

文章编号: 1672-2892(2010)01-0076-04

## 基于直接判决估计和预测估计的语音增强算法

冯炎<sup>1,2</sup>

(1.西南交通大学 信息科学与技术学院, 四川 成都 610031; 2.西藏大学 现代教育技术中心, 西藏 拉萨 850000)

**摘要:**对噪声环境下语音增强算法中的先验信噪比估计提出了一种新的估计算法。该算法结合了直接判决估计和预测估计方法,与传统的基于直接判决的先验信噪比估计算法相比较,当后验信噪比较大时,该算法减少了语音失真;当后验信噪比较小时,算法较明显地减少了音乐噪声。通过研究发现传统的直接判决方法就是该算法的一个特例,该算法可以通过一个权衡因子退化为传统的直接判决方法。实验结果也证实所提出的算法的优越性能。

**关键词:**直接判决估计;预测估计;先验信噪比估计;语音增强

中图分类号: TN912.3

文献标识码: A

## Speech enhancement based on Decision-Directed incorporating predicted estimation

FENG Yan<sup>1,2</sup>

(1.School of Information Science & Technology, Southwest Jiaotong University, Chengdu Sichuan 610031, China;  
2.Modern Education and Technology Center, Tibet University, Lhasa Tibet 850000, China)

**Abstract:** A novel approach combining Decision Directed(DD) estimation and Predicted(PD) estimation was proposed to improve the priori Signal to Noise Ratio(SNR) estimation for speech enhancement in noisy environments. PD estimation incorporated the soft-decision scheme and speech presence uncertainty. In contrast to the conventional DD-based scheme, a less speech distortion was achieved when a posteriori SNR was higher; and a further reduction of musical noise was achieved when a posteriori SNR was lower. It was also found that a special case of the proposed method had been degenerated to the DD estimation with a weighting factor. Experimental results demonstrated the improved performance of the proposed algorithm.

**Key words:** decision-directed estimation; predicted estimation; priori SNR estimation; speech enhancement

语音信号在通信过程中不可避免地受到各种噪声的干扰,这些噪声的存在不仅会使人们产生听觉疲劳,而且破坏语音信号原有的模型参数和声学特性。语音增强是消除这些噪声干扰的重要手段。大多数语音增强算法中的一个重要参数就是先验信噪比<sup>[1-3]</sup>。在实际中常用Ephraim和Malah提出的直接判决<sup>[2]</sup>来估计这个参数,该方法简单可行并且计算有效。然而,直接判决有个缺点,就是它估计出的先验信噪比在语音活动区域仅是后验信噪比延时版本的简单再现。由于这样的延时,该算法在语音起始区域和结束区域对先验信噪比无法准确估计,并且在后验信噪比较低时存在严重的音乐残留噪声和语音失真。在研究中发现,如果通过调节平滑因子来减少音乐噪声,会增加语音失真,同时使延时更严重。

针对以上问题,提出了一种新的先验信噪比估计算法。该算法结合直接判决和预测估计<sup>[4]</sup>,预测估计中采用了软判决方案,并考虑了语音出现的不确定性,从而给出了更为准确的先验信噪比估计。与传统的基于直接判决的先验信噪比估计算法相比较,所提出的算法可以通过不增加延时来减少音乐噪声与减少语音失真。为方便表述,以下用DD表示直接判决算法,PD表示预测估计算法,DDPD表示本文提出的算法。

### 1 问题描述

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  $Y(t)$ ,进行短时离散傅里

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。考虑到语音在带噪语音中出现的 uncertainty, 用  $H_0$  和  $H_1$  分别表示无语音状态和有语音状态,

$$\begin{aligned} H_0 : Y(n, k) &= D(n, k) \\ H_1 : Y(n, k) &= X(n, k) + D(n, k) \end{aligned} \quad (2)$$

根据贝叶斯准则, 给出后验无音概率<sup>[1]</sup> :

$$p(H_0 | Y(n, k)) = \frac{p(Y(n, k) | H_0) p(H_0)}{p(Y(n, k) | H_0) p(H_0) + p(Y(n, k) | H_1) p(H_1)} = \frac{1}{1 + (p(H_1) / p(H_0)) \Lambda(Y(n, k))} \quad (3)$$

式中似然比  $\Lambda(Y(n, k))$ <sup>[5-6]</sup> :

$$\Lambda(Y(n, k)) = \frac{p(Y(n, k) | H_1)}{p(Y(n, k) | H_0)} = \frac{1}{1 + \xi(n, k)} \exp \left[ \frac{\gamma(n, k) \xi(n, k)}{1 + \xi(n, k)} \right] \quad (4)$$

