基于 C_0 复杂度和能量的语音端点检测算法

马伟荣[1],冯宏伟[1],李 宁[2]
MA Wei-rong[1],FENG Hong-wei[1],LI Ning[2]

1.西北大学 信息科学与技术学院,西安 710127
2.大连理工大学 软件学院,辽宁 大连 116023
1.College of Information Science and Technology,Northwest University,Xi'an 710127,China
2.School of Software,Dalian University of Technology,Dalian,Liaoning 116023,China
E-mail:maweirong2002@163.com

MA Wei-rong,FENG Hong-wei,LI Ning.Speech detection algorithm based on C_0 complexity and short-time energy. Computer Engineering and Applications,2009,45(27):143-145.

Abstract: Complexity measure is an important nonlinear property of the signal sequence.Speech detection methods based on complexity measures have nonlinear properties.This paper proposes a new speech detection method,which improves the traditional C_0 complexity and combines with enhanced short-time energy.Simulation results indicate that the method has a strong robust to noise and is able to reliably detect the onset and offset of speech even for low SNR such as 0 dB situation.
Key words: endpoint detection;short-time energy;spectral subtraction;complexity

摘 要:复杂性测度是反映信号序列的一个重要的非线性特征,复杂性测度的语音端点检测技术具有非线性技术的本质特征。对 C_0 复杂度作出改进,并与增强后的短时能量相结合,提出了一种更有效的端点检测算法——C_0 复杂度能量的语音端点检测方法。实验证明,该算法对噪声有很强的鲁棒性,在低信噪比(0 dB)下仍能准确地检测出语音段。
关键词:端点检测;短时能量;谱减法;复杂性测度
DOI:10.3778/j.issn.1002-8331.2009.27.043 **文章编号:**1002-8331(2009)27-0143-03 **文献标识码:**A **中图分类号:**TP391.42

1 引言

语音端点检测就是要把语音和背景区分开来,又称为语音激活检测(VAD),是语音分析、语音合成和语音识别中的一个必要的环节。不准确的检测将会削掉部分音带信息或引入非语言事件,从而增加分析和识别误差[1],正确地检测语音端点检测不仅提高了系统处理效率,同时也能够提高系统的识别率。

从本质上说,语音信号处理方法可分成两大类,其一是基于确定性线性系统理论,其二是基于随机过程理论。目前大多数语音端点方法都属前者,这类方法都有一个基本的假设,即当分段足够小时,非线性系统可以用线性系统来近似,从而产生了诸如线性预测、同态解卷、正交变换等分段线性分析方法[2]。线性方法强调的是稳定、有序和一致性。语音信号是非线性和非平稳的,从物理背景和实验两方面出发,已有许多研究[3]。随着研究的深入,人们也发现传统的分段线性方法存在许多不足,表现为语音识别、说话人识别、语音合成及语音编码系统的性能难以进一步提高,因而人们逐渐将注意力转向非线性信号分析方法的研究。目前对非线性时间序列分析和处理已有很多方法,如李雅普诺夫方法和复杂性分析等。复杂性测度是反映语音信号的一个重要的非线性特征[4],随后又出现了 KC 复杂度、C_1/C_2 复杂度、分区复杂度、涨落复杂度、C_0 复杂度等。KC 的结果是对随机性的一种表述,C_1、C_2 则认为完全随机的运动并不复杂,这些算法都需要对序列做粗粒化操作,而粗粒化过程本身有可能丢失许多有意义的细节,针对这一问题,提出了 C_0 复杂度。C_0 复杂性其实也是随机性的一种表述,不过从算法上避免了粗粒化过程。该文采用 C_0 复杂性测度方法并对其进行改进,提出了 C_0 复杂度能量为特征的新的端点检测方法。

为了说明 C_0 复杂度能量在端点检测中的有效性,还选择能量-过零率、谱熵以及 C_0 复杂度法进行实验数据的对比。

2 端点检测的算法

2.1 谱减法

谱减法是语音增强中最常用的一种算法,能有效处理宽带噪声,其主要思想是在假定加行噪声与短时平稳的语音信号相互独立的条件下,从带噪语音的功率谱中减去噪声功率谱,从而得到较为纯净的语音频谱。因为噪声的功率谱在很短的时间内可认为没有变化,通常用噪声功率谱的平均能量来代替其功率谱。算法原理为[5]:

基金项目:国家高技术研究发展计划(863)(the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z328)。
作者简介:马伟荣(1982-),女,硕士研究生,主要研究领域为多媒体技术;冯宏伟(1964-),男,副教授,主要研究领域为多媒体技术;李宁(1982-),男,硕士研究生,主要研究领域为数据库,数据挖掘。
收稿日期:2008-05-26 **修回日期:**2008-08-06

$$|S_m(f)|^2=\begin{cases}|s_m(f)|^2-\gamma|D_m(f)|^2, |s_m(f)|^2\geqslant|D_m(f)|^2\\ 0.015\times|s_m(f)|^2, |s_m(f)|^2\leqslant|D_m(f)|^2\end{cases} \quad (1)$$

式中:γ 因子的取值为 1 或 2,$|s_m(f)|$为带噪语音信号的能量谱,$|D_m(f)|$为估计的噪声能量谱,$|s_m(f)|$为增强后能量谱。噪声能量谱$|D_m(f)|$估计过程首先假定前几帧(一般取前 10~20 帧)信号为背景噪声,计算这些帧的能量谱,利用这几帧的能量谱的平均值即可估计背景噪声的能量谱初值。

2.2　C_0 复杂性测度及其改进

2.2.1　C_0 复杂性测度

一般认为复杂运动可以是由规则运动和随机运动混合而成的。随机运动所占的份额,就是 C_0 复杂性描述的基础[6]。假设有一复杂运动的时间序列 $x(t)$,它包含了规则运动部分的时间序列及随机运动时间序列。假设规则运动部分时间序列为 $x_1(t)$,它与 $x(t)$的关系为函数 $f(t)$,于是有

$$x_1(t)=f[x(t)] \quad (2)$$

从 $x(t)$中去掉 $x_1(t)$,剩余部分就是随机运动部分。简单地,设有变换 $g(t)$,使得

$$A_0=g[x(t)] \quad (3)$$

$$A_1=g[x(t)-x_1(t)] \quad (4)$$

A_0 代表了整个复杂运动时间序列的某种量度,而 A_1 则代表了随机运动部分时间序列的某种量度。由此,可定义 C_0 复杂性为:

$$C_0=\lim_{t\to\infty}\frac{A_1}{A_0} \quad (5)$$

显然,当 $x_1(t)$在 $x(t)$中所占份额很大时,C_0 趋向于 0。说明系统的动力学行为几乎是规则的不含随机成分。反之,当 $x_1(t)$所占份额很小而随机运动部分时间序列所占的份额很大时,C_0 趋向 1 时,说明系统的动力学几乎是完全随机的。

2.2.2　C_0 复杂性测度改进

非线性特征 C_0 复杂度测度具有较强的抗噪能力,但一定程度上随着噪音的不断加强,复杂性测度的形状保持大致不变,但复杂性测度值变化趋于平稳状态,如图 1 所示,因此提出了 C_0 复杂度能量的新方法,通过引入增强后的语音的能量来增强复杂运动中的规则运动部分 C_0 值,使复杂性测度值变化增强,从而快速有效地进行端点检测。

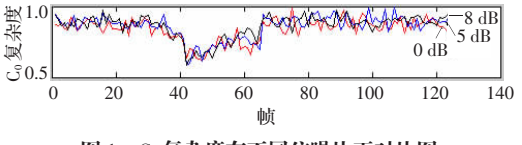

图 1　C_0 复杂度在不同信噪比下对比图

2.2.2.1　C_0 复杂度公式改进

随机运动所占的份额,就是 C_0 复杂性描述的基础[6],这是传统 C_0 复杂性的描述,在文中改进的 C_0 复杂度为规则运动所占的份额,即

$$A_1=g[x_1(t)] \quad (6)$$

代入公式(5),可得改进后的 C_0 复杂度。显然,当 $x_1(t)$在 $x(t)$中所占份额很大时,C_0 趋向于 1。说明系统的动力学行为几乎是规则的不含随机成分的。反之,当 $x_1(t)$所占份额很小而随机运动部分时间序列所占的份额很大时,C_0 趋向 0 时,说明系统的动力学几乎是完全随机的。主要步骤包括如下,其中 $x(t)$为分帧、加窗等预处理后的某频率分量。

(1)对 $x(t)$进行离散的傅立叶变换 $F(\cdot)$,有

$$x(k)=F[x(t)] \quad (7)$$

(2)可求出幅度谱的平均值$\bar{x}$

$$\bar{x}=\frac{1}{N}\sum_{k=1}^{N}x(k),\text{其中},1\leqslant k<N \quad (8)$$

k 为频域变量,N 为 $x(k)$的长度,即 k 的最大值。实际操作中$\bar{x}$还可以乘一个系数 α(α 为大于或等于 1 的常数),这样可适当调整规则部分的标准。大于平均值的频率成分被认为是规则部分的贡献,小于或等于平均值的成分则是随机部分的贡献,只取规则部分的贡献。

$$x'(k)=\begin{cases}x(k), x(k)>\bar{x}\\ 0, x(k)\leqslant\bar{x}\end{cases} \quad (9)$$

