

基于概率耦合的双直接判决先验信噪比估计算法

欧世峰¹, 赵艳磊¹, 宋 鹏², 高 颖¹

(1. 烟台大学光电信息科学技术学院, 山东烟台 264005; 2. 烟台大学计算机与控制工程学院, 山东烟台 264005)

摘 要: 直接判决(DD, Decision-Directed)算法结构简单、音乐噪声抑制能力较好,是当前语音增强领域最为常用的先验信噪比估计方法. 但该算法对于滑动因子的选取数值较为敏感,且估计性能要受到时延问题的限定. 本文首先采用实际的语音和噪声数据,根据音乐噪声残留及输出语音失真两方面的评测标准对 DD 算法中滑动因子的取值问题进行了研究,通过数据分析给出了其较为明确的上下边界值;然后基于语音及噪声信号的复高斯分布模型,采用软判决技术对两个具有不同滑动因子的 DD 算法进行概率耦合,提出了一种具有双 DD 结构的先验信噪比估计算法. 该算法可以充分结合两个具有不同特性 DD 算法的优点,在音乐噪声抑制及限制语音失真等方面均获得了较为理想的输出效果. 多种噪声背景及输入信噪比条件下的仿真结果表明,相对于目前流行的几种先验信噪比估计算法,本文提出算法具有更为优良的估计性能.

关键词: 语音增强; 直接判决算法; 滑动因子; 概率耦合

中图分类号: TN912.3

文献标识码: A

文章编号: 0372-2112 (2020)08-1605-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2020.08.020

Probabilistic Combination Framework of Two Decision-Directed Algorithms for a Priori SNR Estimation

OU Shi-feng¹, ZHAO Yan-lei¹, SONG Peng², GAO Ying¹

(1. School of Science and Technology for Opto-electronic Information, Yantai University, Yantai, Shandong 264005, China;

2. School of Computer and Control Engineering, Yantai University, Yantai, Shandong 264005, China)

Abstract: Due to the low computational complexity and acceptable ability in reducing musical noise effect, the decision-directed (DD) approach is widely used for estimating the a priori signal-noise-ratio (SNR) in many speech enhancement systems. However, the DD approach suffers from the problem of time delay and the performance is very sensitive to the fixed smoothing factor. Firstly, the performance of DD approach in musical noise reduction as well as speech distortion attenuation are analyzed using actual speech and noise data, and the boundary values of smoothing factors are presented in view of the analyzed results. Then, a novel algorithm is proposed, in which two DD approaches with different smoothing factors are probabilistically combined in an attempt to put together the best properties of them. The contribution of either DD approach to the combination is automatically adjusted in accordance with the speech absence probability, which can be computed using the complex Gaussian model and soft decision technique. Experiments are carried out in different noise and input SNR conditions, and the results demonstrate that the proposed algorithm can significantly outperform the popular methods for estimating the a priori SNR.

Key words: speech enhancement; decision-directed algorithm; smoothing factor; probability combination

1 引言

噪声干扰广泛存在于人类工作、生活中的各个环节,这些干扰的存在将严重破坏语音信号原有的声学模型和参数特性,进而大大降低各类语音信号处理设

备或系统的实际输出性能. 语音增强技术是有效抑制背景噪声、提高语音质量及可懂度的常用手段,目前正在被广泛应用于语音信号处理的各个领域. 常用的语音增强算法包括谱减法、最小均方误差算法(MMSE, Minimum Mean-Square Error)及维纳滤波算法等等^[1-3].

这些算法大都利用短时傅里叶变换将带噪语音变换至频率域,然后基于语音和噪声分量的声学模型并结合相应的估计理论来获取纯净语音分量的有效估计.在不同应用环境及条件要求下,各种语音增强算法对于纯净语音的估计效果也不尽相同,但几乎所有算法都存在一共同特点,即它们对纯净语音分量的估计均可以表示成某一增益因子与带噪语音分量的乘积形式.而先验信噪比是构成增益因子最为核心的参数,算法对其估计的精准程度将会对语音增强系统最终的输出效果产生决定性影响^[4],因此,如何有效提高先验信噪比参数的估计效果一直都是语音增强领域的研究热点.

目前,人们对于先验信噪比参数的估计已研究设计出许多成熟有效的算法,其中最为经典、常用的是 Ephraim 与 Malah 提出的直接判决 (DD, Decision Directed) 算法^[2].该算法基于系统对于前一帧先验信噪比及当前帧瞬时信噪比的估计结果,利用滑动因子对两者进行加权组合来获得当前帧先验信噪比的估计. DD 算法最大的特点是计算简单、可有效地抑制输出语音中的音乐噪声,但其估计结果在跟踪语音分量变化时会出现一个帧长的时间延迟,从而会带来一定的语音失真^[5].为了消除 DD 算法的时延问题,众多学者对其进行了深入研究,其中较为著名的是由 Plapous 等人提出的两步噪声消除 (TSNR, Two Step Noise Reduction) 算法^[6].该算法采用“两步走”的策略,通过对 DD 方法输出先验信噪比估计结果的进一步更新和优化,有效地避免了传统 DD 算法产生的时延效应,且同时具备一定的音乐噪声抑制能力.目前,TSNR 算法已被 ITU-T 写入 G. 711. 1 语音编码标准.后来, Yong 等人提出了一种改进的 DD (MDD, Modified DD) 算法^[7],其利用当前帧带噪语音分量来匹配前一帧的增益因子,进而对前一帧的先验信噪比估计结果进行修正,然后再采用 DD 算法获得当前帧的先验信噪比估计. MDD 算法可以提升系统对于语音信号变化的跟踪效果,有效避免了语音的失真,但由于其在估计过程中重复利用了当前帧的带噪语音分量,往往导致估计的先验信噪比对于带噪信号的变化过于敏感,进而产生一定的音乐噪声.在传统 DD 算法中通过引入语音信号的相位信息, Lee 分别利用 MMSE 及最大后验 (MAP, Maximum A Posteriori) 估计理论提出了两种先验信噪比估计算法^[8],有效提升了 DD 算法的整体性能,但该算法在估计过程中将会带来较大的计算量,影响系统的实时运算效率.

另一方面,人们对于 DD 算法的改进还体现在滑动因子的自适应选取方面.较早从事此方面研究的学者 Soon 和 Koh 基于背景噪声为平稳噪声且其能量在帧间变化比较缓慢的假设条件下,提出了一种滑动因子的自适应优选规则,获得了失真度较低的语音增强算

法^[9].但该算法受制于假设条件的限定,无法适用于非平稳噪声环境. Hasan 等人基于 MMSE 估计,结合后验信噪比设计了一种自适应滑动因子 (VSF, Variable Smoothing Factor) 算法^[10],可以有效跟踪语音信号的能量变化,避免 DD 算法固有的时延问题,但该算法在无语音阶段对于先验信噪比估计的波动较大,容易引入较多的音乐噪声. Park 和 Chang 利用后验信噪比的瞬时变化现象,通过设计 Sigmoid 函数获取了一种滑动因子的在线调整方法^[11].该方法可以减少时延效应,并能有效提升输出语音信号的信噪比,但算法在设计过程中引入了四个不同的经验参数,不易调整和设置,同时对于音乐噪声抑制效果也不够理想.基于相同的理念, Yong 等人在 MDD 算法的基础上,通过设计不同的 Sigmoid 函数来对滑动因子进行自适应更新,获得了更为有效的先验信噪比估计结果^[12].但该算法仍需要设置三个经验参数来平衡音乐噪声与语音失真的折中问题.另外,上述所有的自适应算法都没有对滑动因子的设定范围进行深入研究,其多是简单地将其设定于 $[0, 1]$ 区间内,因此,在对先验信噪比估计的过程中往往出现滑动因子过大或过小的情况,进而不可避免地会引入一定的音乐噪声或语音失真.

