

# 基于能量熵统计模型的语音增强算法

邱文武, 蒋建中, 吴 琼  
(解放军信息工程大学信息工程学院, 郑州 450002)

**摘 要:** 针对幅度谱 LOG\_MMSE 估计算法在低信噪比时的残留音乐噪声问题, 提出基于能量熵统计模型的 LOG\_MMSE 估计, 推导其近似解。根据熵的差值动态调整加权因子, 进一步改善弱语音段的去噪效果。实验结果表明, 该算法在改善信噪比和残留音乐噪声方面均优于传统的幅度谱 LOG\_MMSE 算法。

**关键词:** 语音活动检测; 能量熵; 最小均方误差; 噪声估计

## Speech Enhancement Algorithm Based on Energy Entropy Statistics Model

QIU Wenwu, JIANG Jianzhong, WU Qiong  
(Information Engineering College, PLA Information Engineering University, Zhengzhou 450002, China)

**【Abstract】** Aiming at the residual musical noise problem which is caused by the amplitude spectrum LOG\_MMSE estimation algorithm at low SNR, this paper proposes a new speech enhancement algorithm based on energy entropy statistical model. The approximate solution is derived. Weighting factors are dynamically adjusted according to the difference between entropy, to further improve the weak speech segment denoising effect. Experimental results show that the algorithm in the SegSNR improvement and the residual musical noise are better than the traditional amplitude spectrum LOG\_MMSE algorithm.

**【Key words】** Voice Activity Detection(VAD); energy entropy; MMSE; noise estimation

### 1 概述

在日常生活中,人们说话时都会受到来自周围环境的噪声干扰,比如手机铃声、汽车鸣笛声和其他说话人的声音等。为了更清楚地获取说话者所表达的信息,就需要把语音中的各种干扰和噪声滤除。语音增强算法有许多种,其中基于统计模型的算法一直是国内外研究的热点,并且相关技术产品已经在3G手机、IP电话和医疗等方面应用。本文提出了基于能量熵统计模型的LOG\_MMSE算法,并动态调整加权因子,改善了低信噪比时LOG\_MMSE的残留音乐噪声问题。

### 2 基于能量熵统计模型的语音增强

#### 2.1 幅度谱 LOG\_MMSE 估计

高斯统计模型的语音增强算法中假设语音(噪声)的频域幅度谱或者频域的实部和虚部服从高斯分布,且语音和噪声相互独立,而后根据一定的先验知识,采用ML算法、MMSE等方法从带噪语音中估计出纯净语音的频谱幅度或者是其实部和虚部,从而起到语音增强的目的。文献[1]使用MMSE对语音幅度谱进行估计,证明对数域MMSE更适合语音增强。文献[2]表明在同样的统计模型下,幅度相位模型的估计效果要略好于实部和虚部模型,由于可以直接使用带噪语音的相位作为纯净语音的相位估计,因此先对幅度进行估计,而后引入带噪语音的相位,合成增强的语音信号。

假设采样后的带噪语音信号为 $y$ ,由纯净语音信号 $x$ 和不相关的加性噪声 $n$ 组成。对其进行加窗分帧。对每个窗内的信号进行傅里叶变换可知,第 $n$ 帧 $k$ 个点频处的傅里叶变换为:

$$Y(k, n) = X(k, n) + N(k, n) \quad 1 \leq k \leq L \quad (1)$$

其中, $L$ 为帧长。假设语音和噪声都服从高斯分布,采用LOG\_MMSE对幅度进行估计。

可得纯净语音的估计值为:

$$\hat{X}_k = \frac{\xi_k}{1 + \xi_k} \exp\left(\frac{1}{2} \int_{\lambda_k}^{\infty} \frac{e^{-t}}{t} dt\right) Y_k \quad (2)$$

其中, $\lambda_k(k)$ ,  $\lambda_d(k)$ 为语音和噪声的方差; $\lambda_k = \frac{\xi_k}{1 + \xi_k} \lambda_d$ ,  $\xi_k$ 和 $\lambda_k$ 分别为先验和后验信噪比估计,可以通过DD算法求得。经典的DD算法的先验信噪比估计形式为:

$$\hat{\xi}_k = \max[a \frac{\hat{X}_k^2(n-1)}{\lambda_d(k, n-1)} + (1-a) \max[\lambda_k(m) - 1, \xi_{\min}]] \quad (3)$$

