

文章编号: 2095-4980(2013)05-0787-05

基于加权噪声的递归平滑噪声功率谱估计

冯 炎, 安宝坤

(西藏大学 现代教育技术中心, 西藏 拉萨 850000)

摘 要: 提出了一种噪声功率谱估计算法, 该算法对加权后的带噪语音进行递归平滑, 可以持续更新噪声并可应用于非平稳噪声环境中。为了避免在强语音后的弱语音区域出现噪声过估计, 本文提出了用于计算加权函数的投影平滑算法。本文噪声估计算法可以快速跟踪噪声的变化并且没有过估计。实验结果表明, 本文噪声估计算法应用于一个语音增强系统时, 取得了较小的噪声分段估计误差及较好的感知语音质量评价(PESQ)得分。

关键词: 语音增强; 先验信噪比估计; 噪声功率谱估计; 递归平滑

中图分类号: TN912.3

文献标识码: A

doi: 10.11805/TKYDA201305.0787

Recursive averaging noise power spectrum estimation based on weighted noise

FENG Yan, AN Bao-kun

(Modern Education and Technology Center, Tibet University, Lhasa Tibet 850000, China)

Abstract: A noise power spectrum estimation algorithm is presented, which continuously updates the estimated noise by weighted noisy speech for non-stationary noise environments. To avoid noise overestimates in the weak speech regions following strong speech, a projecting smoothing method is developed for weighting factor calculation. The update for varied noise power levels is much faster compared to most other algorithms, and at the same time, noise power spectrum is not overestimated. Experimental results indicate that the proposed noise estimation algorithm yields lower segmental estimation error and better Perceptual Evaluation of Speech Quality(PESQ) scores when integrated in speech enhancement.

Key words: speech enhancement; weighted noise estimation; noise power spectrum estimation; recursive averaging

噪声功率谱估计是语音增强系统的重要组成部分。噪声功率谱的过估计与欠估计都会给语音增强系统带来较大危害。因此, 噪声估计是学者们一直较为关注的问题^[1-7]。Martin^[6]提出了最小值统计跟踪噪声估计算法(Minimum Statistics, MS), 无需进行语音检测, 无论在有声或无声时都对噪声估计进行更新, 但该算法中所估计的噪声功率谱无法快速跟踪噪声的变化, 因而在噪声上升区域出现噪声欠估计, 在连续的强语音后的区域出现噪声过估计。Kato 等人提出的加权噪声估计(Weighted Noise estimation, WN)算法^[7-8]不依赖最小值跟踪算法, 并且简单有效, 能跟踪瞬变噪声, 但该算法无法区别弱语音与噪声, 在弱语音段会出现噪声过估计。为此, 本文提出了投影平滑算法, 该算法可以将强语音后的弱语音与噪声区分开。结合投影平滑算法, 本文在加权噪声估计的基础上提出一种新的递归平滑噪声功率谱估计算法。该算法可避免强语音后弱语音区域出现的噪声过估计, 由于采用了递归平滑, 估计出的噪声功率谱比加权噪声估计算法出现的过估计与欠估计更少。相对于传统的噪声功率谱估计方法, 本文提出的噪声估计算法能提供更准确的噪声估计, 从而为增益函数计算提供更准确的先验信噪比和后验信噪比估计, 使增强后的语音有更好的语音质量和充分的噪声抑制, 进一步减少残留噪声, 且可以解决多种噪声背景下的语音增强问题。

1 本文提出的递归平滑噪声估计算法

递归平滑的思想经常在噪声功率谱的估计中使用, 尤其是在缓慢变化的噪声环境中^[6]。利用相邻帧语音出现

概率具有很强的相关性^[9],通过平滑可以进一步减少语音到非语音段的波动,能够进一步增强跟踪噪声谱突变的能力。本文提出的噪声估计算法主要有 3 个步骤:对带噪语音进行投影平滑;用投影平滑后的带噪语音计算信噪比,进而计算加权函数;对用该加权函数加权后的带噪语音功率谱进行递归平滑得到所估计的噪声功率谱。

1.1 噪声估计

本文用 $x(t)$ 和 $d(t)$ 分别表示纯净语音和不相关的加性噪声,观测到的带噪语音信号为 $y(t)$,进行短时离散傅利叶变换后得到:

$$Y(n,k) = X(n,k) + D(n,k) \quad (1)$$

式中 n 和 k 分别表示时间帧序号和频率点序号。本文问题是从带噪语音谱 $Y(n,k)$ 中估计噪声功率谱,若用 $W(n,k)$ 表示加权函数,在无法确定语音是否出现时,对加权后的带噪语音 $W(n,k)|Y(n,k)|^2$ 进行时域递归平滑更新噪声功率,而在语音确定出现时不更新噪声,即:

当 $\tilde{\gamma}(n,k) \geq \tilde{\gamma}_0$ 时:

$$\hat{\lambda}_D(n,k) = \hat{\lambda}_D(n-1,k) \quad (2)$$

当 $\tilde{\gamma}(n,k) < \tilde{\gamma}_0$ 时:

$$\hat{\lambda}_D(n,k) = a_d \hat{\lambda}_D(n-1,k) + (1-a_d)[W(n,k)|Y(n,k)|^2] \quad (3)$$

式中: a_d ($0 < a_d < 1$) 是平滑参数; $\tilde{\gamma}(n,k)$ 为所估计的信噪比; $\tilde{\gamma}_0$ 为强语音阈值。初始几帧被认为是噪声,为获得满意的噪声估计,对初始几帧进行平均^[3], T_{init} 表示初始帧的大小:

$$n < T_{\text{init}}, \hat{\lambda}_D(n,k) = \frac{1}{n} \sum_{i=1}^n |Y(i,k)|^2 \quad (4)$$

1.2 加权函数计算及信噪比估计

参考 WN 算法,本文中加权函数 $W(n,k)$ 的计算方法如下^[3]:

$$W(n,k) = \begin{cases} 1, & \tilde{\gamma}(n,k) < \tilde{\gamma}_1 \\ \frac{\tilde{\gamma}_2 - \tilde{\gamma}(n,k)}{\tilde{\gamma}_2 - \tilde{\gamma}_1}, & \tilde{\gamma}_1 < \tilde{\gamma}(n,k) < \tilde{\gamma}_0 \\ 0, & \text{others} \end{cases} \quad (5)$$

式中: $\tilde{\gamma}_1$, $\tilde{\gamma}_2$ 和 $\tilde{\gamma}_0$ 是加权函数参数,由本文实验来确定。 $\tilde{\gamma}(n,k)$ 为所估计的信噪比,本文的信噪比估计是通过投影平滑后的带噪语音功率与前一帧的噪声之比来获得:

$$\tilde{\gamma}(n,k) = \frac{S_p(n,k)}{\hat{\lambda}_D(n-1,k)} \quad (6)$$

式中 $S_p(n,k)$ 表示投影平滑后的带噪语音。

通过上面的讨论可知,信噪比估计的准确性决定所估计的噪声谱密度准确性。而信噪比的估计需要从投影平滑后的带噪语音中获得,由此可见,投影平滑是本文算法成功的关键。

1.3 投影平滑算法

为解决强语音后的弱语音区域信噪比估计偏小的问题,本文在信噪比估计问题中考虑了相邻帧间相关性的信息^[7],进而改进了投影平滑方法。若用 $I(n,k)$ 表示投影指示函数,指示是否需要在该区域使用投影平滑来保护弱语音。当 $I(n,k)=1$,表示该区域是强语音后的弱语音区域;当 $I(n,k)=0$,则表示该区域不属于保护的区域。投影平滑后的带噪语音功率谱 $S_p(n,k)$ 求法如下:

当 $I(n,k)=0$

$$S_p(n,k) = |Y(n,k)|^2 \quad (7)$$

当 $I(n,k)=1$

$$S_p(n,k) = \alpha_p(n,k)S_p(n-1,k) + [1-\alpha_p(n,k)]|Y(n,k)|^2 \quad (8)$$

式中 α_p ($0 < \alpha_p < 1$) 表示投影平滑因子。将上面两式合并简化得到投影平滑算法公式:

$$S_p(n,k) = \tilde{\alpha}_p(n,k)S_p(n-1,k) + [1-\tilde{\alpha}_p(n,k)]|Y(n,k)|^2 \quad (9)$$

式中: $\tilde{\alpha}_p(n,k)$ 为时频相关的投影平滑因子,且 $\tilde{\alpha}_p(n,k) = \alpha_p I(n,k)$ 。

投影指示函数计算是投影平滑算法的关键。若前一帧不是所要保护的区域时, 当满足 3 个条件, 即认为当前帧属于保护的区域。这 3 个条件是: 1) 当前帧功率比前一帧小; 2) 前一帧信噪比较高, 即出现强语音; 3) 当前帧信噪比不是很小, 即当前帧属于噪声帧的可能性很小。若前一帧是所要保护的区域时, 仅满足 2 个条件之一, 认为当前帧不属于保护的区域。这 2 个条件是: 1) 前一帧投影平滑后的值比当前帧小, 无法保护当前帧; 2) 当前帧信噪比很小, 即当前帧属于噪声帧的可能性很大。公式描述如下:

当 $I(n-1, k) = 0$ 时:

$$I(n, k) = \begin{cases} 1, & |Y(n-1, k)|^2 > |Y(n, k)|^2 \text{ and } \frac{|Y(n-1, k)|^2}{\hat{\lambda}_D(n-1, k)} > \sigma_{\max} \text{ and } \frac{|Y(n, k)|^2}{\hat{\lambda}_D(n-1, k)} > \sigma_{\min} \\ 0, & \text{others} \end{cases} \quad (10)$$

式中: $\sigma_{\max}$ 表示强语音阈值; $\sigma_{\min}$ 表示噪声阈值。

当 $I(n-1, k) = 1$ 时:

$$I(n, k) = \begin{cases} 0, & \frac{|Y(n, k)|^2}{\hat{\lambda}_D(n-1, k)} \leq \sigma_{\min} \text{ or } S_p(n-1, k) \leq |Y(n, k)|^2 \\ 1, & \text{others} \end{cases} \quad (11)$$

对于初始时刻, 认为不会出现强语音, 因而也就不进行对弱语音的保护, 不进行投影平滑, 即对 $n < T_{\text{init}}$, $I(n, k) = 0$ 。

图 1 给出了 Volvo 汽车噪声中对单一频率点 $k=13$ 处的带噪语音功率谱的投影平滑, 中心频率约为 390 Hz, 输入信噪比为 5 dB。从图 1 可看出, 投影平滑算法可以有效地将强语音后的弱语音功率谱值提高。

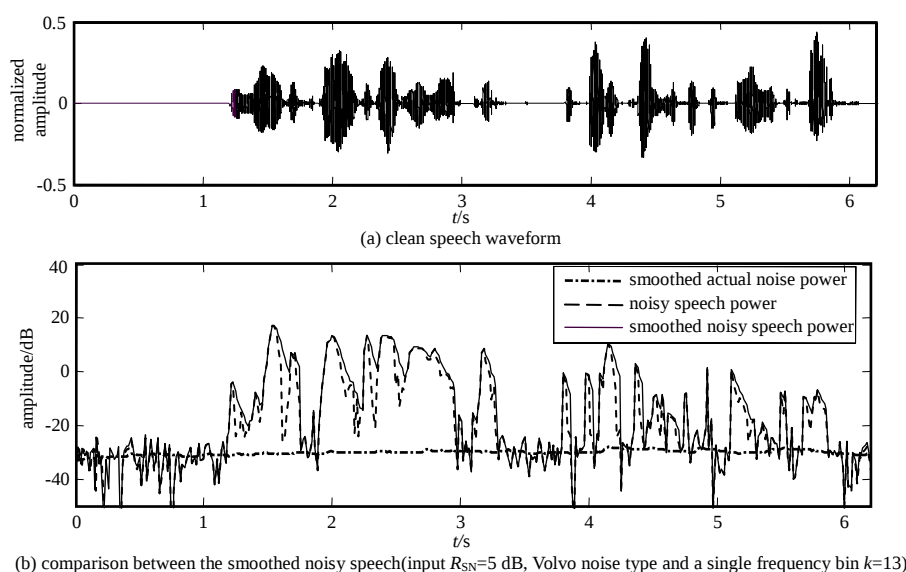


Fig.1 Projecting smoothed noisy speech power

图 1 投影平滑后的带噪语音功率

2 算法步骤与算法复杂度分析

本文中基于加权噪声估计的递归平滑噪声估计算法基本步骤如下:

1) 初始化: $n=1$, $T_{\text{init}}=4$ 。

2) 如果 $n \leq T_{\text{init}}$, 对所有频率点 k , $I(n, k) = 0$, $\hat{\lambda}_D(n, k) = \frac{1}{n} \sum_{i=1}^n |Y(i, k)|^2$ 。

3) 如果 $n > T_{\text{init}}$, 对所有频率点 k : 用式(10)、式(11)计算投影平滑指示函数 $I(n, k)$; 用式(9)求投影平滑后的带噪语音 $S_p(n, k)$; 用式(6)计算信噪比 $\hat{\gamma}(n, k)$; 用式(5)计算加权函数 $W(n, k)$; 用式(3)计算所估计的噪声功率谱 $\hat{\lambda}_D(n, k)$ 。

