一种基于能量聚类分析的句子语音端点检测法

罗世谦,冯子亮,张 恒
(四川大学 计算机学院,四川 成都 610064)

摘 要:针对语音复读系统等背景噪声相对较小且稳定的实际应用环境,提出一种改进的基于时域分析的句子语音端点检测算法。因为在此类应用环境中,对句子语音端点检测的干扰因素较少,且一般需要实现快速的句子语音端点检测。因此,简化了所要使用的语音特征参数,不再使用时域分析中常用的过零率,仅使用语音信号能量特征值辅以聚类分析完成语音端点检测。实验表明,本算法简化了端点检测的过程,可以便捷有效地检测出句子中的语音端点。

关键词:句子语音端点检测;能量特征;聚类分析
中图分类号:TN912.34 **文献标识码**:A **文章编号**:1673-629X(2008)04-0013-03

A Sentential Endpoint Detection Algorithm Based on Energy Eigenvalue and Clustering Analysis

LUO Shi-qian, FENG Zi-liang, ZHANG Heng
(College of Computer Science, Sichuan University, Chengdu 610064, China)

Abstract: To actual application system of low noise, such as voice replayer, an improved sentential endpoint detection algorithm based on time domain analysis was discussed in this paper. Because of less disturbed inspects in this application environment, and needing realize fast speed endpoint detection of sentences, simplify the parameters of voice character and do not use cross-zero rate. By using speech energy eigenvalue and clustering analysis, the endpoint of sentences can be detected very rapidly. Experiments show that the algorithm can simplify the process of sentential endpoint detection and have excellent performance.

Key words: endpoint detection of sentences; energy eigenvalue; clustering analysis

0 引 言

在语音复读系统以及语音识别系统中,句子的语音端点检测非常关键。传统的基于时域分析[1,2]的语音端点检测算法需要考虑短时线性能量、短时对数能量、短时过零率等多种语音特征值[3,4],并通过门限判断句子的端点,在大多数情况下可以获得较好的效果,但存在检测速度慢、效率不高等问题。在一些使用环境相对简单,背景噪声相对较小且稳定的情况中,需要实现快速的句子语音端点检测。

针对背景噪声相对较小且稳定的实际应用环境,在基于时域分析的语音端点检测算法基础上,提出基于能量特征和聚类分析的检测算法。仅使用能量特征一个特征值,通过对所有语音端点间隙值的聚类分析得到句子端点,不仅可以简化端点检测过程,而且对不同语速语音材料有较好的适应性。

收稿日期:2007-07-11
基金项目:国家863计划资助项目(2006AA12A104)
作者简介:罗世谦(1984-),男,四川人,硕士研究生,研究方向为实时软件工程;冯子亮,副教授,研究方向为图形图像等。

1 句子语音端点检测处理

1.1 全部语音端点检测

通过对语音帧的能量特征值的计算,实现包括单词端点和句子端点在内的全部语音端点检测。具体处理包括快速分帧、归一化处理帧、帧能量计算等过程。

(1)快速分帧:传统的分帧算法[4,5]取连续 N 个语音信号作为一帧,如对10kHz的样本频率,典型的帧尺寸为10到20ms,即 $N=100$ 至200,为了保证特征矢量系数的平滑,帧与帧之间有部分样本重叠使用,比如重叠2/3帧,但这样会导致算法复杂度增大,而且必须使用特定窗函数对数据进行处理。文中将对上述过程进行了简化改进,在语音帧分帧时不用重叠,也不用复杂的窗函数进行加窗处理,而是直接采用最基本的矩形窗对语音信号进行分段。鉴于分的段大小不能跨过语音间隙,且所包含部分若丢弃不会对语音播放效果造成影响,所以切取0.1秒语音信号大小为一帧(即一段)来进行分段。

(2)归一化处理帧(每帧中样本值的预处理):为减少样本值的离散程度,减少后续计算的复杂度,对每帧

样本值进行归一化预处理,使样本值都分布在 -1 与 1 之间。首先将所有样本点除以样本幅度最大值(若样本值为正,则除以正的最大值;若样本值为负,则除以负的最大值),则每一个样本点的值都处于[0,1]的区间之内,得以实现归一化。

(3)帧能量计算:因为本算法在分帧后进行能量分析的最小单位是帧,所以需要对一帧的样本数据求和,即以 0.1 秒为段,对数字语音信号的样本值求和。因为样本值反映了此处语音信号的振幅,即能量,所以求和是得到一帧语音信号的能量值。

