Open Journal of Circuits and Systems 电路与系统, 2018, 7(2), 25-35
Published Online June 2018 in Hans. http://www.hanspub.org/journal/ojcs
https://doi.org/10.12677/ojcs.2018.72004

Hans汉斯

Comparison and Analysis of a Priori SNR Parameter Estimation Algorithm in Single Channel Speech Enhancement

Chen Chen, Ying Gao, Wei Liu, Ruirui Han, Shuo Zhang

School of Opto-Electronic Information, Yantai University, Yantai Shandong
Email: claragaoying@126.com

Received: May 10th, 2018; accepted: May 28th, 2018; published: Jun. 4th, 2018

Abstract

The accuracy of a priori signal-to-noise ratio parameter estimation is a key factor in determining the output performance of the speech enhancement system in the noise background. The Decision-Directed (DD) technique is the simplest and straightforward algorithm in the a priori signal-to-noise ratio estimation system. Subsequent algorithms are mostly the further optimization or improvement of this technique. This paper deals with four commonly used direct decision algorithms, Two-step Noise Reduction (TSNR) algorithm, Modified Two-step Noise Cancellation (MTSNR) algorithm, and Convex Combination (CC) algorithm. A priori signal to noise ratio technology was compared and analyzed. The basic design principles of various algorithms were given. The output performance and advantages and disadvantages of the four algorithms were discussed from the aspects of theoretical analysis and experimental simulation.

Keywords

Speech Enhancement, A Priori Signal-to-Noise Ratio, Direct Decision Algorithm, Two-Step Noise Reduction Algorithm

单信道语音增强中先验信噪比参数估计算法的对比分析

陈　晨，高　颖，刘　伟，韩蕊蕊，张　硕

烟台大学，光电信息科学技术学院，山东　烟台
Email: claragaoying@126.com

收稿日期：2018年5月10日；录用日期：2018年5月28日；发布日期：2018年6月4日

文章引用: 陈晨, 高颖, 刘伟, 韩蕊蕊, 张硕. 单信道语音增强中先验信噪比参数估计算法的对比分析[J]. 电路与系统, 2018, 7(2): 25-35. DOI: 10.12677/ojcs.2018.72004

摘　要

先验信噪比参数估计的准确性是决定噪声背景下语音增强系统输出性能的关键因素。直接判决(Decision-Directed, DD)技术是先验信噪比估计体系中最为简易直接的算法，后续算法多为对此技术的进一步优化或改进。本文对常用的直接判决算法、两步噪声消除(Two-step Noise Reduction, TSNR)算法、改进的两步噪声消除(Modified TSNR, MTSNR)算法以及融合耦合因子(Convex Combination, CC)算法等四种先验信噪比技术进行了对比分析，给出了各种算法的基本设计原理，并从理论分析和实验仿真两个方面讨论了四种算法的输出性能及其优缺点。

关键词

语音增强，先验信噪比，直接判决算法，两步噪声消除算法

Copyright © 2018 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

现实中背景噪声的存在往往会对语音增强系统造成较大损伤，尤其在听觉场景复杂的环境中，噪声污染下的原始语音信号给人类带来听觉损伤，严重影响通信质量。因此消除语音通信系统中的背景噪声，增强语音识别的准确率成为研究重点。单信道语音增强是语音信号处理的关键分支，其应用技术的研究具有重要的适用价值，尤其在语音识别，医疗领域，军事通信，数字家电等领域已被广泛应用[1]。为了提高移动环境下的通信品质，实现实时通信，需要在传输到远端之前有效抑制背景噪声[2]。虽然语音增强技术看似只是一个恢复纯净语音的简易过程，但在学术领域涉及到的众多技术和方法是不容小觑的。经过多年探索，涌现出越来越多的语音增强算法，代表性的算法有谱减算法，维纳滤波算法，最小均方误差算法，小波变换算法等等[3]。

研究表明，在几乎所有的语音增强算法中，先验信噪比参数的估计是最为重要的部分之一[2]。先验信噪比是语音增强系统增益因子的函数，而纯净语音谱估计是由带噪语音频谱与系统增益因子的乘积得到，因此先验信噪比估计精度将在较大程度上影响语音增强系统的总体输出性能[4]。先验信噪比估计最经典的算法是由 Ephraim 和 Malah 提出的直接判决算法[5]，该算法以低复杂度及低音乐噪声著称，其应用相当广泛。然而该算法的估计结果会引入一帧的延迟，从而降低了系统降噪的性能。为了改进 DD 算法，Plapous 等人提出了两步噪声消除(TSNR)算法[6]，它两次运用 DD 算法结果，先求出增益因子，再利用增益因子结果进一步修正语音的当前帧先验信噪比，获得基于 TSNR 算法的先验信噪比估计。该算法虽然避免了延迟问题，但是过于依赖增益因子，在应用上产生较大局限，并且导致结果出现短时间的频谱峰值，降低了语音的频谱特性。为了进一步克服算法的不足，近年来又有学者在 TSNR 算法上进行了改进，提出改进的两步噪声消除算法[7]。该算法在消除系统对增益因子依赖的基础上，直接利用 DD 算法估计结果计算当前帧的先验信噪比，大大简化了计算过程，但是该算法会受平滑参数影响，无法自适应于不同的环境。随着算法的改进，近几年有人提出融合耦合因子的先验信噪比估计算法[8]，经过大量实验验证后选取两个大小不同的平滑参数，结合 DD 算法求出不同平滑参数的先验信噪比，再加入一

