

基于 Mel-TEO 的带噪语音端点检测算法

王茂蓉¹, 周 萍¹, 景新幸², 杨 青¹

(1 桂林电子科技大学 电子工程与自动化学院, 广西 桂林 541004;

2 桂林电子科技大学 信息与通信学院, 广西 桂林 541004)

摘 要: 针对短时 TEO 能量算法抗噪性差的缺点, 提出了一种强噪声下的端点检测新算法. 该算法在短时 TEO 能量端点检测的基础上, 增加 Mel 倒谱距离判断环节, 采用先粗判后精判的互补性两级判决机制. 首先利用强抗噪性 Mel 倒谱距离进行端点粗判, 然后再利用体现语音信号时域特征与语音共振峰特性的短时 TEO 能量进行端点精判. 实验表明, 在信噪比相对较低的环境下, 该改进算法与传统的双门限法和短时 TEO 能量相比, 在没有增加运算复杂度的同时提高了检测系统的准确度.

关键词: 语音端点检测; 说话人识别; Teager 能量算子; Mel 倒谱距离; 噪声

中图分类号: TN912.34

文献标识码: A

文章编号: 1000-7180(2016)04-0046-04

Voice Activity Detection Algorithm Based on Mel-TEO in Noisy Environment

WANG Mao-rong¹, ZHOU Ping¹, JING Xin-xing², YANG Qing¹

(1 Department of Electric Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China; 2 Department of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: In order to solve the problem of the poor anti-noise performance, this paper proposes a new voice detection algorithm in noisy environment. On the basis of short-time TEO energy voice detection, increasing the judging section Mel Cepstral distance, we use two-stage decision mechanism with accurate judgment after rough judgment. First, we judge the voice endpoints preliminarily using Mel Cepstral distance which has a good anti-noise performance. Then, we judge the endpoints accurately using short-time TEO energy which reflects time-domain and voice formant characteristics of the speech signal. Experiments show that this algorithm has better capability in low SNR, compared with the traditional double thresholds and short-time TEO energy, without increasing the computational complexity and improving the accuracy of the detection system.

Key words: voice activity detection; speaker recognition; Teager energy operator; Mel cepstral distance; noise

1 引言

端点检测在说话人识别系统中起着至关重要的作用, 不仅减小特征提取的处理时间, 降低了计算复杂度, 而且可让整个系统免于静音部分的噪声干扰^[1], 故其检测精度直接影响整个系统的性能. 好的

特征参数必须满足对语音信号的端点判别能力强, 鲁棒性好, 以及适用于不同的噪声环境^[2], 目前最常见的检测算法有传统双门限法、谱熵法、Mel 倒谱距离法、短时 TEO 能量法等. 其中谱熵法在应用中存在很多不足, 比如说对清音的判别度低, 抗噪声性差; 传统双门限法对清音的检测效果好, 但抗噪性能

收稿日期: 2015-05-23; 修回日期: 2015-07-10

基金项目: 国家自然科学基金资助项目(61363005, 61462017); 广西自然科学基金(2012GXNSFAA053221, 2014GXNSFAA118353); 广西教育厅科研项目(2013YB083); 广西研究生教育创新计划资助项目(YCSZ2015152)

稍差,难以区分微弱的摩擦音以及语音末段的鼻音,只能在噪声相对较小的时获得较好的检测效果.所以,如何进一步提高端点检测的准确度仍为说话人识别领域研究的热门课题.文献[3]首先求取语音各帧的 Mel 倒谱系数,继而估算背景噪声,接着通过滤波得到加权倒谱距离,然后根据其进行端点判别,使得判别的效果得到改善;文献[4]将检测模型分为训练模块和检测模块,运用训练模型判断出语音段和噪声段,利用检测模块去除语音段中的静音部分和噪声,在提高端点检测性能的同时增加了计算复杂度.

以上算法都是对信号进行逐帧处理,并没利用到语音信号的长时信息,因此,为了提高在强噪声中的端点检测性能,针对现有算法不足,本文提出将 Mel 倒谱距离和短时 TEO 能量两种特征联合起来对语音端点判决,采用两级判决机制,先通过抗噪性较好的 Mel 倒谱距离对语音端点进行粗判,判断出语音段和非语音段,然后通过 TEO 进行端点精判.实验表明,在没有增加运算复杂度的同时检测系统的准确度得到了提高.

