

语音识别中双门限端点检测算法的研究

肖宜

武汉理工大学信息工程学院, 湖北武汉 (430070)

E-mail: dogxy@qq.com

摘 要: 语音信号起止点的判别是任何一个语音识别系统必不可少的组成部分。有噪声的情况下, 单纯用短时能量或者短时过零率不能准确检测出语音信号。本文采用短时能量和短时过零率相结合的方法, 利用短时能量和短时过零率两个门限来确定语音信号的起点和终点, 最后通过仿真实现结果。

关键词: 语音识别 端点检测 双门限 仿真

中图分类号:

1. 引言

语音信号起止点的判别是任何一个语音识别系统必不可少的组成部分。因为只有准确的找出语音段的起始点和终止点, 才有可能使采集到的数据是真正要分析的语音信号, 这样不但减少了数据量、运算量 and 处理时间, 同时也有利于系统识别率的改善^[1]。因此端点作为语音分割的重要特征, 在很大程度上影响语音识别系统的性能, 如何在噪声环境下设计一个鲁棒的端点检测算法是一个非常棘手的问题。但是, 单纯用短时能量或者短时过零率不能准确检测出语音信号。本文采用短时能量和短时过零率相结合的方法, 利用短时能量和短时过零率两个门限来确定语音信号的起点和终点, 目的是从采集到的语音信号^[2]中分离出真正的语音信号作为系统处理的对象。

2. 常用的端点检测方法

常用的端点检测方法有以下几种:

(1) 短时平均能量

设 $S(n)$ 为加窗语音信号, 第 t 帧语音的短时平均能量为:

$$Eng(t) = \frac{1}{N} \sum_{n=0}^{N-1} |S_t^2(n)| \quad (1)$$

$$Or \quad Eng(t) = \frac{1}{N} \sum_{n=0}^{N-1} |S_t(n)| \quad (2)$$

其中 N 为分析窗宽度, $S_t(n)$ 为第 t 帧语音信号中的第 n 个点的信号样值。上面两式原理是相同的, 但后式有利于区别小取样值和大取样值, 不因前式取平方造成很大差异^{[3][4]}。

短时平均能量是时域特征参数。把它用于模型参数时, 应进行归一化处理, 本文语音识别系统中取其对数值后使用, 使计算和识别结果均取得了较好的效率和结果。

(2) 短时过零率

短时过零率 ZCR(Zero-Crossing-Rate)为:

$$Z_n = \sum_{m=-\infty}^{+\infty} Sgn[x(m)] - Sgn[x(m-1)] \times W(n-m) \quad (3)$$

其中:

$$\begin{cases} Sgn[x(n)] = 1 & x(n) > NoiseMax(NoiseMax \text{ 为噪声上限}) \\ Sgn[x(n)] = -1 & x(n) < NoiseMin(NoiseMin \text{ 为噪声下限}) \\ Sgn[x(n)] = 0 & otherwise \end{cases} \quad (4)$$

$$\begin{cases} W(n) = \frac{1}{2N} & 0 \leq n \leq N-1 (N \text{ 为一阵声音的长度}) \\ W(n) = 0 & otherwise \end{cases} \quad (5)$$

有噪声的情况下,单纯用短时能量或者短时过零率不能准确检测出语音信号。本课题采用短时能量和短时过零率相结合的方法,利用短时能量和短时过零率两个门限来确定语音信号的起点和终点,目的是从采集到的语音信号中分离出真正的语音信号作为系统处理的对象。

3. 双门限端点检测算法的原理

本文采用短时能量和短时过零率相结合的方法,利用短时能量和短时过零率两个门限来确定语音信号的起点和终点,目的是从采集到的语音信号中分离出真正的语音信号作为系统处理的对象。

在开始进行端点检测之前,首先为短时能量和过零率分别确定两个门限。一个是比较低的门限,其数值比较小,对信号的变化比较敏感,很容易就会被超过。另一个是比较高的门限,数值比较大,信号必须达到一定的强度,该门限才可能被超过。低门限被超过未必就是语音的开始,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信是由于语音信号引起的。

整个语音信号的端点检测可以分成四段:静音、过渡段、语音段、结束。程序中使用一个变量 `status` 来表示当前所处的状态。在静音段,如果能量或过零率超越了低门限,就应该开始标记起始点,进入过渡段。在过渡段中,由于参数的数值比较小,不能确信是否处于真正的语音段,因此只要两个参数的数值回落到低门限以下,就将当前状态恢复到静音状态。而如果在过渡段中两个参数中的任一个超过了高门限,就可以确信进入语音段了。

