

基于分形理论的语音端点检测

冯凯 刘珩

(中国农业大学 信息与电气工程学院, 北京 100083)

摘要 为提高语音端点检测(VAD)在较低信噪比(< 10 dB)下的准确率,提出一种基于短时分形维数的改进算法。结合语音信号的特点,对2种常用的语音信号分形维数计算方法进行了比较和选择,同时采用动态跟随门限值实现语音端点的自适应检测。试验结果表明:对于信噪比6~10 dB的带噪语音,此方法可以实现整段语音的检测,而且具有一定的噪声鲁棒性,系统运行期间能够自适应调整门限值以适应环境噪声的变化,提高了VAD算法的准确率。

关键词 VAD; 分形维数; 动态门限值; 自适应

中图分类号 TN 912.3

文章编号 1007-4333(2006)04-0114-03

文献标识码 A

Study of voice activity detection based on the fractal theory

Feng Kai, Liu Heng

(College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract A reformative algorithm, based on the short-time fractal dimension, is given to improve VAD (voice activity detection) in low signal-to-noise ratio (SNR) environments. Two ways to calculate the fractal dimensions of speech signals in common were compared, and one was selected to fit the speech signals. At the same time, a dynamically updated threshold was applied to adaptively detect the speech segments in noisy speech. The experimental results showed that the method could realize the detection of speech segment in low signal-to-noise ratio (SNR at 6~10dB) and had a good noise robustness.

Key words VAD; fractal dimension; dynamic threshold; adaptive

语音端点检测(voice activity detection, VAD)是语音处理系统的关键问题之一,其准确性会影响到后续处理的结果,如:信号模型参数的确定等。传统的VAD算法主要采用能量、过零率、自相关、模式识别等方法,这些方法以语音信号的单一物理特性为判断依据,当语音信号中混入噪声信号后,特别是当信噪比较低(< 10 dB)时,其相应的物理特征会发生较大改变,从而使传统语音识别方法出现误判或者丢失音节等现象^[1~3]。

声学理论表明语音信号具有分形特性,且其所对应的物理特性并非单一的物理特性,因此可以将分形理论应用于VAD的研究;但目前的研究中往往使用单一的门限值^[4~6],且往往根据经验取得,系统运行时不能跟随背景噪声幅值的变化而变动,导致VAD算法的准确率下降。针对这一问题的改进

算法很多,主要集中于改进分形维数的计算方法^[7~8],目的在于增加其噪声的鲁棒性,但其门限值的选取仍然采用固定值或者有一定变动范围的固定值,从而影响了VAD算法的准确率,限制了算法的适用范围(背景噪声波动较小的情况)。本研究旨在解决单一门限值影响算法适用范围的问题,基于分形维数,提出一种门限值动态跟随算法。

1 分形维数及计算方法

分形维数是描述分形特征的重要参数,是分形信号最重要的特征。分形方法将维数从整数扩大到分数,突破了一般拓扑维数为整数的限制。分形维数的定义多种多样,常用的有计盒维数、信息维数、关联维数、相似维数等^[9]。本研究最终的试验平台为DSP平台,为满足实时性要求,分别采用计盒维数

收稿日期:2006-02-23

作者简介:冯凯,硕士研究生;刘珩,副教授,主要从事高噪声环境中抗噪通信技术研究, E-mail:cauzd@163.com

和关联维数方法进行计算,并根据 Matlab 的模拟试验结果,选择其中较好的一种进行基于 DSP 平台的进一步试验。

1.1 计盒维数

n 维欧氏空间子集 F 的计盒维数定义为

$$D_B = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}$$

即

$$\log N_\delta(F) = D_B \log(1/\delta) \quad (1)$$

式中 $N_\delta(F)$ 表示用边长为 δ 的网格覆盖 F 所需的最小正方形个数。实际应用中通常采用多点直线拟合的方法进行极限运算,以实现 D_B 的快速计算,其数学简化公式为

$$D_B = \frac{M \sum_{i=1}^M \log N_{\delta_i} \log i - \sum_{i=1}^M \log N_{\delta_i} \sum_{i=1}^M \log i}{\left(\sum_{i=1}^M \log i \right)^2 - M \sum_{i=1}^M (\log i)^2} \quad (2)$$

1.2 关联维数

关联维数源自多重分形维数(即 q 阶分形维数),多重分形维数提供了一组连续谱指数来描述系统特征^[1-5]。多重维数表达集合 F 的时候,不仅计算覆盖 F 所需要的方格数,还估算 $P(N_i/\delta)$,即 N_i 个点落入 δ 大小的盒子里的概率。由此可知其分形维数

$$D_q = \frac{1}{1-q} \lim_{\delta \rightarrow 0} \frac{\log \sum_{i=1}^{\max} p_i^q(m/\delta)}{\log \delta} \quad q = 2 \quad (3)$$

