

基于子带二次谱熵的语音端点检测

朱建伟^{1,2}, 孙水发¹, 但志平¹, 雷帮军¹

(1 三峡大学 智能视觉与图像信息研究所, 湖北 宜昌 443002; 2 三峡大学 电气与新能源学院, 湖北 宜昌 443002)

摘 要: 为了提高在强噪声环境下语音端点检测的准确度, 提出基于子带二次谱熵的端点检测算法. 该算法把子带二次谱熵作为端点检测新的特征参数, 首先计算每帧语音信号的二次谱, 再多子带分析, 计算二次谱熵; 引入顺序统计滤波对二次谱熵平滑处理; 将有限状态机判别方法与子带二次谱熵相合, 形成新的语音/噪声判别算法, 有效地解决单门限法易出现的两类误判. 实验表明: 与传统的两种方法相比, 提出的端点检测算法具有准确性高、抗噪性强等优点.

关键词: 端点检测; 子带二次谱熵; 有限状态机; 顺序统计滤波

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-7180(2011)03-0077-04

Voice Activity Detection Based on Sub-band Reprocessed Pectrum Entropy

ZHU Jian-wei^{1,2}, SUN Shui-fa¹, DAN Zhi-ping¹, LEI Bang-jun¹

(1 Institute of Intelligent Vision and Image Information, China Three Gorges University, Yichang 443002, China;
2 College of Electrical Engineering and New Energy Technology, China Three Gorges University, Yichang 443002, China)

Abstract: Voice activity detection (VAD) in strong noise environments is improved by an algorithm based on subband reprocessed spectrum entropy (BRSE). As a new feature parameter for VAD, the reprocessed spectrum is calculated firstly, and reprocessed spectrum entropy is then calculated with multi-sub-band analysis. The order statistics filter is selected to smooth the BRPE. A new voice / noise discrimination algorithm is proposed by combining the finite state machine (FSM) with BRSE. The misdetections caused by single-threshold are reduced greatly. Experimental results show that the proposed algorithm has higher accuracy and stronger robustness than other two methods.

Key words: voice activity detection; subband reprocessed power spectrum entropy; finite state machine; order statistics filter

1 引言

端点检测是对语音信号中的有声片段和无声片段进行分割. 作为语音识别的前端, 准确的端点检测可以提高识别的准确率; 用于语音增强系统中, 可以进行准确的噪声模型估计; 在语音编码领域中可以降低编码的平均比特率并降低功耗. 端点检测的准确与否直接影响着语音信号的后续处理. 典型的

端点检测方法首先计算短时语音特征(如短时能量、过零率等), 再根据门限做出判决. 这些方法在高信噪比环境下比较有效, 但在强噪声环境下, 检测正确率急剧下降. 为了解决这一问题, 许多新算法被提出, 其中基于熵的端点检测方法具有较好的噪声鲁棒性^[1-5].

目前基于谱熵的端点检测, 首先把分帧后的语音信号进行傅氏变换, 求其功率谱计算谱熵. 文献

收稿日期: 2010-04-23; **修回日期:** 2010-07-01

基金项目: 国家自然科学基金项目(60972162); 湖北省自然科学基金项目(2008CDB346); 湖北省教育厅重大项目(Z20081301); 湖北省高等学校优秀中青年科技创新团队计划项目(T201002); 湖北省教育厅中青年项目(Q20101202, Q20101205); 宜昌市科学技术研究与开发项目(A09302-31, A09302-32, A2010-302-10)

[6]在基音检测方法研究中发现二次谱与倒谱具有相同的量纲,但比倒谱具有更好的语音/噪声区分度.本文对每个语音帧求二次谱,提出了子带二次谱熵这一新特征,并经过顺序统计滤波处理,新特征在强噪声环境下的区分度明显高于目前常用的谱熵和传统的短时能量特征.在噪声/语音判别算法中,把子带二次谱熵与有限状态机判别方法^[7]相结合,形成新的判别法.