一些突发性的噪声也可以引起短时能量或过零率的数值很高,但是往往不能维持足够长的时间,如门窗的开关、物体的碰撞等引起的噪声,这些都可以通过设定最短时间门限来判别。当前状态处于语音段时,如果两个参数的数值降低到低门限以下,而且总的计时长度小于最短时间门限,则认为这是一段噪声,继续扫描以后的语音数据。否则就标记好结束端点,并返回。

端点检测程序流程图如下图所示:

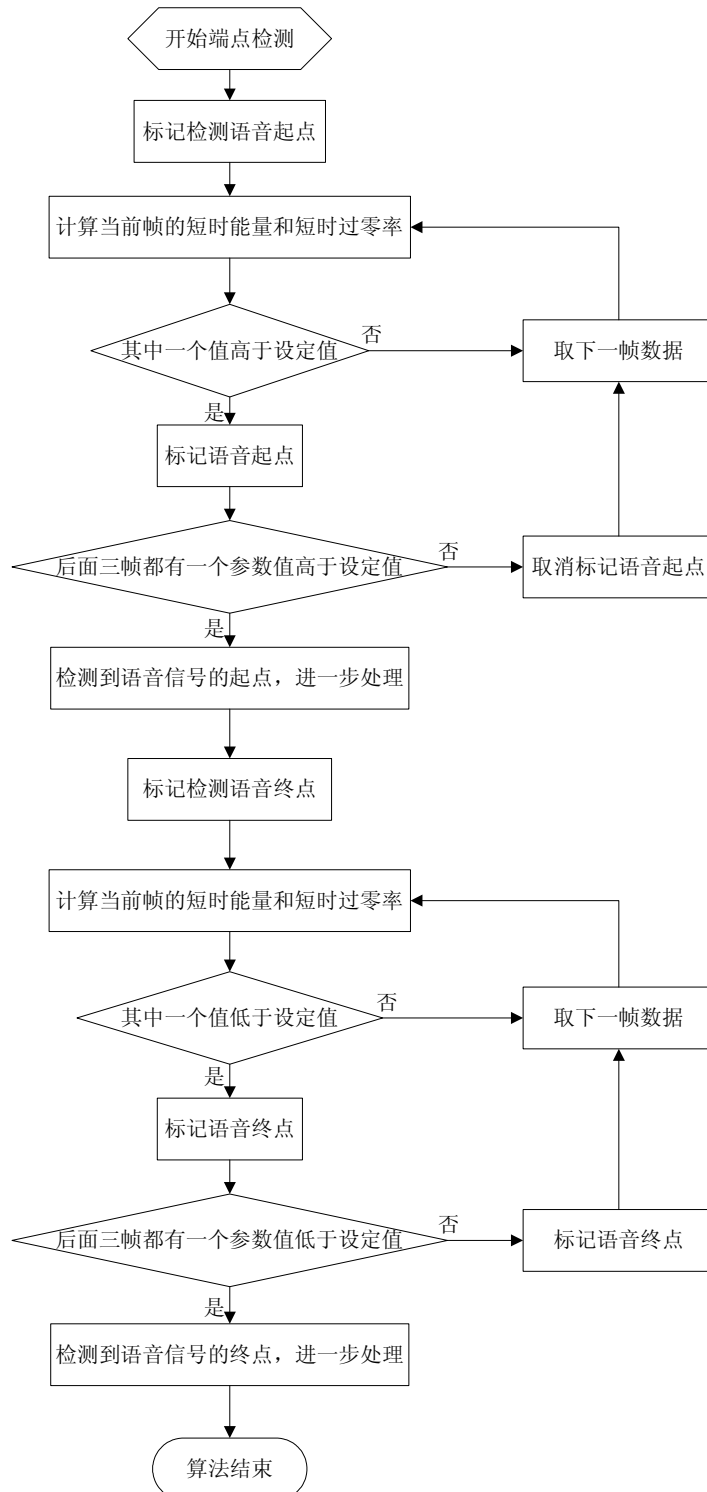


图 1 双门限端点检测程序流程图

4 双门限端点检测算法的仿真实现

在仿真实现中, 我们使用 matlab7.0.4 对女声发音“你好”进行双门限端点检测的仿真。发音的数据是通过麦克风由计算机的声卡录音得到的, 采样频率是 8000Hz。为了进行验证算法的有效性, 我们特意在有环境噪声的条件下进行录音。端点检测的识别效果如下图所示:

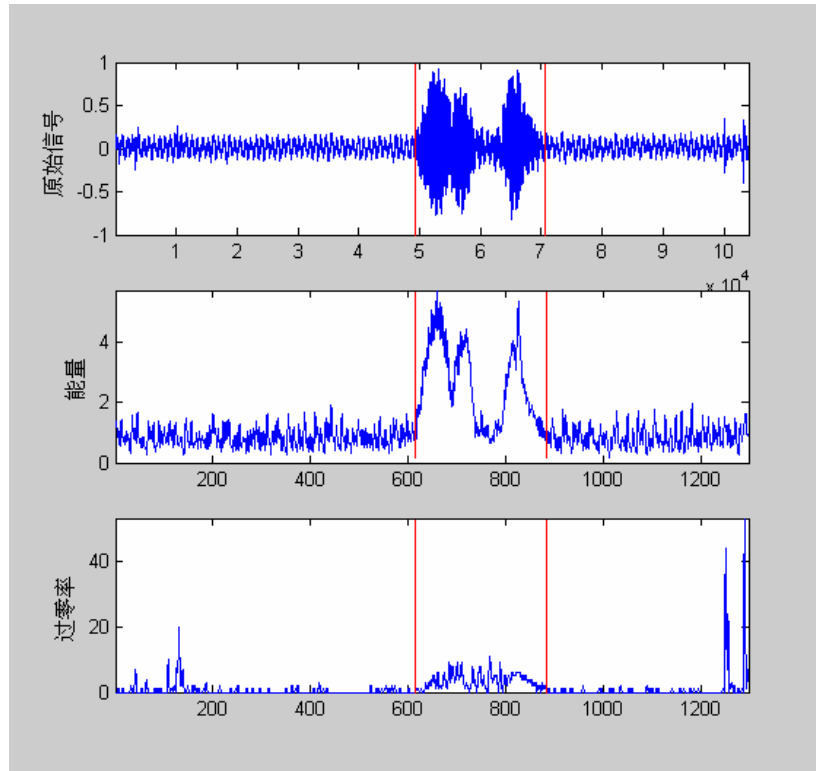


图2 女声发音“你好”的双门限端点检测示意图

从图2我们可以看出,在有环境噪音的情况下,双门限端点检测算法很好的识别了语音信号的起点和终点。将此算法应用于语音识别中能有效去除冗余信息,提高语音识别系统的识别效果^{[5][6]}。

5. 总结

本文设计了一个针对语音识别中端点检测的双门限法端点检测算法,经过理论分析和实验仿真,证明这个算法能有效去除冗余信息,提高语音识别系统的识别效果,具有很现实的重要意义。

参考文献

- [1] 杨行峻, 迟惠生. 语音信号数字处理[M]. 北京: 电子工业出版社, 2003
- [2] 姚天任. 数字语音处理. 湖北: 华中科技大学出版社, 2002
- [3] M.H. Savoji. A Robust Algorithm for Accurate End Pointing of Speech. Speech Communication, 1989, 8(2):45~60
- [4] R.Bhiksha, S.Rita. Classifier-based Non-linear Projection for Adaptive End Pointing of Continuous Speech. Computer Speech & Language, 2003, 17(1):5~26
- [5] 聂敏. 语音识别及其关键技术[J]. 微波与卫星通信, 1999, 4:53~56
- [6] C. Lee, D. hyun, C. Nadeu. Optimizing feature extraction for speech recognition [J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(1):80~86

Research on Dual-Threshold Endpoint Detection Algorithm of Speech Recognition

Xiao Yi

Department of Information Engineering, Wuhan University of Technology, Wuhan, Hubei, China (430070)

Abstract

The beginning and ending points of voice signals are essential components of any speech recognition system. With noise in voice, simply using short-term energy or short-term zero rate can not accurately detect speech signals. In this paper, short-term energy and short-term zero rate are combined in the way that both of them work as thresholds to determine the voice signal starting and ending points. The result is simulated in the final part.

Keywords: *Speech Recognition, Endpoint Detection, Dual-Threshold, Simulation*

作者简介: 肖宜, 男, 1984 年生, 硕士研究生, 主要研究方向是嵌入式与智能信息。