

基于加权最小统计的噪声谱估计改进算法

牛 铜 张连海 屈 丹
(郑州信息科技学院信息科学系 郑州 450002)

摘 要: 针对最小统计的噪声谱估计算法在噪声谱上升时跟踪时延较大的问题, 该文给出了一种加权最小统计的噪声谱估计改进算法, 通过分析加权对最小统计噪声谱估计算法的影响, 采用 3 种简单典型曲线构造权值, 用实验的方法比较得出由余弦曲线构造的权值最优。实验结果表明加权最小统计的噪声谱估计改进算法能够快速跟踪噪声谱的变化, 提高了噪声谱估计的准确性, 改善了增强后的语音质量。

关键词: 语音增强; 噪声谱估计; 加权最小统计

中图分类号: TN912.35

文献标识码: A

文章编号: 1009-5896(2009)05-1166-04

An Improved Noise Spectral Estimation Algorithm Based on the Weighted Minimum Statistics

Niu Tong Zhang Lian-hai Qu Dan

(Dept. of Information Science, Zhengzhou Information Science and Technology Institute, Zhengzhou 450002, China)

Abstract: As the noise spectral estimation based on the minimum statistics introduces significant tracking latency when the noise spectral rises, an improved algorithm based on the weight minimum statistics is presented. Analyzing the influence of weight on the noise spectral estimation based on the minimum statistics, three kinds of typical simple curves are used to compute the weight, and the experiment shows that the weight computed by the cosine curve is the best. The simulation results show that the improved algorithm traces the change of noise spectral quickly in most cases, improves the accuracy of the noise spectral estimation and the quality of speech in the non-stationary noise environment.

Key words: Speech enhancement; Noise spectral estimation; Weighted minimum statistics

1 引言

目前, 大多数语音增强算法需要从带噪信号中估计噪声功率谱密度, 噪声功率谱估计的准确与否, 很大程度上影响着增强算法的性能: 对噪声功率谱的过估计, 会导致语音失真增大; 对噪声功率谱的欠估计, 会产生大量残余噪声, 即音乐噪声^[1]。

通过对带噪语音的统计特性和频谱的研究发现^[2]: 语音帧中, 不同频带内语音分量的强弱是不同的; 不同频带内噪声对语音的影响不同, 即使在同一帧内, 语音存在的概率也不同。基于此现象, 常采用最小值跟踪算法和时间回归平均算法进行噪声谱估计, 最小值跟踪算法的核心思想是跟踪带噪语音功率谱不同频带内的最小值, 经无偏修正后作为对噪声谱的估计; 时间回归平均算法的基本思想是, 当频带内语音存在的概率较小时, 更新噪声功率谱, 否则不更新。

Martin 结合上述两种算法, 提出了一种基于最小统计和最优滤波的噪声谱估计^[1,3]算法, 该算法不需要语音激活检测, 对于非平稳噪声, 具有较好的噪声谱估计性能, 但在噪声功率谱上升阶段有较大的跟踪时延。

本文针对最小统计(Minimum Statistics, MS)算法在噪声上升阶段跟踪时延较大的问题, 提出了一种加权最小统计算法(Weighted Minimum Statistics, WMS), 仿真实验结果表明, WMS 算法较 MS 算法明显减少了在噪声功率谱上升阶段对噪声谱的跟踪时延。

2 最小统计和最优滤波的噪声谱估计算法

假设语音和噪声是统计独立的, 并且噪声谱变化的速度要远小于语音谱变化的速度。

Martin 提出的最小统计和最优滤波噪声谱估计算法, 首先对带噪语音功率谱 $|Y(\lambda, k)|^2$ 采用时频独立的最优滤波参数 $\alpha_{\text{opt}}(\lambda, k)$ 进行一阶平滑滤波, 得到平滑后的功率谱 $P(\lambda, k)$:

$$P(\lambda, k) = \alpha_{\text{opt}}(\lambda, k)P(\lambda - 1, k) + (1 - \alpha_{\text{opt}}(\lambda, k))|Y(\lambda, k)|^2 \quad (1)$$

