t) = s(t) * h_H(t) \quad (1)$$

$$\hat{s}(t) = s(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau \quad (2)$$

$$h_H(t) = \delta(t) + j \frac{1}{\pi t} \quad (3)$$

式中 $j = \sqrt{-1}$,对于窄带载波信号 $s(t) = a(t)\cos(\omega_0 t + \theta_0)$,模 $|\tilde{s}(t)|$ 就等于原信号 $s(t)$ 的包

络信号 $E_H(t)$,即 $E_H(t) = |\tilde{s}(t)| = |a(t)|^{[1-2]}$ 。因此式(3)定义的 $h_H(t)$ 通常称为HT包络滤波器,其Fourier变换为

$$H_H(\omega) = 1 + j[-j\text{sgn}(\omega)] = \begin{cases} 0 & \omega < 0 \\ 2 & \omega \geq 0 \end{cases} \quad (4)$$

这表明, $\tilde{s}(t)$ 实质上就是用 $H_H(\omega)$ 滤出 $s(t)$ 的所有非负频率成分,而其相位保持不变,即

$$\angle H(\omega) = 0 \quad (5)$$

因而用HT方法提取语音包络信号本身就存在许多固有的缺陷^[5]。为此,文[5]中提出采用复解析小波变换(Complex Analytic Wavelet Transform)提取语音信号包络的新方法(简称为CAWT方法)并给出了语音信号包络滤波器 $h(t)$ 的构造条件:

$$\overline{h(t)} = h_r(t) + j\hat{h}_r(t) = \overline{h_r(t)} * h_H(t) \quad (6)$$

式中 $h_r(t)$ 为实偶函数并满足小波允许性条件,即

$$h_r(t) = h_r(-t), \int_{-\infty}^{\infty} h_r(t) dt = 0 \quad (7)$$

取母函数 $\psi(t)$ 为:

$$\begin{cases} \psi(t) = h(t) = \psi_r(t) + j\hat{\psi}_r(t) \\ \hat{\psi}_r(t) = \hat{h}_r(t), \psi_r(t) = \hat{\hat{h}}_r(t) \end{cases} \quad (8)$$

CSCW中Agent的混沌特性,提出了基于对策论、混沌时间序列和混沌神经网络的预测-决策方法,刻画了相应的Agent的行为^[6],以消除资源冲突。

从人脑到社会系统,都具有混沌行为,这表明混沌是普遍的现象。人工智能研究人的思维和交互,其对象本身就是具有混沌特性的复杂系统,因而混沌动力学作为人工智能的研究工具是合适的^[7],并已有了若干进展。由于混沌动力学自身在不断发展和

成熟的过程中,与混沌研究密切相关的一维符号动力学已较完善,而二维符号动力学则正处在突破前夜,因而对于混沌本身的把握并不完善。我们并不认为混沌能解释所有我们希望解决的问题,但它完全可以作为一种新的技术用于智能信息处理中。智能问题的本质与机理尚不清晰,混沌与智能的相结合的研究必定是充满挑战,同时又富有吸引力的。

(参考文献共51篇略)

则实信号 $s(t)$ 在时刻 t 的复解析小波变换为

$$W_a(a, t) = \langle s(\tau), \psi_a(\tau - t) \rangle \\ = a^{-\frac{1}{2}} \int_{-\infty}^{\infty} s(\tau) \psi\left(\frac{\tau - t}{a}\right) d\tau \quad (9)$$

式中 $a > 0$ 为尺度参数。定义在尺度 a 下, 原信号 $s(t)$ 的包络 $E_a(t)$ 为

$$E_a(t) = \sqrt{a} |W_a(a, t)| \quad (10)$$

对于语音(浊音)信号来说, 当 a 较小时, 将提取细节包络, 当 a 较大时, 将提取音节包络, 这体现了小波变换的自适应分析性能, 式(6)~(10)就构成了信号包络提取新方法——CAWT方法的算法基础。

