

基于短时相关性和极值点的语音边界检测

王道成 胡修林 张蕴玉

(华中理工大学电子与信息工程系)

摘要: 针对基于短时能量和过零率相结合的语音边界检测法在实际应用中的不足,提出了一种改进方法——基于短时相关性和极值点的语音边界检测法,并在 MATLAB 环境下进行实验测试,在语音客观评价及语音识别算法中取得了良好的效果.

关键词: 语音边界检测;短时自相关函数;MATLAB

中图分类号: TP912.34 **文献标识码:** A **文章编号:** 1000-8616(2000)12-0013-03

语音边界检测经常采用的一种方法是基于短时能量和过零率相结合的两级判别法^[1,2],它实施简单,计算量相对较小,因而得到广泛的应用.但这种语音边界检测法比较粗糙,对于语音识别等需要高精度的检测语音边界的应用领域来说,这种检测方法是不足的,难以达到理想的效果.例如汉语数字“7”发音,其浊音部分 [i] 能量较大,用这种方法可以很容易将其检测出来.但它的清音部分 [q],不但幅度较小,而且上下包络的中心围绕着时间轴上下波动,此时过零率也不能够表征语音信号的频率特性,所以在这种情况下用上述的检测方法不能准确地检测出语音段起止的位置.在总结传统的语音边界检测方法的基础上,本文提出了一种基于短时自相关性和极值点数的语音边界检测法,这种方法的突出优点是继承传统方法的优点,同时能够克服它的缺点.

1 基于短时相关性和极值点的语音边界检测方法

方法的基本思想是首先通过相关性检测出语音的大致位置,然后再利用极值点数来检测,将语音比较精确的检测出来.设语音信号为 $s(m)$,用这种改进的检测算法进行相关值计算时,先用一个位于 n 的移动窗 $w(n-m)$ 加在语音信号 $s(m)$ 上,选出一段语音信号 $s_{wn}(m)$, $s_{wn}(m) = s(m)w(n-m)$. 本文采用了矩形窗,表达式如下:

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

式中 N 为帧长.

$s_{wn}(m)$ 的自相关函数称为语音信号 $s(m)$ 的

短时自相关函数 $R_n(k)$, 有

$$R_n(k) = \sum_{m=-} s_{wn}(m) s_{wn}(m+k). \quad (2)$$

因窗的宽度有限(等于 N),上式可写为

$$R_n(k) = \sum_{m=0}^{N-1-k} s_{wn}(m) s_{wn}(m+k). \quad (3)$$

因为 $R_n(k)$ 在 $k=0$ 处, $R_n(0)$ 等于加窗信号的平方和,也就是短时能量

$$R_n(0) = \sum_{m=0}^{N-1} s_{wn}^2(m). \quad (4)$$

一般来说,在强噪声情况下,短时能量不利于区分语音与噪音.但语音前后采样点之间的相关性很大,而噪声通常是随机的,前后之间一般没有相关性或者说相关性很小.因此,用 $L=1 \sim T$ (T 为经验值,本试验取 $T=9$) 的短时自相关值的平均值表示信号的相关值

$$\bar{R} = \frac{1}{T} \sum_{k=1}^T R_n(k).$$

当语音信号随时间上下起伏时,用平均过零率去检测清音比较困难.为此,可用短时极值点数代替平均过零率来二次检测语音边界.所谓极值点就是语音时域波形的局部最大值和最小值.当语音信号随时间上下起伏时,信号波形包络中心偏离时间轴,信号的过零率将发生变化,此时过零率不能很好地反映语音信号的频率特性.虽然波形位置发生了偏移,但是信号波形的极值点数一般不会发生变化,因此用极值点数取代过零率将更好地反映语音信号的真实特性.边界检测算法步骤如下:

a. 进行语音预处理,如滤波、归一化、分帧等,由式 (1) ~ (4) 计算各帧的平均自相关值、极

收稿日期: 2000-07-03.

作者简介: 王道成(1976-),男,硕士研究生;武汉,华中理工大学电子与信息工程系(430074).

值点数;设置自相关上门限值 R_H ,用平均自相关值检测语音的大致位置 N_s 和 N_e .

b. 利用自相关下门限值 R_L ,在步骤 a 的基础上,由 N_s 点向前和由 N_e 点向后搜索语音的起止点 N_{s1} 和 N_{e1} .

c. 利用每帧的极值点数及其上下门限 (P_H 和 P_L),在 N_{s1} 点继续向前搜索语音边界端点,直到某一帧的极值点数 $P > P_H$ 或者 $P < P_L$,找到语音新的起始点 N_{s2} (见图 1).

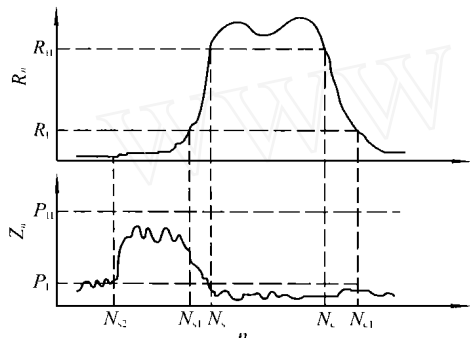


图 1 基于相关性和极值点数相结合的语音边界检测

为了避免将同一个语音段切分成两个或多个语音段,在上面检测的基础之上,对检测出的语音段进行合并,当一语音段的结尾处距下一语音段开始位置小于某一距离值时,可以将这两段语音合并成一段.因为某些语音可能有一个或一个以上的自相关峰值,用相关性进行判别时就可能将它们切分成不同的语音段,但是它们之间的距离通常很短,一般小于 70 ms.

