

基于分形维数与 TEO 的语音端点检测算法

周 萍, 张上鑫, 郑凯鹏

(桂林电子科技大学 电子工程与自动化学院, 广西 桂林 541000)

摘 要: 为了改善短时 TEO 能量的抗噪性和分形检测自适应性, 提出了一种能适应多种噪声环境的端点检测算法. 该算法利用能表征能量分布的频域上的分形维数和 TEO 的消除零均值噪声、表征语音共振峰特性的特点, 在短时 TEO 能量法基础上增加分形维数计算来进行二次判决噪声段和语音段. 与传统端点检测算法相比, 该算法容易实现且准确度高, 实验表明, 在不同的背景噪声下, 该算法相比单一端点检测的算法鲁棒性更好, 弥补了单一算法抗噪性不好的缺点, 能适应低信噪比下的多种噪声环境.

关键词: 端点检测; 能量算子; 分形维数; 噪声

中图分类号: TN912.34

文献标识码: A

文章编号: 1000-7180(2017)06-0021-05

Speech Voice Activity Detection Algorithm Based on Fractal Dimension and TEO

ZHOU Ping, ZHANG Shang-xin, ZHENG Kai-peng

(Department of Electric Electric Engineering and Automation,
Guilin University of Electronic Technology, Guilin 541000, China)

Abstract: In order to improve the ability of anti-noise and fractal detection of short time TEO energy, a new algorithm about voice activity detection is proposed which can adapt to a variety of noise environments. The algorithm use the feature of energy distribution characterization of the fractal dimension in the frequency domain and TEO characteristics of zero mean noise elimination and characterization of formant characteristics, add the calculation of fractal dimension to the twice judge noise and voice part based on short-time energy method TEO. Compared with the traditional voice activity detection algorithm, this algorithm is easy to realize and with a high accuracy. The experiments show that in different background noises, the algorithm has a better robust performance compare to the single voice activity detection algorithm, and improve the performance of single algorithm for anti-noise. it can adapt to a variety of noise under the low SNR environment.

Key words: voice activity detection; energy operator; fractal dimension; noise

1 引言

语音的端点检测用于准确的判定语音的起始点和终止点对于语音识别有着十分关键的作用. 在实际的应用中, 语音识别很大一部分的错误都是由端点检测不准确引起的, 噪声对语音信号的影响以及说话人的语气和发声都会对端点检测的鲁棒性提出高要求. 目前的语音端点检测方法有短时能量与过

零率结合的双门限法、短时 TEO 能量法、谱熵法、梅尔倒谱距离法、基于分形语音端点检测等. 其中短时 TEO 能量法原理简单, 运算量小, 容易在硬件上实现, 能很好的表征信号的共振峰信息, 消除零均值噪声^[1]. 分形可以用来描述混沌信号, 语音便是一种混沌状态的信号^[2], 所以能够适用低信噪比的环境, 而且阈值稳定^[3]. 但是对于某些特定噪声, 分形算法也会检测失效.

为了提高在噪声语音中端点检测的性能,针对现在算法缺陷取长补短,本文提出基于分形维数与短时 TEO 能量相结合的判决机制.实验表明,两种算法的结合使端点检测的准确度有了较大的提高且计算量不大.

2 算法原理

2.1 时域上分形维数的计算

分形维数有几种定义,比如 Hausdorff 的和广义的以及盒维数的^[4],其中盒维数相对其它计算量较少.以下介绍的也是盒分形维数的计算.分形维数就是定量地表示自相似的随机形状和现象的最基本的量,其实就是用分形维数来对信号进行度量,反应信号的复杂度和不规则度.越是有规律性的信号,其分形维数越小^[3].而语音信号和噪声的维数都是在 1 和 2 之间.因为对于几何维数,1 维的是直线,2 维的是平面,所以信号处于直线与平面之间.信号就像平面上的曲线,可以用分形维数来测量.在语音段中韵母的声音信号波形有较大的周期性和规律性,而白噪声相对来说就没有这么规律,即对于分形维数的结果,白噪声比语音信号的要高.

对于一个连续信号 $x(t)$,采样量化后的离散信号为 $x(t_i)$ 对应的数字信号 $x(i)$.设信号采样点之间间隔为 Δ ,则:

$$t_i = i \cdot \Delta, \quad i = 1, 2, \dots, K \quad (1)$$

设正方形网格大小为 Δ ,且 $\Delta > 0$,存在一个有限实数 d ,对于 $x(i)$ 上的时间点 t_i 到 t_{i+1} 之间的曲线段用 Δ 去测量,共需要 $|x(i+1) - x(i)|/\Delta$ 个网格来覆盖.如图 1 所示,时间连续的信号 $x(t)$ 的波形可以用 N 个大小为 Δ 的方形网格沿着波形曲线图表示.