其中 λ 代表帧号, k 代表第 k 个频率分量, $|Y(\lambda, k)|^2$ 采用周期图法得到, $\alpha_{\text{opt}}(\lambda, k)$ 则是在最小均方误差准则下得到的^[3], 然后采用 MS 算法得到连续 D 帧的带噪语音功率谱序列 $\{P(\lambda - D + 1, k), P(\lambda - D + 2, k), \dots, P(\lambda, k)\}$ 的最小值 $P_{\text{min}}(\lambda, k)$, 最后根据 $P_{\text{min}}(\lambda, k)$ 的方差和 D 计算无偏修正因子 $B_{\text{min}}(\lambda, k)$ ^[3], 对 $P_{\text{min}}(\lambda, k)$ 无偏修正, 得到噪声功率谱的估计 $\hat{\sigma}_N^2(\lambda, k)$, 即

$$\delta_N^2(\lambda, k) = B_{\min}(\lambda, k) P_{\min}(\lambda, k) \quad (2)$$

该算法的特点是在噪声功率谱下降时, 能够快速地跟踪到噪声功率谱的变化, 但在噪声功率谱的上升阶段, 最大会产生连续 $2D$ 帧的跟踪时延^[3]。Martin 提出了一种改进的最小值跟踪算法, 将连续 D 帧分成 U 个子窗, 每个子窗内包含有 V 帧, 显然有 $D = UV$ 。计算每个子窗内的最小值 $P_{i,\min}(k)$, 其中, $i = 1, 2, \dots, U$ 。然后找出序列 $\{P_{i,\min}(k)\}_{i=1,2,\dots,U}$ 的最小值作为 $P_{\min}(\lambda, k)$ 。文献[3]指出, 该算法最大会产生 $D + V$ 帧的跟踪时延。

3 加权最小统计的改进算法

对 Martin 提出的最小统计算法的研究发现, 造成算法不能及时跟踪噪声功率谱变化的主要原因是: 在寻找最小值时, 不同子窗内的最小值被赋予了相同的权值, 如图 1 中的实线所示。本文采用一种加权最小统计的改进算法, 寻找加权序列 $\{w_i P_{i,\min}(k)\}_{i=1,2,\dots,U}$ 的最小值 $w_j P_{j,\min}(k)$, 其中 w_i 表示第 i 个权值, $j \in [1, U]$ 且 $j \in Z$, 并将 $P_{j,\min}(k)$ 作为最小值 $P_{\min}(\lambda, k)$, 即 $P_{\min}(\lambda, k) = P_{j,\min}(k)$, 如图 1 中的虚线所示。

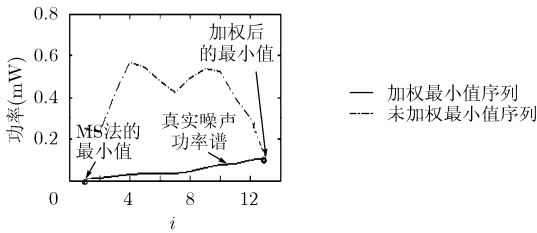


图 1 加权最小统计法避免欠估计示意图

显然权值序列 $\{w_i\}_{i=1,2,\dots,U}$ 在这种基于加权改进的最小值跟踪算法中起到了关键的作用。

权值序列实质上反映了序列 $\{P_{i,\min}(k)\}_{i=1,2,\dots,U}$ 中的元素在时间上相对的关联程度, 所以 $\{w_i\}_{i=1,2,\dots,U}$ 必是单调递减的, 且 w_i 仅与子窗间的时间间隔有关, 与子窗的起始时间无关。但序列 $\{P_{i,\min}(k)\}_{i=1,2,\dots,U}$ 的上升趋势不一定是由噪声引起, 语音的出现也会引起序列 $\{P_{i,\min}(k)\}_{i=1,2,\dots,U}$ 的上升。因此, 权值序列还需防止错误的更新噪声功率谱, 导致对噪声谱的过估计。

假设噪声每秒变化速率不超过 R (dB/s), 且当前帧所在子窗内的最小值所对应的权值等于 1, 并认为带噪语音功率谱变化速率超过 R (dB/s) 的情况, 均是由于语音存在而引起的。