2 算法原理

2.1 Mel 频率倒谱系数

文献[5]等提出的 Mel 倒谱系数,是在非线性的 Mel 尺度上对其功率谱求对数并进行 DCT 变换获得的.由于实际频率带在 Mel 尺度上呈现等间隔的指数分布,相比于倒谱上线性分开的特点,能更好地描述语音的频率特性,故本文选择用其来判别出有声段和无声段.其中,MFCC 特征参数是在 Mel 频率尺度上提取出来的,两者之间的转关系换如式(1)所示.

$$\text{Mel}(f) = 1127 \ln(1 + f/700) \quad (1)$$

式中, f 为频率,单位为 Hz.

提取 MFCC 参数的具体步骤^[6]如图 1 所示.首先对信号进行预处理,然后经离散 FFT 变换,使其从时域转换到频域,接着送入 Mel 滤波器组滤波,求取对数能量,接着对输出向量作 DCT 求倒谱,求得 MFCC,其中本文 MFCC 阶数选为 12 阶,用于后续倒谱距离的计算.



图1 提取 MFCC 参数

2.2 Mel 倒谱距离

目前经常用到的距离测度主要为相关系数距离,欧式、明式距离等,本文通过应用欧式距离来计算 Mel 倒谱距离,其定义式为

$$d_{cep}(c_n, \bar{c}) = \frac{1}{q} \sum_{n=1}^q (c_n - \bar{c}^2) \quad (2)$$

式中, $q = 12$ 代表 Mel 倒谱阶数, c_n 代表 n 帧的倒谱, $\bar{c}$ 代表噪声的倒谱.

2.3 TEO 原理

Teager 能量算子 (Teager Energy Operator, TEO),能够很好地表征信号的共振峰信息,TEO 能在抑制背景噪声中消除零均值噪声,起到信号增强的作用,且计算量不大,对其求取相对简单,同时进行信号特征提取的作用.文献[7]提供了计算 TEO 能量的公式.时间信号连续的情况下,TEO 定义为

$$\phi[x_t(\tau)] = (x_t'(\tau))^2 - x_t(\tau)x_t''(\tau) \quad (3)$$

式中, $x_t'(\tau) = dx_t(\tau)/d\tau$, $x_t(\tau)$ 为连续的时域信号.而对离散时间信号,TEO 定义为

$$\phi[x_t(\tau)] = x_t(\tau)^2 - x_t(\tau-1)x_t(\tau+1) \quad (4)$$

式中, $x_t(\tau)$ 是离散时域信号在 τ 时刻的值.通过式(4)可得,TEO 输出的信号,只跟它前后各一个时刻的样本点有关.

其原理框图如图 2 所示. TEO 的具体求取过程是首先求出 TEO,并对其进行预处理,求各帧信号的 TEO 能量,其计算公式为式(5),接着对 TEO 平滑滤波,然后利用判决机制进行端点判断.

$$(E_{\text{TEO}})_t = \sum_{\tau=1}^{L-1} |x_t(\tau)|^2 \quad (5)$$

式中, $(E_{\text{TEO}})_t$ 为第 t 帧 TEO 能量, τ 是每帧样点数; L 为帧长,本文取为 256.

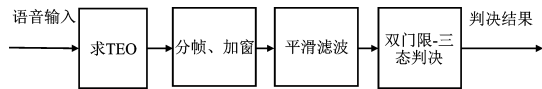


图2 TEO 原理框图

现有环境中不可能完全消除噪声干扰,语音中或多或少都会掺杂一些噪声,所要做的就是尽可能降低噪声扰动而引起的错误判断,增强系统的稳定性,故本文在检测前加入了平滑滤波的环节,即对 TEO 进行平滑滤波,计算如式(6)所示.

$$\overline{E[\phi(n)]} = \frac{E[\phi(n-1)] + E[\phi(n)] + E[\phi(n+1)]}{3} \quad (6)$$

从而在一定程度上降低因噪声的存在而引起的

系统误判,提高了检测系统的准确性.

3 算法实现

由于基于能量特征的端点检测系统,抗噪声效果相对比较差,而这种缺点在实际应用中是无法避免的.为此,本文提出的 Mel 倒谱距离和短时 TEO 能量的互补性的两级联合算法来进行端点检测,此算法可以很大程度地解决仅仅依靠能量算法的不足,提高了系统检测准确率,其系统框图如图 3 所示.

