第37卷 第14期 计 算 机 工 程 2011年7月
Vol.37 No.14 Computer Engineering July 2011

·人工智能及识别技术· 文章编号：1000—3428(2011)14—0140—03 文献标识码：A 中图分类号：TN912.3

基于噪声倒谱阈值频谱估计的语音活动检测

李 宇[1,2]，郭雷勇[1,2]，谭洪舟[2]

(1. 广东药学院医药信息工程学院，广州 510006；2. 中山大学信息科学与技术学院，广州 510275)

摘 要： 针对低方差频谱估计的语音活动检测(VAD)中Welch频谱估计方法计算量大的问题，提出利用倒谱阈值方法估计VAD中的噪声功率谱。该方法在静音时期为噪声的倒谱设置阈值，利用快速傅里叶变换计算频谱，再更新VAD中的判决阈值。算法复杂度分析与仿真结果表明，该方法的检测性能与Welch方法相当，计算量降低约18%，同时降低整个VAD的时间复杂度。

关键词： 语音活动检测；频谱估计；倒谱阈值方法；功率谱密度；快速傅里叶变换

Voice Activity Detection Based on Noise Cepstrum Thresholding Spectral Estimation

LI Yu[1,2], GUO Lei-yong[1,2], TAN Hong-zhou[2]

(1. College of Medical Information Engineering, Guangdong Pharmaceutical University, Guangzhou 510006, China;
2. School of Information Science and Technology, Sun Yat-sen University, Guangzhou 510275, China)

【Abstract】 Aiming at high computational complexity in Voice Activity Detection(VAD) which uses low variance spectral estimation, this paper proposes a method using noise cepstrum thresholding spectral estimation to compute noise Power Spectral Density(PSD). In speech absent period, a threshold is set for smoothing the noise cepstrum, then noise PSD is estimated by using Fast Fourier Transform(FFT), and detecting threshold in VAD system is updated. Computational complexity analysis and simulation results indicate that compared with Welch method, the computation of the method is reduced by 18%, and total time complexity of VAD is decreased.

【Key words】 Voice Activity Detection(VAD); spectral estimation; cepstrum thresholding method; Power Spectral Density(PSD); Fast Fourier Transform(FFT)

DOI: 10.3969/j.issn.1000-3428.2011.14.046

1 概述

语音活动检测(Voice Activity Detection, VAD)是指在噪声环境中检测语音的存在与否，通常用于语音编码、语音增强等语音处理系统中，起到降低语音编码速率、节省通信带宽、减少移动设备能耗、提高识别率等作用。早先具有代表性的VAD方法有ITU-T的G.729 Annex B。近年来，基于统计模型的VAD方法相继出现。文献[1]应用语音增强算法中的判决定向参数估计方法，利用似然比测试，提出基于统计模型的VAD算法。同类型的还有基于似然比测试的VAD改进算法[2]。文献[3]不考虑语音信号的统计信息，定义信噪比测量，利用一致最优测试确定阈值的自适应更新，用以检测活动语音。该方法不需要估计先验信噪比和后验信噪比，具有较低的计算耗时和更少的存储空间，其检测性能优于G.729 Annex B修正版。然而其使用的噪声模型为对称Gaussian模型，与某些噪声(如Babble)的信噪比测量统计分布的拖尾特性有一定的误差，因此，该方法不利于处理Babble等非平稳噪声。随后，文献[4]利用Rayleigh模型对噪声的功率谱密度(Power Spectral Density, PSD)直方图建模，给出新的判决阈值更新表达式。该方法适用于非平稳噪声，其正确识别率高于文献[3]的方法，并且复杂度较低。文献[3-4]这一类将低方差频谱估计和自适应阈值相结合的VAD方法需要对噪声PSD进行精确估计，因此，需要采用Welch等低方差估计方法，但Welch方法本身复杂度较高，且窗跨度的选择没有直观的方法可循。