从以上理论分析可见, CAWT方法在保留了HT方法的精髓——提取信号幅度包络的同时, 既能克服HT方法固有缺点, 又融进了小波分析的优势。

2 CAWT包络提取的数字算法研究

对于时间抽样间隔为 Δ 的语音数字信号 $s[n]$, 需要将式(10)以同样的时间抽样间隔 Δ 离散化, 即 $E_a[n] = E_a(n\Delta)$

$$= \sqrt{a} |(s[m], \psi_a((m-n)\Delta))| \\ = \Delta \left| \sum_{m=-\infty}^{\infty} s[m] \cdot \psi\left(\frac{m-n}{a}\Delta\right) \right|$$

取尺度 a 为正整数 d , 即 $a = d, d \in Z^+$, 并且假设母函数 $\psi(t)$ 的支撑区间在 $[-N\Delta, N\Delta]$ ($N \in Z^+$) 内, 即滤波器 $h[n]$ 的长度 $L = 2N + 1$, 则有

$$A0: E_d[n] = \Delta \left| \sum_{m=-dN}^{dN} s[m] \cdot \psi\left(\frac{m-n}{d}\Delta\right) \right| \\ = \Delta \left| \sum_{m=-dN}^{dN} s[m+n] \cdot \psi\left(\frac{m}{d}\Delta\right) \right| \quad (11)$$

上式就构成了CAWT方法的原始数字算法A0, 其计算量随尺度 d 成正比地增加。为了快速计算 $E_d[n]$, 必须简化计算公式(11), 根据CAWT方法及其语音(浊音)数字信号特点, 本文提出如下几种快速算法。

算法 A1 取 $m = kd + i, m, k \in Z, i = 0, 1, \dots, d-1$, 则式(11)可改写成:

$$A1: E_d[n] \approx E1_d[n] = \Delta \left| \sum_{k=-N}^{N-1} \sum_{i=0}^{d-1} s[kd+i+n] \cdot \psi\left(\left(k + \frac{i}{d}\right)\Delta\right) \right| \quad (12)$$

算法 A2 假设 $\psi(t)$ 是光滑的, 并且在区间 $[k\Delta, (k+1)\Delta]$ ($k \in Z$) 内有:

$$\psi\left(k + \frac{i}{d}\Delta\right) \cong \psi(k\Delta) \\ -N \leq k \leq N, i = 0, 1, \dots, d-1 \quad (13)$$

则(12)可简化为:

$$A2: E_d[n] \approx E2_d[n] = \Delta \left| \sum_{k=-N}^N \psi(k\Delta) \sum_{i=0}^{d-1} s[kd+i+n] \right| \quad (14)$$

(在算法A2中, 更为合理的假设应当取 $\psi\left(k + \frac{i}{d}\Delta\right) \approx \psi\left(\left(k + \frac{d-1}{2d}\right)\Delta\right)$, 但这增加了滤波器的复杂度。)

算法 A3 对于数字语音信号, 特别是浊音情形, 相邻抽样值之间的相关性很大^{[3][7]}, 因此对于较小的尺度参数 d , 可以简单地假设:

$$\sum_{i=0}^{d-1} s[kd+i+n] \cong ds\left[kd+n + \left\lceil \frac{d-1}{2} \right\rceil\right] \\ -N \leq k \leq N \quad (15)$$

从而可得到更为简单的算法A3:

$$E_d[n] \approx E3_d[n] \\ = \Delta d \left| \sum_{k=-N}^N s\left[kd + \left\lceil \frac{d-1}{2} \right\rceil + n\right] \cdot \psi(k\Delta) \right| \quad (16)$$

如果将计算某一固定时刻点 n 的 D 级包络值 $E_d[n]$ ($d = 1, 2, \dots, D$) 所需的复数乘法与加法次数的总和定义为算法的复数量(即计算量) C , 则算法A0, A1, A2和A3的复数量分别为:

$$A0: C0(D) = \sum_{d=1}^D 2(2dN+1) \\ = 2D[(D+1)N+1] \\ A1: C1(D) = \sum_{d=1}^D 2 \times 2dN = 2D(D+1)N \\ A2: C2(D) = \sum_{d=1}^D (2N+1)(d+1) \\ = D(D+3)(N+0.5) \\ A3: C3(D) = \sum_{d=1}^D 2(2N+1) = 2D(2N+1)$$