4) 设置 $n = n+1$, 并且进入第 2) 进行下一帧的噪声估计。

为了更好地理解本文的噪声估计算法的优势, 图 2 描述了本文的算法、WN 和 MS 算法估计的噪声功率谱与

真实的噪声功率谱的性能对比。实验条件与图 1 一样。仿真结果表明本文的噪声估计算法能解决 WN 算法在弱语音时噪声过估计的问题,避免了 MS 算法在持续强语音后的噪声过估计问题。从图还可以看出噪声功率谱突然上升或者突然下降时,本文算法能够很好地跟踪噪声变化,所估计出的噪声功率谱更接近于真实噪声。

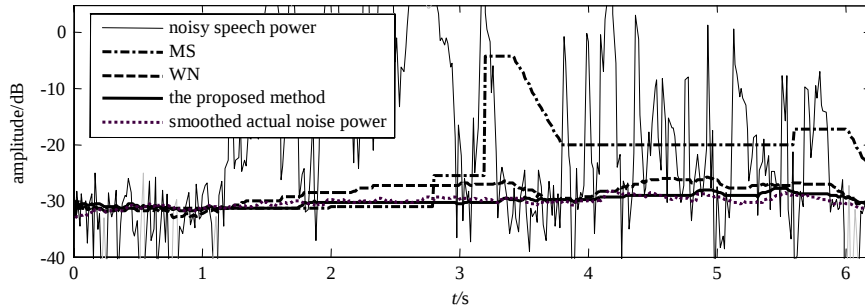


Fig.2 Noise power estimation obtained by the proposed method, MS and WN(input $R_{SN}=5$ dB, Volvo noise type and a single frequency bin $k=13$)

图 2 本文提出的方法与 MS、WN 算法的噪声功率谱估计对比(输入信噪比为 5 dB 的 Volvo 噪声, 频率点 $k=13$)

MS 算法时间复杂度: $T(n)=16+n+n+8\times 129n+5n+2=O(n)$;

WN 算法时间复杂度: $T(n)=14+n+n+8+9\times 129(n-4)+n+2=O(n)$;

本文所提算法时间复杂度: $T(n)=19+n+n+12+17\times 129(n-4)+2=O(n)$ 。

经分析发现, MS 算法、WN 算法与本文所提算法时间复杂度都是线性阶函数。

3 仿真实验与结果分析

为衡量上述算法性能,将本文算法应用到最小均方误差语音增强系统中^[9],并对实验结果进行讨论。实验中所采用语音段由取自 NOIZEUS 语音库的 6 句不同的语音组成,其中有 3 个男声和 3 个女声。这些语音用 16 kHz 采样, 16 bit 编码, 噪声是取自 Noisex92 噪声库的白噪声(White)、汽车噪声(Volvo)以及工厂噪声(factory), 分别将这 3 种噪声与上述语音段合成信噪比为 0 dB, 5 dB, 10 dB 和 15 dB 的带噪语音。对语音信号应用短时傅里叶变换时, 分帧的帧长为 256 点, 帧间重叠 128 点, 为防止分帧时的截断效应, 对分帧的语音信号进行“加窗”处理, 所使用的窗函数为汉明窗。

先验信噪比估计中参数设定^[10]: $\alpha=0.98$, $\xi_{\min}=-25$ dB。MS 算法的参数^[6]: $\alpha=0.85$, $D=100$, $U=4$, $V=25$, $\beta=1.5$ 。WN 算法的参数设定为^[7]: $T_{\text{init}}=4$, $L_z=20$, $\tilde{\gamma}_1=0$ dB, $\tilde{\gamma}_2=10$ dB, $\theta_z=7$ dB。

本文噪声估计算法的实验参数: $T_{\text{init}}=4$, $\alpha_p=0.5$, $\alpha_d=0.95$, $\sigma_{\min}=1.5$, $\sigma_{\max}=8$, $\tilde{\gamma}_0=5$, $\tilde{\gamma}_1=1.5$, $\tilde{\gamma}_2=10$ 。

表 1 给出了本文算法与 WN 算法、MS 算法的相对估计误差对比实验, 相对估计误差值越小说明所估计的噪声谱越接近实际噪声。从实验数据可看出, 不同的噪声环境和不同的输入信噪比的试验中, 相对于 WN 算法与 MS 算法, 本文算法均取得了较低的相对估计误差, 而 WN 算法与 MS 算法各有优劣。说明本文算法估计出的噪声功率谱最接近真实噪声。

为了进一步衡量语音的质量, 对增强语音进行 PESQ 得分实验。PESQ 是 ITU-T P.862 推荐的语音质量评价标准^[11], 分值是 0~4.5, 数值越大说明语音质量越好。表 2 给出了分别使用本文算法及 WN 算法、MS 算法增强后的语音的 PESQ 得分对比。仿真结果表明, 与 MS 及 WN 算法相比, 本文算法在不同的噪声环境和不同的噪声等级水平中均取得了较高的 PESQ 得分。