个耦合因子进行折衷，最终得到新算法的先验信噪比估计。该算法在对原始语音低损伤的情况下，有效滤除了背景噪声，同时减少了语音失真。

本文对当今具有代表性的先验信噪比算法进行了详尽研究，给出了其设计原理，并通过理论分析和实验验证讨论了各个算法的优缺点，同时给出了先验信噪比估计算法今后的改进方向。本文的结构如下：第二部分简单描述了语音增强算法在 DFT 域的基本理论，第三部分回顾了几种应用广泛的先验信噪比估计算法，进行了理论分析与对比，第四部分通过实验仿真的语谱图，时域波形图以及三种客观评价标准：分段信噪比(Segmental SNR, SegSNR)，短时客观可懂度(Short Time Objective Intelligibility, STOI)和对数谱距离(Log-Spectral Distortion, LSD)结果定量分析，进一步比较几种算法的优缺点，最后做出总结并对改进方向进行分析。

2. DFT 域语音增强算法基本理论

假定在 t 时刻的带噪语音信号为 $y(t)$，其由互不相关的原始纯净语音信号 $x(t)$和加性噪声 $n(t)$叠加而成[9]，即：

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

将该时域语音信号变换到 DFT 域，表示如下：

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

式中 $Y_{m,k}$，$X_{m,k}$ 和 $N_{m,k}$ 分别表示带噪语音频谱、纯净语音频谱和噪声谱，m，k 分别表示帧索引和频率。

语音增强的目的是滤除背景噪声并从带噪语音谱中尽可能地提取出纯净语音谱分量。一般情况下，纯净语音谱的估计可由一个非线性增益函数与带噪语音谱的乘积得到[10]，即为：

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

其中 $G_{m,k}$ 是增益函数。其作用是对带噪语音信号进行衰减以得到估计的纯净语音谱。由于估计的纯净语音谱与原始纯净语音谱之间的代价函数形式不同，因此会产生不同形式的增益因子。但是几乎所有形式的增益因子都是先验信噪比与后验信噪比的二元函数，表示为：

$$G_{m,k}=f\left(\xi_{m,k},\eta_{m,k}\right) \tag{4}$$

其中先验信噪比和后验信噪比的定义如下：

$$\xi_{k,m}=\frac{E\left\{\left|X_{k,m}\right|^2\right\}}{E\left\{\left|N_{k,m}\right|^2\right\}}=\frac{\lambda_X(m,k)}{\lambda_N(m,k)} \tag{5}$$

$$\eta_{k,m}=\frac{\left|Y_{k,m}\right|^2}{\lambda_N(m,k)} \tag{6}$$

其中 $\lambda_N(m,k)$表示噪声方差，它可由语音活动检测技术在无语音区检测更新，多假设其为已知条件。在此基础上我们可见，增益函数在主要依赖于系统对先验信噪比参数的估计结果。由于维纳滤波语音增强算法的系统增益因子仅为先验信噪比参数的函数，故不失一般性，本文在对先验信噪比算法进行对比分析时，语音增强系统的增益因子选取如下[11]：

$$G_{m,k}^{WF}=\frac{\xi_{m,k}}{1+\xi_{m,k}} \tag{7}$$

通过合适的算法计算 $\xi_{m,k}$ 后，结合式(3)和式(7)可获得维纳滤波语音增强系统的输出 $\hat{X}_{m,k}$，再将其通过 N

点 IDFT 变换至时域即可最终得到纯净语音信号的估计。

3. 几种先验 SNR 估计算法

如前文所述，先验信噪比为语音增强算法的核心参数，本节将重点讨论几种常用的先验信噪比估计算法。

将前一帧语音谱幅度中第 k 个分量的纯净语音信号估计用 $\hat{X}_{m-1,k}$ 表示，则 DD 算法可表示为[5]：

$$\hat{\xi}_{m,k}^{\mathrm{DD}}=\alpha\frac{\left|\hat{X}_{m-1,k}\right|^2}{\lambda_N\left(m,k\right)}+\left(1-\alpha\right)\max\left\{\eta_{m,k}-1\right\} \tag{8}$$

