Digital Signal Processing 30 (2014) 154–164

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

A new *a priori* SNR estimator based on multiple linear regression technique for speech enhancement

Soojeong Lee, Chungsoo Lim, Joon-Hyuk Chang

Department of Electronic Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong, Seoul 133-791, Republic of Korea

ARTICLE INFO

Article history:
Available online 18 April 2014

Keywords:
Speech enhancement
A priori SNR estimation
Multiple linear regression
Gaussian mixture model

ABSTRACT

We propose a new approach to estimate the *a priori* signal-to-noise ratio (SNR) based on a multiple linear regression (MLR) technique. In contrast to estimation of the *a priori* SNR employing the decision-directed (DD) method, which uses the estimated speech spectrum in previous frame, we propose to find the *a priori* SNR based on the MLR technique by incorporating regression parameters such as the ratio between the local energy of the noisy speech and its derived minimum along with the *a posteriori* SNR. In the experimental step, regression coefficients obtained using the MLR are assigned according to various noise types, for which we employ a real-time noise classification scheme based on a Gaussian mixture model (GMM). Evaluations using both objective speech quality measures and subjective listening tests under various ambient noise environments show that the performance of the proposed algorithm is better than that of the conventional methods.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Speech enhancement is a key factor in various speech communication systems such as robust speech recognition, mobile communication, and speech coding due to background acoustic noise [1–11]. The *a priori* signal-to-noise ratio (SNR) is one of the most crucial parameters in speech enhancement areas [12–16]. Actually, the *a priori* SNR should be carefully estimated for the reductions of musical noise and speech distortion within the minimum mean squared error (MMSE)-based spectral gain estimation [17,18]. However, accurate estimation of the *a priori* SNR is actually difficult in non-stationary noise environments [19]. Thus, over the past few years, several studies have been performed to estimate the *a priori* SNR [12–19]. For example, Ephraim and Malah found that the performance of the speech enhancement could be significantly degraded due to the inaccurate *a priori* SNR estimation [1] so that they firstly used a maximum likelihood (ML) method to estimate the *a priori* SNR. They also introduced the decision-directed (DD) approach, which is used to estimate the *a priori* SNR, based on the definition of the *a priori* SNR and its relationship with the *a posteriori* SNR [1]. Alternatively, Cohen proposed a noncausal (NC) *a priori* SNR estimator that uses future spectral signals to estimate the spectral variance of the clean speech signal [13]. However, this approach has limited applications because noncausal estimation always has an additional delay [20]. Recently, Park and Chang proposed a novel *a priori* SNR estimator by employing the sigmoid type function [6]. Unfortunately, this scheme does not consider a diversity of noise environments on noise variation [21]. More recently, Suhadi et al. proposed a data-driven approach that employs two trained neural networks to estimate the *a priori* SNR [20]. However, this approach requires a substantial training process for estimating the *a priori* SNR, which cannot be robust estimator under varying noise environments.

Among the previous works, the DD approach of Ephraim and Malah is often preferred for estimating the *a priori* SNR because it is the most computationally efficient method with acceptable performance. However, the DD approach also has a drawback such as slow responses to speech onsets [14,20]. Indeed, the DD estimation for the true *a priori* SNR is composed of the estimation of the *a priori* SNR obtained from the previous frames and that of the current *a posteriori* SNR. The weight of the *a priori* SNR estimator obtained in the previous frame is substantially larger than that of the *a posteriori* SNR estimator based on the current frame [14]. Therefore, this characteristic of the DD approach cause a delay in *a priori* SNR especially in speech onsets, possibly degrading the quality of the enhanced speech signal. In particular, Breithaupt and Martin [22] analyzed that the DD-based SNR estimator in low SNR conditions has limits in reducing noise in terms of preservation of speech onsets and the suppression of musical noise. In addition, it is discovered that subband-weighting rule can be separately trained from the various noise environments and stored in a look-up table [23]. For example, Erkelens et al. [24] proposed a data-driven weighting method, which uses a training step under white noise with known

E-mail address: jchang@hanyang.ac.kr (J.-H. Chang).

http://dx.doi.org/10.1016/j.dsp.2014.04.001
1051-2004/© 2014 Elsevier Inc. All rights reserved.

spectral variance to address the bias problem especially in a low SNR. However, it has a tendency to offer inaccurate noise estimate especially in various noise conditions and is computationally inefficient.

For this reason, an efficient methodology to estimate the *a priori* SNR is needed while allowing its application to a variety of noise environments without a considerable training process. Thus, we propose a novel approach to estimate the *a priori* SNR using multiple linear regression (MLR) [25–27]. In our proposed approach, we apply the MLR to overcome the aforementioned problem in the DD algorithm because the MLR does not use the estimation of the *a priori* SNR obtained from processing the previous frame as the DD. Specifically, the MLR can estimate the best-fitting surface of a suitable function that relates the independent and dependent variables [27]. We use the ratio between the local energy of the noisy speech and its derived minimum S_r [8], which is known to have similar characteristics to the estimation of the *a priori* SNR [28] as an independent variable along with the estimated *a posteriori* SNR and we use the true SNR as a dependent variable. In our training process, the regression coefficients are estimated, which represents the best-fitting surface between the independent and dependent variables, so that the estimated *a priori* SNR fits better the true SNR than the conventional estimators. In the testing process, assignment of the regression coefficients is performed according to various noise types determined by a real-time noise classification scheme using the Gaussian mixture model (GMM) [21,29]. The performance of the proposed algorithm is evaluated using extensive objective and subjective speech signal quality measures in various noise environments. The experimental results reveal that the proposed method shows better performance than the conventional methods.