为了比较三种简化算法计算复杂程度的改善情况, 定义A1, A2和A3相对原始算法A0的相对计算复杂度(以A0的计算复杂度为100%)分别为

$$A0: c0(D) = 100\% \\ A1: c1(D) = C1(D)/C0(D) \\ = \frac{N(D+1)}{N(D+1)+1} \times 100(\%)$$

$$A2: c2(D) = C2(D)/C0(D) \\ \approx \frac{1}{2} \left[1 + \frac{1N+D+1}{2(ND+N+1)} \right] \cdot 100(\%)$$

$$A3: c3(D) = C3(D)/C0(D) \\ = \frac{2N+1}{(D+1)N+1} \times 100(\%)$$

相对计算复杂度曲线如图1所示。

3 快速算法的误差分析

定义 1 算法 A_i ($i = 1, 2, 3$) 相对原始算法A0在尺度 d 下, 时刻 n 时的绝对误差值为 $e_i(d, n) = |E_d(n) - E1_d(n)|$, 则有

$$e_i(d, n) = \Delta |\psi(N\Delta)| |s[dN+n]| \quad (17)$$

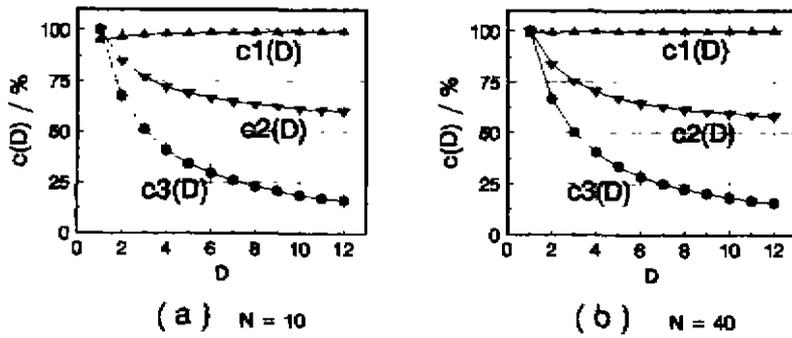


图1 计算复杂度曲线

$$e_1(d, n) = \left| \Delta \left| \sum_{m=-dN}^{dN} s[m+n] \cdot \psi\left(\frac{m}{d}\Delta\right) \right| - \Delta \left| \sum_{m=-dN}^{dN} s[m+n] \cdot \psi\left(\left\lceil \frac{m}{d} \right\rceil \Delta\right) \right| \right| \leq \Delta \left| \sum_{m=-dN}^{dN} s[m+n] \left[\psi\left(\frac{m}{d}\Delta\right) - \psi\left(\left\lceil \frac{m}{d} \right\rceil \Delta\right) \right] \right| \leq \Delta \sum_{m=-dN}^{dN} \left| \psi\left(\frac{m}{d}\Delta\right) - \psi\left(\left\lceil \frac{m}{d} \right\rceil \Delta\right) \right| \cdot |s[m+n]| \quad (18)$$

$$e_2(d, n) = |E_d(n) - E_{2d}(n) + E_{2d}(n) - E_{3d}(n)| \leq e_2(d, n) + e_{23}(d, n) \quad (19)$$

$$e_{23}(d, n) = |E_{2d}(n) - E_{3d}(n)| \leq \Delta \sum_{k=-N}^N |\psi(k\Delta)| \cdot |A_d(k, n)| \quad (20)$$