(4)语音端点初步判断:根据一帧静音(没有绝对的静音,此静音是指较微弱的平稳的背景噪音)语音段能量的经验值(通过公认的经验值与实验综合得出),依次比较相邻两帧的能量。

* 若第一帧能量大于此经验值,且第二帧能量小于此经验值,则第一帧末尾(即第二帧开始)就是一个语音结束点(因为视作第二帧开始进入了静音区间)。

* 若第一帧能量小于此经验值,且第二帧能量大于此经验值,则第一帧末尾(即第二帧开始)就是一个语音开始点(因为视作第二帧开始进入了有声区间)。从而判断出语音端点(此时包括了单词端点和句子端点)。

1.2 句子语音端点判断

通过对语音间隙值的聚类分析,实现句子语音端点的判断。聚类分析是句子端点检测过程中最重要的一步。因为要面对的语音材料语速各异,用聚类分析得到具体语音材料的句子间隙门限值,能有效地提高语音端点检测法对不同语音材料的适应性。

具体过程包括语音间隙计算、语音间隙初始分类、语音平均间隙计算、聚类分析处理、聚类迭代等,进而得到句子间隙的门限,实现句子语音端点的检测和判断。处理过程如图 1 所示。

(1)得到全部语音间隙值:因为语音端点的排列是语音开始点(s_ begin)和语音结束点(p_ begin)交替出现。并以 s_ begin 开始,以 p_ begin 结束。可见用 s_ begin 减去之前相邻出现的 p_ begin 即得到所有语音间隙值。

(2)语音间隙值初始分类:首先以一个语音间隙经验值来对所有的语音间隙值进行初始分类(此经验值通过多次实验得出,实验最终采用的是 200ms)。初始分类为两类,可按间隙值大小分为句子间隙类和单词间隙类,此时这两个类仅是粗略地对语音间隙值的区分。

(3)语音平均间隙值计算:算出两个类中所有间隙值的平均值(相当于类的重心),作为下面判别间隙值属于哪个类的判别标准(每一次聚类分析过程后都要再重新计算两个平均间隙值,得到聚类后新类的判别标准)。

(4)聚类分析处理:以上述两个平均间隙值作为相应类的判别标准,依次对两个类中的所有间隙值进行判断。若某个间隙值与所在类平均间隙值的差值大于此间隙值与另一个类的平均间隙值的差值,表明该间隙值离另一类的重心更近,则将此间隙值从原来的类中移出,并放入另一个类中。相反则该间隙值仍保留在原来的类中。

(5)聚类迭代过程:反复进行上述(3)和(4)过程,直到不会再有间隙值需要从一个类移到另一个类,表明每个类中的间隙值都紧绕在所在类的重心周围,离所在类的平均间隙值更近。

(6)句子语音端点检测门限:聚类分析循环过程结束以后,可以得到用于判断句子间隙与单词间隙的门限值,利用此门限值来对之前得到的语音端点进行判断。

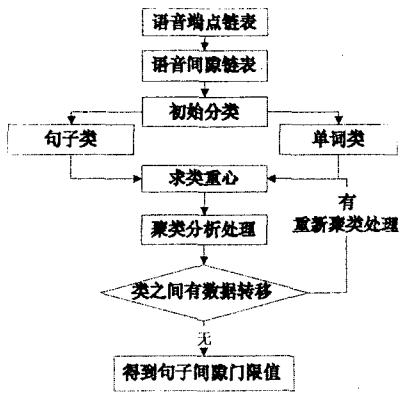

图 1 聚类分析处理流程

2 实验及结果分析

2.1 有效性测试

根据 wav 文件特点[6],基于 Direct Sound 平台编程搭建好实验平台。随机选取某英语课文中的 5 段语音数据作为实验内容,用以验证文中端点检测方法的有效性[6],其结果如表 1 所示。

从实验结果可以看出,文中的句子语音端点检测算法的检测成功率较高,是一种在较低且稳定的噪声环境下有效的句子语音端点检测方法。

表 1 英语课文句子检测结果

语音素材	A	B	C	D	E
检测句数	18	21	8	16	20
实际句数	19	22	10	18	22
检测率%	94.7	95.5	80.0	88.9	90.1

2.2 不同噪音环境下语速适应性测试

随机选择某英语课文两段语音数据，这两段语音分别由男女声朗读，环境噪音不同，语速也明显不同，以测试聚类分析针对不同语音材料的适应性，实验结果如表 2 所示。

表 2　句子端点门限值

语音素材	A	B
检测门限值	800ms	1100ms

实验结果显示，文中的方法可以对不同语速的语音，进行相应的句子端点门限检测，具有一定的自适应性，如图 2 所示，素材 A 的门限为 800ms，其中的间隙 a 为单词间隙，b 为句子间隙；如图 3 所示素材 B 的门限为 1100ms，其中的 a 为单词间隙，b 为句子间隙。