2 子带二次谱熵及其使用

2.1 子带二次谱熵的提出

功率谱二次处理(二次谱)具有较好区分语音和噪声的功能^[5]:(1)浊音二次谱处理结果是一列相关函数串;(2)二次谱分辨率与倒谱一样具有时间量纲;(3)浊音二次谱具有很大的直流分量,噪声则无;(4)具有较好的噪声鲁棒性.因为端点检测的目的就是为了区分有声段和无声段,而二次谱具有以上特点,所以本文提出了基于子带二次谱熵的端点检测方法.

对具有 M 帧,每帧 N 个点的语音信号求 FFT,再取模得到每帧信号“功率谱”,其中 k 为频率分量, l 为当前分析的帧.“功率谱”加引号是由于其为取模,并不是取模平方.

$$X_{\text{energy}}(k, l) = |x(k, l)|, 1 \leq k \leq N \quad (1)$$

对“功率谱”做 IFFT 后取模,即得到“二次谱”.二次谱与倒谱具有一样的量纲,是关于 $N/2 + 1$ 点对称的^[6].

$$X_{\text{RE}}(k, l) = | \text{IFFT} \{ X_{\text{energy}}(k, l) \} |, 1 \leq k \leq N/2 \quad (2)$$

每个二次谱能量占整个二次谱能量的概率为

$$p(k, l) = \frac{X_{\text{RE}}(k, l)}{\sum_{l=1}^M \sum_{k=1}^{N/2} X_{\text{RE}}(k, l)} \quad (3)$$

每一帧二次谱熵为

$$H(l) = \sum_{k=1}^{N/2} p(k, l) \cdot \log(1/p(k, l)) \quad (4)$$

文献[2]研究表明每一个功率谱点的幅度易受噪声的干扰,在低信噪比时参数性能会有所下降,提出多子带分析,从而克服了在噪声环境中功率谱的敏感性.本文也对二次谱做多带分析,假设帧长为 256 个采样点,每 4 个二次谱点为一个子带,共 $N_b = 32$ 个子带,每个子带二次谱定义为

$$X_{\text{REB}}(m, l) = \sum_{k=1+(m-1)*4}^{1+(m-1)*4+3} X_{\text{RE}}(k, l), 1 \leq m \leq N_b \quad (5)$$

相应的子带二次谱概率和子带二次谱熵定义为

$$Pb(m, l) = \frac{X_{\text{REB}}(m, l)}{\sum_{l=1}^M \sum_{m=1}^{N_b} X_{\text{RE}}(m, l)} \quad (6)$$

$$Hb(l) = \sum_{m=1}^{N_b} Pb(m, l) \cdot \log(1/Pb(m, l)) \quad (7)$$

2.2 顺序统计滤波

顺序统计滤波是图像处理中常用的滤波方法. Ramirze^[8] 等人首先将其用在语音端点检测中,设有 $L = 2N + 1$ 个序列升序排列为 $\{x(l-N), \dots, x(l), \dots, x(l+N)\}$, 顺序统计滤波器的输出是 L 个信号序列的线性组合: $y(l) = \sum_{i=1}^L a_i \cdot x(i)$, 系数 a_i 决定了滤波器的特征, 中值滤波器就是顺序统计滤波器的一种. 本文顺序统计滤波器实现步骤如下:

(1) 计算式(7)时对每一帧的子带熵暂不累加, 取长度为 $L = 2N + 1$ 的子带二次谱熵 $E_b(m, l-N), \dots, E_b(m, l), \dots, E_b(m, l+N)$, 其中 $E_b(m, l)$ 表示当前分析的是第 l 帧, 第 m 个子带;

(2) 对上步中的 L 个熵升序排序, 找出序列中第 h 个最大值, 记为 $E_{b(h)}(m, l)$;

(3) 第 l 帧的第 k 个子带二次谱熵滤波输出结果定义为

$$E_b(m, l) = (1 - \lambda) * E_{b(h)}(m, l) + \lambda E_{b(h+1)}(m, l) \quad 1 \leq m \leq N_b$$

式中, $h = \lfloor \lambda L \rfloor$, $(0 < \lambda < 1, L = 2N + 1)$, λ 称为顺序统计滤波器采样分位数, 本文 λ 取 0.85.