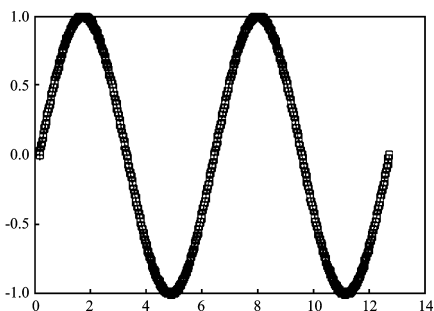


图 1 信号被方形网格覆盖

当 Δ 足够小时,大小为 Δ 的方形网格对时间信号 $x(t)$ 中的一帧曲线段进行覆盖测量,共需要 $N(\Delta)$ 个网格,则测量值 $N(\Delta)\Delta^d$ 不依赖 Δ 的大小,这样的 d 就是这一帧的曲线段的分形维数,计算由

以下公式给出:

$$\lim_{\Delta \rightarrow 0} N(\Delta)\Delta^d = C \quad (2)$$

式中, C 为常数,然后两边取对数为:

$$\lim_{\Delta \rightarrow 0} \lg N(\Delta) = \lim_{\Delta \rightarrow 0} (-d \cdot \lg \Delta + \lg C) \quad (3)$$

式(3)中 d 为时间信号 $x(t)$ 图形的分形维数, $\lg C$ 和 d 都是未知数,故需要联立方程组求解,由于 Δ 是可变的,因此可以通过改变 Δ 的值来建立方程以便求解 $\lg C$ 和 d :

$$\lim_{\Delta \rightarrow 0} [\lg N(k\Delta)] = \lim_{\Delta \rightarrow 0} [-d \cdot \lg(k\Delta) + \lg C] \quad (4)$$

K 个不同的网格得到 K 个方程,任意两个大小不同的网格 $j\Delta$ 和 $l\Delta$ 各自生成的方程,组成方程组相减可得:

$$d = \frac{\lg(N(j\Delta)/N(l\Delta))}{\lg(l/j)} \quad (5)$$

K 个方程的任意两个方程联立方程组的个数有 $K(K-1)/2$ 个,然后对 $K(K-1)/2$ 个方程组求出的分形维数 d 求平均值:

$$d = \frac{1}{K(K-1)/2} \sum_{j,l} \frac{\lg(N(j\Delta)/N(l\Delta))}{\lg(l/j)}, \quad 1 \leq j < l \leq K \quad (6)$$

通过式(6)的计算可以得出 $x(i)$ 波形上的一帧曲线段盒维数的平均值.

2.2 频域上的分形维数

通过频域分析可以对语音信号的频率特性有直观的了解,短时频域上的分形维数计算与时域上的类似,用 Δ 大小的网格覆盖波形 $x(n)$ 中频点 n 和频点 $n+1$ 之间的曲线段,需要 $|x(n+1) - x(n)|/\Delta$ 个网格.通过对相邻频点计算网格数来获得每帧需要的网格数同样记为 $N(\Delta)$,一帧的频域分形维数公式和时域上的基本一样,不同的是要经过傅里叶变换到频域才进行分形维数计算.频域上算出的分形维数阈值更加稳定,算法鲁棒性更高.设置门限值,计算每帧的分形维数,凡是大于门限阈值的标记噪声帧,小于门限阈值的则标记语音帧.

2.3 TEO 原理

TEO(Teager Energy Operator, Teager 能量算子)能够实时跟踪信号的能量变化,对调幅信号包络以及调频信号瞬间的频率变化很是敏感的非线性算子.并且对不同类型的信号都能显示不同的 TEO 运算结果.能很好表征信号共振峰信息,消除零均值噪声起到信号增强的作用.

文献[5]给出了 TEO 的运算公式,在连续时间信号 $x(t)$ 中,TEO 定义为:

$$\psi[x(t)] = (x'(t))^2 - x(t)x''(t) \quad (7)$$

式(7)中, $x'(t) = dx(t)/dt$. 从公式中可以看出 TEO 可以用来表示信号瞬时能量值,但是式(7)计算量太大. 对于离散时间信号,TEO 定义为:

$$\psi[x(\tau)] = x(\tau)^2 - x(\tau-1)x(\tau+1) \quad (8)$$

式(8)中, $x(\tau)$ 是离散信号在 τ 时的采样值. 从公式中可以看出 TEO 输出的信号,只跟当前信号采样点以及前后各一个采样点有关.

