·多媒体技术及应用· 文章编号:1000-3428(2016)12-0268-04 文献标志码:A 中图分类号:TP37

基于多尺度样本熵与阈值的语音端点检测

王 波,于凤芹

(江南大学 物联网工程学院,江苏 无锡 214122)

摘 要: 针对样本熵对突变噪声敏感导致的误检问题,提出一种改进的语音端点检测算法。该算法在时域采用尺度因子对语音信号进行多尺度变换,计算各尺度下的样本熵和阈值,统计样本熵大于门限阈值的尺度个数并与总尺度个数进行比较,实现语音端点检测。实验结果表明,该算法能够较好地消除样本熵对突变噪声的敏感性,并且与近似熵和样本熵检测算法相比,在低信噪比条件下具有更高的检测准确率。

关键词: 多尺度样本熵;多尺度变换;语音端点检测;阈值;近似熵

中文引用格式: 王 波,于凤芹. 基于多尺度样本熵与阈值的语音端点检测[J]. 计算机工程,2016,42(12):268-271.

英文引用格式: Wang Bo,Yu Fengqin. Speech Endpoint Detection Based on Multi-scale Sample Entropy and Threshold[J]. Computer Engineering,2016,42(12):268-271.

Speech Endpoint Detection Based on Multi-scale Sample Entropy and Threshold

WANG Bo,YU Fengqin

(School of Internet of Things Engineering,Jiangnan University,Wuxi,Jiangsu 214122,China)

【Abstract】 In order to overcome the defect that sample entropy can be falsely detected due to its sensitivity to the suddenly changing noise,this paper proposes a speech endpoint detection algorithm. This algorithm does the multi-scale transform for the speech signal in the time domain. The sample entropy and threshold of different scales can be calculated. The number of the sample entropy which is greater than the threshold of corresponding scale is counted and compared with the number of total scale to realize speech endpoint detection. Experimental results show that this algorithm can eliminate the mutation noise sensitivity of the sample entropy,and the detection accuracy is well improved in the low Signal Noise Ratio(SNR) conditions,compared with approximate entropy and sample entropy detection algorithms.

【Key words】 multi-scale sample entropy; multi-scale transform; speech endpoint detection; threshold; approximate entropy

DOI: 10.3969/j.issn.1000-3428.2016.12.045

0 概述

语音端点检测是计算机通过语音的声学特征将带噪语音区分为语音段和非语音段的方法,它被广泛应用于语音识别[1]、语音编码[2]、语音传输[3]和语音增强[4]等领域,是语音处理领域的基础。语音端点检测有很多种,目前主要分为2类:1)基于模型[5]的方法,通过对语音和噪声模型训练得到背景噪声和语音的统计信息。这类方法主要建立在理想的实验室条件下,要求预先得知背景噪声和语音的统计信息,但是实际情况是无法预知的。2)基于特征的方法[6],通过提取语音与背景噪声不同的声学特征来区分背景噪声和语音。这类方法主要有基于短时能量、短时过零率和谱熵方法。文献[7]针对能量特征易被噪声掩盖的特征,提出基于谱熵的端点检测方法,熵只与能量的随机性有关,而与能量的幅值无关,所以,可以很好地区分语音和非语音,对噪声具有一定的鲁棒性,而且避免了大量运算。文献[8]给出子带谱熵的概念,并结合自适应子带选择方法[9],提出一种自适应子带谱熵端点检测方法。针对现有特征方法在强噪声环境下无法准确检测端点的问题,文献[10]通过采用近似熵来表征语音的复杂程度,在强噪声环境下获得了较好的检测效果,但是一致性较差。文献[11]在近似熵基础上进行改进,提出一种不计算自身匹配的统计量,即样本熵。样本熵作为近似熵[12-13]上的改进算法,具有较好的一致

作者简介: 王 波(1991—),男,硕士研究生,主研方向为语音信号处理;于凤芹,教授、博士。

收稿日期: 2015-12-21 **修回日期:** 2016-01-25 **E-mail:** xzyzwb@163.com

性及对丢失数据的不敏感性。针对样本熵对突变噪声信号比较敏感的问题,文献[14]给出一种多尺度样本熵的交通流复杂性分析,通过对信号多尺度变换,较好地消除了突变噪声信号的影响。本文将多尺度样本熵引入语音端点检测,采用尺度因子提取语音分别在多个尺度上的特征并实现检测,通过综合多个尺度检测结果作为最终检测结果,以降低单一尺度检测带来的误差。

