

- 22 Helton Hideraldo Biscaro, Antonio Castelo Filho, Luis Gustavo Nonato, et al. A topological approach for surface reconstruction from sample points. *The Visual Computer*, 2007,23(9):793–801.
- 7 Ruzzeli AG, Jurdak R, O'Hare GMP. On the RFID wake-up impulse for multi-hop sensor networks. *Proc. of 1st ACM Workshop on Convergence of RFID and Wireless Sensor Networks and their Applications (2007)*. 2007.
- 8 段翠琴,孙晶晶,周端.WSN 中能量有效的分区聚簇多跳路由算法. *计算机工程*,2010,36(14):1000–3428.
- 9 李岩,张曦煌,李彦中.LEACH-EE-基于 LEACH 协议的高校聚类路由算法. *计算机应用*,2007,27(5):1103–1105.
- 10 Yuh-Ren Tsai. Coverage-Preserving Routing Protocols for Randomly Distributed Wireless Sensor Networks, *IEEE Trans. on Wireless Communications*, April 2007,6(4):1240–1245.

语音端点检测研究^①

邓艳容¹, 景新幸¹, 杨海燕¹, 杨运泽²

¹(桂林电子科技大学 信息与通信学院, 桂林 541004)

²(桂林电子科技大学 机电工程学院, 桂林 541004)

摘要: 论文对噪声环境下的语音端点检测方法进行了系统地研究。由于传统的短时能量和短时过零率双门限算法在低信噪比条件下不能检测出语音信号的端点, 对此, 本文提出了一种改进的谱熵-双门限算法, 文中给出了改进算法的实现框图, 并用 MATLAB 进行了算法仿真, 仿真结果表明该算法具有一定的鲁棒性, 在较低信噪比下仍能准确地区分有用信号和噪声, 从而验证了该算法的有效性和高效性。

关键词: 语音端点检测; 谱熵; 双门限

Research on Endpoint Detection of Speech

DENG Yan-Rong¹, JING Xin-Xing¹, YANG Hai-Yan¹, YANG Yun-Ze²

¹(School of Information and communication, Guilin University of Electronic Technology, Guilin 541004, China)

²(School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: Systemic research has been done on the endpoint detection algorithms of speech in the noise environment. traditional method of short-time energy and short-time zero-crossing rate of double threshold can not detect speech signal endpoint in low SNR. To solve this problem, A improved algorithm of spectral entropy-double threshold is proposed in this paper, the diagram of realization of this algorithm is presented and simulated in MATLAB. Simulation results show that the algorithm perform well on anti-noise, they can still accurately distinguish between voice and noise, so the effectiveness and high efficiency of the improved algorithm is proved.

Key words: speech endpoint detection; spectral entropy; double threshold

1 引言

语音端点检测是指从一段语音信号中准确地找出语音信号的起始点和结束点, 其目的是为了使有效的语音信号和无用的噪声信号得以分离, 在语音识别、语音增强、回声抵消等系统中得到广泛应用。目前端

点检测方法大体上可以分成两类^[1], 一类是基于模式识别的方法, 由于该方法自身复杂度高, 运算量大, 因此很难被应用到实时语音信号系统中去; 另一类方法是基于阈值的方法, 该方法根据语音信号和噪声信号的不同特征, 提取每一段语音信号的特征, 然后把这些特征值与设定的阈值进行比较, 从而达到语音端点检测的目的, 该方法原理简单, 运算方便, 被人们

广泛使用。

本文主要对基于阈值的方法进行研究,首先对传统的短时能量和短时过零率双门限端点检测法进行简单的介绍,其次针对该方法在强噪声环境下检测准确率急剧下降这一缺陷进行改进,提出了一种具有鲁棒性的抗噪能力强的谱熵-双门限端点检测算法。

① 基金项目:国家自然科学基金(60961002);桂林电子科技大学博士科研启动基金(UF08018Y);

教学团队建设—微电子学专业主干课程教学团队(ZJT0121A)

收稿时间:2011-11-01;收到修改稿时间:2011-12-07

然后对每一帧分别求其能量。短时能量是随机参数,在无声段,清音段和浊音段,它们分别具有不同的概率密度函数,一般浊音的短时能量最大,清音能量次之,无声的短时能量最小。

n 时刻某语音信号 $\{x(n)\}$ 的短时平均能量定义为:

$$E_n = \sum_{m=-\infty}^{\infty} x^2(m) \cdot h(n-m) \quad (1)$$

其中 $h(n)$ 为可移动的有限长度的窗函数,是低通滤波器的单位冲激响应,用来实现分帧处理。

2.2 短时过零率

短时过零率表示一帧语音信号波形穿过横轴(零电平)的次数,即样本改变符号的次数。语音信号 $X_n(m)$ 的短时过零率 Z_n 可以定义为:

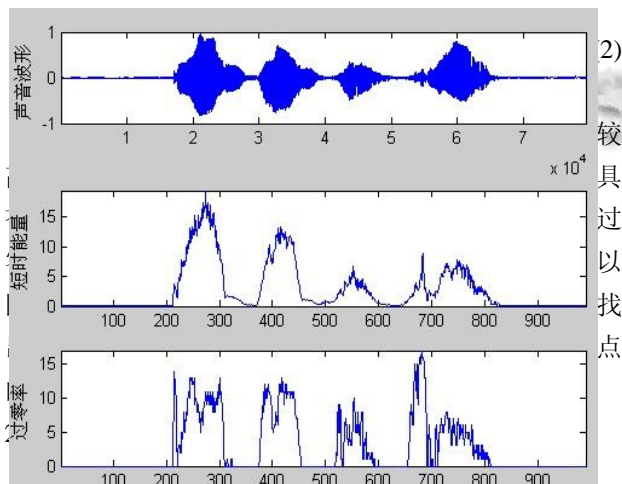


图 1 语音信号的短时能量和过零率波形图

2 传统短时能量和过零率双门限端点检测

2.1 短时平均能量

语音和噪声的主要区别在于它们的能量上,因此能量是常用的音频特征参数,是语音信号最直观的表现^[2]。语音信号一般可分为无声段(静音段)、清音段和浊音段。由于语音是不稳定的,所以常采用“短时能量”。所谓短时能量,就是先对语音信号进行分帧处理,

双门限法是利用短时能量和过零率的乘积进行检测的。其基本思想是为短时能量和过零率分别确定两个门限,一个为较低的门限,对信号的变化比较敏感,另一个是较高的门限。当低门限被超过时,很有可能是由于很小的噪声所引起的,未必是语音的开始,当高门限被超过并且在接下来的时间段内一直超过低门限时,则意味着语音信号的开始。该方法充分利用过零率和能量的特点,用过零率检测清音,用短时能量检测浊音,两者结合起来即可检测语音端点。下图为某一语音信号的短时能量和过零率波形图。

3 改进的谱熵-双门限端点检测

传统的短时能量-短时过零率的检测方法虽然计算简单,但是在低信噪比环境下,检测准确率急剧下降,特别是在噪声很大时,完全不能检测出语音端点。为了提高端点检测准确率,解决语音信号能量小且易被淹没这一特点,本节提出谱熵-双门限端点检测算法并对其进行改进。

3.1 谱熵语音端点检测原理

熵用来表示任何一种能量在空间中分布的均匀程度,是一个系统混乱程度的度量,与能量特征相比,

谱熵的变化很小，所以在某种程度上熵还具有鲁棒性的特征，具有一定的抗噪能力，谱熵语音端点检测方法主要是检测谱的平坦程度，有用语音段相对于背景噪声而言，它的能量主要集中在某几个频段，起伏突变比较大，熵值小，而噪声信号在整个频带内分布相对比较平坦，其熵值比较大，因此可以利用这种差异来区分有用语音段和噪声段，从而达到语音端点检测的目的。

谱熵的计算首先通过快速傅立叶变换(FFT)得到每一帧信号的频谱，其中每个频谱向量的系数表明了该帧信号在该频率点的大小分布，然后计算每个频谱分量在每帧总能量中所占的比例，将其作为信号能量集中在某频率点的概率，其概率密度函数定义为：

$$p_i = s(f_i) / \sum_{k=1}^M s(f_k) \quad i = 1, \dots, M \quad (3)$$

式中 $s(f_i)$ 是 f_i 的能量， p_i 是相应的概率密度， M 是 FTF 中频率成分的所有点数。则相应的每一帧的谱熵定义如下：

$$H = -\sum_{k=1}^M P_k \cdot \log P_k \quad (4)$$

3.2 改进的谱熵-双门限端点检测原理

为了更好地进行语音端点检测，本文采用语音信号的短时功率谱构造语音信息谱熵，从而更好的对语音段和噪声段进行区分，同时在进行端点检测之前先用谱减法进行一定的降噪处理^[3]。谱减法即为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。