此外,结合数据驱动、噪声分类、多元线性回归、倒谱分析等技术,人们相继提出了多种新型的先验信噪比估计算法^[13-17].如: Suhadi 等采用两个人工神经网络对先验信噪比进行分阶段训练,设计了一种基于数据驱动的先验信噪比估计算法; Choi 与 Chang 利用高斯混合模型对背景噪声进行辨识,并以此为依据设计最优参数来获取先验信噪比的有效估计;同时, Chang 还在噪声分类的基础上采用多元线性回归技术对先验信噪比参数的估计问题进行了研究; Elshamy 则采用倒谱分析技术提出了一种瞬时先验信噪比估计方法.这些算法由于融合了神经网络等较为复杂的处理过程,其对先验信噪比参数估计的准确性有了一定提升,但它们也都存在各自的应用局限性,如需要大量数据进行训练、运算量较大等.

本文将重点围绕最为常用的 DD 先验信噪比估计算法开展研究,首先根据音乐噪声残留及输出语音失真两个方面的测评结果对滑动因子的取值问题进行了分析,给出了其较为明确的上下边界值;然后基于语音及噪声的统计模型,采用软判决技术对两个具有不同滑动因子的 DD 算法进行概率耦合,获取了一种双 DD 先验信噪比估计算法.该算法可以充分结合两个不同滑动因子 DD 算法的优点,在语音静默阶段具备大滑动因子较为平滑的估计特性,而在语音活动阶段又具有小滑动因子较为快速的瞬时信噪比跟踪速度,从而有效地抑制了音乐噪声的生成、减小了语音增强系统对

于纯净语音信号的损伤. 在多种噪声背景及输入信噪比条件下, 本文采用三种评价准则对提出算法的估计性能进行了验证, 其结果有效表明了提出算法对带噪语音信号在语音质量与可懂度等方面的改善效果.

2 频率域语音增强

分别用 $y(n)$ 、 $x(n)$ 和 $v(n)$ 表示时域中的带噪语音、纯净语音及加性噪声信号, 并假设纯净语音与噪声信号互不相关, 有:

$$y(n) = x(n) + v(n), \quad E\{x(i)v(j)\} = 0 \quad \forall i, j \quad (1)$$

对式(1)进行 N 点短时傅里叶变换, 可得:

$$Y_{m,k} = X_{m,k} + V_{m,k}, \quad k = 0, \dots, N-1 \quad (2)$$

这里, k 和 m 分别表示频点及帧索引 ($m = 0, 1, \dots$); $Y_{m,k}$ 、 $X_{m,k}$ 与 $V_{m,k}$ 则分别表示信号 $y(n)$ 、 $x(n)$ 和 $v(n)$ 在频率域中的第 m 帧、第 k 个谱分量. 定义 $X_{m,k}$ 与其估计值 $\hat{X}_{m,k}$ 之间的失真函数为 $d(X_{m,k}, \hat{X}_{m,k})$, 则根据贝叶斯估计理论, 可以获得纯净语音分量 $X_{m,k}$ 在失真函数 $d(X_{m,k}, \hat{X}_{m,k})$ 下的最优估计为^[18]:

$$\hat{X}_{m,k} = \arg \min_{X_{m,k}} E\{d(X_{m,k}, \hat{X}_{m,k}) | Y_{m,k}\} \quad (3)$$

根据所选取的失真函数的不同, 通过推导上式可获得多种形式的估计算法, 如短时谱估计算法、维纳滤波算法、对数谱估计算法等^[19]. 这些算法虽然在抑制背景噪声、减少语音损伤等方面各有优劣, 应用环境也不尽相同, 但它们都存在一共同特点, 即对纯净语音分量的估计均可以表示成某一增益因子与带噪语音分量的乘积形式^[18]:

$$\hat{X}_{m,k} = G(m, k) \cdot Y_{m,k} \quad (4)$$

其中, $G(m, k)$ 表示增益因子, 为先验信噪比 $\xi(m, k)$ 与后验信噪比 $\gamma(m, k)$ 的函数:

$$G(m, k) = F\{\xi(m, k), \gamma(m, k)\} \quad (5)$$

其中,

$$\xi(m, k) = \frac{E\{|X_{m,k}|^2\}}{\lambda_v(m, k)}, \quad \gamma(m, k) = \frac{|Y_{m,k}|^2}{\lambda_v(m, k)} \quad (6)$$

这里, $\lambda_v(m, k) = E(|V_{m,k}|^2)$ 表示噪声分量方差, 其可采用语音激活检测或最小统计 (MS, Minimum Statistics) 算法进行估计或更新; 而后验信噪比 $\gamma(m, k)$ 则可以利用 $Y_{m,k}$ 结合噪声方差 $\lambda_v(m, k)$ 直接计算获得. 结合以上三式不难看出, 先验信噪比是构成语音增强系统中增益因子的核心参数, 不管是选取何种增益因子的语音增强系统, 算法对先验信噪比参数的估计精度将会在很大程度上决定这些语音增强系统的实际输出效果.

在各类语音增强系统中, 目前最为常用的先验信噪比估计算法是 Ephraim 与 Malah 提出的 DD 算法, 其对当前帧先验信噪比参数的估计形式如下^[2]:

$$\hat{\xi}_{DD}(m, k) = \beta \frac{|\hat{X}_{m-1,k}|^2}{\lambda_v(m-1, k)} + (1-\beta) \max\{\gamma(m, k) - 1, 0\} \quad (7)$$

其中, $\hat{X}_{m-1,k}$ 为前一帧纯净语音谱分量的估计, $\beta \in [0, 1]$ 为滑动因子, 为获得较为理想的输出效果, 其值一般设置为 0.98. DD 算法的特点是易于实现、可有效抑制输出语音中的音乐噪声, 但其估计结果在跟踪瞬时信噪比变化时多会产生一帧的时延, 从而不可避免地带来部分语音失真. 为消除 DD 算法的时延效应, Plapous 等人基于“两步走”的策略提出了 TSNR 算法^[6]. 该算法的估计过程可简单分为以下两步: 第一步, 利用式(7)中 DD 方法计算 $\hat{\xi}_{DD}(m, k)$; 第二步, 结合 $\hat{\xi}_{DD}(m, k)$ 与增益因子联合获得当前帧先验信噪比的估计 $\hat{\xi}_{TSNR}(m, k)$.