其中,求先验和后验信噪比需要知道当前帧和前一帧的噪声方差,噪声估计的方法有很多,一种方法是采用VAD算法<sup>[3-4]</sup>,区分出语音段和噪声段,而后在噪声段进行噪声估计更新。另一种是可以在全程对噪声进行跟踪估计,但是会引入一定的延迟误差。本文采用文献[3]提出的基于子带熵的VAD检测算法。

#### 2.2 能量熵统计模型的LOG\_MMSE估计

以上的基于幅度谱的LOG\_MMSE估计与MMSE估计算法相比语音损伤更低,噪声滤除更彻底,但是在低信噪比条

作者简介:邱文武(1981-),男,硕士,主研方向:语音信号处理;蒋建中,副教授;吴琼,硕士

收稿日期:2010-05-27 E-mail: qww11234@sina.com

件下,会残留令人讨厌的音乐噪声。基于文献[3],本文将鲁棒性更好的能量熵作为估计因子,从带噪声语音的能量熵中估计出纯净语音的能量熵,而后得到增强后的语音。具体做法如下:

先求每个点频处的能量:

$$E(k, n) = |Y(k, n)|^2 \quad 1 \leq k \leq \frac{L}{2} \quad (4)$$

第  $k$  帧第  $n$  个点频处的能量占总能量的百分比为:

$$P(k, n) = \frac{E(k, n)}{E_{sum}} \quad 1 \leq k \leq \frac{L}{2} \quad (5)$$

其中,  $E_{sum}$  为总能量。而后求熵:

$$H(k, n) = -P(k, n) \log(P(k, n)) \quad (6)$$

应用 LOG\_MMSE 求信号熵的估计:

$$\hat{H}_k = \exp(E\{\log H | Y\}) \quad (7)$$

化简可知:

$$\hat{H}_k = \frac{\int_0^\infty x_k^2 \log(x_k) \exp\left(-\frac{x_k^2}{\lambda_k}\right) I_0\left(2x_k \sqrt{\frac{\lambda_k}{\lambda_k}}\right) dx_k}{\int_0^\infty x_k \exp\left(-\frac{x_k^2}{\lambda_k}\right) I_0\left(2x_k \sqrt{\frac{\lambda_k}{\lambda_k}}\right) dx_k} \quad (8)$$

由于方程没有确切解,只能求其近似解。

由式(5)可知,  $P(k, n)$  的值很小,进行泰勒展开,当  $x$  很小时近似有:

$$\ln(p+1) \approx p - \frac{p^2}{2} + \frac{p^3}{3} - \frac{p^4}{4} \quad 0 < p \leq 1 \quad (9)$$

对  $(p+1)\ln(p+1)$  进行泰勒展开可知:当语音段足够长时( $> 3$  s)  $x$  的值小于 0.001,满足:

$$(p+1)\ln(p+1) \approx p \quad (10)$$

由推导可知,式(5)近似满足自由度为 1 的卡方分布,由熵的定义可知,熵应满足自由度为  $M$  的卡方分布,当  $M$  足够大时满足中小极限定理,认为熵近似满足高斯分布,可得到式(8)的近似解:

$$\hat{H}_k = \frac{\xi_k}{1 + \xi_k} \exp\left(\frac{1}{2} \int_{\lambda_k}^\infty \frac{e^{-t}}{t} dt\right) H_k \quad (11)$$

之后根据  $\hat{H}_k$  反求出  $\hat{p}(n, k)$ 。反解式(6),得:

$$\hat{p}(n, k) = -\hat{H}/\text{lambertw}(-\hat{H}) \quad (12)$$

$\text{lambertw}$  函数在区间  $(0, 1)$  内解并不唯一,但在  $(0, 1/e)$  内有唯一解。由于  $\hat{p}(n, k) < 0.001$ ,因此取其在区间  $(0, 1/e)$  的解,根据式(4)、式(5)可求出  $\hat{x}_k$ 。

### 3 基于熵差的改进加权

DD 算法之所以降低甚至消除了音乐噪声是因为其采用了波动性较小的先验信噪比和后验信噪比进行加权,在文献[3]中  $\alpha$  的值控制了残留噪声的程度和语音损伤的程度。当  $\alpha$  较大时,后验信噪比起的作用很小,因此,波动性小,音乐噪声降低甚至完全消除。但是由于在语音的起始段由于使用前一帧的先验信噪比估计,引入了一定的延迟,因此出现噪声的过估计,使语音信号被过度滤除。特别是在突发的短时高能量语音帧,这种损失更为明显。为了减小延迟又提出了许多改进的 DD 算法<sup>[56]</sup>。本文将含噪语音分为:语音段,噪声段和过渡段,并且在各段分别采用不同的  $\alpha$  值来调整先验和后验信噪比的比重来达到改善跟踪的目的。在时间序列上噪声帧、语音帧和过渡帧间隔分布,为了加快噪声帧和语音帧的滤噪效果,适当增加  $\alpha$  的取值。反之,减小  $\alpha$  的取值。其具体

步骤为:

**Step1** 根据式(4)~式(6)求各频点的熵,而后对其进行帧间平滑和帧内平滑。

**Step2** 采用文献[3]中子带熵的 VAD 检测算法,求各帧的熵,而后根据初始噪声帧的熵确定语音和噪声初始门限。

$$Th_x = w_1(u + c \times d) \quad (13)$$

$$Th_n = w_2(u + c \times d) \quad (14)$$

其中,  $u$  为前 5 帧的均值;  $d$  为方差;  $w_1, w_2, c$  为经验常数。根据 VAD 的结果动态调整语音和噪声门限,使其在噪声帧不断更新。

**Step3** 在过度帧根据  $H(k, n)$  的波动性调整  $\alpha$  的取值

$$\alpha = \alpha_1 - (H(n, k) - \frac{1}{2L} \sum_{k-L}^k H(n, k+L)) \quad (15)$$

其中,  $\alpha_1 = 0.98$ ;  $L$  的大小根据当前语速决定。式(15)用于判断当前过度帧及其前后帧的波动性大小,波动性大则减小  $\alpha$ ,反之增大  $\alpha$ 。图 1 和图 2 分别为加入 0 dB 白噪声后采用权值固定( $\alpha = 0.98$ )和本文算法的去噪结果。

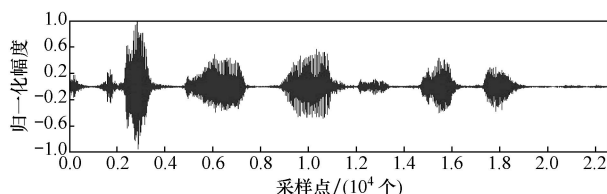


图 1 权值固定( $\alpha = 0.98$ )去噪时域结果

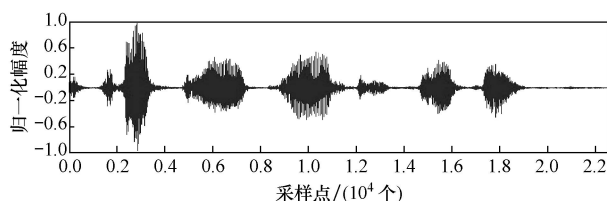


图 2 本文算法去噪时域结果

可以看出:本文提出的算法在弱语音段的结果优于固定权值算法,实际测试表明其整体信噪比也有所改善。

### 4 实验仿真分析

仿真中纯净语音信号来自 NOIZEUS 数据库,时长 2.24 s,采样频率为 8 000 Hz,噪声信号分别选来自 Noisex92 数据库的白噪声,色噪声和高频噪声。测 5 dB、0 dB 和 -5 dB 3 种情况。帧长度为 256,采用汉明窗,帧间重叠 50%。 $w_1 = 1.15$ 、 $w_2 = 1.25$ 、 $L = 1$ ,采用分段信噪比(SegSNR)对语音增强结果进行比较。

由表 1 可见,本文算法分段信噪比改善要优于传统的 LOG\_MMSE 估计算法,实际听辨证实残留音乐噪声明显减小。

表 1 2 种算法的 SegSNR 比较

输入 SNR	dB					
	白噪声		粉噪声		hf 噪声	
	Log_MMSE	本文算法	Log_MMSE	本文算法	Log_MMSE	本文算法
5	4.96	5.29	4.73	4.91	4.17	4.72
0	5.28	5.92	4.93	5.21	4.59	4.97
-5	5.42	6.34	5.06	5.44	4.62	5.20

### 5 结束语

本文提出了一种能量熵的 LOG\_MMSE 估计算法,降低了幅度谱 LOG\_MMSE 算法的残留音乐噪声问题。但是算法

(下转第 233 页)

2.0, 本文提出算法的加速比为 2.34, 较文献[5]的方法效率提高 17%, 说明基于功能模块的并行算法解码器效率得到提高。

图 5 中还比较了采用不同码率和熵编码方式的码流采用基于功能模块的并行解码器的加速比。由图可知, 随着码率的增加, CAVLC 和 CABAC 格式码流的加速比均下降, 这是由于高码率带来更高的运算量。而 CAVLC 模式较 CABAC 模式算法复杂度低 10%~15%, 因此 CAVLC 模式码流的加速比一直高于同码率下 CABAC 格式码流的加速比。

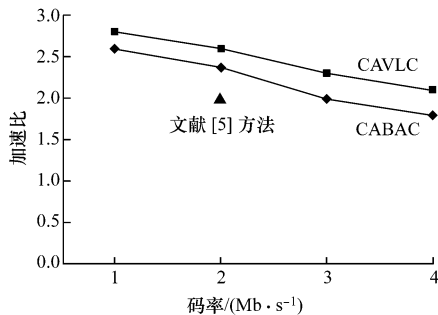


图 5 不同码率码流加速比

## 5 结束语

本文提出一种基于功能模块的并行解码算法, 了解码器内部模块的划分方法, 将解码器较均衡的划分为 3 个功

能模块, 设计了三核并行解码器结构, 分析了模块间数据的依赖关系, 简化了三核间数据交互。实验数据表明, 本文提出的基于功能模块的并行解码算法并行程度更高, 加速比优于基于可变模块划分等解码方法。

## 参考文献

- [1] Ravasi M. A Computational Complexity Comparison of MPEG4 and JVT Codes[S]. 2002.
- [2] Tol E B. Mapping of H. 264 Decoding on a Multiprocessor Architecture[C]//Proc. of SPIE' 03. Santa Clara, USA: [s. n.], 2003.
- [3] Seinster F H. A Macroblocklevel Analysis on the Dynamic Behavior of an H. 264 Decoder[C]//Proc. of ISCE' 07. [S. l.]: IEEE Press, 2007.
- [4] Tullsen D M. Limitations of Cache Prefetching on a Bus based Multiprocessor[C]//Proc. of the 20th Annual International Symposium on Computer Architecture. [S. l.]: IEEE Press, 1993.
- [5] Nishihara K. Parallelization of H. 264 Video Decoder for Embedded Multicore Processor[C]//Proc. of 2008 IEEE International Conference on Multimedia and Expo. [S. l.]: IEEE Press, 2008.

编辑: 陈 文

(上接第 228 页)

- [3] Spinsante S, Gambi E, Chiaraluce F, et al. Complexity Issues in the Introduction of High Definition in H. 264/AVC Based on Videoconferencing[C]//Proc. of 2006 International Conference on Software, Telecommunications and Computer Networks. [S. l.]: IEEE Press, 2006.
- [4] 宁 华, 梅 铮, 李锦涛. 基于 Slice 的 H. 264 并行视频编码算法[J]. 计算机工程, 2005, 31(4): 181-182.
- [5] Liu Yang, Li Zhengguo, Soh Yeng-Chai. Region of Interest

based Resource Allocation for Conversational Video Communication of H. 264/AVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(1): 134-139.

- [6] Ma Zhan, Zhang Zhongbo, Wang Yao. Complexity Modeling of H. 264 Entropy Decoding [C]//Proc. of International Conference on Image Processing. San Diego, USA: [s. n.], 2008.

编辑: 陈 文

(上接第 230 页)

运算量大, 不利于实时处理, 这是下一步的研究方向。

## 参考文献

- [1] Ephraim Y, Malah D. Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator[J]. IEEE Trans. on Acoust Speech Signal Process, 1984, 32(6): 1109-1121.
- [2] Erkelens J S, Hendriks R C, Heusdens R, et al. Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients with Generalized Gamma Priors[J]. IEEE Trans. on Audio, Speech, and Language Processing, 2007, 15(6): 1741-1752.
- [3] Wu Bingfei, Wang Kurr Ching. Robust Endpoint Detection

Algorithm Based on the Adaptive Band partitioning Spectral Entropy in Adverse Environments[J]. IEEE Trans. on Speech and Audio Processing, 2005, 13(5): 762-774.

- [4] 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.
- [5] Cohen I. Relaxed Statistical Model for Speech Enhancement and a Prior SNR Estimation[J]. IEEE Trans. on Speech and Audio Processing, 2005, 13(5): 870-881.
- [6] Plapous C, Marro C, Scalart P. Reliable A Posteriori Signal-to-noise Ratio Features Selection [C]//Proc. of 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, USA: [s. n.], 2005.

编辑: 金胡考