

基于 Matlab 的语音端点检测方法浅析

邢亚从

(苏州市职业大学 江苏 苏州 215000)

【摘 要】: 语音的端点检测在语音的编码、语音识别、语音增强、说话人识别中起着非常重要的作用,直接影响着后续工作的正确率。本文介绍了端点检测的基本方法:基于短时能量的、基于短时过零率的、双门限检测的方法,并通过 Matlab 仿真对双门限检测的方法进行分析。

【关键词】: 语音端点检测、过零率、Matlab

端点检测是语音识别中非常重要的一步。所谓语音端点检测,就是从一段给定的语音信号中找出语音的起始点和结束点。在语音识别系统中,正确、有效的进行端点检测不仅可以减少计算量和缩短处理时间,而且能排除无声段的噪声干扰、提高语音识别的正确率。研究表明[1],即使是在安静的环境下,语音识别系统一半以上的错误可能主要来源于端点检测。除此之外,在语音合成、编码等系统中,高效的端点检测也直接影响甚至决定着系统的主要性能。因此,端点检测的效率、质量在语音处理系统中显得至关重要,广泛开展端点检测实现手段方面的研究,有一定的现实意义[2]。

1、短时能量法

信号 $\{x(n)\}$ 的短时能量定义为:

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

令 $h(n)=w^2(n)$, 则有:

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

其中 $h(n)$ 是低通滤波器的单位冲激响应。

语音信号的短时平均幅度定义为:

$$M_n = \sum_{m=-\infty}^{\infty} |x(m)| \cdot w(n-m) \quad (3)$$

E_n 和 M_n 都反映信号强度,但其特性有所不同。

语音和噪声的区别可以体现在它们的能量上,语音段的能量比噪声段能量大,语音段的能量是噪声段能量叠加语音声波能量的和。在信噪比很高时,如果环境噪声和系统输入噪声比较小,以至能够保证系统的信噪比相当高,那么只要计算输入信号的短时能量或短时平均幅度就能够把语音段和噪声背景区分开。这是仅基于短时能量的端点检测方法。但低信噪比情形下,此算法就将失效[3-4]。

2、短时平均过零率法

信号 $\{x(n)\}$ 的短时平均过零率定义为[5]:

$$Z_n = \frac{1}{2} \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| \cdot w(n-m) \quad (4)$$

一般取

$$w(n) = \begin{cases} \frac{1}{2N} & 0 \leq n \leq N-1 \\ 0 & \text{其它} \end{cases} \quad (5)$$

信号的过零率是其频率量的一种简单度量,窄带信号尤其如此,其中,当信号为单一正弦波时,过零率为信号频率的两倍。对于采样率为 F_s , 频率为 F_0 的正弦波数字信号,平均每个样本的过零率为 $2F_0/F_s$ 。

过零率有两类重要的应用:第一,用于粗略地描述信号的频谱特性,就是用多带滤波器将信号分为若干个通道,对各通道进行短时平均过零率和短时能量的计算,即可粗略地估计频谱特性。第二,用于判别清音和浊音、有语和无语。从上面提到的定义出发计算过零率容易受低频干扰,特别是 50Hz 交流干扰的影响。解决这个问题的办法,一个是做高通滤波器或带通滤波,减

小随机噪声的影响;另一个有效方法是对上述定义做一点修改,设一个门限 T ,将过零率的含义修改为跨过正负门限,如图 1 所示。

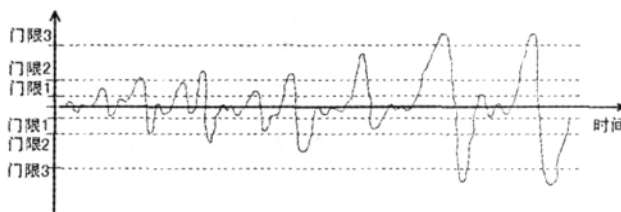


图 1 门限过零率

于是有定义:

$$Z_n = \frac{1}{2} \sum_{m=-\infty}^{\infty} \{ |\text{sgn}[x(m)-T] - \text{sgn}[x(m-1)-T]| + |\text{sgn}[x(m)+T] - \text{sgn}[x(m-1)+T]| \} \cdot w(n-m) \quad (6)$$