式中 $A_d(d, n) = \sum_{i=0}^{d-1} (s[kd+i+n] - s[kd + \lceil \frac{d-1}{2} \rceil + n])$ 。对于汉语(元音)浊音段信号,可以简单地假设相邻样本之间线性相关,相关系数为 $R(i) = 1 - |i|r$ (r 为依赖于信号 $s[n]$ 的常数),故当 d 较小时有

$$A_d(k, n) \approx -r \sum_{i=-\lceil \frac{d-1}{2} \rceil}^{\lceil \frac{d-1}{2} \rceil} |i|s[kd + \lceil \frac{d-1}{2} \rceil + n] \quad (21)$$

假设当 $-dN \leq m \leq dN$ 时,信号 $s[m+n]$ 是平稳的,其绝对均值为常数并取为 μ ,即 $E\{|s[m+n]|\} = \mu, n \in Z, -dN \leq m \leq dN$ (22) 则可作如下定义:

定义2 算法A1、A2和A3相对原始算法A0在尺度 d 下,时刻 n 时的相对极大绝对误差均值分别定义为(假设 $\mu = 1$)

$$M_1(d) = E\{\text{Max}[e_1(d, n)]\} = \Delta |\psi(N\Delta)| (\%) \quad (23)$$

$$M_2(d) = E\{\text{Max}[e_2(d, n)]\} = \Delta \sum_{m=-dN}^{dN} \left| \psi\left(\frac{m}{d}\Delta\right) - \psi\left(\left\lceil \frac{m}{d} \right\rceil \Delta\right) \right| (\%) \quad (24)$$

$$M_3(d) = E\{\text{Max}[e_3(d, n)]\} = M_2(d) + \Delta \sum_{k=-N}^N |\psi(k\Delta)| r \sum_{i=-\lceil \frac{d-1}{2} \rceil}^{\lceil \frac{d-1}{2} \rceil} |i| (\%) \quad (25)$$

可见 $M_1(d)$ 、 $M_2(d)$ 仅与母函数 $\psi(t)$ 有关,而 $M_3(d)$ 与母函数 $\psi(t)$ 和信号的性质都有关。因 $\psi(N\Delta) \approx 0$,故有 $M_1(d) \approx 0$ 。

结论 相对原始算法A0来说,算法A1并无实质的改变。

算法A2在假设式(13)成立的条件下,即仅对母函数 $\psi(t)$ 的光滑性有一定的要求,通过简化各级滤波器系数,减少复数乘法次数来降低计算复杂度。由A2产生的误差是很低的,当 $d \leq 8$ 时,各级实际的相对误差 $Me_2(d) < 3\%$,而其计算量当 $D = 8$ 时却减少了大约40%。算法A2与信号性质无关,因此算法A2对非语音信号包络提取也是有效的。

算法A3在算法A2的基础上,根据语音数字信号相邻样本之间相关性强的特点,在假设式(15)成立条件下,通过减少各级实数加法次数来降低计算复杂度。由算法A3产生的误差随级数 d (即尺度)增大急剧增长。当 $d \leq 4$ 时,理论上相对极大误差均值 $M_3(d) < 10\%$,实际的音节相对误差在实验中几乎都有 $Me_3(d) \leq 5\%$,而其计算量当 $D = 4$ 时,却减少了大约60%,因此,当 D 较小时,A3不失为一种有效的语音信号包络提取快速数字算法。

参考文献

[1] [加] 海金, 通信系统, 北京: 中国铁道出版社, 1983
 [2] 陈元亨, 信息与信号理论基础, 北京: 高等教育出版社, 1989
 [3] 张绪省, 失聪盛, 成晓雄等, 信号包络提取方法, 电子科学学刊, 1997, 19(1): 120~123
 [4] 杨行峻, 迟惠生, 语音信号数字处理, 北京: 电子工业出版社, 1995
 [5] 袁晓, 张红雨, 虞厥邦, 语音信号包络滤波器设计, 电子科技大学学报, 待发表
 [6] 吴宗济, 林茂灿等, 实验语音学教程, 北京: 高等教育出版社, 1989
 [7] L. R. Rabiner, R. W. Schaffer. "Digital Processing of Speech Signals" Prentice-Hall, Inc. 1978