基于短时能量和过零率分析的语音端点检测方法研究

刘波，聂明新，向俊涛
武汉理工大学信息工程学院，湖北武汉 (430070)
E-mail：ngulb@126.com

摘　要：短时能量分析和过零率分析作为语音信号时域分析中最基本的方法，应用相当广泛，特别是在语音信号端点检测方面。由于在语音信号端点检测方面这两种方法通常是独立使用的，在端点检测的时候很容易漏掉重要的信息。本文将这两种方法结合起来，利用MATLAB 工具对其进行了分析。实验结果表明，检测的效果好于分别使用其中一种方法的情况。

关键词：端点检测，短时能量 过零率 门限

1 引言

近年来，在语音信号处理领域，关于语音信号中端点检测及判定的研究越来越重要。作为语音识别的前提工作，有效的端点检测方法不仅可以减少数据的存储量和处理时间，而且可以排除无声段的噪声干扰，使语音识别更为准确。目前的语音信号端点检测算法比较多，有短时能量，短时过零率分析，自相关法等等，其中以短时能量和短时过零率用的最多。大多文献和教材都是把它们分别进行介绍，由于它们各有其优缺点，分别使用作为语音端点检测的手段难免会漏掉很多有用的信息，因此，笔者将这两种方法结合起来进行分析，在判断清浊音及静音方面可以起到互补的作用，从语音信号的短时能量和过零率分析的特点出发，加以门限值来分析将两种方法相结合应用的效果，最后通过 Matlab 进行了仿真。

2 语音信号短时能量和过零率的特征

语音一般分为无声段，清音段和浊音段。一般把浊音认为是一个以基音周期为周期的斜三角脉冲串，把清音模拟成随机白噪声。由于语音信号是一个非平稳态过程，不能用处理平稳信号的信号处理技术对其进行分析处理。但由于语音信号本身的特点，在 10~30ms 的短时间范围内，其特性可以看作是一个准稳态过程，即具有短时性。因此采用短时能量和过零率来对语音进行端点检测是可行的。

信号的短时能量定义为：设语音波形时域信号为 $x(l)$、加窗分帧处理后得到第 n 帧语音信号为 $x_n(m)$，则 $x_n(m)$ 满足下式：

$$x_n(m) = w(m)x(n+m) \quad 0 \le m \le N-1 \qquad (2\text{-}1)$$

$$w(m) = \begin{cases} 1, & m = 0 \sim (N-1) \\ 0, & m = \text{其他值} \end{cases}$$

其中，$n = 0, 1T, 2T, \cdots$，并且 N 为帧长，T 为帧移长度。

设第 n 帧语音信号 $x_n(m)$ 的短时能量谱用 E_n 表示，则其计算公式如下[4][5][6]：

$$E_n = \sum_{m=0}^{N-1} x_n^2(m) \qquad (2\text{-}2)$$

如图 1 所示为英文单词“eat”的短时能量。

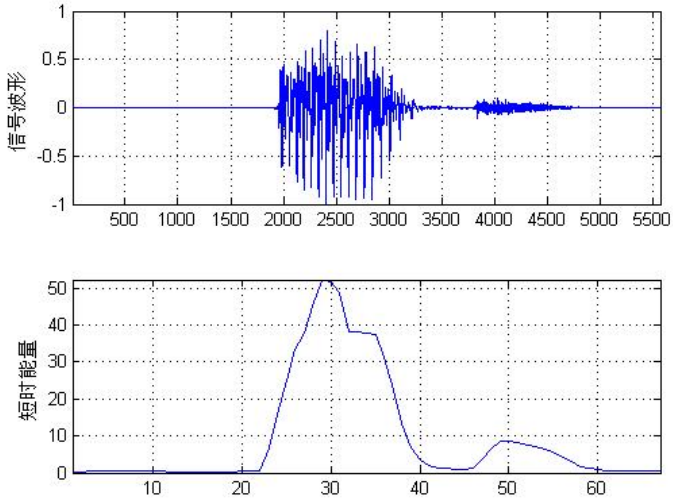

图 1 英文单词“eat”的短时能量

语音和噪声的区别可以体现在他们的能量上，语音段的能量比噪声段的能量大，如果环境噪声和系统输入的噪声比较小，只要计算输入信号的短时能量就能够把语音段和噪声背景区分开，除此之外，用基于能量的算法来检测浊音通常效果也是比较理想的，因为浊音的能量值比清音大得多，可以判断浊音和清音之间过渡的时刻[3]，但对清音来说，效果不是很好，因此还需要借助短时过零率来表征。

短时过零率表示一帧语音中语音信号波形穿过横轴（零电平）的次数。它可以用来区分清音和浊音，这是因为语音信号中的高频段有高的过零率，低频段过零率较低。

定义语音信号 $x_n(m)$ 的短时过零率 Z_n 为[4][5][6]：

$$Z_n = \frac{1}{2}\sum_{m=0}^{N-1} |\operatorname{sgn}[x_n(m)] - \operatorname{sgn}[x_n(m-1)]| \qquad (2\text{-}3)$$

式中，sgn[] 是符号函数，即：

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

如图 2 所示为英文单词“eat”的短时过零率。

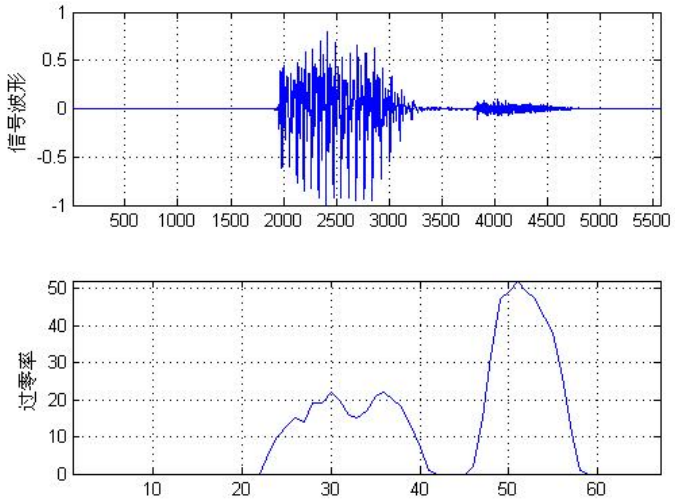

图 2 英文单词“eat”的短时过零率

从两幅图可以看出，短时能量可以近似为互补的情况，短时能量大的地方过零率小，短时能量小的地方过零率较大。

3 基于短时能量和过零率的检测方法

尽管基于短时能量和过零率的检测方法各有其优缺点，但是若将这两种基本方法相结合起来使用也可以实现对语音信号可靠的端点检测。无声段的短时能量为零，清音段的短时能量又比浊音段的短时能量大，而在过零率方面，理想的情况是无声段的过零率为零，浊音段的过零率比清音段的过零率要大的多，因此，假设有一段语音，如果某部分短时能量和过零率都为零或者为很小的值，就可以认为这部分为无声段，如果该部分语音短时能量很大但是过零率很小，则认为该部分语音为浊音段，如果该部分短时能量很小但是过零率很大，则认为该部分语音为清音段。正如前面提到，语音信号具有短时性，因此在对语音信号进行分析时，需要将语音信号以 30ms 为一段分为若干帧来进行分析，则两帧起始点之间的间隔为 10ms。[1][2]

为防止误判以及无声段过零率太大，设$tmp1$和$tmp2$为相邻两个采样点，则同时满足$tmp1*tmp2<0$和$tmp1-tmp2$的绝对值大于$delta$时才算一次过零，除此之外，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T_{EL}和T_{ZL}，其数值较小，对信号的变化比较敏感，很容易就会被超过。另一个是较高的门限T_{EH}和T_{ZH}，数值较大，信号必须达到一定的强度，该门限才可能被超过。低门限被超过有可能是时间很短的噪声引起的，高门限被超过则可以基本确定是由语音信号引起的，如图 3 所示。

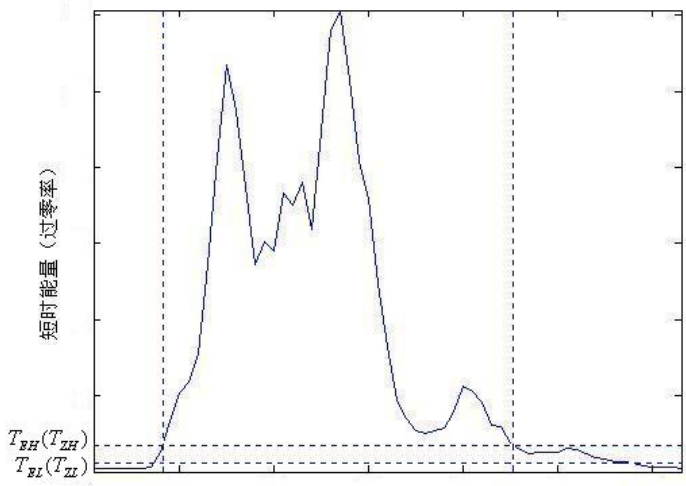

图 3 端点检测方法示意图

4. Matlab 实验仿真

笔者以英语单词“fine”为例，首先将语音信号进行归一化，把幅度限制在[-1，1]之间，以便处理方便，利用 MATLAB 语句：

[Y, Fs, BITS]=wavread ('fine.wav');

可知其采样频率为 8KHz，再对输入的语音信号进行分帧，所以 30ms 的帧长对应 240 个采样点，而 10ms 的帧移对应 80 点。则可以利用 matlab 语音分析工具箱对字母“fine”的采样序列分帧的过程为：

```
>> [Y, Fs, BITS]=wavread ('fine.wav');
>> y = enframe (Y, 240, 80);
>> whos y Y
  Name    Size    Bytes Class     Attributes
  Y       17375x1   139000        double
  y       215x240    412800       double
```