如图 2 所示的步骤得出 TEO 检测的结果,首先根据公式对输入信号计算 TEO 能量,然后对所得 TEO 结果加汉明窗和分出每一帧的能量值,接着平滑滤波处理 TEO,根据既定的判决机制来端点判断.

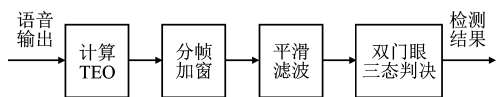


图 2 TEO 端点检测原理框图

实际环境中难免都会存在噪声,语音中也掺杂着各类噪声的干扰,为了达到减少噪声尖刺或抖动的干扰带来的误判,提高系统性能的目的,在检测前进行平滑滤波的处理:

$$\begin{aligned} E[\psi(\tau)] = \\ \frac{E[\psi(\tau-1)] + E[\psi(\tau)] + E[\psi(\tau+1)]}{3} \end{aligned} \quad (9)$$

通过平滑滤波的环节来平滑 TEO 能量曲线,从而一定程度上减少因为噪声导致的错误判断,提高检测系统的准确性.

3 算法实现

分形维数体现了信号波形的规律性和细节丰富性^[3]. 越是规律和细节不丰富的信号其分形维数越小,比如语音信号. 白噪声的频域能量分布是均匀而没有规律性,故其分形维数明显高于语音,所以用分形维数计算可以很好区分. 但是有些噪声,比如坦噪声细节不丰富且规律性较强,还有某些具有噪声行为特征的辅音,如 d, m, n , 在低信噪比的条件下很难区别它们,本文利用频域上的分形维数能更好的在频域上区分噪声和语音的分布,利用 TEO 来实时跟踪能量的变化并且更好细致区分一些具有噪声特征的辅音. 由于基于能量特征的端点检测系统,抗噪性相对比较差,这种缺点又是难以避免的^[6]. 所以本文提出的基于频域的分形维数和短时 TEO 能量结合的互补性质的联合算法来进行端点检测,能有效解决单独使用分形维数或者能量算法的不足,互

补性地提高端点检测准确率. 系统框图如图 3 所示.

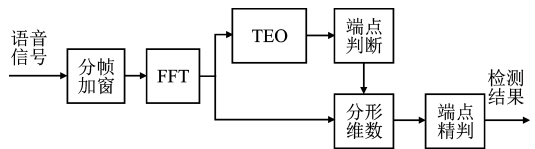


图 3 本文算法系统原理框

3.1 TEO 粗判

我们把语音信号检测的结果分为静音段和过渡段以及语音段三种,通过判断机制把有效语音信号区分开来,为后续语音的识别打好基础. 设置低门限阈值 T_1 和高门限阈值 T_2 , 低门限相对较小,容易被超越. 高门限的设置是为了让语音信号强度达到一定程度才被超越. 当信号刚开始没有超过低门限时,标记为静音段,超过低门限时,标记信号进入过渡段,在进入过渡段连续五帧内一直低于高门限或者低于低门限则标记为静音段. 如果信号超过高门限则标记信号进入语音段,并且如果进入语音段的信号连续三帧都超过高门限,则往前回溯最近进入过渡段的时刻,标记为起始点. 而语音段的终止点的判断则由信号从语音段跌落到静音段来决定,即信号强度低于低门限连续三帧则信号进入静音段,标记低于低门限的时刻为语音终止点.

根据上述的判断条件,我们只要计算出高低门限的阈值就可以实现 TEO 的端点判断. 低门限的值一般要稍微高于噪声,但是噪声不一定是平稳的,而且环境不同噪声的值也不同,所以取信号前几帧来预估噪声平均值,会造成端点检测的准确性下降,这种情况只适用于处于平稳噪声甚至静音的环境中,所以 TEO 端点检测的门限值要适应环境的变化,取动态自适应门限. 计算整个信号段的平均能量值 $\bar{E}$ 和最低能量值 $E_{\min}$, 通过两个值来实时调整高低门限的阈值,起到自适应的作用. 门限的计算公式为:

$$\begin{cases} T_1 = \max[0.4E_{S\min} + 0.6\bar{E}_{\min}, 0.3\bar{E} + 0.7E_{\min}] \\ T_2 = \max[T_1 + E_{S\max} - E_{S\min}, T_1 + (\bar{E} + E_{\min})/10] \end{cases} \quad (10)$$

式中, $E_{S\min}$ 和 $E_{S\max}$ 为信号前后各五帧共十帧的静音段的最小和最大的 TEO 能量值,然后根据低门限 T_1 和高门限 T_2 来判断语音起始点和终止点.

