一种基于基音频率的实时性端点检测方法

A Real-time Endpoint Detection Method Based on Pitch Frequency

王秀坤[1]， 李宁[2]， 魏燚潍[2]，戴维[3]
WANG XIUKUN LI NING WEI YIWEI DAI WEI

(1大连理工大学计算机科学与工程系,2大连理工大学软件学院,3大连泰康科技有限公司)

摘要：端点检测是语音识别中的一项关键技术，端点检测的准确性对语音识别的结果有很大影响。本文提出一种引入自适应门限的基于基音频率的检测算法，并对文中提及的几种算法的实时性进行了实验与分析。大量实验表明，在低信噪比下，该算法具有较高的准确率和很好的实时性，适用于鲁棒性、实时性的语音识别。

关键词：端点检测；基音频率；自适应门限；实时性

中图分类号：TP391.4 **文献标识码**：A

Abstract: Endpoint detection of speech is a curial problem in speech recognition. Endpoint detection has an important effect on speech recognition. This paper suggests an effective endpoint detection algorithm based on pitch frequency. Self-adaptive threshold are applied to the algorithm. This paper also analyzes real-time performance of the algorithm referred. Experimental results show the proposed algorithm has high accuracy and good real-time performance under lower SNR. The algorithm can be available in robust and real-time speech recognition.

Keywords: Endpoint detection; Pitch frequency; Self-adaptive threshold; Real-time

1 引言

语音的端点检测在语音识别和语音编码中起着重要的作用，准确实时的端点检测可以减少运算量和处理时间，从而直接影响着后续工作的效率。常用的端点检测方法有能量结合过零率方法、谱熵法、短时频带方差法[1,3,4]、基于自相关函数分析方法、LPC 系数法[2]以及多特征联合的方法[4]等等。这些方法在无噪声或是信噪比较高的情况下表现出很高的准确率，但是在低信噪比下很多算法的性能急剧下降，这是因为包括短时能量，谱熵能量在内的很多语音的特征参数对噪声相当敏感。此外有些算法时间复杂度很高，需要大量烦琐的运算，虽然达到了一定的准确率，但却不满足实时性的要求。

本文经研究发现，几乎所有的端点检测算法都倾向于如何提高检测准确率，而忽视检测方法的实时性。检测的准确率是一个检测方法很重要的衡量标准，但对于很多语音识别系统，检测识别的速度也同等重要。故本文提出一种基于基音频率的检测方法，并对方法的准确性和实时性进行了分析和实验，通过与常用的检测方法相比较，发现该方法在低信噪比下仍然具有很高的准确率，同时具有很好的实时性，完全可以满足一般的鲁棒性，实时性语音识别。

2 算法原理：

2.1 短时能量和平均过零率

短时能量是语音的端点检测中一个常用的参数，计算短时能量可以用下面的公式[5]：

$$E_n = \sum_{m=-\infty}^{+\infty} [x(m)\omega(n-m)]^2 = \sum_{m=-\infty}^{+\infty} x^2(m)h(n-m) = x^2(n)h(n) \qquad (1)$$

其中 $h(n) = \omega^2(n)$，即它是数据窗的平方。

为了减少电源以及直流分量的影响，本文采用改进的过零率计算方法：

$$Z_n = \sum_{m=-\infty}^{+\infty} |\mathrm{sgn}[x(m)-th]-\mathrm{sgn}[x(m-1)-th]|\,\omega(n-m) \qquad (2)$$

其中 th 为很小的一个阈值。

2.2 基音能量参数的计算

设输入语音信号的最高频率为 F_{high} Hz,快速傅立叶变换点数为 N， E_n 为短时能量。由信号处理原理可知，傅立叶变换后频谱是关于中心频点对称的，即关于 $\frac{N}{2}+1$ 点对成,则频率分辨率 $R_{speech}=\frac{M}{\frac{N}{2}}=\frac{2}{MN}$，令基音频率范围为 $F_{\min}-F_{\max}(F_{\min}=60,F_{\max}=480)$，可计算出每帧信号对应基音频率的点数范围是 $\left\{\left\lfloor\frac{F_{\min}}{R_{speech}}\right\rfloor,\left\lfloor\frac{F_{\max}}{R_{speech}}\right\rfloor\right\}$,计算可得每帧基音子带能量为：

$$E_{frame}=\sum_{n=\left\lfloor\frac{F_{\min}}{R_{speech}}\right\rfloor}^{\left\lfloor\frac{F_{\max}}{R_{speech}}\right\rfloor}E_n=\sum_{n=\left\lfloor\frac{F_{\min}*MN}{2}\right\rfloor}^{\left\lfloor\frac{F_{\max}*MN}{2}\right\rfloor}E_n \qquad (3)$$