实际运算中常用以下方法进行简化

$$D_2 = \lim_{r \rightarrow 0} \frac{\log C(r, m)}{\log r} \quad (4)$$

式中:关联积分 $C(r, m) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N H(d_{i,j}, r)$,

$$H(d_{i,j}, r) = \begin{cases} 1 & d_{i,j} < r \\ 0 & d_{i,j} \geq r \end{cases}, d_{i,j} = |x_i - x_j|。$$

噪声语音信号序列是语音信号和噪声信号的加性序列。语音起始端点检测就是在可测量的数据中判断语音起止点。由于噪声语音信号在语音段,特别是语音中韵母声音信号中的图形较高频噪声信号(如白噪声)图形有较大的周期性和规则性,导致语音信号的分形维数与噪声的分形维数有所不同,因此,可以用信号的短时分形维数来提取语音段信号和噪声信号的特征,作为语音端点检测的依据。

2 基于分形维数的改进算法

不同的维数方法在维数的区分上有所不同:计盒维数中,语音信号的分形维数要小于噪声的分形维数;关联维数中,语音信号的分形维数要大于噪声的分形维数,其 Matlab 仿真结果见图 1。可以看出,对于背景噪声为白噪声,信噪比为 15 dB 的波形信号,由于关联维数充分考虑到集合中点的分布信息,描述了结构中点的分布密度,所以其计算结果的波动相对较小,对于背景噪声的鲁棒性较计盒维数好;又考虑到硬件平台的计算能力,故采用关联分形维数作为算法设计中的维数计算方法。

由于白噪声是一种近似理想化的噪声模型,实际的环境噪声往往与语音有很多类似之处,且当系统从一种环境进入另外一种环境时,其噪声的分形维数值将发生变化,如果采用单一的门限值,将需要手动调整以满足现有的条件。对于同一环境,如行驶中的车辆,路面的变化而产生的环境噪声也是一个动态变化的过程,如果靠手动调整门限值,将会给操作带来很大麻烦。本文中采用一种动态更新门限值的算法来解决这个问题,其原理是:由连续帧背景

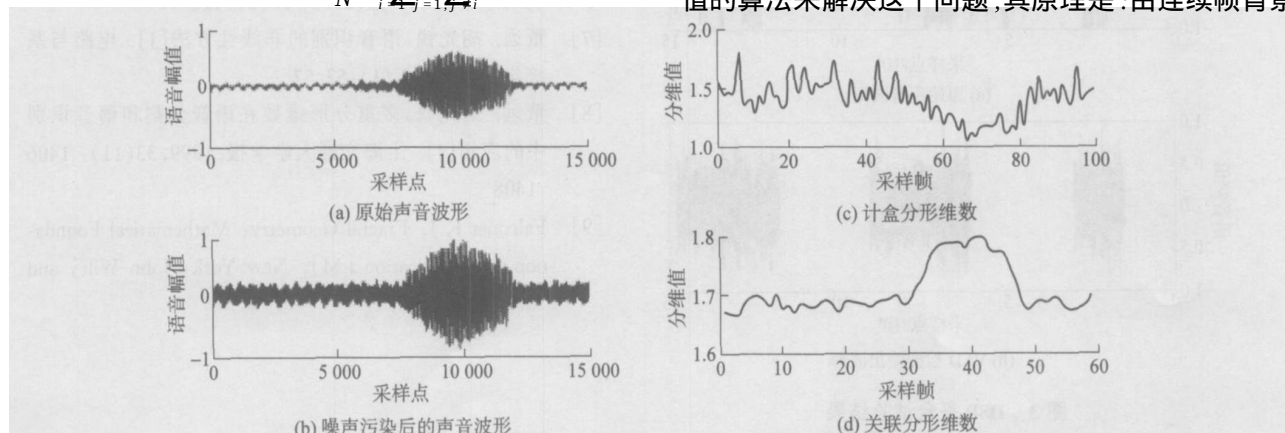


图1 分形维数计算结果

Fig. 1 Comparison of different fractal dimensions

噪声得到分形维数的门限值 $|T|$ 和变动门限值 $|T'|$,当连续帧分形维数小于 $|T|$ 时被认为是噪声,并根据当前若干帧的幅值更新门限值 $|T|$ 和幅值变动范围 $|T'|$ 值;当连续帧分形维数大于 $|T|$ 且幅值变动范围小于 $|T'|$ 时,认为环境噪声发生大的变动,由此更新 $|T|$ 值以适应当前环境的变化。其流程见图2。