(4) 计算第 l 帧经顺序统计滤波后的二次谱熵:

$$Hb(l) = \sum_{k=1}^{N_b} E_b(m, l)$$

图 1(a) 为信噪比为 0 db 时语音波形, (b) 是子带二次谱熵, (c) 为经顺序统计滤波后的子带二次谱熵. 可以看出, 滤波后波形平滑, 语音/噪声区分度增强.

2.3 基于有限状态机的判别

通过子带二次谱熵计算和顺序统计滤波后, 每帧信号可以得到一个子带二次谱熵 $Hb(l)$. 在噪声与有声段分类过程中, 文献[3]均预先设定一个门限阈值, 当 $Hb(l)$ 高于这个阈值时判为语音, 否则为噪声. 然而这种单一门限的判别方容易产生以下两类误判: 一是把高能瞬间噪声(如关门声)误判为有效语音; 二是把语间停顿(如 six/siks/)误判为多段语音. 基于有限状态机的判别方法^[7]能够有效地解决上述两类误判, 该检测方法把语音信号分为静音态、起始预备态、有效数据态、结束预备态 4 个状态和 7

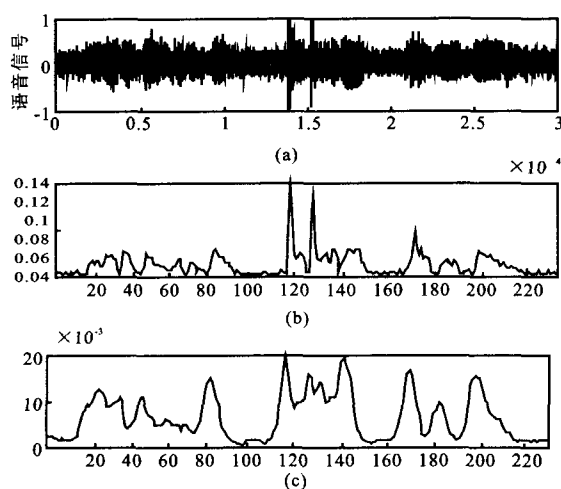


图1 顺序统计滤波后的子带二次谱熵波形图

个主要事件(如图2所示). 这种检测方法是基于短时能量和过零率两种特征设计的,在低信噪比时这两种特征易失效,端点检测精度也会大大降低. 本文把子带二次谱熵特征与有限状态机判别机制相结合对噪声与有声段进行分类. 与文献[6]不同的是,用子带二次谱熵特征替代短时能量与过零率特征.

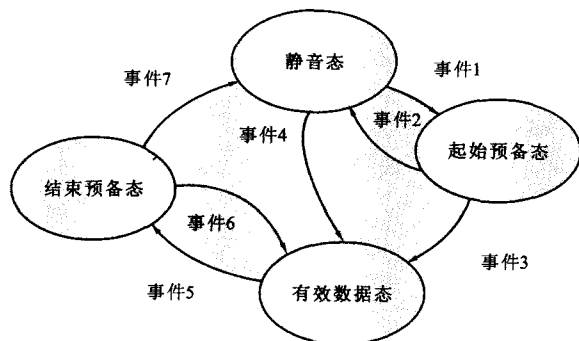
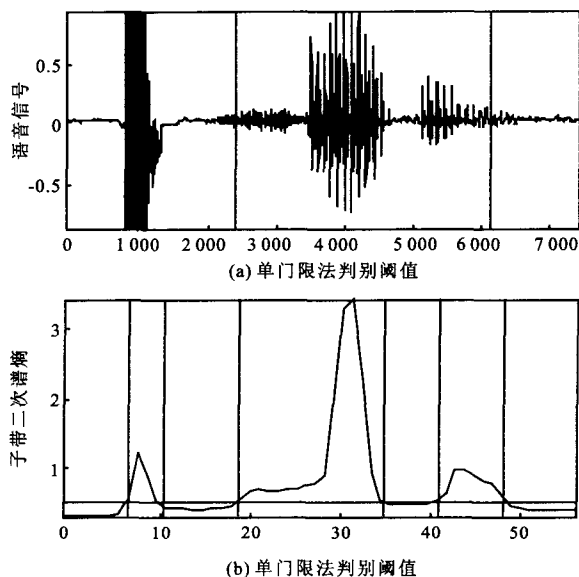