$$G_{DD}(m, k) = \frac{\hat{\xi}_{DD}(m, k)}{\hat{\xi}_{DD}(m, k) + 1}, \quad (8)$$

$$\hat{\xi}_{TSNR}(m, k) = \frac{|G_{DD}(m, k) Y_{m,k}|^2}{\lambda_v(m, k)}$$

其中 $G_{DD}(m, k)$ 为系统的维纳滤波增益因子. TSNR 算法实际上是将两个连续 DD 方法相互串联的结果, 其在第二个 DD 算法中结合当前帧带噪语音分量对第一个 DD 算法的输出结果进行优化, 从而有效提升了算法对于语音信号变化的跟踪速度. 近来, Yong 等人提出了一种改进的 DD(MDD)算法^[7], 其通过当前帧带噪语音分量与前一帧增益因子的匹配, 来修正前一帧先验信噪比的估计结果, 再结合常规 DD 算法获得当前帧的先验信噪比估计, MDD 算法的计算过程如下:

$$\hat{\xi}_{MDD}(m, k) = \beta \frac{|G(m-1, k) \cdot Y_{m,k}|^2}{\lambda_v(m-1, k)} + (1-\beta) \max\{\gamma(m, k) - 1, 0\} \quad (9)$$

其中, $G(m-1, k)$ 表示第 $m-1$ 帧的增益因子. MDD 算法在估计过程中两次利用了当前帧带噪语音分量的相关信息, 从而可以提升算法对于语音变化的跟踪效果. 另一种针对传统 DD 算法的改进方案是设计一定的规则对滑动因子进行在线调整, 较有代表性的方法是 Hasan 等人提出的 VSF 算法^[10], 该算法仍以式(7)的 DD 算法为基础, 依据 MMSE 估计准则设计滑动因子的自适应实时计算规则如下:

$$\beta(m, k) = \frac{1}{1 + \left(\frac{\xi(m, k) - \hat{\xi}(m-1, k)}{\hat{\xi}(m, k)} \right)^2} \quad (10)$$

其中, $\hat{\xi}(m-1, k)$ 表示算法在第 $m-1$ 帧的先验信噪比估计. 由于当前帧先验信噪比 $\xi(m, k)$ 未知, 在对 $\beta(m, k)$ 的实际计算过程中采用 $\xi(m, k) \approx \max\{\gamma(m, k) - 1, 0\}$ 对其进行近似. 由于式(10)算法在最小均方误差意义下实现了对于滑动因子的自适应调整, 其在跟踪瞬时信噪比时可以提供较为快速的反应速度. 以上几种改进方案有效避免了 DD 算法固有的时延问题, 减少了语音失真, 但也存在其各自的缺点, 如 TSNR 算法的性能对于系统所选增益因子的形式具有一定依赖性, 而

MDD 和 VSF 算法则对于音乐噪声的抑制方面不够理想等。

3 双 DD 耦合先验信噪比估计算法

上文对频率域语音增强的基本过程进行了简单概述,重点围绕先验信噪比参数的估计问题,给出了经典 DD 算法以及几种较具代表性的改进算法。这些改进算法可以不同程度地改善 DD 算法的时延问题,提供更小的语音失真输出效果,但它们对于音乐噪声的残留问题多没有进行深入分析,个别改进算法在避免 DD 算法时延问题的同时,往往可能会产生较多的音乐噪声。而评价一个先验信噪比估计算法的优劣需要从音乐噪声残留、语音失真大小两方面来综合衡量,单纯地考虑某一方面而忽略另一方面都是不合适的。本节将围绕 DD 算法中滑动因子的优选问题进行探讨,从音乐噪声残留、语音失真两方面来分析滑动因子的取值情况对于系统整体性能的影响,进而获取 DD 算法滑动因子的边界数值,并以此为基础设计具有双 DD 结构的先验信噪比估计算法,以期在多种背景噪声环境下获得更为有效的语音增强效果。

重写 DD 算法的估计过程如下^[20]:

$$\hat{\xi}_{DD}(m, k) = \beta \hat{\xi}_{DD}(m-1, k) + (1-\beta) \hat{\xi}_{ML}(m, k) \quad (11)$$

其中,

$$\hat{\xi}_{DD}(m-1, k) = \frac{|\hat{X}(m-1, k)|^2}{\lambda_v(m-1, k)}, \quad (12)$$

$$\hat{\xi}_{ML}(m, k) = \max\{\gamma(m, k) - 1, 0\}$$

它们分别表示 DD 算法在第 $m-1$ 帧的先验信噪比估计以及当前帧瞬时信噪比的最大似然 (ML, Maximum Likelihood) 估计。结合以上两式可以看出, DD 算法对于先验信噪比的估计主要包含两大部分,即前一帧的先验信噪比估计以及当前帧的瞬时信噪比估计,算法通过滑动因子来分配此两部分在整个估计过程的比例。数值较大的滑动因子 (β 接近于 1) 表示算法更加侧重于以前时刻的估计结果,利用当前帧瞬时信噪比的信息较少,因而先验信噪比的估计曲线较为平滑,在无语音阶段不易产生数值较大的奇异点,非常利于音乐噪声的抑制。但由此带来的负面影响也较为明显,即算法在语音存在阶段无法快速跟踪语音信号的能量变化,导致系统对纯净语音信号造成损伤,形成语音失真。而另一方面,如选用数值较小的滑动因子,算法则更加强当前时刻瞬时信噪比的更新,估计结果能够根据语音信号的变化快速做出反应,但无语音区的估计结果波动较大,容易产生较多的音乐噪声^[5]。为更清晰地说明此问题,图 1 在白噪声背景下 (输入信噪比为 10dB), 给出了不同滑动因子 DD 算法在频率 625Hz 处

对于先验信噪比的估计结果 (为不失一般性,本文以下仿真中均采用维纳滤波估计作为语音增强系统的增益因子,先验信噪比的最小值设为 $\xi_{\min} = -30\text{dB}$)。图中 1 至 15 帧为无语音区,此阶段内带噪语音信号中只包含背景噪声;15 帧以后语音信号开始出现,直至第 30 帧皆为语音活动区域。从图 1 中不难看出,在无语音区,滑动因子的数值越大,其对先验信噪比参数的估计曲线就越平滑,较少出现估计数值较大的奇异值,因而不会在输出语音中造成较多音乐噪声残留;但在语音出现阶段,数值较大滑动因子的估计结果在跟踪瞬时信噪比时则存在着明显的时间延迟问题,且滑动因子数值越大,时延效应越突显,从而导致语音失真现象也就越严重。反之,选择数值较小的滑动因子则带来恰恰相反的估计结果。因此,对于采用固定滑动因子的 DD 算法而言,无论滑动因子如何选值算法都无法同时避免语音失真现象的产生以及音乐噪声的生成。

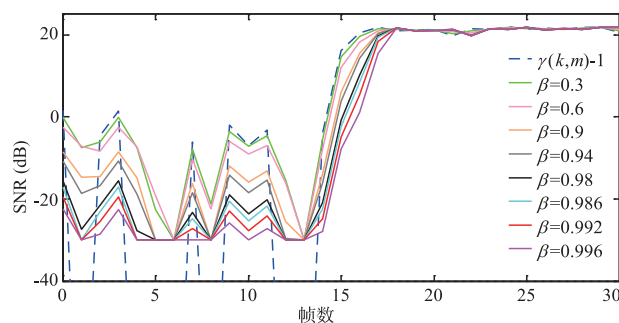


图1 DD算法先验信噪比估计曲线图

解决上述问题一个比较理想的方案是在无语音区选择数值较大的滑动因子,而在语音活动阶段则自动选取较小的滑动因子,从而使得算法在消除语音失真现象的同时又可以对音乐噪声进行有效抑制。但这种解决方案势必会带来两个问题:(1)在语音静默及活动阶段如何确定滑动因子的具体取值,如按照 β 的取值范围 $[0, 1]$ 而简单地将其设定为一个接近于 0 或 1 的数值是否与实际情况相符合?(2)如何将语音静默及活动阶段获得的先验信噪比估计结果进行有效融合,以综合其各自优势取得更为理想的输出效果?本文 3.1 节与 3.2 节将主要针对上述两个问题进行深入研究,通过实验分析和算法设计,以期获得更为优良的先验信噪比估计问题解决途径。

