Open Journal of Circuits and Systems 电路与系统, 2018, 7(3), 75-83
Published Online September 2018 in Hans. http://www.hanspub.org/journal/ojcs
https://doi.org/10.12677/ojcs.2018.73010

Hans汉斯

A New Single Channel Speech Enhancement Algorithm Based on Improved A-Priori SNR

Chen Chen, Ying Gao, Shun Zhang, Ruirui Han, Shuo Zhang

School of Opto-Electronic Information, Yantai University, Yantai Shandong
Email: 215302236@qq.com

Received: Aug. 11th, 2018; accepted: Aug. 24th, 2018; published: Aug. 31st, 2018

Abstract

Aiming at the problem of "music noise residue" existing in most speech enhancement algorithms, a new a-priori SNR estimation algorithm is proposed. Since the accuracy of the a-priori SNR estimation determines the overall performance of the speech enhancement system, the Convex-Combination (CC) algorithm is the most widely used a-priori SNR estimation algorithm. Although its real-time performance and distortion are small, its ability to suppress music noise is lacking. In order to solve this defect, this paper will improve the part of the maximum likelihood estimation in a-priori SNR estimation, and recursively smooth the a-posteriori signal-to-noise ratio by incorporating smoothing parameters, instead of the a-posteriori signal-to-noise ratio in the maximum likelihood estimation. The simulation results show that the proposed algorithm has better music noise suppression ability than CC algorithm.

Keywords

Speech Enhancement, A-Priori Signal-to-Noise Ratio, Fusion Coupling Factor, Maximum Likelihood Estimation

基于改进先验信噪比的新型单声道语音增强算法

陈　晨，高　颖，张　顺，韩蕊蕊，张　硕

烟台大学，光电信息科学技术学院，山东　烟台
Email: 215302236@qq.com

收稿日期：2018年8月11日；录用日期：2018年8月24日；发布日期：2018年8月31日

摘　要

针对多数语音增强算法中存在的“音乐噪声残留”问题，提出一种新型先验信噪比估计算法。由于先验

文章引用: 陈晨, 高颖, 张顺, 韩蕊蕊, 张硕. 基于改进先验信噪比的新型单声道语音增强算法[J]. 电路与系统, 2018, 7(3): 75-83. DOI: 10.12677/ojcs.2018.73010

信噪比的估计准确度决定语音增强系统的整体性能，而融合耦合因子(CC, Convex-Combination)算法是应用最广的先验信噪比估计算法。虽然其实时性强且失真小，但其抑制音乐噪声能力欠缺。为解决这一缺陷，本文将改进先验信噪比估计中的最大似然估计部分，通过融入平滑参数将后验信噪比递归平滑，代替最大似然估计中的后验信噪比。经仿真实验结果证明，所提出的算法相对于CC算法具有更好的音乐噪声抑制能力。

关键词

语音增强，先验信噪比，融合耦合因子，最大似然估计

1. 引言

语音交流是人与人，人与机器之间沟通最方便快捷的媒介之一。但是移动通信过程中总是无法避免地出现由非交流者带来的外界噪音，如各种交通工具产生的交通噪声，工厂设备产生的工厂噪声，电子热噪声，环境噪声等等。正是由于这些形形色色的噪声干扰，使得接收端语音识别系统受到损伤，准确性大大降低，严重影响了语音通信系统的质量和可理解性。因此，在语音处理领域抑制噪声干扰的语音增强技术应运而生，不断受到学者的高度重视[1]。同时，在研究过程的深入渗透下，语音增强算法应用领域广阔延伸，如：移动电话，助听器设备，军事窃听技术，语音编码与合成技术等[2]。

在过去的五十年历史中，为了更好地适应科学领域发展，大量的短时频域语音增强算法已经逐渐衍生出来并得到了广泛应用。其中较为著名的算法有：谱减算法[3]、短时谱估计算法[4]、子空间算法[5]等。由于噪声信号的随机性和非平稳特性，很多算法在真实环境下的运行效果会受到阻碍，因此单声道语音增强算法面临着亟待攻克的问题。众多算法中，经研究发现几乎所有的语音增强算法都与增益因子息息相关，而增益因子又是先验信噪比和后验信噪比的二元函数[6]，由于后验信噪比在算法中是已知参数，因而一个准确的先验信噪比估计在增强结果中扮演着关键性的角色。应用性较广的先验信噪比估计算法有直接判决(DD，Decision-Directed)算法[4]，两步噪声消除(TSNR，Two-Step Noise Reduction)算法[7]、改进的直接判决(MDD，Modified Decision Directed)算法[8]，融合耦合因子(CC，Convex-Combination)算法[9]等。

DD 算法由于计算简洁且容易实现，是迄今最为普及的先验信噪比估计算法。由于该算法中纯净语音谱与噪声谱相互正交的不合理假设，以及采用最大似然估计算法估计当前帧先验信噪比过程中在跟踪后验信噪比时引起一帧的时延，使得该算法音乐噪声较大。针对此缺点，后续有人提出一系列改进算法。其中运行效果较理想的是融合耦合因子算法。该算法在 DD 算法的基础上引入两个不同平滑参数取值的先验信噪比估计，融入一个耦合参数进行调和，在实际和估计的先验信噪比中建立代价函数求出耦合参数真实值，最终得到新的先验信噪比估计值。该算法有效避免了时延问题，能够实时跟踪信噪比的快速变化，同时失真程度大大降低。但是由于该算法在 DD 算法估计中采用最大似然法对后验信噪比估计，以此代替当前帧的先验信噪比估计。这使得在无语音活动区产生较大波动，输出的语音信号残留孤立峰值居多，继而产生“音乐噪声”。为了解决上述问题，本文将融合耦合因子算法进行改进，用递归平滑的方式计算后验信噪比，代替传统的瞬时后验信噪比并带入最大似然估计中，有效减少了信号的波动，

同时音乐噪声抑制能力有所提升。

本文首先介绍了语音增强算法的基本理论，并对经典的融合耦合因子的先验信噪比估计算法进行了理论分析；其次，提出改进算法并做出理论和公式推导；最后，用 Matlab 进行实验仿真，分析实验结果验证理论部分，并作出总结。

2. 语音增强算法的基本理论

在语音信号的短时平稳特性下假设原始纯净语音信号与噪声信号是不相关的，则有[10]：

$$y(t)=x(t)+d(t) \tag{1}$$

其中，$y(t)$ 代表带噪语音信号，$y(t)$ 与 $d(t)$ 分别表示纯净语音信号和噪声。

对等式两侧分别进行 STFT 变换，将其转换到频域中：

$$Y_{m,k}=X_{m,k}+D_{m,k} \tag{2}$$

其中，$Y_{m,k}$，$X_{m,k}$，$D_{m,k}$ 分别表示为带噪语音谱、纯净语音谱和噪声谱。m 和 k 表示为帧数和频率。

任何语音增强算法均可以表示为增益因子与带噪语音谱的乘积，即为：

$$\hat{X}_{m,k}=G_{m,k}\cdot Y_{m,k} \tag{3}$$

由于不同算法拥有不同形式的增益因子，而只有维纳滤波算法增益因子不受其他参数影响，仅与先验信噪比有关，为了方便且不失一般性，一般采用维纳滤波算法表示的增益因子[11]：

$$G_{m,k}=\frac{\xi_{m,k}}{1+\xi_{m,k}} \tag{4}$$

最终结合带噪语音相位对纯净语音谱进行 ISTFT 变换可以得到估计的纯净语音信号。

先验信噪比估计通常采用 DD 算法，其定义为：

$$\hat{\xi}_{m,k}^{DD}=\lambda\hat{\xi}_{m-1,k}+(1-\lambda)\cdot\hat{\xi}_{m,k}^{ML} \tag{6}$$

其中 λ 表示平滑参数，取值范围在 0 到 1 之间。$\hat{\xi}_{m-1,k}$ 表示前一帧先验信噪比估计值，$\hat{\xi}_{m,k}^{ML}$ 表示最大似然估计下的当前帧先验信噪比，表示为 $\hat{\xi}_{m,k}^{ML}=\max\{\eta_{m,k}-1,0\}$，$\eta_{m,k}$ 为后验信噪比。

由上式可见，该估计值分为两部分：前部分是上一帧的先验信噪比估计值，后部分是当前帧先验信噪比的估计值，平滑参数在两部分中起到调节作用。当取值趋于 1 时，估计值由上一帧的估计结果决定，会出现帧延迟现象，并带来较为严重的语音失真。取值趋向 0 时，估计值则主要由最大似然方法的估值决定，此时在静音区波动剧烈，进一步引发音乐噪声。

由此可见，传统 DD 算法计算简单，能够很好地滤除背景噪音。但由于采用 ML 算法在追踪后验信噪比时产生一帧的延时，带来恼人的音乐噪声，导致估计效果准确性降低。为此，有学者提出一种融合耦合因子的先验信噪比估计算法，即将平滑参数取大值和小值的优点结合，融入一个耦合参数来达到实时性跟踪。

平滑因子取值分别为 a 和 b ($a>b$)的两个 DD 算法估计的先验信噪比分别为：

$$\hat{\xi}_{m,k}^{1}=a\hat{\xi}_{m-1,k}+(1-a)\hat{\xi}_{m,k}^{ML} \tag{7}$$

$$\hat{\xi}_{m,k}^{2}=b\hat{\xi}_{m-1,k}+(1-b)\hat{\xi}_{m,k}^{ML} \tag{8}$$

其中 $\hat{\xi}_{m,k}^{1}$ 近似于 DD 算法对前一帧先验信噪比的估计值，$\hat{\xi}_{m,k}^{2}$ 的取值近似于最大似然方法估计的当前帧先验信噪比估计值。在两个先验信噪比中加入一个取值范围在[0,1]之间的耦合因子，控制该算法的估计值

取值情况，则 CC 算法定义如下：

$$\hat{\xi}_{m,k}^{CC}=\mu\hat{\xi}_{m,k}^{1}+(1-\mu)\hat{\xi}_{m,k}^{2}=(\mu a-\mu b+b)\hat{\xi}_{m-1,k}+(1-\mu a+\mu b-b)\hat{\xi}_{m,k}^{ML} \tag{9}$$

为得到自适应耦合参数，在实际先验信噪比与先验信噪比估计值之间的最小均方误差准则下建立一个代价函数：

$$J=E\left\{\left(\hat{\xi}_{m,k}^{CC}-\xi_{m,k}\right)^{2}\right\} \tag{10}$$

通过对代价函数求偏导数并运用最大似然估计方法得到的当前帧的先验信噪比估计代替先验信噪比真实值，得到该耦合参数[12]：

$$\mu=\frac{(1-b)\left(\hat{\xi}_{m,k}^{ML}+1\right)^{2}-b\left(\hat{\xi}_{m-1,k}-\hat{\xi}_{m,k}^{ML}\right)^{2}}{(a-b)\left[\left(\hat{\xi}_{m-1,k}-\hat{\xi}_{m,k}^{ML}\right)^{2}+\left(\hat{\xi}_{m,k}^{ML}+1\right)^{2}\right]} \tag{11}$$

将耦合因子带入定义式(9)，可得到 CC 算法的先验信噪比估计，进一步求出增益因子，与带噪语音谱相乘后再进行 IDFT 变换即可得到增强后的时域语音信号。该算法通过自适应地结合两个具有不同平滑参数取值的 DD 算法，有效减少了失真，具有实时跟踪性能。但由于最大似然估计算法对当前帧先验信噪比估计，静音区波动较大，易产生音乐噪声。

3. 改进先验信噪比的新型算法

为解决上述缺点，可以在算法中使用平滑处理的后验信噪比来代替 ML 算法估计的后验信噪比。由于最大似然估计取决于后验信噪比的值，为了减少快速波动，将后验信噪比估计值进行递归平滑，即为：

$$\hat{\gamma}_{m,k}=\beta\cdot\hat{\gamma}_{m-1,k}+(1-\beta)\cdot\min\left[\gamma_{m,k},20\right] \tag{12}$$