式中 m 为帧数，max(.)表示求最大值的函数，用于确保值的非负性。α 表示取值范围在 0 到 1 之间的平滑参数。当取值接近于 0 时，先验信噪比估计结果近似于最大似然估计方法得到的当前帧的先验信噪比估计，而当取值接近于 1 时，估计结果近似于前一帧的先验信噪比估计，所以平滑参数为这两部分的平衡参数。按照文献[6]中的分析，一般将 α 的值设置为 0.98。$\hat{\xi}_{m,k}^{\mathrm{DD}}$ 带入系统增益因子公式，得到 DD 算法的增益函数，

$$G_{m,k}^{\mathrm{DD}}=\frac{\hat{\xi}_{m,k}^{\mathrm{DD}}}{\hat{\xi}_{m,k}^{\mathrm{DD}}+1} \tag{9}$$

则 DD 算法增强语音谱为

$$\hat{X}_{m,k}^{\mathrm{DD}}=G_{m,k}^{\mathrm{DD}}\times Y_{m,k} \tag{10}$$

DD 算法应用相当广泛，它计算简单并且可以有效抑制音乐噪声，但其缺点有以下几点：

1、DD 算法的估计结果在很大程度上依赖于平滑参数取值大小，系统难以适应不同的环境，导致估计结果出现偏差。

2、系统增益函数匹配的先验信噪比是前一帧的语音谱估计而不是当前帧，导致无法实时跟踪瞬时信噪比。

3、在语音起始和结束的阶段，先验信噪比无法快速改变以适应系统性能，造成增强后的语音在听觉上有混响，在语音活动期间降低了噪声消除性能[9]。

为了避免 DD 算法中出现的延时问题，Plapous 等人基于各种假设和理论提出了 TSNR 算法[6]，这个算法利用 DD 算法的估计结果，分两步进行先验信噪比的估计计算。具体的估计过程如下：

第一步，利用 DD 算法得到先验信噪比的估计结果 $\hat{\xi}_{m,k}^{\mathrm{DD}}$，将该结果带入维纳滤波增益函数中计算出系统增益因子 $G_{m,k}^{\mathrm{DD}}=\hat{\xi}_{m,k}^{\mathrm{DD}}/\hat{\xi}_{m,k}^{\mathrm{DD}}+1$；第二步，对先验信噪比的估计进行细化，去除 DD 算法的偏差，从而去除混响效应。结合带噪语音谱与噪声方差，将系统增益因子带入求先验信噪比定义的公式中，得到基于此算法的先验信噪比估计：

$$\hat{\xi}_{m,k}^{\mathrm{TSNR}}=\frac{\left|G_{m,k}^{\mathrm{DD}}\cdot Y_{m,k}\right|^2}{\lambda_N\left(m,k\right)}=\frac{\left|\hat{X}_{m,k}\right|^2}{\lambda_N\left(m,k\right)} \tag{11}$$

将 TSNR 算法计算出的先验信噪比估计结果带入系统增益因子，进一步得到 TSNR 算法增强的语音谱。由此可见，TSNR 算法实际上是两次运用 DD 算法结果，先求出系统增益因子，再利用增益因子的结果进一步修正当前帧语音的先验信噪比。在瞬时信噪比突变之前已经估计出了下一帧的先验信噪比，来代替当前帧的信噪比，这种超前估计有效解决了 DD 算法中出现的延时问题，同时在一定程度上减少了语音失真。但是，TSNR 算法估计的先验信噪比在无语音阶段波动较大，通常会在短时间内产生谱峰，

这将破坏频谱异常值，同时这种算法计算复杂度相对较高，过于依赖增益因子，因此系统无法适应不同的环境，进而降低语音增强系统的性能。

为了进一步克服算法的不足，近年来又有学者在 TSNR 算法上进行了改进，提出改进的两步噪声消除算法[7]。由于 TSNR 算法在估计纯净语音谱时采用了系统增益因子，造成计算量增加，为了简化计算，该算法直接利用 DD 算法估计的先验信噪比求 $\left|\hat{X}_{m,k}\right|^2$。假定纯净语音和噪声均服从零均值的复高斯分布，$F_{m,k}$ 和 $D_{m,k}$ 分别代表纯净语音幅度谱和带噪语音幅度谱，$\Phi_{m,k}$ 和 $\Psi_{m,k}$ 分别为纯净语音分量相位和带噪语音分量的相位。通过最小化最小均方误差意义下的纯净语音短时谱能量及估计值之间的贝叶斯风险函数 $W=E\left(\left|X_{m,k}\right|^2-\left|\hat{X}_{m,k}\right|^2\right)^2$，可得到纯净语音幅度平方谱估计：

$$\hat{F}^2=\frac{\int_0^{\infty}\int_0^{2\pi}\frac{F^3}{\pi^2\lambda_X\lambda_N}\exp\left(-\frac{\left|Y-F\mathrm{e}^{j\Phi}\right|^2}{\lambda_N}\right)\exp\left(-\frac{F^2}{\lambda_X}\right)\mathrm{d}F\mathrm{d}\Phi}{\int_0^{\infty}\int_0^{2\pi}\frac{F}{\pi^2\lambda_X\lambda_N}\exp\left(-\frac{\left|Y-F\mathrm{e}^{j\Phi}\right|^2}{\lambda_N}\right)\exp\left(-\frac{F^2}{\lambda_X}\right)\mathrm{d}F\mathrm{d}\Phi} \tag{12}$$