表 1 三种算法的相对估计误差

Table1 Average relative estimation error of the three algorithms

noise type	input R_{SN}/dB	proposed method	WN	MS
white	0	0.059 3	0.091 3	0.124 7
	5	0.067 9	0.102 5	0.197 2
	10	0.066 8	0.124 2	0.845 8
	15	0.075 8	0.145 0	6.885 2
factory	0	0.122 4	0.151 5	0.174 6
	5	0.155 8	0.192 0	0.186 5
	10	0.180 7	0.255 4	0.318 1
	15	0.147 2	0.347 8	1.554 6
Volvo	0	0.241 6	0.261 7	0.244 3
	5	0.241 9	0.259 2	0.244 9
	10	0.242 7	0.259 1	0.253 6
	15	0.242 2	0.259 7	0.344 8

表 2 三种算法的 PESQ 得分

Table2 PESQ scores obtained by the three algorithms

noise type	input R_{SN}/dB	proposed method	WN	MS
white	0	1.896	1.858	1.851
	5	2.310	2.270	2.246
	10	2.650	2.622	2.583
	15	2.920	2.898	2.847
factory	0	2.018	1.987	1.955
	5	2.384	2.356	2.308
	10	2.691	2.670	2.616
	15	2.977	2.959	2.885
Volvo	0	3.423	3.387	3.202
	5	3.772	3.733	3.386
	10	4.044	4.020	3.506
	15	4.244	4.223	3.575

4 结论

本文的投影平滑方法可以将强语音后的弱语音与噪声区分开,并结合投影平滑算法提出了基于加权噪声估计的递归平滑噪声功率谱估计算法。本文的噪声功率谱估计算法可避免强语音后弱语音区域出现的噪声过估计。由于采用了递归平滑,估计出的噪声功率谱比加权噪声估计算法出现的过估计与欠估计更少。相对于传统的噪声功率谱估计方法,本文提出的噪声估计算法简单并能提供更准确的噪声估计,进而会使增强后的语音有更好的语音质量和充分的噪声抑制,且可以解决多种噪声背景下的语音增强问题。

参考文献:

- [1] Hao J, Attias H, Nagarajan S, et al. Speech enhancement, gain, and noise spectrum adaptation using approximate Bayesian estimation[J]. IEEE Transactions on Audio, Speech and Language Processing, 2009, 17(1): 24–37.
- [2] 安扣成. 基于先验信噪比估计和增益平滑的语音增强[J]. 计算机应用, 2012, 32(z1): 29–31. (AN Koucheng. Speech enhancement based on priori SNR estimation and smoothing gain[J]. Journal of Computer Applications, 2012, 32(z1): 29–31.)
- [3] Ephraim Y, Cohen I. Recent advancements in speech enhancement[M]// The Electrical Engineering Handbook, 3rd ed. Boca Raton, FL: CRC, 2004.
- [4] 冯炎. 基于直接判决估计和预测估计的语音增强算法[J]. 信息与电子工程, 2010, 8(1): 76–7979. (FENG Yan. Speech enhancement based on Decision-Directed incorporating predicted estimation[J]. Information and Electronic Engineering, 2010, 8(1): 76–79.)
- [5] 冯炎, 尼玛扎西. 基于频带间相关性的加权噪声功率谱估计[J]. 信息与电子工程, 2010, 8(4): 431–435. (FENG Yan, NYIMA Trashi. Weighted noise power spectrum estimation based on the inter-frequency correlation[J]. Information and Electronic Engineering, 2010, 8(1): 76–79.)
- [6] Martin R. Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics[J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(5): 504–512.
- [7] Kato M, Sugiyama A, Serizawa M. Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA[C]// Proceedings of IWAENC2001. [s.n.], 2001: 183–186.
- [8] Benesty Jacob, Makino Shoji, CHEN Jingdong. Speech enhancement[M]. Berlin: Springer, 2005: 115–133.
- [9] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1984, 32(6): 1109–1121.
- [10] Cappé O. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor[J]. IEEE Transactions on Speech and Audio Processing, 1994, 2(2): 345–349.
- [11] ITU-T P.862. Perceptual Evaluation of Speech Quality(PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs[S], 2001.

作者简介:



冯 炎(1979–), 男, 山西省吕梁市人, 硕士, 讲师, 主要研究方向为信号与信息处理、藏文信息处理. email: fy4528@163.com.

安宝坤(1978–), 男, 河北省唐山市人, 硕士, 讲师, 主要研究方向为计算机网络、藏文信息处理等.