用于汉语语音信号端点检测与切分的有效方法*

郭 巧 张立伟 陆际联
北京理工大学机器人研究中心 (北京 100081)

摘 要 文章给出了计算机辅助汉语教学系统中语音端点信号的检测和清浊音信号的切分方法:采用短时相对能频积对汉语语音信号的端点进行检测;采用短时相对能频比的方法对语音信号的清浊音进行切分。这两种方法的使用与现有方法相比可以有效地提高汉语语音信号切分的成功率,实验结果表明正确率可达到95%以上。文中通过实验验证了所提出的汉语语音信号切分方法是有效的和可行的。它基本上能够满足计算机辅助汉语教学系统在线切分汉语语音信号的需要,比已有的语音信号切分方法的切分效果有显著提高,为下一步提高语音信号的识别率奠定了基础。

关键词 语音信号处理 语音信号切分 计算机辅助教学

An Effective Method for Capsheaf Detection and Phoneme Separation of Chinese Speech Signal

Guo Qiao Zhang Liwei Lu Jilian
(Robotics Research Center,Beijing Institute of Technology,Beijing 100081)

Abstract: In order to make the capsheaf detection of Chinese speech signal and separation of its syllable and consonant-vowel segment more efficiently,This paper defines a short-time relative EFP (Energy-Frequency-Product) and EFQ (Energy-Frequency-Quotient).It uses the short-time relative EFP to make the capsheaf detection of the Chinese speech signal, and also use the short-time relative EFQ to separate its syllable and consonant-vowel segment.The experiment results show that the correct rates of the two methods can reach 95% or more.The experiment results show that the two methods are useful and efficient,and also that it can satisfy the requirements of the on-line recognition task of the Chinese speech signal in the multimedia Chinese teaching program.

Keywords: Speech Signal Processing,Speech Signal Recognition,Computer Aided Teaching

1 引言

随着我国改革开放和对外合作的不断深化,商务往来、文化交流、来华旅游等活动日益频繁,越来越多的外国人需要学习汉语。外国人学习汉语存在的一个主要问题是发音不易准确。传统的汉语教学方法有着许多不足之处。诸如,课堂教学通常受时间、地点以及教师教学水平的限制;广播电视教学和录音录象教学等教学手段不够灵活,无法及时地分析学习者存在的问题,因此也就不能及时地、有针对性地反馈指导意见。

随着计算机技术的飞速发展和数字信号处理技术(特别是语音信号处理技术)的日臻完善,使得采用计算机进行对外汉语教学成为可能。为了构造能够在线自动评价语音学习效果的计算机辅助汉语多媒体语音教学系统,需要有好的语音识别系统。而好的语音识别系统又必须建立在正确地进行语音信号端点检测与切分的基础上。

大家知道,汉语语音识别中并不是所有的语音信号都是有用信息。通过对语音信号的波形分析可以看到,语音信号中开始很长一段属于无声段,每段语音的最后一段也是无声段。如果在语音识别时把这两部分包括在内,会增加很大的工作量同时还增加了语音识别的难度。作为语音识别的基础,语音信号端点的检测不但是必须的,而且是语音识别的关键之一。

在语音识别中,一段语音信号的端点检测完毕后,需要对这一段语音信号作进一步的切分处理,也就是要对它进行音节及其声韵母的切分。切分后的语音信号才能被识别。可以说没有正确的语音切分就没有正确的语音识别。

2 语音信号的分类及其各类的特点

语音信号一般可分为无声段、清音段和浊音段。无声段的平均能量最低,浊音段的平均能量最高,清音段的平均能量居于两者之间。在噪声较低的环境下,清音段的平均能量一般比无声段的能量高出几倍到几十倍,而浊音段的平均能量则能高出几十倍至上百倍,应用平均能量基本上能粗略地将它们分开。

对于语音信号的这三部分来说,另一个同等重要的特征参数是它们的过零率。清音段的过零率大多数情况下最高;无声段的过零率变化范围较大,一般情况下比浊音段低一点,但有时会比浊音段稍高一些或者差不多。

*该项目由国家自然科学基金资助。

作者简介:郭巧,博士,教授。主要研究领域为智能信息处理、网络技术、智能控制理论及其应用等。张立伟,硕士。主要研究领域为语音信号处理。陆际联,教授。主要研究领域为智能系统及其运动控制等。

3 特征参数的定义和提取

语音信号通过麦克风输入到声卡，声卡通过一定的采样频率把连续的语音信号变成数字信号，这些数字信号的频率和精度可以根据需要设定。这里，采用采样频率 16KHz、精度 16bits。通过声卡采样后得到的数字信号用 s(n)表示。

由于录音和发声的间隔，正常情况下语音信号的前 100ms 是无声段，所以可以提取这段语音信号的平均能量、平均过零率、它们的乘积(称为能频积)和它们之比(称为能频比)作为进行粗略判断时的特征参数。又由于某些声母发声短促，用振幅的平方表示能量时数值过大，因此，在切分和端点检测过程中，采用方窗，窗的长度为 5ms，用振幅的绝对值表示能量 X(i)，加窗后的语音信号为 S_w。具体的实现过程如下：

$X(i)=|S_w(i)|=|S(K^*I+i)|$，$i=0\sim(I-1)$；

$I=16000*0.005=80$，$k=0\sim14$

$E_n(k)=\sum_{i=0}^{I-1}X(i)$，$E_n(k)$表示第(k+1)帧语音信号的能量；

$SE_n=\sum_{k=0}^{14}En(k)$，$SE_n$ 表示前 15 帧语音信号的总能量；

$EN=SE_n/15$，EN 表示前 15 帧语音信号的平均能量；

$$Z_r(k)=\frac{1}{2}\left(\sum_{i=0}^{I-1}|sgn(X(i))-sgn(X(i-1))|\right),$$

$$sgn(x)=\begin{cases}1 & x\geq 0\\ -1 & x<0\end{cases}$$

$Z_r(k)$表示第(k+1)帧语音信号的过零率；

$SZ_r=\sum_{k=0}^{14}Z_r(k)$，$SZ_r$ 表示前 15 帧语音信号的总过零率；

$ZR=SZ_r/15$，ZR 表示前 15 帧语音信号的平均过零率；

$A_r=E_n(k)*Z_r(k)$，A_r 表示一帧语音信号能量与其过零率的的乘积，即短时能频积；

AR=EN＊ZR，AR 表示平均能量与平均过零率的乘积，即平均短时能频积；

$B_r=E_n(k)/Z_r(k)$，B_r 表示一帧语音信号总能量与其过零率的比值，称为短时能频比；

BR=EN/ZR，BR 表示平均能量与平均过零率的比值，称为平均短时能频比。

根据以上步骤所得到的特征参数，用各自合理的系数做一个加权运算后得到四个相对值。用这些相对值作为端点检测和切分粗判的阈值。

4 初步的端点检测和音素切分

这里采用以下步骤来初步进行端点检测和音素切分：

(1) 设定合理阈值

判断语音开始：帧能量阈值 1.5＊EN、过零率阈值 2＊Z_r、能频积阈值 2＊AR；

判断语音信号中清音和浊音切分的的阈值分别为：能量阈值 5.6＊EN、过零率阈值 0.5＊ZR≤Z_r≤3.5＊ZR、能频比阈值 B_r≥5＊BR；

判断语音结束：由于一段语音信号结束时末尾经常带有比较大的噪声或者有比较长的拖音，所以应用上述的特征参数作为阈值可能会造成一些错误。鉴于此原因，可利用结尾后十帧(在正常的语音信号中这部分肯定是无声段)来提取特征参数设定阈值(具体计算与上述相同)。三个特征参数(能量、过零率和能频积)的阈值系数分别设为 2、2、2。

(2) 粗判语音开头

若连续三帧的能量、过零率或者能频积大于自己相对应的阈值，则粗判该帧为语音的开头，转(3)；否则重复(2)。

(3) 粗判清音和浊音的切分点

连续采集十帧，若连续十帧的能量、过零率和能频比都超过各自所设定的阈值范围，则粗判此帧为语音信号中该音节浊音的开始，转(4)；否则重复(3)。

(4) 粗判语音音节结尾

若连续五帧的能量、过零率和能频积均小于所设定的阈值，则粗判此帧为这个音节的结尾，转(5)；否则重复(4)。

(5) 粗判下一个音节的开头

重复(2)，判断是否已经连续取了 30 帧。若小于 30 帧，且满足转(3)的条件，继续进行，否则转(6)。

(6) 粗判结束。

5 精确的端点检测和切分

由于声韵母发声时的不稳定性和连续语音有音节连读的现象，使得粗判的结果不太精确，有时甚至不能将音节切分开。因此，通过用粗判得到的第一个音节作为相对样本，对它取帧能量和过零率的平均值，分别用 Energy 和 Zero 表示。令能量阈值为 Energy 乘以某一系数，同时能频积和能频比的阈值也做相应的变化。经过多次实验证明，判断语音信号的开头和结尾可以采用粗判的结果，它与进一步精确判断的结果是一致的。在判断清音和浊音的分界点、音节以及音节间的结尾和开头时，需要进一步设定阈值。具体的阈值设定为：判断清音和浊音的分界点时，能量阈值为 0.5＊Energy，能频比阈值为 0.5，过零率与上面介绍的相同；判断音节与音节之间的开头和结尾时，能量阈值分别为 0.15＊Energy 和 0.2＊Energy，过零率阈值不变，能频积阈值为 0.3＊Zero＊Energy (判断开始)，能频比阈值系数为 0.2(判断结尾)。

6 实验结果与结论

在已有的文献中，处理汉语语音信号的端点检测和清浊音切分问题通常采用固定特征值方法。这里采用相对特征值进行端点检测和清浊音切分处理。实验结果表明，该方法对于单音节、说话速度比较稳定的多音节的头尾和清浊音节的判别效果相当好。表 1 和表 2 分别给出了针对单音节和连续语音信号的两种方法实验统计结果。图 1 和图 2 给出了两个时域语音波形切分效果的例子。这里，统计用实验样本大部分采用中科院自动化所提供的裸文件语音数据库。