由上面的讨论可知, 权值序列 $\{w_i\}_{i=1,2,\dots,U}$ 应该保证算法能够及时更新噪声功率谱, 当带噪语音上升速率大于 R (dB/s) 时, $\{w_i\}_{i=1,2,\dots,U}$ 应该保证算法不再更新噪声功率谱。

因此, 对于帧长 W , 帧移 SF , 采样速率 fs , 最小统计

算法采用连续 D 帧, 分成 U 个子窗, 每个子窗有 V 帧, 噪声最大上升速率为 R (dB/s) 的情况, 权值序列中的最大值 w_0 为

$$w_0 = (10^{R/10} + 1) \times (SF \times (U \times V - 1) + W) / fs \quad (3)$$

这里采用曲线构造的方法得到权值序列, 用于构造权值的曲线简称为权值曲线。由于限制了权值序列的取值范围为 $[1, w_0]$, 对常用的曲线 sigmoid, 余弦曲线和线性曲线函数作相应的修正后可得权值 w_i 的 3 种表达式:

sigmoid

$$w_i = \frac{w_0 - 1}{1 + \exp(i - U/2 + 1)} + 1 \quad (4)$$

余弦曲线

$$w_i = \frac{1}{2}(w_0 - 1) \left[\cos\left(\frac{\pi}{U} i\right) + 1 \right] + 1 \quad (5)$$

线性

$$w_i = w_0 - \frac{(w_0 - 1)}{U} i \quad (6)$$

其中 $i = 1, \dots, U$ 。

当 $U = 12, V = 8, fs = 8\text{kHz}, W = 256, SF = 64, R = 15$ 时, 3 种权值曲线的变化趋势如图 2 所示。

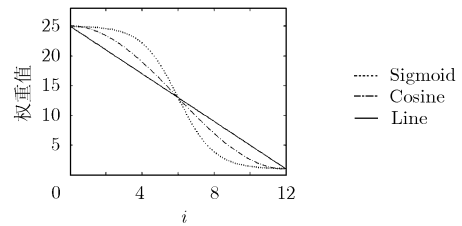


图 2 权值曲线

下面以线性曲线构造权值为例, 讨论加权改进最小统计算法的性能。

首先讨论只有噪声存在的情况。当噪声功率谱上升时, 考虑极限情况, 设噪声功率谱以 R (dB/s) 的速度上升, 所以当噪声功率以最大速度上升时, 经加权后的子窗最小值序列的变化曲线为直线, 这时加权后序列的最小值对应的值表示当前帧内的噪声功率谱; 当噪声功率谱不变或下降时, 显然加权后的算法将及时更新噪声功率谱。总的来说, WMS 算法在仅有噪声的情况下, 无论噪声功率谱是上升、下降还是保持不变, 都将根据当前帧所在子窗内的最小值更新噪声功率谱。

当带噪语音段为语音和噪声的叠加时, 引起带噪语音功率谱上升的原因主要有两个: 语音信号的出现和噪声信号功率谱的上升。由于语音信号与噪声信号相比较, 语音信号功率谱变化速度快。对于仅仅由语音引起带噪语音功率谱上升的情况, 经加权后的子窗最小值序列的变化曲线是递增的, 此时, 噪声谱估计值保持不变; 当语音信号功率谱在窗长 D 内升高后又降为零的情况下, 由于加权的缘故, 可能会在

语音信号功率谱的下降沿错误地更新噪声谱，产生对噪声谱的过估计。如图3所示。

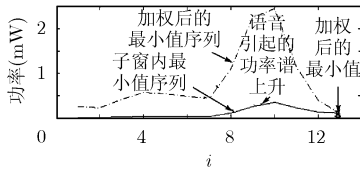


图3 加权最小值跟踪算法过估计举例

通过以上的分析可知，WMS产生过估计的主要原因是：当带噪声语音功率谱下降时，算法错误地更新了噪声功率谱。为解决此问题，当带噪声语音功率谱下降时采用以下策略控制对噪声功率谱的更新，减少过估计的出现：比较当前帧与上一帧的噪声谱估计值的大小，若当前帧大于上一帧的噪声功率谱估计值，则不更新噪声功率谱；否则，更新噪声功率谱。

4 实验及性能比较

本文通过实验的方法确定权值曲线及R的取值。同时为了验证改进算法对增强后语音质量的影响，采用文献[1]中的语音增强算法进行对比，通过客观测试和语谱图比较的方法，评估WMS算法与MS算法的性能。