为了简便，这里省略了 m，k。通过化简上式再带入 DD 算法估计出的先验信噪比，求出纯净语音信号幅度平方谱估计，最终得到该算法的先验信噪比估计表示如下

$$\hat{\xi}_{m,k}^{\text{M-TSNR}}=\frac{\hat{\xi}_{m,k}^{\text{DD}}}{\hat{\xi}_{m,k}^{\text{DD}}+1}+\frac{\left(\hat{\xi}_{m,k}^{\text{DD}}\left|Y_{m,k}\right|\right)^2}{\left(\hat{\xi}_{m,k}^{\text{DD}}+1\right)^2\lambda_N} \tag{13}$$

该算法计算相对简单，并且可以有效跟踪瞬时信噪比的变化，实现实时性，消除了残余噪声，改善了语音系统的性能。但是这种算法在很大程度上要依赖于 DD 算法的估计结果，其固定的平滑参数在不同应用环境和信噪比的情况下性能会受到限制。

由于 DD 算法受平滑参数牵制，平滑参数大小设置不当会引发音乐噪声及语音失真问题，平滑参数过大时，音乐噪声的抑制能力加强，但语音失真更严重，平滑参数过小效果则相反。为了进一步提升算法的性能，近年来有学者根据平滑参数取值对语音系统性能的影响程度选取了两个大小不同的平滑参数，并分别带入 DD 算法中得到两个先验信噪比估计结果，将结果相结合并融入一个耦合因子 δ，在无语音段耦合因子为 0，语音突变阶段取 1，两个平滑参数一个取大值一个取小值。提出融合耦合因子的先验信噪比估计算法[8]，定义如下

$$\hat{\xi}_{m,k}^{\text{CC}}=\delta\hat{\xi}_{m,k}^{1}+(1-\delta)\hat{\xi}_{m,k}^{2} \tag{14}$$

为计算自适应耦合因子 δ，在真实先验信噪比与估计的先验信噪比之间的最小均方误差准则下建立一个代价函数：

$$J=E\left\{\left(\hat{\xi}_{m,k}^{\text{CC}}-\xi_{m,k}\right)^2\right\} \tag{15}$$

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 $\xi_{m,k}$，得到该耦合因子：

$$\delta=\frac{(1-b)\left\{\max\left\{\eta_{m,k}-1,0\right\}+1\right\}^2-b\left\{\hat{\xi}_{m-1,k}-\max\left\{\eta_{m,k}-1,0\right\}\right\}^2}{(a-b)\left[\left\{\hat{\xi}_{m\text{-}1,k}-\max\left\{\eta_{m,k}-1,0\right\}\right\}^2+\left\{\max\left\{\eta_{m,k}-1,0\right\}+1\right\}^2\right]} \tag{16}$$

将耦合因子带入 CC 算法定义式(12)，可得到 CC 算法的先验信噪比，进一步求出维纳滤波增益因子，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。该算法的优势是可以自适应地结合两个具有不同平滑参数的 DD 算法，在无语音区自动地选取平滑参数较大的 DD 算法，而在语音存在区域则选取较小平滑参数的 DD 算法，其结果是即有效抑制了音乐噪声的产出，又避免了输出语音的失真。

4. 仿真实验结果分析

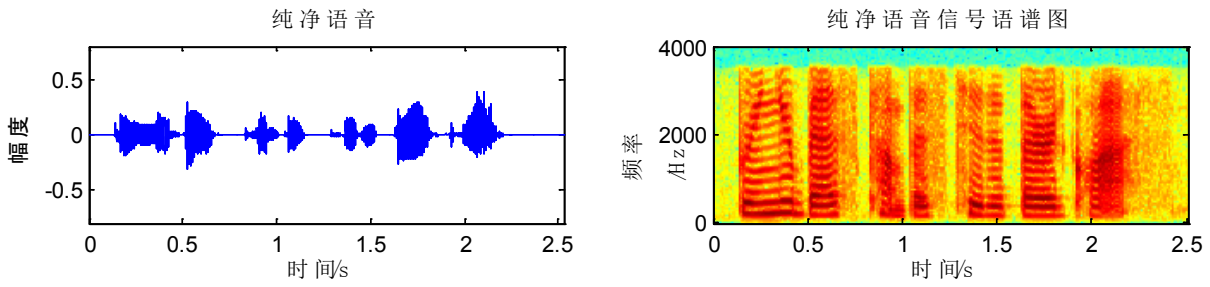
为了更好的对比几种算法的性能，采用 MATLAB 软件仿真对以上几种先验信噪比算法的输出结果进行了仿真验证，得到四种算法的时域波形图和语谱图，以及三种不同客观评价标准下的输出数值。通过仿真实验结果对比，得到四种算法的优劣顺序，验证了前面部分的理论分析。