式中:  $\gamma(n, k)$  表示后验信噪比<sup>[2]</sup>;  $\xi(n, k)$  表示先验信噪比<sup>[2]</sup>, 且

$$\gamma(n, k) = \frac{|Y(n, k)|^2}{\lambda_D(n, k)}, \quad \xi(n, k) = \frac{\lambda_X(n, k)}{\lambda_D(n, k)} \quad (5)$$

式中  $\lambda_X(n, k)$  和  $\lambda_D(n, k)$  分别表示纯净语音和噪声的功率谱, 本文采用 R Martin 的最小统计<sup>[7]</sup>来估计  $\lambda_D(n, k)$ 。

多数语音增强算法均依赖于先验信噪比  $\xi(n, k)$ , 假设  $\lambda_D(n, k)$  已知, 但  $\lambda_X(n, k)$  未知, 故  $\xi(n, k)$  也是未知的, 并需要从带噪语音中来估计。

## 2 DDPD 算法描述

### 2.1 PD 算法

文献[4]中, 利用预测估计来进行语音活动检测, 本文根据预测估计的方法设计出针对语音增强算法中的先验语音信噪比估计:

$$\hat{\xi}^{\text{PD}}(n+1, k) = \frac{\zeta |\hat{X}(n, k)|^2}{\lambda_D(n, k)} + \frac{(1-\zeta) E \left[ |X(n, k)|^2 | Y(n, k) \right]}{\lambda_D(n, k)} \quad (6)$$

式中:  $\zeta$  是权衡因子;  $\hat{X}(n, k)$  是当前帧语音信号估计值。因为是用当前帧的值来预测下一帧的值, 因而称作预测估计。在 PD 估计方法中, 用软判决<sup>[4]</sup>求上式中的期望值:

$$E \left[ |X(n, k)|^2 | Y(n, k) \right] = E \left[ |X(n, k)|^2 | Y(n, k), H_0 \right] p(H_0 | Y(n, k)) + E \left[ |X(n, k)|^2 | Y(n, k), H_1 \right] p(H_1 | Y(n, k)) \quad (7)$$

式中  $p(H_1 | Y(n, k)) = 1 - p(H_0 | Y(n, k))$ , 且  $E \left[ |X(n, k)|^2 | Y(n, k), H_0 \right] = 0$ ,

$$E \left[ |X(n, k)|^2 | Y(n, k), H_1 \right] = \left( \frac{1}{1 + \tilde{\xi}(n, k)} \right) |\hat{X}(n, k)|^2 + \left( \frac{\tilde{\xi}(n, k)}{1 + \tilde{\xi}(n, k)} \right)^2 |Y(n, k)|^2 \quad (8)$$

式中  $\tilde{\xi}(n) = |\hat{X}(n, k)|^2 / \lambda_D(n, k)$ 。

### 2.2 DD 算法

前一帧估计出的纯净语音用  $\hat{X}(n-1, k)$  表示, 用  $\alpha$  表示平滑因子, DD 算法<sup>[2]</sup>描述如下:

$$\hat{\xi}^{\text{DD}}(n, k) = \alpha \frac{|\hat{X}(n-1, k)|^2}{\lambda_D(n-1, k)} + (1-\alpha) \max\{\gamma(n, k) - 1, 0\} \quad (9)$$

式中:  $\alpha$  ( $0 < \alpha < 1$ ) 表示平滑参数;  $\max(\cdot)$  是求最大值的函数。

### 2.3 DD 和 PD 相结合的估计算法

为了克服 DD 方法在语音活动区域残留音乐噪声大的缺点, 我们引入基于软判决方案的 PD 方法, 并将 PD 算法融入 DD 算法。用  $\hat{\xi}^{\text{PD}}(n, k)$  表示 PD 算法估计出的先验信噪比, 新提出的 DDPD 算法表示为:

$$\hat{\xi}^{\text{DDPD}}(n, k) = \alpha \hat{\xi}^{\text{PD}}(n, k) + (1 - \alpha) \max\{\gamma(n, k) - 1, 0\} \quad (10)$$

### 2.4 DDPD 算法与 DD 算法的关系

通过观察式(6)、式(9)、式(10)发现, 当  $\zeta = 1$  时代入式(6)可得:

$$\hat{\xi}^{\text{PD}}(n+1, k) = \frac{|\hat{X}(n, k)|^2}{\lambda_D(n, k)} \quad (11)$$