实验中以NOIZEUS带噪语音料库^[4]和文献[5]中的典型非平稳噪声文件，作为本实验的语料。

这里，真实噪声功率谱 $\sigma_N^2(\lambda, k)$ 计算方法如下^[3]：

$$\sigma_N^2(\lambda, k) = \alpha\sigma_N^2(\lambda - 1, k) + (1 - \alpha)|D(\lambda, k)|^2 \quad (7)$$

其中 $\alpha = 0.9$ ， $D(\lambda, k)$ 表示噪声的DFT系数。

本文采用分段估计误差^[5](LOG-Err_{seg})评价改进噪声谱估计算法估计的准确性，并采用分段信噪比^[5](SNR_{seg})、感知语音质量评测^[6](PESQ)和改进的PESQ(mPESQ)来衡量^[7]噪声谱估计算法对语音增强系统的影响。

PESQ是目前常用的语音客观质量评测方法之一，mPESQ则是从信号失真度(Sig)、背景干扰度(Bak)和整体听觉质量(Ovl)3个方面对语音质量评估。

4.1 权值曲线的确定

通过对纯净语音信号的上升速率进行直方图统计，发现语音信号上升速率主要集中在25~55(dB/s)，如图4所示。因此，当R < 25时的取值都是合理的。

R选取不同的值，分别采用Sigmoid曲线，余弦曲线和线性曲线构造权值，采用本文描述的增强系统对典型非平稳

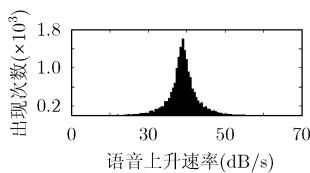


图4 语音上升速率统计直方图

噪声文件^[5]进行增强，通过非正式主观听辨，发现随着R取值的增大，增强后残余噪声越小，语音清音部分越不明显。上述3种权值构造曲线分别在R取21,21,20时增强后的语音质量较好。分别采用LOG-Err_{seg}和PESQ对噪声谱估计算法进行评价，如表1所示。

表1 权值曲线间的性能比较

权值曲线	R	LOG-Err _{seg}	PESQ
Sigmoid	21	1.6481	2.1392
Cosine	21	1.5914	2.1797
Line	20	1.5721	2.1795

表1说明线性曲线在R = 20时性能最优，但实验中发现该权重曲线在平稳噪声环境下会产生严重的过估计，如图5所示，这在实际中是不允许的。

4.2 改进算法性能测试

根据上面讨论的结果，这里采用R = 21时的余弦曲线作为权值曲线，其他参数设置如下U = 12，V = 8，fs = 8kHz，W = 256，SF = 64。WMS算法和MS算法均采用相同的参数。

WMS与MS算法相比较，WMS在噪声上升的阶段能够更好地跟踪到噪声的变化，MS算法有明显的时延现象，如图6所示。图6中给出了WMS算法和MS算法在k = 25的情况下噪声功率谱跟踪性能，对应频率分量为781.25 Hz，该频率分量在大部分时间内包含有语音能量。

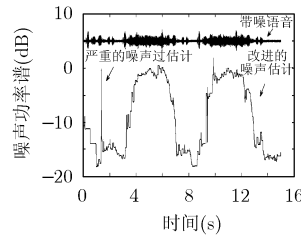


图5 R = 20时线性权值曲线过估计

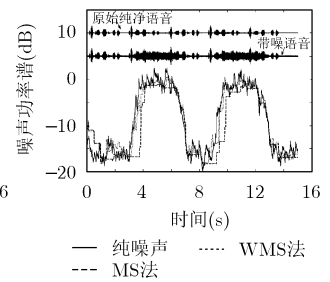


图6 WMS与MS算法跟踪性能比较

为了更进一步地说明WMS和MS算法性能之间的区别，将两种算法对噪声谱的估计值和增强后的语音进行比较，如表2所示。

表2 WMS算法与MS算法性能比较

	WMS	MS
LOG-Err _{seg}	1.6712	2.4581
mPESQ		
Sig	2.2559	2.1075
Bak	2.2036	2.0859
Ovl	2.0737	1.9038