1　多尺度样本熵

1.1　语音多尺度变换

语音多尺度变换是通过不同的尺度因子将语音变换到不同尺度上。对于一个给定长度的语音序列 $u=\{u(1),u(2),\cdots,u(N)\}$,用尺度因子 τ 对其进行尺度变换,得到尺度 τ 下新的语音序列。

$$x_j^{\tau}=\frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau}u(i),1\leqslant j\leqslant N/\tau \tag{1}$$

通过式(1)可以计算各个尺度下的语音序列 x^{τ}。对于带噪语音来说,小尺度可比较完整地保存语音信息,大尺度可以通过相邻 τ 个数据点求均值来平滑突变信号。

1.2　多尺度样本熵的计算

多尺度样本熵[14]是样本熵的改进算法,在多个尺度下计算时间序列的样本熵。样本熵[11]是用非负数表示时间序列复杂度的特征,而多尺度样本熵是在多个尺度上反映时间序列的复杂度。

对于 τ 尺度下的语音序列 x^{τ},按照顺序组成 m 维矢量 $\boldsymbol{Y}^{\tau}(i)=[x^{\tau}(i),x^{\tau}(i+1),\cdots,x^{\tau}(i+m-1)]$,$i=1\sim N/\tau-m+1$。对每一个 i 计算矢量 $\boldsymbol{Y}^{\tau}(i)$ 与其余矢量 $\boldsymbol{Y}^{\tau}(j)$ 之间的距离。

$$d[\boldsymbol{Y}^{\tau}(i),\boldsymbol{Y}^{\tau}(j)]=\max_{k=0}^{m-1}[|x^{\tau}(i+k)-x^{\tau}(j+k)|] \tag{2}$$

按照给定的相似容限 $r(r>0)$,矢量 $\boldsymbol{Y}^{\tau}(i)$ 与其余矢量 $\boldsymbol{Y}^{\tau}(j)$ 之间的相似度 $B_i^m(r,\tau)$ 为:

$$B_i^m(r,\tau)=\frac{1}{N-m}num\{d[\boldsymbol{Y}^{\tau}(i),\boldsymbol{Y}^{\tau}(j)]<r\},$$
$$j=1,2,\cdots,N-m+1;j\neq i \tag{3}$$

其中,num 表示取数量,然后计算 τ 尺度下的相似度均值为 $B^m(r,\tau)$。

$$B^m(r,\tau)=\frac{1}{N-m+1}\sum_{i=1}^{N-m+1}B_i^m(r,\tau) \tag{4}$$

将维数 m 变成 $m+1$,重复以上过程,计算出 $B^{m+1}(r,\tau)$,最终得到 τ 尺度下的样本熵值。

$$SampEn(m,r,\tau)=\lim_{N\to\infty}\left\{-\ln\left(\frac{B^{m+1}(r,\tau)}{B^m(r,\tau)}\right)\right\} \tag{5}$$

在实际计算中,N 取有限值时,采用式(6)来计算样本熵:

$$SampEn(m,r,\tau,N)=-\ln\left(\frac{B^{m+1}(r,\tau)}{B^m(r,\tau)}\right) \tag{6}$$

由文献[15]可知,在一般情况下,$m=1$ 或 2,$r=(0.1\sim0.25)\delta$(其中,δ 是语音序列 x^{τ} 的标准方差)时计算得到的样本熵具有较合理的统计特性。基于此,本文计算样本熵取 $m=2$,$r=0.2\delta$。同时 N 在 100 ~ 1 000 范围内,样本熵是比较稳定的,因此,本文 N 取 512。

对于一个语音序列,如果它的复杂度高于另一个语音序列,则它在各个尺度下的样本熵都大于另一个语音序列的样本熵。同时,由文献[14]可知在小尺度下的样本熵对突变噪声信号比较敏感,随着尺度因子的增大,将会对突变噪声具有较好的平滑效果。对于本文来说,当尺度因子过大时,将会产生语音过平滑。因此,本文的尺度因子根据实验设置为 1,2,3,4,5。

2　基于多尺度样本熵的阈值计算

语音端点检测的准确性和门限阈值的准确确定有很大关系,太大或太小的阈值都会产生误判。为了更好地适应噪声的变化,门限需根据带噪语音的样本熵值来确定。一般来说采用双门限来检测端点。通过高门限检测起点,防止较低门限,导致将噪声误判为语音。通过低门限检测终点,防止较高门限,导致将语音误判为噪声。由于本文是根据语音的多个尺度计算样本熵,因此要根据不同的尺度设置不同的高低门限。将 τ 尺度下高低门限定义为:

$$\begin{aligned}T1_{\tau}&=\max(SampEn(m,r,\tau,N))\\&+\lambda_1[\max(SampEn(m,r,\tau,N))\\&-\min(SampEn(m,r,\tau,N))]\end{aligned} \tag{7}$$

$$\begin{aligned}T2_{\tau}&=\max(SampEn(m,r,\tau,N))\\&+\lambda_2[\max(SampEn(m,r,\tau,N))\\&-\min(SampEn(m,r,\tau,N))]\end{aligned} \tag{8}$$