首先是纯净语音信号，带噪语音信号和四种算法下增强的语音信号的时域波形图和语谱图，如图 1 所示。实验中选取 Noise x-92 数据库中的 Pink 噪声作为背景噪声，输入信噪比为 10 dB。纯净语音信号和背景噪声的采样频率均为 8 kHz，帧长为 256，采用汉明窗对时域信号进行分帧加窗处理，帧重叠为 50%。前三种算法的平滑参数均为 0.98，CC 算法的两个平滑参数分别为 0.992 和 0.6。

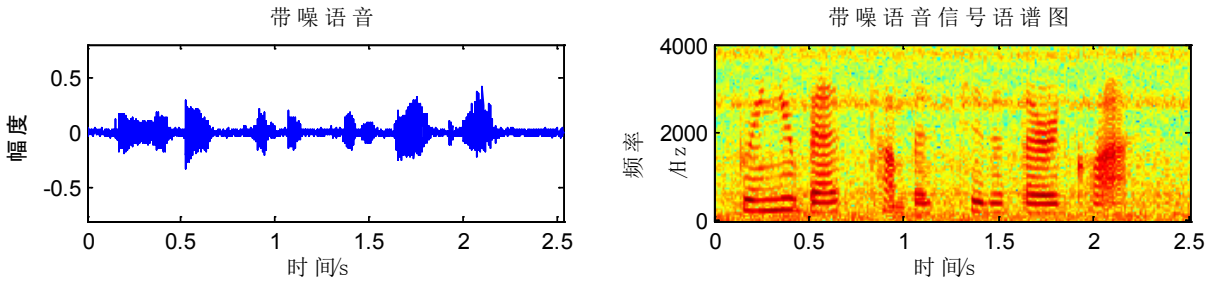
从以上时域波形图可看出，几种算法都能有效的消除背景噪声，但是也都在一定程度上对初始语音信号造成损伤。相较而言，MTSNR 算法和 CC 算法对原始语音的损伤程度更小，尤其是对于较小幅度的纯净语音信号而言损伤更小。从语谱图的结果中可看出，几种算法对背景噪声的消除和语音失真的改进效果有所不同。DD 算法增强后的语音由于帧延迟问题的存在依然残留较多的背景噪声，且语音失真较严重，而 TSNR 算法和 MTSNR 算法以及 CC 算法对 DD 算法进行了改进以后相比有效减少了背景噪声，且语音失真明显减少。相较而言，CC 算法的语谱图最接近于原始纯净语音信号，与前面理论分析的结果一致。

为了更加细致准确的对几种算法的性能进行定量分析，本文在不同背景噪声和不同信噪比环境下对几种算法进行了三种客观评价标准的测量。采用的客观评价标准有短时客观可懂度(STOI)，分段信噪比(SegSNR)和对数谱距离(LSD)。其中 STOI 是评价增强语音可懂度的指标，通过对比纯净语音分段信噪比是对每一帧的语音信号进行处理，通过将每一帧信号的信噪比求和取平均来评价语音增强的结果，其值越大说明算法的处理性能越好[12]。和带噪语音信号的短时时域包络的相关系数，来表示语音的真实可懂度。STOI 值越大，语音的可懂度越高，说明算法的性能越好。总帧数用 M 表示，帧长度和帧索引分别为 N 和 m，其定义公式如下：

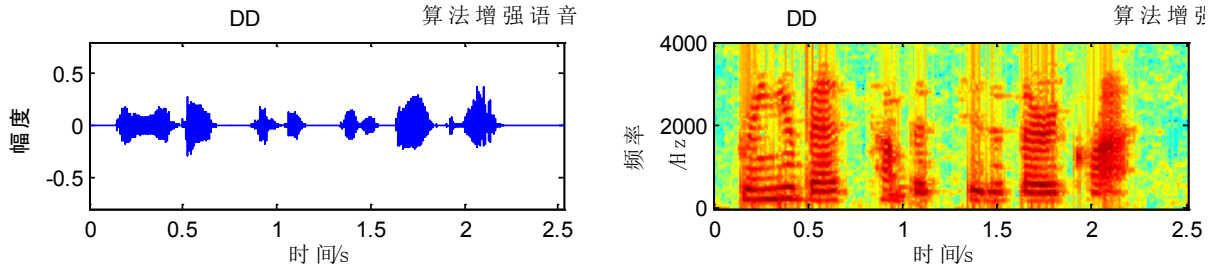
$$\mathrm{SegSNR}=\frac{1}{M}\sum_{m=0}^{M-1}\left\{10\log_{10}\frac{\sum_{n=0}^{N-1}s^2(n,m)}{\sum_{n=0}^{N-1}\left[s(n,m)-\hat{s}(n,m)\right]^2}\right\} \tag{17}$$