2. Review of the *a priori* SNR estimation

As mentioned in the previous section, in the DD algorithm, the true *a priori* SNR is estimated based on the estimation of the *a priori* SNR obtained from processing the previous frame and that of the *a posteriori* SNR based on the current frame. For the derivation of the DD, let $x(n)$ and $d(n)$ denote clean speech and uncorrelated additive noise signals, respectively. The observed noisy speech signal $y(n)$ is the sum of the clean speech signal $x(n)$ and the noise signal $d(n)$, where n is a discrete-time index. Applying a short-time Fourier transform (STFT), we then have

$$Y_k(t) = X_k(t) + D_k(t) \tag{1}$$

where $k(= 1, 2, \cdots, K)$ is the frequency bin, and t is the frame index. The *a posteriori* SNR $\gamma_k(t)$ and the *a priori* SNR $\xi_k(t)$ are defined by

$$\gamma_k(t) \triangleq \frac{|Y_k(t)|^2}{\lambda_{d,k}(t)}, \tag{2}$$

$$\xi_k(t) \triangleq \frac{\lambda_{x,k}(t)}{\lambda_{d,k}(t)}, \tag{3}$$

where $\lambda_{x,k}(t)$ and $\lambda_{d,k}(t)$ are the variances of the clean speech and the noise, respectively. In time, $\hat{\xi}_k(t)$ could be estimated using the DD approach [1] as follows:

$$\hat{\xi}_k(t) = \alpha_\xi \frac{|\hat{X}_k(t-1)|^2}{\hat{\lambda}_{d,k}(t-1)} + (1-\alpha_\xi) F\big[\hat{\gamma}_k(t) - 1\big] \tag{4}$$

where $\hat{X}_k(t-1)$ denotes the estimated clean speech spectrum of the previous frame, and $F[x] = x$ if $x \geq 0$; otherwise $F[x] = 0$. Here, the weighting factor, α_ξ $(0 \leq \alpha_\xi \leq 1)$, which controls the trade-off between the noise suppression and the transient signal distortion, is generally set very close to 1 (i.e., $\alpha_\xi = 0.99$). The DD algorithm can effectively eliminate the residual musical noise [14,20]. However, this algorithm also has major problems due to employing the previous frame and using the constant weighting factor very close to 1. Specifically, these characteristic of the DD leads to slow responses in speech onsets [14,20].

Since a speech enhancement algorithm is eventually based on the newly derived $\hat{\xi}_k(t)$ and $\hat{\gamma}_k(t)$, the estimated clean speech spectrum $\hat{X}_k(t)$ is obtained using the MMSE-based method [1] which is the optimal spectral magnitude estimator because it is obtained by minimizing the mean-squared error between the estimated spectral amplitude and true amplitude.

$$\hat{X}_k(t) = G\big(\hat{\xi}_k(t), \hat{\gamma}_k(t)\big) \cdot Y_k(t), \tag{5}$$

where $G(\cdot,\cdot)$ is the noise suppression rule according to the MMSE as given by

$$G\big(\hat{\xi}_k(t), \hat{\gamma}_k(t)\big) = \frac{\sqrt{\pi}}{2} \sqrt{\frac{\hat{\xi}_k(t)}{\hat{\gamma}_k(t)(1+\hat{\xi}_k(t))}} \times C\left[\frac{\hat{\gamma}_k(t)\hat{\xi}_k(t)}{1+\hat{\xi}_k(t)}\right], \tag{6}$$

with

$$C\big[\nu_k(t)\big] = \exp\left(\frac{-\nu_k(t)}{2}\right)\left[1 + \nu_k(t) \cdot I_0\left(\frac{\nu_k(t)}{2}\right) + \nu_k(t) \cdot I_1\left(\frac{\nu_k(t)}{2}\right)\right], \tag{7}$$

in which I_0 and I_1 denote the modified Bessel functions of zero*th* and first order, respectively. Also, $\nu_k(t)$ is given by

$$\nu_k(t) = \frac{\hat{\xi}_k(t)}{1+\hat{\xi}_k(t)} \hat{\gamma}_k(t) \tag{8}$$

Note that it is discovered that the gain $G(\cdot,\cdot)$ is dominantly affected by $\hat{\xi}_k(t)$ especially in low SNR conditions so that we should estimate $\hat{\xi}_k(t)$ accurately [12].

3. The proposed *a priori* SNR estimation based on MLR

3.1. Parameter estimation

In the proposed method, derivation of a new *a priori* SNR begins from the following equation, which is also used in the DD approach [1].

$$\xi_k(t) = \frac{E\{|X_k(t)|^2\}}{\lambda_{d,k}(t)}. \tag{9}$$

In time, $\xi_k(t)$ can be rewritten based on the ML estimate such that

$$\hat{\xi}_k(t) = \frac{|Y_k(t)|^2}{\lambda_{d,k}(t)} - 1 \tag{10}$$

where the *a priori* SNR can be thus expressed using the *a posteriori* SNR $\gamma_k(t)$ such that $\hat{\xi}_k(t) = \gamma_k(t) - 1$ and used in the DD approach to find estimator $\hat{\xi}_k(t)$.

On the other hand, we can also consider the smoothed local energy of the noisy speech in order to acquire the ratio between the local energy of the noisy speech and its derived minimum $S_r(= S_k/S_{\min,k})$ based on the following equation [8]:

$$S_k(t) = \alpha_s S_k(t-1) + (1-\alpha_s)\big|Y_k(t)\big|^2, \tag{11}$$

where $\alpha_s(= 0.98)$ denotes the smoothing parameter. Based on Eq. (10), the numerator in Eq. (10) is obtained from the noisy power spectrum and the denominator is acquired by the noise power spectrum. Note that $b_k(t)S_{\min,k}(t)$ is called the minimum

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات