ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 30,No. 2,Mar. 2015,pp. 359－364
DOI:10.16337/j.1004-9037.2015.02.014

http://sjcj.nuaa.edu.cn
E-mail:sjcj@nuaa.edu.cn
Tel/Fax:＋86-025-84892742

基于隐马尔可夫模型的非监督噪声功率谱估计

许春冬[1,2,3]　战　鸽[1]　应冬文[1]　李军锋[1]　颜永红[1]

(1.中国科学院声学研究所语言声学与内容理解重点实验室,北京,100190; 2.江西理工大学信息工程学院,赣州,341000; 3.北京理工大学信息与电子学院,北京,100081)

摘　要:噪声功率谱估计是语音增强算法的基本组成部分,传统算法大多采用启发式的估计方法,因而不能保证噪声估计值的统计最优。提出了一种基于极大似然的非监督噪声功率谱估计方法,采用隐马尔可夫模型(Hidden Markov model, HMM)在每个子带建立语音和非语音对数功率谱的统计模型,模型包含语音和非语音两个高斯分量,其中非语音高斯分量的均值表示噪声功率谱估计值,根据最大期望(Expectation maximization, EM)算法得到包括噪声均值在内的 HMM 参数集。针对语音信号可能出现的长时缺失,对 HMM 引入了一些约束条件,保证了模型的稳定性。实验表明,该方法获得的极大似然噪声估计优于基于启发式的经典方法获得的噪声估计。

关键词:语音增强;噪声功率谱估计;隐马尔可夫模型;极大似然准则;模型约束

中图分类号:TN912.3　　**文献标志码**:A

Unsupervised Noise Power Estimation Using Hidden Markov Model

Xu Chundong[1,2,3], Zhan Ge[1], Ying Dongwen[1], Li Junfeng[1], Yan Yonghong[1]

(1. Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China; 2. School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, 341000, China; 3. School of Information and Electronics, Beijing Institute of Technology, Beijing, 100081, China)

Abstract: Noise estimation is a fundamental part of speech enhancement. Most traditional methods are heuristic which can not enable the optimal estimation. An unsupervised noise power estimation is presented based on maximum likelihood. A log-power statistical model is constructed using hidden Markov model (HMM) in each subband. This model comprises speech and nonspeech Gauss components, and the mean value of nonspeech Gauss component is the estimation of noise power. Moreover, speech may be long-term absent, some constraints are introduced to this model for stability. The experiments validate that the proposed method can obtain the maximum likelihood noise estimation and outperforms conventional heuristic methods.

Key words: speech enhancement; noise power estimation; hidden Markov model; maximum likelihood criterion; model constraints

---

基金项目:国家重点基础研究发展计划("九七三"计划)(2013CB329302)资助项目;国家自然科学基金(61271426,10925419,90920302,61072124,11074275,11161140319)资助项目;中国科学院战略性先导科技专项(XDA06030100,XDA06030500)资助项目;中国科学院重点部署(KGZD-EW-103-2)资助项目;江西理工大学科研基金(NSFJ2015-G21)资助项目。
收稿日期:2015-01-13;**修订日期**:2015-01-28

引 言

基于单麦克风的语音增强算法广泛应用于噪声环境下的语音编码、语音通信以及语音识别等领域，以降低背景噪声对语音处理系统的干扰[1]。而噪声功率谱估计是大多数单通道语音增强算法的重要组成部分，明显影响语音增强算法的性能。传统的语音增强算法多采用语音活动性检测(Voice activity detection, VAD)来判决语音的存在和缺失，只在语音缺失阶段估计噪声。该方法只能在平稳噪声环境下取得良好的噪声估计性能。

为提高复杂噪声环境下的噪声估计效果，在过去的几十年，出现了大量的噪声估计算法[1-7]。其中比较经典的算法有最小统计(Minimum statistics, MS)算法[5]、最小控制递归平均(Minima controlled recursive averaging, MCRA)算法[6]以及改进的最小控制递归平均(Improved minima controlled recursive averaging, IMCRA)算法[7]。与 VAD 方法不同，即使在语音活动期间，这些算法仍然能够在不包含语音信号的频率成分上更新噪声功率谱。这些方法虽然提高了复杂噪声环境下的噪声估计性能，但它们大多在本质上属于启发式的估计方法，不能在统计上保证噪声估计的最优。同时，由于传统算法大多要求噪声起始条件，在实际应用中可能会出现问题[8]。

基于以上考虑，本文提出一种基于极大似然的无监督噪声功率谱估计方法。对每个子带上的对数功率谱包络构建一个隐马尔可夫模型(Hidden Markov model, HMM)来描述语音和噪声的分布。同时通过最大期望(Expectation maximization, EM)学习方法，得到 HMM 模型的参数集，其中噪声分布的均值为噪声功率谱的最优估计。为了保证语音信号在长时缺失的情况下模型的稳定性，给出了一些约束条件。

1 基于隐马尔可夫模型的对数功率谱建模

本文关注单个子带的噪声功率谱估计问题，为不失一般性，首先在高信噪比频带采用隐马尔可夫模型建立对数功率谱包络的统计模型。

设 $\boldsymbol{x}\triangleq\{x_1,\cdots,x_L\}$ 表示一个长为 L 的观察序列，假设语音和非语音对数功率谱均服从高斯分布，语音状态和非语音状态间的谱序列动态转移可以采用马尔科夫链描述。这里的语音表示由纯净语音和噪声混合而成，而非语音表示无纯净语音叠加的信号，因此可以将带噪语音分为语音和非语音两个分量。λ 表示通过时间序列 $\boldsymbol{x}$ 估计的 HMM 参数集。设 $\boldsymbol{s}\triangleq\{s_1,\cdots,s_l,\cdots,s_L\}$ 表示对应于 $\boldsymbol{x}$ 的状态序列，其中 $s_l=1$ 和 $s_l=0$ 分别表示第 l 个样本的语音出现和缺失，对应于语音和非语音两种状态。对应 HMM 的概率密度函数可表示为

$$p(\boldsymbol{x}\mid\lambda)=\sum p(\boldsymbol{x},\boldsymbol{s}\mid\lambda)=\sum p(\boldsymbol{s}\mid\lambda)p(\boldsymbol{x}\mid\boldsymbol{s},\lambda) \tag{1}$$

式中：$p(\boldsymbol{s}|\lambda)$为状态序列 $\boldsymbol{s}$ 的概率，可表示为

$$p(\boldsymbol{s}\mid\lambda)=\prod_{t=1}^{L}a_{s_{t-1},s_t} \tag{2}$$

式中：a_{s_{t-1},s_t} 表示从 $t-1$ 时刻的状态 s_{t-1} 到 t 时刻的状态 s_t 的转移概率。$p(\boldsymbol{x}|\boldsymbol{s},\lambda)$是给定状态序列 $\boldsymbol{s}$ 的概率密度函数，可表示为

$$p(\boldsymbol{x}\mid\boldsymbol{s},\lambda)=\prod_{t=1}^{L}p(x_t\mid s_t,\lambda)\triangleq\prod_{t=1}^{L}f(x_t\mid s_t,\lambda) \tag{3}$$

式中：$f(x_t|s_t,\lambda)$表示在给定状态 s_t 和参数集 λ 下观察数据 x_t 的概率密度函数，可表示为

$$f(x_t\mid s_t=i,\lambda)=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) \tag{4}$$

式中：μ_i 和 σ_i^2 分别是给定状态 $s_t=i$ 高斯分布的均值和方差。

给定一个观察值序列 $\boldsymbol{x}$，HMM 参数集 $\lambda'=\{\boldsymbol{\pi},\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{\sigma}^2\}$ 的极大似然估计[9]可由式(5)计算

$$\lambda' = \arg\max_{\lambda} \ln \sum_{s} p(\boldsymbol{x}, \boldsymbol{s} \mid \lambda) \tag{5}$$

式中:$\boldsymbol{a}$ 为二阶转移概率矩阵,$\boldsymbol{\mu} \triangleq \{\mu_0, \mu_1\}$,$\boldsymbol{\sigma}^2 \triangleq \{\sigma_0^2, \sigma_1^2\}$,$\boldsymbol{\pi} \triangleq \{\pi_0, \pi_1\}$。需要指出,这里的 $\boldsymbol{\pi} \triangleq \{\pi_i\} \triangleq \{\pi_0, \pi_1\}$ 特指模型的初始状态分布。

2 隐马尔可夫模型的参数估计

估计 HMM 参数的常用批处理方法是 EM 算法[10],通过 EM 算法采用长为 L 的观察序列 $\boldsymbol{x}$ 估计模型参数。参数估计是在极大似然准则的指导下通过多次迭代直至收敛,从而得到模型的参数估值。在某次迭代中假定上次估计的参数集 λ 已知,求解一个更优的 λ' 的计算方法可通过 Q 函数实现,Q 函数定义为

$$Q(\lambda') \triangleq E\{\log p(\boldsymbol{x}, \boldsymbol{s} \mid \lambda) \mid \boldsymbol{x}, \lambda\} = \sum_{t=1}^{L} \Psi_{t|\lambda}(\lambda') + \sum_{i} \gamma_{1|\lambda}(i) \log \pi_i \tag{6}$$

其中,$\Psi_{t|\lambda}(\lambda')$ 表示为

$$\Psi_{t|\lambda}(\lambda') = \sum_{i} \sum_{j} \xi_{t|\lambda}(i,j) \log a_{ij} + \sum_{i} \gamma_{t|\lambda}(i) \log \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \tag{7}$$

式(6)和式(7)中的 a_{ij},μ_i,σ_i^2,π_i 为 λ' 中未知参数,λ 为上一次迭代中产生的已知参数。$\xi_{t|\lambda}(i,j) = p(s_{t-1} = i, s_t = j \mid x_t, \lambda)$ 表示条件转移概率,$\gamma_{t|\lambda}(i) = p(s_t = i \mid x_t, \lambda)$ 表示语音出现/缺失的概率。

根据 EM 算法,得到 λ' 中的未知参数估计值,分别表示为

$$\mu_i = \frac{\sum_{t} \gamma_t(i) x_t}{\sum_{t} \gamma_t(i)} \tag{8}$$

$$\sigma_i^2 = \frac{\sum_{t} \gamma_t(i) (x_t - \mu'_i)^2}{\sum_{t} \gamma_t(i)} \tag{9}$$

$$a_{ij} = \frac{\sum_{t} \xi_t(i,j)}{\sum_{t} \gamma_t(i)} \tag{10}$$

$$\pi_i = \gamma_0(i) \tag{11}$$

式中:$\gamma_t(i)$ 和 $\xi_t(i,j)$ 由 λ 得到。在计算完成后,可以采用 λ' 替换 λ 进行下一次迭代直至收敛,从而求得最终的参数集估计值。

3 隐马尔可夫模型的约束

上述模型适合于高信噪比(Signal to noise ratio, SNR)频带,要求语音和非语音必须同时出现。然而,在低 SNR 环境下,语音信号容易受到噪声的掩蔽,甚至某些子带只存在非语音分量,这时语音状态难以建模。因此,需要完善模型,使其适合于各种噪声环境。本文通过在二元状态 HMM 模型中添加限制条件来解决该问题。

语音和非语音在能量域有不同的分布。一般来说,低能量值为非语音,而高能量值为语音。同时,非语音能量相对语音能量更为平稳,相应的有非语音能量方差小于语音能量方差。根据以上分析,可以描述出某子带带噪语音对数谱包络的二元分布图。

图 1 为理想 SNR 条件下的带噪语音对数功率谱包络分布直方图。可以看出,语音和非语音状态之间满足一定的分布关系。当处于低 SNR 时,语音信号可能长时缺失。这时需要维持这种二元分布关系来保证模型的有效性。

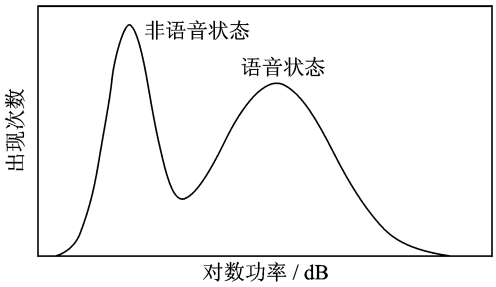

图 1 带噪语音对数谱包络分布直方图

Fig. 1 Histogram of noisy speech log-power distribution

为使低信噪比频带能满足图 1 所示分布关系，首先建立语音和非语音状态间的约束关系

$$\mu_1 = \max\{\mu_1, \mu_0 + \tau\} \tag{12}$$

式中：$\tau > 0$。其次，方差服从另外一种约束关系

$$\sigma_1^2 = \max\{\sigma_0^2, \sigma_1^2\} \tag{13}$$

以上两个约束关系是解决语音长时缺失问题的关键，如果没有均值约束，当二元状态 HMM 随着大量的噪声功率持续更新，则语音均值减少并逐步接近噪声均值，HMM 最终失去区分语音/噪声分量的能力。均值约束使语音均值至少比噪声均值大 τ 分贝，因此，当语音信号缺失时，区分能力得以保持；语音缺失情况下，语音状态被转换成均值和方差分别为 $\mu_0 + \tau$ 和 σ_0^2 的一个虚拟状态。τ 的取值对算法性能有一定的影响，它满足以下关系

$$\tau = \theta\sigma_0 \tag{14}$$

式中：θ 为常量因子，取 $\theta = 2.5$。可以看出，τ 随 σ_0 的变化而作出相应调整。

4 算法性能评价

实验采用 Noisex-92 噪声数据库[11] 中的 White 噪声、F16 噪声以及 Babble 噪声。纯净语音选自 TIMIT 数据库[12]，选用 10 人的语音数据，其中男声和女声各占 50%，每段纯净语音由 TIMIT 数据库中的两个短句构成，共测试 10 段长句。将噪声添加语音中得到 SNR 分别为 0，5 和 10 dB 的带噪语音信号。信号采样率均为 16 kHz，帧长为 16 ms，帧移为 8 ms，窗函数选择汉宁窗。测试数据集共 90 组，分别由 MS 算法、IMCRA 算法以及本文提出的算法进行测试。其中 HMM 算法实现过程如下：

(1) 采取加窗和快速傅里叶变换(Fast Fourier transformation, FFT)获得幅度谱序列 $\{y_1, y_2, \cdots, y_L\}$，平滑幅度序列可以获得对数功率谱序列 $\{x_1, x_2, \cdots, x_L\}$；

(2)针对长为 L 的观察序列 $\boldsymbol{x}$，通过约束聚类和 EM 批处理算法，采用对数功率谱序列在每个子带建立 HMM 模型；

(3)由式(8～11)计算参数集估计值；

(4)由式(12～14)对模型进行相关约束；

(5)由算法中的非语音均值求得噪声功率谱估计值。

为进一步评估算法性能，本文将提出的算法与当前主流的 MS 算法和 IMCRA 算法进行比较。提高语音质量需要尽量降低背景噪声和“音乐噪声”，而提高语音可懂度需要尽可能降低语音畸变。因此本文将测试不同噪声条件下语音增强的分段信噪比以及语音信号对数谱失真。

第 1 个评价指标为分段信噪比(Segmental signal to noise ratio, SegSNR)[13] 定义为

$$\mathrm{SegSNR}=\frac{1}{M}\sum_{m=1}^{M}10\log_{10}\left(\frac{\sum_{n=1}^{N}z_{m,n}^{2}}{\sum_{n=1}^{N}\left[z_{m,n}-\hat{z}_{m,n}\right]^{2}}\right) \tag{15}$$

式中:M 表示含有语音信号的帧数量;N 为每帧采样数;$z_{m,n}$ 为第 m 帧中第 n 个样点的纯净语音信号;$\hat{z}_{m,n}$ 为增强语音信号。SegSNR 值越大,则语音质量越好,相应的语音舒适度也更好。

第 2 个评价指标为语音信号对数谱失真((Log-spectral distortion, LSD)[14]定义为

$$\mathrm{LSD}=\frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{N/2+1}\sum_{k=0}^{N/2}\left[20\log_{10}|v_{m,k}|-20\log_{10}|\hat{v}_{m,k}|\right]^{2}\right]^{1/2} \tag{16}$$

式中:$v_{m,k}$ 和 $\hat{v}_{m,k}$ 分别表示第 m 帧中第 k 个频点的纯净语音和增强语音的幅度谱。对数谱失真越小,则增强后的语音可懂度越好。

从表 1 可以看出,HMM 算法的分段信噪比高于 MS 算法和 IMCRA 算法。从表 2 可以看出,HMM 算法的对数谱失真低于 MS 算法和 IMCRA 算法。因此,在复杂噪声环境下,本文算法对一些噪声类型具有更好的适应性和增强后的语音质量。

表 1 不同噪声环境下的分段信噪比

Table 1 Segmental SNR under various noise conditions dB

SNR	White 噪声			F16 噪声			Babble 噪声		
	MS	IMCRA	HMM	MS	IMCRA	HMM	MS	IMCRA	HMM
0	3.76	4.12	4.75	3.56	3.40	3.98	3.43	3.59	3.89
5	6.32	6.48	7.19	5.91	5.76	6.37	5.98	6.06	6.37
10	9.23	9.35	10.10	8.97	8.84	9.52	9.13	9.02	9.53

表 2 不同噪声环境下的对数谱失真

Table 2 Log-spectral distortion under various noise conditions dB

SNR	White 噪声			F16 噪声			Babble 噪声		
	MS	IMCRA	HMM	MS	IMCRA	HMM	MS	IMCRA	HMM
0	10.12	7.17	5.77	7.84	7.06	5.97	8.83	8.36	8.22
5	6.99	5.70	5.30	6.35	5.49	4.98	6.64	5.75	5.58
10	5.02	4.96	4.72	4.39	4.38	4.03	4.92	4.29	3.81

5 结束语

本文提出了一种基于极大似然准则的非监督噪声估计方法,算法在理论上保证了噪声估计值的统计最优,实验也证实了最优估计相对于启发式估计的优势。针对语音缺失对 HMM 的约束,保证了模型的稳定性。同时,由于模型的非监督性,噪声估计器不需要传统算法在句子开头部分的"噪声起始"假设,算法的实用性得到增强。

参考文献:

[1] Yuan Wenhao, Lin Jiajun, An Wei, et al. Noise estimation based on time-frequency correlation for speech enhancement[J]. Applied Acoustics, 2013, 74(5): 770-781.

[2] 赵胜跃, 戴蓓蒨. 基于最小统计噪声估计的信号子空间语音增强[J]. 数据采集与处理, 2007, 22(4): 453-457.
Zhao Shengyue, Dai Beiqian. Subspace speech enhancement based on minimum statistics noise estimation[J]. Journal of Data Acquisition and Processing, 2007, 22(4): 453-457.

[3] Zhong L, Rafik A G, Richard M D. Noise estimation using speech/non-speech frame decision and subband spectral tracking

[J]. Speech Communication, 2007, 49: 542-557.

[4] 余耀, 赵鹤鸣. 非平稳噪声环境下的噪声功率谱估计方法[J]. 数据采集与处理, 2012, 27(4): 486-489.
Yu Yao, Zhao Heming. New noise estimation method for highly non-stationary noise environments[J]. Journal of Data Acquisition and Processing, 2012, 27(4): 486-489.

[5] Martin R. Bias compensation methods for minimum statistics noise power spectral density estimation[J]. Signal Processing, 2006, 86: 1215-1229.

[6] Cohen I. Noise estimation by minima controlled recursive averaging for robust speech enhancement[J]. IEEE Signal Process Letters, 2002, 9(1):12-15.

[7] Cohen I. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging[J]. IEEE Transaction on Audio, Speech, and Language Processing, 2003, 11(5): 466-475.

[8] Quoc V Le. Building high-level features using large scale unsupervised learning [C]//Proc ICASSP13. Vancouver, Canada: IEEE Signal Processing Society, 2013: 8595-8598.

[9] Frédéric P, Yacine C, Ovarlez J P, et al. Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2008, 56(1):34-48.

[10] Ying D, Yan Y, Dang J, et al. Voice activity detection based on an unsupervised learning framework[J]. IEEE Transaction on Audio, Speech, and Language Processing, 2011, 19(8):2624-2633.

[11] Loizou P C. Speech enhancement: Theory and practice[M]. New York: CRC Press, 2007: 460-461.

[12] Donwen Y, Yonghong Y. Noise estimation using a constrained sequential hidden Markov model in the log-spectral domain [J]. IEEE Trans on Audio, Speech, and Language Processing, 2013, 21(6):1145-1157.

[13] Hu Y, Loizou P C. Evaluation of objective quality measures for speech enhancement[J]. IEEE Transaction on Audio, Speech, and Language Processing, 2008, 16(1): 229-238.

[14] Donwen Y, Masashi U, Xugang L, et al. Speech enhancement based on noise eigenspace projection[J]. IEICE Transactions on Information and Systems, 2009, 92(5): 1137-1145.

作者简介:**许春冬**(1976-),男,博士研究生,研究方向:语音及音频信号处理,E-mail:xcd201@qq.com;**战鸽**(1990-),男,硕士研究生,研究方向:语音及音频信号处理;**应冬文**(1975-),男,研究员,研究方向:语音增强、语音识别、声源定位等;**李军锋**(1979-),男,研究员,研究方向:语音增强、三维音频;**颜永红**(1967-),男,研究员,研究方向:语音识别、语音增强,E-mail:yanyonghong@hccl.ioa.ac.cn。