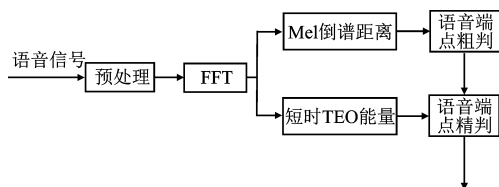


图3 本文算法系统原理框图

3.1 Mel 倒谱距离粗判

假设语音前 10 帧为背景噪声,通过对其求倒谱系数均值来估算噪声倒谱系数,记为向量 c . 接着通过式(7)求得每帧语音信号和噪声的倒谱距离,继而获得倒谱距离的曲线轨迹,确定两个门限阈值即低门限阈值 G_1 和高门限阈值 G_2 .

$$G_i = d_{cep} \cdot k_i, i = 1, 2 \quad (7)$$

式中, d_{cep} 代表噪声倒谱距离, $k_2 > k_1$, 均是门限系数. 最后,利用每帧语音信号的倒谱距离判断语音起始点,若其倒谱距离超过 G_1 ,就标记为起始点进入过渡段,若期间倒谱距离低于 G_1 ,则表明当前状态是静音段;若连续的 3 帧倒谱距离高于 G_2 ,方可判断为已经进入语音段,结束点的判决类似.

3.2 短时 TEO 能量精判

与 Mel 倒谱距离粗判类似,在端点检测之前,先要确定两个门限阈值,即高低门限阈值. 阈值的选择应该遵循以下原则:一是低门限阈值要相对较小,能够感知微弱变化的信号,较容易被超越;二是高门限阈值要相对大一些,只有当语音信号强度到达一定程度时,方能将其超越. 需要注意的是,当有突发的尖峰噪声存在时,低门限阈值有可能被超越,故仅凭低门限阈值是不能判断语音的起始点,只有当高门限阈值被超越才能判断为进入语音段,即该帧记为起始点. 一般情况下,语音段的前后各 5 帧为没有语音信号,可以假设为静音段,仅存在噪声,根据上述假设我们可以很容易地计算出高低门限的阈值.

通过多次实验得出,一方面由于语音段的前后

各 5 帧信号能量忽高忽低,导致端点检测准确度降低,另一方面由于对于不同的语音段其高低门限阈值也是不同的,如若采用一成不变的门限阈值,不仅不能增强系统的检测性能,反而适得其反. 为了解决上述两种情况,在计算门限阈值时,我们加入整个语音段的平均能量值 $\bar{E}$ 和最小能量值 $E_{\min}$,实时调整高低门限阈值,得出动态门限阈值,其计算公式如下:

$$\begin{cases} E_1 = \max[0.4E_{Tmx} + 0.6E_{\min}, 0.3\bar{E} + 0.7E_{\min}] \\ E_2 = \max[E_1 + E_{Tmx} - E_{Tmm}, E_1 + (\bar{E} + E_{\min})/10] \end{cases} \quad (8)$$

式中, E_{Tmm} 、 E_{Tmx} 分别为前后 10 帧静音段的最小、最大 TEO 能量,最后根据高低门限阈值判断语音的起始点和终止点,方法和 Mel 倒谱距离类似.

由于在实际环境中几乎不存在纯净的语音,都或多或少地掺杂着噪声,本文充分利用了 Mel 倒谱距离的抗噪性能首先对语音信号进行端点粗判,然后又结合体现语音共振峰特性和消除零均值噪声影响的 TEO 能量,进行端点精判,又为了上述两种特征的计算方便,故在没有增加计算复杂度的同时提高了端点检测的准确性,改善了系统的性能.

4 结果与分析

4.1 识别结果

实验采用噪声库中 4 种背景噪声(White, Babble, Factory, Volvo),在 4 种不同信噪比下(10 dB, 5 dB, 0 dB, -5 dB)下进行端点检测. 测试语音的采样频率 8 kHz,采样精度 16 bit,并给出双门限法、短时 TEO 能量和本文算法的各自的检测精度. 首先我们采用同一信噪比(-5 dB)同一句话作为测试,用传统双门限法、短时 TEO 能量和本文算法对语音信号进行端点检测,由检测结果可以看出,相对于双门限法和短时 TEO 能量,本文算法的端点准确度更高,检测结果输出的线条更加平滑,说明对噪声的过滤性更好,更容易读取其特性,为进行说话人识别的研究奠定了基础.

4.2 结果分析

通过上述实验,证明本文所提算法是可行的. 本文方法不仅原理简单,且计算量相对较小,在信噪比较低时检测准确度也能较高. 表 1 列出三种算法在相同背景噪声(Babble)不同信噪比环境中的检测率. 表 2 列出三种方法在同一信噪比不同背景噪声下检测率的比较. 图 4 所示为综合噪声下三种方法

平均检测率的对比图。

表 1 三种算法在相同背景噪声(Babble)
不同信噪比环境中的检测率

信噪比	检测算法		
	传统双门限法	短时 TEO 能量法	本文算法
无噪声	0.98	0.98	0.97
10 dB	0.89	0.89	0.92
5 dB	0.82	0.85	0.90
0 dB	0.76	0.82	0.87
-5 dB	0.66	0.79	0.83

表 2 三种方法在同一信噪比(-5 dB)
不同背景噪声下检测率的比较

算法	White	Babble	Factory	Volvo
双门限法	0.69	0.62	0.64	0.62
TEO 法	0.85	0.78	0.83	0.81
本文算法	0.86	0.88	0.90	0.85

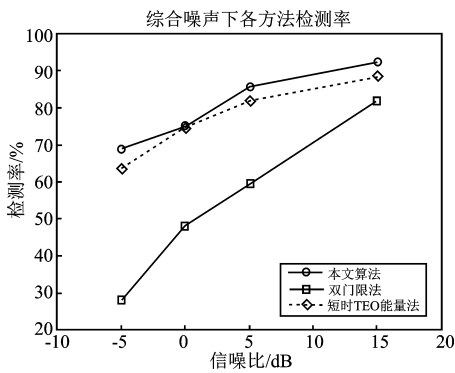


图 4 综合噪声下三种算法平均检测率的对比

由表 1 可得,在各性噪比下本文算法的检测率均比短时 TEO 能量和传统双门限要高,原因是在短时 TEO 能量条件下加入了 Mel 倒谱距离后,增加了抗噪性能,使得检测率得到提高.从表 2 可得到,本文方法在不同背景噪声下检测率也是最高的,且计算复杂度相对较低,相较于传统双门限法,本文所提算法的检测率有将近 26%的提升,但其检测复杂度基本相当.通过表 1、表 2、图 4,本文算法在各种不同环境下的检测率相对于传统双门限法有将近 15%的提升,相较于短时 TEO 能量法有将近 5%的提升,从而进一步说明本文所提算法是可行的.

5 结束语

本文通过对 Mel 倒谱距离和短时 TEO 能量进行分析,将二者组成联合特征,采用两级判决机制对语音进行端点检测.实验结果表明,联合特征弥补了单一特征抗噪性差的不足,得到了更高的检测率,鲁棒性更好.但由于本文选择的语音长度较短,对连续语音的检测率有待进一步提高,这将是下一阶段研究的重点.

参考文献:

[1] 赵力. 语音信号处理[M]. 北京:机械工业出版社,2012.

[2] 柯晶晶,周萍,景新幸,等. 差分 and 加权 Mel 倒谱混合参数应用于说话人识别[J]. 微电子学与计算机,2014,31(9):89-91.

[3] 陈振锋,吴蔚澜,刘加,等. 基于 Mel 倒谱特征顺序统计滤波的语音端点检测算法 [J]. 中国科学院大学学报,2014,31(4):254-259.

[4] Cho Namgook, Kim Eun-Kyoung. Enhanced voice activity detection using higher order statistics[J]. IEEE Trans Consumer Electronics,2011,57(1):196-202.

[5] 王伟,邓辉文. 基于 MFCC 参数和 VQ 的说话人识别系统[J]. 仪器仪表学报,2006,27(6):2253-2255.

[6] 刘伯森,卢志茂. 基于希尔伯特-黄变换的低信噪比语音端点检测[J]. 吉林大学学报:工学版,2011,41(3):844-848.

[7] Kaiser J F. On a simple algorithm to calculate the energy of a signal[C]// IEEE International Conference on Acoustics , Speech and Signal Processing. Albuquerque, NM :IEEE,1990:381-384.

作者简介:

王茂蓉 女,(1990-),硕士研究生.研究方向为语音识别与反蓄意模仿. E-mail:13737726138@163.com.

周 萍 女,(1961-),硕士,教授.研究方向为语音识别与智能控制.

景新幸 男,(1960-),博士,教授.研究方向为语音识别及其混合集成电路设计等.

杨 青 女,(1976-),硕士,副教授.研究方向为语音识别算法与仿真.