3.1 滑动因子取值范围

首先,在滑动因子取值对于音乐噪声的影响方面,考虑到音乐噪声的生成主要来源于无语音阶段算法对于先验信噪比参数的过估计,故将背景噪声直接作为输入信号送至语音增强系统的输入端,通过比较不同滑动因子 DD 算法下系统输出信号谱分量幅度的分布,即可较为清晰地分析其各自音乐噪声的残留情况。图 2

给出的是 White、Volvo、Pink 及 Babble 四种噪声背景下语音增强系统输出信号分量归一化幅度的对数分布直方图(LH, Log-Histogram), 滑动因子 β 的取值分别设为 $\beta = 0.3, 0.6, 0.9, 0.94, 0.98, 0.986, 0.992, 0.996$. 综合图中结果可以看出, 当滑动因子选取的数值越大时, 输出信号中数值较大的奇异点越少, 信号的分布越接近于其原始的噪声信号分布, 从而说明算法对于先验信噪比参数的过估计较少, 音乐噪声不明显; 而当选择的

滑动因子数值较小时, 输出信号中则会产生较多的奇异点, 其分布往往拖着一个长长的“尾巴”, 算法存在多处对于先验信噪比参数的过估计, 易生成较多的音乐噪声. 同时, 从图中结果也不难看出, 当滑动因子数值增大至 0.992 及以上时, 四种噪声背景下系统输出信号谱分量的分布曲线与其原始的噪声信号分布已基本重合, 进而说明在此取值阶段滑动因子的数值变化对于音乐噪声生成情况的影响已非常不显著.

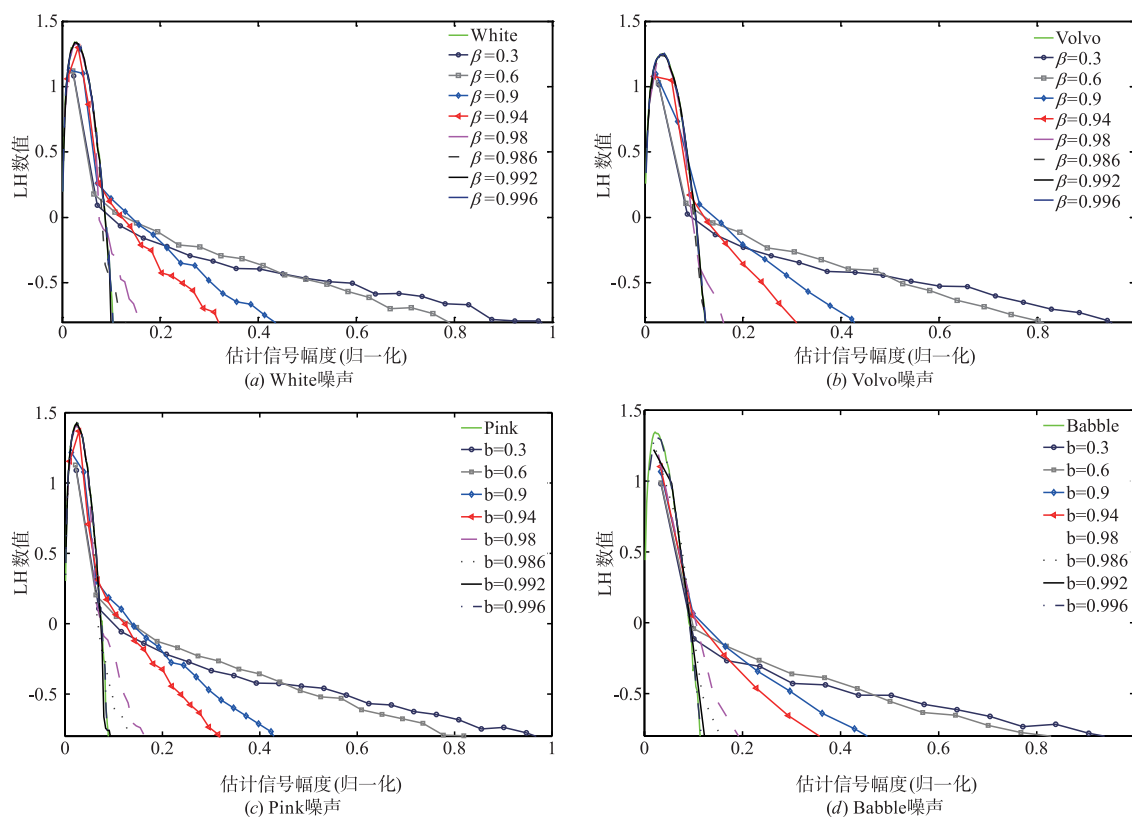


图2 不同噪声背景下输出信号对数分布直方图

另一方面, 语音增强系统在对带噪语音的处理过程中, 其在消除背景噪声的同时不可避免地会对纯净语音信号造成损伤, 从而在输出端形成语音失真. 由于板仓-斋藤 (IS, Itakura-Saito) 距离具有尺度不变性特点, 且其输出结果与接受者的主观感受相符合, 故在测量语音增强系统输出信号失真程度的过程中, 本节将 IS 距离作为主要测度^[21], 其值越小, 说明输出信号中的语音失真越少, 越接近于原始的纯净语音. 实验中背景噪声选择 White、Volvo、Pink 及 Babble 四种不同类型的噪声信号, 所用语音数据为 30 段纯净的语音信号, 将消除静音区后的纯净语音混入不同的噪声信号后, 分别产生输入信噪比为 0、5、10 和 15 dB 的带噪语音. 图 3 给出的是不同噪声背景及输入信噪比条件下, 语音增强系统输出语音与纯净语音之间的 IS 距离 (30 段纯净语音 IS 计算结果的平均值), 其中滑动因子的取值为 $\beta =$

0.1, 0.2, ..., 0.999. 从图 3 中可以看出, 平滑因子的取值对于语音失真的大小具有一定影响, β 取值越大, 语音失真情况越为严重, 这与图 1 中所获结论一致. 但从图 3 中同时不难看出, 滑动因子对于语音失真的影响大致可分为两个阶段, 即当滑动因子较大时 ($\beta > 0.7$), 其取值情况对于输出语音 IS 距离的影响较大, 语音失真程度随着滑动因子的变大迅速增加; 而当 β 较小时 ($\beta \leq 0.6$), 输出语音的 IS 距离数值基本呈现为水平状态, 滑动因子的数值选择对于输出语音的 IS 距离几乎没有影响. 也就是说, 在平滑因子取值范围为 $\beta \leq 0.6$ 的区间内, DD 算法输出语音中的语音失真情况变化较小, 基本上不受平滑因子取值的影响. 需要特别说明的是, 在其他噪声如 Factory、M109、Destroyer 等背景及不同输入信噪比下所得结果与上述分析相一致, 其 IS 距离计算结果及对数分布直方图本文将不再一一进行描

述和讨论.