其中,$T1_{\tau}$ 和 $T2_{\tau}$ 分别表示 τ 尺度下的高门限值和低门限值;max 和 min 分别表示取最大值和最小值;λ_1 和 λ_2 是可调参数,确定了 λ_1 和 λ_2 就确定了门限阈值 $T1_{\tau}$ 和 $T2_{\tau}$。本文经过大量实验表明,当 λ_1 和 λ_2 分别取 0.32 和 0.16 时具有较好的识别性能。

3　算法实现步骤

本文基于多尺度样本熵与阈值的语音端点检测流程如图 1 所示,具体步骤如下:

1)输入语音,帧长为 512,帧移为 200,对语音信号进行分帧,加汉明窗,计算总帧数和确定最大尺度因子 Max,本文最大尺度因子为 $Max=5$。

2)分别对每帧语音信号进行 1 ~ Max 的尺度变换,并计算每个尺度下的样本熵值为 $SampEn(m,r,\tau,N)$。

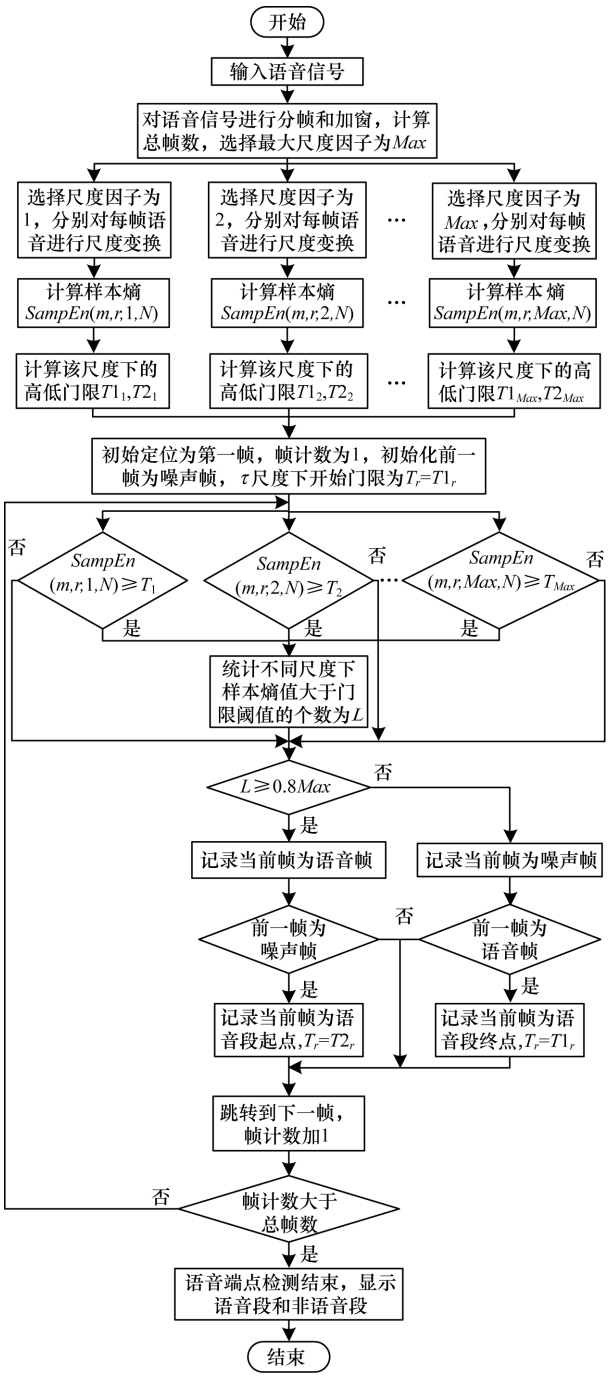

图 1 基于多尺度样本熵与阈值的语音端点检测流程

3)计算不同尺度下的门限阈值 $T1_\tau$ 和 $T2_\tau$,设定 τ 尺度下的开始门限 $T_\tau = T1_\tau$,初始化前一帧为噪声帧,初始定位为第一帧,帧计数为 1。

4)用不同尺度下的样本熵 $SampEn(m,r,\tau,N)$ 与对应尺度下的阈值 T_τ 进行比较,然后统计大于 T_τ 的尺度下样本熵的个数为 L。

5)比较 L 与 $0.8Max$ 的大小,Max 为总尺度个数。如果 $L \geqslant 0.8Max$,则表示当前帧为语音帧,跳转到第 6)步;否则当前帧为噪声帧,跳转到第 7)步。

6)判读前一帧是否为噪声帧。如果前一帧为噪声帧,则记录当前帧为语音段起点,同时尺度 τ 下的阈值 $T_\tau = T2_\tau$。判断结束后跳转到第 8)步。

7)判断前一帧是否为语音帧。如果前一帧为语音帧,则记录当前帧为语音段终点,同时在尺度 τ 下的阈值 $T_\tau = T1_\tau$。

8)跳转到下一帧,帧计数加 1,同时判断帧计数与总帧数的大小。如果帧计数大于总帧数,那么语音端点检测结束,显示语音段和噪声段;否则跳转到第 4)步。

4 仿真实验结果与分析

语音选自标准 TIMIT 语音库中 160 条语音,语音采样频率为 16 000 Hz,16 bit 量化。噪声选自标准 NOISEM_92 噪声库中的 White 噪声、Babble 噪声和 Volvo 噪声,分别生成 -10 dB,0 dB,5 dB 和 10 dB的带噪语音。尺度因子 τ 为 1,2,3,4,5。

图 2 给出了本文算法和文献[11]算法在 0 dB Volvo 噪声环境下的端点检测对比图。

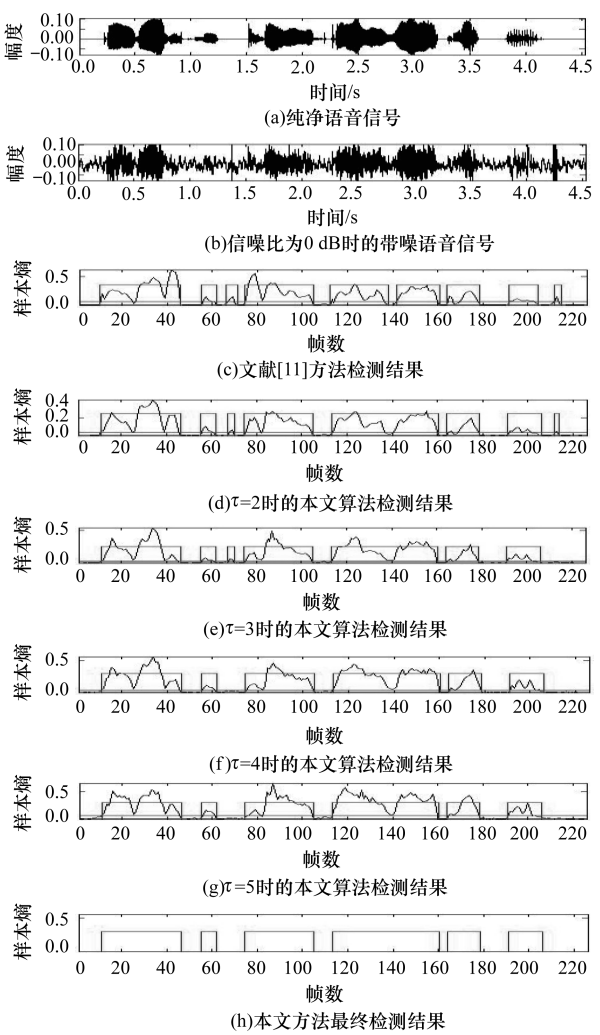

图 2 本文算法与文献[11]算法的端点检测比较

图 2(c)~图 2(h)中曲线表示对应尺度下的样本熵,直线表示对应的端点检测结果。图 2(d)~

图 2(g)分别为本文算法在尺度因子 $\tau=2,3,4,5$ 时的端点检测结果,当尺度因子 $\tau=1$ 时检测结果和文献[11]算法检测结果一样。图 2(h)为本文算法最终端点检测结果。由图 2(b) ~ 图 2(h)对比可知,文献[11]算法易受突变噪声影响将噪声误判为语音,并将语音分割为几段。而本文算法却能很好地区分语音和噪声,同时也可以看出本文算法随着尺度因子的增大对突变噪声具有较好的平滑效果。