其中，β 代表经验平滑常数，取值为 0.6，$\hat{\gamma}_{m-1,k}$ 表示前一帧的后验信噪比估计值，$\gamma_{m,k}$ 表示瞬时后验信噪比，即为带噪语音功率谱与噪声谱估计的比值求得，最小值函数是为了限制后验信噪比估计值的上限，最大不能超过 13 dB (=10lg(20))，同时避免信号的过度衰减[13]。过去的实验研究发现，后验信噪比的平滑处理了改进了带噪语音功率谱在均方误差意义下的估计[13]。

将平滑处理后的后验信噪比估计值带入 ML 算法后再代替公式(7)和(8)中的 $\hat{\xi}_{m,k}^{ML}$，即：

$$\hat{\xi}_{m,k}^{1}=a\hat{\xi}_{m-1,k}+(1-a)\hat{\gamma}_{m,k}^{ML} \tag{13}$$

$$\hat{\xi}_{m,k}^{2}=b\hat{\xi}_{m-1,k}+(1-b)\hat{\gamma}_{m,k}^{ML} \tag{14}$$

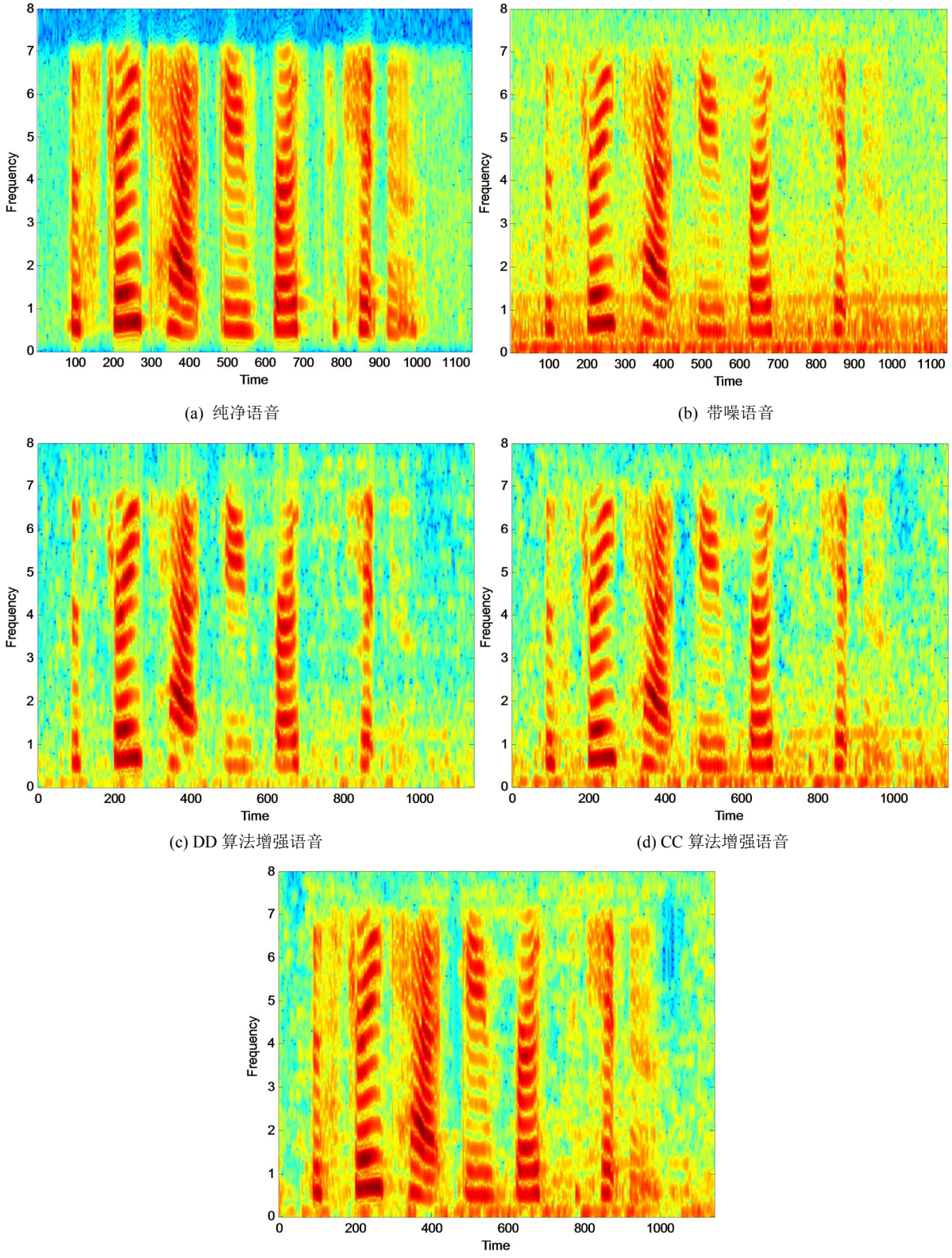
其中 $\hat{\gamma}_{m,k}^{ML}=\max\left\{\hat{\gamma}_{m,k}-1,0\right\}$ 为最大似然估计方法得到的当前帧先验信噪比估计值。