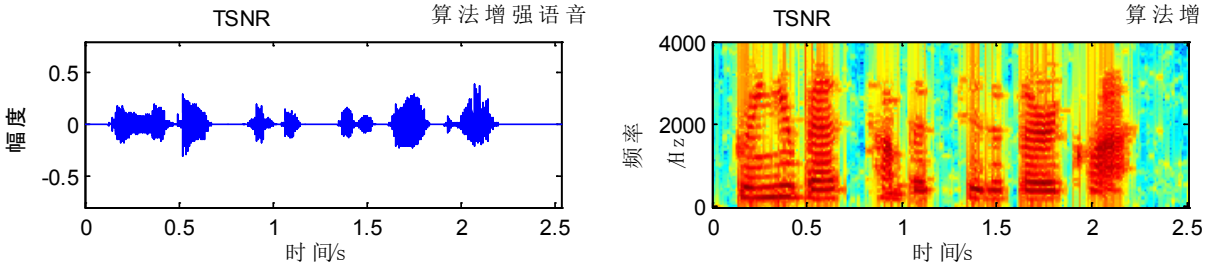
(a) 纯净语音信号时域波形图和语谱图



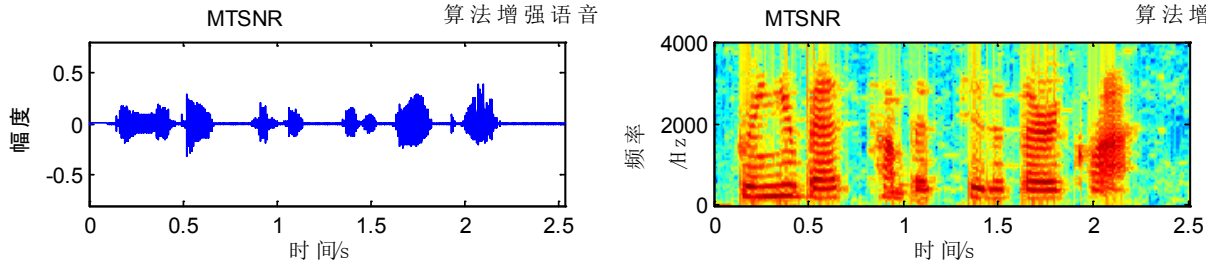
(b) 带噪语音信号时域波形图和语谱图



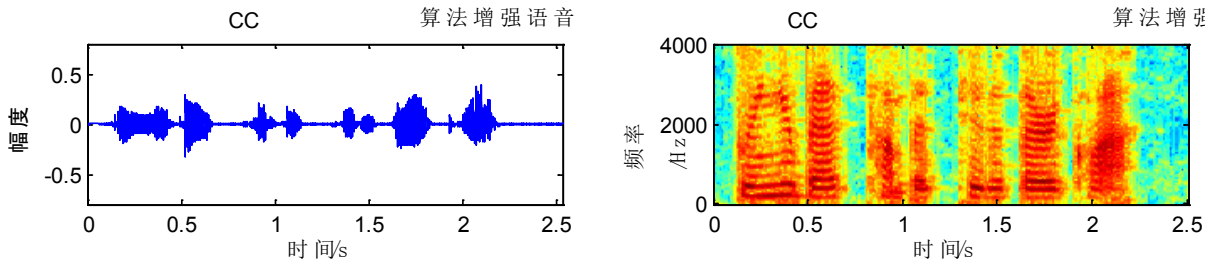
(c) 基于 DD 算法的增强语音信号时域波形图和语谱图



(d) 基于 TSNR 算法的增强语音信号时域波形图和语谱图



(e) 基于 MTSNR 算法的增强语音信号时域波形图和语谱图



(f) 基于 CC 算法的增强语音信号时域波形图和语谱图

Figure 1. The time domain figures and the spectrum of speech signal of different algorithms under pink noise (SNR = 10 dB)
图 1. 粉色噪声下不同算法的语音信号时域图和语谱图(SNR = 10 dB)

对数谱距离表示增强后的语音信号与纯净语音之间的接近程度，其值越小，说明增强后的语音越接近原始纯净语音，增强效果越好，对数谱距离定义如下[13]：

$$\mathrm{LSD}=\left\{\frac{1}{M}\sum_{m=0}^{M-1}\frac{1}{K}\sum_{k=0}^{K-1}\left[10\log_{10}\left(\frac{\left|\hat{S}_{m,k}\right|^2}{\left|S_{m,k}\right|^2}\right)\right]^2\right\}^{1/2} \tag{18}$$