图2 端点检测状态转换图

图3(a)为一瞬时噪声和英文数字 seven 的语音波形, (b)为经顺序统计滤波后的子带二次谱熵波形, 图中的横线为单门限法判别阈值. 可以看出, 单门限法把瞬时噪声误判为语音段, 且把 seven 单词折成两部分. 有限状态机判别法可以很好的解决这两类误判, 图3(a)为判别结果, 剔除噪声部分, 保留了语音停顿.

3 实验结果分析

为了证明本文提出的基于有限状态机的子带二次谱熵端点检测方法的准确性和抗噪性, 实验分为两组: 第一组实验首先对实验语音进行端点手工标注, 再将不同方法检测结果与标注进行比较; 第二组实验将不同方法端点检测后的语音进行说话人识别, 通过比较识别率来比较不同方法的优劣. 本文实

图3 单门限法与有限状态机的判别结果比较
实验中的噪声为标准的 noise 92 中的白噪声.

3.1 实验一

实验一所采用的语音库为 20 人的录音, 录音内容为日常用语, 采样频率为 8 000 Hz, 精度为 16 Bit, 每人语音中共有 10 段短语或孤立词. 首先用 cooledit 软件对语音端点手工标注, 然后采用传统的短时能量、过零率^[6], 目前流行的子带谱熵^[3]和本文方法对上述语音进行端点检测, 最后把这三种方法的检测结果与手工标注的结果比较. 由图4可知, 在低信噪比时, 子带二次谱熵比子带谱熵和短时能量更能较好地地区别语音与噪声.

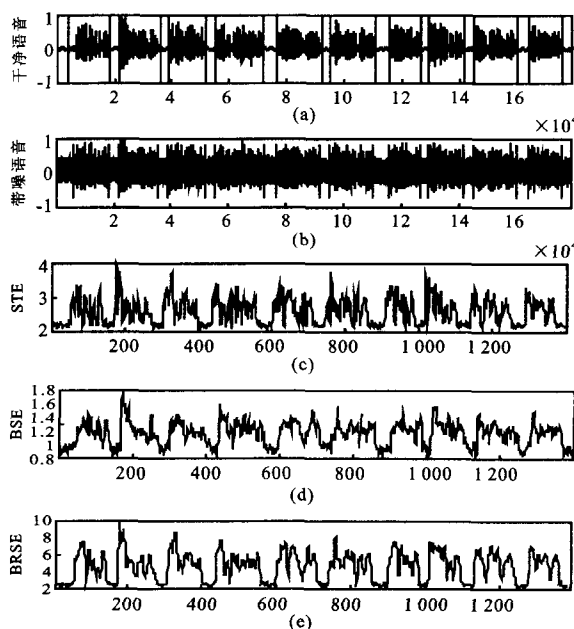


图4 低信噪比时三种方法特征图

在与标注的结果比较时,如果检测结果在前后 5 帧的范围之内,认为检测结果正确,否则认为错误.本文实验中,帧长设为 256 个采样点,帧移为 128 个采样点.表 1 所示为三种检测方法正确率比较,在信噪比为 20 dB 和 10 dB 时,子带谱熵(BSE)的检测准确率要高于短时能量(STE)和过零率,但在 5 dB 和 0 dB 时,子带谱熵的准确率不及短时能量,这是由于采用的语音/噪声判别方法造成的,正如上文分析到,文献[6]的基于有限状态机的判别方法要优于文献[3]中的单门限判别.基于子带二次谱熵的检测方法无论是在低信噪比还是在高信噪比时,均要优于其他两种方法.