为进一步减小噪声对端点检测的影响，采取把带噪语音信号划分成多个频带进行处理^[4]，这样频带的能量谱定义为：

$$G(n, m) = \frac{\sum_{k=1+5(n-m)}^{1+5(n-1)+3} |X(k, m)|^2}{\sum_{k=1}^N E(k, m)} \quad 1 \leq n \leq N \quad (5)$$

$$G(m) = \sum_{k=1}^N E(k, m) \quad (6)$$

其中 N 为频带个数，每 5 点为一个频带，则 $G(n, m)$ 的概率密度为：

$$P(n, m) = \frac{G(n, m)}{G(m)} \quad (7)$$

则多带谱熵为：

$$H(m) = -\sum_{n=1}^N P(n, m) \log P(n, m) \quad (8)$$

最后根据多带能量谱熵确定双门限 $T1$ 和 $T2$ ，且 $T1 > T2$ 。检测时使用变量 $status$ 表示当前状态， $status=0$ 时设置为静音， $status=1$ 时为可能处于语音段， $status=2$ 时确定进入语音段， $status=3$ 时为语音结束， $count$ 为根据 $status$ 的改变的计时长度计数，具体检测流程如下所示^[5]：

① 最开始静音段时 $status$ 为 0， $count$ 为 0，如果谱熵值超过低门限 $T2$ ，就开始标记起点，进入过渡段， $status$ 变为 1， $count$ 加 1，过渡段时当谱熵值降到较低门限 $T2$ 以下，则认为当前状态恢复到静音状态，此时 $status$ 又变为 0， $count$ 也变为 0；

② 如果过渡段中谱熵值的大小超过较高门限 $T1$ ，则认为进入语音段，此时 $status$ 变为 2。

③ 当处于语音段时，如果谱熵值小于较低门限 $T2$ ，而且总的计时长度小于最短时间长度，则认为是一段噪音，继续扫描语音段，否则标记为语音的终点。

综上所述，改进的谱熵-双门限算法原理如下图 2 所示：

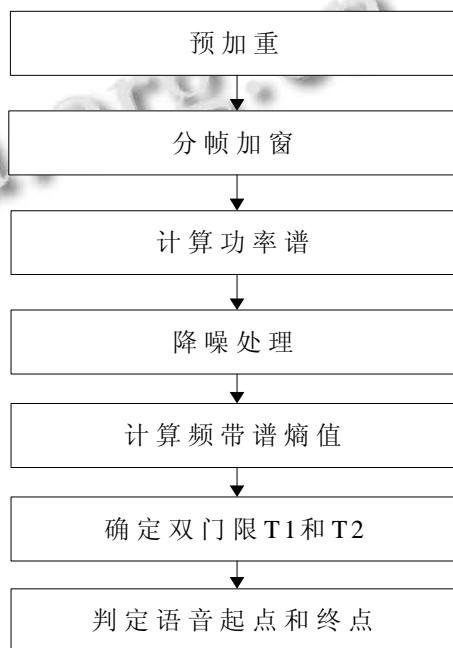


图 2 改进的谱熵-双门限算法流程图

4 实验仿真

4.1 实验环境及参数设置

为了验证算法的性能,在 MATLAB 软件下进行仿真实验。语音信号为标准普通话“我爱北京”,在实验室录制完成,采样频率为 16KHz,采样精度为 16bit,采用汉明窗进行分帧,帧长为 1024 点,帧移为 512 点,通过特性为 $(1-0.94Z^{-1})$ 的滤波器预加重。利用伪随机加性高斯白噪声对纯净语音信号进行加噪处理,从而形成不同信噪比 (SNR) 的带噪语音信号,然后对其进行端点检测,并改变信噪比值,比较在高信噪比和低信噪比条件下两种算法的端点检测结果。

4.2 性能分析

本实验采用传统的短时能量-短时过零率双门限端点检测法和改进的谱熵-双门限端点检测法分别对语音信号进行端点检测,两种算法采用同一段原始语音信号,在不同信噪比情况下的端点检测结果如下图所示,其中红色竖线表示语音起点,绿色竖线表示语音终点。

从图 3 和图 4 可以看出在高信噪比 (SNR=62.7) 条件下,传统的短时能量-短时过零率双门限端点检测法和改进的谱熵-双门限端点检测法检测效果都很好,都能较准确地检测出语音端点,但是随着信噪比的降低,传统的短时能量-短时过零率双门限端点检测效果明显变差,从图 5 可以看出在低信噪比 (SNR=22.7) 时,传统的短时能量-短时过零率双门限端点检测法完全不能检测出语音信号的端点,而在相同条件下改进的谱熵-双门限端点检测法仍能较好地检测出语音端点,如图 6 所示。由此可见,传统的短时能量-短时过零率双门限端点检测法虽然简单易实现,但抗噪性太差,在信噪比较低的环境下基本失去检测能力,而改进的谱熵-双门限端点检测法具有较强的抗噪声能力,且检测结果精确高效。