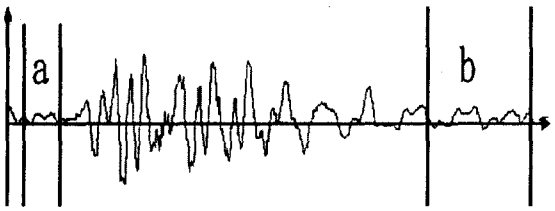

图 2　语音素材 A 的门限值

3 结束语

此检测法相比常规的语音端点检测方法，最显著的是简化了所要考虑的特征值，不再使用过零率特征值（过零数门限）等常规特征值。能量门限也只采用了单能量门限。并且针对语音复读系统等背景噪声相对较小且稳定的实际应用环境，在实现过程上进行了简化和改进，从而在语音端点检测的实现过程上更加简单方便，实验表明能针对不同语速的语音材料较准确地判断出特定的句子语音门限。对于语音复读机软件等具有较低背景噪音的环境中的句子语音端点检测是一种简单有效的检测方法。

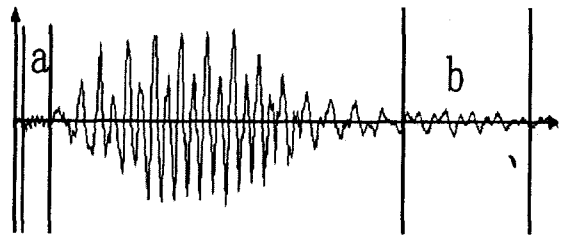

图 3　语音素材 B 的门限值

参考文献：

[1] 古井贞熙. 数字声音处理[M]. 朱家新，张国海，易武秀，译. 北京：人民邮电出版社，1993.

[2] 胡广书. 数字信号处理[M]. 北京：清华大学出版社，2003.

[3] Rabiner L, Juang Biing - Hwang. Fundamentals of Speech Recognition[M]. [s.l.]: PTR Prentice - Hall, Inc, 1993.

[4] 吴亚栋. 语音识别基础[D]. 上海：上海交通大学，1999.

[5] 李祖鹏，姚佩阳. 一种语音段起止端点检测方法[J]. 电讯技术，2000(3): 68 - 70.

[6] 张新宇. Windows 声音应用程序开发指南[M]. 西安：西安电子科技大学出版社，2003.

（上接第 12 页）

所示。从结果可以看出：在验证时，使用了偏序规约的方法，搜索了全部的空间状态，并使用了 never claim，搜索深度最大为 41，共搜索了 42 个状态，77 个转换，其中 32 重复检测，发现错误为 0 个。从而验证了考试会话模型满足期望的性质 1。

4 结　论

引入了验证 CSCW 系统静态需求的一种策略。根据 CSCW 系统的特点，使用角色访问控制描述，利用模型检测的工具 SPIN，结合具体的实例，验证了 CSCW 系统中任务流模型的具体性质。模型检测具有高度自动化，覆盖全部状态和能够生成反例的特点，把这种方法应用于验证软件设计的正确性将有更加广阔的前途。

参考文献：

[1] 郑庆华. CSCW 建模与实现方法[J]. 计算机学报，1998，21(8): 270 - 276.

[2] Joost - Pieter K. Concepts Algorithms and Tools for Model Checking[M]. [s.l.]: [s.n.], 1999.

[3] 古天龙，蔡国永. 网络协议的形式化分析与设计[M]. 北京：电子工业出版社，2003.

[4] 李成锴. 基于角色的 CSCW 系统访问控制模型[J]. 软件学报，2000，11(7): 931 - 937.

[5] Tripathi A, Ahmed T, Kumar R. Specification of Secure Distributed Collaboration Systems[C]// In IEEE International Symposium on Autonomous Distributed Systems (ISADS). Pisa, Italy: IEEE Computer Society, 2003.

[6] Tripathi A, Ahmed T, Kumar R, et al. Design of a Policy - Driven Middleware for Secure Distributed Collaboration[C]// In Proc. of International Conference on Distributed Computing Systems 2002. Vienna, Austria: IEEE Computer Society, 2002: 393 - 400.

[7] Holzmann G J. The SPIN Model Checker[M]. [s.l.] Addison - Wesley, 2003.