表 1 不同信噪比时端点检测正确率

	0 db	5 db	10 db	20 db
STE ^[7]	74.5	89	92.5	97.5
BSE ^[3]	63.5	83.5	95.5	98
BRSE	91.5	95	99	99.5

3.2 实验二

与语音识别不同的是,说话人识别根据说话人的语音特征来区别说话人的身份,而噪声不含这种区分信息.所以,对于一段语音,如果端点检测算法把噪声误判为语音,会使识别率降低;如果把语音误判为噪声,则训练和识别的有用语音段就会减少,也会使识别率降低.实验二所采用的语音库为 200 人(100 男,100 女)的英文电话信道语音库.分别对不同信噪比的语音采用上述三种方法进行端点检测,再对语音进行分帧,提取 20 维的 MFCC 参数作为说话人识别的特征参数.训练和识别时采用高斯混合模型.

实验结果如图 5 所示,基准方法是不经过端点检测,对上述语音库,随机截取 20 s 作为训练语音,随机截取 3 s 作为识别语音.其他三种方法均为先对截取的训练和识别语音端点检测处理,再进行说话人识别.实验结果表明:经过端点检测后,说话人识别率有较大的提高;对于三种检测方法,子带二次谱熵法的性能要优于其他两种方法.

4 结束语

本文提出了子带二次谱熵这一新的端点检测特征,并对它顺序统计滤波处理,在低信噪比时仍有很好的噪声/语音区分度.把基于有限状态机的噪声/语音判别方法,与子带二次谱熵相结合,提出了基于有限状态机的子带二次谱熵端点检测算法.与传统

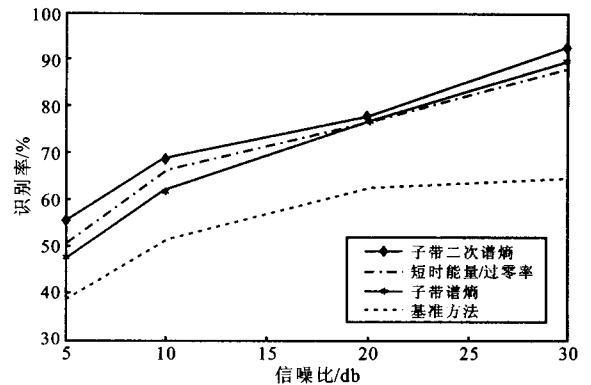


图 5 三种端点检测方法对识别率的影响

的短时能量法和目前普遍使用的谱熵法相比,本文方法具有较高的准确性和较好的鲁棒性.

参考文献:

- [1] Shen J L, Hung J W, Lee L S. Robust entropy-based endpoint detection for speech recognition in noisy environment [J]. ICSLP, 1998:232-238.
- [2] Wu Bingfei, Wang Kunching. IEEE Robust Endpoint Detection Algorithm Based on the Adaptive Band-Partitioning Spectral Entropy in Adverse Environments[J]. IEEE Trans Speech Audio Process 2005, 13(5): 762-775.
- [3] 郭丽惠,何昕,张亚昕,等. 基于顺序统计滤波的实时语音端点检测算法[J]. 自动化学报, 2008, 34(4): 419-425.
- [4] 钱彦旻,刘加. 基于交叉熵顺序统计滤波的语音端点检测算法[J]. 清华大学学报:自然科学版, 2009, 49(10): 87-90.
- [5] 韩韬,王玲,刘辉. 一种应用于语音识别的端点检测改进方法[J]. 微电子学与计算机, 2008, 25(5): 146-149.
- [6] 张天骐,张战,林孝康,等. 一种语音信号基音检测的功率谱二次处理方法[J]. 计算机工程, 2006, 32(1): 1-3.
- [7] 黄秋安,姜波,汪秉文. 基于有限状态机的汉语数字语音端点检测[J]. 湖北大学学报:自然科学版, 2004, 26(1): 35-38.
- [8] Ramirez J, Segura J C, Benitez A, et al. An effective subband OSF-based VAD with noise reduction for robust speech recognition[J]. IEEE Transaction on Speech and Audio Processing, 2005, 13(6): 1119-1129.

作者简介:

朱建伟 男, (1985-), 硕士研究生. 研究方向为说话人识别.

孙水发 男, (1977-), 博士, 副教授. 研究方向为多媒体信息处理、信息隐藏等.