由实验结果图和表中统计结果可知，该实验系统所采用的汉语语音切分方法可以满足多媒体汉语教学系统实验平台的任务需要，其切分的正确率与已有方法相比有显著提高。

(收稿日期：1999 年 5 月)

表 1　两种方法对于单音节端点检测与清浊音切分正确率的统计结果

声母类型 \ 统计类别			采用固定的特征值		采用相对的特征值	
			端点检测	清浊音切分	端点检测	清浊音切分
清音	塞爆音	b,d,g	98.5	85.6	100	98.5
		p,t,k	98.8	82.5	100	99
	塞擦音	Z,zhj	98.5	85.8	100	99.2
		C,chq	99	81.8	100	99.6
	擦音	f,s,sh,hx	99	88.6	100	99.2
浊音	m,n r,l		99.1	(能量比其后的韵母要低很多)	100	

表 2　两种方法对于连续语音信号的端点检测、音节切分以及清浊音的切分正确率的统计结果

声母类型 \ 统计类别			采用固定的特征值			采用相对的特征值		
			端点检测	音节切分	清、浊音切分	端点检测	音节切分	清、浊音切分
清音	塞爆音	b,d,g	98.8	90.5	84.5	100	99.5	98.8
		p,t k	98.5	99	84.2	100	99.6	99
	塞擦音	z,zhj	98.6	90.6	85.5	100	99.5	99.6
		c,chq	99.2	92.5	82.2	100	99.7	99.5
	擦音	f,s,sh,hx	98.6	90.6	85.6	100	99.6	99.8
浊音	m,n r,l		99.5	92.5	(能量比较低)	100	99.6	

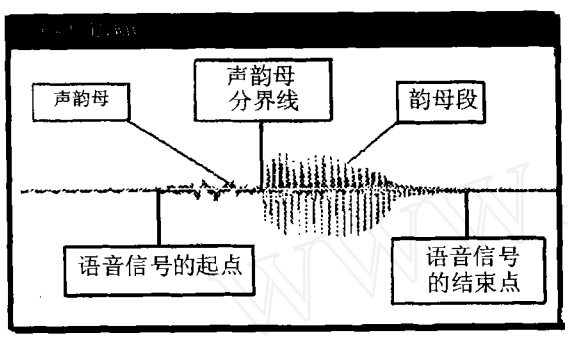

图 1　单音节“前”的端点检测和切分

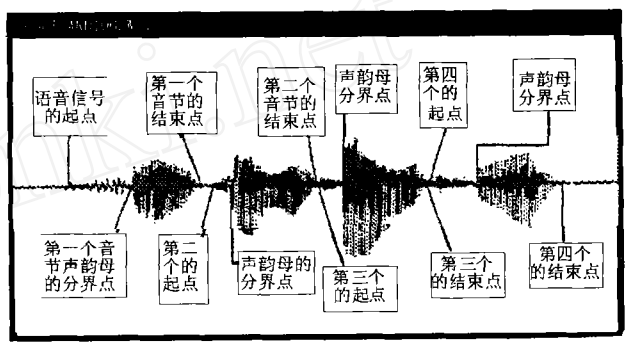

图 2　短语“似曾相识”的端点检测和切分

参考文献

1.杨行峻,迟惠生.语音信号数字处理.电子工业出版社,1995

2.陈永彬.语音信号处理.上海交通大学出版社,1991

3.黄昌宁,夏莹.语言信息处理专论.清华大学出版社,1996

4.朱民雄.计算机语音技术.北京航空航天大学出版社,1992

5.Ma Bin,Huang Taiyi,Xu Bo,et al.Context-dependent Acoustic Models for Chinese Speech Recognition.IEEE International Conference on Acoustics,Speech and Signal Processing,1996:455-458

北大青鸟发布 GIS 新产品

北大青鸟通用地理信息系统开发平台产品发布会于 2000 年 4 月 18 日在北京友谊宾馆举行。此次发布的青鸟 Geo-Union Enterprise Server 5.0 是支持 Web 网络应用的企业级地理信息系统开发环境。它是国内首个采用构件构架技术开发完全基于商业数据库存储空间数据的系统,支持异构分布式环境,具有动态可伸缩结构、多层索引和缓存结构。

Geo-Union 系列通用地理信息系统基础软件已经具有 14 年的历史，此次发布最新版本 5.0 版是基于以往多年的开发成果和青鸟 JBCASE 软件工程环境,结合最新网络和软件技术,面向 Web 应用开发,面向企业级大型专业 GIS 应用,可以作为各种信息基础设施的基础平台。

Geo-Union Enterprise Server 5.0 除了具有完整的 GIS 空间数据管理和空间分析的基础功能外，还有专门设计的网络加速器和开放数据结构,具有杰出的网络运行性能,支持用户自定义数据结构,支持三维数据结构扩展,支持 GPS 移动目标,可以广泛应用于资源、环境、国防、公安、城市公共设施管理、交通、测绘、电力、电信等与地理信息有关的系统。Geo-Union Enterprise Server 5.0 的推出,使得国内有了可以替代、以至超过国外同类产品的自主版权大型 GIS 应用系统开发环境。