4. 仿真结果比较

为了进一步证明改进算法相对于传统算法的优越性，通过 Matlab 实验仿真得到语谱图和客观评价标准数据进行对比。纯净语音来自于语音库中选取的 6 段语音(其中 3 段男声；3 段为女声)，5 种噪声(White, Pink, Buccaneer2, F16, M109)来自于 Noisex-92 噪声库，输入信噪比分别为 0 dB，5 dB，10 dB，15 dB。实验中采用汉明窗进行加窗分帧处理，采样频率 8kHz，帧长为 256，重叠率 50%，λ 取 0.98，a 和 b 分别为 0.99 和 0.60。

图 1 中(a)至(e)分别为纯净语音信号，带噪语音信号，DD 算法，CC 算法以及改进算法下增强的语音信号的语谱图对比，其中背景噪声为 M109 噪声，信噪比水平为 10 dB。

针对以上仿真语谱图不难看出：三个算法都能有效消除背景噪声，但是 DD 算法在消除背景噪声的



(a) 纯净语音

(b) 带噪语音

(c) DD 算法增强语音

(d) CC 算法增强语音

(e) 改进算法增强语音

Figure 1. The spectrum of speech signal of different algorithms under M109 noise (SNR = 10 dB)
图 1. M109 噪声下不同算法的语音信号语谱图(SNR = 10 dB)

基础上语音失真更为严重，尤其是在信噪比水平较低的环境下失真更显著。CC 算法语音失真情况虽然较之 DD 算法有所提升，但是最终效果不如改进算法理想。同时，三种算法都对原始语音造成不同程度的损伤，相较来说，改进算法与原始纯净语音语谱图更加接近，即改进算法对纯净语音损伤程度最小。即改进算法的去噪能力更彻底，因此验证了理论部分的分析。

为了客观定量比较分析三种算法的性能，对三种算法增强后语音的质量、失真程度和可懂度等各种标准进行测试，常用的评价标准有分段信噪比(SegSNR) [14]，短时客观可懂度(STOI) [15]和对数谱距离(LSD) [16]等。分段信噪比大小表征算法滤除噪声的能力，数值越大说明算法残余音乐噪声越少。STOI 是与人的听力特性最契合的评价标准，数值越大表明增强语音质量越好。LSD 表明增强语音和原始纯净语音的接近程度，数值越小说明失真程度越小，即增强效果越好。表 1~表 3 即为三个算法在四种信噪比水平和五种背景噪声下的客观评价数值情况。

通过三个表可得，在不同背景噪声环境和输入信噪比条件下，改进算法的 SegSNR 和 STOI 数据最高，CC 算法次之，DD 算法的数值最低。说明改进算法可以更大限制地抑制背景噪声，提高增强后语音的信噪比水平，增强后的语音可懂度更高。在 LSD 输出数据上，改进算法比其他两个算法数据值更小，说明改进算法增强后的语音与原始语音更接近，同时和语谱图的结果相吻合。综上所述，改进算法比 CC 算

Table 1. The SegSNR data comparison table of the four algorithms
表 1. 四种算法的 SegSNR 数据对比表

噪声类型	输入信噪比	分段信噪比(SegSNR)		
		DD算法	CC算法	改进算法
White	0 dB	4.3907	5.8005	7.1560
	5 dB	5.8434	6.6359	7.8063
	10 dB	7.0984	7.7320	8.6589
	15 dB	9.1980	10.811	11.2345
Pink	0 dB	5.1565	5.2808	6.6384
	5 dB	6.1975	6.2242	7.4263
	10 dB	7.5781	7.6449	8.6095
	15 dB	10.2558	10.7721	11.338
Buccaneer2	0 dB	4.1343	4.9211	6.1662
	5 dB	5.2256	5.7149	6.8133
	10 dB	6.8214	7.7139	8.5504
	15 dB	8.7949	10.4505	10.8979
F16	0 dB	5.2623	5.3991	6.7674
	5 dB	6.4201	6.5044	7.7754
	10 dB	7.8656	7.9391	9.0580
	15 dB	10.5363	10.9798	11.6437
M109	0 dB	6.2730	6.8603	8.6496
	5 dB	7.4082	7.9176	9.3838
	10 dB	8.9422	9.4135	10.6136
	15 dB	11.4692	12.5741	13.3508

Table 2. The STOI data comparison table of the four algorithms
表 2. 四种算法的 STOI 数据对比表

噪声类型	输入信噪比	短时客观可懂度(STOI)		
		DD算法	CC算法	改进算法
White	0 dB	72.2350	75.9131	77.5240
	5 dB	79.8583	81.7923	82.8082
	10 dB	85.4420	85.7764	86.5817
	15 dB	89.7984	89.6146	90.4553
Pink	0 dB	73.0863	76.5339	77.2854
	5 dB	79.3534	82.4027	83.0000
	10 dB	85.9826	87.1785	87.8651
	15 dB	90.3885	90.4289	91.0242
Buccaneer2	0 dB	70.8405	75.0766	75.7672
	5 dB	76.9853	80.9993	81.5867
	10 dB	83.5680	85.1948	85.3056
	15 dB	88.4355	88.8371	88.8654
F16	0 dB	76.2078	79.5544	80.7624
	5 dB	82.5660	85.0770	86.1478
	10 dB	88.1948	89.2560	90.5462
	15 dB	91.7398	92.1378	93.2722
M109	0 dB	76.2916	79.1938	79.9584
	5 dB	82.0582	84.7726	85.9888
	10 dB	87.9739	88.0564	89.5696
	15 dB	90.3163	90.3722	91.5371

Table 3. The LSD data comparison table of the four algorithms
表 3. 四种算法的 LSD 数据对比表

噪声类型	输入信噪比	对数谱距离(LSD)		
		DD算法	CC算法	改进算法
White	0 dB	6.1261	5.7423	5.0706
	5 dB	5.5611	5.3879	4.8288
	10 dB	5.4087	5.1224	4.6612
	15 dB	5.0679	4.3266	4.0306
Pink	0 dB	5.2437	5.1120	4.4945
	5 dB	4.9327	4.9294	4.3898
	10 dB	4.8238	4.4972	4.0886

Continued

	15 dB	4.4760	3.7501	3.5135
Buccaneer2	0 dB	5.9493	5.4718	4.8267
	5 dB	5.6324	5.2619	4.6566
	10 dB	5.3918	4.6995	4.2640
	15 dB	5.0920	4.0812	3.8244
F16	0 dB	4.9231	4.8705	4.3207
	5 dB	4.7659	4.6650	4.2287
	10 dB	4.6171	4.3696	4.0078
	15 dB	4.2148	3.6763	3.4605
M109	0 dB	4.3395	3.4064	3.1264
	5 dB	4.1441	3.4095	3.1283
	10 dB	4.1441	3.0336	2.8718
	15 dB	3.6170	2.5228	2.4159

