

基于小波分析的语音端点检测算法研究与仿真

陈宝远, 梁伟明

(哈尔滨理工大学 测控技术与通信工程学院, 黑龙江 哈尔滨 150040)

摘 要: 以小波变换及多分辨分析为理论基础, 对语音端点检测中小波系数方差算法和子带平均能量算法进行了分析和研究, 利用语音和噪声的频域差别, 对这两种算法进行了优化, 并应用于端点检测系统中, 有效地改善了小波系数方差算法耗时长、实时性差的缺点, 并克服了子带平均能量算法只对高斯白噪声检测效果好的局限性, 提高了语音端点检测系统的实用性. 通过 MATLAB 软件仿真的实验结果表明, 采用优化算法的系统实现了语音端点检测准确性和快速性的最佳匹配, 达到了此类检测设备的实用要求.

关键词: 端点检测; 小波变换; 系数方差; 子带平均能量

中图分类号: TN93 **文献标识码:** A **文章编号:** 1007-2683(2009)01-0051-04

The Algorithm Research of Speech Endpoint Detection and Simulated Based on the Wavelet Analysis

CHEN Bao-yuan, LIANG Wei-ming

(School of Measure-control Technology and Communication Engineering, Harbin University of Science and Technology, Harbin 150040, China)

Abstract: This paper studies the algorithm of sub-band average-energy and the variance of the wavelet coefficients in speech endpoint detection, which were on the basis of the principle of Wavelet transform and Multi-resolution Analysis. It also optimizes the two algorithms. The proposed algorithm applied to the endpoint detection system improves the drawback of former algorithm in Long-Running and real-time badness which makes use of the variance of speech and noise in frequency characteristics. It also overcomes the limitation of only gaussian white noise is detected in the latter algorithm and improves applicability in speech endpoint detection. The simulation in the MATLAB software shows that system used optimization algorithm achieves optimal accurate and rapid matching. It meets the practical requirements of this kind of detecting equipment.

Key words: endpoint detection; wavelet transform; parameter variance; sub-band average-energy

1 引 言

语音端点检测的目的就是在复杂应用环境下, 从信号流中分辨出语音信号和非语音信号, 并确定语音信号的开始及结束. 小波理论的出现, 就可以实现将语音数据从时域变换到频域, 根据频域特征来

进行端点检测.

小波变换的基本思想是通过伸缩平移一组小范围的小波基, 使其时频窗口均随频率的变化而变化, 以实现低对低频分量采用大视窗分析, 对高频分量采用小视窗分析. 这种特性, 使小波变换对信号的处理具有自适应性. 小波变换中的多分辨分析 (Multi-resolution Analysis), 使小波变换在时域和频域都能

够表征信号的局部特征,可以对信号进行精细分析.

本文分析了语音端点检测中应用较广的子带平均能量算法和小波系数方差算法.在语音端点检测的实际应用中,前一种检测算法只对含有高斯白噪声的被测语音检测效果好,在含有其他噪声时检测准确性差,而后一种运算复杂,对系统要求高.所以,这两种算法都没有很好地解决应用设备对语音端点检测的实际要求.本文针对会场环境主要为白噪声(据统计,有色噪声出现的几率为5~10%)的特点,在对两种算法进行研究的基础上提出了优化算法,通过对信号能量的前期判断确定应用算法,设计出语音端点检测方案,不仅能正确地识别语音段的起始端点,而且减少了计算量,提高了运算速度,从而提高了检测系统的实时性.应用该方案能够有效解决音频信号的控制矩阵中准确性与快速性之间的矛盾^[1].

2 基于小波变换的端点检测算法

2.1 小波子带平均能量检测算法

如果待检测信号主要由白噪声组成,那么由于白噪声功率谱特性,这个能量序列应该为一个常数序列,其方差为零.如果信号 $x(m)$ 主要由语音组成,那么这个能量序列是一个数值变化的序列,其方差不会近似为零.如果被检测信号由白色噪声和语音组成,那么,不论白色噪声能量有多大,由于语音的存在,这个能量序列依然是一个数值变化的序列,其方差不会近似为零.因此,可以根据方差的大小来区分噪声帧和(含噪的)语音帧,从而检测语音段的起止端点^[2].

语音端点检测的门限值为

$$UT = 200 \frac{\sum_{i=1}^N x^2(n)}{\alpha} \quad (1)$$

其中: $x(n)$ 为离散语音信号时间序列; N 为帧长; UT 为端点检测的门限值; α 为测量的标准值.根据统计可知,绝大多数语音数据的能量分布在 $2 \times 10^6 \sim 8 \times 10^6$ 之间,因此取 $\alpha = 6 \times 10^6$.

根据上述思想,对每帧小波变换后的语音数据进行计算并判断.如果方差大于式(1)计算出来的门限值,即认为这一帧之后的语音数据为有效语音数据,记为起始帧.如果在起始帧存在的前提下,方差小于式(1)计算出来的门限值,就认为这一帧为结束帧.

小波子带平均能量方差的语音端点检测方法对语音信号中常见的高斯白噪声具有高的检测率,在信噪比降低的情况下也能得到较好的结果,而且具有算法简单、快速、系统资源占用小的特点.但它是针对白噪声设计的,在遇到其他突发性的噪声或有色噪声时并不能获得好的检测结果^[3].

2.2 小波系数方差的端点检测方法

语音信号是统计自相似的随机过程,它的统计特性在时域内不随波形的扩充或压缩而变化,因此它具有 $1/f$ 过程的特性.根据这一特性可知,对语音信号进行小波变换后,每一子带小波内的小波系数也一定具有相同的统计特性.因此,可以把经小波变换后获得的小波系数方差作为统计特性的参数,进行端点检测^[3].

在端点检测中,经常利用贝叶斯分类方法对小波系数方差的统计特性进行分类.假设有一离散语音信号为 $f[n]$,经小波变换后它的小波系数为 f_k ,它的方差如式(2)所示.

$$(\sigma_f)^2 = \frac{1}{N} \sum_{k \in N} (f_k - E(f_k))^2 \quad (2)$$

根据 $1/f$ 过程的特性,信号经过小波变换后获得的小波系数可以近似看作是零均值的随机变量,则式(2)变为

$$(\sigma_f)^2 = \frac{1}{N} \sum_{k \in N} (f_k)^2 \quad (3)$$

对噪音、清音和纯净语音分别提取小波系数方差作为先验知识,如式(4)所示.

$$(\sigma_n)^2 = \frac{1}{N} \sum_{k \in N} (n_k)^2, \quad (\sigma_q)^2 = \frac{1}{N} \sum_{k \in N} (q_k)^2, \quad (\sigma_c)^2 = \frac{1}{N} \sum_{k \in N} (c_k)^2 \quad (4)$$

式中: $(\sigma_n)^2$ 、 $(\sigma_q)^2$ 、 $(\sigma_c)^2$ 分别表示噪声、清音和不含噪声的纯净语音小波系数的方差, σ_n 、 σ_q 、 σ_c 分别表示噪声、清音和纯净语音的小波系数, N 为小波系数的总数.

V_0 表示预先提取的噪声的方差参数, V_1 表示预先提取的纯净语音的方差参数, V_2 表示清音的方差参数.根据贝叶斯分类原理,其分别如式(5)~(7).

$$P(\xi_k^m | V_0) = \prod_{m \in M} \prod_{k \in N(m)} P(\{\xi_k^m\} | V_0) = \prod_{m \in M} \prod_{k \in N(m)} \frac{1}{\sqrt{2\pi}(\sigma_n^m)^2} \exp \left\{ -\frac{(\xi_k^m)^2}{2(\sigma_n^m)^2} \right\} \quad (5)$$

$$P(\xi_k^m | V_1) = \prod_{m \in M} \prod_{k \in N(m)} P(\{\xi_k^m\} | V_1) =$$

$$\prod_{m \in M, k \in N(m)} \frac{1}{\sqrt{2\pi} (\sigma_c^m)^2} \exp \left\{ -\frac{(\xi_k^m)^2}{2(\sigma_c^m)^2} \right\} \quad (6)$$

$$P(\xi_k^m | V_2) = \prod_{m \in M} \prod_{k \in N(m)} P(\{\xi_k^m\} | V_2) = \prod_{m \in M, k \in N(m)} \frac{1}{\sqrt{2\pi} (\sigma_q^m)^2} \exp \left\{ -\frac{(\xi_k^m)^2}{2(\sigma_q^m)^2} \right\} \quad (7)$$

其中: M 为小波分层的层数; m 为各个不同的小波层; $N(m)$ 表示第 m 层小波系数的总数. $P(\{\xi_k^m\} | V_0)$ 是信号 $s(t)$ 为噪声的概率, $P(\{\xi_k^m\} | V_1)$ 是信号 $s(t)$ 为语音的概率, $P(\{\xi_k^m\} | V_2)$ 是信号 $s(t)$ 为清音的概率.

通过式 (5)、(6) 分别进行计算, 如果 $P(\{\xi_k^m\} | V_0) < P(\{\xi_k^m\} | V_1)$ 则判断为语音, 否则计算式 (7). 如果 $P(\{\xi_k^m\} | V_0) < P(\{\xi_k^m\} | V_2)$ 则标识此帧为语音, 否则标识此帧为噪声.

对于小波系数方差的语音端点检测算法, 适应性较好, 对于白噪声及有色噪声以及其他噪声都有较好的检测结果, 但算法复杂、系统资源占用较多, 对系统运算速度的要求很高, 很难达到实时性要求. 这样就限制了算法的应用^[4].

3 算法的优化

根据前面的分析可知, 这两种方法相比各有优缺点, 而且在不同的应用环境下能发挥各自的特点来完成端点检测. 因此本文提出一种语音端点检测的综合方法, 根据对噪声的估计来选择算法对语音进行端点检测, 利用子带平均能量方差简单高效的特点做静音条件下和平稳噪声条件下的端点检测, 而如果检测到信号中具有有色噪声或者其他噪声时, 采用小波系数方差的检测方法.

噪音环境可以简单地分为以下三类: 准静音环境; 平稳噪声环境; 有色噪声环境, 也称为非平稳噪声环境. 在准静音环境下, 频域内信号各频率分量幅度都很低. 设定信号的采样频率为 F , 平稳噪声的频谱幅度从 1Hz 到 $F/2$ 的分布没有大的变化. 而有色噪声的幅度从 1Hz 到 $F/2$ 的分布有很大变化, 与语音信号相似. 对语音的频谱特性和能量的描述如表 1 所示^[5].

根据上述分析, 本文提出的最终优化算法如下:

1) 对待测信号进行采样, 量化. 进行分帧处理, 分帧后各段信号如下:

$$x_1(m), x_2(m), x_3(m), \dots, x_N(m) \quad (8)$$

表 1 语音的频谱特性

语音分类	频谱 / kHz	能量情况
浊辅音 (浊音)	低频 0.1~0.4	较高
	中频 0.4~0.6	较高
元音 (浊音)	低频 0.1~0.4	较高
	中频 0.64~0.8	较低
清辅音	高频 > 3.5	较高

2) 对上述数据进行小波变换, 变换结果为

$$X_1(m), X_2(m), X_3(m), \dots, X_N(m) \quad (9)$$

3) 提取前五帧语音信号检测其频率分布, 计算各个频带能量幅值, 即

$$E_i = \frac{1}{N_i} \sum_{i=1}^N X_i^2(m) = C^2 \quad (10)$$

其中: N_i 是第 i 个信号 $X_i(m)$ 的长度, $i=1, 2, \dots, N$.

4) 如果检测到信号在频域内分布均匀或各个子带内的能量幅度值很小, 认为语音处于准静音或平稳噪声的环境下, 用小波子带平均能量的方法进行语音处理.

计算这个能量序列的均值 E 和方差 D , 即

$$E = \frac{1}{N} \sum_{i=1}^N E_i^2 \quad (11)$$

$$D = \frac{1}{N} \sum_{i=1}^N |E_i - E|^2 \quad (12)$$

其中: $E_i (i=1, 2, \dots, N)$ 表示信号 $x(m)$ 的小波子带平均能量序列^[6]. 将能量序列通过一个中心滤波器, 进行一次中心消波, 这样可以使得能量序列的特征更好地表现出来. 根据计算出来的方差, 利用式 (1) 确定的门限值判断语音和噪声, 并进行标识.

5) 如果检测到信号在频域内分布均匀或各个子带内的能量幅度值很大, 用小波系数方差的算法进行语音处理. 利用 2.2 节中的贝叶斯方法进行分类, 通过式 (7)、(8)、(9) 的概率判断含语音帧和非语音帧.

6) 对 4)、5) 检测所得的结果进行后续处理. 最小语音持续帧长为 8 帧, 最小噪声持续帧长为 5 帧, 一些界限如小于定义的最小持续时间时就被舍弃. 这样做可以减少语音端点的误检率.

为了增加算法的精确度, 对阈值进行了自适应处理, 实验采取语音信号前三帧子带频域方差的平均值的两倍作为阈值. 每当信号从语音段进入噪声段时, 重新取四帧语音的参数平均值作为阈值. 在不同的应用情况下, 也可以根据实际环境的背景噪声特性来确定阈值.

4 仿真及结果

实验是在不同的噪声条件下按照优化算法使用 MATLAB 进行仿真的. 首先, 使用软件对语音信号进行抽样和量化, 与白噪声和其他噪声相混合. 在所有实验中, 语音信号被分为每帧含有 220 样点的帧. 对每个语音文件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. 图 1 为未加噪声的语音信号, 图 2 为加入高斯噪声后的语音信号, 图 3 为加入混杂噪声后的语音信号, 图 4、图 5 为经过上述综合方法处理后的输出结果.

实验同时使用子带平均能量与小波方差系数方法, 以及上述优化综合的实现方法对相同语音进行了端点检测处理. 实验采集了大量语音, 在信噪比 $SNR = 15\text{db}$ 的条件下, 分别在纯净语音 (实验室采集)、高斯白噪声和混杂噪声的情况下进行检测. 检测率比较结果如表 2 所示^[7].

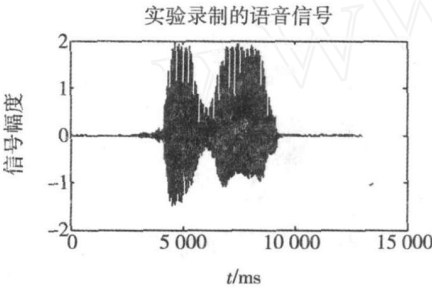


图 1 原始语音信号

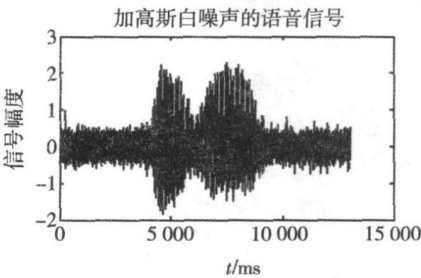


图 2 含高斯白噪声的语音信号

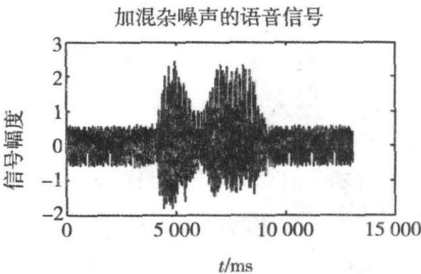


图 3 含混杂噪声的语音信号

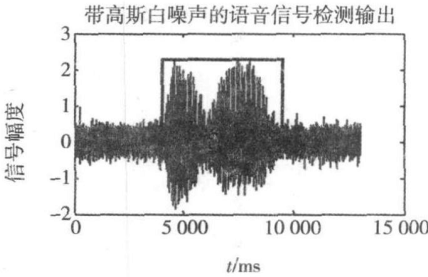


图 4 含高斯白噪声的信号检测

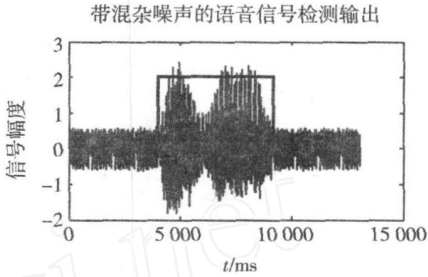


图 5 含混杂噪声的信号检测

表 2 语音端点检测率 (%)

检测算法	纯语音	含高斯白噪声	含混杂噪声
子带平均能量方差	98.3	84.3	75.1
小波系数方差	97.4	98.2	86.5
综合算法	98.2	97.9	86.3

5 结 语

从实验结果可以看出, 综合算法在对语音起止端点的检测中, 不仅减少系统的资源占用, 在 100 ms 的检测要求下, 对含混杂噪声信号的检测率高达 86.3% 以上, 准确性较原检测算法提高了 11.2%.

通过大量实验数据表明: 本算法充分利用子带平均能量方差算法和小波系数方差算法各自的优点, 实现了语音端点检测系统中语音起止端点检测实时性和准确性的最佳匹配, 达到了既节约系统资源, 又可取得可靠检测效果的目的.

参 考 文 献:

[1] 王 岩, 苏健民, 于慧伶, 等. 小波分析在语音识别系统中的研究与应用 [J]. 林业劳动安全, 2007 (1): 39 - 41.
[2] 董 力, 陈宏钦, 马争鸣. 基于小波变换的语音段起止端点检测算法 [J]. 中山大学学报 (自然科学版), 2005 (3): 116

表 3 向运行有 Echo, Finger服务的 RedHat 主机发送 Echo相关的攻击数据

状 态	误报率 / (%)
警报未经验证	90.2
警报经过验证	3.5

表 4 Snort和 Improved DS的比较

项目	规则数	检测时间 /s	检测速率 /k (bit/s)	误报率 / (%)
Snort	2 308	20	4.8	88.6
Improved DS 平均	1 200	16	8.2	5

表 1中,改进后的基于特征 DS的平均规则数来自统计平均值.这里本文运行 DS两天,每隔一小时统计一下当前实际匹配的规则数目.最后得到 DS的运行期间的平均规则数.检测速率定义为匹配速率与平均匹配规则数的比值.从实验结果中可以看到,改进后的 DS在检测速率和误报率方面都有很大提高.

5 结 语

本文从 DS要了解自己所保护的网路这一思想

出发,提出了一种结合网络安全扫描技术的 DS系统,对匹配规则进行精简,对警报进行验证,根据实验结果,该系统提高了匹配速率,同时也大大降低了误报率.下一步要做的工作是利用本文提出系统中的警报数据库进行更深一步的警报融合^[4],使得我们的 DS系统在进一步降低误报率和漏报率的同时,能够为网络管理员提供更高层次上的综合报警和预警.

参 考 文 献:

[1] 张绍莲. FICoM:一种防火墙和入侵检测协同工作的模型 [J]. 计算机工程与应用, 2002 (17): 187 - 189.
[2] 洪 宏,张玉清,胡予濮,等. 网络安全扫描技术研究 [J]. 计算机工程, 2004, 30 (10): 54 - 54.
[3] BRIAN Caswell, JAY Beale, JAMES C Foster, et al Snort2.0 Intrusion Detection [M]. 北京:国防工业出版社, 2004: 140 - 142.
[4] GULA R. Correlating DS Alerts with Vulnerability Information [R]. Tenable Network Security, 2002.

(编辑:付长缨)

(上接第 54页)

[3] 马道钧,陈天策,高 婕. 语音端点检测方法的分析与实现 [J]. 北京电子科技学院学报, 2007 (4): 67 - 68.
[4] ZHANG Xueying, ZHAO Zhefeng, ZHAO Gaofeng A Speech End-point Detection Method Based on Wavelet Coefficient Variance and Sub - band Amplitude Variance [J]. Proceedings of the First International Conference on Innovative Computing, Information and Control, 2006 (3): 83 - 86.
[5] 王达伟,曹俊兴. 一种基于噪音类型估计的端点检测系统 [J]

计算机工程与应用, 2005, 19: 87 - 89.
[6] 刘晓明,覃 胜,刘宗行,等. 语音端点检测的仿真研究 [J]. 系统仿真学报, 2005, 8: 1974 - 1975.
[7] 葛哲学,沙 威. 小波分析理论与 MATLAB R2007实现 [M]. 北京:电子工业出版社, 2007.

(编辑:付长缨)