综合以上分析结果不难看出:当采用 DD 算法进行先验信噪比估计时,小的滑动因子可以有效减小语音失真,但由于其数值设置过小时系统同时会生成更多的音乐噪声,因此在实际应用时,无需将滑动因子设置

在非常接近于 0 的较小区域内, $\beta = 0.6$ 是较为合适的一个选择;而在抑制音乐噪声生成方面,也不必把滑动因子设置为一个无限接近于 1 的常数,当 β 取值为 0.992 时,语音增强系统输出语音中已基本不再包含音乐噪声.

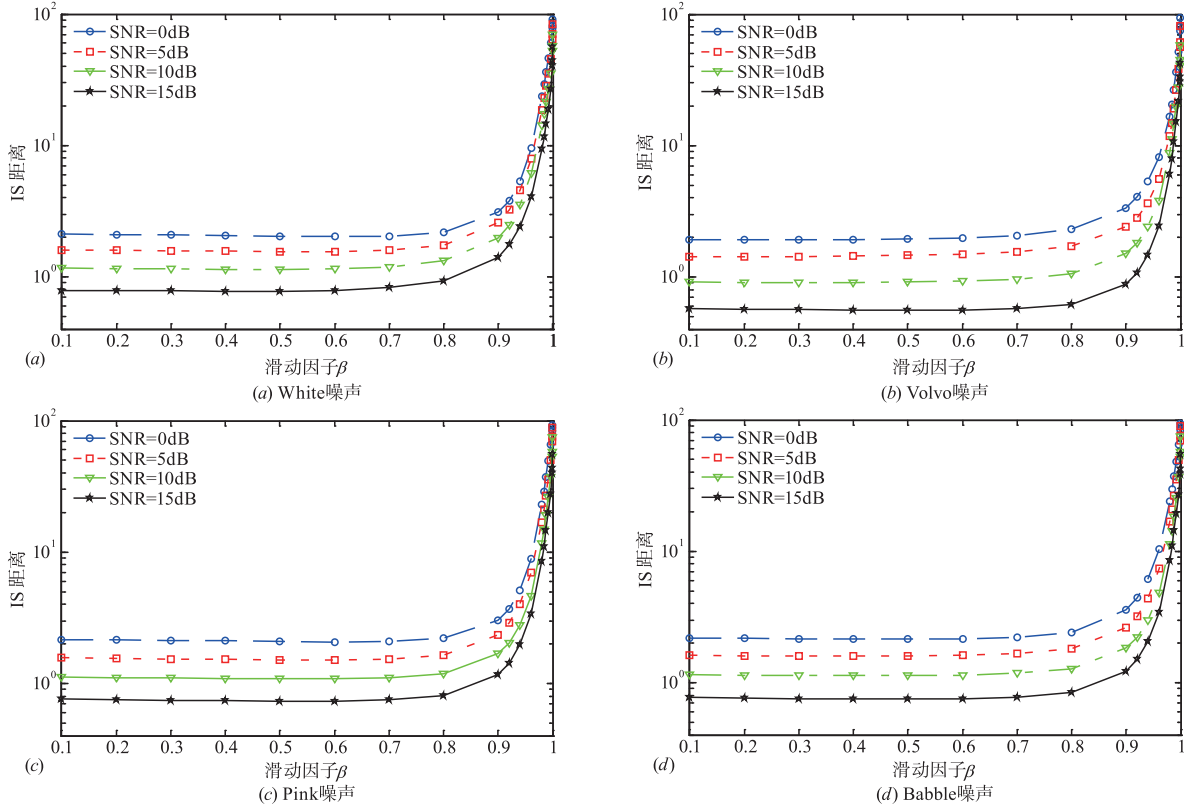


图3 不同噪声背景下输出语音的IS距离

3.2 耦合算法设计

为使先验信噪比估计算法在无语音区域具有较好音乐噪声抑制能力,同时又能在语音存在阶段尽可能地减少对纯净语音的损伤,本文考虑结合两个不同滑动因子 DD 算法的优势,并基于软判决技术对其估计结果进行自适应耦合,以进一步提高算法对于先验信噪比参数的估计性能,获得更为理想的语音增强效果.

基于式(11)和(12),设计两个具有不同滑动因子的 DD 算法分别为:

$$\hat{\xi}_1(m, k) = \beta_1 \cdot \hat{\xi}_1(m-1, k) + (1 - \beta_1) \cdot \hat{\xi}_{ML}(m, k) \quad (13)$$

$$\hat{\xi}_2(m, k) = \beta_2 \cdot \hat{\xi}_2(m-1, k) + (1 - \beta_2) \cdot \hat{\xi}_{ML}(m, k) \quad (14)$$

其中, β_1 和 β_2 分别表示两 DD 算法各自的滑动因子,且要求 $\beta_1 > \beta_2$,即第一个 DD 算法在无语音阶段的音乐噪声抑制能力要优于第二个算法,而在语音活动阶段,第二个 DD 算法对于纯净语音信号的损伤程度则更小.

对于当前时刻的带噪语音分量 $Y_{m,k}$,其存在两个基

本状态,即^[22]:

$$H_0: Y_{m,k} = V_{m,k} \quad (15)$$

$$H_1: Y_{m,k} = X_{m,k} + V_{m,k} \quad (16)$$

其中, H_0 表示语音不存在状态,当前时刻带噪语音中只有背景噪声; H_1 则表示语音存在状态,当前时刻 $Y_{m,k}$ 中包含纯净语音与背景噪声.

用 $p(H_0 | Y_{m,k})$ 表示已知带噪语音分量 $Y_{m,k}$ 条件下语音不存在的软判决概率,则考虑将式(13)与式(14)的估计结果进行概率耦合,可得:

$$\hat{\xi}_{PC}(m, k) = p(H_0 | Y_{m,k}) \hat{\xi}_1(m, k) + p(H_1 | Y_{m,k}) \hat{\xi}_2(m, k) \quad (17)$$

其中, $\hat{\xi}_{PC}(m, k)$ 为本文将两个 DD 算法进行组合后获得的先验信噪比估计结果, $p(H_1 | Y_{m,k})$ 则表示当前时刻语音信号的存在概率,显然有:

$$p(H_1 | Y_{m,k}) = 1 - p(H_0 | Y_{m,k}) \quad (18)$$

为计算 $Y_{m,k}$ 条件下语音不存在的软判决概率 $p(H_0 | Y_{m,k})$,假设频率域中纯净语音及噪声信号的谱分量均服从零均值的复高斯分布,有:

$$p(X_{m,k}) = \frac{1}{\pi\lambda_x(m,k)} \times \exp\left(-\frac{|X_{m,k}|^2}{\lambda_x(m,k)}\right) \quad (19)$$

$$p(V_{m,k}) = \frac{1}{\pi\lambda_v(m,k)} \times \exp\left(-\frac{|V_{m,k}|^2}{\lambda_v(m,k)}\right) \quad (20)$$

