基于倒谱特征的带噪语音端点检测

胡光锐,韦晓东
(上海交通大学电子工程系,上海 200030)

摘 要: 在语音识别系统中产生错误识别的原因之一是端点检测有误差. 在高信噪比情况下,正确地确定语音的端点并不困难. 然而,大多数实际的语音识别系统需工作在低信噪比情况下,一些常规的端点检测方法,例如基于能量的端点检测方法在噪声环境下不能有效地工作. 本文利用倒谱特征来检测语音端点,提出了带噪语音端点检测的两个算法,第一个算法利用倒谱距离代替短时能量作为判决的门限,第二个算法改进了基于隐马尔柯夫模型(HMM)的语音检测以适应噪声的变化,实验结果表明本方法可得到高正确率的带噪语音端点检测.

关键词: 语音识别; 端点检测; 倒谱距离
中图分类号: TN912.34 **文献标识码**: A **文章编号**: 0372-2112 (2000) 10-0095-03

Endpoint Detection of Noisy Speech Based on Cepstrum

HU Guang-rui, WEI Xiao-dong
(*Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai* 200030, *China*)

Abstract: A major cause of errors in automatic speech recognition (ASR) systems is the inaccurate detection of the beginning and ending boundaries of test and reference patterns. Accurate determination of endpoints of speech is not very difficult if the SNR is high. Unfortunately, most practical ASR systems must work with a small SNR, and the conventional speech detection methods based on some simple features such as energy cannot work well in noisy environments. In this paper, cepstrum is used as the feature to detect the voice activity. Two algorithms for endpoint detection of noisy speech signal are proposed. The first one takes the cepstral distance as the decision thresholds instead of short-time energy. The second approach modified the HMM-based speech detector to make it adaptive to the change of noise. The experiments show high accurate rates can be obtained.

Key words: speech recognition; endpoint detection; cepstral distance

1 引言

在语音识别系统中,正确确定语音段端点不仅减少计算量,而且可以提高语音识别的正确率,语音端点检测也是语音自适应增强算法与语音编码系统的重要部分. 语音端点检测的传统方法通常采用语音的短时能量,这些方法在高信噪比(SNR)时具有良好的性能,而在低信噪比时性能很差. 然而,语音处理系统通常工作在不同的噪声环境下,在语音处理系统中采用的端点检测应当适应最不利的情况,在实际应用中达到好的性能. 本文给出了两种基于倒谱的语音端点检测方法. 由于倒谱距离对声环境具有更好的鲁棒性,因此第一种方法采用倒谱距离来代替短时能量作为检测门限,第二种方法是基于 HMM 的语音检测的改进方法,这种方法采用更新噪声模型来改进算法性能. 实验结果表明,本文提出的算法具有优越的性能.

2 语音端点检测方法

语音端点检测的算法步骤如下:

(1) 语音信号分成相邻有重迭的语音段,称为语音帧;

(2) 对每一语音帧,选取一种特征向量;

(3) 采用一种判决准则,例如门限判定或模式分类,来检测语音帧与非语音帧;

(4) 对上述结果进行后处理,得到语音的全部区间.

2.1 基于能量的端点检测

通常的语音端点检测方法采用测试信号的短时能量或短时对数能量作为特征参数,并采用门限判定法来检测语音[1,2]. 在这些方法中,当测试信号帧的短时能量超过噪声能量门限并持续一段时间,则第一次超过能量门限的点被判定为语音段的起点. 而当测试信号帧的能量低于另一个噪声能量门限并持续一定时间,就可测定语音段的终止点. 噪声能量门限的估计对这种方法的性能影响很大.

在低噪声环境下,如 SNR 大于 20dB 时,这种方法具有很好的性能. 然而,实际的语音识别系统常应用于不同的环境. 例如,在汽车中 SNR 通常只有几个 dB. 在低 SNR 环境下,由于难以确定适当的门限值,基于能量的端点检测不能很好地工作,这种方法也会舍弃一些低能量的清音语音(摩擦音,如/f/,/s/),而且,难以处理非平衡噪声. 在有些算法中,一些其它

收稿日期:1999-06-03;修回日期:2000-04-24
基金项目:国家自然科学基金(No. 69272007)

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的特征参数,如过零率,音调等被采用,以改进端点检测,但这些算法在高噪声情况下仍然不具备好的性能.

2.2 基于倒谱的端点检测

倒谱能很好表示语音的特征,因此在大多数语音识别系统中选择倒谱系数作为输入特征矢量[3]. 在噪声环境下,短时能量与其它特征参数都不能很好地区分语音段与非语音段,因此采用倒谱系数来作为端点检测的参数.