另一方面,在语音边界检测的过程中有时会遇到短时突发性的干扰噪声,它们的能量值比较大,短时自相关性也会超过自相关上门限值 R_H ,这样就会将这种噪声误识别为语音.观察研究表明这种干扰噪声持续时间很短,一般小于 50 ms,而一般一个字的发音通常在 200 ms ~ 500 ms.为了消除这种干扰,可以用检测后的起止长度来判别它是不是语音.

2 几个门限参数大小的确定

需要设定的几个门限参数值:自相关门限的上下门限、极值点数的上下门限.

自相关门限值上限与下限.由于语音的响度不一,故自相关门限没有固定值,需要针对具体的语音进行动态设定,或者首先对语音数据进行归一化处理,然后再设立固定门限值.本文采用动态门限法.首先求出各帧自相关值的大小,然后再求

出它们的平均值 R_A ,实验表明,当相关长度 $T = 9$,自相关门限上限 $HIGH_REL_THRESHOLD = R_A$;下限 $LOW_REL_THRESHOLD = R_A/8$ 时检测效果最好.

极值点数的上下门限.与自相关值一样,在不同的语音环境或者对于不同人的发音(尤其是男声和女声),平均极值点数也是不断变化,有些甚至差别很大,所以,对于极值点数门限的设定也需要采用动态门限法.首先计算各帧的极值点数,然后求出它们的平均极值点数 P_A .实验表明,极值点数门限的上限 $HIGH_POLE_THRESHOLD = 5 P_A$,下限 $HIGH_POLE_THRESHOLD = 2 P_A$.

3 检测结果分析比较

实验表明,通过以上方法的检测,可将各种信噪比下的语音边界较为准确的检测出来,尤其对于小信噪比语音信号(最小可达 -5 dB 以下).表 1 是三种检测结果的对比.从中可以看出:基于相关性和极值点数的检测方法检测的结果最逼近人工检测的结果.相对于基于能量和过零率的检测方法,这种检测结果更为精确,尤其是对于清音部分,例如汉语数字 3,4 和 7 等,检测结果与人工检测结果误差 1 ~ 2 帧(约 10 ~ 20 ms),在常用的语音处理领域已经能够满足边界检测的要求.

表 1 三种检测结果的比较(检测结果以起止帧表示)

数字语音	传统方法	改进方法	人工检测
0	70 ~ 106	69 ~ 110	69 ~ 110
1	162 ~ 190	159 ~ 193	160 ~ 193
2	252 ~ 285	250 ~ 284	250 ~ 285
3	343 ~ 371	333 ~ 374	332 ~ 375
4	441 ~ 465	421 ~ 466	422 ~ 466
5	520 ~ 552	513 ~ 558	514 ~ 557
6	608 ~ 645	607 ~ 644	607 ~ 645
7	709 ~ 736	695 ~ 738	694 ~ 737
8	794 ~ 820	793 ~ 826	793 ~ 825
9	888 ~ 915	884 ~ 919	883 ~ 919

基于相关性和极值点数的边界检测方法具有传统方法无法比拟的优越性,不仅可将浊音部分准确检测出来,而且能将清音部分较为准确地检测出来,还可避免传统方法中的虚检和漏检现象.

参 考 文 献

- [1] 姚天任. 数字语音处理. 武汉:华中理工大学出版社, 1992.
- [2] 陈斐利,朱 杰. 一种新的基于自相关相似距离的语音信号端点检测方法. 上海交通大学学报, 1999, 33

(9) : 97 ~ 99

Speech Signal Endpoint Detection Based on Relativity and Pole

Wang Daocheng Hu Xiulin Zhang Yunyu

Abstract : An improved algorithm of speech endpoint detection is presented which is based on Relativity and Pole. Using the MATLAB platform , the new algorithm has been tested carefully , and it can be used in the speech quality objective measurement and speech recognition field successfully.

Key words : speech endpoint detection ; short-time auto-correlation function ; MATLAB

Wang Daocheng Postgraduate ; Dept. of Electronics & Information Eng. , HUST, Wuhan 430074 , China.

(上接第 7 页) MAD 次数的不确定而引起的. 图 2 和图 3 分别给出了 Miss American 序列中第 15 帧的原图和解码重构图像.



图 2 原图

综上所述可以看出 :SASE 算法比 ESA 算法大大地减少了计算量 ,提高了运算速度 ,而复原图像的主客观质量非常接近于 ESA 算法 ,是一种较好的快速块运动估计搜索算法.



图 3 解码图像

参 考 文 献

- [1] Li W, Salari E. Successive Elimination Algorithm for Motion Estimation. IEEE Trans. on Image Processing, 1995, 4 (1) : 105 ~ 107
- [2] Liu Bede, Zaccaritin A. New Fast Algorithms for the Estimation of Block Motion Vectors. IEEE Trans on CSVT, 1993, 3(2) : 148 ~ 157

A New Algorithm for Block Motion Estimation of Video Coding

Wang Chao Peng Jiaxiong

Abstract : A sub-sampling alternating search algorithm with successive elimination (SASE) used for block motion estimation is proposed. Quarter sub-sampling alternating algorithm is used to implement exhaustively search in the reference areas. The experiments show which performance is closed to exhaustive search algorithm (ESA) and its computation requirement is a quarter of ESA. In the course of any sub-sampling block computation , an inequality is employed to successive location elimination so that the computation requirement can be reduced and the search efficiency can be improved.

Key words : image recognition ; motion estimation ; MPEG1 ; MPEG2 ; video code

Wang Chao Postgraduate ; Dept. of Contr. Sci. & Eng. , HUST, Wuhan 430074 , China.