截取带噪语音、MS 和 WMS 增强后的语音在噪声上升阶段(大约长 2s)的语谱图,如图 7 所示,可以明显地看出,MS 与 WMS 算法相比较,增强后的语音在噪声功率谱的上升阶段有明显的残余噪声。因此,对于加权改进的噪声谱估计算法,无论是估计误差还是增强后的语音质量都明显的要优于 MS 算法。

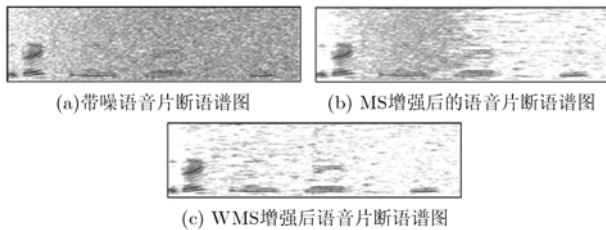


图 7 MS 与 WMS 算法增强后的语音片断语谱图比较

选取 NOIZEUS 语料库中的 4 种噪声、两种低信噪比下的文件,分别采用 MS 和 WMS 算法处理,采用分段估计误差和分段信噪比来比较两算法的性能,实验结果如表 3 所示。

表 3 不同噪声、信噪比下算法性能比较

噪声类型	SNR	LOG-Err _{seg}		SNR _{seg} (dB)	
		MS	WMS	MS	WMS
airport	0 dB	2.36	2.00	-2.01	-1.92
	5 dB	3.07	2.73	-0.21	0.23
babble	0 dB	2.90	2.30	-2.23	-1.91
	5 dB	3.22	2.90	0.32	0.73
car	0 dB	2.98	2.76	-1.56	-0.95
	5 dB	1.81	1.83	1.66	1.67
station	0 dB	2.53	2.31	-1.59	-1.01
	5 dB	2.12	1.97	1.65	1.82

表 3 说明改进算法在对非平稳的噪声跟踪性能较 MS 算法稍好,在近似平稳的噪声环境下,两算法的性能相似。

综上所述,改进算法不仅减少了噪声谱估计的误差,在某些噪声环境下(如噪声强度剧烈变化的噪声环境)也改善了语音增强算法的效果。

5 结束语

本文针对最小统计算法在噪声谱估计中的缺点,提出了一种基于加权改进的最小值跟踪算法(WMS),并给出了一种较优的权值计算方法。该算法能够改善最小统计算法在噪声上升阶段的跟踪时延,实验结果表明,无论是噪声谱的估计误差还是增强后语音的质量,尤其是在非平稳的噪声环境下,改进算法较原算法性能都有改善。

参考文献

- [1] Martin R, Malah D, and Richard V, *et al.* A noise reduction preprocessor for mobile voice communication [J]. *EURASIP Journal on Applied Signal Processing*, 2004, 2004(1): 1046-1058.
- [2] Loizou P C. *Speech Enhancement: Theory and Practice* [M]. Boca Raton, FL: CRC Press, 2007, Chapter 9.
- [3] Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics [J]. *IEEE Trans. on Speech, and Audio Processing*, 2001, 9(5): 504-512.
- [4] Hu Y and Loizou P C. Subjective comparison of speech enhancement algorithms [C]. *Proceedings of ICASSP-2006*, Toulouse, France, May 2006, vol.I: 153-156.
- [5] Hendriks R C, Jensen J, and Heusdens R. DFT Domain subspace based noise tracking for speech enhancement [C]. *INTERSPEECH 2007*, Antwerp, Belgium, August 2007: 830-834.
- [6] ITU. Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codes [S]. *ITU-T Recommendation P. 862*, 2000.
- [7] Hu Y and Loizou P C. Evaluation of objective quality measures for speech enhancement [J]. *IEEE Trans. on Audio, Speech, Language Process.*, 2008, 16(1): 229-238.

牛 铜: 男, 1983 年生, 硕士生, 研究方向为语音信号处理、语音增强。

张连海: 男, 1971 年生, 副教授, 硕士生导师, 研究方向为语音信号处理。

屈 丹: 女, 1974 年生, 讲师, 博士, 研究方向为语音信号处理。