其中，M 表示信号总帧数，m 为帧索引，K 和 k 分别表示语音帧长度和频点索引。

表 1~表 3 分别是对四种算法在六种背景噪声及三种输入信噪比水平下的短时客观可懂度和分段信噪比以及对数谱距离的取值情况。实验中选取 10 段纯净语音信号(5 段男声，5 段女声)作为测试数据，六种背景噪声均取自 Noise x-92 数据库，分别为 Pink 噪声，F16 噪声，Babble 噪声，white 噪声，M109 噪声和 Buccaneer2 噪声。在 5 dB，10 dB，15 dB 三种不同输入信噪比水平下进行实验仿真。表 1 是对四种算法在六种背景噪声及三种不同输入信噪比水平下的短时客观可懂度数据对比情况。在多种背景噪声环境和不同输入信噪比条件中，CC 算法的短时客观可懂度数值最高。STOI 是最符合人的听力特性的评价指标，其数值越大，表明语音信号增强的效果越理想。其他三种算法性能的优劣顺序依次为 MTSNR 算法，TSNR 算法，DD 算法。DD 算法在四种算法中增强效果最差，与时域波形图和语谱图分析结果一致。

Table 1. The STOI data comparison table of the four algorithms
表 1. 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	TSNR算法	MTSNR算法	CC算法
Pink	5 dB	76.6465	76.9760	78.7201	81.8246
	10 dB	82.7413	83.3708	84.7816	86.9183
	15 dB	87.7343	88.4643	89.4282	90.7700
F16	5 dB	78.0234	78.9158	80.2810	82.0189
	10 dB	84.0742	84.8826	85.7080	86.9714
	15 dB	88.3106	88.9572	89.6886	90.6344
Babble	5 dB	75.6156	76.1836	77.1963	79.1251
	10 dB	83.1912	83.8829	84.7243	85.9524
	15 dB	88.3324	88.7441	89.4067	90.4754
White	5 dB	76.9185	76.9480	78.5630	80.8341
	10 dB	82.5107	82.9600	84.5153	86.5193
	15 dB	87.6356	88.2637	89.4288	90.7903
M109	5 dB	81.4599	82.1766	83.2136	85.0790
	10 dB	86.4158	87.4440	88.4011	89.6771
	15 dB	90.8658	91.3544	92.1515	93.3784
Buccaneer2	5 dB	74.7512	74.9318	76.4619	79.2696
	10 dB	81.4295	81.8784	83.2401	85.1083
	15 dB	86.3910	86.9127	87.9595	89.4078

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

噪声类型	输入信噪比	输出SegSNR (dB)平均值			
		DD算法	TSNR算法	MTSNR算法	CC算法
Pink	5 dB	7.7595	7.8926	8.5593	8.8403
	10 dB	10.1003	10.2671	10.8837	11.2559
	15 dB	12.2246	12.4408	12.9993	13.4082
F16	5 dB	7.7804	7.9011	8.5319	8.8273
	10 dB	10.0120	10.2014	10.7824	11.1357
	15 dB	12.0817	12.3005	12.8168	13.2014
Babble	5 dB	6.3294	6.3447	6.9169	7.3786
	10 dB	8.7320	8.7570	9.3361	10.0693
	15 dB	11.0451	11.1620	11.7020	12.3966
White	5 dB	7.8036	7.9204	8.5914	8.8175
	10 dB	10.0043	10.1989	10.8263	11.2054
	15 dB	11.9357	12.1830	12.7365	13.1820
M109	5 dB	9.5445	9.7064	10.4104	10.5049
	10 dB	11.9818	12.1678	12.8649	13.0092
	15 dB	14.1505	14.3819	14.9797	15.1215
Buccaneer2	5 dB	7.5264	7.6248	8.2693	8.4364
	10 dB	9.7695	9.9275	10.5146	10.9007
	15 dB	11.6525	11.8350	12.3897	12.8825

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	TSNR算法	MTSNR算法	CC算法
Pink	5 dB	6.1382	5.6772	4.8901	4.6162
	10 dB	5.3530	5.1095	4.4512	4.1719
	15 dB	4.8693	4.7812	4.0778	3.8250
F16	5 dB	5.8571	5.3979	4.7203	4.4425
	10 dB	5.0454	4.8131	4.2468	3.9493
	15 dB	4.6380	4.5195	3.9139	3.6643
Babble	5 dB	5.9760	5.5552	4.7150	4.4304
	10 dB	4.8488	4.5689	4.0396	3.6463
	15 dB	4.1656	4.0181	3.5108	3.1951
White	5 dB	6.3569	6.0757	5.1953	4.9998
	10 dB	5.7116	5.4606	4.7743	4.5086
	15 dB	5.3542	5.3115	4.5084	4.3221
M109	5 dB	4.6216	4.2704	3.8028	3.4746
	10 dB	3.8918	3.6464	3.2585	2.9590
	15 dB	3.3437	3.1878	2.8373	2.5657
Buccaneer2	5 dB	6.3685	6.0362	5.1764	5.0710
	10 dB	5.6340	5.3730	4.7070	4.4706
	15 dB	5.1720	5.1147	4.3839	4.1650