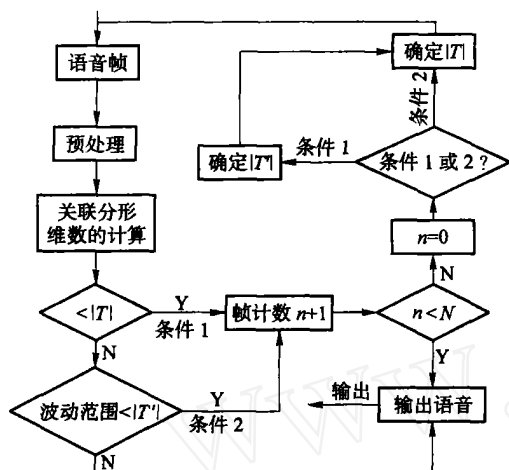


图2 改进后的算法流程图

Fig. 2 Flow Diagram of improved algorithm

3 试验结果与分析

采用 TI 公司的硬件平台 (TMS320C5402 + AIC231) 进行试验,背景噪声为时速 120 km 汽车的噪声,语音信号为正常交谈声,信噪比 6 ~ 10 dB,结果见图3。由于采用的送话器具有一定抗噪功能,

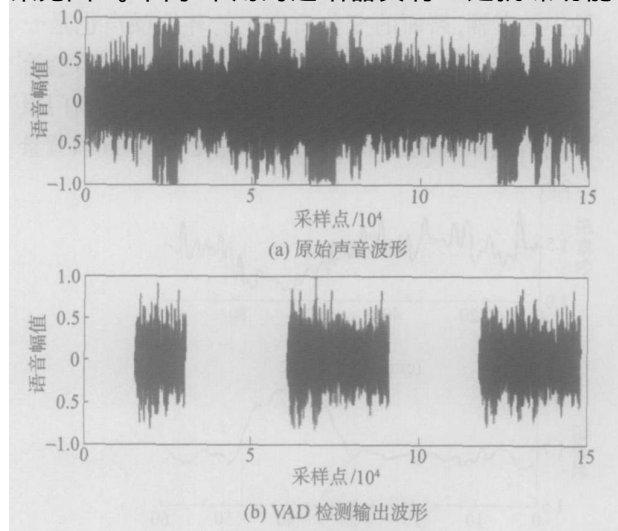


图3 DSP平台试验结果

Fig. 3 Results from DSP plat

所以处理后的语音幅值会有所减小。可以看出,对于信噪比为 6 ~ 10 dB 的信号,该方法能够比较准确地起到语音检测的作用,而且其延时很短。

4 结论

在语音信号维数计算方法上,关联维数的计算中考虑了更多细节因素,其计算结果较计盒维数具有更好的稳定性,能够较全面地反映语音信号的特征。试验结果表明,对于信噪比为 6 ~ 10 dB 的语音信号,改进后的基于分形维数的算法能够实现语音端点的准确检测。该算法在较低信噪比时的识别准确率优于传统算法。

该算法的不足之处为会出现清音丢字情况,且对于波动过大的背景噪声其处理能力仍然有限,有待进一步改进和提高。

参考文献

- [1] 杨行峻,迟惠生. 语音信号数字处理[M]. 北京:电子工业出版社,1995:8~21
- [2] Quatieri T F. 离散时间语音信号处理——原理与应用[M]. 赵胜辉,等译. 北京:电子工业出版社,2004
- [3] Marzinik M. Kollmeier B. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics[J]. IEEE Transaction on Speech and Audio Processing, 2002, 10(2): 109~118
- [4] 费珍福,王树勋,何凯. 分形理论在语音信号端点检测及增强中的应用[J]. 吉林大学学报, 2005, 23(1): 139~142
- [5] 雷涛,姚明海,马洪蕊. 基于 MFCC 与分形维数混合参数的语音识别[J]. 控制工程, 2005, 12(5): 93~98
- [6] 陈亮,张雄伟. 基于分形维实现语音分割和增强[J]. 北京邮电大学学报, 2003, 26(6): 112~114
- [7] 董远,胡光锐. 语音识别的非线性方法[J]. 电路与系统学报, 1998, 3(1): 52~57
- [8] 董远,胡光锐. 多重分形维数在语音分割和语音识别中的应用[J]. 上海交通大学学报, 1999, 33(11): 1406~1408
- [9] Falconer KJ. Fractal Geometry: Mathematical Foundation and Application [M]. New York: John Wiley and Sons, 1990