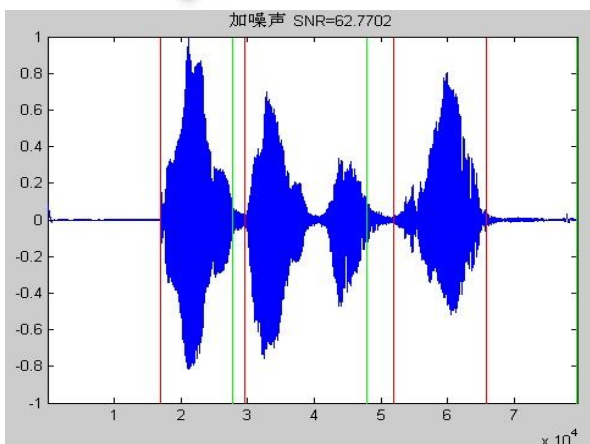


图 3 SNR=62.7 时短时能量-过零率端点检测波形图

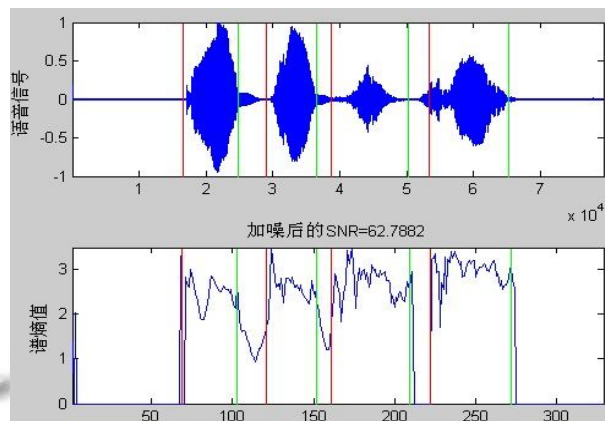


图 4 SNR=62.7 时改进的谱熵-双门限端点检测波形图

5 结论

本文在传统的短时能量-短时过零率双门限端点检测算法的基础上,提出了一种改进的谱熵-双门限端点检测法,该算法充分地利用了熵的抗噪性,而且在端点检测之前进行了降噪处理。通过理论分析和实验仿真结果表明,改进的谱熵-双门限端点检测法相对传统的短时能量-短时过零率双门限端点检测法而言,算法稳健,抗噪性能强,鲁棒性好,从而验证该方法是一种行之有效的方法。

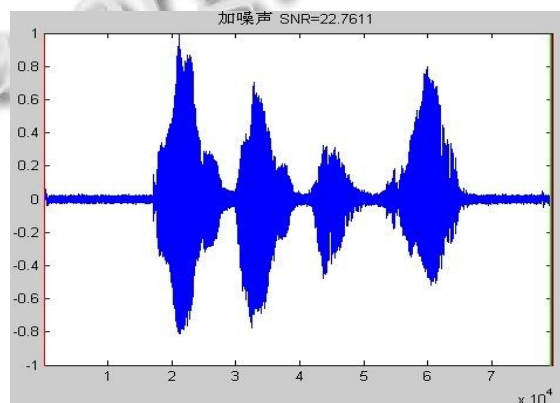


图 5 SNR=22.7 时短时能量-过零率端点检测波形图

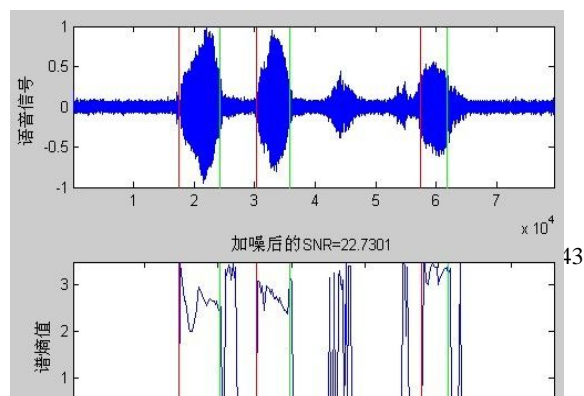


图 6 SNR=22.7 时改进的谱熵-双门限端点检测波形图

参考文献

- 1 He Suning Yu, Juebang. A. Novel Chinese Continuous Speech Endpoint Detection Method Based on Time Domain Features of the Word Structure. IEEE Int. Conf. on Commun. Circuits and Systems and West Sino Expositions, 2002. 992–996.
- 2 Qi Li, Jinsong zheng, Tsai A, et al. Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition. IEEE Transactions on Speech and Audio Processing, 2002, 10(3): 146–152.
- 3 赵丽霞. 噪声环境下基于特征的语音端点检测研究. 长沙: 湖南大学, 2010.
- 4 李晋. 语音信号端点检测算法研究. 长沙: 湖南师范大学, 2006.
- 5 何强, 何英. MATLAB 扩展编程. 北京: 清华大学出版社, 2002.