法的增强效果更具有优越性，进一步证实了理论部分。

5. 结论

由于各种杂乱噪声的干扰，涌现出越来越多单声道语音算法，高精度的先验信噪比估计值对语音增强系统的性能好坏起到关键性作用。由于传统的先验信噪比估计算法在跟踪后验信噪比过程中采用极大似然估计方法会出现一帧的延时，并产生音乐噪声。为了切实解决这种弊端，本文通过递归平滑的方式对后验信噪比进行估计，得到新的增强算法。最后实验仿真结果证明了改进算法有更实时的跟踪信噪比变化的性能以及更好的增强效果。

参考文献

[1] 刘伟, 陈晨, 高颖. 一种融合相位信息先验信噪比估计算法的研究[J]. 电声技术, 2017, 41(11/12): 84-87.

[2] Cho, J.W. and Park, H.M. (2016) Independent Vector Analysis Followed by HMM-Based Feature Enhancement for Robust Speech Recognition. *Signal Process*, **120**, 200-208. https://doi.org/10.1016/j.sigpro.2015.09.002

[3] Boll, S.F. (1979) Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Transactions on Acoustics*, *Speech*, *and Signal Processing*, **27**, 113-120. https://doi.org/10.1109/TASSP.1979.1163209

[4] Ephraim, Y. and Malah, D. (1984) Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Transaction on Acoustic Speech Signal Process*, **32**, 1109-1121. https://doi.org/10.1109/TASSP.1984.1164453

[5] Ephraim, Y. and Harry, L.V.T. (1995) A Signal Subspace Approach for Speech Enhancement. *IEEE Transactions on Speech and Audio Processing*, **3**, 251-266. https://doi.org/10.1109/89.397090

[6] 孙海东. 基于新型先验信噪比估计的语音增强算法研究[D]: [硕士学位论文]. 烟台: 烟台大学, 2015.

[7] Plapous, C. and Marro, C. (2006) Improved Signal-to-Noise Ratio Estimation for Speech Enhancement. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, **14**, 2098-2108. https://doi.org/10.1109/TASL.2006.872621

[8] Yong, P.C., Nordholm, S. and Dam, H.H. (2013) Optimization and Evaluation of Sigmoid Function with A Priori SNR Estimate for Real-Time Speech Enhancement. *Speech Communications*, **55**, 358-376. https://doi.org/10.1016/j.specom.2012.09.004

[9] Shen, S., Ou, S., Wei, J., *et al*. (2017) A Priori SNR Estimator Based on a Convex Combination of Two DD Ap-

proaches for Speech Enhancement. 2016 *IEEE International Conference on Signal and Image Processing*, Beijing, 13-15 August 2016, 750-754.

[10] Hasan, T. and Hasan, Md.K. (2010) MMSE Estimator for Speech Enhancement Considering the Constructive and Destructive Interference of Noise. *IEI Signal Processing*, **4**, 1-4. https://doi.org/10.1049/iet-spr.2008.0114

[11] 陈国明. 语音增强技术研究[D]: [博士学位论文]. 南京: 东南大学, 2007.

[12] 沈锁金. 语音增强技术中的先验信噪比估计算法研究[D]: [硕士学位论文]. 烟台: 烟台大学, 2017.

[13] Lu, Y. and Loizou, P.C. (2008) A Geometric Approach to Spectral Subtraction. *Speech Communication*, **50**, 453. https://doi.org/10.1016/j.specom.2008.01.003

[14] Sun, H., Ou, S., Liu, R., *et al.* (2015) A Variable Momentum Factor Algorithm for a Priori SNR Estimation in Speech Enhancement. 2014 7*th International Congress on Image and Signal Processing*, Dalian, 14-16 October 2014, 888-892.

[15] Taal, C.H., Hendriks, R.C., Heusdens, R., *et al.* (2011) An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech. *IEEE Transactions on Audio Speech & Language Processing*, **19**, 2125-2136. https://doi.org/10.1109/TASL.2011.2114881

[16] Pei, C.Y., Nordholm, S. and Hai, H.D. (2013) Optimization and Evaluation of Sigmoid Function with A Priori SNR Estimate for Real-Time Speech Enhancement. *Speech Communication*, **55**, 358-376. https://doi.org/10.1016/j.specom.2012.09.004