本文应用倒谱频谱估计方法[5]对VAD中噪声的PSD进行估计，再利用高斯阈值更新准则[3]检测活动语音，实现了复杂度更低的VAD。

2 基于倒谱阈值的平滑非参数估计方法

文献[5]提出一种基于倒谱阈值的平滑非参数估计方法，它可以自动、有效地获得平稳信号的非参数频谱估计。

对于 $\forall p$，设 $\Phi_p > 0$ 是信号的功率谱密度。当 Φ_p 用周期图估计 $\hat{\Phi}_p$ 时，依据信号 $\{x(t)\}$ 倒谱的定义，得到倒谱 c_k 的估计值 $\hat{c}_k$：

$$\hat{c}_k = \frac{1}{N}\sum_{p=0}^{N-1}\ln(\hat{\Phi}_p)\,\mathrm{e}^{\mathrm{j}\omega_k p}, k = 0,1,\mathrm{L}\ ,N-1 \tag{1}$$

对于元素 $\hat{c}_k$，估计矢量 $\overline{\boldsymbol{c}} = \left[\overline{c}_0\overline{c}_1\mathrm{L}\ \overline{c}_{\frac{N}{2}}\right]^{\mathrm{T}}$ 与其有以下关系：

$$\begin{cases}\overline{c}_0 = \hat{c}_0 + \gamma \\ \overline{c}_k = \hat{c}_k\end{cases} \tag{2}$$

其中，$\gamma = 0.577\ 126\mathrm{L}$ 是欧拉常数；$k = 1,2,\mathrm{L}\ ,\frac{N}{2}$。

该方法与小波阈值有所不同，同样是平滑 $\{\ln\Phi\}$，但是通过倒谱估计 $\{\overline{c}_k\}$ 设置阈值起到平滑 $\{\ln\Phi\}$ 的作用。利用一致最优测试，可以推断 c_k 是否为0或接近于0，从而设置 $\overline{c}_k$ 为

基金项目： 国家自然科学基金资助项目(60874060)

作者简介： 李 宇(1977－)，男，讲师、博士，主研方向：语音、医学信号处理；郭雷勇，讲师、博士；谭洪舟，教授、博士、博士生导师

收稿日期： 2010-12-10 **E-mail：** liyu33@gmail.com

0。对倒谱估计 $\{\bar{c}_k\}$ 的估计量为：

$$\hat{c}_k=\begin{cases}0 & \text{if } |\bar{c}_k|\leqslant\dfrac{\mu\pi}{(d_kN)^{1/2}}\\ \bar{c}_k & \text{otherwise}\end{cases} \tag{3}$$

其中，$k=0$ 或 $k=\dfrac{N}{2}$ 时，$d_k=3$，$k=1,2,\text{L}\ ,\dfrac{N}{2}-1$ 时，$d_k=3$；μ 起到控制虚警率的作用，μ 的选择要依据经验，当 $N<500$ 时，对于窄带信号，采用 $\mu=2$，而对于宽带信号，采用 $\mu=4$。

频谱估计可以通过对 $\{\hat{c}_k\}$ 即设置阈值后的倒谱估计 $\{\bar{c}_k\}$ 进行快速傅里叶变换(Fast Fourier Transform, FFT)得到，如式(4)所示：

$$\hat{\Phi}_p=\exp\left[\sum_{k=0}^{N-1}\hat{c}_k\mathrm{e}^{-\mathrm{i}\omega_pk}\right],\ p=0,1,\text{L}\ ,N-1 \tag{4}$$

对 $\hat{\Phi}_p$ 进行缩放，得到：

$$\hat{\hat{\Phi}}_p=\hat{\alpha}\hat{\Phi}_p,\ p=0,1,\text{L}\ ,N-1 \tag{5}$$

其中，比例因子为：

$$\hat{\alpha}=\min_{\alpha}\sum_{P=0}^{N-1}\left(\hat{\Phi}_p-\alpha\hat{\Phi}_{pp}\right)^2$$

求解得：

$$\hat{\alpha}=\frac{\sum_{P=0}^{N-1}\hat{\Phi}_p\hat{\Phi}_P}{\sum_{P=0}^{N-1}\hat{\Phi}_P^2}$$