2.3 自适应门限值的更新

设 $V(i)$ 是保存语音判别结果的向量，信号输入开始后初始化高门限和低门限分别为：$E_{thresholdhigh}=\lambda*NE$， $E_{thresholdlow}=\delta*NE$ （λ，δ 为经验值，需要经过大量实验得出，文中取 λ=1.40，δ=1.01， NE 为初始噪声能量，为前 10 帧基音子带能量的平均值)，即：

$$NE=\frac{1}{10}\sum_{m=1}^{10}\sum_{n=\left\lfloor\frac{F_{\min}*MN}{2}\right\rfloor}^{\left\lfloor\frac{F_{\max}*MN}{2}\right\rfloor}E_n \qquad (4)$$

初始化门限后检测时需要对噪声值门限 NE 进行更新，更新的方法为：

$$\begin{cases} NE=\beta*NE+(1-\beta)*E_{frame}(i) & E_{frame}(i)<E_{thresholdlow} \\ NE=\alpha*NE+(1-\alpha)*E_{frame}(i) & E_{thresholdhigh}>E_{frame}(i)\geq E_{thresholdlow} \end{cases} \qquad (5)$$

公式（5）中 α 和 β 为调节系数，合适的选取有助于得到更优的门限值，本文中取 $\alpha=0.1$， $\beta=0.9$。$E_{frame}(i)$ 为基音能量参数，可通过公式（6）计算。

$$E_{frame}(i)=\sum_{n=\left\lfloor\frac{F_{\min}*MN}{2}\right\rfloor}^{\left\lfloor\frac{F_{\max}*MN}{2}\right\rfloor}E_n \qquad (6)$$

2.4 语音帧的判定

$$\begin{cases} V(i)=1 & E_{frame}(i) \geq E_{thresholdhigh} \\ V(i)=0 & E_{frame}(i) < E_{thresholdhigh} \end{cases} \tag{7}$$

由于很多情况下语音都是以清音开始，清音能量很小，这时如果利用能量检测会出现漏检，所以如检测到第一帧语音信号，可以向前回溯（一般取 7-12 帧，实验中取 10 帧），利用过零率判断是否存在清音。当基音能量判断出第一帧语音帧 $E_{frame}(i)$ 之后，回溯前 10 帧计算过零率 $ZCR(j), j=i-1,i-2,.....i-10$，如果 $ZCR(j) > ZCR_{th}$，其中 ZCR_{th} 为过零率门限，则 $V(j)=1$。由于在检测过程中对于可能是语音帧的情况在向量中都置为 1，所以在语音信号输入完毕后，要对 $V(m)$ 向量进行中值滤波，以保证检测结果的准确性，滤波公式如下：

$$\tilde{V}(m)=\frac{V(m-1)+V(m)+V(m+1)}{3} \tag{8}$$

对于中值滤波的结果，采取公式（9）进行更新，得到最后的语音帧向量：

$$\tilde{V}_{speech}(m)=\begin{cases} 1 & \tilde{V}(m) \leq 0.5 \\ 0 & \tilde{V}(m) > 0.5 \end{cases} \tag{9}$$

语音帧判定方法：

$$\tilde{V}_{speech}(m)=\begin{cases} \text{语音帧} & \tilde{V}_{speech}(m)=1 \\ \text{非语音帧} & \tilde{V}_{speech}(m)=0 \end{cases} \tag{10}$$

根据公式（10）即可判断出语音帧，算法过程如图 1：

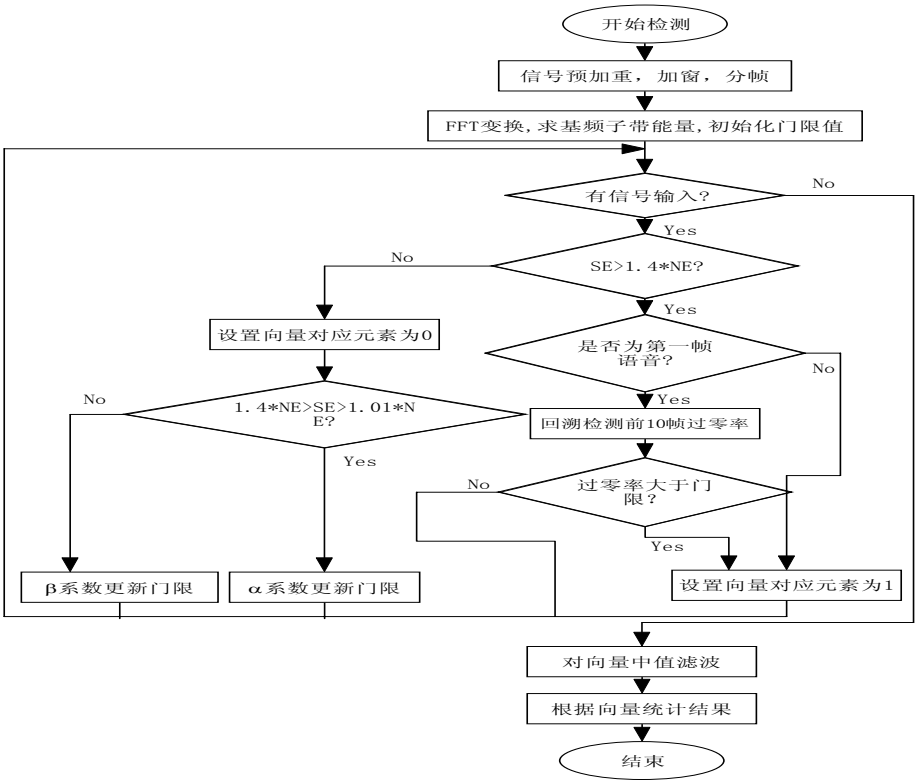

图 1 本文算法的程序流程图

3 实验结果及分析

3.1 准确性

实验中统计了在 20dB,10dB,5dB,0dB 信噪比下，本文算法与能量-过零率，频带方差方法的 matlab 仿真结果（见表 1），采样频率为 11.025KHz,16bit 量化，帧长 160，帧移 80，所用噪声来自 NOISE92 标准噪声库。由于语音识别系统准确性与多种因素相关，如用整个识别过程来检测其识别率会一定程度上受到其他部分算法效率的影响，故实验中采取手工标定端点来判断检测结果的准确性。表中检测结果单位为帧，ER 代表能量-过零率方法，TFV 代表频带方差法，PE 代表本文的基音能量法。在坦克噪声、工厂噪声和 F16 噪声下语音数据采用手工切分得出的真实语音段分别是：92-127，87-132，96-120。由表 1 可见，本文提出的基音能量方法与另两种方法相比，具有一定的鲁棒性。在高信噪比时，与两种方法检测效果基本相当，但随着信噪比降低，另外两种方法准确性下降很快，而本文所提方法则可以在很低的信噪比下达到很好的效果，这是因为充分利用了噪声和人的基音在频率分布上的差别。

表 1 不同信噪比下三种方法的检测结果

		20dB	10dB	5dB	0dB
坦克噪声	ER	94-123	95-125	90-130	-
	TFV	94-122	95-124	91-128	90-132
	PE	94-121	95-121	94-124	95-126
工厂噪声	ER	87-131	86-135	83-139	-
	TFV	87-131	87-132	86-135	84-136
	PE	87-132	87-132	87-134	86-135
F16 噪声	ER	96-118	93-123	92-127	-
	TFV	96-118	96-120	94-124	93-127
	PE	96-119	96-119	96-121	96-121

3.2 实时性

与其他方法如频带方差法相比，本方法无需计算频带方差这样复杂的参数，只需要能量参数即可，这极大的减少了计算量，加快了检测的速度。本文对实时性也进行了实验，对几种常用的检测方法执行时间进行了测量，发现本文方法要比方差法，谱熵法等快 30%左右，而一些算法例如频带方差法等等，虽检测效果基本接近，但速度明显过慢，无法满足一些对实时性要求较高的语音识别系统，本文所提出的方法在满足一定准确性的同时，可以更快的检测出语音的端点，目前已成功应用到开发的系统中。

4 结论

本文所提出的方法可以准确而且快速的检测出语音的端点，在工厂噪声，坦克噪声，f16 等噪声下都表现出了很好的效果，具有一定的鲁棒性和很强的实时性，可以应用到一些对实时性要求很高的语音识别系统中。

本文的创新之处：简化了传统基频检测中的复杂计算，保证了准确性的同时也有非常好的实时性。

参考文献

[1]王炳锡,屈丹,彭煊.实用语音识别基础[M].北京:国防工业出版社,2004.5.

[2]Qi.Y,Huntbr.Voiced-unvoiced-silence classification of speech using hybrid features and a network classifier[J].IEEE Transactions on Speech and Audio Processing,1993,1(2):250-255.

[3]李祖鹏,姚佩阳.一种语音段起止端点检测新方法[J].电讯技术,2000,21(3):68-70.

[4]江官星,王建英.一种改进的检测语音端点方法[J].微计算机信息,2006,5-1:138-139.

[5]焦蓬蓬,沈廷根,宋雪桦,吴斌.一种典型的语音端点检测方法的研究[J].微计算机信息,2008,2-1:217-218.

作者简介：王秀坤（1945.11-），女，大连理工大学，教授，博士生导师；李宁，男（1982.8-），大连理工大学，硕士研究生； 魏燚潍，男（1982.8-）大连理工大学，硕士研究生。戴维（1959.11-），男，大连理工大学，教授，博士，大连泰康科技有限公司董事长。

Biography: Wang XiuKun(1945.11-), Female, DaLian University of Technology, Professor, Doctoral supervisor.

（116023 大连 大连理工大学 电子与信息工程学院 计算机科学与工程系）王秀坤

（116023 大连 大连理工大学 软件学院 软件工程系）李宁 魏燚潍

（116023 大连 大连市高新园区黄浦路 624 号 泰康科技有限公司）戴维