

基于双门限两级判决的语音端点检测方法

路青起, 白燕燕

(西安工业大学 电子信息工程学院, 陕西 西安 710032)

摘 要 端点检测是语音信号处理过程中的重要步骤, 其准确性直接影响语音信号处理的速度和结果。因此对于端点检测方法, 特别是在噪声环境下的端点检测研究, 一直是语音信号处理中的热点。文中针对声纹识别系统所作的端点检测前端处理, 对比了利用短时能量和短时平均过零率进行端点检测的方法, 运用 Matlab 实现了双门限法端点检测的编程和仿真。仿真结果表明, 端点检测准确时识别率为 93%。

关键词 端点检测; 语音信号处理; 短时能量; 过零率; 双门限法

中图分类号 TN912; TP312 **文献标识码** A **文章编号** 1007-7820(2012)01-013-04

A Speech Endpoint Detection Algorithm Based on Dual-threshold Two Sentences

LU Qingqi, BAI Yanyan

(Institute of Electronic and Information Engineering, Xi'an Technological University, Xi'an 710032, China)

Abstract Endpoint detection of speech signal processing is a very important step, which directly affects the accuracy of the speed and results in voice signal processing. The research on endpoint detection, especially in noisy environments, has always been the hot spots in voice signal processing. Aiming at the front-end detection, which is made for the voice recognition system, this paper compares the use of short-term energy with the use of short-term average zero-crossing rate for endpoint detection. It achieves a double-threshold method of endpoint detection programming and simulation by Matlab. Simulation results show that an accuracy of as high as 93% for endpoint detection can be achieved.

Keywords endpoint detection; speech signal processing; short-term energy; zero crossing rate; dual-threshold

端点检测是语音信号处理中的一个基本问题, 其目的是从包含语音的一段信号中确定出语音的起点及结束点。有效的端点检测不仅能使处理时间减到最少, 而且能抑制无声段的噪声干扰, 提高语音信号质量。有学者用一个多话者的数字语音识别系统做实验。首先对所有记录的语音用手工找出准确的端点, 得到辨识率; 然后逐帧加大端点检测的误差, 在每次加大误差的同时得到识别率。结果表明, 端点检测准确时识别率为 93% 的系统, 当端点检测误差在 ± 60 ms 时, 识别率降低了 3%; 在 ± 90 ms 时, 降低了 10%; 而当误差进一步加大时, 识别率急剧下降^[1]。语音端点检测的成败, 直接影响着后续工作的正确率乃至整个语音识别系统的成败。

端点检测的常用方法有: 能量阈值、基音检测、频谱分析、倒谱分析及 LPC (Linear Prediction Coefficients) 预测残差等。其中基于能量和过零率的双门限判决法最为常用。在许多语音信号处理任务中需要判断, 一段输入信号中哪些是语音段, 哪些是无声段, 哪

些是噪声, 哪些是语音。在一些语音识别或低速语音编解码器应用中, 对于已经判别为语音段的部分, 还需进一步判断清音和浊音。这些问题可以称为有声或无声判决, 以及更细致的无声 (S)、清音 (U)、浊音 (V) 判决^[2]。能够实现这些判决的依据在于, 不同性质语音的各种短时参数具有不同的概率密度函数, 以及相邻的若干帧语音具有一致的语音特性, 不会在无声、清音和浊音之间随意跳变。文章比较了利用短时能量和短时平均过零率进行端点检测的方法, 并运用双门限两级判决法在 Matlab 中。

1 语音信号的短时能量分析

语音信号的能量随时间变化比较明显, 一般清音部分的能量比浊音部分的能量小。语音实现了端点检测的编程和仿真。

信号的短时能量分析给出了反应这些幅度变化的一个合适的描述方法。对于信号 $x(n)$, 短时能量的定义如下

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

式 (1) 中 $h(n) = \omega^2(n)$, E_n 表示在信号的第 n 个点开

收稿日期: 2011-08-20

作者简介: 路青起 (1955—), 男, 副教授。研究方向: 信号与信息处理, 语音信号处理。白燕燕 (1983—), 女, 硕士研究生。研究方向: 信号与信息处理, 语音信号处理。