可见，包含语音采样的一维数组经过处理后得到二维数组，表示总帧数为 215 帧，每帧 240 个采样。在计算短时能量之前，首先将语音信号通过一个一阶高通滤波器进行预加重，主要是去除低频干扰。而后使幅度归一化求其能量和过零率，在进行求过零率的时候首先要设定一个门限值。本文通过在经验值 0.02 的基础上，通过不断地进行实验和细微调整，得到在门限值为 $delta$ =0.03 的时候效果最好。设置短时能量的两个门限分别为 T_{EH} =4 和 T_{EL} =1，设置过零率的两个门限值分别为 T_{ZH} =4 和 T_{ZL} =2，得到过零率波形如图 4，

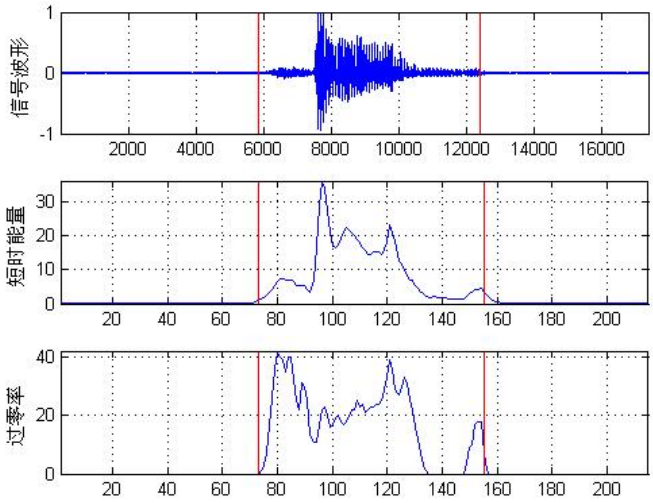

图 4 单词“fine”的短时能量和过零率

单词当中的[f]音为清音，可以看作是随机的白噪声，它的短时能量比较低，过零率次数却比较多。从图上可以看出，当语音信号的能量值超过T_{EH}时且过零率超过T_{ZH}时，这就表示正式进入了语音段，在语音段结束时，通过判断能量值小于T_{EH}且过零率小于T_{ZH}来判断语音段的结束，如图中红线所标识，由于设置了门限值，使得出现误判的情况相对减少，因此，可用来区分清音，浊音和无声段。结果表明，在信噪比很低时，短时能量方法的检测效果不太理想，结合过零率的方法以后，检测的效果有明显的改善。

5 结论

短时能量分析和过零率分析作为语音信号时域分析中最基本的方法。但是很多情况表明使用单一的一种方法并不能得到理想的检测结果，这是因为短时能量分析是通过能量的高低来区分清音和浊音，不容易确定语音信号片段的起始点；而过零率分析仅仅是表明清音的过零率高于浊音，对噪声的存在比较敏感，如果背景中有反复穿越坐标轴的随机噪声，会产生大量的虚假过零率，影响检测结果。对于背景噪声和清音的区分则显得无能为力。将这两种方法结合起来，通过短时能量分析去除高频环境噪声的干扰，用过零率分析去除低频的干扰，检测效果较好。但综合考虑后，由于这两种方法本身的局限性以及过零率门限值和短时能量门限值的选取，使得检测的范围和精度仅限于单个单词，而对整个句子的检测还达不到令人满意的效果。

参考文献

[1] 何强，何英. MATLAB 扩展编程[M]，北京：清华大学出版社，2002，

[2] 刘羽. 语音端点检测及其在 Matlab 中的实现[J]. 计算机时代，2005（8）：25-26，

[3] 张仁志. 基于短时能量的语音端点检测算法研究[J]. 语音技术，2005（7）：52-54，

[4] 胡航. 语音信号处理[M]. 黑龙江：哈尔滨工业大学，2000.5，

[5] 张雄伟. 现代语音处理技术及应用[M]. 北京：机械工业出版社，2003.8，

[6] 赵力. 语音信号处理[M]. 北京：机械工业出版社，2003.3

Research on Endpoints Detection of Speech Signal based on Short-time Energy and Zero-crossing counts

Liu Bo, Nie Mingxin，Xiang Juntao
School of information, WuHan University of Technology, HuBei, PRC (430070)

Abstract

As the most basic method of speech signal analysis in Time-domain, short-time analysis and zero-crossing counts analysis are widely used, especial in endpoints detection of speech signal, but it is easily to ignore important information because of their used separately. This paper discussed the two methods together, and take an analysis by MATLAB,. The result indicated that the effects is better than using any one of them separately..

Keywords：*endpoints detection, short-time energy, zero-crossing counts, threshold*

作者简介：刘波，男，1981 年生，硕士研究生，主要研究方向：语音信号处理。