(3)对规则部分贡献的频谱 $x'(k)$作傅立叶反变换 $F^{-1}(\cdot)$,即得 $x_1(t)$。所以有

$$x_1(t)=F^{-1}[x'(k)] \quad (10)$$

至此,求得了 $x_1(t)$,即规则部分时间序列。

(4)利用公式

$$A_0=\int_0^{\infty}|x(t)|\mathrm{d}t \quad (11)$$

$$A_1=\int_0^{\infty}|x_1(t)|\mathrm{d}t \quad (12)$$

并代入式(5)求 C_0 复杂度。

2.2.2.2　短时能量

采用同求 C_0 时一样的窗函数、帧长、帧移对语音信号进行加窗分帧,计算每一帧的短时能量:

$$E_i=\sum_{i=1}^{n}s_k^2 \quad (13)$$

式中 s_k 为去噪后的采样值。

2.2.2.3　C_0 复杂度能量

$$C_0E=C_0*E \quad (14)$$

2.3　端点检测的步骤

(1)对语音信号进行加窗分帧,帧间 50%重叠。

(2)利用前 20 帧估计噪声谱,应用谱减法对每一帧进行频域谱减增强,得到增强后的能量谱。

(3)进行 FFT 变换,求得每帧改进后的 C_0 复杂度值。

(4)根据公式(14)求得每帧的 C_0 复杂度能量,然后采用自适应双门限判决准则,检测出语音段。

3　实验与结果分析

为了检测算法的有效性,进行大量实验。实验中,纯净语音信号采样频率是 8 000 Hz,量化精度为 16 bit;语音信号加 Hamming 窗,窗长 256、帧间重叠 128,FFT 变换长度为 256,手工加入白噪声。图 2 给出了一段语音在信噪比为 5 dB 时谱熵、C_0 复杂度以及 C_0 复杂度能量的特征示意图。

实验采用自适应门限法,得到基于 C_0 复杂度能量(所述方法)、C_0 复杂度、谱熵以及短时能量-过零率的统计信息。如表 1 所示,表中数据表示端点检测的正确率(百分数),其中正确率是通过检测到的位置与手工划分的端点位置对比来确定的。

图 3 给出了一个单语音文件不同端点检测算法的检测结果图示(原始语音中为手工标定的端点位置,其他为某种端点检测算法检测出的端点位置,下同)。

为了突出提出算法的优越性,给出其在低信噪比 0 dB、

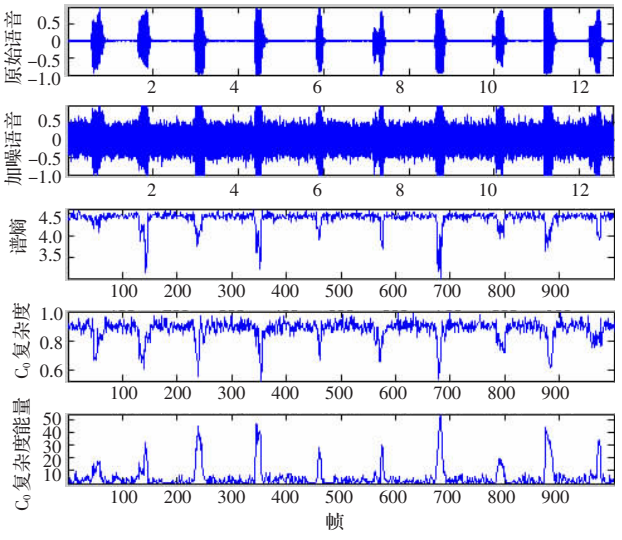

图 2　5 dB 带噪连续语音的三种方法特征曲线示意图

表 1　各种特征在不同信噪比下正确率对比表（%）

	0 dB	2 dB	8 dB	12 dB	15 dB
C_0 复杂度能量	89.4	92.8	96.4	97.5	98.9
C_0	78.2	80.3	89.7	94.8	95.5
谱熵	69.4	72.6	84.0	93.6	95.1
短时能量-过零率	32.1	45.6	63.7	72.6	80.0