表 1 给出了文献[10]算法、文献[11]算法和本文算法在 3 种不同噪声和 4 种信噪比条件下的检测准确率。由表 1 可知,本文算法相对于文献[10]算法和文献[11]算法在低信噪比条件下检测准确率有较大提升,这是因为通过引入多尺度变换消除了小尺度样本熵对突变噪声的敏感性,同时削弱了单一尺度样本熵计算带来的误差。

表 1　3 种算法在不同信噪比下的端点检测正确率

噪声	信噪比/dB	检测正确率/%		
		文献[10]算法	文献[11]算法	本文算法
White 噪声	-10	56.57	60.36	65.42
	0	73.46	78.65	82.56
	5	82.47	85.45	86.21
	10	91.56	90.32	90.45
Babble 噪声	-10	50.35	54.25	61.35
	0	67.84	70.21	77.43
	5	78.42	79.38	82.45
	10	85.31	86.54	87.56
Volvo 噪声	-10	55.43	61.53	67.86
	0	74.54	79.68	83.56
	5	82.86	86.34	88.25
	10	92.64	93.56	94.51

5　结束语

本文针对样本熵的语音端点检测算法对突变噪声比较敏感的问题,提出一种多尺度样本熵的语音端点检测算法。通过对语音信号进行粗粒化变换,计算多个尺度下的样本熵,来减弱样本熵对突变信号的敏感性,同时降低单一尺度样本熵计算带来的误差。仿真实验结果表明,本文算法能够较好地消除由突变噪声引起的误检,同时其在低信噪比条件下的检测准确率相对于样本熵算法有较大提升。下一步将研究尺度选择机制,通过自动选择尺度来降低算法计算量。

参考文献

[1] Dahl G E, Dong Yu, Li Deng, et al. Context-dependent Pre-trained Deep Neural Network for Large-vocabulary Speech Recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 30-42.

[2] Voran S, Catellier A. Multiple-description Speech Coding Using Speech-polarity Decomposition[C]// Proceedings of IEEE Global Telecommunications Conference. Washington D. C., USA: IEEE Press, 2010: 1-6.

[3] Lee B K, Chang J H. Packet Loss Concealment Based on Deep Neural Networks for Digital Speech Transmission[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2016, 24(2): 378-387.

[4] Rezayee A, Gazor S. An Adaptive KLT Approach for Speech Enhancement[J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(2): 87-95.

[5] Petsatodis T, Boukis C, Talantzis F, et al. Convex Combination of Multiple Statistical Models with Application to VAD[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(8): 2314-2327.

[6] Ghosh P K, Tsiartas A. Robust Voice Activity Detection Using Long-term Signal Variability[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(3): 600-613.

[7] Shen Jialin, Hung J W, Lee L S. Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments[C]//Proceedings of International Conference on Spoken Language Processing. Sydney, Australia: [s. n.], 1998: 232-238.

[8] Wu Bingfei, Wang Kun-ching. Robust Endpoint Detection Algorithm Based on the Adaptive Band-partitioning Spectral Entropy in Adverse Environments[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 762-775.

[9] Wu Gin-der, Lin Chin-teng. Word Boundary Detection with Mel-scale Frequency Bank in Noisy Environment[J]. IEEE Transactions on Speech and Audio Processing, 2000, 8(5): 541-554.

[10] 雷雄国,曾以成,李　凌. 基于近似熵的语音端点检测[J]. 声学技术,2007,26(1):121-125.

[11] 赵　欢,王纲金,胡　炼,等. 车载环境下基于样本熵的语音端点检测方法[J]. 计算机研究与发展,2011,48(3):471-476.

[12] Cai Chaofeng, Guo Shuting, Ren Jingying, et al. Approximate Entropy Analysis on the Electroencephalogram Signal Evoked by Mental Tasks[C]// Proceedings of IEEE Symposium on Electrical and Electronics Engineering. Washington D. C., USA: IEEE Press, 2012: 52-54.

[13] 洪　波,唐庆天,杨福生,等. 近似熵、互近似熵的性质、快速算法及其在脑电与认知研究中的初步应用[J]. 信号处理,1999,15(2):100-108.

[14] 向郑涛,陈宇峰,李昱瑾,等. 基于多尺度熵的交通流复杂性分析[J]. 物理学报,2014,63(3):1-9.

[15] Richman J S, Moorman J R. Physiological Time-series Analysis Using Approximate Entropy and Sample Entropy[J]. American Journal of Physiology Heart and Circulatory Physiology, 2000, 278(6): 2039-2049.

编辑　陆燕菲