3 基于倒谱阈值噪声频谱估计的 VAD

3.1 信号条件分析

低方差频谱估计和自适应阈值相结合的 VAD 方法其运算主要集中在噪声帧和含噪语音帧的 PSD 计算。倒谱阈值平滑非参数估计方法相比周期图法具有更低的方差，适合计算噪声的 PSD。在语音增强、编码等处理中常需要计算噪声以及语音的倒谱。一般都假定噪声或语音的倒谱矢量服从具有零均值的高斯统计分布，这符合文献[6]提出的基于检测类方法的使用条件。

3.2 算法复杂度分析

对于 Welch 方法，将 N 点信号序列分为长 L 的子段，段间重叠 50%，则子段数为 $\dfrac{2N}{L}$。设 L 点的 FFT 需要乘法运算 $\dfrac{L\text{lb}(L)}{2}$ 次，由文献[7]可知，Welch 方法实时运算时共需要 $4N+L+N\text{lb}(L)$ 次乘法。倒谱阈值方法运算主要包括先对输入的 N 点信号序列求倒谱，再与阈值比较，最后由 FFT 计算得出频谱，由式(1)、式(4)、式(5)可得其需要 $6N+\dfrac{3N\text{lb}(N)}{2}$ 次乘法。在实际工程使用时，噪声倒谱可以直接从语音处理系统的其他模块获得，这样可以省去求倒谱的 $3N+N\text{lb}(N)$ 次乘法，因此，实际系统中需要的乘法运算为 $3N+\dfrac{N\text{lb}(N)}{2}$ 次。两者的乘法数量之差为：

$$4N+L+N\text{lb}(L)-3N-\frac{N\text{lb}(N)}{2}=N+N\text{lb}(\frac{\sqrt{N}}{10})+L \tag{6}$$

实际使用 Welch 方法(如 Matlab 的 welch()函数)一般分为 8 段，段间重叠 50%，则 $L=\dfrac{2N}{9}$，一般有 $L\geqslant\dfrac{N}{10}$，取 $L=\dfrac{N}{10}$，代入上式得 $N(\dfrac{\text{lb}(N)}{2}-2.222)$，要使该式大于 0，必须满足：$\text{lb}(N)>4.444$。

语音处理中用到的帧数 N 一般大于 80，则上式表明倒谱阈值方法需要的乘法数量少于 Welch 方法，所以，倒谱阈值方法的复杂度比 Welch 方法低。例如取 $N=160$、$L=16$，则实际运算时倒谱阈值方法与 Welch 方法的乘法数分别为 1 066 与 1 296，前者比后者节省约 18%的乘法运算量。

4 仿真实验及结果分析

实验采用倒谱阈值方法和 Welch 方法估计噪声频谱，结合 Gaussian(或 Rayleigh)模型更新准则，进行语音活动检测。整个检测流程如下：含噪语音帧信号通过倒谱阈值方法或 Welch 方法计算出信号的频谱，再计算信噪比测量，然后比较信噪比测量与平均阈值之间的大小。若信噪比测量大于平均阈值，则初判决为语音，反之，初判决为静音。初判决后再通过时滞策略得出 VAD 的最终判决。当最终判决为静音时，利用前面 2 种方法计算的噪声频谱更新平均阈值。流程中各变量的计算与具体的 VAD 信号框图可以参考文献[3-4]。

分别在 3 种信噪比及 3 种噪声环境下采用 Welch 方法和本文方法评测 VAD 的性能。实验用到的噪声库与语音数据及其格式设置与文献[4]相同。除了用于信噪比测量方差进行时间指数平均的 $\alpha_{var_{sk}(f_l)}=0.4$ 外，实验中 Welch 方法以及 VAD 中的参数设置与文献[3]相同。

利用倒谱阈值法求出的频谱有 80 维，与 Welch 方法不同，需要设置的参数如表 1 所示。

表 1 本文 VAD 方法的参数

参数	取值
阈值的最大值	3.20
阈值的最小值	0.65
PSD 均值最小值	0.000 1
信噪比测量的平滑因子	0.75
信噪比测量方差的平滑因子	0.35
阈值的平滑因子	0.75
虚警率 P_{FA}	0.05