由于纯净语音与噪声信号互不相关,有:

$$\lambda_y(m,k) = \lambda_x(m,k) + \lambda_v(m,k) \quad (21)$$

其中, $\lambda_y(m,k) = E(|Y_{m,k}|^2)$ 和 $\lambda_x(m,k) = E(|X_{m,k}|^2)$ 分别表示带噪语音与纯净语音分量的方差. 结合以上三式易得 H_0 和 H_1 两种状态下带噪语音分量的条件概率分别为^[23]:

$$p(Y_{m,k} | H_0) = \frac{1}{\pi\lambda_v(m,k)} \times \exp\left(-\frac{|Y_{m,k}|^2}{\lambda_v(m,k)}\right) \quad (22)$$

$$p(Y_{m,k} | H_1) = \frac{1}{\pi(\lambda_v(m,k) + \lambda_x(m,k))} \times \exp\left(-\frac{|Y_{m,k}|^2}{\lambda_v(m,k) + \lambda_x(m,k)}\right) \quad (23)$$

根据贝叶斯公式,有:

$$p(H_0 | Y_{m,k}) = \frac{p(Y_{m,k} | H_0)p(H_0)}{p(Y_{m,k} | H_0)p(H_0) + p(Y_{m,k} | H_1)p(H_1)} \quad (24)$$

其中, $p(H_0)$ 与 $p(H_1)$ 分别表示语音不存在及语音存在的先验概率. 对上式进行进一步优化,可得:

$$p(H_0 | Y_{m,k}) = \frac{1}{1 + \rho\Lambda(m,k)} \quad (25)$$

式中 $\rho = p(H_1)/p(H_0)$, $\Lambda(m,k)$ 为似然函数,其计算公式为:

$$\begin{aligned} \Lambda(m,k) &= \frac{p(Y_{m,k} | H_1)}{p(Y_{m,k} | H_0)} \\ &= \frac{1}{1 + \xi(m,k)} \times \exp\left(\frac{\gamma(m,k)\xi(m,k)}{1 + \xi(m,k)}\right) \end{aligned} \quad (26)$$

综合式(13)、(14)以及式(17)、(25)即可获得本文提出算法的先验信噪比估计结果,其中 β_1 和 β_2 依据 3.1 节分析结果分别选取为 0.992 和 0.6. 由于式(26)中 $\xi(m,k)$ 未知,实际运算过程中采用算法前一帧的估计结果 $\hat{\xi}_{PC}(m-1,k)$ 对其进行替代,同时假设 H_0 和 H_1 的先验概率相同,即 $\rho = 1$. 为验证本文提出算法(PC, Probabilistic Combination)的先验信噪比估计性能,图 4 给出了白噪声背景下多个算法的先验信噪比估计曲线,其仿真环境与图 1 相同. 通过图 4 仿真结果不难看出,在 1~15 帧的无语音阶段,本文 PC 算法与 DD 算法($\beta = 0.992$)的估计结果基本一致,均在 -25dB 左右浮动,估计曲线较为平滑、波动较小,因此系统可在消除背景噪声的同时有效地抑制音乐噪声的生成. 而在 15 帧以后的语音存在阶段,本文 PC 算法则可以快速地根

据瞬时信噪比的变化做出反应,其估计曲线与 DD 算法($\beta = 0.6$)的波动情况相一致,基本上不存在时延效应,从而有效避免了系统对于纯净语音信号的损伤. 以上结果表明,本文采用软判决技术提出的双 DD 算法耦合方案可以有效结合两个不同滑动因子 DD 算法的优点,在无语音阶段可以拥有大滑动因子较为稳定的平滑特性,而在语音出现阶段又具有小滑动因子快速的跟踪速度,从而有效地消除了传统 DD 算法中音乐噪声和语音失真间的矛盾关系,确保算法在具有较小语音损伤的同时,又能抑制音乐噪声的生成. 作为参考,图 4 同时给出了传统 DD 算法($\beta = 0.98$ 时最优)、VSF 算法、MDD 算法以及 TSNR 算法的先验信噪比估计结果. 通过对比可以发现,几种参考算法在语音出现阶段的估计结果相对于传统 DD 算法均有不同程度的改善,但改进效果均不如 PC 算法;而在无语音阶段,参考算法的估计曲线相对于 PC 算法均存在一定程度的波动(由于采用了两步优化,TSNR 算法在此阶段的估计数值较小,故在图中没有对其最小值进行限定),说明在对音乐噪声的抑制方面,PC 算法的实际应用效果更为出色.

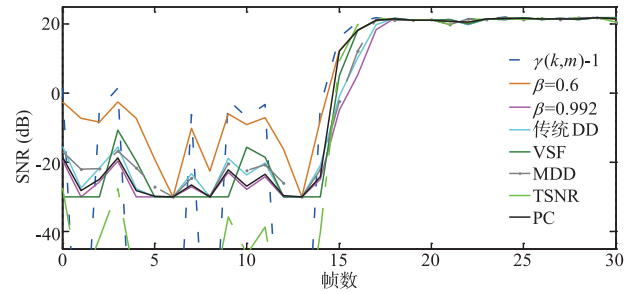


图4 几种算法先验信噪比估计曲线图

4 仿真实验与结果分析

为进一步表明本文提出算法的有效性,本节通过仿真实验来对其在语音增强系统中的应用效果进行验证,并将仿真结果与传统 DD 算法($\beta = 0.98$)、VSF 算法、MDD 算法以及 TSNR 算法的结果进行对比分析. 仿真中采用 30 段纯净语音信号作为测试语音数据,数据来源于 <https://ecs.utdallas.edu/loizou/>. 背景噪声选取 6 种不同类型的噪声信号,它们依次为 White、Pink、Babble、Destroyer、Volvo 以及 Factory 噪声,仿真中所用噪声信号同前文实验均取自于噪声库 <http://spib.ece.rice.edu/>,噪声估计则均采用 MS 算法. 纯净语音及噪声信号的采样频率均为 8kHz,将不同类型噪声叠加到纯净语音信号上,分别产生输入信噪比为 0、5、10 和 15dB 的带噪语音. 仿真中选取语音帧的长度为 256,帧间重叠 50%,各个算法的参数取值同图 4 仿真环境. 为了衡量五种算法输出语音信号的主观接受程度、失真情况和可懂度大小,采用语音感知质量(PESQ, Percep-

tual Evaluation of Speech Quality)、对数谱距离 (LSD, Log-Spectral Distortion) 以及短时客观可懂度 (STOI, Short-Time Objective Intelligibility) 三种测量标准来评估算法的整体性能^[24,25].

表 1 给出的是不同噪声背景及输入信噪比下, 五种算法输出语音的 PESQ 评估结果. PESQ 是 ITU-T 在 P. 862 标准中提出的语音质量评估方法^[24], 由于综合考虑了滤波器、信号时延、失真等因素的影响, 该评估模型结果与主观质量评估结果具有较好的一致性, 因此被广泛用于衡量增强语音信号的主观接受程度. 其数值越大, 说明算法输出语音的质量越高、越容易被受听者主观所接受. 通过比较表 1 中五种算法的 PESQ 结果可以发现, 相对于传统的 DD 算法, 其他四种改进方法的 PESQ 数值都有不同程度的提升, 说明 VSF、MDD、TSNR 以及 PC 算法均可以有效提高先验信噪比参数的估计精度、改善传统 DD 算法的整体性能. 而相对其他三种改进算法而言, PC 算法的输出 PESQ 数值最高 (表 1 中加黑数值表示同等条件下的最佳结果), 从而表明 PC 算法的整体性能更为优良, 经过该算法处理后的语音信号具有更好的输出质量、更加符合接受者的主观感受.