这样计算的过零率就有一定的抗干扰能力了。即使存在小的随机噪声,只要它不使信号越过正、负门限所构成的带,就不会产生虚假的过零率。在语音识别前端检测时还可采用多门限过零率,进一步改善检测效果。

3、双门限检测法

语音端点检测方法可采用测试信号的短时能量或短时对数能量、联合过零率等特征参数,并采用双门限判定法来检测语音端点,即利用过零率检测清音,用短时能量检测浊音,两者配合。首先为短时能量和过零率分别确定两个门限,一个是较低的门限值数值较小,对信号的变化比较敏感,很容易超过;另一个是较高的门限,数值较大。低门限被超过未必是语音的开始,有可能是很短的噪声引起的,高门限被超过并且接下来的自定义时间段内的语音超过低门限,意味着信号开始[5]。

算法过程如下:

(1)、在开始阶段要做预加重和分帧的处理,讲语音信号分成一帧一帧的,分帧处理有利于对语音信号进行准确的分析,并且能够提高识别率,这时再分别求出每帧的短时能量和短时过零率。

(2)、接着要设置初始化参数,比如最大静音长度,这是一个经验值,用来判断语音段是否结束,论文中是根据大量的语音样本的长短设置的一个经验值。另外,短时能量和短时过零率的门限也要设置初始值等。

(3)、判断当语音在静音段或者是过渡段时,如果语音信号的短时能量值大于短时能量的高门限,或者语音信号的短时过零率大于短时过零率的高门限,那么就确认进入了语音段,如果短时能量的值大于短时能量的低门限或者过零率的值大于过零率的低门限,那么语音处于过渡段,否则,语音仍就处于静音段。

(4)、当语音信号在语音段时,判断如果短时能量的值大于短时能量的低门限或者短时过零率的值大于短时过零率的低门限,那么语音信号仍然处于语音段。

(5)、如果静音长度小于设置的最大静音长度,那么就表明语

音还尚未结束，还在语音段，如果语音的长度小于最小噪声长度，那么认为语音太短，此时是噪声，同时判断语音处于静音段；否则语音就进入结束段。

4、仿真结果

本文对双门限检测法进行编程仿真，结果如图 2 和表 1~3。

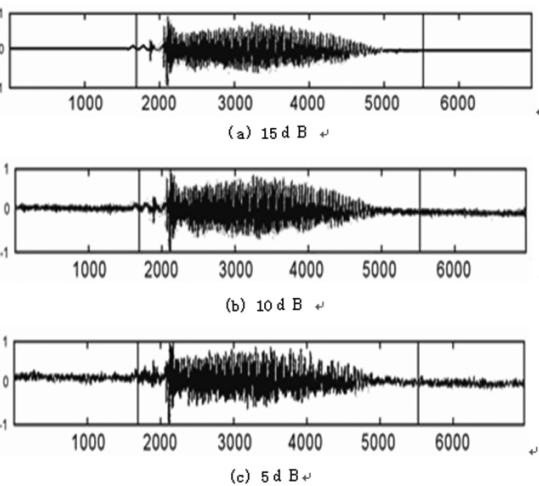


图 2 双门限检测方法检测结果

SNR	孤立词库		连续词库	
	start	end	start	end
30dB	100	100	94	95
15dB	81	83	74	73
10dB	73	74	57	56

表 1 白噪声下的双门限检测方法的检测准确率(%)

SNR	孤立词库		连续词库	
	start	end	start	end
30dB	97	95	94	95
15dB	77	76	74	73
10dB	70	69	57	56

表 2 工厂噪声下的双门限检测方法的检测准确率(%)