性能评估沿用文献[3]的 5 个客观评测指标，分别为：

(1)正确检测率 *Correct*；

(2)前端语音判为噪声的误判率 *FEC*；

(3)中间语音判为噪声的误判率 *MSC*；

(4)尾端噪声判为语音的误判率 *OVER*；

(5)非活动时段的噪声判为语音的误判率 *NDS*。

在以上实验条件下评测 2 种噪声估计方法对 VAD 性能的影响和时间复杂度，实验结果如表 2、表 3 所示。

表 2 Welch VAD 方法的检测结果

测试环境		评测指标				
噪声	信噪比/dB	*Correct*/(%)	*FEC*/(%)	*MSC*/(%)	*OVER*/(%)	*NDS*/(%)
White	5	93.62	2.21	2.88	1.16	0.12
	10	94.51	1.88	1.67	1.79	0.14
	15	95.19	1.59	0.88	2.21	0.14
Babble	5	84.14	1.26	1.07	3.71	9.83
	10	87.98	1.26	0.42	3.56	6.79
	15	92.01	1.35	0.25	3.08	3.31
Factory1	5	82.21	1.43	2.61	2.96	10.8
	10	83.68	1.23	0.89	3.81	10.4
	15	86.63	1.13	0.32	3.83	8.08

表 3 本文 VAD 方法的检测结果

测试环境		评测指标				
噪声	信噪比/dB	*Correct*/(%)	*FEC*/(%)	*MSC*/(%)	*OVER*/(%)	*NDS*/(%)
White	5	92.88	2.35	3.84	0.80	0.13
	10	94.24	1.97	2.27	1.39	0.13
	15	95.04	1.70	1.26	1.86	0.14
Babble	5	87.18	1.48	2.38	1.97	7.00
	10	87.63	1.23	0.67	3.18	7.29
	15	86.79	1.01	0.20	4.21	7.78
Factory1	5	86.00	2.07	5.41	1.07	5.45
	10	88.96	1.64	2.02	1.94	5.44
	15	89.50	1.32	0.58	2.94	5.66

倒谱 VAD 方法在白高斯噪声环境下的 3 个信噪比都能获得与 Welch VAD 方法相当的 *Correct* 值；对 Babble 噪声在高信噪比情况下，反而下降严重，*Correct* 值在 15 dB 时有 5.21 个百分比的下降，这是由于 Babble 属于非平稳噪声；在 Factory1 噪声环境下，*Correct* 值在 5 dB~15 dB 时平均高出 4 个百分比。图 1~图 4 给出了采用本文方法分别对同一段含 Babble 噪声语音和 Factory1 噪声语音在 5 dB 信噪比环境下的检测输出和平均阈值更新情况。对于非平稳信号的 Babble 噪声和 Factory1 噪声，出现噪声误检的情况明显增加，这也可以从表 2、表 3 的 *OVER*、*NDS* 参数看出。