始加窗函数时的短时能量。可以看出,短时能量可以看作语音信号的平方经过一个线性滤波器的输出,该线性滤波器的单位冲击响应为 $h(n)$,如图1所示。

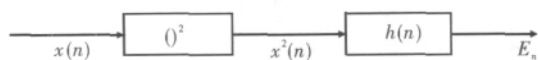


图1 语音信号的短时平均能量实现框图

冲激响应 $h(n)$ 的选择,或者说窗函数的选择决定了短时能量表示方法的特点。为反应窗函数选择对短时能量的影响,假设式中的 $h(n)$ 非常长,且为恒定幅度,那么 E_n 随时间的变化将很小,这样的窗就等效为很窄的低通滤波器。 $h(n)$ 对 $x^2(n)$ 的平滑作用较显著,无法反映语音的时变特性。反之,若 $h(n)$ 序列长度 N 过小,那么等效窗不能提供足够的平滑,以至于语音振幅瞬时变化的许多细节仍被保留,从而看不出振幅包络的变化规律。通常 N 的选择与语音的基音周期相联系,一般要求窗长为几个基音周期的数量级。

短时能量主要用于区分清音和浊音,因为浊音比清音的能量大得多;其次可以用短时能量对有声段和无声段进行判定,对声母和韵母分界,以及对连字分界等。在语音识别系统中,短时能量一般也作为特征中的一维参数来表示语音信号能量的大小和超音段信息^[1]。

短时能量的一个主要问题是 E_n 对信号电平值过于敏感。由于需要对信号进行平方运算,在定点实现时容易产生溢出。解决问题的方法是采用短时平均幅度来表示能量的变化,其公式是

$$M_n = \sum_{m=-\infty}^{\infty} |x(m)| \omega(n-m) = \sum_{m=n}^{n+N-1} |x_{\omega(m)}| \quad (2)$$

与短时能量相比,短时平均幅度相当于用绝对值之和代替了平方和,简化了运算,窗长 N 对平均幅度函数的影响与短时能量的分析结论是一致的。但平均幅度函数没有平方运算,因此其动态范围比短时能量小,接近于标准能量计算的动态范围的平方根。图2是“西安工业大学”的波形图。

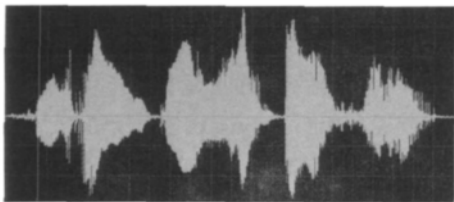


图2 “西安工业大学”波形图

图3是语音“西安工业大学”的短时能量分析图,图4是“西安工业大学”的短时平均幅度的分析图。

从图3和图4中可以看到,尽管短时平均幅度也可用于区分清音和浊音,无声和有声,但二者之间的幅度差就不如短时能量那么明显。

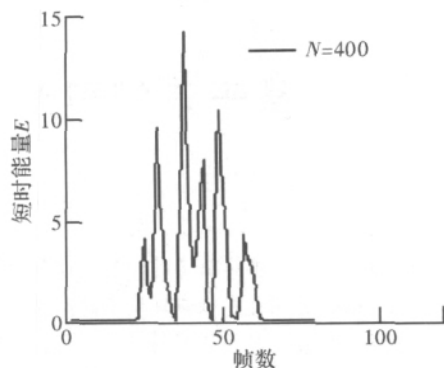


图3 短时平均能量

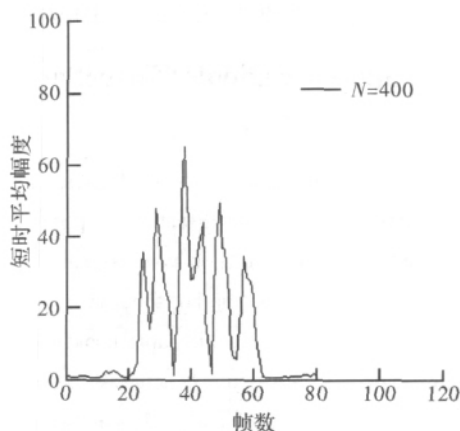


图4 短时平均幅度

2 短时平均过零率

短时平均过零率是语音信号时域分析中的一种特征参数。它是指每帧内信号通过零值的次数。对于连续语音信号,可以考察其实域波形通过时间轴的情况。对于离散信号,短时平均过零率实际就是信号采样点符号变化的次数。短时平均过零率仍可以在一定程度上反映其频谱性质,可以通过短时平均过零率获得谱特性的一种粗略估计。短时平均过零率的公式为

$$Z_n = \sum_{m=-\infty}^{\infty} |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]| \omega(n-m) = |\operatorname{sgn}[x(n)] - \operatorname{sgn}[x(n-1)]| * \omega(n) \quad (3)$$

其中, sgn 是符号函数即

$$\operatorname{sgn}[x(n)] = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases} \quad (4)$$