表 1 五种算法输出语音 PESQ 结果对比

噪声类型	输入 SNR	输出 PESQ				
		DD ($\beta=0.98$)	VSF	TSRN	MDD	PC
White	0dB	2.02	2.09	2.19	2.14	2.21
	5dB	2.36	2.46	2.53	2.49	2.58
	10dB	2.64	2.77	2.84	2.83	2.91
	15dB	3.01	3.03	3.10	3.11	3.18
Pink	0dB	2.12	2.23	2.25	2.20	2.31
	5dB	2.48	2.59	2.61	2.57	2.68
	10dB	2.82	2.89	2.92	2.92	3.01
	15dB	3.10	3.13	3.18	3.19	3.28
Babble	0dB	1.82	1.90	1.98	1.97	1.98
	5dB	2.27	2.34	2.40	2.37	2.42
	10dB	2.67	2.69	2.78	2.75	2.81
	15dB	3.00	3.02	3.12	3.10	3.16
Destroyer	0dB	2.11	2.26	2.29	2.25	2.33
	5dB	2.53	2.60	2.66	2.63	2.70
	10dB	2.86	2.90	2.97	2.96	3.01
	15dB	3.18	3.19	3.24	3.24	3.31
Volvo	0dB	3.54	3.56	3.74	3.73	3.79
	5dB	3.67	3.73	3.95	3.96	4.01
	10dB	4.10	4.11	4.11	4.12	4.15
	15dB	4.22	4.22	4.24	4.23	4.29
Factory	0dB	2.58	2.64	2.70	2.67	2.76
	5dB	2.91	2.94	3.00	3.00	3.07
	10dB	3.21	3.22	3.27	3.28	3.37
	15dB	3.49	3.54	3.55	3.56	3.66

表 2 给出的是各种仿真环境下五种先验信噪比估计算法输出语音的 LSD 结果. LSD 通过计算每一帧输出语音与纯净语音信号之间的接近程度, 进而获得两者在整体上的差距. LSD 的数值越小, 表示输出语音与原始的纯净语音越接近, 语音增强算法对于语音信号的损伤越小, 其计算公式如下:

$$\text{LSD} = \left(\frac{1}{M} \sum_{m=0}^{M-1} \frac{1}{K} \sum_{k=0}^{K-1} \left(10 \log_{10} \left(\frac{|\hat{X}_{m,k}|^2}{|X_{m,k}|^2} \right) \right)^2 \right)^{1/2} \quad (27)$$

其中, M 表示待测语音信号的总帧数, K 表示语音帧长度. 对比表 2 中的数据结果可以发现, 传统 DD 算法的 LSD 数值最大, 这是由于该算法的先验信噪比估计结果在跟踪语音能量变化时存在一帧的时间延迟问题, 从而在语音增强过程中将不可避免地会对语音信号产生较大的损伤, 导致输出语音中包含数量较多的语音失真. 在传统 DD 算法的基础上, 其他四种算法采用不同的优化策略都对其固有的时延问题进行了改善, 因此其他四种算法输出 LSD 的数值相对 DD 算法均有不同程度的减小. 相较于 VSF、MDD 和 TSNR 算法, 本文 PC 算法的输出 LSD 在绝大多数环境下都是最低的 (表 2 中加黑数值表示同等条件下的最佳结果), 说明 PC

表 2 五种算法输出语音 LSD 结果对比

噪声类型	输入 SNR	输出 LSD				
		DD ($\beta=0.98$)	VSF	TSRN	MDD	PC
White	0dB	6.55	6.17	6.50	5.22	5.11
	5dB	5.58	5.32	5.41	4.60	4.50
	10dB	4.87	4.68	4.71	4.06	3.98
	15dB	4.41	4.23	4.32	3.65	3.58
Pink	0dB	6.58	6.14	6.42	4.90	4.91
	5dB	5.42	5.13	5.10	4.22	4.21
	10dB	4.58	4.37	4.38	3.69	3.66
	15dB	4.00	3.83	3.84	3.25	3.22
Babble	0dB	6.79	6.27	6.47	4.96	4.95
	5dB	5.24	4.89	4.86	4.10	4.04
	10dB	4.16	3.94	3.83	3.45	3.33
	15dB	3.46	3.31	3.25	2.97	2.83
Destroyer	0dB	6.20	5.80	5.86	4.66	4.66
	5dB	5.01	4.76	4.70	3.94	3.91
	10dB	4.11	3.93	3.83	3.33	3.30
	15dB	3.46	3.33	3.30	2.86	2.81
Volvo	0dB	1.54	1.46	1.46	1.45	1.32
	5dB	1.25	1.19	1.16	1.16	1.06
	10dB	1.01	0.96	0.95	0.92	0.86
	15dB	0.85	0.80	0.79	0.76	0.72
Factory	0dB	4.53	4.31	4.23	3.51	3.50
	5dB	3.63	3.50	3.42	2.98	2.92
	10dB	2.94	2.86	2.74	2.51	2.42
	15dB	2.41	2.37	2.30	2.12	2.02

算法对于先验信噪比参数的估计性能要更优一些,依据其设计的语音增强系统在消除背景噪声干扰的同时,对语音频率成分的损伤程度要小于其它参考算法。

评价一个语音增强系统输出语音的优劣需从语音质量与可懂度两方面综合进行考虑。语音质量表示人耳听觉对噪声干扰的忍受程度,而语音可懂度则表示人耳对所听语音内容的理解程度,两者之间并没有明确的因果关系。前文给出的 PESQ 与 LSD 测度主要用来评估各种算法输出语音的质量,而在语音可懂度方面本文则考虑采用 STOI 评估方法^[25]。STOI 是将待测信号进行重采样后,通过比较纯净语音和增强语音信号短时包络的相关系数,并计算所有频点及语音帧中间变量的平均值来表示语音信号的可懂度。STOI 的取值范围为 $[0 \sim 1]$,其数值越大则表示被测语音信号的可懂度越高。表 3 给出的是不同噪声背景以及输入信噪比条件下五种算法输出语音的 STOI 评估结果。通过表 3 中数据可以看出,本文 PC 算法的 STOI 结果相对于其他参考算法均有不同程度的提高,说明 PC 算法输出语音的可懂度最高,其语音内容更容易被接受者所理解。