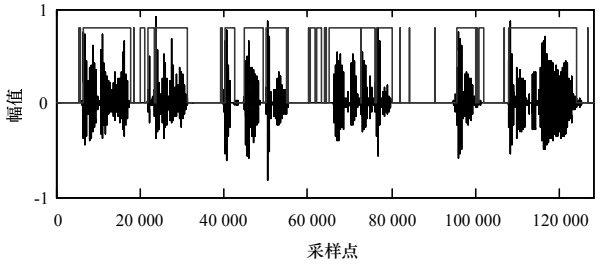


图 1 5 dB Babble 噪声下本文方法的 VAD 输出

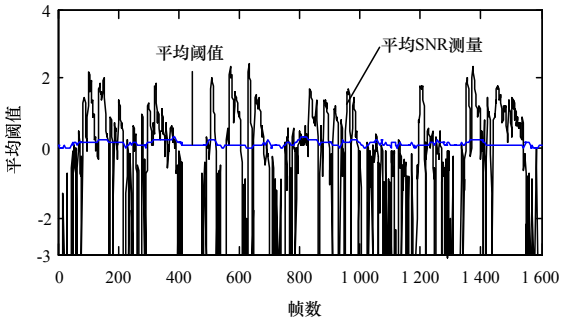


图 2 5 dB Babble 噪声下的平均阈值更新情况

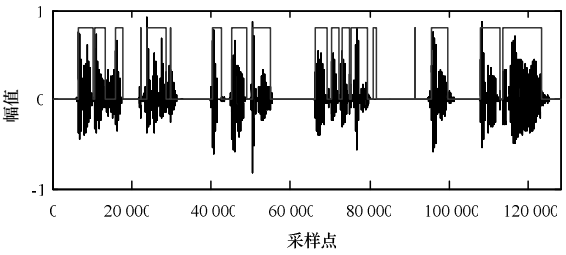


图 3 5 dB Factory1 噪声下本文方法的 VAD 输出

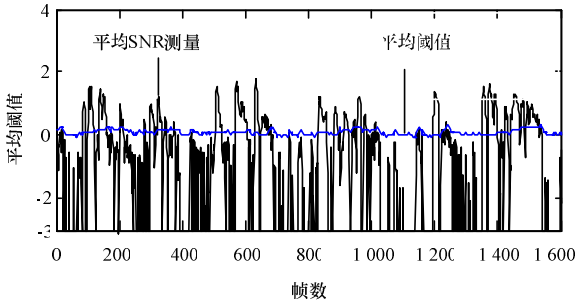


图 4 5 dB Factory1 噪声下的平均阈值更新情况

此外，还对两者的运算耗时进行了比较。在 Matlab2008a、Windows XP 和联想启天计算机(E2200 CPU、2 GB 内存)软硬件环境下，分别编程实现 Welch 方法的 miniWelch()函数、倒谱方法 Cepspec()函数和 Matlab 自带的 welch()函数。用这 3 个函数对 225 s 的 Babble 噪音进行频谱估计，分别耗时 1.653 s、3.368 s 与 132.493 s。实验结果与上一节的复杂度分析相符，验证了倒谱阈值噪声估计的低复杂度。welch()函数采用冗余设计，因此，运算耗时较长。

总的来看，2 种方法的检测性能相当，倒谱 VAD 方法有更低的算法复杂度。但其对于非平稳噪声性能不够理想，这是由于其阈值的确定是依据信号是平稳的并服从高斯分布这一假设，而非平稳噪声的分布与这一假设偏差较大。

5 结束语

本文采用倒谱阈值估计噪声频谱，并将其应用于基于统计模型自适应阈值更新准则的 VAD 中。由于实际语音系统中信号倒谱可从其他模块获得，因此提高了语音处理系统中倒谱运算的复用性。理论分析和仿真实验证明，该方法在不损失检测性能的同时减少了整个 VAD 系统的计算量。下一步将展开基于倒谱框架的语音增强算法研究，构建一个复用性高的语音处理系统。

参考文献

[1] Sohn J, Kim N S, Sung W. A Statistical Model-based Voice Activity Detection[J]. IEEE Signal Processing Letters, 1999, 6(1): 1-3.

[2] 李燕诚, 崔慧娟, 唐 昆. 基于似然比测试的语音激活检测算法[J]. 计算机工程, 2009, 35(10): 214-216.

[3] Davis A, Nordholm S, Togneri R. Statistical Voice Activity Detection Using Low-variance Spectrum Estimation and an Adaptive Threshold[J]. IEEE Trans. on Audio, Speech, and Language Processing, 2006, 14(2): 412-424.

[4] 李 宇, 陈建铭, 谭洪舟, 等. 基于 Rayleigh 噪声统计分布的有音区检测[J]. 信号处理, 2009, 25(11): 1809-1813.

[5] Stoica P, Sandgren N. Smoothed Nonparametric Spectral Estimation via Cepstrum Thresholding[J]. IEEE Signal Processing Magazine, 2006, 23(6): 34-45.

[6] Stoica P, Sandgren N. Total-variance Reduction via Thresholding: Application to Cepstral Analysis[J]. IEEE Trans. on Signal Processing, 2007, 55(1): 66-72.

[7] Welch P D. The Use of Fast Fourier Transform for The Estimation of Power Spectra: A Method Based on Time Averaging over Short Modified Periodograms[J]. IEEE Trans. on Audio and Electroacoustics, 1967, 15(2): 70-73.

编辑 张 帆