根据式(9)~式(11)可知, 此时的 DDPD 算法退化为 DD 算法。DDPD 算法其实是 DD 算法的扩展, 它能够提供更比 DD 算法更灵活的残余音乐噪声控制能力。

## 3 实验及结果分析

将新提出的 DDPD 算法和传统的 DD 算法分别应用到最小均方误差语音增强<sup>[2]</sup>系统中, 并计算在不同噪声环境和不同信噪比下的分段信噪比增益和 PESQ(Perceptual Evaluation of Speech Quality)得分。PESQ 是 ITU-T P.862 建议的语音质量感性评价标准。试验中采用 3 种类型的噪声分别是: 白噪声(white)、工厂噪声(factory1)和人群嘈杂声(babble), 这些噪声取自 NOISEX92 数据库, 噪声信号与语音信号的采样频率均为 8 kHz, 使用不同的信噪比分别是: 5 dB, 10 dB, 15 dB。纯净语音取自 TIMIT 数据库。仿真中选取语音帧为 256 个采样点, 帧间叠接 50%。噪声功率  $\lambda_D(n, k)$  采用 R. Martin 的最小统计<sup>[7]</sup>来估计。 $\alpha$ 、 $\zeta$  和  $p(H_0)$  的取值需要根据具体实验来确定, 实验中,  $\alpha$  取值为 0.98,  $\zeta$  取值为 0.3,  $p(H_0)$  取值为 0.062 5。

表 1、表 2 给出了实验结果。从表 1 可看出, DDPD 算法的分段信噪比增益在较低的输入信噪比时明显优于 DD 算法, 主要是因为 DDPD 算法的残留音乐噪声较小的缘故。而在较高的输入信噪比时, 因 DDPD 算法抑制噪声的能力减弱, 使得 DDPD 算法的优势减弱, 这可以通过增加权衡因子  $\zeta$  的取值来改善。从表 2 可看出, DDPD 算法的 PESQ 得分优于 DD 算法, 在输入信噪比较高时, 因有较少的语音失真, DDPD 算法的优势更明显。而输入信噪比较低时有语音失真, 使得 DDPD 算法的优势减弱。

在较低的输入信噪比时, DDPD 算法具有较高抑制音乐噪声的能力, 残留音乐噪声较小, 但同时也抑制了弱小语音, 带来了语音失真。在较高的输入信噪比时, DDPD 算法抑制噪声的能力较差, 但同时也减少了语音失真。减少音乐噪声与减少语音失真是一对矛盾, 在实际应用中, 我们可以通过合理的调节权衡因子  $\zeta$  的取值来获得满意的语音增强效果。

## 4 结论

本文分析了传统的 DD 算法及 PD 算法, 且在 PD 算法中采用软判决方案, 并成功地将 PD 算法融入 DD 算法, 提出了 DDPD 算法。该算法较明显地减少了音乐噪声。通过研究发现该算法可以通过一个权衡因子  $\zeta$  退化为 DD 算法。DDPD 能够提供比 DD 算法更灵活的残余音乐噪声控制。实验结果也证实本文所提出的算法的优越性能。本文的 DDPD 算法中, 权衡因子  $\zeta$  取值是固定的, 如果该因子的取值能随先验信噪比自适应地变化, 将会提升 DDPD 算法的性能。

表 1 DD 算法与 DDPD 算法的分段信噪比增益对比  
Table1 Segmental SNR improvement obtained by using the DD approach and DDPD approach

noise type	input $R_{\text{SN}}/\text{dB}$	DD	DDPD
white	5	3.336 9	4.419 7
	10	3.163 0	4.195 5
	15	2.745 6	3.681 0
babble	5	1.072 7	1.423 1
	10	0.886 7	1.243 2
	15	0.651 3	0.931 5
factory1	5	2.106 0	2.979 1
	10	1.949 6	2.733 3
	15	1.718 1	2.410 9

表 2 DD 算法与 DDPD 算法的 PESQ 得分对比  
Table2 PESQ scores obtained by using the DD approach

noise type	input $R_{\text{SN}}/\text{dB}$	PESQ score	
		DD	DDPD
white	5	2.041 0	2.249 0
	10	2.375 0	2.554 0
	15	2.642 0	2.842 0
babble	5	2.178 0	2.198 0
	10	2.492 0	2.533 0
	15	2.769 0	2.841 0
factory1	5	2.004 0	2.103 0
	10	2.352 0	2.468 0
	15	2.671 0	2.766 1