2.2.1 倒谱距离测量方法 信号的复倒谱定义为信号能量谱密度函数 $S(w)$ 的对数的傅里叶级数,$\log S(w)$ 的傅里叶级数表示式为

$$\log S(w) = \sum_{n=-\infty}^{+\infty} c_n e^{-jnw} \tag{1}$$

式中 $c_n = c_{-n}$ 为实数,通常称为倒谱系数,且

$$c_0 = \frac{1}{2\pi}\int_{\pi}^{\pi} \log S(w)\, dw \tag{2}$$

对于一对谱密度函数 $S(w)$ 与 $S'(w)$,利用 Parseval 定理,用谱的倒谱距离表示对数谱的均方距离为:

$$d_{cep}^2 = \frac{1}{2\pi}\int_{\pi}^{\pi} |\log S(w) - \log S'(w)|\, dw = \sum_{n=-\infty}^{+\infty} (c_n - c_n')^2 \tag{3}$$

式中,c_n 与 c_n' 分别表示谱密度 $S(w)$ 和 $S'(w)$ 的倒谱系数. 对数谱的均方距离表示两个信号谱的差别,故可用来作为一个判决参数. 实际上,由于 c_0 包含信号能量信息,基于能量的端点检测可以看作倒谱距离的一个特例.

倒谱距离的测量法步骤类似于基于能量的端点检测,但将倒谱距离代替短时能量来作为门限. 首先,假定前几帧信号是背景噪声,计算这些帧的倒谱系数矢量,利用前几帧倒谱矢量的平均值可估计背景噪声的倒谱矢量,噪声倒谱矢量的近似值可按下述规则进行更新,即当前帧被认为是非语音帧:

$$\bar{c} = p\bar{c} + (1-p)\, c_t \tag{4}$$

式中 $\bar{c}$ 为噪声倒谱矢量的近似值,c_t 为当前测试帧的倒谱矢量,p 为调节参数.

式(3)表示的倒谱距离可以利用式(5)来近似计算[3]:

$$d'_{cep} = 4.3429\sqrt{(c_0 - c_0')^2 + 2\sum_{n=1}^{p}(c_n - c_n')^2} \tag{5}$$

式中 c_n' 为对应于 $\bar{c}$ 的噪声倒谱系数,计算所有测试帧与背景噪声之间的倒谱距离可得到倒谱距离轨迹. 类似于基于能量的端点检测过程,利用倒谱距离轨迹可检测语音的端点.

图 1 给出一带噪语音信号的倒谱距离轨迹与短时能量曲线比较的例子. 一段录有三个词的语音被人为地加入白噪声,原始语音几乎被白噪声掩蔽,其信噪比小于 - 10dB,在如此恶劣的情况下,利用倒谱距离轨迹仍可找到词与词、词与背景噪声之间的边界,而基于短时能量的方法无法作为判决的准则.

事实上,这一方法类似于基于能量的检测器,仍籍助于门限判决,门限的选择对性能有重要影响. 当语音信号存在严重的谱失真时会给端点检测带来困难,难以选择适当的门限. 另外,当存在非平稳噪声,例如开关门的声音,电话铃声及其它声音时,倒谱距离很小以致于难以区分处理语音与非平稳噪声. 为了克服以上这些缺点,将采用模式分类法进行端点检测[2].

2.2.2 基于改进的 HMM 的端点检测 语音端点检测可采用统计模式分类方法,隐马尔柯夫模型(HMM)也可以象倒谱系数那样作为语音特征的统计模型. 基于 HMM 的方法已进行试验以用于语音端点检测,并得到高的正确率[2]. 原始的想法来自一些词的分割算法,在一段不受限制的语音中检测出一个或多个词的存在位置. 语音端点检测可看作为分割词的一个特例. 在这种 HMM 语音检测器中,一个为词作标记的连续 HMM 和一个为背景噪声作标记的连续 HMM 被训练来分别表示一般语音与噪声的特征. 训练采用基于 Baum-Welch 算法的倒谱向量来进行. HMM 与一个语法模型相连接,在端点检测阶段,对带噪语音进行预处理以得到输入特征矢量,每一矢量由倒谱系数,倒谱系数的增量或时间导数以及当前帧的短时能量增量等组成,然后引入 Viterbi 解码,按照模型参数与输入语音特征流得到与正发生的语音非常相似的语音,Viterbi 解码器给出语音的端点. 这种方法的基本系统结构与通常的语音识别器相同[3,4].