SNR	孤立词库		连续词库	
	start	end	start	end
30dB	98	97	94	95
15dB	74	73	74	73
10dB	63	65	57	56

表 3 办公室噪声下的双门限检测方法的检测准确率(%)

从图 2 和表 1~3 中多样本的实验可以看出，这种方法在背景噪声幅度保持恒定、且远低于语音信号幅度时，并且对孤立字的最小帧数、最大帧数、句子间间隙的最小帧数，以及人为的突发性音节帧数有充分先验知识的条件下，可以十分有效地准确检测出语音信号的端点。尤其在用过零率方法作辅助处理来调整检测后的端点时，测出的端点位置是比较准确的。从表 1~3 中可以明显看出，该方法检测孤立词端点的准确性远远高于连续语音；在高信噪比的情况，对孤立词检测比较准确。然而，当该类算法在信噪比较低的情况下，检测性能开始恶化。在更恶劣的情况下，信噪比小于 5dB 的，甚至完全不能检测出其端点。而且该方法在背景噪是工厂噪声、办公室噪声或白噪声时，噪声的过零率均不相同，所以很难通过经验值得到合适的门限，有时与语音某些音节的过零率相重叠，也很难作为一种辅助的判据。

5、总结

本文介绍了端点检测的基本方法：基于短时能量的、基于短时过零率的、双门限检测的方法，并通过 Matlab 仿真对双门限检测的方法进行分析。通过分析，双门限结合了短时能量和短时过零率的优点，其精确度和误检率都有了进一步的改善。但是，其中还是有很多不足的地方：双门限的误检率还是很高，因为在时域范围内数据的很多特征是无法准确决定声音的端点的。而且双门限方法不具备抗噪能力。

参考文献：

1.Junqua J C. Robustness and Cooperative Multimodel Man-machine Communication Applications [M]. Proc. Second Venaco Workshop and ESCA ETRW. 1991. 9.
2. 张震宇. 基于 Matlab 的语音端点检测实验研究[J]. 浙江科技学院学报, 2007, 19(3):197-201
3. 易克初, 田斌, 付强. 语音信号处理(speech Singal Proceessing)[M], 北京: 国防工业出版社, 2000
4.K.Yao,E.shi,ect.,Residual noise compensation for robust speech recognition nonstationary noise[C]. Proc.ICASS,2000,(2):1125-1128
5.Gevren S,Xie Fei.A comparative study of speech detection methods[C], EijROSPEECH,Gereee,1015-1020,1997

(上接第 50 页)

习学生保持联系，了解监控实习过程；3. 组织未能进入企业的学生，利用企业提供的工程案例进行虚拟项目的实习；4. 指派教师指导实习；5. 收集整理实习资料，建立相应的资料库、项目库、开发规范文档库等。

4. 结束语

建立教学实习基地是高等学校的常规工作，因为它在人才培养中的地位越来越重要。目前，许多城市大力发展以信息产业为核心现代服务外包业，各级政府出台了相应的软件人才培养的具体政策，鼓励校企合作培养软件人才，为高等学校搞好教学实习基地建设提供良好机遇，学校学院对实习基地建设非常重视。如何给学生提供实践机会，让他们尽早接触到企业，实实在在地参与到实际项目的开发中去，缩短学校教学与社会需求之

间距离，需要我们不断研究解决。建立与企业紧密结合的校内外实习基地是其中一种行之有效的方法。

参考文献：

1. 中国计算机科学与技术学科教程 2000 研究组. 中国计算机科学与技术学科教程[M]北京:清华大学出版社,2002.
2. 教育部软件工程学科课程体系研究课题组. 中国软件工程学科教程[M]北京:清华大学出版社,2005.
3. 丁宇辰. 计算机专业实践教学体系探索 [J]. 南京工程学院学报(社会科学版),2005,4: 46-49
4. 单志龙. 浅谈计算机专业实践教学体系的改革 [J]. 科教文汇,2008,8: 81-82