图5给出了短时平均过零率的计算过程。



图5 短时平均过零率的计算

首先对语音信号序列 $x(n)$ 进行成对处理,检查是否有过零现象,若有符号变化,则表示有一次过零现象;然后进行一阶差分计算,取绝对值;最后进行低通滤波。

短时平均过零率可以在一定程度上反映频率的高低,因此在浊音段,一般具有较低的过零率,而在清音段具有较高的过零率,可用于初步判断清浊音。

在矩形窗条件下,式可以简化为

$$Z_n = \sum_{m=n}^{n+N+1} |\operatorname{sgn}[x_\omega(m)] - \operatorname{sgn}[x_\omega(m-1)]| \quad (5)$$

图6和图7均是“西安工业大学”语音波形图与短时平均过零率。

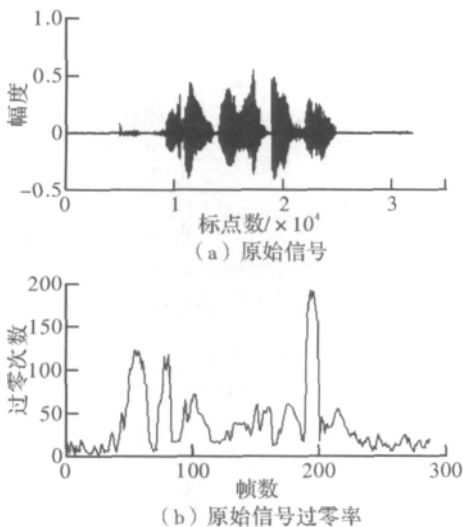


图6 “西安工业大学”语音波形图与短时平均过零率

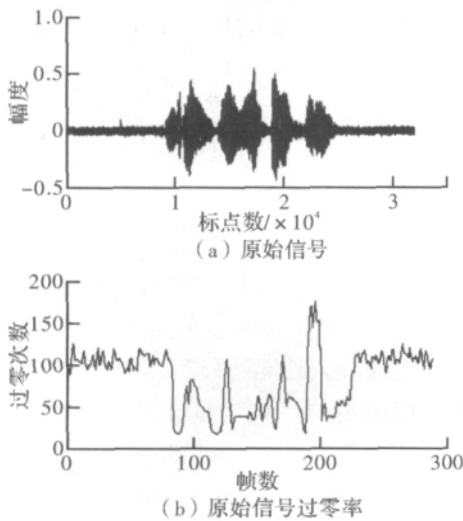


图7 加入噪声的“西安工业大学”语音波形图与短时平均过零率

从图中可以看出,在背景噪声较小的情况下,短时能量较准确,但当背景噪声较大时,短时平均过零率可以获得较好的检测效果。有些发音仅用短时能量或短时平均过零率来判断起点和终点比较困难,由于短时平均幅度更适合检测浊音,而短时过零率更适合检测清音,因此将短时平均幅度和短时平均过零率相结合进行端点检测是一种有效的端点检测方法。

3 双门限判决法

在进行端点检测前,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,其数值较小,对信号的变化较敏感,很容易被超过。另一个是较高的门限,数值较大,信号必须达到一定的强度,该门限才可能被超过。

基于Matlab程序实现能量与过零率的端点检测算法步骤如下:

(1) 语音信号 $x(n)$ 进行分帧处理,每一帧记为, $n=1, 2, \dots, N$, n 为离散语音信号时间序列, N 为帧长, i 表示帧数。

(2) 计算每一帧语音的短时能量,得到语音的短时帧能量

$$E_i = \sum_{n=1}^N s_i^2(n) \quad (6)$$

(3) 计算每一帧语音的过零率,得到短时过零率

$$Z_i = \sum_{n=1}^N |\operatorname{sgn}[s_i(n)] - \operatorname{sgn}[s_i(n-1)]| \quad (7)$$

其中

$$\operatorname{sgn}[s_i(n)] = \begin{cases} 1 & s_i(n) \geq 0 \\ 0 & s_i(n) < 0 \end{cases} \quad (8)$$

(4) 考察语音的平均能量,设置一个较高的门限 T_1 ,用以确定语音开始,然后再根据背景噪声的平均能量确定一个较低的门限 T_2 ,用以确定第一级中的语音结束点, $T_2 = \alpha_1 E_N$, E_N 为噪声段能量的平均值,完成第一级判决。第二级判决同样根据背景噪声的平均过零率 Z_N ,设置一个门限 T_3 ,用于判断语音前端的清音和后端的尾音。 α_1, β_1 为经过大量实验得到的经验值。图8是语音“西安工业大学”的端点检测图。

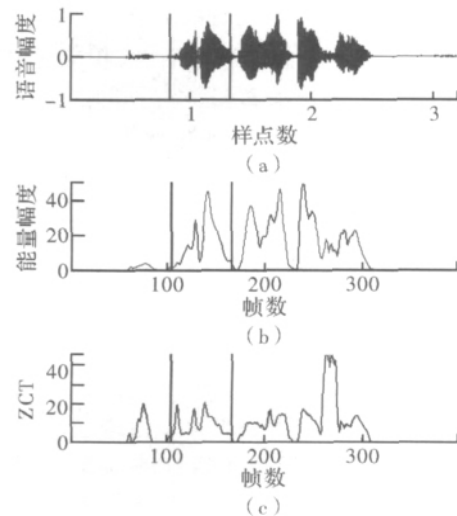


图8 语音“西安工业大学”的端点检测

(下转第19页)

参考文献

- [1] 赵国庆. 雷达对抗原理[M]. 西安: 西安电子科技大学出版社 2003.
- [2] Wiley R G. 电子情报—雷达信号截获与分析[M]. 吕跃广, 译. 北京: 电子工业出版社 2008.
- [3] 李忠良. 雷达侦察设备测量信号极化的方法[J]. 上海航天, 1997(6): 22-28.
- [4] 王文英. 对极化分集雷达的对抗技术研究[D]. 西安: 西安电子科技大学 2007.
- [5] 牛中奇, 朱满座, 卢智远, 等. 电磁场理论基础[M]. 西安: 西安电子科技大学出版社 2003.
- [6] 曾清平. 雷达极化技术与极化信息应用[M]. 北京: 国防工业出版社 2006.
- [7] 吴大正. 信号与线性系统分析[M]. 4 版. 北京: 高等教育出版社 2000.
- [8] 丁鹭飞, 耿富录. 雷达原理[M]. 西安: 西安电子科技大学出版社 2000.
- [9] 庄钊文. 电磁波极化状态的表征[J]. 北京: 国防工业出版社, 1999.
- [10] 李明译. 双极化曲折天线[J]. 电子侦察干扰, 1991.
- [11] 李晋文. 曲折天线分析与设计[D]. 长沙: 国防科技大学, 1999.
- [12] 庄钊文, 肖顺平, 王雪松. 雷达极化信息处理及其应用[M]. 北京: 国防工业出版社, 1999.
- [13] 肖顺平. 宽带极化雷达目标识别的理论与应用[D]. 长沙: 国防科技大学, 1995.
- [14] 杨少华, 都佰胜. 雷达信号处理干扰仿真平台[J]. 电子科技 2011 24(1): 75-77.

(上接第 15 页)

图 8 可以看到检测结果: 将短时平均能量和短时平均过零率结合, 进行端点检测, 可以很好地检测语音的开始和结束。虽然这种方法可以有效地区分语音信号中的清音和噪声, 但它难区分语音信号中的浊音和噪声, 也存在检测时漏掉某些音素这一问题。因此对于端点检测仍需进一步研究。

4 结束语

针对声纹识别系统所作的端点检测前端处理, 运用 Matlab 实现了双门限法端点检测的编程和仿真, 仿真结果显示, 端点检测准确识别率为 93%。

参考文献

- [1] 赵力. 语音信号处理[M]. 北京: 机械工业出版社 2010.
- [2] 韩纪庆, 张磊, 郑铁然. 语音信号处理[M]. 北京: 清华大学出版社 2004.
- [3] 何强, 何英. Matlab 扩展编程[M]. 北京: 清华大学出版社 2002.
- [4] 张雪英. 数字信号处理及 Matlab 仿真[M]. 北京: 电子工业出版社 2010.
- [5] 尹巧萍, 吴海宁, 赵力. 含噪语音信号端点检测方法的研究[C]. 成都: 2008 促进中西部发展声学学术交流会议论文集 2008.
- [6] 刘建, 郑方, 吴文虎. 基于幅度差平方和函数的基音周期提取算法[J]. 清华大学学报: 自然科学版, 2006 46(1): 74-77.
- [7] 张仁志, 崔慧娟. 基于短时能量的语音端点检测算法研究[J]. 电声技术, 2005(7): 52-59.
- [8] 陈艳. 改进的独立分量分析在语音分离中的应用[J]. 电子科技 2009 22(12): 87-91.

欢迎刊登广告

请访问: www.dianzikeji.org

E-mail: dzkj@mail.xidian.edu.cn

联系电话: 029-88202440

传真: 029-88202440