表 3 五种算法输出语音 STOI 结果对比

噪声类型	输入 SNR	输出 STOI (%)				
		DD ($\beta = 0.98$)	VSF	TSRN	MDD	PC
White	0dB	65.32	65.45	66.63	66.59	67.24
	5dB	75.17	75.60	76.94	77.01	77.83
	10dB	83.29	83.57	84.75	84.99	85.66
	15dB	88.75	88.97	89.95	90.03	90.81
Pink	0dB	65.06	65.18	66.17	66.56	67.40
	5dB	73.88	74.51	75.26	75.89	76.49
	10dB	82.51	83.04	84.26	84.91	85.56
	15dB	88.71	89.09	90.15	90.21	91.02
Babble	0dB	61.04	61.08	61.27	61.87	62.31
	5dB	72.24	72.69	73.74	74.24	74.71
	10dB	81.80	82.51	83.87	84.03	84.54
	15dB	88.82	89.24	90.57	90.65	90.94
Destroyer	0dB	64.74	64.77	65.21	66.01	67.03
	5dB	72.97	73.26	73.98	74.86	75.51
	10dB	79.73	80.17	80.71	81.21	81.75
	15dB	86.08	86.54	87.12	87.57	88.05
Volvo	0dB	92.04	92.46	92.92	92.88	93.35
	5dB	94.49	94.55	94.67	94.70	94.92
	10dB	95.83	95.88	96.07	96.13	96.31
	15dB	96.86	96.92	97.01	97.11	97.26
Factory	0dB	74.47	74.69	75.52	76.10	76.80
	5dB	81.54	81.97	82.75	82.93	83.86
	10dB	88.51	88.87	89.79	90.25	90.89
	15dB	92.87	93.06	93.72	93.88	94.21

5 结束语

DD 算法是目前语音增强系统中最为常用的先验

信噪比估计方法之一,但该算法在对先验信噪比估计的过程中存在一帧的时间延迟,且其估计性能受滑动因子选取数值的影响较大。针对 DD 算法存在的缺陷问题,本文采用实际的语音及噪声数据,对不同滑动因子 DD 算法的估计性能进行了深入研究,依据实验结果给出了滑动因子较为明确的边界值;然后基于信号的复高斯统计模型,利用软判决技术对具有不同滑动因子的 DD 算法进行有效耦合,以结合其各自的优势来突破单一固定滑动因子 DD 算法的性能局限。文中在不同噪声及信噪比环境下对多种算法的估计性能进行了实验验证,其结果表明,相对于当前流行的几种先验信噪比估计算法,基于本文算法设计的语音增强系统在语音质量和可懂度两方面均可对带噪语音具有不同程度的提升。

参考文献

- [1] Kuldeep P, Kamil W and Belinda S. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain[J]. *Speech Communication*, 2010, 52(5): 450 – 475.
- [2] Ephraim Y and Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator[J]. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 1984, 32(6): 1109 – 1121.
- [3] Jahangir A and Douglas O. Perceptual improvement of Wiener filtering employing a post-filter[J]. *Digital Signal Processing*, 2011, 21: 54 – 65.
- [4] Mohamed D and Pascal S. Reducing over and under-estimation of the a priori SNR in speech enhancement techniques[J]. *Digital Signal Processing*, 2014, 32: 124 – 136.
- [5] 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.
- [6] Plapous C, Marro C and Scalart P. Improved signal-to-noise ratio estimation for speech enhancement[J]. *IEEE Transactions on Audio, Speech and Language Processing*, 2006, 14(6): 2098 – 2108.
- [7] Yong P C, Nordholm S and Dam H. Optimization and evaluation of sigmoid function with a priori SNR estimate for real-time speech enhancement[J]. *Speech Communication*, 2013, 55(2): 358 – 376.
- [8] Lee Y K, Gue P J, Keun L Y, et al. Speech enhancement using phase-dependent a priori SNR estimator in Log-Mel spectral domain[J]. *ETRI Journal*, 2014, 36(5): 721 – 729.
- [9] Soon I Y and Koh S N. Low distortion speech enhancement[J]. *IEE Proceedings-Vision Image and Signal Processing*,

- 2000, 147(3):247–253.
- [10] Hasan M K, Salahuddin S and Khan M R. A modified a priori SNR for speech enhancement using spectral subtraction rules [J]. IEEE Signal Processing Letters, 2004, 11(4):450–453.
- [11] Park Y S and Chang J H. A novel approach to a robust a priori SNR estimator in speech enhancement [J]. IEICE transactions on Communication, 2007, E90-B(8):2182–2185.
- [12] Nahma L, Yong P C, Dam H, et al. Improved a priori SNR estimation in speech enhancement [A]. In Proceedings of 23rd Asia-Pacific Conference on Communications [C]. Ho Chi Minh, Vietnam; 2017. 1–5.
- [13] Suhadi S, Last C and Fingscheidt T. A data-driven approach to a priori SNR estimation [J]. IEEE Transactions on Audio, Speech and Language Processing, 2010, 19(1):186–195.
- [14] Choi J H and Chang J H. On using acoustic environment classification for statistical model-based speech enhancement [J]. Speech Communication, 2012, 54(3):477–490.
- [15] Lee S, Lim C and Chang J H. A new a priori SNR estimator based on multiple linear regression technique for speech enhancement [J]. Digital Signal Processing, 2014, 30:154–164.
- [16] Elshamy S, Madhu N, Tirry W, et al. Instantaneous a priori SNR estimation by cepstral excitation manipulation [J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2017, 25(8):1592–1605.
- [17] Xu Z, Elshamy S and Tim F. A priori SNR estimation using discriminative non-negative matrix factorization [A]. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Calgary, AB, Canada; 2018. 661–665.
- [18] Ou S, Song P and Gao Y. Phase-sensitive decision-directed SNR estimator for single-channel speech enhancement [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2017, 31(8):1758003–1–16.
- [19] Shin H S, Fingscheidt T and Kang H G. A priori SNR estimation using air and bone-conduction microphones [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(11):2015–2025.
- [20] Délier J, Proakis J and Hansen J. Discrete-time processing of speech signals [M]. Wiley-Interscience, 1993.
- [21] Park T J and Chang J H. Dempster-shafer theory for enhanced statistical model-based voice activity detection [J]. Computer Speech and Language, 2018, 47:47–58.
- [22] Chang J, Jo Q, Kim D K, et al. Global soft decision employing support vector machine for speech enhancement [J]. IEEE Signal Processing Letters, 2008, 16(1):57–60.
- [23] Hu Y and Loizou P C. Evaluation of objective quality measures for speech enhancement [J]. IEEE Transactions on Audio, Speech and Language Processing, 2008, 16(1):229–238.
- [24] Taal C H, Hendriks R C, Heusdens R and Jensen J. An algorithm for intelligibility prediction of time-frequency weighted noisy speech [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(7):2125–2136.

作者简介



欧世峰 (通信作者) 男, 1979 年生于山东巨野, 2008 年毕业于吉林大学, 获工学博士学位. 现为烟台大学光电信息科学技术学院副院长, 教授, 硕士生导师. 主要研究方向为语音信号处理.

E-mail: ousufeng@126.com



赵艳磊 女, 1995 年生于山东滨州, 2018 年毕业于烟台大学, 获工学学士学位. 现为烟台大学光电信息科学技术学院硕士研究生. 主要研究方向为语音信号处理.



宋 鹏 男, 1983 年生于山东莱阳, 2014 年毕业于东南大学, 获工学博士学位. 现为烟台大学计算机与控制工程学院副教授, 硕士生导师. 主要研究方向为语音信号处理.



高 颖 女, 1978 年生于辽宁大连. 2008 年毕业于吉林大学, 获工学博士学位. 现为烟台大学光电信息科学技术学院副教授, 硕士生导师. 主要研究方向为现代信号处理理论及应用.