3.2 短时频域分形维数精判

短时频域的分形维数也是需要设置阈值,设置一个噪声阈值 G , 低于噪声阈值的为语音段,反之则为

噪声段. 通过 TEO 的粗判之后, 可以得出第一次的语音段和非语音段了. 接下来再对粗判之后的语音段进行分形维数端点检测, 先把语音段前后各十帧的被 TEO 判断为噪声段的信号进行分形维数计算, 作为短时频域的噪声阈值 G . 语音段中低于阈值 G 的被二次判断为语音段, 否则丢弃为噪声段.

实际环境中总是会存在各种不同的噪声, 本文利用 TEO 能量来粗判表征语音共振峰特性和消除零均值噪声影响, 然后利用频域的分形维数的抗噪性来二次判决, 两种方法的结合有效地提高了系统的性能.

4 实验结果与分析

4.1 测试

本文对提出的方法在电脑上用 matlab 进行实验, 实验用电脑的麦克风采集声音信号样本, 以 wav 文件格式保存, 8 k 赫兹的采样率和 16 位的位宽, 取帧长 256, 帧移 128, 噪声则使用的是 NoiseX-92 噪声库中常见的噪声, 包括白噪声 (White), 背景说话噪声 (Babble), 坦克内部噪声 (M109), 工厂环境噪声 (Factory). 录一段语音信号, 将语音信号混合各种噪声, 然后用传统的短时能量和过零率结合的双门限法和短时 TEO 能量以及本文所提出的算法对语音信号进行检测对比, 在不同 SNR 下比较本文的端点检测算法与传统的端点检测算法的结果.

4.2 实验结果

语音内容为“桂林”, 图 4 至图 7 为原始语音加了不同噪声在信噪比 -5 dB 下用本文算法检测的结果.

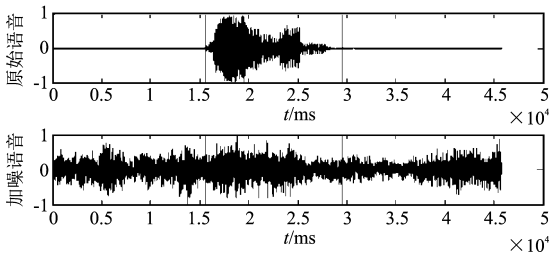


图 4 加了 Babble 噪声在 -5 dB 下的检测结果

4.3 结果分析

从实验过程和结果可以看出, 本文算法原理简单容易实现, 运算量小, 在硬件上也能轻松实现, 抗干扰能力强, 判决门限自适应, 即使低 SNR 的不同背景噪声, 本文提出的算法也有较高的准确率. 该算法相比双门限法, 从运算处理时间上这两个算法其实差不多, 但是在低信噪比下准确率最高可以高于

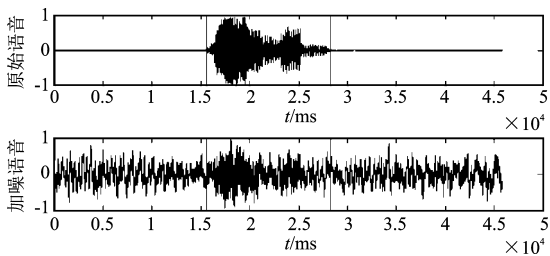


图 5 加了 Factory 噪声在 -5 dB 下的检测结果

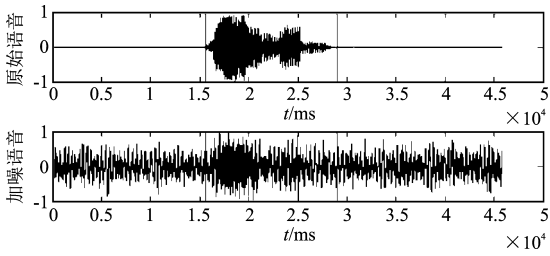


图 6 加了 M109 噪声在 -5 dB 下的检测结果

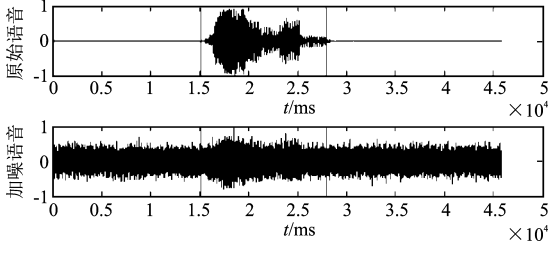


图 7 加了 White 噪声在 -5 dB 下的检测结果

传统双门限 27%. 相对短时 TEO 能量法有明显提高, 不同的背景噪声下, 本文提出的算法检测率都是最高的, 其中在平稳的 Factory 环境下检测率最高, 在 Babble 环境下检测率最低, 这是由于 Babble 就是不同的人说话的一个背景噪声, 所以会与语音信号有所相似从而产生误判. 本文算法与传统双门限法以及短时 TEO 能量法的检测率比较结果见表 1.

表 1 三种算法在不同背景噪声同一信噪比 (-5 dB) 下检测的正确率比较

算法	环境噪声			
	Babble	Factory	White	M109
双门限	0.11	0.51	0.16	0.34
TEO	0.56	0.77	0.72	0.76
本文算法	0.85	0.91	0.88	0.90

由表 1 可以得出, 各种背景噪声下, 本文提出的算法检测率比双门限和单一 TEO 能量法要高, 尤其是平稳噪声下检测率比不平稳的噪声如 Babble 的环境下要高. 这是因为在 TEO 的基础上加入了

频域分形维数的计算,使得噪声和语音之间的区分更加明显,抗噪性得到了提高.

5 结束语

针对单一端点检测低信噪比下检测率下降的缺陷,本文通过分析分形维数和 TEO 的原理,研究分形维数在时域和频域上的使用,将分形维数和短时 TEO 能量结合在一起,采用两级判决的机制对语音进行端点检测.实验结果表明,联合算法弥补了单一算法的抗噪性不足以及改善了整个系统的准确度.即使在低信噪比的情况下也有良好的性能.

参考文献:

[1] 李杰,周萍,杜志然. 短时 TEO 能量在带噪语音端点检测中的应用[J]. 计算机工程与应用,2013,49(12): 144-147.

[2] 张雪英. 数字信号处理及 MATLAB 仿真[M]. 北京: 电子工业出版社,2010:207-215.

[3] 刘悦,王晓婷. 短时频域分形端点检测算法[J]. 微电子学与计算机,2015,32(9):81-84.

[4] 陈建安. 分形维数的定义及测定方法[J]. 电子科技, 1999,1(4):44-46.

[5] Kaiser J F. On a simple algorithm to calculate the energy of a signal[C]//IEEE International conference on Acoustics,Speech and Signal Processing. Albuquerque, NM: IEEE,1990:381-384.

[6] 王茂蓉,周萍,景新幸,等. 基于 Mel-TEO 的带噪语音端点检测算法[J]. 微电子学与计算机,2016,33(4): 46-49.

作者简介:

周 萍 女,(1961-),硕士,教授.研究方向为语音识别与智能控制.

张上鑫(通讯作者) 男,(1992-),硕士研究生.研究方向为语音识别. E-mail:503216199@qq. com.

郑凯鹏 男,(1992-),硕士研究生.研究方向为语音识别.

(上接第 20 页)

[3] Tsai M C, Cheng C H, Yang C M. An all-digital high-precision built-in delay time measurement circuit [C] // 26th IEEE VLSI Test Symposium (vts 2008). San Diego, California,IEEE, 2008: 249-254.

[4] Lee M K, Teng W K, Krishnasamy R K, et al. Delay-line based embedded memory access time measurement: Circuit, implementation and characterization techniques [C]// Quality Electronic Design (ASQED), 2012 4th Asia Symposium on. Penang, Malaysia,IEEE, 2012: 259-264.

[5] Murakami D, Kuwabara T. A digitally programmable delay line and duty cycle controller with picosecond resolution[C] // Bipolar Circuits and Technology Meeting, Proceedings of the 1991. Minneapolis,

USA, IEEE, 1991: 218-221.

作者简介:

姜 彬 男,(1992-),硕士.研究方向为集成电路工程、数字电路设计. E-mail:seu_jiangbin@163. com.

张敏敏 女,(1994-),硕士.研究方向为集成电路工程、数字电路设计.

王 琴 女,(1975-),副教授.研究方向为大规模集成电路设计、先进集成电路设计方法.

蒋剑飞 男,(1979-),助理研究员.研究方向为高速集成电路设计.

毛志刚 男,(1962-),博士,教授,博士生导师.研究方向为大规模集成电路设计、可重构体系结构设计、集成电路与元器件可靠性等.