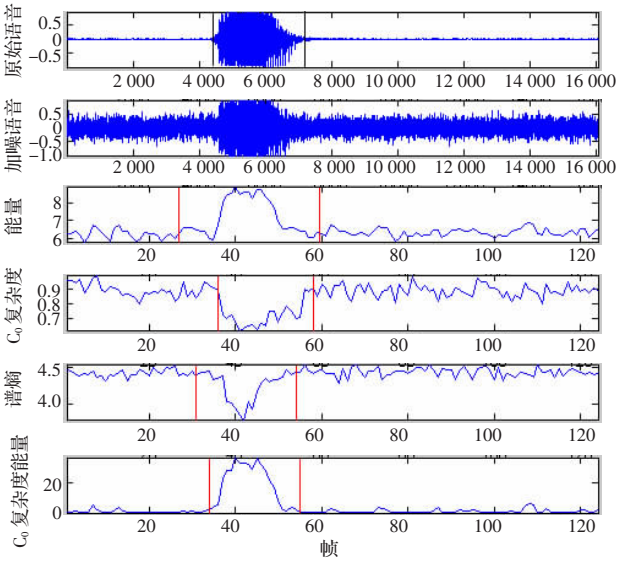

图 3　*SNR*=6 dB 时不同算法检测结果比较

2 dB、8 dB、12 dB、15 dB 下检测效果图示如图 4。

从表 1 可以看出,当 SNR 在 10 dB 以下时,提出的方法仍能准确地检测出端点。实验中还发现,单个语音端点检测的准确率要比连续语音的端点检测准确率高;在 0 dB 的 SNR 下,

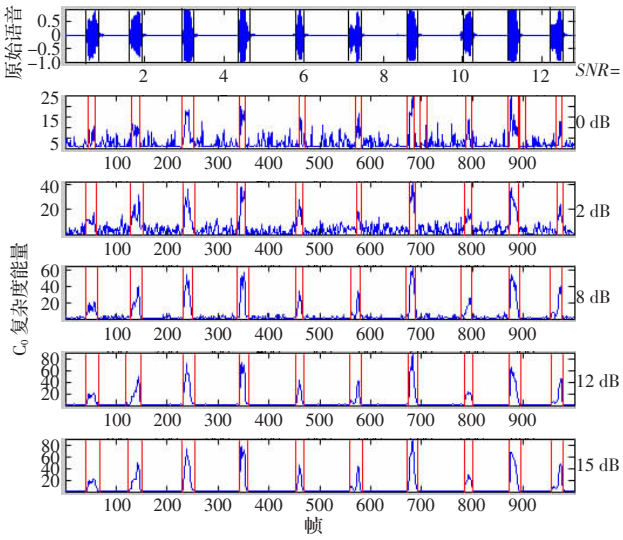

图 4　不同信噪比下 C_0 复杂度能量端点检测结果比较

单个音检测出的端点虽然将原始语音删减了部分,但通过视听辨析还保留有原音的主干部分。经实验分析,C_0 复杂度能量对噪音的鲁棒性要明显优于 C_0 复杂度、谱熵和能量。

4　结论

应用非线性 C_0 复杂度进行端点检测的方法,与其他常规方法相比,具备如下三个特征:端点检测的准确率高;检测算法具有对各种噪声的鲁棒性,抗干扰能力强;算法简单,易于实现,端点检测实时性强。

C_0 复杂度能量在充分发挥传统非线性 C_0 复杂度端点检测方法三个特征的同时,也充分利用了谱减法增强语音的优势,将两者有效结合,对复杂运动中规则运动部分很显著地增强,在特征参数上突显语音部分,从而可以更快速地进行端点检测。实验证明在低信噪比环境中,该算法对语音段、噪音和静音有很好的区分,适合于鲁棒语音识别。

参考文献:

[1] 吕秀良,范影乐.基于排列组合熵的语音端点检测技术研究[J].计算机工程与应用,2008,44(1):240-242.

[2] 范影乐,武传艳.基于复杂度的语音端点检测技术研究[J].传感技术学报,2006,19(3):750-753.

[3] Thompason C,Mulpor A,Mehta V.Transition to chaos in acoustically driven flow[J].J Acoust Soc Am,1991,90(4):2097-2103.

[4] Crutchfield J P,Young K.Inferring statistical complexity[J].Physical Review Letters,1989,63(10):245-250.

[5] 赵力.语音识别处理[M].北京:机械工业出版社,2003.

[6] 徐京华,吴祥宝.以复杂度测度刻划人脑皮层上的信息传输[J].中国科学:B 辑,1994,24(1):57-62.