表 2 表示四种算法在不同客观条件下的输出分段信噪比的数据对比表格。由表格中的数据可看出，对比下的四种先验信噪比估计算法中 DD 算法的输出 SegSNR 数值最小，均低于其他三种算法。分段信噪比是表征带噪语音信号抑制噪声性能优劣的重要标准，数值越大，表明算法对背景噪声抑制能力越强，增强效果越理想。在噪声抑制能力中，TSNR 算法改进了 DD 算法的缺陷，但是效果不大。而 MTSNR 算法和 CC 算法相较之下分段信噪比数值更高，能在很大程度上抑制背景噪声，CC 算法的抑制效果最为显著。

表 3 分别表示为 DD 算法，TSNR 算法，MTSNR 算法和 CC 算法这四种先验信噪比估计算法的输出 LSD 数据对比表。对表中数据分析可知：不同环境下 CC 算法的 LSD 数据均小于其他三种算法。对数谱距离数值越小，说明算法中增强后的语音越接近原始语音，即对原始语音的损伤程度越小。由该表可得，CC 算法增强后的语音失真程度最小，其次是 MTSNR 算法，TSNR 算法，DD 算法。

综合以上三个表格输出数据可看出，TSNR 算法在抑制语音失真性能方面有效改进了 DD 算法，但是未有效消除背景噪声。MTSNR 算法在满足了实时性的同时，有效抑制了背景噪声，但是由于受平滑参数的牵制，增强效果也没有达到理想水平。而 CC 算法在四种算法中性能最优，无论是语音失真还是音乐噪声方面，都达到了较理想的增强效果。

5. 结论

本文主要对比研究了几种单信道语音增强系统中先验信噪比的估计算法，首先说明了先验信噪比估计对语音增强系统性能的重要影响，然后介绍了 DD 算法、TSNR 算法、MTSNR 算法和 CC 算法在 DFT 域的基本理论，并给出了先验信噪比与增益因子的函数关系式，最后运用仿真实验得出时域图和语谱图以及两种客观评价标准数据对比分析了几种算法的性能，从实验上论证了理论的正确性，也进一步突显出先验信噪比估计对语音增强系统性能的重要性。近几年，深度神经网络算法在学术界应用较为普遍，已被顺利引入到语音增强领域中。与此同时，基于改进相位估计的语音增强算法也有很大的发展潜力，对于先验信噪比参数的估计有显著的作用。今后的研究中可以考虑将改进的相位估计算法和深度神经网络算法相融合，估计出准确度更高的先验信噪比，以增强语音系统的整体性能，从而对纯净语音的估计效果达到更优。

基金项目

烟台大学 2017 年研究生科技创新基金重点项目(YDZD1711)。

参考文献

[1] Schwerin, B. and Paliwal, K. (2014) Using STFT Real and Imaginary Parts of Modulation Signals for MMSE-Based Speech Enhancement. *Speech Communication*, **58**, 49-68. https://doi.org/10.1016/j.specom.2013.11.001

[2] Fang, Y., Liu, G. and Guo, J. (2011) Speech Enhancement Based on Modified a Priori SNR Estimation. *Frontiers of Electrical & Electronic Engineering in China*, **6**, 542-546. https://doi.org/10.1007/s11460-011-0181-8

[3] Xia, B. and Bao, C. (2014) Wiener Filtering Based Speech Enhancement with Weighted Denoising Auto-Encoder and Noise Classification. *Speech Communication*, **60**, 13-29. https://doi.org/10.1016/j.specom.2014.02.001

[4] 沈锁金, 魏静, 高颖. 基于新型先验信噪比估计的语音增强算法的对比研究[J]. 中国集成电路, 2016, 210(11): 41-45.

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

[6] 沈锁金, 欧世峰, 刘伟, 魏静. 基于先验信噪比估计语音增强算法的对比分析[J]. 烟台大学学报(自然科学与工程版), 2017, 30(2): 298-305.

[7] 欧世峰, 赵晓晖. 改进型先验信噪比估计语音增强算法[J]. 吉林大学学报: 工学版, 2009, 39(3): 787-791.

[8] Shen, S., Ou, S., Wei, J., *et al.* (2017) A Priori SNR Estimator Based on a Convex Combination of Two DD Approaches for Speech Enhancement. *IEEE International Conference on Signal and Image Processing*, 750-754.

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

[10] Plapous, C., Marro, C. and Scalart, P. (2006) Improved Signal-to-Noise Ratio Estimation for Speech Enhancement. *IEEE Transactions on Acoustics*, *Speech*, *and Signal Processing*, **14**, 2098-2108.

[11] 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.

[12] 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

[13] Abramson, A. and Cohen, I. (2010) Simultaneous Detection and Estimation Approach for Speech Enhancement. *IEEE Transactions on Audio Speech & Language Processing*, **15**, 2348-2359. https://doi.org/10.1109/TASL.2007.904231

Hans汉斯

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2327-0853，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：http://www.hanspub.org/Submission.aspx
期刊邮箱：ojcs@hanspub.org