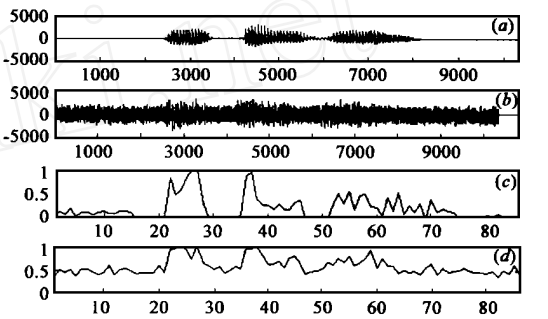

图 1 带噪语音倒谱距离轨迹与短时能量曲线的比较

(a) 原始语音波形; (b) 加入白噪声后的带噪语音波形图; (c) 带噪语音归一化倒谱距离轨迹图; (d) 带噪语音归一化短时能量曲线

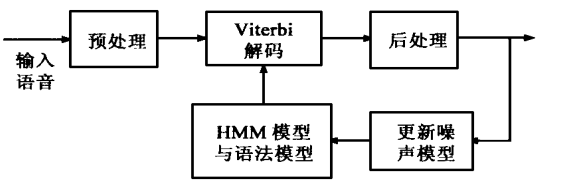

图 2 一种改进的 HMM 语音端点检测器

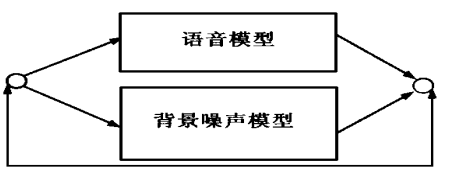

图 3 采用的语法模型

本文提出一种改进的 HMM 语音端点检测器,新方法采用一种更新噪声模型来改进系统的性能. 这种检测器示于图 2,系统中采用的语法模型示于图 3. 在图 2 中,预处理器给出输入倒谱向量流,Viterbi 解码方案按照语法模型及背景噪声模型参数来识别一段语音. 然后,后处理器通过状态序列从

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Viterbi 解码器得到处于噪声状态的语音帧,检测到的第一个小段被设置为语音的起点,而最后一个小段被设置为语音的终点.

在实际应用时,背景噪声是变化的.当预训练模型和实际情况失配时,将引入误差.为了减少这些误差,噪声模型将对处理作重新估计.在给定时间 T 内,根据先前在解码时确定的噪声帧所提供的信息来更新噪声模型.假定噪声 HMM 模型在 M 个混合高斯概率密度函数时具有一个状态,更新规则如下:

(1) 给定时间 T,从已经检测白噪声帧计算平均倒谱向量 $\bar{c}$;

(2) 在观察概率密度函数的 M 个混合高斯概率密度函数中找到一个均值 $\bar{\mu}_i$,其序号为 i,和 $\bar{c}$ 具有最小汉明距离的函数,可利用式(5)计算距离;

(3) 通过更新第 i 个观察概率密度函数 $\bar{\mu}_i$ 来重新估计噪声模型,即

$$\bar{\mu}_i = p\bar{\mu}_i + (1 - p)\bar{c} \qquad (6)$$

式中 p 为一调节因子,类似于式(4).通过更新,背景噪声模型将和测试环境的动态变化更好地匹配.

3 实验结果

语音端点检测方法是在不同的噪声条件下进行测试的.首先,语音信号经 8000Hz 抽样和 16bit 量化后,与不同电平的白噪声或实际的汽车噪声相混合.在所有实验中,语音信号被分为 240 采样的帧,相邻帧有 50 %重叠.每帧采用 12 阶 LPC 倒谱系数,对每个语音文本通过手工标号以区分语音与背景噪声,可作为测试端点检测正确率的标准.表 1 给出实验结果的简表.

表 1　语音端点检测测试实验结果

检测器	正确率	SNR = 15dB (白噪声)	SNR = 5dB (白噪声)	SNR = 0dB (白噪声)	SNR = 5dB (汽车噪声)
Energy	P(A/S)	0.98	0.76	0.64	0.70
	P(A/N)	0.99	0.60	0.51	0.50
	P(A)	0.98	0.71	0.59	0.61
CDM	P(A/S)	0.99	0.96	0.92	0.92
	P(A/N)	0.99	0.80	0.70	0.76
	P(A)	0.99	0.90	0.81	0.86
HMM	P(A/S)	0.97	0.94	0.94	0.89
	P(A/N)	0.98	0.73	0.69	0.60
	P(A)	0.97	0.87	0.95	0.81

在表 1 中,Energy 表示基于能量对数的端点检测器,CDM 表示基于倒谱距离测量的端点检测器,HMM 表示基于 HMM 的端点检测器,P(A/S) 表示语音检测的正确率,P(A/N) 表示非语音检测的正确率,P(A) 表示总的检测正确率.

4 结论

基于倒谱特征的语音端点检测方法在不利的环境下比通常的基于能量的端点检测方法的鲁棒性好.这一特性使其适合实际应用的需要,如在噪声环境下的语音增强与鲁棒语音识别等.

参考文献:

[1] S. Van Gerven, Fei Xie. a comparative study of speech detection methods [A]. EUROSPEECH'97 [C], 1997.

[2] Gregory B. Tucher, A. S. Spanisas, P. C. Leizou. An HMM-based endpoint detector for computer communication applications [EB]. From http://www.eas.asu.eud/~trcsip.

[3] L. R. Rabiner, B. H. Juang. Fundamentals of speech recognition [Z]. Murray Hill, New Jersey, USA, 1993.

[4] C. H. Lee, F. K. Soong, K. K. Paliwal. Automatic Speech and speaker recognition-advanced topics [M]. Kluwer Academic Publishers, 1996.

作者简介:

胡光锐　上海交通大学电子工程系教授,博士生导师,主要的研究兴趣包括:语音识别,神经网络,通信系统抗干扰研究,协同学和混沌神经网络在语音识别中的应用等.

韦晓东　1999 年 7 月获上海交通大学电子工程系通信与信息系统博士学位.目前研究方向包括:语音识别,发言者识别及鲁棒语音识别等.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net